1

Linear transformations

1.1 Vectors

A sequence of n numbers that are complex in general, denoted $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$, is called a vector in n-dimensional linear vector space $V^{(n)}$ with the components v_i; $i = 1, 2, \ldots, n$. The coordinates of a point with respect to a coordinate system in n-dimensional space can also be considered as a vector in $V^{(n)}$. Vectors obey the following rules.

(1) Two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ are said to be equal if their corresponding components are equal, i.e. $\boldsymbol{v} = \boldsymbol{u}$, if $v_i = u_i$ for all i.
(2) Addition of two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ is also a vector with components $(\boldsymbol{v} + \boldsymbol{u})_i = v_i + u_i$.
(3) The product $c\boldsymbol{v}$ of a number c with a vector $\boldsymbol{v}$ is a vector whose components are c times the components of $\boldsymbol{v}$, i.e. $(c\boldsymbol{v})_i = cv_i$.

A vector is a null vector if all its components vanish.

Vectors of a set $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(k)}$ are said to be *linearly independent* if there exists no relationship of the form

$$a_1\boldsymbol{v}^{(1)} + a_2\boldsymbol{v}^{(2)} + \cdots + a_k\boldsymbol{v}^{(k)} = 0$$

excluding the trivial case in which all coefficients $a_1, a_2, \ldots, a_k$ are zero. If these vectors are linearly dependent, there exists a non-zero coefficient, say $a_1 \neq 0$, then $\boldsymbol{v}^{(1)}$ is expressed in terms of the remaining vectors as follows:

$$\boldsymbol{v}^{(1)} = -\frac{a_2}{a_1}\boldsymbol{v}^{(2)} - \frac{a_3}{a_1}\boldsymbol{v}^{(3)} - \cdots - \frac{a_k}{a_1}\boldsymbol{v}^{(k)}$$

There exist no more than n vectors in $V^{(n)}$ which are linearly independent; in fact, if we introduce n linearly independent vectors of the form

$$\begin{aligned} \boldsymbol{e}^{(1)} &= (1, 0, 0, \ldots, 0) \\ \boldsymbol{e}^{(2)} &= (0, 1, 0, \ldots, 0) \\ &\cdots \quad \cdots \quad \cdots \\ \boldsymbol{e}^{(n)} &= (0, 0, 0, \ldots, 1) \end{aligned} \tag{1.1.1}$$

then an arbitrary vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ is expressed in terms of the set of n vectors as follows:

$$\boldsymbol{v} = v_1\boldsymbol{e}^{(1)} + v_2\boldsymbol{e}^{(2)} + \cdots + v_n\boldsymbol{e}^{(n)}$$

The set $e = \{\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(2)}, \ldots, \boldsymbol{e}^{(n)}\}$ is called *the natural basis* of the n-dimensional vector space $V^{(n)}$.

Another important elementary concept is the scalar product of two vectors. *The scalar product* of two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ is a number defined by

$$(\boldsymbol{v}, \boldsymbol{u}) = v_1 u_1 + v_2 u_2 + \cdots + v_n u_n \tag{1.1.2}$$

There exists another scalar product called *the Hermitian scalar product* defined by

$$\langle \boldsymbol{v}, \boldsymbol{u} \rangle = v_1^* u_1 + v_2^* u_2 + \cdots + v_n^* u_n = (\boldsymbol{v}^*, \boldsymbol{u}) \tag{1.1.3}$$

where v_i^* denotes the complex conjugate of v_i.

Let c be a number. Then the Hermitian scalar product is linear in the second factor but is 'antilinear' in the first factor:

$$\langle \boldsymbol{v}, c\boldsymbol{u} \rangle = c\langle \boldsymbol{v}, \boldsymbol{u} \rangle, \qquad \langle c\boldsymbol{v}, \boldsymbol{u} \rangle = c^* \langle \boldsymbol{v}, \boldsymbol{u} \rangle \tag{1.1.4}$$

whereas the simple scalar product is linear in both factors.

When $(\boldsymbol{v}, \boldsymbol{u}) = 0$ or $\langle \boldsymbol{v}, \boldsymbol{u} \rangle = 0$ we say that two vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ are orthogonal under the ordinary scalar product or the Hermitian scalar product, respectively. A vector that satisfies $(\boldsymbol{v}, \boldsymbol{v}) = 0$ but $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \neq 0$ is said to be *self-orthogonal* or *isotropic*.

Isotropic vectors were introduced by Cartan (1913) to formulate the 'spinor algebra.' Obviously, if $\langle \boldsymbol{v}, \boldsymbol{v} \rangle = 0$, then $\boldsymbol{v}$ is a null vector.

1.2 Linear transformations and matrices

Consider a set of n linear equations

$$\begin{aligned} y_1 &= A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ y_2 &= A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ &\cdots \quad \cdots \quad \cdots \\ y_n &= A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n \end{aligned}$$

which may be written in a more compact form

$$y_j = \sum_{i=1}^{n} A_{ji}x_i; \qquad j = 1, 2, \ldots, n$$

The set of equations is said to form a linear transformation in n variables. It may be regarded as a point transformation that brings a point x to another point y in $V^{(n)}$. The point transformation is a mapping of the n-dimensional space $V^{(n)}$ into itself and is completely defined by the set of n^2 quantities $\{A_{ji}\}$, which can be any complex numbers. We therefore associate with the transformation the array of numbers called an *$n \times n$ matrix*,[1] denoted $A = \|A_{ji}\|$. Then the set of equations is written as follows:

$$y = Ax \tag{1.2.1}$$

Two matrices $A = \|A_{ji}\|$ and $B = \|B_{ji}\|$ are said to be equal if and only if $A_{ji} = B_{ji}$ for all j and i.

We define addition of two $n \times n$ matrices by the rule

$$(A + B)_{ji} = A_{ji} + B_{ji}$$

[1] An $n \times n$ matrix is called a square matrix. Hereafter we are primarily concerned with square matrices unless otherwise specified.

in accordance with the vector addition $z = Ax + Bx = (A + B)x$. Then addition is commutative and associative, for $A + B = B + A$ and $A + (B + C) = (A + B) + C$.

Now let $z_k = \sum_j B_{kj} y_j$ be a second transformation by an $n \times n$ matrix $B = \|B_{kj}\|$. The effect of two transformations consecutively by A and then by B produces a third transformation

$$z_k = \sum_j B_{kj} A_j = \sum_{ji} B_{kj} A_{ji} x_i; \qquad \text{i.e. } z = BAx$$

Therefore we define the product P of two matrices by the rule

$$P_{ki} = \sum_j B_{kj} A_{ji} \qquad \text{or} \qquad P = BA$$

The rule is analogous to the rule for the product of two determinants. If we denote the determinant of a matrix A by $\det A$, we have

$$\text{if } P = BA, \qquad \text{then} \qquad \det P = \det B \det A \tag{1.2.2}$$

Note that $\det B \det A = \det A \det B$ since determinants are numbers. However, the matrix product BA need not equal AB in general; e.g. see (1.2.5b). Thus, matrix multiplication need not be commutative. Multiplication is, however, associative: $A(BC) = (AB)C$, so that the product is simply written ABC.

Another important characteristic property of a matrix A is the *trace* of A defined by the sum of the diagonal elements

$$\operatorname{tr} A = \sum_i A_{ii}$$

Then, the trace of a product ABC equals the trace of the product BCA, because

$$\operatorname{tr} ABC = \sum_{ijk} A_{ij} B_{jk} C_{ki} = \sum_{ijk} B_{jk} C_{ki} A_{ij} = \operatorname{tr} BCA$$

That is, the trace of a product of matrices is invariant under a cyclic permutation of the matrices.

The matrix with unity in each position in the leading diagonal and zero elsewhere is called the unit matrix, denoted $\mathbf{1} = \|\delta_{ij}\|$, where $\delta_{ij} = 0$ $(i \neq j)$ and $\delta_{ii} = 1$. The unit matrix corresponds to the identity transformation. A matrix of the form $D = \|d_i \delta_{ij}\|$ that has diagonal elements $d_1, d_2, \ldots, d_n$ and zero elsewhere is called a diagonal matrix. It is denoted by

$$D = \operatorname{diag}[d_1, d_2, \ldots, d_n] \tag{1.2.3a}$$

then

$$\det D = d_1 d_2 \ldots d_n, \qquad \operatorname{tr} D = d_1 + d_2 + \cdots + d_n$$

Two diagonal matrices always commute and their product gives a diagonal matrix; in fact, let $D' = \operatorname{diag}[d'_1, d'_2, \ldots, d'_n]$ be another diagonal matrix, then

$$DD' = \operatorname{diag}[d'_1 d_1, d'_2 d_2, \ldots, d'_n d_n] = D'D \tag{1.2.3b}$$

If the determinant of a matrix is not zero, then the matrix is said to be non-singular. If the coefficient matrix A of (1.2.1) is non-singular, then the equation (1.2.1) may be solved for x; i.e. there exists the inverse transformation from y to x, which may be written as

$$x = A^{-1}y$$

where A^{-1} is called the inverse of A and satisfies

$$A^{-1}A = AA^{-1} = 1 \tag{1.2.4}$$

If, however, $\det A = 0$, there exists no inverse of A and the matrix A is said to be singular. When $y = 0$, on the other hand, the equation (1.2.1) has non-null solutions for x, if and only if A is singular; this has an important application in the theory of matrix diagonalization. Note that, for a product of non-singular matrices, we have $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

The product of a number c and a matrix $A = \|A_{ij}\|$ is a matrix whose elements are c times the elements of A:

$$cA = \|cA_{ij}\|$$

Accordingly, a number commutes with any matrix. Note, however, that

$$\det cA = c^n \det A$$

where n is the dimensionality of A.

The well-known examples of matrices in two dimensions are the *Pauli spin* matrices defined by

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \qquad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{1.2.5a}$$

Their determinants are all equal to -1, and their traces are all zero. They also anticommute with each other:

$$\sigma_x\sigma_y = -\sigma_y\sigma_x, \qquad \sigma_y\sigma_z = -\sigma_z\sigma_y, \qquad \sigma_z\sigma_x = -\sigma_x\sigma_z \tag{1.2.5b}$$

Their squares are all equal to the unit matrix **1**; i.e. the σ_i are all involutional satisfying $x^2 = 1$. Therefore, they form a set of anticommuting matrices satisfying, with $\sigma_x = \sigma_1$, $\sigma_y = \sigma_2$ and $\sigma_z = \sigma_3$,

$$[\sigma_\nu, \sigma_\mu]_+ = \sigma_\nu\sigma_\mu + \sigma_\mu\sigma_\nu = 2\delta_{\nu\mu}; \qquad \nu, \mu = 1, 2, 3 \tag{1.2.6}$$

Furthermore, $\sigma_1\sigma_2\sigma_3 = \mathrm{i}$ so that $\sigma_1\sigma_2 = \mathrm{i}\sigma_3$. A set of matrices that satisfy anti-commutation relations as given by (1.2.6) is said to form *a Clifford algebra*.

Now, consider a set G of all non-singular matrices M in n dimensions. Then a product of two members of G is also a non-singular matrix belonging to G. This property is called *the group property* of G or the closure of G. Summarizing the properties of G discussed in this section, we may conclude that the set G satisfies the following properties:

1. The group property:[2] if $M_1, M_2 \in G$ then $M_1M_2 \in G$.
2. The associative property for the product: $(M_1M_2)M_3 = M_1(M_2M_3)$.
3. Existence of the unit matrix **1**: $\mathbf{1}M = M\mathbf{1}$ for any $M \in G$.
4. Existence of the inverse M^{-1} for any $M \in G$: $M^{-1}M = MM^{-1} = \mathbf{1}$.

[2] $M \in G$ means M belongs to the set G.

A set of elements that satisfies these four properties is said to form *a group*. Thus the set of non-singular $n \times n$ matrices forms a group called *the group of general linear transformations* in n dimensions, denoted $GL(n)$. Full discussion of the axiomatic system of a group will be given later in Chapter 3. Here, we need the group axiom simply to characterize a set of matrices.

1.2.1 Functions of a matrix

Let us define powers of a non-singular matrix A by

$$A^0 = 1, \qquad A^1 = A, \qquad A^{n+1} = A^n A, \qquad A^{-n} = (A^{-1})^n$$

where n is an integer. Then, corresponding to a function $f(z)$ of a scalar variable z that can be expanded in powers of z,

$$f(z) = \sum_{n=-\infty}^{\infty} c_n z^n = c_0 + c_1 z + c_2 z^2 + \cdots + c_{-1} z^{-1} + c_{-2} z^{-2} + \cdots \tag{1.2.7a}$$

one can define a function of a matrix A as follows:

$$f(A) = \sum_{n=-\infty}^{\infty} c_n A^n \tag{1.2.7b}$$

Such a function $f(A)$ commutes with any other function $g(A)$ of A. An elementary example of a function of A is the exponential function

$$f(sA) = \exp(sA) = 1 + sA + s^2A^2/2! + \cdots + s^nA^n/n! + \cdots$$

where s is a scalar parameter. One can differentiate the function $\exp(sA)$ with respect to s regarding A as a constant and obtain

$$f'(sA) = \mathrm{d}[\exp(sA)]/\mathrm{d}s = A\exp(sA)$$

where $f'(z)$ is the derivative of $f(z)$ with respect to z. This can be checked by the expansion.

In the special case of a diagonal matrix D

$$D = \mathrm{diag}\,[d_1, d_2, \ldots, d_n]$$

the function $f(D)$ of D is simply given by a diagonal matrix with elements $f(d_1)$, $f(d_2), \ldots, f(d_n)$, i.e.

$$f(D) = \mathrm{diag}\,[f(d_1), f(d_2), \ldots, f(d_n)] \tag{1.2.7c}$$

This follows simply from the definition (1.2.7b). Thus, for example,

$$\exp\left(\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}\right) = \begin{bmatrix} \mathrm{e}^{d_1} & 0 \\ 0 & \mathrm{e}^{d_2} \end{bmatrix} \tag{1.2.7d}$$

For the Pauli spin σ_z defined by (1.2.5a), we have

$$\exp(\lambda\sigma_z) = \begin{bmatrix} \mathrm{e}^{\lambda} & 0 \\ 0 & \mathrm{e}^{-\lambda} \end{bmatrix} \tag{1.2.7e}$$

where λ is a constant.

1.2.2 Special matrices

From a given matrix $A = \|A_{ij}\|$ one can define a new matrix by complex conjugation or by transposition of the rows and columns. The complex conjugate A^* and *the transpose* $A^\sim$ of A are defined by the elements

$$(A^*)_{ij} = A^*_{ij}, \qquad (A^\sim)_{ij} = A_{ji} \tag{1.2.8a}$$

where A^*_{ij} is the complex conjugate of the element A_{ij}. For a product of matrices we have

$$(ABC)^* = A^*B^*C^*, \qquad (ABC)^\sim = C^\sim B^\sim A^\sim$$

Note that a transformation of a vector x by a matrix A can be written in the following two ways:

$$\sum_j A_{ij}x_j = \sum_j x_j A^\sim_{ji}$$

If one combines the above two operations defined by (1.2.8a), one obtains *the adjoint* or *Hermitian conjugate matrix* defined by

$$A^\dagger = A^{\sim *} = A^{*\sim} \tag{1.2.8b}$$

These special matrices come in when we describe the transformations of the scalar products, $(\boldsymbol{v}, \boldsymbol{u})$ and $\langle \boldsymbol{v}, \boldsymbol{u} \rangle$,

$$(A\boldsymbol{v}, \boldsymbol{u}) = (\boldsymbol{v}, A^\sim \boldsymbol{u}), \qquad \langle A\boldsymbol{v}, \boldsymbol{u} \rangle = \langle \boldsymbol{v}, A^\dagger \boldsymbol{u} \rangle \tag{1.2.9}$$

which can be verified by writing out the respective scalar product. By assuming various relationships between a matrix A and its complex conjugate, transpose, adjoint and inverse, one can obtain special kinds of matrices. For example, if $A^\dagger = A$, A is said to be *self-adjoint* or *Hermitian*. Further examples will be introduced, however, later in Section 1.5 along with their transformation properties.

1.2.3 Direct products of matrices

Frequently, we encounter the concept of a direct product $P = A \otimes B$ of two square matrices. It is defined by their elements as follows: let $A = \|A_{ij}\|$ and $B = \|B_{ks}\|$ then

$$P = A \otimes B = \|A_{ij}B_{ks}\|$$

so that the elements of P may be expressed by

$$P_{ik,js} = A_{ij}B_{ks};$$
$$i, j = 1, 2, \ldots, d_A; \qquad k, s = 1, 2, \ldots, d_B \tag{1.2.10a}$$

where the respective dimensionalities d_A and d_B of A and B need not be the same and the dimensionality of the direct product P equals $d_A d_B$. By definition, the product of two similar direct products $P = A \otimes B$ and $P' = A' \otimes B'$ is given by

$$PP' = AA' \otimes BB'$$

where the dimensionalities of A and A' should be the same and so should those of B and B'. Note the trace of a direct product $A \otimes B$ equals the product of the traces of A and B:

$$\text{tr}\, A \otimes B = (\text{tr}\, A)(\text{tr}\, B)$$

Moreover, the transpose of a direct product is given by

$$(A \otimes B)^{\sim} = A^{\sim} \otimes B^{\sim}$$

The direct product $P = A \otimes B$ can also be expressed by *a super matrix* (whose elements are matrices) as follows:

$$A \otimes B = \begin{bmatrix} A_{11}B & A_{12}B & \dots & A_{1n}B \\ A_{21}B & A_{22}B & \dots & A_{2n}B \\ . & . & \dots & . \\ A_{n1}B & A_{n2}B & \dots & A_{nn}B \end{bmatrix} \tag{1.2.10b}$$

where $n = d_A$. For example, for the Pauli spin matrices,

$$\sigma_x \otimes \sigma_y = \begin{bmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{bmatrix}$$

1.2.4 Direct sums of matrices

Let A and B be two square matrices, then their direct sum is defined by

$$S = A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \tag{1.2.11}$$

It is a generalization of a diagonal matrix: it is a super matrix whose diagonal elements are square matrices while the remaining elements are null matrices. The respective dimensionalities d_A and d_B of A and B need not be the same and the dimensionality of the direct sum $A \oplus B$ equals $d_A + d_B$. For example, when $d_A = 2$ and $d_B = 3$ we have

$$A \oplus B = \begin{bmatrix} A_{11} & A_{12} & 0 & 0 & 0 \\ A_{21} & A_{22} & 0 & 0 & 0 \\ 0 & 0 & B_{11} & B_{12} & B_{13} \\ 0 & 0 & B_{21} & B_{22} & B_{23} \\ 0 & 0 & B_{31} & B_{32} & B_{33} \end{bmatrix}$$

Let $S' = A' \oplus B'$ be another direct sum that has a similar shape to that of S of (1.2.11), then their matrix product is given by

$$SS' = (A \oplus B)(A' \oplus B') = AA' \oplus BB'$$

Thus, e.g. for a polynomial of S

$$\sum c_n S^n = \sum_n c_n (A \oplus B)^n = \left(\sum_n c_n A^n\right) \oplus \left(\sum_n c_n B^n\right)$$

More generally, let $f(z)$ be a function of z defined by (1.2.7a), then

$$f(A \oplus B) = f(A) \oplus f(B) \tag{1.2.12}$$

i.e. a function of a direct sum $A \oplus B$ equals the direct sum of the functions $f(A)$ and $f(B)$. This means that a functional notation f may be regarded as a linear operator for a direct sum. For example,

$$\exp(A \oplus B) = \exp A \oplus \exp B \tag{1.2.13a}$$

so that (1.2.7d) can be rewritten as follows:

$$\exp(d_1 \oplus d_2) = \exp d_1 \oplus \exp d_2 \tag{1.2.13b}$$

The determinant and trace of a direct sum are given by

$$\det(A \oplus B) = \det A \det B$$
$$\operatorname{tr}(A \oplus B) = \operatorname{tr} A + \operatorname{tr} B \tag{1.2.14}$$

1.3 Similarity transformations

Let us consider the effect of a change of the coordinate system in $V^{(n)}$ on the linear transformation $y = Ax$ given by (1.2.1). Let $x' = (x'_1, x'_2, \ldots, x'_n)$ and $y' = (y'_1, y'_2, \ldots, y'_n)$ be the new coordinates of the two points originally defined by x and y respectively. Then, there exists a non-singular matrix S such that

$$x = Sx', \qquad y = Sy' \tag{1.3.1}$$

Substitution of these into $y = Ax$ yields the new transformation

$$y' = S^{-1}ASx'$$

The matrix defined by

$$A' = S^{-1}AS \tag{1.3.2}$$

is called the transform of the matrix A by the matrix S. Then, A is also the transform of A' by S^{-1}. The matrices which are transforms of one another are called equivalent matrices, while the transformation itself is called *the similarity transformation* or *equivalent transformation* by S.

The equivalent matrices have many properties in common. For example, their determinants are equal and so are their traces:

$$\det S^{-1}AS = \det S^{-1} \det A \det S = \det A \tag{1.3.3}$$
$$\operatorname{tr} S^{-1}AS = \operatorname{tr} ASS^{-1} = \operatorname{tr} A \tag{1.3.4}$$

We say that the determinant and trace of a matrix are *invariant* under a similarity transformation. Another important property of the similarity transformation is that

$$(S^{-1}AS)^n = S^{-1}A^nS \tag{1.3.5}$$

Suppose that $f(A)$ is a function of A that can be expanded in powers of A. Then we have

$$S^{-1}f(A)S = f(S^{-1}AS) \tag{1.3.6}$$

1.3.1 Functions of a matrix (revisited)

Suppose that the matrix A is diagonalized by the similarity transformation

$$S^{-1}AS = \Lambda = \operatorname{diag}[\lambda_1, \lambda_2, \ldots, \lambda_n]$$

Then we have, from (1.3.6) and (1.2.7c),

$$S^{-1}f(A)S = f(\Lambda) = \operatorname{diag}[f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)]$$

so that

$$f(A) = Sf(\Lambda)S^{-1} \tag{1.3.7}$$

This gives $f(A)$ explicitly as an $n \times n$ matrix. Even if $f(z)$ cannot be expanded in powers of z, we may define the function $f(A)$ of the matrix A by (1.3.7), provided that A can be diagonalized by a similarity transformation. According to this definition the functions like $\sqrt{A}$ and $\ln A$ become meaningful, even though they cannot be expanded in powers of A. In the next section we shall discuss the condition for a matrix to be diagonalized.

1.4 The characteristic equation of a matrix

Let A be an $n \times n$ matrix that can be diagonalized by a similarity transformation. Then there exists a non-singular matrix T such that

$$T^{-1}AT = \Lambda = \mathrm{diag}\,[\lambda_1, \lambda_2, \ldots, \lambda_n] \tag{1.4.1a}$$

This can be written in the form

$$AT = T\Lambda \qquad \text{or} \qquad \sum_j A_{kj}T_{ji} = T_{ki}\lambda i \tag{1.4.1b}$$

Let us regard the matrix T as a set of n column vectors and write the matrix elements in the form $T^i_j \equiv T_{ji}$ and define the ith column vector by

$$\boldsymbol{t}^{(i)} = (T^i_1, T^i_2, \ldots, T^i_n); \qquad i = 1, 2, \ldots, n \tag{1.4.2}$$

Then we arrive at the eigenvalue problem of the matrix A

$$A\boldsymbol{t}^{(i)} = \lambda_i \boldsymbol{t}^{(i)}; \qquad i = 1, 2, \ldots, n \tag{1.4.3}$$

where λ_i is called *an eigenvalue* of A and $\boldsymbol{t}^{(i)}$ is called *the eigenvector* of A belonging to the eigenvalue λ_i.

Conversely, if the eigenvalue problem (1.4.3) of A is solvable, i.e. if it provides a set of n linearly independent eigenvectors of A, then one constructs a non-singular matrix T by

$$T = [\boldsymbol{t}^{(1)}, \boldsymbol{t}^{(2)}, \ldots, \boldsymbol{t}^{(n)}] \tag{1.4.4}$$

which diagonalizes the matrix A by (1.4.1a).

The condition for the existence of the non-null eigenvector $\boldsymbol{t}^{(i)}$ of A is that the coefficient matrix $[\lambda_i\mathbf{1} - A]$ of (1.4.3) is singular for each λ_i. This implies that λ_i is a root of the following nth-order polynomial equation of x:

$$\begin{aligned} D^{(n)}(x) &= \det(x\mathbf{1} - A) \\ &= x^n + a_1x^{n-1} + \cdots + a_n \\ &= (x - \lambda_1)(x - \lambda_2)\ldots(x - \lambda_n) = 0 \end{aligned} \tag{1.4.5}$$

where $\det(x\mathbf{1} - A)$ is called *the secular determinant* of A. The coefficients a_1, a_2, $\ldots$, a_n are determined by the elements of the matrix A. This equation is called *the characteristic (or secular) equation* of the matrix A (even if A cannot be diagonalized), and its (*characteristic*) roots provide the eigenvalues of A. For an $n \times n$ matrix A, there exist exactly n characteristic roots, some of which may be repeated. One sees immediately that equivalent matrices satisfy the same characteristic equation, since

$$\det(x\mathbf{1} - S^{-1}AS) = \det[S^{-1}(x\mathbf{1} - A)S] = \det(x\mathbf{1} - A)$$

This means that all the characteristic roots and hence every coefficient of the characteristic equation are invariant under any similarity transformation. Note, in particular, that two invariants of A that we know already are given by the coefficients as follows: using (1.4.1a) and (1.4.5),

$$\operatorname{tr} A = \sum_i \lambda_1 = -a_1, \qquad \det A = \prod_i \lambda_i = (-1)^n a_n$$

The eigenvectors $\boldsymbol{t}^{(i)}$ and $\boldsymbol{t}^{(j)}$ belonging to different eigenvalues λ_i and λ_j are linearly independent. Suppose that $\boldsymbol{t}^{(i)}$ and $\boldsymbol{t}^{(j)}$ are linearly dependent, then there exist non-null coefficients such that

$$c_i \boldsymbol{t}^{(i)} + c_j \boldsymbol{t}^{(j)} = 0$$

By applying $(A - \lambda_j \mathbf{1})$ to both sides of this equation from the left we obtain $c_i(\lambda_i - \lambda_j) = 0$, which yields $c_i = 0$ for $\lambda_i \neq \lambda_j$, and hence $c_j = 0$ also. Thus, if all eigenvalues of A are different then all eigenvectors of A are linearly independent. This means that the matrix T defined by (1.4.4) with these eigenvectors is non-singular so that A can be diagonalized by T. On the other hand, if some of the eigenvalues of A are degenerate, the eigenvalue problem need not provide n linearly independent eigenvectors to form a non-singular transformation matrix T.

Before establishing the necessary and sufficient condition for a matrix to be diagonalized by a similarity transformation we shall show the following fundamental theorem for a square matrix.

Theorem 1.4.1. Every matrix satisfies its own characteristic equation.

This theorem follows from the very definition of the characteristic equation given by (1.4.5): since x commutes with A, we can substitute $x = A$ into (1.4.5) and obtain $D^{(n)}(A) = \det(A - A) = 0$, i.e.

$$D^{(n)}(A) = A^n + a_1 A^{n-1} + \cdots + a_n \mathbf{1} = 0 \tag{1.4.6}$$

1.4.1 Diagonalizability and projection operators

We shall now discuss the condition for a matrix to be diagonalized by a similarity transformation. Suppose that the characteristic equation of a matrix A has the form

$$D^{(n)}(x) = \prod_{i=1}^{r} (x - \lambda_i)^{n_i} = 0 \tag{1.4.7}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_r$ are all distinct roots with degeneracies $n_1, n_2, \ldots, n_r$ respectively. If the matrix A is diagonalized by a similarity transformation, i.e. $T^{-1}AT = \Lambda$, then the diagonal matrix Λ can be written in the form

$$\Lambda = \operatorname{diag}[\overbrace{\lambda_1, \ldots \lambda_1}^{n_1 \text{ times}}, \overbrace{\lambda_2, \ldots \lambda_2}^{n_2 \text{ times}}, \ldots, \overbrace{\lambda_r, \ldots \lambda_r}^{n_r \text{ times}}] \tag{1.4.8}$$