Cambridge University Press
978-1-108-41741-9 — Matrix Analysis and Applications
Xian-Da Zhang
Excerpt
More Information

1

Introduction to Matrix Algebra

In science and engineering, we often encounter the problem of solving a system of linear equations. Matrices provide the most basic and useful mathematical tool for describing and solving such systems. Matrices not only have many basic mathematics operations (such as transposition, inner product, outer product, inverse, generalized inverse etc.) but also have a variety of important scalar functions (e.g., a norm, a quadratic form, a determinant, eigenvalues, rank and trace etc.). There are also special matrix operations, such as the direct sum, direct product, Hadamard product, Kronecker product, vectorization, etc.

In this chapter, we begin our introduction to matrix algebra by relating matrices to the problem of solving systems of linear equations.

1.1 Basic Concepts of Vectors and Matrices

First we introduce the basic concepts of and notation for vectors and matrices.

1.1.1 Vectors and Matrices

Let $\mathbb{R}$ (or $\mathbb{C}$) denote the set of real (or complex) numbers. An m-dimensional *column vector* is defined as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}. \tag{1.1.1}$$

If the ith component x_i is a real number, i.e., $x_i \in \mathbb{R}$, for all $i = 1, \ldots, m$, then $\mathbf{x}$ is an m-dimensional *real vector* and is denoted $\mathbf{x} \in \mathbb{R}^{m\times 1}$ or simply $\mathbf{x} \in \mathbb{R}^m$. Similarly, if $x_i \in \mathbb{C}$ for some i, then $\mathbf{x}$ is known as an m-dimensional *complex vector* and is denoted $\mathbf{x} \in \mathbb{C}^m$. Here, $\mathbb{R}^m$ and $\mathbb{C}^m$ represent the sets of all real and complex m-dimensional column vectors, respectively.

An m-dimensional *row vector* $\mathbf{x} = [x_1, \ldots, x_m]$ is represented as $\mathbf{x} \in \mathbb{R}^{1\times m}$ or

$\mathbf{x} \in \mathbb{C}^{1\times m}$. To save space, an m-dimensional column vector is usually written as the transposed form of a row vector, denoted $\mathbf{x} = [x_1, \dots, x_m]^T$.

An $m \times n$ *matrix* is expressed as

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}]_{i=1,j=1}^{m,n}. \tag{1.1.2}$$

The matrix $\mathbf{A}$ with (i, j)th real entry $a_{ij} \in \mathbb{R}$ is called an $m \times n$ *real matrix*, and denoted by $\mathbf{A} \in \mathbb{R}^{m\times n}$. Similarly, $\mathbf{A} \in \mathbb{C}^{m\times n}$ is an $m \times n$ *complex matrix*.

An $m\times n$ matrix can be represented as $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ where its column vectors are $\mathbf{a}_j = [a_{1j}, \dots, a_{mj}]^T$, $j = 1, \dots, n$.

The *system of linear equations*

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}\right\} \tag{1.1.3}$$

can be simply rewritten using vector and matrix symbols as a *matrix equation*

$$\mathbf{Ax} = \mathbf{b}, \tag{1.1.4}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}. \tag{1.1.5}$$

In modeling physical problems, the matrix $\mathbf{A}$ is usually the symbolic representation of a physical system (e.g., a linear system, a filter, or a wireless communication channel). There are three different types of vector in science and engineering [232]:

(1) *Physical vector* Its elements are physical quantities with magnitude and direction, such as a displacement vector, a velocity vector, an acceleration vector and so forth.
(2) *Geometric vector* A directed line segment or arrow is usually used to visualize a physical vector. Such a representation is called a geometric vector. For example, $\mathbf{v} = \overrightarrow{AB}$ represents the directed line segment with initial point A and the terminal point B.
(3) *Algebraic vector* A geometric vector can be represented in algebraic form. For a geometric vector $\mathbf{v} = \overrightarrow{AB}$ on a plane, if its initial point is $A = (a_1, a_2)$ and its terminal point is $B = (b_1, b_2)$, then the geometric vector $\mathbf{v} = \overrightarrow{AB}$ can

be represented in an algebraic form $\mathbf{v} = \begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \end{bmatrix}$. Such a geometric vector described in algebraic form is known as an algebraic vector.

Physical vectors are those often encountered in practical applications, while geometric vectors and algebraic vectors are respectively the visual representation and the algebraic form of physical vectors. Algebraic vectors provide a computational tool for physical vectors.

Depending on the different types of element value, algebraic vectors can be divided into the following three types:

(1) *Constant vector* Its entries are real constant numbers or complex constant numbers, e.g., $\mathbf{a} = [1, 5, 4]^T$.
(2) *Function vector* It uses functions as entries, e.g., $\mathbf{x} = [x^1, \ldots, x^n]^T$.
(3) *Random vector* Its entries are random variables or signals, e.g., $\mathbf{x}(n) = [x_1(n), \ldots, x_m(n)]^T$ where $x_1(n), \ldots, x_m(n)$ are m random variables or random signals.

Figure 1.1 summarizes the classification of vectors.

- Vectors
  - Physical vectors
  - Geometric vectors
  - Algebraic vectors
    - Constant vectors
    - Function vectors
    - Random vectors

Figure 1.1 Classification of vectors.

Now we turn to matrices. An $m \times n$ matrix $\mathbf{A}$ is called a *square matrix* if $m = n$, a *broad matrix* for $m < n$, and a *tall matrix* for $m > n$.

The *main diagonal* of an $n \times n$ matrix $\mathbf{A} = [a_{ij}]$ is the segment connecting the top left to the bottom right corner. The entries located on the main diagonal, $a_{11}, a_{22}, \ldots, a_{nn}$, are known as the (main) *diagonal elements*.

An $n \times n$ matrix $\mathbf{A} = [a_{ij}]$ is called a *diagonal matrix* if all entries off the main diagonal are zero; it is then denoted by

$$\mathbf{D} = \mathbf{Diag}(d_{11}, \ldots, d_{nn}). \tag{1.1.6}$$

In particular, a diagonal matrix $\mathbf{I} = \mathbf{Diag}(1, \ldots, 1)$ is called an *identity matrix*, and $\mathbf{O} = \mathbf{Diag}(0, \ldots, 0)$ is known as a *zero matrix*.

A vector all of whose components are equal to zero is called a *zero vector* and is denoted as $\mathbf{0} = [0, \ldots, 0]^T$.

An $n \times 1$ vector $\mathbf{x} = [x_1, \ldots, x_n]^T$ with only one nonzero entry $x_i = 1$ constitutes a *basis vector*, denoted $\mathbf{e}_i$; e.g.,

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \tag{1.1.7}$$

Clearly, an $n \times n$ identity matrix $\mathbf{I}$ can be represented as $\mathbf{I} = [\mathbf{e}_1, \ldots, \mathbf{e}_n]$ using basis vectors.

In this book, we use often the following matrix symbols.

$\mathbf{A}(i,:)$ means the ith row of $\mathbf{A}$.

$\mathbf{A}(:,j)$ means the jth column of $\mathbf{A}$.

$\mathbf{A}(p:q, r:s)$ means the $(q-p+1) \times (s-r+1)$ *submatrix* consisting of the pth row to the qth row and the rth column to the sth column of $\mathbf{A}$. For example,

$$\mathbf{A}(3:6, 2:4) = \begin{bmatrix} a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \\ a_{52} & a_{53} & a_{54} \\ a_{62} & a_{63} & a_{64} \end{bmatrix}.$$

A matrix $\mathbf{A}$ is an $m \times n$ *block matrix* if it can be represented in the form

$$\mathbf{A} = [\mathbf{A}_{ij}] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \cdots & \mathbf{A}_{mn} \end{bmatrix},$$

where the $\mathbf{A}_{ij}$ are matrices. The notation $[\mathbf{A}_{ij}]$ refers to a matrix consisting of block matrices.

1.1.2 Basic Vector Calculus

Basic vector calculus requires vector addition, vector multiplication by a scalar and vector products.

The *vector addition* of $\mathbf{u} = [u_1, \ldots, u_n]^T$ and $\mathbf{v} = [v_1, \ldots, v_n]^T$ is defined as

$$\mathbf{u} + \mathbf{v} = [u_1 + v_1, \ldots, u_n + v_n]^T. \tag{1.1.8}$$

Vector addition has the following two main properties:

- *Commutative law* $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
- *Associative law* $(\mathbf{u} + \mathbf{v}) \pm \mathbf{w} = \mathbf{u} + (\mathbf{v} \pm \mathbf{w}) = (\mathbf{u} \pm \mathbf{w}) + \mathbf{v}$.

The *vector multiplication* of an $n \times 1$ vector $\mathbf{u}$ by a scalar α is defined as

$$\alpha\mathbf{u} = [\alpha u_1, \ldots, \alpha u_n]^T. \tag{1.1.9}$$

The basic property of vector multiplication by a scalar is that it obeys the distributive law:

$$\alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}. \tag{1.1.10}$$

The *inner product* (or *dot product* or *scalar product*) of two real or complex $n \times 1$ vectors $\mathbf{u} = [u_1, \ldots, u_n]^T$ and $\mathbf{v} = [v_1, \ldots, v_n]^T$, denoted $\langle \mathbf{u}, \mathbf{v} \rangle$, is defined as the real number

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T\mathbf{v} = u_1v_1 + \cdots + u_nv_n = \sum_{i=1}^{n} u_iv_i, \tag{1.1.11}$$

or the complex number

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^H\mathbf{v} = u_1^*v_1 + \cdots + u_n^*v_n = \sum_{i=1}^{n} u_i^*v_i, \tag{1.1.12}$$

where $\mathbf{u}^H$ is the complex conjugate transpose, or Hermitian conjugate, of $\mathbf{u}$.

The inner product of two vectors $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^H\mathbf{v}$ has several important applications; for example, it may be used to measure the size (or length) of a vector, the distance between vectors, the neighborhood of a vector and so on. We will present these applications later.

The *outer product* (or *cross product*) of an $m \times 1$ real vector and an $n \times 1$ real vector, denoted $\mathbf{u} \circ \mathbf{v}$, is defined as the $m \times n$ real matrix

$$\mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}^T = \begin{bmatrix} u_1v_1 & \cdots & u_1v_n \\ \vdots & \vdots & \vdots \\ u_mv_1 & \cdots & u_mv_n \end{bmatrix}; \tag{1.1.13}$$

if $\mathbf{u}$ and $\mathbf{v}$ are complex then the outer product is the $m \times n$ complex matrix

$$\mathbf{u} \circ \mathbf{v} = \mathbf{u}\mathbf{v}^H = \begin{bmatrix} u_1v_1^* & \cdots & u_1v_n^* \\ \vdots & \vdots & \vdots \\ u_mv_1^* & \cdots & u_mv_n^* \end{bmatrix}. \tag{1.1.14}$$

In signal processing, wireless communications, pattern recognition, etc., for two $m\times 1$ data vectors $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ and $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^T$, the $m \times m$ *autocorrelation matrix* is given by $\mathbf{R}_{xx} = E\{\mathbf{x}(t) \circ \mathbf{x}(t)\} = E\{\mathbf{x}(t)\mathbf{x}^H(t)\}$ and the $m \times m$ *cross-correlation matrix* is given by $\mathbf{R}_{xy} = E\{\mathbf{x}(t) \circ \mathbf{y}(t)\} = E\{\mathbf{x}(t)\mathbf{y}^H(t)\}$, where E is the expectation operator. Given the sample data $x_i(t), y_i(t), i = 1, \ldots, m$, $t = 1, \ldots, N$, the sample autocorrelation matrix $\hat{\mathbf{R}}_{xx}$ and the sample cross-correlation matrix $\hat{\mathbf{R}}_{xy}$ can be respectively estimated by

$$\hat{\mathbf{R}}_{xx} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{x}^H(t), \quad \hat{\mathbf{R}}_{xy} = \frac{1}{N}\sum_{t=1}^{N} \mathbf{x}(t)\mathbf{y}^H(t). \tag{1.1.15}$$

1.1.3 Basic Matrix Calculus

Basic matrix calculus requires the matrix transpose, conjugate, conjugate transpose, addition and multiplication.

Definition 1.1 If $\mathbf{A} = [a_{ij}]$ is an $m \times n$ matrix, then its *transpose* $\mathbf{A}^T$ is an $n \times m$ matrix with the (i, j)th entry $[\mathbf{A}^T]_{ij} = a_{ji}$. The *conjugate* of $\mathbf{A}$ is represented as $\mathbf{A}^*$ and is an $m \times n$ matrix with (i, j)th entry $[\mathbf{A}^*]_{ij} = a^*_{ij}$, while the *conjugate or Hermitian transpose* of $\mathbf{A}$, denoted $\mathbf{A}^H \in \mathbb{C}^{n \times m}$, is defined as

$$\mathbf{A}^H = (\mathbf{A}^*)^T = \begin{bmatrix} a^*_{11} & a^*_{21} & \cdots & a^*_{m1} \\ a^*_{12} & a^*_{22} & \cdots & a^*_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a^*_{1n} & a^*_{2n} & \cdots & a^*_{mn} \end{bmatrix}. \tag{1.1.16}$$

Definition 1.2 An $n \times n$ real (complex) matrix satisfying $\mathbf{A}^T = \mathbf{A}$ ($\mathbf{A}^H = \mathbf{A}$) is called a *symmetric matrix* (*Hermitian matrix*).

There are the following relationships between the transpose and conjugate transpose of a matrix:

$$\mathbf{A}^H = (\mathbf{A}^*)^T = (\mathbf{A}^T)^*. \tag{1.1.17}$$

For an $m \times n$ block matrix $\mathbf{A} = [\mathbf{A}_{ij}]$, its conjugate transpose $\mathbf{A}^H = [\mathbf{A}^H_{ji}]$ is an $n \times m$ block matrix:

$$\mathbf{A}^H = \begin{bmatrix} \mathbf{A}^H_{11} & \mathbf{A}^H_{21} & \cdots & \mathbf{A}^H_{m1} \\ \mathbf{A}^H_{12} & \mathbf{A}^H_{22} & \cdots & \mathbf{A}^H_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^H_{1n} & \mathbf{A}^H_{2n} & \cdots & \mathbf{A}^H_{mn} \end{bmatrix}.$$

The simplest algebraic operations with matrices are the addition of two matrices and the multiplication of a matrix by a scalar.

Definition 1.3 Given two $m \times n$ matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$, *matrix addition* $\mathbf{A} + \mathbf{B}$ is defined by $[\mathbf{A} + \mathbf{B}]_{ij} = a_{ij} + b_{ij}$. Similarly, *matrix subtraction* $\mathbf{A} - \mathbf{B}$ is defined as $[\mathbf{A} - \mathbf{B}]_{ij} = a_{ij} - b_{ij}$.

By using this definition, it is easy to verify that the addition and subtraction of two matrices obey the following rules:

- *Commutative law* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.
- *Associative law* $(\mathbf{A} + \mathbf{B}) \pm \mathbf{C} = \mathbf{A} + (\mathbf{B} \pm \mathbf{C}) = (\mathbf{A} \pm \mathbf{C}) + \mathbf{B}$.

Definition 1.4 Let $\mathbf{A} = [a_{ij}]$ be an $m \times n$ matrix, and α be a scalar. The product $\alpha\mathbf{A}$ is an $m \times n$ matrix and is defined as $[\alpha\mathbf{A}]_{ij} = \alpha a_{ij}$.

Definition 1.5 Consider an $m \times n$ matrix $\mathbf{A} = [a_{ij}]$ and an $r \times 1$ vector $\mathbf{x} = [x_1, \ldots, x_r]^T$. The product $\mathbf{Ax}$ exists only when $n = r$ and is an $m \times 1$ vector whose entries are given by

$$[\mathbf{Ax}]_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, m.$$

Definition 1.6 The *matrix product* of an $m \times n$ matrix $\mathbf{A} = [a_{ij}]$ and an $r \times s$ matrix $\mathbf{B} = [b_{ij}]$, denoted $\mathbf{AB}$, exists only when $n = r$ and is an $m \times s$ matrix with entries

$$[\mathbf{AB}]_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \quad i = 1, \ldots, m;\ j = 1, \ldots, s.$$

Theorem 1.1 *The matrix product obeys the following rules of operation:*

(a) *Associative law of multiplication* *If* $\mathbf{A} \in \mathbb{C}^{m\times n}, \mathbf{B} \in \mathbb{C}^{n\times p}$ *and* $\mathbf{C} \in \mathbb{C}^{p\times q}$, *then* $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.
(b) *Left distributive law of multiplication* *For two* $m \times n$ *matrices* $\mathbf{A}$ *and* $\mathbf{B}$, *if* $\mathbf{C}$ *is an* $n \times p$ *matrix then* $(\mathbf{A} \pm \mathbf{B})\mathbf{C} = \mathbf{AC} \pm \mathbf{BC}$.
(c) *Right distributive law of multiplication* *If* $\mathbf{A}$ *is an* $m \times n$ *matrix, while* $\mathbf{B}$ *and* $\mathbf{C}$ *are two* $n \times p$ *matrices, then* $\mathbf{A}(\mathbf{B} \pm \mathbf{C}) = \mathbf{AB} \pm \mathbf{AC}$.
(d) *If* α *is a scalar and* $\mathbf{A}$ *and* $\mathbf{B}$ *are two* $m\times n$ *matrices then* $\alpha(\mathbf{A}+\mathbf{B}) = \alpha\mathbf{A}+\alpha\mathbf{B}$.

Proof We will prove only (a) and (b) here, while the proofs of (c) and (d) are left to the reader as an exercise.

(a) Let $\mathbf{A}_{m\times n} = [a_{ij}], \mathbf{B}_{n\times p} = [b_{ij}], \mathbf{C}_{p\times q} = [c_{ij}]$, then

$$\begin{aligned}[\mathbf{A}(\mathbf{BC})]_{ij} &= \sum_{k=1}^{n} a_{ik}(\mathbf{BC})_{kj} = \sum_{k=1}^{n} a_{ik}\left(\sum_{l=1}^{p} b_{kl} c_{lj}\right) \\ &= \sum_{l=1}^{p}\sum_{k=1}^{n}(a_{ik} b_{kl}) c_{lj} = \sum_{l=1}^{p}[\mathbf{AB}]_{il} c_{lj} = [(\mathbf{AB})\mathbf{C}]_{ij}\end{aligned}$$

which means that $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.

(b) From the rule for matrix multiplication it is known that

$$[\mathbf{AC}]_{ij} = \sum_{k=1}^{n} a_{ik} c_{kj}, \quad [\mathbf{BC}]_{ij} = \sum_{k=1}^{n} b_{ik} c_{kj}.$$

Then, according to the matrix addition rule, we have

$$[\mathbf{AC} + \mathbf{BC}]_{ij} = [\mathbf{AC}]_{ij} + [\mathbf{BC}]_{ij} = \sum_{k=1}^{n} (a_{ik} + b_{ik}) c_{kj} = [(\mathbf{A} + \mathbf{B})\mathbf{C}]_{ij}.$$

This gives $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$. □

Generally speaking, the product of two matrices does not satisfy the commutative law, namely $\mathbf{AB} \neq \mathbf{BA}$.

Another important operation on a square matrix is that of finding its inverse.

Put $\mathbf{x} = [x_1, \ldots, x_n]^T$ and $\mathbf{y} = [y_1, \ldots, y_n]^T$. The matrix–vector product $\mathbf{Ax} = \mathbf{y}$ can be regarded as a *linear transform* of the vector $\mathbf{x}$, where the $n \times n$ matrix $\mathbf{A}$ is called the *linear transform matrix*. Let $\mathbf{A}^{-1}$ denote the *linear inverse transform* of the vector $\mathbf{y}$ onto $\mathbf{x}$. If $\mathbf{A}^{-1}$ exists then one has

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}. \tag{1.1.18}$$

This equation can be viewed as the result of using $\mathbf{A}^{-1}$ to premultiply the original linear transform $\mathbf{Ax} = \mathbf{y}$, giving $\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{y} = \mathbf{x}$, which means that the *linear inverse transform matrix* $\mathbf{A}^{-1}$ must satisfy $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. Furthermore, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ should be invertible as well. In other words, after premultiplying $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ by $\mathbf{A}$, we get $\mathbf{Ax} = \mathbf{AA}^{-1}\mathbf{y}$, which should be consistent with the original linear transform $\mathbf{Ax} = \mathbf{y}$. This means that $\mathbf{A}^{-1}$ must also satisfy $\mathbf{AA}^{-1} = \mathbf{I}$.

On the basis of the discussion above, the inverse matrix can be defined as follows.

Definition 1.7 Let $\mathbf{A}$ be an $n \times n$ matrix. The matrix $\mathbf{A}$ is said to be invertible if there is an $n \times n$ matrix $\mathbf{A}^{-1}$ such that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, and $\mathbf{A}^{-1}$ is referred to as the *inverse matrix* of $\mathbf{A}$.

The following are properties of the conjugate, transpose, conjugate transpose and inverse matrices.

1. The matrix conjugate, transpose and conjugate transpose satisfy the distributive law:
$$(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*, \quad (\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T, \quad (\mathbf{A}+\mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H.$$
2. The transpose, conjugate transpose and inverse matrix of product of two matrices satisfy the following relationship:
$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T, \quad (\mathbf{AB})^H = \mathbf{B}^H\mathbf{A}^H, \quad (\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$
in which both $\mathbf{A}$ and $\mathbf{B}$ are assumed to be invertible.
3. Each of the symbols for the conjugate, transpose and conjugate transpose can be exchanged with the symbol for the inverse:
$$(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*, \quad (\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T, \quad (\mathbf{A}^H)^{-1} = (\mathbf{A}^{-1})^H.$$
The notations $\mathbf{A}^{-*} = (\mathbf{A}^{-1})^*, \mathbf{A}^{-T} = (\mathbf{A}^{-1})^T$ and $\mathbf{A}^{-H} = (\mathbf{A}^{-1})^H$ are sometimes used.
4. For any $m \times n$ matrix $\mathbf{A}$, the $n \times n$ matrix $\mathbf{B} = \mathbf{A}^H\mathbf{A}$ and the $m \times m$ matrix $\mathbf{C} = \mathbf{AA}^H$ are Hermitian matrices.

1.1.4 Linear Independence of Vectors

Consider the system of linear equations (1.1.3). It can be written as the matrix equation $\mathbf{Ax} = \mathbf{b}$. Denoting $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, the m equations of (1.1.3) can be written as

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}.$$

This is called a *linear combination* of the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

Definition 1.8 A set of n vectors, denoted $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, is said to be *linearly independent* if the matrix equation

$$c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n = \mathbf{0} \tag{1.1.19}$$

has only zero solutions $c_1 = \cdots = c_n = 0$. If the above equation may hold for a set of coefficients that are not all zero then the n vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are said to be *linearly dependent.*

An $n \times n$ matrix $\mathbf{A}$ is *nonsingular* if and only if the matrix equation $\mathbf{Ax} = \mathbf{0}$ has only the zero solution $\mathbf{x} = \mathbf{0}$. If $\mathbf{Ax} = \mathbf{0}$ exists for any nonzero solution $\mathbf{x} \neq \mathbf{0}$ then the matrix $\mathbf{A}$ is *singular.*

For an $n \times n$ matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, the matrix equation $\mathbf{Ax} = \mathbf{0}$ is equivalent to

$$\mathbf{a}_1x_1 + \cdots + \mathbf{a}_nx_n = \mathbf{0}. \tag{1.1.20}$$

From the above definition it follows that the matrix equation $\mathbf{Ax} = \mathbf{0}$ has a zero solution vector only, i.e., the matrix $\mathbf{A}$ is nonsingular, if and only if the column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of $\mathbf{A}$ are linearly independent. Because of importance of this result, it is described in a theorem below.

Theorem 1.2 *An $n \times n$ matrix $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ is nonsingular if and only if its n column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly independent.*

To summarize the above discussions, the nonsingularity of an $n \times n$ matrix $\mathbf{A}$ can be determined in any of the following three ways:

(1) its column vectors are linearly independent;
(2) for the matrix equation $\mathbf{Ax} = \mathbf{b}$ there exists a unique nonzero solution;
(3) the matrix equation $\mathbf{Ax} = \mathbf{0}$ has only a zero solution.

1.1.5 Matrix Functions

The following are five common *matrix functions*:

Cambridge University Press
978-1-108-41741-9 — Matrix Analysis and Applications
Xian-Da Zhang
Excerpt
More Information

1. *Triangle matrix function*

$$\sin \mathbf{A} = \sum_{n=0}^{\infty} \frac{(-1)^n \mathbf{A}^{2n+1}}{(2n+1)!} = \mathbf{A} - \frac{1}{3!}\mathbf{A}^3 + \frac{1}{5!}\mathbf{A}^5 - \cdots \tag{1.1.21}$$

$$\cos \mathbf{A} = \sum_{n=0}^{\infty} \frac{(-1)^n \mathbf{A}^{2n}}{(2n)!} = \mathbf{I} - \frac{1}{2!}\mathbf{A}^2 + \frac{1}{4!}\mathbf{A}^4 - \cdots . \tag{1.1.22}$$

2. *Logarithm matrix function*

$$\ln(\mathbf{I} + \mathbf{A}) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\mathbf{A}^n = \mathbf{A} - \frac{1}{2}\mathbf{A}^2 + \frac{1}{3}\mathbf{A}^3 - \cdots . \tag{1.1.23}$$

3. *Exponential matrix function* [311], [179]

$$\mathrm{e}^{\mathbf{A}} = \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n = \mathbf{I} + \mathbf{A} + \frac{1}{2}\mathbf{A}^2 + \frac{1}{3!}\mathbf{A}^3 + \cdots \tag{1.1.24}$$

$$\mathrm{e}^{-\mathbf{A}} = \sum_{n=0}^{\infty} \frac{1}{n!}(-1)^n\mathbf{A}^n = \mathbf{I} - \mathbf{A} + \frac{1}{2}\mathbf{A}^2 - \frac{1}{3!}\mathbf{A}^3 + \cdots \tag{1.1.25}$$

$$\mathrm{e}^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{1}{2}\mathbf{A}^2t^2 + \frac{1}{3!}\mathbf{A}^3t^3 + \cdots . \tag{1.1.26}$$

4. *Matrix derivative* If the entries a_{ij} of the matrix $\mathbf{A}$ are the functions of a parameter t then the derivative of the matrix $\mathbf{A}$ is defined as follows:

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \dot{\mathbf{A}} = \begin{bmatrix} \frac{\mathrm{d}a_{11}}{\mathrm{d}t} & \frac{\mathrm{d}a_{12}}{\mathrm{d}t} & \cdots & \frac{\mathrm{d}a_{1n}}{\mathrm{d}t} \\ \frac{\mathrm{d}a_{21}}{\mathrm{d}t} & \frac{\mathrm{d}a_{22}}{\mathrm{d}t} & \cdots & \frac{\mathrm{d}a_{2n}}{\mathrm{d}t} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{d}a_{m1}}{\mathrm{d}t} & \frac{\mathrm{d}a_{m2}}{\mathrm{d}t} & \cdots & \frac{\mathrm{d}a_{mn}}{\mathrm{d}t} \end{bmatrix}. \tag{1.1.27}$$

- The derivative of a matrix exponential function is given by

$$\frac{\mathrm{d}\mathrm{e}^{\mathbf{A}t}}{\mathrm{d}t} = \mathbf{A}\mathrm{e}^{\mathbf{A}t} = \mathrm{e}^{\mathbf{A}t}\mathbf{A}. \tag{1.1.28}$$

- The derivative of matrix product is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{A}\mathbf{B}) = \frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t}\mathbf{B} + \mathbf{A}\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t}, \tag{1.1.29}$$

where $\mathbf{A}$ and $\mathbf{B}$ are matrix functions of the variable t.