1

Mathematical Preliminaries of Neurocomputing

Studies of systems of nonlinear ordinary differential equations, difference equations, linear algebra and nonlinear progranuning are fundamental to the analysis and synthesis of artificial neural networks (ANNs).

This chapter summarizes the mathematical background which may be helpful in the study of artificial neural networks and provides a convenient source for repeated reference to standard facts. On a first reading the chapter should be read with a minimum of attention to detail in order not to discourage the reader who may be impatient to get to ANNs applications. The purpose of this chapter is to provide a selective list of mathematical definitions, notations and results which will be frequently used. For a detailed exposition the reader can consult texts on mathematical analysis, algebra and optimization (nonlinear programming). This chapter is more referential than formally complete, and rather serves as a collection of notations and irnportant results used elsewhere in the book and collected here for better readability.

1.1 Linear Matrix Algebra

1.1.1 Matrix Representations and Notations

Matrices and Vectors

A matrix is a rectangular array of entries (elements) which are usually numbers or functions.

Let A be an $m \times n$ matrix given by

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{1.1}$$

which has m rows and n columns. The matrix $\boldsymbol{A}$ of dimension $m \times n$ will be denoted by $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ if the entries are real-valued or $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ if the entries are complex [1].

Other algebraic and topological notations which will be needed later are summarized in the list of symbols and notations presented in the Appendix.

The matrix $\boldsymbol{A}$ can be expressed compactly as

$$\boldsymbol{A} = [a_{ij}]_{m \times n}, \tag{1.2}$$

where a_{ij} is the entry from the i-th row in the j-th column. An $m \times 1$ matrix is called a column vector and a $1 \times n$ matrix is called a row vector. If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, then the n columns of $\boldsymbol{A}$ are denoted by $\boldsymbol{a}_i \in \mathbb{R}^{m \times 1} = \mathbb{R}^m$ $(i = 1, 2, \ldots, n)$. Hence $\boldsymbol{A}$ can be expressed in terms of its columns by

$$\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]. \tag{1.3}$$

In our notation, every vector is considered to be a column vector. For column vectors we customarily use bold lower case letters.

Transpose

The transpose of any matrix is obtained by rewriting all columns as rows. The transpose of the matrix $\boldsymbol{A}$ is denoted $\boldsymbol{A}^{\mathrm{T}}$. The transpose $\boldsymbol{A}^{\mathrm{T}}$ of the $m \times n$ matrix $\boldsymbol{A}$ is an $n \times m$ matrix the (i, j)-th entry of which is a_{ji}. It follows that

$$(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A} \tag{1.4}$$

and

$$(\boldsymbol{A}\,\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}}, \tag{1.5}$$

where

$$\boldsymbol{A} \in \mathbb{R}^{m \times n}, \quad \boldsymbol{B} \in \mathbb{R}^{n \times p},$$

$$\boldsymbol{A}\,\boldsymbol{B} = \left[\sum_{k=1}^{n} a_{ik}\, b_{kj} \right]_{m \times p}.$$

Symmetric, Diagonal and Unit Matrices

Let $\boldsymbol{A}$ be an $n \times n$ matrix. If $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$, then $\boldsymbol{A}$ is called symmetric. If $a_{ij} = 0$ for all $i \neq j$, then $\boldsymbol{A}$ is a diagonal matrix and is represented as

$$\boldsymbol{A} = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$

If all diagonal elements of a diagonal matrix are units, then the matrix is an identity (unit) matrix. The $n \times n$ identity matrix is denoted by $\boldsymbol{I}_n$ and its k-th column by $\boldsymbol{e}_k$. When the dimension is clear from the context we simply write $\boldsymbol{I}$.

Indexing and Partitioning of Matrices

The most elementary component of the matrix $\boldsymbol{A}$ is the element a_{ij}. Sometimes it is convenient to use more sophisticated elements of $\boldsymbol{A}$. If one or more entries of a matrix $\boldsymbol{A}$ form a matrix, then $\boldsymbol{A}$ is a partitioned matrix. A convenient way to indicate the block dimensions of a partitioned matrix is as follows:

$$\boldsymbol{A} = \begin{matrix} \\ m \\ n \end{matrix} \overset{\begin{matrix} p & q \end{matrix}}{\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}}, \tag{1.6}$$

with

$$\boldsymbol{A}_{11} \in \mathbb{R}^{m \times p}, \quad \boldsymbol{A}_{12} \in \mathbb{R}^{m \times q}, \quad \boldsymbol{A}_{21} \in \mathbb{R}^{n \times p}, \quad \boldsymbol{A}_{22} \in \mathbb{R}^{n \times q}.$$

In the product of partitioned matrices the submatrices are multiplied according to the same rules as the entries in matrix products, for example

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_1 \\ \boldsymbol{B}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}\,\boldsymbol{B}_1 + \boldsymbol{A}_{12}\,\boldsymbol{B}_2 \\ \boldsymbol{A}_{21}\,\boldsymbol{B}_1 + \boldsymbol{A}_{22}\,\boldsymbol{B}_2 \end{bmatrix}. \tag{1.7}$$

Of course, the submatrices must have compatible dimensions so that the matrix products are defined. If

$$\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{m \times n},$$

$$\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n] \in \mathbb{R}^{p \times n},$$

then

$$\boldsymbol{A}\,\boldsymbol{B}^{\mathrm{T}} = \sum_{i=1}^{n} \boldsymbol{a}_i\, \boldsymbol{b}_i^{\mathrm{T}}. \tag{1.8}$$

1.1.2 Inner and Outer Product

The inner product of two n-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ is obtained as the product of the row vector $\boldsymbol{x}^{\mathrm{T}}$ and the column vector $\boldsymbol{w}$, i.e. for real entries

$$\langle \boldsymbol{x}, \boldsymbol{w} \rangle := \boldsymbol{x}^{\mathrm{T}} \boldsymbol{w} = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} = \sum_{i=1}^{n} w_i\, x_i, \tag{1.9}$$

$$\boldsymbol{x}, \boldsymbol{w} \in \mathbb{R}^{n\times 1}$$

and for complex entries

$$\langle \boldsymbol{x}, \boldsymbol{w} \rangle := \bar{\boldsymbol{x}}^{\mathrm{T}} \boldsymbol{w} = \sum_{i=1}^{n} \bar{x}_i w_i , \tag{1.10}$$

$$\boldsymbol{x}, \boldsymbol{w} \in \mathbb{C}^{n\times 1}$$

where $\bar{\boldsymbol{x}}^{\mathrm{T}}$ is the complex conjugate transpose of $\boldsymbol{x}$ and $\bar{x}_i$ the complex conjugate of x_i. If the two vectors are variable in time, then the inner product on the linear space of continuous real functions in the interval $t_1 < t < t_2$ is defined by

$$\langle \boldsymbol{x}(t), \boldsymbol{w}(t) \rangle := \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \boldsymbol{x}^{\mathrm{T}}(t) \boldsymbol{w}(t) \, \mathrm{d}t$$

$$= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \sum_{i=1}^{n} x_i(t) w_i(t) \, \mathrm{d}t . \tag{1.11}$$

The notation $\rangle \boldsymbol{x}, \boldsymbol{w} \langle$ will be used to denote the outer product defined as

$$\rangle \boldsymbol{x}, \boldsymbol{w} \langle := \boldsymbol{x}\, \boldsymbol{w}^{\mathrm{T}} = \begin{bmatrix} x_1 w_1 & \cdots & x_1 w_n \\ \vdots & & \vdots \\ x_m w_1 & \cdots & x_m w_n \end{bmatrix}, \tag{1.12}$$

where $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{w} \in \mathbb{R}^n$.

1.1.3 Linear Independence of Vectors

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ be a set of vectors in $\mathbb{R}^m$ and $\alpha_1, \alpha_2, \ldots, \alpha_n$ be a set of scalars. The vector

$$\boldsymbol{c} := \sum_{i=1}^{n} \alpha_i \boldsymbol{a}_i$$

is said to be a linear combination of the vectors $\{\boldsymbol{a}_i\}$. A set of vectors $\{\boldsymbol{a}_i\}$ is a linearly independent set if and only if the linear combination $\boldsymbol{c}$ which equals the zero vector $\boldsymbol{0}$, is characterized by $\alpha_i = 0$ $(i = 1, \ldots, n)$. Mathematically, the linear independence will be written as

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{a}_i = \boldsymbol{0} \iff \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 . \tag{1.13}$$

Otherwise, a nontrivial combination of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ is the zero vector and the set of vectors $\{\boldsymbol{a}_i\}$ is said to be linearly dependent. The columns of the matrix

$\boldsymbol{A} \in \mathbb{R}^{m\times n}$ are linearly independent if and only if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is a nonsingular matrix (of full rank, cf. Section 1.1.4). The rows of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ are linearly independent if $\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}$ is a nonsingular matrix.

1.1.4 Rank of a Matrix

The rank of the matrix $\boldsymbol{A}$ is the maximum number of linearly independent columns, or, equivalently, the maximum number of linearly independent rows. The rank of the matrix $\boldsymbol{A}$ will be denoted by rank $(\boldsymbol{A})$. It is said that an $m \times n$ matrix $\boldsymbol{A}$ is of full rank if

$$\text{rank } (\boldsymbol{A}) = \min \{m, n\} .$$

It can be shown that for any matrix $\boldsymbol{A}$

$$\text{rank } (\boldsymbol{A}) = \text{rank } (\boldsymbol{A}^{\mathrm{T}}) = \text{rank } (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \text{rank } (\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}) . \tag{1.14}$$

1.1.5 Positive and Negative Definite Matrices

A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is said to be positive definite (negative definite) if the quadratic form $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x}$ satisfies

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} > 0 \quad (< 0) \quad \text{for all } \boldsymbol{x} \neq \boldsymbol{0},\ \boldsymbol{x} \in \mathbb{R}^n . \tag{1.15}$$

$\boldsymbol{A}$ is said to be positive (negative) semidefinite, if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{x} \geqq 0$ $(\leqq 0)$ for all $\boldsymbol{x} \in \mathbb{R}^n$. The matrices $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}$ are positive semidefinite for any $\boldsymbol{A}$. For a symmetric positive (negative) semidefinite matrix the eigenvalues are real and nonnegative (nonpositive). For a positive definite matrix all eigenvalues are positive [1,6].

1.1.6 The Inverse and Pseudoinverse of Matrices

If the matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{A}\,\boldsymbol{B} = \boldsymbol{B}\,\boldsymbol{A} = \boldsymbol{I}_n$, then the matrix $\boldsymbol{B}$ is the inverse of the matrix $\boldsymbol{A}$ and denoted by $\boldsymbol{A}^{-1}$. If $\boldsymbol{A}^{-1}$, exists then the matrix $\boldsymbol{A}$ is said to be nonsingular, otherwise $\boldsymbol{A}$ is singular.

For every matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, a unique matrix $\boldsymbol{A}^{+}$ exists which is called pseudoinverse (or Moore-Penrose generalized inverse) satisfying the following conditions [1]:

(i) $\boldsymbol{A}\,\boldsymbol{A}^{+}\,\boldsymbol{A} = \boldsymbol{A}$,

(ii) $\boldsymbol{A}^{+}\,\boldsymbol{A}\,\boldsymbol{A}^{+} = \boldsymbol{A}^{+}$,

(iii) $(\boldsymbol{A}\,\boldsymbol{A}^{+})^{\mathrm{T}} = \boldsymbol{A}\,\boldsymbol{A}^{+}$,

(iv) $(\boldsymbol{A}^{+}\,\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{A}^{+}\,\boldsymbol{A}$.

In the special case that the matrix $\boldsymbol{A}$ is a square nonsingular matrix, the pseudoinverse of $\boldsymbol{A}$ is simply its inverse, i.e. $\boldsymbol{A}^{+} = \boldsymbol{A}^{-1}$. Sometimes the pseudoinverse is defined as

$$\boldsymbol{A}^{+} = \lim_{\nu \to 0} (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \nu^2 \boldsymbol{I})^{-1}\,\boldsymbol{A}^{\mathrm{T}} = \lim_{\nu \to 0} \boldsymbol{A}^{\mathrm{T}}\,(\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}} + \nu^2 \boldsymbol{I})^{-1} \tag{1.16}$$

or

$$\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\,(\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}})^{-1} \tag{1.17}$$

if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ or $\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}$ is nonsingular.

The pseudoinverse matrix can be interpreted by the set of linear equations

$$\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}\,, \tag{1.18}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{x} \in \mathbb{R}^{n}$, $\boldsymbol{b} \in \mathbb{R}^{m}$, rank $(\boldsymbol{A}) = n$ and $m > n$. The vector

$$\boldsymbol{x}^{*} = \boldsymbol{A}^{+}\,\boldsymbol{b} \tag{1.19}$$

solves the above set of equations in the sense that the scalar error (objective) function

$$E(\boldsymbol{x}) = \frac{1}{2}\,(\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}\,(\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{b}) \tag{1.20}$$

becomes minimum for $\boldsymbol{x}^{*} = \boldsymbol{A}^{+}\,\boldsymbol{b}$. The minimum value of the error function is

$$E(\boldsymbol{x}^{*}) = \boldsymbol{b}^{\mathrm{T}}\,[\boldsymbol{I} - \boldsymbol{A}\,(\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A})^{-1}\,\boldsymbol{A}^{\mathrm{T}}]\,\boldsymbol{b}\,. \tag{1.21}$$

If α is a scalar, then its pseudoinverse can be determined as

$$\alpha^{+} = \begin{cases} \alpha^{-1}, & \text{if } \alpha \neq 0\,, \\ 0\,, & \text{if } \alpha = 0\,. \end{cases}$$

Some useful formulas for pseudoinverse matrices are listed in the following [1,10]:

(i) $(\boldsymbol{A}^{+})^{+} = \boldsymbol{A}$,

(ii) $(\alpha\,\boldsymbol{A})^{+} = \alpha^{-1}\,\boldsymbol{A}^{+}$, if $\alpha \neq 0$,

(iii) $(\boldsymbol{A}^{+})^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}})^{+}$,

(iv) $\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A})^{+}\,\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\,(\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}})^{+}$,

(v) $\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}}\,(\boldsymbol{A}^{+})^{\mathrm{T}} = \boldsymbol{A}$,

(vi) $\boldsymbol{A}^{+}\,\boldsymbol{A}\,\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}$,

(vii) $(\boldsymbol{A}^{+})^{\mathrm{T}}\,\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A} = \boldsymbol{A}$,

(viii) $\boldsymbol{A}^{\mathrm{T}}\,\boldsymbol{A}\,\boldsymbol{A}^{+} = \boldsymbol{A}^{\mathrm{T}}$,

(ix) rank $(\boldsymbol{A}^{+})$ = rank $(\boldsymbol{A})$ = rank $(\boldsymbol{A}^{\mathrm{T}})$.

1.1.7 Orthogonality, Unitary Matrices, Conjugate Vectors

A set of vectors $\{\boldsymbol{q}_i\}$ $(\boldsymbol{q}_i \in \mathbb{R}^{N}\ \forall\, i\,,\ i = 1,2,\ldots,n)$ is orthogonal if

$$\boldsymbol{q}_i^{\mathrm{T}}\,\boldsymbol{q}_j = 0 \tag{1.22}$$

whenever $i \neq j$ and orthonormal if $\boldsymbol{q}_i^{\mathrm{T}}\,\boldsymbol{q}_j = \delta_{ij}$, where δ_{ij} is the Kronecker symbol.

A square matrix $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is called orthogonal if its columns represent an orthonormal set, i.e. $\boldsymbol{Q}^{\mathrm{T}}\,\boldsymbol{Q} = \boldsymbol{Q}\,\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{I}_n$. [1] An important property of orthogonal matrices is that they have no effect on inner products, e.g.

$$\|\,\boldsymbol{Q}\,\boldsymbol{x}\,\|_2 = \|\,\boldsymbol{x}\,\|_2 \quad \text{and} \quad (\boldsymbol{Q}\,\boldsymbol{x})^{\mathrm{T}}\,(\boldsymbol{Q}\,\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\,\boldsymbol{y}\,. \tag{1.23}$$

For $m \neq n$, a matrix $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ is called orthogonal if $\boldsymbol{Q}^{\mathrm{T}}\,\boldsymbol{Q} = \boldsymbol{I}_n$. The complex analogue of an orthogonal matrix is the unitary matrix. A complex matrix $\boldsymbol{Q}$ is unitary if $\overline{\boldsymbol{Q}}^{\mathrm{T}}\,\boldsymbol{Q} = \boldsymbol{I}$, where $\overline{\boldsymbol{Q}}^{\mathrm{T}}$ denotes the transposed complex conjugate.

Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^{n}$ are $\boldsymbol{A}$-conjugate (or simply mutually conjugate with respect to the matrix $\boldsymbol{A}$) if $\boldsymbol{x}^{\mathrm{T}}\,\boldsymbol{A}\,\boldsymbol{y} = 0$ [1,9,10].

1.1.8 Eigenvalues and Eigenvectors

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a given square matrix. If there exists a scalar λ and a nonzero vector $\boldsymbol{v}$ such that

$$\boldsymbol{A}\,\boldsymbol{v} = \lambda\,\boldsymbol{v}\,, \tag{1.24}$$

then λ is called an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{v}$ is the corresponding eigenvector. All eigenvalues λ_i $(i = 1,2,\ldots,n)$ (some of which may be equal) can be obtained by solving the characteristic equation of $\boldsymbol{A}$, i.e.

[1] This means that the matrix $\boldsymbol{Q}$ is orthogonal if and only if $\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}^{-1}$.

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0. \tag{1.25}$$

If the matrix $\boldsymbol{A}$ is normal, i.e. $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, then it can be factorized into

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}, \tag{1.26}$$

where $\boldsymbol{V}$ is an $n \times n$ orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix which has all the eigenvalues of $\boldsymbol{A}$ as its diagonal elements. This means that a normal matrix $\boldsymbol{A}$ is orthogonally diagonalizable [1] and $\boldsymbol{\Lambda} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}$.

The set of all eigenvalues of a matrix is its spectrum. The spectral radius of the matrix $\boldsymbol{A}$ is defined as [1]

$$\rho(\boldsymbol{A}) := \max_i |\lambda_i(\boldsymbol{A})|, \tag{1.27}$$

where $\lambda_i(\boldsymbol{A})$ denote the eigenvalues of $\boldsymbol{A}$.

If the matrix $\boldsymbol{A}$ is nonsingular, then all its eigenvalues are nonzero and the eigenvalues of $\boldsymbol{B} = \boldsymbol{A}^{-1}$ are the reciprocals of the eigenvalues of $\boldsymbol{A}$.

If the matrix $\boldsymbol{A}$ is symmetric, then all the eigenvalues are real numbers. These eigenvalues correspond to n distinct eigenvectors. The eigenvectors of a symmetric matrix form an orthogonal set of vectors, i.e.

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = 0 \quad \text{for} \quad i \neq j, \tag{1.28a}$$

$$\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_i \neq \boldsymbol{0}. \tag{1.28b}$$

The maximum and minimum eigenvalues of the matrix $\boldsymbol{A}$ satisfy

$$\lambda_{\max}(\boldsymbol{A}) = \max_{\boldsymbol{v} \neq \boldsymbol{0}} \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}} \quad \text{and} \quad \lambda_{\min}(\boldsymbol{A}) = \min_{\boldsymbol{v} \neq \boldsymbol{0}} \frac{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}}{\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}}. \tag{1.29}$$

If λ is an eigenvalue, any nontrivial solution of $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$ is an eigenvector of $\boldsymbol{A}$ corresponding to λ. The pair $(\lambda, \boldsymbol{v})$ is called an eigenpair consisting of the eigenvalue λ and the corresponding eigenvector $\boldsymbol{v}$. Note that if $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A}$, then it is also any multiple of $\boldsymbol{v}$ with a nonzero multiplying factor. To eliminate the multiplicity of eigenpairs the eigenvector is usually normalized to unit length, i.e. $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v} = 1$.

If all the eigenvalues of a symmetric matrix $\boldsymbol{A}$ are strictly positive, then the matrix is positive definite. The matrix $\boldsymbol{A}$ is positive semidefinite (respectively, negative semidefinite) if all the eigenvalues of $\boldsymbol{A}$ are nonnegative (respectively, nonpositive). The matrix $\boldsymbol{A}$ is indefinite if $\boldsymbol{A}$ has at least one positive eigenvalue and

[1] In general, a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is diagonalizable, i.e. there exists a nonsingular matrix $\boldsymbol{V}$ so that $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}$ and $\boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}$, if and only if $\boldsymbol{A}$ has n linearly independent eigenvectors. In that case $\boldsymbol{V} = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_n]$ where $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ are the linearly independent eigenvectors of $\boldsymbol{A}$.

at least one negative eigenvalue.

1.1.9 Vector and Matrix Norms

Norms are nonnegative scalars which are used as a measure of length, size or distance depending on the context. The p-norm (also called L_p-norm) of the $n \times 1$ vector $\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ is defined as

$$\|\boldsymbol{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \tag{1.30}$$

where $p \in \mathbb{R}^+$. The parameter p usually takes one of the values 1, 2 or ∞ and the corresponding norms are called the one-, two- or infinity-norm, respectively. The latter is separately defined below [1].

The one-norm, also called absolute value or L_1-norm, is defined as

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|. \tag{1.31}$$

The two-norm, also called the Euclidean or L_2-norm, represents the ordinary vector length and can be defined as

$$\|\boldsymbol{x}\|_2 = (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})^{1/2} = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}. \tag{1.32}$$

For the Euclidean vector norm the Cauchy-Schwarz inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| = |\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leqq \|\boldsymbol{x}\|_2^{1/2} \|\boldsymbol{y}\|_2^{1/2} \tag{1.33}$$

holds where we have equality if the vector $\boldsymbol{x}$ is collinear with the vector $\boldsymbol{y}$ (i.e. $\boldsymbol{x} = \alpha\boldsymbol{y}$ for some scalar α). In some applications the weighted Euclidean norm defined by

$$\|\boldsymbol{x}\|_{\boldsymbol{Q}} := (\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})^{1/2} \tag{1.34}$$

is used where $\boldsymbol{Q}$ is a symmetric positive definite matrix. The infinity-norm, also called the L_∞- or Chebyshev norm, can be determined as

$$\|\boldsymbol{x}\|_\infty := \max_i |x_i|.$$

Any vector norm satisfies the following three conditions:

(i) $\|\boldsymbol{x}\| \geqq 0$ where the equality holds only for $\boldsymbol{x} = \boldsymbol{0}$,

(ii) $\| \alpha \boldsymbol{x} \| = | \alpha | \, \| \boldsymbol{x} \|$ for any scalar α,

(iii) $\| \boldsymbol{x} + \boldsymbol{y} \| \leqq \| \boldsymbol{x} \| + \| \boldsymbol{y} \|$ (the triangular inequality).

The distance between two vectors defined as

$$\| \boldsymbol{x} - \boldsymbol{y} \|_p = \left(\sum_{i=1}^{n} | x_i - y_i |^p \right)^{1/p}, \quad p \in \mathbb{R}^+ \tag{1.35}$$

is called the distance or measure of similarity in the Minkowski metric [1,10].

The norm of a matrix can be defined in different ways, but generally it is compatible with a vector norm in the sense that

$$\| \boldsymbol{A} \boldsymbol{x} \| \leqq \| \boldsymbol{A} \| \, \| \boldsymbol{x} \|$$

where the first and the last terms are vector norms and $\| \boldsymbol{A} \|$ is a matrix norm. A matrix norm is said to be consistent with the vector norm if it is defined by

$$\| \boldsymbol{A} \| = \max_{\boldsymbol{x}} \| \boldsymbol{A} \boldsymbol{x} \|$$

subject to the constraint $\| \boldsymbol{x} \| = 1$. The vector norms for $p = 1, 2$ and ∞ induce three norms for the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$:

$$\| \boldsymbol{A} \|_1 := \max_j \sum_{i=1}^{m} | a_{ij} | \quad \text{(the maximum absolute column sum)},$$

$$\| \boldsymbol{A} \|_2 := (\lambda_{\max})^{1/2} \quad \text{(the square root of the maximum eigenvalue of } \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \text{ or } \boldsymbol{A} \boldsymbol{A}^{\mathrm{T}}\text{)},^{1)}$$

$$\| \boldsymbol{A} \|_\infty := \max_i \sum_{j=1}^{n} | a_{ij} | \quad \text{(the maximum absolute row sum)}.$$

An alternative norm is the Frobenius norm which is not directly induced by a vector norm. In fact, the Frobenius norm is defined as the square root of the sum of the squares of all elements of the matrix $\boldsymbol{A}$, i.e. by

$$\| \boldsymbol{A} \|_F := \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$$

1.1.10 Singular Value Decomposition (SVD)

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Then real orthogonal matrices

1) The matrix two-norm $\| \ \|_2$ is sometimes known as the spectral norm.

$$\boldsymbol{U} := [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m],$$

$$\boldsymbol{V} := [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$$

exist such that

$$\boldsymbol{U}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V} = \text{pseudo-diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_p), \tag{1.36}$$

where

$$p = \min(m, n),$$

$$\sigma_1 \geqq \sigma_2 \geqq \sigma_3 \geqq \cdots \geqq \sigma_p \geqq 0.$$

The real nonnegative numbers $\sigma_1, \sigma_2, \ldots$ are called the singular values of $\boldsymbol{A}$. Taking into account that the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal, the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ can be written as the product

$$\boldsymbol{A} = \boldsymbol{U} \boldsymbol{S} \boldsymbol{V}^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}} \tag{1.37}$$

in which $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n \times n}$, and $\boldsymbol{S} \in \mathbb{R}^{m \times n}$ is the real pseudo-diagonal $m \times n$ matrix

$$\boldsymbol{S} := \text{pseudo-diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_p). \tag{1.38}$$

The SVD reveals many important properties of the matrix $\boldsymbol{A}$. Introducing r as the index of the smallest singular value, i.e.

$$r = \begin{cases} i \quad \text{with } \sigma_i \neq 0,\ \sigma_{i+1} = 0, & \text{if } \sigma_p = 0, \\ p, & \text{if } \sigma_p \neq 0, \end{cases}$$

the rank can be expressed by $\text{rank}\,(\boldsymbol{A}) = r$.

A specific and important special case of the SVD is obtained when $\boldsymbol{A}$ is a symmetric nonnegative definite matrix. In this case the matrix $\boldsymbol{S}$ takes the form

$$\boldsymbol{S} = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n), \tag{1.39}$$

where $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n \geqq 0$ are the real eigenvalues of $\boldsymbol{A}$ corresponding to the eigenvectors of $\boldsymbol{v}_i$ [1].

The singular values σ_i of the matrix $\boldsymbol{A}$ are the square roots of the nonzero eigenvalues of $\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A}$ or $\boldsymbol{A} \boldsymbol{A}^{\mathrm{T}}$. From the SVD of the matrix $\boldsymbol{A}$ its norms can easily be deduced. Specifically

$$\| \boldsymbol{A} \|_2 = \sigma_1, \tag{1.40}$$

$$\| \boldsymbol{A} \|_F = (\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_p^2)^{1/2} . \tag{1.41}$$

The pseudoinverse $\boldsymbol{A}^+$ of $\boldsymbol{A}$ is related to the SVD of $\boldsymbol{A}$ by the formula

$$\boldsymbol{A}^+ = \boldsymbol{V}\, \boldsymbol{S}^+\, \boldsymbol{U}^{\mathrm{T}} \tag{1.42}$$

in which $\boldsymbol{S}^+$ is obtained from $\boldsymbol{S}$ by replacing each positive diagonal entry $\sigma_i > 0$ by its reciprocal. One important use of the SVD is the solution of a system of linear equations (in the sense of the minimum L_2-norm)

$$\boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{b} ,$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^m$. Substituting the SVD of $\boldsymbol{A}$ yields

$$\boldsymbol{U}\, \boldsymbol{S}\, \boldsymbol{V}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{b} .$$

Partitioning the matrices in the form

$$\boldsymbol{U} = [\boldsymbol{U}_1\ \boldsymbol{U}_2] , \quad \boldsymbol{V} = [\boldsymbol{V}_1\ \boldsymbol{V}_2] ,$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} , \quad \boldsymbol{S}_0 = \operatorname{diag}(\sigma_1 , \sigma_2 , \ldots , \sigma_r) ,$$

we can obtain

$$\boldsymbol{x}^* = [\boldsymbol{V}_1 \quad \boldsymbol{V}_2] \begin{bmatrix} \boldsymbol{S}_0^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}} \\ \boldsymbol{U}_2^{\mathrm{T}} \end{bmatrix} \boldsymbol{b} = \boldsymbol{V}_1\, \boldsymbol{S}_0^{-1}\, \boldsymbol{U}_1^{\mathrm{T}}\, \boldsymbol{b} = \sum_{i=1}^{r} \sigma_i^{-1} \boldsymbol{v}_i\, \boldsymbol{u}_i^{\mathrm{T}} \boldsymbol{b} . \tag{1.43}$$

The result so obtained solves the minimization problem

$$\min_{\boldsymbol{x}} \| \boldsymbol{A}\, \boldsymbol{x} - \boldsymbol{b} \|_2^2 \tag{1.44}$$

called a linear least-squares problem.

1.1.11 Condition Numbers

The condition number cond $(\boldsymbol{A})$ of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is defined as

$$\operatorname{cond}_p (\boldsymbol{A}) := \| \boldsymbol{A} \|_p \, \| \boldsymbol{A}^+ \|_p , \tag{1.45}$$

where any norm can be chosen, i.e. $p = 1, 2, \infty$ or the Frobenius norm, and $\boldsymbol{A}^+$ is the pseudoinverse of $\boldsymbol{A}$. It can be shown that

$$\operatorname{cond} (\boldsymbol{A}) \geqq 1 . \tag{1.46}$$

Matrices with a small condition number are usually referred to as being well-conditioned and those with a large condition number as being ill-conditioned. The condition number arises naturally in describing the sensitivity of the solution of a system of linear equations. When solving the matrix equation $\boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{b}$ the error in the solution can be magnified by an amount as large as cond $(\boldsymbol{A})$ times the norm of the error of the data. Thus the error can be large for an ill-conditioned problem and small for a well-conditioned one.

For convenience and simplicity consider a system of linear equations $\boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a nonsingular matrix. We want to investigate how the solution vector $\boldsymbol{x}$ is affected by small perturbations of the data in the vector $\boldsymbol{b}$ and the matrix $\boldsymbol{A}$. Consider a simultaneous perturbation in the vector $\boldsymbol{b}$ and the matrix $\boldsymbol{A}$, i.e.

$$(\boldsymbol{A} + \Delta \boldsymbol{A})(\boldsymbol{x} + \Delta \boldsymbol{x}) = \boldsymbol{b} + \Delta \boldsymbol{b} . \tag{1.47}$$

Since $\boldsymbol{A}$ is nonsingular we can write

$$\Delta \boldsymbol{x} = (\boldsymbol{A} + \Delta \boldsymbol{A})^{-1} (\Delta \boldsymbol{b} - \Delta \boldsymbol{A}\, \boldsymbol{x}) \simeq \boldsymbol{A}^{-1} (\Delta \boldsymbol{b} - \Delta \boldsymbol{A}\, \boldsymbol{x}) . \tag{1.48}$$

Taking the norm in this equation and considering $\boldsymbol{A}\, \boldsymbol{x} = \boldsymbol{b}$ we obtain

$$\| \Delta \boldsymbol{x} \| = \| \boldsymbol{A}^{-1} (\Delta \boldsymbol{b} - \Delta \boldsymbol{A}\, \boldsymbol{x}) \|$$

$$\leqq \| \boldsymbol{A}^{-1} \| \, [\, \| \Delta \boldsymbol{b} \| + \| \Delta \boldsymbol{A} \| \, \| \boldsymbol{x} \| \,]$$

and

$$\| \boldsymbol{b} \| \leqq \| \boldsymbol{A} \| \, \| \boldsymbol{x} \| .$$

Hence we get

$$\frac{\| \Delta \boldsymbol{x} \|}{\| \boldsymbol{x} \|} \leqq \| \boldsymbol{A}^{-1} \| \left[\frac{\| \Delta \boldsymbol{b} \|}{\| \boldsymbol{x} \|} + \| \Delta \boldsymbol{A} \| \right]$$

$$= \| \boldsymbol{A}^{-1} \| \, \| \boldsymbol{A} \| \left[\frac{\| \Delta \boldsymbol{b} \|}{\| \boldsymbol{A} \| \, \| \boldsymbol{x} \|} + \frac{\| \Delta \boldsymbol{A} \|}{\| \boldsymbol{A} \|} \right]$$

$$\leqq \| \boldsymbol{A}^{-1} \| \, \| \boldsymbol{A} \| \left[\frac{\| \Delta \boldsymbol{b} \|}{\| \boldsymbol{b} \|} + \frac{\| \Delta \boldsymbol{A} \|}{\| \boldsymbol{A} \|} \right]$$

$$= \operatorname{cond} (\boldsymbol{A}) \left\{ \frac{\| \Delta \boldsymbol{b} \|}{\| \boldsymbol{b} \|} + \frac{\| \Delta \boldsymbol{A} \|}{\| \boldsymbol{A} \|} \right\} . \tag{1.49}$$

This inequality gives only an upper bound of the error, i.e. a worst possible error in the solution of $\boldsymbol{x}$.

In most applications the two-norm (L_2-norm) condition number of an $m \times n$ matrix $\boldsymbol{A}$ is used. If the singular values of $\boldsymbol{A}$ are denoted by σ_i then the two-norm condition number is determined as