# A. Matrixs

## A.1 Basic Operations

For a real matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ , the element in the i -th row and j -th column is denoted as $(\mathbf{A})ij = Aij$ . The **transpose** of matrix $\mathbf{A}$ is written as $\mathbf{A}^T$ , where $(\mathbf{A}^T)ij = Aji$ . Clearly:

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T, \tag{A.1}$$

$$(\mathbf{A}\mathbf{B})^T = \mathbf{B}^T \mathbf{A}^T. \tag{A.2}$$

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ , if $m = n$ , it is called an n -order square matrix. Let $\mathbf{I}_n$ denote the n -order identity matrix. The inverse matrix $\mathbf{A}^{-1}$ satisfies $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ . It can be easily shown:

$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T, \tag{A.3}$$

$$(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{A.4}$$

For an n -order square matrix $\mathbf{A}$ , its **trace** is the sum of the elements on its main diagonal, i.e., $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} A_{ii}$ . The following properties hold:

$$\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A}), \tag{A.5}$$

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}), \tag{A.6}$$

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A}), \tag{A.7}$$

$$\mathrm{tr}(\mathbf{A}\mathbf{B}\mathbf{C}) = \mathrm{tr}(\mathbf{B}\mathbf{C}\mathbf{A}) = \mathrm{tr}(\mathbf{C}\mathbf{A}\mathbf{B}). \tag{A.8}$$

The **determinant** of an n -order square matrix $\mathbf{A}$ is defined as:

$$\det(\mathbf{A}) = \sum_{\sigma \in S_n} \mathrm{par}(\sigma) A_{1\sigma_1} A_{2\sigma_2} \dots A_{n\sigma_n}, \tag{A.9}$$

where $S_n$ is the set of all n-order permutations, and $\mathrm{par}(\sigma)$ takes the value -1 or +1 , depending on whether $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ is an **odd permutation** or an **even permutation**—that is, the number of inversions is odd or even, respectively.

For an n -order square matrix $\mathbf{A}$ , the determinant has the following properties:

$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}), \tag{A.10}$$

$$\det(\mathbf{A}^T) = \det(\mathbf{A}), \tag{A.11}$$

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}), \tag{A.12}$$

$$\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}, \tag{A.13}$$

$$\det(\mathbf{A}^n) = \det(\mathbf{A})^n. \tag{A.14}$$

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ , the **Frobenius norm** is defined as:

$$\|\mathbf{A}\|F = \left(\mathrm{tr}(\mathbf{A}^T\mathbf{A})\right)^{1/2} = \left(\sum i = 1^m \sum_{j=1}^{n} A_{ij}^2\right)^{1/2}. \tag{A.15}$$

It is easy to see that the Frobenius norm of a matrix corresponds to the $L_2$ -norm of the matrix when treated as a vector.

## A.2 Derivative

The derivative of a vector $\mathbf{a}$ with respect to a vector $\mathbf{x}$ , and the derivative of $\mathbf{x}$ with respect to $\mathbf{a}$ , are both vectors. The i -th component is given by:

$$\left(\frac{\partial \mathbf{a}}{\partial \mathbf{x}}\right)_i = \frac{\partial a_i}{\partial x}, \tag{A.16}$$

$$\left(\frac{\partial \mathbf{x}}{\partial \mathbf{a}}\right)_i = \frac{\partial x}{\partial a_i}. \tag{A.17}$$

Similarly, for a matrix $\mathbf{A}$ , the derivative with respect to a vector $\mathbf{x}$ , and the derivative of $\mathbf{x}$ with respect to $\mathbf{A}$ , are both matrices. The element in the i -th row and j -th column is given by:

$$\left(\frac{\partial \mathbf{A}}{\partial \mathbf{x}}\right)_{ij} = \frac{\partial A_{ij}}{\partial x}. \tag{A.18}$$

$$\left(\frac{\partial \mathbf{x}}{\partial \mathbf{A}}\right)_{ij} = \frac{\partial x}{\partial A_{ij}}. \tag{A.18}$$

For a function $f(\mathbf{x})$ , assuming the elements of the vector $\mathbf{x}$ are differentiable, the first-order derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$ is a vector, with the i-th component defined as:

$$(\nabla f(\mathbf{x}))_i = \frac{\partial f(\mathbf{x})}{\partial x_i}. \tag{A.20}$$

The second-order derivative of $f(\mathbf{x})$ with respect to $\mathbf{x}$ is called the **Hessian matrix**, where the i -th row and j -th column element is defined as:

$$\left(\nabla^2 f(\mathbf{x})\right)_{ij} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}. \tag{A.21}$$

The derivatives of vectors and matrices satisfy the **product rule**. Specifically:

$$\frac{\partial(\mathbf{x}^T\mathbf{a})}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}, \tag{A.22}$$

$$\frac{\partial(\mathbf{AB})}{\partial \mathbf{x}} = \frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial \mathbf{x}}. \tag{A.23}$$

Here $\mathbf{a}$ is a constant vector compared to $\mathbf{x}$

From $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and Equation (A.23), the derivative of the inverse matrix can be expressed as:

$$\frac{\partial \mathbf{A}^{-1}}{\partial \mathbf{x}} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{A}^{-1}. \tag{A.24}$$

If the target of differentiation is the elements of matrix $\mathbf{A}$ , we have:

$$\frac{\partial \mathrm{tr}(\mathbf{AB})}{\partial A_{ij}} = B_{ji}, \tag{A.25}$$

$$\frac{\partial \mathrm{tr}(\mathbf{AB})}{\partial \mathbf{A}} = \mathbf{B}^T. \tag{A.26}$$

We further have:

$$\frac{\partial \mathrm{tr}(\mathbf{A}^T\mathbf{B})}{\partial \mathbf{A}} = \mathbf{B}, \tag{A.27}$$

$$\frac{\partial \mathrm{tr}(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{I}, \tag{A.28}$$

$$\frac{\partial \mathrm{tr}(\mathbf{ABA}^T)}{\partial \mathbf{A}} = \mathbf{A}(\mathbf{B} + \mathbf{B}^T). \tag{A.29}$$

From Equations (A.15) and (A.29), we obtain:

$$\frac{\partial \|\mathbf{A}\|_F^2}{\partial \mathbf{A}} = \frac{\partial \mathrm{tr}(\mathbf{AA}^T)}{\partial \mathbf{A}} = 2\mathbf{A}. \tag{A.30}$$

The **chain rule** is an essential tool for computing derivatives of composite functions. Simply put, for functions f , g , and h , where $f(\mathbf{x}) = g(h(\mathbf{x}))$ , we have:

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial g(h(\mathbf{x}))}{\partial h(\mathbf{x})} \cdot \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}. \tag{A.31}$$

For example, when computing the following, treating $\mathbf{Ax} - \mathbf{b}$ as a whole simplifies the calculation:

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{Ax} - \mathbf{b})^T \mathbf{W}(\mathbf{Ax} - \mathbf{b}) = \frac{\partial(\mathbf{Ax} - \mathbf{b})}{\partial \mathbf{x}} \cdot 2\mathbf{W}(\mathbf{Ax} - \mathbf{b}) = 2\mathbf{A}^T \mathbf{W}(\mathbf{Ax} - \mathbf{b}).. \quad \text{(A.32)}$$

## A.3 Singular Value Decomposition

Any real matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be decomposed as:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad \text{(A.33)}$$

where:

- $\mathbf{U} \in \mathbb{R}^{m \times m}$ is an **orthogonal matrix** (unitary matrix in the real case), satisfying $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ ,
- $\mathbf{V} \in \mathbb{R}^{n \times n}$ is also an **orthogonal matrix**, satisfying $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ ,
- $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal matrix with elements $(\mathbf{\Sigma})_{ii} = \sigma_i$ and all other elements equal to $0$ . The singular values $\sigma_i$ are **non-negative real numbers** that satisfy: $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$.