DIFFERENTIATION WITH RESPECT TO A VECTOR

The first derivative of a scalar-valued function $f(\mathbf{x})$ with respect to a vector $\mathbf{x} = [x_1 \; x_2]^T$ is called the gradient of $f(\mathbf{x})$ and defined as

$$\nabla f(\mathbf{x}) = \frac{d}{d\mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{bmatrix} \tag{C.1}$$

Based on this definition, we can write the following equation.

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{y} = \frac{\partial}{\partial \mathbf{x}} \mathbf{y}^T \mathbf{x} = \frac{\partial}{\partial \mathbf{x}} (x_1 y_1 + x_2 y_2) = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{y} \tag{C.2}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{x} = \frac{\partial}{\partial \mathbf{x}} (x_1^2 + x_2^2) = 2 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2\mathbf{x} \tag{C.3}$$

Also with an $M \times N$ matrix A, we have

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A \mathbf{y} = \frac{\partial}{\partial \mathbf{x}} \mathbf{y}^T A^T \mathbf{x} = A\mathbf{y} \tag{C.4a}$$

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{y}^T A \mathbf{x} = \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A^T \mathbf{y} = A^T \mathbf{y} \tag{C.4b}$$

where

$$\mathbf{x}^T A \mathbf{y} = \sum_{m=1}^{M} \sum_{n=1}^{N} a_{mn} x_m y_n \tag{C.5}$$

Especially for a square, symmetric matrix A with $M = N$, we have

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T A \mathbf{x} = (A + A^T)\mathbf{x} \xrightarrow{\text{if } A \text{ is symmetric}} 2A\mathbf{x} \tag{C.6}$$

The second derivative of a scalar function $f(\mathbf{x})$ with respect to a vector $\mathbf{x} = [x_1\ x_2]^T$ is called the Hessian of $f(\mathbf{x})$ and is defined as

$$H(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \frac{d^2}{d\mathbf{x}^2} f(\mathbf{x}) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 \end{bmatrix} \tag{C.7}$$

Based on this definition, we can write the following equation:

$$\frac{d^2}{d\mathbf{x}^2} \mathbf{x}^T A \mathbf{x} = A + A^T \xrightarrow{\text{if } A \text{ is symmetric}} 2A \tag{C.8}$$

On the other hand, the first derivative of a vector-valued function $\mathbf{f}(\mathbf{x})$ with respect to a vector $\mathbf{x} = [x_1\ x_2]^T$ is called the Jacobian of $f(\mathbf{x})$ and is defined as

$$J(\mathbf{x}) = \frac{d}{d\mathbf{x}} \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix} \tag{C.9}$$