1 Jacobian Derivatives

Let $\boldsymbol{y}$ be an m-dimensional vector consisting of y_1 through y_m:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

Its transposed vector is given by

$$\boldsymbol{y}^{\mathrm{T}} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix}.$$

Assume that elements y_1 through y_m depend on scalar x. Partial derivatives of $\boldsymbol{y}$ and $\boldsymbol{y}^{\mathrm{T}}$ with respect to x are defined as follows:

$$\frac{\partial \boldsymbol{y}}{\partial x} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x} \\ \dfrac{\partial y_2}{\partial x} \\ \vdots \\ \dfrac{\partial y_m}{\partial x} \end{bmatrix}, \quad \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial x} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x} & \dfrac{\partial y_2}{\partial x} & \cdots & \dfrac{\partial y_m}{\partial x} \end{bmatrix}.$$

Let y be a scalar depending on an n-dimensional vector $\boldsymbol{x}$. Assume that $\boldsymbol{x}$ consists of x_1 through x_n:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Its transposed vector is given by

$$\boldsymbol{x}^{\mathrm{T}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}.$$

Partial derivatives of y with respect to $\boldsymbol{x}$ and $\boldsymbol{x}^{\mathrm{T}}$ are defined as follows:

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} \\ \dfrac{\partial y}{\partial x_2} \\ \vdots \\ \dfrac{\partial y}{\partial x_n} \end{bmatrix}, \quad \frac{\partial y}{\partial \boldsymbol{x}^{\mathrm{T}}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} & \dfrac{\partial y}{\partial x_2} & \cdots & \dfrac{\partial y}{\partial x_n} \end{bmatrix}.$$

Assume that m-dimensional vector $\boldsymbol{y}$ depends on n-dimensional vector $\boldsymbol{x}$. Partial derivative of $\boldsymbol{y}$ with respect to $\boldsymbol{x}^{\mathrm{T}}$ is defined as

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial \boldsymbol{y}}{\partial x_1} & \frac{\partial \boldsymbol{y}}{\partial x_2} & \cdots & \frac{\partial \boldsymbol{y}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{1}$$

or equivalently

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial y_1}{\partial \boldsymbol{x}^{\mathrm{T}}} \\ \frac{\partial y_2}{\partial \boldsymbol{x}^{\mathrm{T}}} \\ \vdots \\ \frac{\partial y_m}{\partial \boldsymbol{x}^{\mathrm{T}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}. \tag{2}$$

Note that the above two equations are equivalent to each other, resulting in an $m \times n$ matrix. Partial derivative of $\boldsymbol{y}^{\mathrm{T}}$ with respect to $\boldsymbol{x}$ is defined as

$$\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial x_1} \\ \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial x_2} \\ \vdots \\ \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{3}$$

or equivalently

$$\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \boldsymbol{x}} & \frac{\partial y_2}{\partial \boldsymbol{x}} & \cdots & \frac{\partial y_m}{\partial \boldsymbol{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}. \tag{4}$$

Note that the above two equations are equivalent to each other, resulting in an $n \times m$ matrix. The above equations directly yield the followings:

$$\left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^{\mathrm{T}}}\right)^{\mathrm{T}} = \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}}, \quad \frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left(\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}}.$$

Assume that vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ depend on vector $\boldsymbol{x}$. Let A be a constant matrix that defines a quadratic form $\boldsymbol{y}^{\mathrm{T}} A \boldsymbol{z}$. Since $\boldsymbol{y}^{\mathrm{T}} A \boldsymbol{z} = \boldsymbol{z}^{\mathrm{T}} A^{\mathrm{T}} \boldsymbol{y}$, the gradient vector of the quadratic form with respect to n-dimensional vector $\boldsymbol{x}$ is described as

$$\frac{\partial (\boldsymbol{y}^{\mathrm{T}} A \boldsymbol{z})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}} A \boldsymbol{z} + \frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{x}} A^{\mathrm{T}} \boldsymbol{y}. \tag{5}$$

Partial derivatives $\partial \boldsymbol{y}^{\mathrm{T}}/\partial \boldsymbol{x}$ and $\partial \boldsymbol{z}^{\mathrm{T}}/\partial \boldsymbol{x}$ are given in (3) or (4). Note that the above equation provides an n-dimensional gradient vector. Additionally, the above equation yields

$$\frac{\partial (\boldsymbol{y}^{\mathrm{T}} A \boldsymbol{y})}{\partial \boldsymbol{x}} = 2 \frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}} A \boldsymbol{y}. \tag{6}$$

Recall that matrix A that defines quadratic form $\boldsymbol{y}^{\mathrm{T}} A \boldsymbol{y}$ should be symmetric.

2 Time Derivatives

Let y be a scalar depending on an n-dimensional vector $\boldsymbol{x}$ consisting of x_1 through x_n. Assume that x_1 through x_n depend on time. Time derivative of y is then described as:

$$\begin{aligned} \dot{y} &= \frac{\partial y}{\partial x_1}\frac{\mathrm{d}x_1}{\mathrm{d}t} + \frac{\partial y}{\partial x_2}\frac{\mathrm{d}x_2}{\mathrm{d}t} + \cdots + \frac{\partial y}{\partial x_n}\frac{\mathrm{d}x_n}{\mathrm{d}t} \\ &= \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \frac{\partial y}{\partial \boldsymbol{x}^{\mathrm{T}}} \dot{\boldsymbol{x}} \end{aligned} \tag{7}$$

or equivalently

$$\dot{y} = \left(\frac{\partial y}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} \dot{\boldsymbol{x}} = \left(\frac{\partial}{\partial \boldsymbol{x}} y\right)^{\mathrm{T}} \dot{\boldsymbol{x}}. \tag{8}$$

The first-order partial derivative

$$\frac{\partial y}{\partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} \\ \dfrac{\partial y}{\partial x_2} \\ \vdots \\ \dfrac{\partial y}{\partial x_n} \end{bmatrix} \tag{9}$$

is referred to as gradient vector.

Noting that $\partial y/\partial x_k$ depends on $\boldsymbol{x}$, time derivative of $\partial y/\partial x_k$ is described as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial y}{\partial x_k} = \left(\frac{\partial}{\partial \boldsymbol{x}}\frac{\partial y}{\partial x_k}\right)^{\mathrm{T}} \dot{\boldsymbol{x}}.$$

Time derivative of vector $\partial y/\partial \boldsymbol{x}$ is then described as:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial y}{\partial \boldsymbol{x}} &= \begin{bmatrix} \dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\partial y}{\partial x_1} \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\partial y}{\partial x_2} \\ \vdots \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\dfrac{\partial y}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \left(\dfrac{\partial}{\partial \boldsymbol{x}}\dfrac{\partial y}{\partial x_1}\right)^{\mathrm{T}} \dot{\boldsymbol{x}} \\ \left(\dfrac{\partial}{\partial \boldsymbol{x}}\dfrac{\partial y}{\partial x_2}\right)^{\mathrm{T}} \dot{\boldsymbol{x}} \\ \vdots \\ \left(\dfrac{\partial}{\partial \boldsymbol{x}}\dfrac{\partial y}{\partial x_n}\right)^{\mathrm{T}} \dot{\boldsymbol{x}} \end{bmatrix} = \begin{bmatrix} \left(\dfrac{\partial}{\partial \boldsymbol{x}}\dfrac{\partial y}{\partial x_1}\right)^{\mathrm{T}} \\ \left(\dfrac{\partial}{\partial \boldsymbol{x}}\dfrac{\partial y}{\partial x_2}\right)^{\mathrm{T}} \\ \vdots \\ \left(\dfrac{\partial}{\partial \boldsymbol{x}}\dfrac{\partial y}{\partial x_n}\right)^{\mathrm{T}} \end{bmatrix} \dot{\boldsymbol{x}} \\
&= \begin{bmatrix} \dfrac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\dfrac{\partial y}{\partial x_1} \\ \dfrac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\dfrac{\partial y}{\partial x_2} \\ \vdots \\ \dfrac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\dfrac{\partial y}{\partial x_n} \end{bmatrix} \dot{\boldsymbol{x}} = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}\frac{\partial y}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}} = \frac{\partial^2 y}{\partial \boldsymbol{x}^{\mathrm{T}}\partial \boldsymbol{x}}\dot{\boldsymbol{x}}.
\end{aligned}$$

The second-order partial derivative

$$\frac{\partial^2 y}{\partial \boldsymbol{x}^{\mathrm{T}} \partial \boldsymbol{x}} = \begin{bmatrix} \dfrac{\partial^2 y}{\partial x_1 \partial x_1} & \dfrac{\partial^2 y}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 y}{\partial x_n \partial x_1} \\ \dfrac{\partial^2 y}{\partial x_1 \partial x_2} & \dfrac{\partial^2 y}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 y}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 y}{\partial x_1 \partial x_n} & \dfrac{\partial^2 y}{\partial x_2 \partial x_n} & \cdots & \dfrac{\partial^2 y}{\partial x_n \partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2 y}{\partial x_1 \partial x_1} & \dfrac{\partial^2 y}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 y}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 y}{\partial x_2 \partial x_1} & \dfrac{\partial^2 y}{\partial x_2 \partial x_2} & \cdots & \dfrac{\partial^2 y}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 y}{\partial x_n \partial x_1} & \dfrac{\partial^2 y}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 y}{\partial x_n \partial x_n} \end{bmatrix} \tag{10}$$

is referred to as Hessian matrix. Hessian matrix is symmetric.

Differentiating (8) with respect to time t yields the second-order time derivative:

$$\begin{aligned} \ddot{y} &= \left(\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial y}{\partial \boldsymbol{x}} \right)^{\mathrm{T}} \dot{\boldsymbol{x}} + \left(\frac{\partial y}{\partial \boldsymbol{x}} \right)^{\mathrm{T}} \ddot{\boldsymbol{x}} \\ &= \dot{\boldsymbol{x}}^{\mathrm{T}} \left(\frac{\partial^2 y}{\partial \boldsymbol{x}^{\mathrm{T}} \partial \boldsymbol{x}} \right) \dot{\boldsymbol{x}} + \left(\frac{\partial y}{\partial \boldsymbol{x}} \right)^{\mathrm{T}} \ddot{\boldsymbol{x}}. \end{aligned}$$

In summary,

$$\begin{aligned} y &= y(\boldsymbol{x}), \\ \dot{y} &= \left(\frac{\partial y}{\partial \boldsymbol{x}} \right)^{\mathrm{T}} \dot{\boldsymbol{x}}, \\ \ddot{y} &= \left(\frac{\partial y}{\partial \boldsymbol{x}} \right)^{\mathrm{T}} \ddot{\boldsymbol{x}} + \dot{\boldsymbol{x}}^{\mathrm{T}} \left(\frac{\partial^2 y}{\partial \boldsymbol{x}^{\mathrm{T}} \partial \boldsymbol{x}} \right) \dot{\boldsymbol{x}}. \end{aligned}$$

The first-order time derivative $\dot{y}$ includes the first-order time derivative $\dot{\boldsymbol{x}}$. The second-order time derivative $\ddot{y}$ includes the second-order time derivative $\ddot{\boldsymbol{x}}$ as well as a quadratic form with respect to $\dot{\boldsymbol{x}}$. Gradient vector given in (9) and Hessian matrix given in (10) characterize the above time derivatives.