Matrix Differentiation

CS4243 Computer Vision and Pattern Recognition

Leow Wee Kheng

Department of Computer Science
School of Computing
National University of Singapore

Matrix Differentiation

This set of slides summarizes some commonly used matrix differentiation formulae.

In the following,

- $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}. \tag{1}$$

- $\mathbf{x}$ and $\mathbf{y}$ are $n \times 1$ column matrices (i.e., vectors)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \tag{2}$$

(1) $E = \mathbf{x}^\top \mathbf{x}$

$$E = \mathbf{x}^\top \mathbf{x} = \sum_i x_i^2 = x_1^2 + \cdots + x_n^2. \tag{3}$$

$$\frac{\partial E}{\partial x_i} = 2x_i. \tag{4}$$

So,

$$\frac{\partial E}{\partial \mathbf{x}} = \begin{bmatrix} \partial E/\partial x_1 \\ \vdots \\ \partial E/\partial x_n \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 2\,\mathbf{x}. \tag{5}$$

(2) $E = \mathbf{x}^\top \mathbf{y}$

$$E = \mathbf{x}^\top \mathbf{y} = x_1 y_1 + \cdots + x_n y_n. \tag{6}$$

$$\frac{\partial E}{\partial x_i} = y_i. \tag{7}$$

So,

$$\frac{\partial E}{\partial \mathbf{x}} = \mathbf{y}. \tag{8}$$

(3) $E = (\mathbf{x}^\top \mathbf{y})^2$

$$E = (\mathbf{x}^\top \mathbf{y})^2 = (x_1 y_1 + \cdots + x_n y_n)^2. \tag{9}$$

$$\frac{\partial E}{\partial x_i} = 2(x_1 y_1 + \cdots + x_n y_n)\, y_i. \tag{10}$$

So,

$$\frac{\partial E}{\partial \mathbf{x}} = \begin{bmatrix} \partial E / \partial x_1 \\ \vdots \\ \partial E / \partial x_n \end{bmatrix} = 2\, \mathbf{x}^\top \mathbf{y}\, \mathbf{y}. \tag{11}$$

(4) **Ax**

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n \end{bmatrix}. \tag{12}$$

Let

$$s_i = [\mathbf{A}\mathbf{x}]_i = a_{i1}x_1 + \cdots + a_{in}x_n = \sum_j a_{ij}x_j. \tag{13}$$

Then,

$$\frac{\partial s_i}{\partial x_j} = a_{ij}. \tag{14}$$

So,

$$\frac{\partial(\mathbf{A}\mathbf{x})}{\partial \mathbf{x}^\top} = \begin{bmatrix} \partial s_1/\partial x_1 & \cdots & \partial s_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial s_n/\partial x_1 & \cdots & \partial s_n/\partial x_n \end{bmatrix} = \mathbf{A}. \tag{15}$$

And,

$$\frac{\partial (\mathbf{A}\mathbf{x})^\top}{\partial \mathbf{x}} = \mathbf{A}^\top. \tag{16}$$

(5) $E = \|\mathbf{A}\mathbf{x}\|^2$

$$E = \|\mathbf{A}\mathbf{x}\|^2 = (\mathbf{A}\mathbf{x})^\top(\mathbf{A}\mathbf{x}) = \sum_k \left[\sum_j a_{kj} x_j\right]^2. \tag{17}$$

$$\frac{\partial E}{\partial x_i} = 2 \sum_k \sum_j a_{ki} a_{kj} x_j = 2 \sum_j \left[\sum_k a_{ki} a_{kj}\right] x_j. \tag{18}$$

That is,

$$\frac{\partial E}{\partial x_i} = 2 \sum_j \left[\mathbf{A}^\top \mathbf{A}\right]_{ij} x_j. \tag{19}$$

Thus,

$$\frac{\partial E}{\partial \mathbf{x}} = 2\mathbf{A}^\top \mathbf{A}\mathbf{x}. \tag{20}$$

(6) $E = \|\mathbf{A}\mathbf{x}\|^2$

$$E = \sum_i s_i^2 \tag{21}$$

where

$$s_i = \sum_j a_{ij} x_j. \tag{22}$$

$$\frac{\partial E}{\partial a_{ij}} = 2 s_i \frac{\partial s_i}{\partial a_{ij}} = 2 s_i x_j. \tag{23}$$

Thus,

$$\frac{\partial E}{\partial \mathbf{A}} = 2 \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} [x_1 \cdots x_n] = 2\mathbf{A}\mathbf{x}\,\mathbf{x}^\top. \tag{24}$$

(7) $E = \|\mathbf{B}\mathbf{A}\mathbf{x}\|^2$

Let $\mathbf{y} = \mathbf{A}\mathbf{x}$. Then,

$$E = \|\mathbf{B}\mathbf{y}\|^2. \tag{25}$$

$$\frac{\partial E}{\partial \mathbf{B}} = 2\mathbf{B}\mathbf{y}\mathbf{y}^\top = 2\mathbf{B}\mathbf{A}\mathbf{x}\,\mathbf{x}^\top\mathbf{A}^\top. \tag{26}$$

(8) $E = \|\mathbf{BAx}\|^2$

Let $\mathbf{C} = \mathbf{BA}$. Then,

$$E = \|\mathbf{Cx}\|^2. \tag{27}$$

$$\frac{\partial E}{\partial \mathbf{x}} = 2\mathbf{C}^\top\mathbf{Cx} = 2\mathbf{A}^\top\mathbf{B}^\top\mathbf{BAx}. \tag{28}$$

(9) $E = \|\mathbf{BAx}\|^2$

$$E = \|\mathbf{BAx}\|^2 = \sum_k \left[\sum_l \sum_m b_{kl} a_{lm} x_m \right]^2 . \tag{29}$$

$$\begin{aligned}
\frac{\partial E}{\partial a_{ij}} &= 2 \sum_k \left[\sum_l \sum_m b_{kl} a_{lm} x_m \right] b_{ki} x_j \\
&= 2 \sum_l \sum_m \left[\sum_k b_{ki} b_{kl} \right] a_{lm} x_m x_j \\
&= 2 \sum_l \sum_m \left[\mathbf{B}^\top \mathbf{B} \right]_{il} a_{lm} x_m x_j \\
&= 2 \left[\mathbf{B}^\top \mathbf{BAx} \right]_i x_j .
\end{aligned} \tag{30}$$

So,

$$\frac{\partial E}{\partial \mathbf{A}} = 2\mathbf{B}^\top \mathbf{BAxx}^\top . \tag{31}$$