Vector/Matrix Derivatives

Premmi

May 6, 2016

Proposition 1 Let

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1}$$

where the dimension of $\mathbf{y}$ is $\mathfrak{m} \times 1$, $\mathbf{A}$ is $\mathfrak{m} \times \mathfrak{n}$, $\mathbf{x}$ is $\mathfrak{n} \times 1$ and $\mathbf{A}$ is independent of $\mathbf{x}$, then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A} \tag{2}$$

Proof. Since the $\boldsymbol{i}$th element of $\mathbf{y}$ is given by

$$y_i = \sum_{k=1}^{n} a_{ik} x_k \tag{3}$$

it follows that

$$\frac{\partial y_i}{\partial x_j} = a_{ij} \tag{4}$$

for all $\mathfrak{i} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{m}$ and $\mathfrak{j} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{n}$. Hence

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{5}$$

$$\tag{6}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{7}$$

$$= \mathbf{A} \tag{8}$$

□

$\boxed{\textbf{Hence, if } \mathbf{y} = \mathbf{Ax} \textbf{ then } \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}.}$

Proposition 2 Let the scalar α be defined as

$$\alpha = \mathbf{y}^{\mathbf{T}}\mathbf{Ax} \tag{9}$$

where the dimension of $\mathbf{y}$ is $\mathfrak{m} \times 1$, $\mathbf{A}$ is $\mathfrak{m} \times \mathfrak{n}$, $\mathbf{x}$ is $\mathfrak{n} \times 1$ and $\mathbf{A}$ is independent of $\mathbf{x}$ and $\mathbf{y}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^{\mathbf{T}}\mathbf{A} \tag{10}$$

and

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x^T A^T} \tag{11}$$

Proof. Let us define

$$\mathbf{w^T} = \mathbf{y^T A} \tag{12}$$

and so

$$\alpha = \mathbf{w^T x} \tag{13}$$

Hence, by *Proposition 1* we have

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{w^T} = \mathbf{y^T A} \tag{14}$$

which proves the first result. Since α is a scalar,

$$\alpha = \alpha^T = \mathbf{x^T A^T y} \tag{15}$$

and applying *Proposition 1* again, we get

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x^T A^T} \tag{16}$$

which proves the second result. $\square$

$$\boxed{\textbf{Hence, if } \alpha = \mathbf{y^T A x} \textbf{ then } \frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y^T A} \textbf{ and } \frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x^T A^T}.}$$

Proposition 3 Let the scalar α be expressed as a quadratic form given by

$$\alpha = \mathbf{x^T A x} \tag{17}$$

where the dimension of $\mathbf{x}$ is $\mathfrak{n} \times 1$, $\mathbf{A}$ is $\mathfrak{n} \times \mathfrak{n}$ and $\mathbf{A}$ is independent of $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x^T} \left(\mathbf{A} + \mathbf{A^T}\right) \tag{18}$$

Proof. By definition

$$\alpha = \mathbf{x^T A x} \tag{19}$$

$$= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{20}$$

$$= \begin{bmatrix} a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n & \cdots & a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{21}$$

$$\begin{aligned} &= \left(a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n\right) x_1 \\ &\quad + \left(a_{12}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n\right) x_2 \\ &\quad + \left(a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{nn}x_n\right) x_n \end{aligned} \tag{22}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \tag{23}$$

Differentiating with respect to the $\boldsymbol{k}$th element of $\mathbf{x}$ we get

$$\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{jk} x_j + \sum_{i=1}^{n} a_{ki} x_i \tag{24}$$

for all $\mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{n}$ and consequently,

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x^T A} + \mathbf{x^T A^T} = \mathbf{x^T}\left(\mathbf{A} + \mathbf{A^T}\right) \tag{25}$$

Note:

Here $\frac{\partial \alpha}{\partial \mathbf{x}}$ is a **row vector**. If we want to write it as a **column vector** then

$$\left(\frac{\partial \alpha}{\partial \mathbf{x}}\right)^T = \left(\mathbf{A} + \mathbf{A^T}\right)\mathbf{x} \tag{26}$$

□

$$\boxed{\textbf{Hence, if } \alpha = \mathbf{x^T A x} \textbf{ then } \frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x^T}\left(\mathbf{A} + \mathbf{A^T}\right).}$$

Proposition 4 If $\mathbf{A}$ is a symmetric matrix and

$$\alpha = \mathbf{x^T A x} \tag{27}$$

where the dimension of $\mathbf{x}$ is $\mathfrak{n} \times 1$, $\mathbf{A}$ is $\mathfrak{n} \times \mathfrak{n}$ and $\mathbf{A}$ is independent of $\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x^T A} \tag{28}$$

and

$$\frac{\partial^2 \alpha}{\partial \mathbf{x}^2} = 2\mathbf{A} \tag{29}$$

Proof. If $\mathbf{A}$ is symmetric then $\mathbf{A} = \mathbf{A^T}$. So from *Proposition 3* it follows that

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x^T}\left(\mathbf{A} + \mathbf{A}\right) = 2\mathbf{x^T A} \tag{30}$$

which proves the first result.

Note:

Here $\frac{\partial \alpha}{\partial \mathbf{x}}$ is a **row vector**. If we want to write it as a **column vector** then

$$\left(\frac{\partial \alpha}{\partial \mathbf{x}}\right)^T = 2\mathbf{A}\mathbf{x} \tag{31}$$

The Hessian of the quadratic form is given by

$$\frac{\partial^2 \alpha}{\partial x_j x_k} = \frac{\partial}{\partial x_j}\left[\frac{\partial \alpha}{\partial x_k}\right] = \frac{\partial}{\partial x_j}\left[2\sum_{i=1}^{n} a_{ki}x_i\right] = 2a_{kj} = 2a_{jk} \tag{32}$$

for all $\mathbf{j} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{n}$ and $\mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{n}$. Hence

$$\frac{\partial^2 \alpha}{\partial \mathbf{x}^{\mathbf{T}} \partial \mathbf{x}} = 2\mathbf{A} \tag{33}$$

□

Hence, if A is symmetric and $\alpha = \mathbf{x}^{\mathbf{T}}\mathbf{A}\mathbf{x}$ then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}^{\mathbf{T}}\mathbf{A} \textbf{ and } \frac{\partial^2 \alpha}{\partial \mathbf{x}^{\mathbf{T}} \partial \mathbf{x}} = 2\mathbf{A}.$$

Proposition 5 Let

$$\mathbf{C} = \mathbf{A}\mathbf{B} \tag{34}$$

where the dimension of $\mathbf{A}$ is $\mathfrak{m} \times \mathfrak{n}$, $\mathbf{B}$ is $\mathfrak{n} \times \mathfrak{p}$, $\mathbf{C}$ is $\mathfrak{m} \times \mathfrak{p}$ and the elements of $\mathbf{A}$ and $\mathbf{B}$ are functions of the elements x_n of the vector $\mathbf{x}$ of dimension $\mathfrak{n} \times 1$. Then,

$$\frac{\partial \mathbf{C}}{\partial x_n} = \frac{\partial \mathbf{A}}{\partial x_n}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial x_n} \tag{35}$$

Proof. By definition, the *(m, p)*-th element of the matrix $\mathbf{C}$ is given by,

$$c_{mp} = \sum_{j=1}^{n} a_{mj}b_{jp} \tag{36}$$

Applying the product rule for differentiation to (36) yields,

$$\frac{\partial c_{mp}}{\partial x_n} = \sum_{j=1}^{n} \left(\frac{\partial a_{mj}}{\partial x_n} b_{jp} + a_{mj} \frac{\partial b_{jp}}{\partial x_n} \right) \tag{37}$$

for all $\mathbf{m} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{m}$ and $\mathbf{p} = \mathbf{1}, \mathbf{2}, \ldots, \mathfrak{p}$. Hence,

$$\frac{\partial \mathbf{C}}{\partial x_n} = \frac{\partial \mathbf{A}}{\partial x_n}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial x_n} \tag{38}$$

□

$$\boxed{\textbf{Hence, if } \mathbf{C} = \mathbf{AB} \textbf{ then, } \frac{\partial \mathbf{C}}{\partial x_n} = \frac{\partial \mathbf{A}}{\partial x_n}\mathbf{B} + \mathbf{A}\frac{\partial \mathbf{B}}{\partial x_n}\textbf{.}}$$