Matrix Differentiation

CS5240 Theoretical Foundations in Multimedia

Leow Wee Kheng

Department of Computer Science
School of Computing
National University of Singapore

Matrix Derivatives

There are 6 common types of matrix derivatives:

Type	Scalar	Vector	Matrix
Scalar	$\dfrac{\partial y}{\partial x}$	$\dfrac{\partial \mathbf{y}}{\partial x}$	$\dfrac{\partial \mathbf{Y}}{\partial x}$
Vector	$\dfrac{\partial y}{\partial \mathbf{x}}$	$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$	
Matrix	$\dfrac{\partial y}{\partial \mathbf{X}}$		

Derivatives by Scalar

Numerator Layout Notation

$$\frac{\partial y}{\partial x}$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}$$

Denominator Layout Notation

$$\frac{\partial y}{\partial x}$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix} \equiv \frac{\partial \mathbf{y}^\top}{\partial x}$$

Derivatives by Vector

Numerator Layout Notation

$$\frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1} \quad \cdots \quad \frac{\partial y}{\partial x_n} \right]$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y_m}{\partial x_1} & \cdots & \dfrac{\partial y_m}{\partial x_n} \end{bmatrix}$$

$$\equiv \frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top}$$

Denominator Layout Notation

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_1} \\ \vdots \\ \dfrac{\partial y}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y_1}{\partial x_n} & \cdots & \dfrac{\partial y_m}{\partial x_n} \end{bmatrix}$$

$$\equiv \frac{\partial \mathbf{y}^\top}{\partial \mathbf{x}}$$

Derivative by Matrix

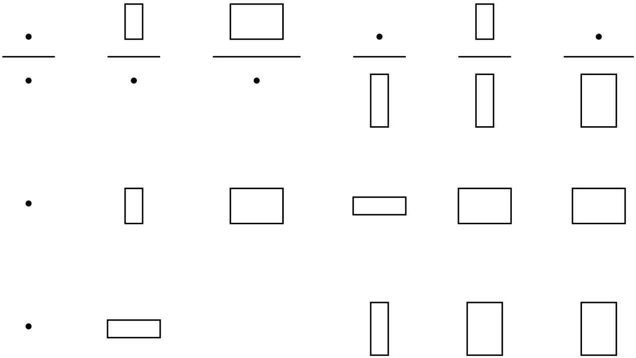
Numerator Layout Notation

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1n}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} \equiv \frac{\partial y}{\partial \mathbf{X}^\top}$$

Denominator Layout Notation

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} \equiv \frac{\partial y}{\partial \mathbf{X}}$$

Pictorial Representation



numerator layout

denominator layout

Caution

- Most books and papers don't state which convention they use.
- Reference [2] uses both conventions but clearly differentiate them.

$$\frac{\partial y}{\partial \mathbf{x}^\top} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix} \qquad \frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}^\top} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \qquad \frac{\partial \mathbf{y}^\top}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

- It is best not to mix the two conventions in your equations.
- We adopt **numerator layout** notation.

Commonly Used Derivatives

Here, scalar a, vector $\mathbf{a}$ and matrix $\mathbf{A}$ are not functions of x and $\mathbf{x}$.

(C1) $$\frac{d\mathbf{a}}{dx} = \mathbf{0} \quad \text{(column matrix)}$$

(C2) $$\frac{da}{d\mathbf{x}} = \mathbf{0}^\top \quad \text{(row matrix)}$$

(C3) $$\frac{da}{d\mathbf{X}} = \mathbf{0}^\top \quad \text{(matrix)}$$

(C4) $$\frac{d\mathbf{a}}{d\mathbf{x}} = \mathbf{0} \quad \text{(matrix)}$$

(C5) $$\frac{d\mathbf{x}}{d\mathbf{x}} = \mathbf{I}$$

(C6) $$\frac{d\,\mathbf{a}^\top\mathbf{x}}{d\mathbf{x}} = \frac{d\,\mathbf{x}^\top\mathbf{a}}{d\mathbf{x}} = \mathbf{a}^\top$$

(C7) $$\frac{d\mathbf{x}^\top\mathbf{x}}{d\mathbf{x}} = 2\,\mathbf{x}^\top$$

(C8) $$\frac{d(\mathbf{x}^\top\mathbf{a})^2}{d\mathbf{x}} = 2\,\mathbf{x}^\top\mathbf{a}\,\mathbf{a}^\top$$

(C9) $$\frac{d\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{A}$$

(C10) $$\frac{d\,\mathbf{x}^\top\mathbf{A}}{d\mathbf{x}} = \mathbf{A}^\top$$

(C11) $$\frac{d\,\mathbf{x}^\top\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)$$

Math Notation

We represent a vector $\mathbf{x}$ as a column matrix

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}.$$

Its transpose $\mathbf{x}^\top$ is a row matrix

$$\mathbf{x}^\top = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}.$$

Consider two vectors $\mathbf{x}$ and $\mathbf{y}$ with the same number of components. Their inner product $\mathbf{x}^\top\mathbf{y}$ is actually a 1×1 matrix:

$$\mathbf{x}^\top\mathbf{y} = [\, s \,]$$

where

$$s = \sum_{i=1}^{m} x_i y_i.$$

For notational inconvenience, we usually drop the matrix and regard the inner product as a scalar, i.e.,

$$\mathbf{x}^\top\mathbf{y} = \sum_{i=1}^{m} x_i y_i.$$

Derivatives of Scalar by Scalar

(SS1) $$\frac{\partial(u+v)}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}$$

(SS2) $$\frac{\partial uv}{\partial x} = u\frac{\partial v}{\partial x} + v\frac{\partial u}{\partial x}$$ (product rule)

(SS3) $$\frac{\partial g(u)}{\partial x} = \frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial x}$$ (chain rule)

(SS4) $$\frac{\partial f(g(u))}{\partial x} = \frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial x}$$ (chain rule)

Derivatives of Vector by Scalar

(VS1) $$\frac{\partial a\mathbf{u}}{\partial x} = a\frac{\partial \mathbf{u}}{\partial x}$$

where a is not a function of x.

(VS2) $$\frac{\partial \mathbf{A}\mathbf{u}}{\partial x} = \mathbf{A}\frac{\partial \mathbf{u}}{\partial x}$$

where $\mathbf{A}$ is not a function of x.

(VS3) $$\frac{\partial \mathbf{u}^\top}{\partial x} = \left(\frac{\partial \mathbf{u}}{\partial x}\right)^\top$$

(VS4) $$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial x} = \frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{v}}{\partial x}$$

(VS5) $$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial x} = \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial x} \qquad \text{(chain rule)}$$

with consistent matrix layout.

(VS6) $$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} = \frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}} \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial x} \qquad \text{(chain rule)}$$

with consistent matrix layout.

Derivatives of Matrix by Scalar

(MS1) $$\frac{\partial a\mathbf{U}}{\partial x} = a\frac{\partial \mathbf{U}}{\partial x}$$

where a is not a function of x.

(MS2) $$\frac{\partial \mathbf{AUB}}{\partial x} = \mathbf{A}\frac{\partial \mathbf{U}}{\partial x}\mathbf{B}$$

where $\mathbf{A}$ and $\mathbf{B}$ are not functions of x.

(MS3) $$\frac{\partial (\mathbf{U}+\mathbf{V})}{\partial x} = \frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{V}}{\partial x}$$

(MS4) $$\frac{\partial \mathbf{UV}}{\partial x} = \mathbf{U}\frac{\partial \mathbf{V}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x}\mathbf{V} \quad \text{(product rule)}$$

Derivatives of Scalar by Vector

(SV1) $$\frac{\partial au}{\partial \mathbf{x}} = a\frac{\partial u}{\partial \mathbf{x}}$$

where a is not a function of $\mathbf{x}$.

(SV2) $$\frac{\partial (u+v)}{\partial \mathbf{x}} = \frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}$$

(SV3) $$\frac{\partial uv}{\partial \mathbf{x}} = u\frac{\partial v}{\partial \mathbf{x}} + v\frac{\partial u}{\partial \mathbf{x}} \qquad \text{(product rule)}$$

(SV4) $$\frac{\partial g(u)}{\partial \mathbf{x}} = \frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}} \qquad \text{(chain rule)}$$

(SV5) $$\frac{\partial f(g(u))}{\partial \mathbf{x}} = \frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{x}} \qquad \text{(chain rule)}$$

(SV6) $$\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial \mathbf{x}} = \mathbf{u}^\top \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \qquad \text{(product rule)}$$

where $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ and $\dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ are in numerator layout.

(SV7) $$\frac{\partial \mathbf{u}^\top \mathbf{A} \mathbf{v}}{\partial \mathbf{x}} = \mathbf{u}^\top \mathbf{A} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^\top \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \qquad \text{(product rule)}$$

where $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ and $\dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ are in numerator layout,

and $\mathbf{A}$ is not a function of $\mathbf{x}$.

Derivatives of Scalar by Matrix

(SM1) $$\frac{\partial au}{\partial \mathbf{X}} = a\frac{\partial u}{\partial \mathbf{X}}$$

where a is not a function of $\mathbf{X}$.

(SM2) $$\frac{\partial (u+v)}{\partial \mathbf{X}} = \frac{\partial u}{\partial \mathbf{X}} + \frac{\partial v}{\partial \mathbf{X}}$$

(SM3) $$\frac{\partial uv}{\partial \mathbf{X}} = u\frac{\partial v}{\partial \mathbf{X}} + v\frac{\partial u}{\partial \mathbf{X}} \qquad \text{(product rule)}$$

(SM4) $$\frac{\partial g(u)}{\partial \mathbf{X}} = \frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{X}} \qquad \text{(chain rule)}$$

(SM5) $$\frac{\partial f(g(u))}{\partial \mathbf{X}} = \frac{\partial f(g)}{\partial g}\frac{\partial g(u)}{\partial u}\frac{\partial u}{\partial \mathbf{X}} \qquad \text{(chain rule)}$$

Derivatives of Vector by Vector

(VV1) $$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} = a\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial a}{\partial \mathbf{x}}$$ (product rule)

(VV2) $$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} = \mathbf{A}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$$

where $\mathbf{A}$ is not a function of $\mathbf{x}$.

(VV3) $$\frac{\partial(\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$$

(VV4) $$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$$ (chain rule)

(VV5) $$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} = \frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$$ (chain rule)

Notes on Denominator Layout

In some cases, the results of denominator layout are the transpose of those of numerator layout. Moreover, the chain rule for denominator layout goes from right to left instead of left to right.

	Numerator Layout Notation	Denominator Layout Notation
(C7)	$\frac{d\mathbf{a}^\top\mathbf{x}}{d\mathbf{x}} = \mathbf{a}^\top$	$\frac{d\mathbf{a}^\top\mathbf{x}}{d\mathbf{x}} = \mathbf{a}$
(C11)	$\frac{d\mathbf{x}^\top\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{x}^\top(\mathbf{A}+\mathbf{A}^\top)$	$\frac{d\mathbf{x}^\top\mathbf{A}\mathbf{x}}{d\mathbf{x}} = (\mathbf{A}+\mathbf{A}^\top)\mathbf{x}$
(VV5)	$\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial\mathbf{x}} = \frac{\partial\mathbf{f}(\mathbf{g})}{\partial\mathbf{g}}\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial\mathbf{x}}$	$\frac{\partial\mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial\mathbf{x}} = \frac{\partial\mathbf{u}}{\partial\mathbf{x}}\frac{\partial\mathbf{g}(\mathbf{u})}{\partial\mathbf{u}}\frac{\partial\mathbf{f}(\mathbf{g})}{\partial\mathbf{g}}$

Derivations of Derivatives

(C6) $$\frac{d\,\mathbf{a}^\top\mathbf{x}}{d\mathbf{x}} = \frac{d\,\mathbf{x}^\top\mathbf{a}}{d\mathbf{x}} = \mathbf{a}^\top$$

(The not-so-hard way)

Let $s = \mathbf{a}^\top\mathbf{x} = a_1x_1 + \cdots + a_nx_n$. Then, $\dfrac{\partial s}{\partial x_i} = a_i$. So, $\dfrac{ds}{d\mathbf{x}} = \mathbf{a}^\top$.

(The easier way)

Let $s = \mathbf{a}^\top\mathbf{x} = \displaystyle\sum_i a_ix_i$. Then, $\dfrac{\partial s}{\partial x_i} = a_i$. So, $\dfrac{ds}{d\mathbf{x}} = \mathbf{a}^\top$.

(C7) $$\frac{d\mathbf{x}^\top\mathbf{x}}{d\mathbf{x}} = 2\,\mathbf{x}^\top$$

Let $s = \mathbf{x}^\top\mathbf{x} = \displaystyle\sum_i x_i^2$. Then, $\dfrac{\partial s}{\partial x_i} = 2x_i$. So, $\dfrac{ds}{d\mathbf{x}} = 2\,\mathbf{x}^\top$.

(C8) $$\frac{d(\mathbf{x}^\top \mathbf{a})^2}{d\mathbf{x}} = 2\,\mathbf{x}^\top \mathbf{a}\,\mathbf{a}^\top$$

Let $s = \mathbf{x}^\top \mathbf{a}$. Then, $\dfrac{\partial s^2}{\partial x_i} = 2s\dfrac{\partial s}{\partial x_i} = 2\,s\,a_i$. So, $\dfrac{ds^2}{d\mathbf{x}} = 2\,\mathbf{x}^\top \mathbf{a}\,\mathbf{a}^\top$.

(C9) $$\frac{d\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{A}$$

(The hard way)

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

(The easy way)

Let $\mathbf{s} = \mathbf{A}\mathbf{x}$. Then, $s_i = \displaystyle\sum_j a_{ij}x_j$, and $\dfrac{\partial s_i}{\partial x_j} = a_{ij}$. So, $\dfrac{d\mathbf{s}}{d\mathbf{x}} = \mathbf{A}$.

(C10) $$\frac{d\,\mathbf{x}^\top \mathbf{A}}{d\mathbf{x}} = \mathbf{A}^\top$$

Let $\mathbf{y}^\top = \mathbf{x}^\top \mathbf{A}$, and $\mathbf{a}_j$ denote the j-th column of $\mathbf{A}$. Then, $y_i = \mathbf{x}^\top \mathbf{a}_j$.

Applying (C6) yields $\dfrac{dy_i}{d\mathbf{x}} = \mathbf{a}_j^\top$. So, $\dfrac{d\mathbf{y}^\top}{d\mathbf{x}} = \mathbf{A}^\top$.

(C11) $$\frac{d\,\mathbf{x}^\top \mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)$$

Apply (SV6) to $\dfrac{d\,\mathbf{x}^\top \mathbf{A}\mathbf{x}}{d\mathbf{x}}$ and obtain $\mathbf{x}^\top \dfrac{d\mathbf{A}\mathbf{x}}{d\mathbf{x}} + (\mathbf{A}\mathbf{x})^\top \dfrac{d\mathbf{x}}{d\mathbf{x}}$,

Next, apply (C9) to the first part of the sum, and obtain

$\mathbf{x}^\top \mathbf{A} + (\mathbf{A}\mathbf{x})^\top$, which is $\mathbf{x}^\top(\mathbf{A} + \mathbf{A}^\top)$.

(Need to prove SV6—Homework.)

Derivatives of Trace

For variable matrix $\mathbf{X}$ and constant matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$,

(DT1) $$\frac{\partial\, \mathrm{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}$$

(DT2) $$\frac{\partial\, \mathrm{tr}(\mathbf{X}^k)}{\partial \mathbf{X}} = k\mathbf{X}^{k-1}$$

(DT3) $$\frac{\partial\, \mathrm{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial\, \mathrm{tr}(\mathbf{X}\mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}$$

(DT4) $$\frac{\partial\, \mathrm{tr}(\mathbf{A}\mathbf{X}^\top)}{\partial \mathbf{X}} = \frac{\partial\, \mathrm{tr}(\mathbf{X}^\top\mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}^\top$$

(DT5) $$\frac{\partial\, \mathrm{tr}(\mathbf{X}^\top\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{X}^\top(\mathbf{A} + \mathbf{A}^\top)$$

(DT6) $$\frac{\partial \operatorname{tr}(\mathbf{X}^{-1}\mathbf{A})}{\partial \mathbf{X}} = -\mathbf{X}^{-\top}\mathbf{A}\mathbf{X}^{-\top}$$

(DT7) $$\frac{\partial \operatorname{tr}(\mathbf{AXB})}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{BAX})}{\partial \mathbf{X}} = \mathbf{BA}$$

(DT8) $$\frac{\partial \operatorname{tr}(\mathbf{AXBX}^{\top}\mathbf{C})}{\partial \mathbf{X}} = \mathbf{BX}^{\top}\mathbf{CA} + \mathbf{B}^{\top}\mathbf{X}^{\top}\mathbf{A}^{\top}\mathbf{C}^{\top}$$

Derivatives of Determinant

For variable matrix $\mathbf{X}$ and constant matrices $\mathbf{A}$, $\mathbf{B}$,

(DD1) $$\frac{\partial\,|\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\,\mathbf{X}^{-1}$$

(DD2) $$\frac{\partial|\mathbf{X}^k|}{\partial \mathbf{X}} = k|\mathbf{X}^k|\mathbf{X}^{-1}$$

(DD3) $$\frac{\partial \log|\mathbf{X}|}{\partial \mathbf{X}} = \mathbf{X}^{-1}$$

(DD4) $$\frac{\partial|\mathbf{AXB}|}{\partial \mathbf{X}} = |\mathbf{AXB}|\mathbf{X}^{-1}$$

(DD5) $$\frac{\partial|\mathbf{X}^\top\mathbf{AX}|}{\partial \mathbf{X}} = 2|\mathbf{X}^\top\mathbf{AX}|\mathbf{X}^{-1}$$

Derivations of Derivatives

(DD1) $\quad \dfrac{\partial\,|\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\,\mathbf{X}^{-1}$

For an $n \times n$ matrix $\mathbf{X}$, Laplace's formula expresses $|\mathbf{X}|$ as

$$|\mathbf{X}| = \sum_{i=1}^{n} (-1)^{i+j} x_{ij} M_{ij},$$

where M_{ij} is a minor, which is the determinant of the sub-matrix of $\mathbf{X}$ obtained by removing the ith row and jth column.

The adjugate $\mathrm{adj}(\mathbf{X})$ is the transpose of the matrix consisting of the cofactors $(-1)^{i+j} M_{ij}$:

$$\mathrm{adj}(\mathbf{X})_{ji} = (-1)^{i+j} M_{ij},$$

which is independent of x_{ij}, element (i, j) of $\mathbf{X}$.

So,

$$|\mathbf{X}| = \sum_{i=1}^{n} x_{ij} \mathrm{adj}(\mathbf{X})_{ji}.$$

Then,

$$\frac{\partial |\mathbf{X}|}{\partial x_{ij}} = \mathrm{adj}(\mathbf{X})_{ji},$$

and, in numerator layout,

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \mathrm{adj}(\mathbf{X}).$$

$|\mathbf{X}|$ and $\mathrm{adj}(\mathbf{X})$ are also related by

$$|\mathbf{X}|\, \mathbf{I} = \mathrm{adj}(\mathbf{X})\, \mathbf{X}.$$

So,

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\, \mathbf{X}^{-1}.$$

(DD2) $$\frac{\partial|\mathbf{X}^k|}{\partial\mathbf{X}} = k|\mathbf{X}^k|\mathbf{X}^{-1}$$

Note that $|\mathbf{X}^k| = |\mathbf{X}|^k$. So,

$$\frac{\partial|\mathbf{X}^k|}{\partial\mathbf{X}} = \frac{\partial|\mathbf{X}|^k}{\partial\mathbf{X}} = k|\mathbf{X}|^{k-1}|\mathbf{X}|\mathbf{X}^{-1} = k|\mathbf{X}^k|\mathbf{X}^{-1}.$$

(DD3) $$\frac{\partial\log|\mathbf{X}|}{\partial\mathbf{X}} = \mathbf{X}^{-1}$$

$$\frac{\partial\log|\mathbf{X}|}{\partial\mathbf{X}} = \frac{1}{|\mathbf{X}|}\frac{\partial|\mathbf{X}|}{\partial\mathbf{X}} = \mathbf{X}^{-1}.$$

(DD4) $$\frac{\partial|\mathbf{AXB}|}{\partial\mathbf{X}} = |\mathbf{AXB}|\mathbf{X}^{-1}$$

Note that $|\mathbf{AXB}| = |\mathbf{A}||\mathbf{X}||\mathbf{B}|$. So,

$$\frac{\partial|\mathbf{AXB}|}{\partial\mathbf{X}} = |\mathbf{A}||\mathbf{X}|\mathbf{X}^{-1}|\mathbf{B}| = |\mathbf{AXB}|\mathbf{X}^{-1}.$$

$$\text{(DD5)} \qquad \frac{\partial|\mathbf{X}^\top\mathbf{A}\mathbf{X}|}{\partial\mathbf{X}} = 2|\mathbf{X}^\top\mathbf{A}\mathbf{X}|\mathbf{X}^{-1}$$

Note that $|\mathbf{X}^\top| = |\mathbf{X}|$. So,

$$\frac{\partial|\mathbf{X}^\top\mathbf{A}\mathbf{X}|}{\partial\mathbf{X}} = |\mathbf{X}|\mathbf{X}^{-1}|\mathbf{A}||\mathbf{X}| + |\mathbf{X}^\top||\mathbf{A}||\mathbf{X}|\mathbf{X}^{-1} = 2|\mathbf{X}^\top\mathbf{A}\mathbf{X}|\mathbf{X}^{-1}.$$

(DT1) $$\frac{\partial\,\mathrm{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}$$

For an $n \times n$ matrix $\mathbf{X}$,

$$\mathrm{tr}(\mathbf{X}) = \sum_{i=1}^{n} x_{ii}.$$

Then,

$$\frac{\partial\,\mathrm{tr}(\mathbf{X})}{\partial x_{ij}} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

So,

$$\frac{\partial\,\mathrm{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}.$$

(DT2) $$\frac{\partial\,\mathrm{tr}(\mathbf{X}^k)}{\partial \mathbf{X}} = k\mathbf{X}^{k-1}$$

$$\begin{aligned}
(\mathbf{X}^k)_{ij} &= \sum_{l_1} \cdots \sum_{l_{k-1}} x_{il_1} x_{l_1 l_2} \cdots x_{l_{k-1}j} \\
\mathrm{tr}(\mathbf{X}^k) &= \sum_{l_0} \sum_{l_1} \cdots \sum_{l_{k-1}} x_{l_0 l_1} x_{l_1 l_2} \cdots x_{l_{k-1} l_0}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\,\mathrm{tr}(\mathbf{X}^k)}{\partial x_{ij}} = \; & \sum_{l_2} \cdots \sum_{l_{k-1}} \overbrace{x_{jl_2} x_{l_2 l_3} \cdots x_{l_{k-1}i}}^{k-1 \text{ factors}} + && (l_0 = i, l_1 = j) \\
& \cdots + && (k-2 \text{ terms}) \\
& \sum_{l} \cdots \sum_{l_{k-2}} x_{jl_1} x_{l_1 l_2} \cdots x_{l_{k-2}i} && (l_{k-1} = i, l_0 = j)
\end{aligned}$$

So, in numerator layout,

$$\frac{\partial\,\mathrm{tr}(\mathbf{X}^k)}{\partial \mathbf{X}} = k\mathbf{X}^{k-1}.$$

(DT3) $$\frac{\partial\,\mathrm{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \frac{\partial\,\mathrm{tr}(\mathbf{XA})}{\partial \mathbf{X}} = \mathbf{A}$$

$$\begin{aligned}
(\mathbf{AX})_{ij} &= \sum_k a_{ik} x_{kj} \\
\mathrm{tr}(\mathbf{AX}) &= \sum_l \sum_k a_{lk} x_{kl} \\
\frac{\partial\,\mathrm{tr}(\mathbf{AX})}{\partial x_{ij}} &= a_{ji},
\end{aligned}$$

So, in numerator layout,

$$\frac{\partial\,\mathrm{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \mathbf{A}.$$

Caution: $\mathrm{tr}(\mathbf{AB}) \neq \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})$.

$$\text{(DT4)} \qquad \frac{\partial \operatorname{tr}(\mathbf{A}\mathbf{X}^\top)}{\partial \mathbf{X}} = \frac{\partial \operatorname{tr}(\mathbf{X}^\top \mathbf{A})}{\partial \mathbf{X}} = \mathbf{A}^\top$$

Similar to the derivation for DT3,

$$\begin{aligned} (\mathbf{A}\mathbf{X}^\top)_{ij} &= \sum_k a_{ik} x_{jk} \\ \operatorname{tr}(\mathbf{A}\mathbf{X}) &= \sum_l \sum_k a_{lk} x_{lk} \\ \frac{\partial \operatorname{tr}(\mathbf{A}\mathbf{X}^\top)}{\partial x_{ij}} &= a_{ij} \end{aligned}$$

So, in numerator layout,

$$\frac{\partial \operatorname{tr}(\mathbf{A}\mathbf{X}^\top)}{\partial \mathbf{X}} = \mathbf{A}^\top.$$

(DT5) $$\frac{\partial \operatorname{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{X}^\top(\mathbf{A} + \mathbf{A}^\top)$$

$$\begin{aligned}
(\mathbf{X}^\top \mathbf{A}\mathbf{X})_{ij} &= \sum_k \sum_l x_{ki} a_{kl} x_{lj} \\
\operatorname{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X}) &= \sum_r \sum_k \sum_l x_{kr} a_{kl} x_{lr} \\
\frac{\partial \operatorname{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X})}{\partial x_{ij}} &= \sum_l a_{il} x_{lj} + \sum_k x_{kj} a_{ki} \\
&= \sum_k x_{kj} a_{ki} + \sum_l x_{lj} a_{il}.
\end{aligned}$$

So, in numerator layout,

$$\frac{\partial \operatorname{tr}(\mathbf{X}^\top \mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{X}^\top \mathbf{A} + \mathbf{X}^\top \mathbf{A}^\top = \mathbf{X}^\top(\mathbf{A} + \mathbf{A}^\top).$$

(DT7) DT7 is a direct result of DT3.

(DT8) $$\frac{\partial\,\mathrm{tr}(\mathbf{AXBX}^\top\mathbf{C})}{\partial\mathbf{X}} = \mathbf{BX}^\top\mathbf{CA} + \mathbf{B}^\top\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}^\top$$

$$\begin{aligned}
(\mathbf{AXBX}^\top\mathbf{C})_{ij} &= \sum_p\sum_q\sum_r\sum_s a_{ip}x_{pq}b_{qr}x_{sr}c_{sj} \\
\mathrm{tr}(\mathbf{AXBX}^\top\mathbf{C}) &= \sum_k\sum_p\sum_q\sum_r\sum_s a_{kp}x_{pq}b_{qr}x_{sr}c_{sk}.
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\,\mathrm{tr}(\mathbf{AXBX}^\top\mathbf{C})}{\partial x_{ij}} &= \sum_k\sum_r\sum_s a_{ki}b_{jr}x_{sr}c_{sk} + \sum_k\sum_p\sum_q a_{kp}x_{pq}b_{qj}c_{ik} \\
&= \sum_k\sum_r\sum_s b_{jr}x_{sr}c_{sk}a_{ki} + \sum_k\sum_p\sum_q b_{qj}x_{pq}a_{kp}c_{ik}
\end{aligned}$$

So, in numerator layout,

$$\frac{\partial\,\mathrm{tr}(\mathbf{AXBX}^\top\mathbf{C})}{\partial\mathbf{X}} = \mathbf{BX}^\top\mathbf{CA} + \mathbf{B}^\top\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}^\top.$$

References

1. J. E. Gentle, *Matrix Algebra: Theory, Computations, and Applications in Statistics*, Springer, 2007.
2. H. Lütkepohl, *Handbook of Matrices*, John Wiley & Sons, 1996.
3. K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, 2012. `www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf`
4. Wikipedia, *Matrix Calculus*. `en.wikipedia.org/wiki/Matrix_calculus`