Matrix derivatives, Single entry matrix and derivatives of $\boldsymbol{X}, \boldsymbol{X}^\top\boldsymbol{X}, \det\boldsymbol{X}, \mathrm{logdet}\boldsymbol{X}$ and $\mathrm{logdet}\boldsymbol{X}^\top\boldsymbol{X}$

Andersen Ang

ECS, Uni. Southampton, UK

andersen.ang@soton.ac.uk

Homepage angms.science

Version: January 20, 2024
First draft: July 14, 2017

Content

$\dfrac{\partial \boldsymbol{Y}}{\partial x}$ and $\dfrac{\partial y}{\partial \boldsymbol{X}}$

Single entry matrix $\boldsymbol{J}^{ij}$ as column & row selector

$$\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot \boldsymbol{X}^\top$$

$$\frac{\partial \mathrm{logdet} \boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{X}^\top$$

$$\frac{\partial \mathrm{logdet} \boldsymbol{X}^\top \boldsymbol{X}}{\partial \boldsymbol{X}} = 2\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

Derivatives between scalar and matrix

- **(Derivative of a matrix wrt. scalar)** For a matrix $\boldsymbol{Y} \in \mathbb{R}^{m\times n}$ and a scalar variable $x \in \mathbb{R}$

$$\frac{\partial \boldsymbol{Y}}{\partial x} = \frac{\partial}{\partial x}\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{pmatrix}. \tag{1}$$

- **(Derivative of a scalar wrt. matrix)** For a scalar $y \in \mathbb{R}$ and a matrix $X \in \mathbb{R}^{m\times n}$

$$\frac{\partial y}{\partial \boldsymbol{X}} = \frac{\partial y}{\partial \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}} = \begin{pmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{pmatrix}. \tag{2}$$

- What about vector: vector is just a special case of matrix.

Single Entry matrix $\boldsymbol{J}^{ij}$

- $\boldsymbol{X} \in \mathbb{R}^{m \times n}$, what is $\dfrac{\partial \boldsymbol{X}}{\partial \boldsymbol{X}}$?

- For simplicity, consider $\dfrac{\partial \boldsymbol{X}}{\partial X_{ij}}$, where X_{ij} is the ith-row jth-column element of $\boldsymbol{X}$, which is a scalar.

- Now by (1), for $(i, j) = (1, 1)$:

$$\frac{\partial \boldsymbol{X}}{\partial X_{11}} = \begin{pmatrix} \frac{\partial X_{11}}{\partial X_{11}} & \frac{\partial X_{12}}{\partial X_{11}} & \dots & \frac{\partial X_{1n}}{\partial X_{11}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial X_{m1}}{\partial X_{11}} & \frac{\partial X_{m2}}{\partial X_{11}} & \dots & \frac{\partial X_{mn}}{\partial X_{11}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

  which is a zero matrix except a one at the position $(1, 1)$.

- Matrix $\boldsymbol{J}^{ij}$ denotes the single entry matrix with all zeros except a one in the position (i, j).

Single entry matrix $\boldsymbol{J}^{ij}$ as column & row selector

- Right-multiplication with $\boldsymbol{J}^{ij}$ = put ith column to the jth column.

$$\boldsymbol{A}\boldsymbol{J}^{21} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} b_1 & 0 & 0 \\ b_2 & 0 & 0 \\ b_3 & 0 & 0 \end{pmatrix}$$

  If $i = j$, right-multiplication with $\boldsymbol{J}^{ii}$ = keep ith column.

- Left-multiplication with $\boldsymbol{J}^{ij}$ = put jth row to ith row.

$$\boldsymbol{J}^{31}\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ a_1 & b_1 & c_1 \end{pmatrix}$$

  If $i = j$, left-multiplication with $\boldsymbol{J}^{ii}$ = keep ith row.

- Reference: Matrix Cookbook (444) and (445)

Matrix derivative using single entry matrix

- Using single entry matrix gives

$$\frac{\partial \boldsymbol{X}}{\partial X_{ij}} = \boldsymbol{J}^{ij} \tag{3}$$

$$\frac{\partial \boldsymbol{X}^\top}{\partial X_{ij}} = \boldsymbol{J}^{ji} \tag{4}$$

$$\frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial X_{ij}} = \boldsymbol{X}^\top \boldsymbol{J}^{ij} + \boldsymbol{J}^{ji} \boldsymbol{X} \tag{5}$$

These are derivatives of matrix wrt. scalar X_{ij}, not matrix $\boldsymbol{X}$.

- Reference: Matrix Cookbook (73), (80), (456), (457), (458)

Illustrate $\frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial X_{ij}} = \boldsymbol{X}^\top \boldsymbol{J}^{ij} + \boldsymbol{J}^{ji} \boldsymbol{X}$ on $\boldsymbol{X} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

$$= \begin{pmatrix} a_1^2 + a_2^2 + a_3^2 & a_1 b_1 + a_2 b_2 + a_3 b_3 & a_1 c_1 + a_2 c_2 + a_3 c_3 \\ b_1 a_1 + b_2 a_2 + b_3 a_3 & b_1^2 + b_2^2 + b_3^2 & b_1 c_1 + b_2 c_2 + b_3 c_3 \\ c_1 a_1 + c_2 a_2 + c_3 a_3 & c_1 b_1 + c_2 b_2 + c_3 b_3 & c_1^2 + c_2^2 + c_3^2 \end{pmatrix}$$

$$\frac{\partial \boldsymbol{X}^\top \boldsymbol{X}}{\partial a_3} = \begin{pmatrix} 2a_3 & b_3 & c_3 \\ b_3 & 0 & 0 \\ c_3 & c_3 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} a_3 & 0 & 0 \\ b_3 & 0 & 0 \\ c_3 & 0 & 0 \end{pmatrix} + \begin{pmatrix} a_3 & b_3 & c_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \boldsymbol{X}^\top \boldsymbol{J}^{31} + \boldsymbol{J}^{13} \boldsymbol{X}$$

Single entry matrix with trace

- For $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{J} \in \mathbb{R}^{m\times n}$

$$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{J}^{ij}) \;=\; \mathrm{Tr}(\boldsymbol{J}^{ji}\boldsymbol{A}) \;=\; (\boldsymbol{A}^\top)_{ij} \tag{6}$$

- For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ $\boldsymbol{J} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$

$$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{J}^{ij}\boldsymbol{B}) = (\boldsymbol{A}^\top\boldsymbol{B}^\top)_{ij} \tag{7}$$

$$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{J}^{ji}\boldsymbol{B}) = (\boldsymbol{B}\boldsymbol{A})_{ij} \tag{8}$$

- Illustration of (6): $\boldsymbol{X} = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$, then $\mathrm{Tr}(\boldsymbol{X}\boldsymbol{J}^{23}) = \mathrm{Tr}\begin{pmatrix} 0 & 0 & a_2 \\ 0 & 0 & b_2 \\ 0 & 0 & c_2 \end{pmatrix} = c_2 = (\boldsymbol{X}^\top)_{23}$

- Reference : Matrix Cookbook equations (450-452)

Application of $\boldsymbol{J}^{ij}$ in deriving matrix derivatives

- Jacobi's formula relates the derivative of determinant of a matrix to the derivative of the matrix

$$\frac{\partial \det \boldsymbol{X}}{\partial x} = \det \boldsymbol{X} \cdot \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial x}\Big). \tag{J}$$

  Note that

  - $\det \boldsymbol{X}$, x and $\det \boldsymbol{X} \cdot \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial x}\Big)$ are all scalars.
  - $m = n$ here for $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ since det only works on square matrices

- To find $\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}}$, use (2) gives

$$\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} \overset{(2)}{=} \begin{bmatrix} \frac{\partial \det \boldsymbol{X}}{\partial X_{11}} & \dots & \frac{\partial \det \boldsymbol{X}}{\partial X_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \det \boldsymbol{X}}{\partial X_{n1}} & \dots & \frac{\partial \det \boldsymbol{X}}{\partial X_{nn}} \end{bmatrix} \overset{(J)}{=} \begin{bmatrix} \det \boldsymbol{X} \cdot \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial X_{11}}\Big) & \dots & \det \boldsymbol{X} \cdot \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial X_{1n}}\Big) \\ \vdots & \ddots & \vdots \\ \det \boldsymbol{X} \cdot \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial X_{n1}}\Big) & \dots & \det \boldsymbol{X} \cdot \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial X_{nn}}\Big) \end{bmatrix}$$

- Reference : Matrix Cookbook equations (46)

Derivative of $\det \boldsymbol{X}$ wrt. matrix $\boldsymbol{X}$

- **Theorem [Matrix Cookbook equation 49]** $\dfrac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = (\det \boldsymbol{X})\boldsymbol{X}^\top$
- **Proof** Take out the common factor $\det \boldsymbol{X}$ from the last page, we get

$$\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \begin{bmatrix} \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\dfrac{\partial \boldsymbol{X}}{\partial X_{11}}\Big) & \dots & \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\dfrac{\partial \boldsymbol{X}}{\partial X_{1n}}\Big) \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\dfrac{\partial \boldsymbol{X}}{\partial X_{n1}}\Big) & \dots & \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\dfrac{\partial \boldsymbol{X}}{\partial X_{nn}}\Big) \end{bmatrix} \overset{(3)}{=} \det \boldsymbol{X} \begin{bmatrix} \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\boldsymbol{J}^{11}\Big) & \dots & \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\boldsymbol{J}^{1n}\Big) \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\boldsymbol{J}^{n1}\Big) & \dots & \mathrm{Tr}\Big(\boldsymbol{X}^{-1}\boldsymbol{J}^{nn}\Big) \end{bmatrix}$$

$\mathrm{Tr}(\boldsymbol{A}\boldsymbol{J}^{ij}) \overset{(6)}{=} (\boldsymbol{A}^\top)_{ij}$ gives $\mathrm{Tr}\Big(\boldsymbol{X}^{-1}\boldsymbol{J}^{ij}\Big) = (\boldsymbol{X}^{-1})_{ji}$ and thus $\dfrac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \begin{bmatrix} (\boldsymbol{X}^{-1})_{11} & \dots & (\boldsymbol{X}^{-1})_{n1} \\ \vdots & \ddots & \vdots \\ (\boldsymbol{X}^{-1})_{1n} & \dots & (\boldsymbol{X}^{-1})_{nn} \end{bmatrix}$

Linear Algebra 101: $(\boldsymbol{X}^{-1})_{ij}$ is the cofactor C_{ij} of $\boldsymbol{X}$ divided by $\det \boldsymbol{X}$

$$\begin{bmatrix} (\boldsymbol{X}^{-1})_{11} & \dots & (\boldsymbol{X}^{-1})_{n1} \\ \vdots & \ddots & \vdots \\ (\boldsymbol{X}^{-1})_{1n} & \dots & (\boldsymbol{X}^{-1})_{nn} \end{bmatrix} = \frac{1}{\det \boldsymbol{X}} \begin{bmatrix} C_{11} & \dots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \dots & C_{nn} \end{bmatrix}^\top = \boldsymbol{X}^\top$$

Hence we proved $\dfrac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot \boldsymbol{X}^\top$. $\square$

Derivative of $\mathrm{logdet}\boldsymbol{X}$ wrt. matrix $\boldsymbol{X}$

- **Theorem [Matrix cookbook equation 57]** For positive definite matrix $\boldsymbol{X}$, then $\dfrac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{X}^\top$

  Proof Use the previous result and chain rule

  $$\frac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \boldsymbol{X}} = \frac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \det \boldsymbol{X}} \frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}}$$

  Note that $\dfrac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \boldsymbol{X}}$ and $\dfrac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}}$ are matrices and $\dfrac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \det \boldsymbol{X}}$ is scalar.

  So the expression is "matrix = scalar × matrix". Furthermore,

  - $\det \boldsymbol{X}$ is scalar so $\dfrac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \det \boldsymbol{X}}$ is just simply $\dfrac{1}{\det \boldsymbol{X}}$.

  Combines the two gives

  $$\frac{\partial \mathrm{logdet}\boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{X}^\top. \tag{9}$$

  □

Chain Rule with Frobenius inner product

- **Chain Rule** Let $\boldsymbol{U} = f(\boldsymbol{X})$ be a matrix-valued function of matrix variable $\boldsymbol{X}$ and $g(\boldsymbol{U})$ be a scalar-valued function of matrix variable $\boldsymbol{U}$, then we have

$$\frac{\partial g(\boldsymbol{U})}{\partial x} = \frac{\partial g(f(\boldsymbol{X}))}{\partial x} = \mathrm{Tr}\Big\{\Big(\frac{\partial g(\boldsymbol{U})}{\partial \boldsymbol{U}}\Big)^\top \frac{\partial \boldsymbol{U}}{\partial x}\Big\}. \tag{10}$$

- Important: note that the expression $\dfrac{\partial g(\boldsymbol{U})}{\partial x} = \dfrac{\partial g(\boldsymbol{U})}{\partial \boldsymbol{U}} \dfrac{\partial \boldsymbol{U}}{\partial x}$ is **wrong**

  1. $g(\boldsymbol{U})$ and x are scalar $\implies \dfrac{\partial g(\boldsymbol{U})}{\partial x}$ is scalar
  2. $\boldsymbol{U}$ is matrix $\implies \dfrac{\partial g(\boldsymbol{U})}{\partial \boldsymbol{U}}$, $\dfrac{\partial \boldsymbol{U}}{\partial x}$ and the product $\dfrac{\partial \boldsymbol{g}(U)}{\partial \boldsymbol{U}} \dfrac{\partial \boldsymbol{U}}{\partial x}$ are matrices.
  3. Matrix $\neq$ scalar (unless the matrix is of size 1-by-1).
  4. It is the trace operator turns the matrix into a scalar to fit the equality.
  5. As $\mathrm{Tr}(\boldsymbol{A}^\top \boldsymbol{B}) = \|\boldsymbol{AB}\|_F$, equation (10) tells that for the derivative of a scalar-valued function wrt. a scalar, when a matrix is introduced by chain rule, Frobenius inner product has to be applied to the result of the chain rule (which is a matrix) to change it back to scalar.

Derivative of $\operatorname{logdet} \boldsymbol{X}^\top \boldsymbol{X}$ wrt. matrix $\boldsymbol{X}$

- Let $g(\boldsymbol{X}) = \operatorname{logdet} f(\boldsymbol{X})$ and $f(\boldsymbol{X}) = \boldsymbol{X}^\top \boldsymbol{X}$ (so $g(\boldsymbol{X}) = \operatorname{logdet} \boldsymbol{X}^\top \boldsymbol{X}$)

- Consider the derivative wrt. x. As g and x are scalar so $\dfrac{\partial g}{\partial x}$ should be a scalar, and it is wrong to write

$$\frac{\partial g(\boldsymbol{X})}{\partial x} = \frac{\partial \operatorname{logdet} \boldsymbol{X}^\top \boldsymbol{X}}{\partial \det \boldsymbol{X}^\top \boldsymbol{X}} \frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial \boldsymbol{X}^\top \boldsymbol{X}} \frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}$$

as $\dfrac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial \boldsymbol{X}^\top \boldsymbol{X}}$ and $\dfrac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}$ are matrices.

- The correct way is to apply (10):

$$\begin{aligned}
\frac{\partial \operatorname{logdet} \boldsymbol{X}^\top \boldsymbol{X}}{\partial x} &= \frac{\partial \operatorname{logdet} \boldsymbol{X}^\top \boldsymbol{X}}{\partial \det \boldsymbol{X}^\top \boldsymbol{X}} \operatorname{Tr}\Big[\Big(\frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial \boldsymbol{X}^\top \boldsymbol{X}}\Big)^\top \frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big] \\
&= \frac{1}{\det \boldsymbol{X}^\top \boldsymbol{X}} \operatorname{Tr}\Big[\Big(\frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial \boldsymbol{X}^\top \boldsymbol{X}}\Big)^\top \frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big] \\
&\overset{\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot \boldsymbol{X}^{\top}}{=} \frac{1}{\det \boldsymbol{X}^\top \boldsymbol{X}} \operatorname{Tr}\Big[\det(\boldsymbol{X}^\top \boldsymbol{X})\Big(\boldsymbol{X}^\top \boldsymbol{X}\Big)^{-1} \frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big] \\
&= \operatorname{Tr}\Big[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \frac{\partial \det \boldsymbol{X}^\top \boldsymbol{X}}{\partial x}\Big]
\end{aligned}$$

Derivative of $\mathrm{logdet}\boldsymbol{X}^\top\boldsymbol{X}$ wrt. matrix $\boldsymbol{X}$ ... 2

Now put $x = X_{ij}$, we have

$$
\begin{aligned}
\frac{\partial \mathrm{logdet}\boldsymbol{X}^\top\boldsymbol{X}}{\partial X_{ij}} &= \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\frac{\partial \det \boldsymbol{X}^\top\boldsymbol{X}}{\partial X_{ij}}\Big] \\
\text{by } (5) \quad &= \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}(\boldsymbol{X}^\top\boldsymbol{J}^{ij} + \boldsymbol{J}^{ji}\boldsymbol{X})\Big] \\
&= \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{J}^{ij}\Big] + \mathrm{Tr}\Big[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{J}^{ji}\boldsymbol{X}\Big] \\
\text{by } (6,8) \quad &= \Big[(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\Big]^\top_{ij} + \Big[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big]_{ij} \\
&= \Big[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big]_{ij} + \Big[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big]_{ij} \\
&= 2\Big[\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\Big]_{ij}
\end{aligned}
$$

By (2), we finally have

$$
\frac{\partial \mathrm{logdet}\boldsymbol{X}^\top\boldsymbol{X}}{\partial \boldsymbol{X}} = 2\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = 2(\boldsymbol{X}^\dagger)^\top
$$

where $\boldsymbol{X}^\dagger = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$ is the left inverse of $\boldsymbol{X}$

Last page - summary

- Derivatives involving matrices : $\frac{\partial Y}{\partial x}$ and $\frac{\partial y}{\partial \boldsymbol{X}}$
- Single entry matrix : $\frac{\partial \boldsymbol{X}}{\partial X_{ij}} = \boldsymbol{J}^{ij}$ and some of its properties.
- Showed the following
  - $\frac{\partial \det \boldsymbol{X}}{\partial \boldsymbol{X}} = \det \boldsymbol{X} \cdot \boldsymbol{X}^\top$
  - $\frac{\partial \mathrm{logdet} \boldsymbol{X}}{\partial \boldsymbol{X}} = \boldsymbol{X}^\top$
  - $\frac{\partial \mathrm{logdet} \boldsymbol{X}^\top \boldsymbol{X}}{\partial \boldsymbol{X}} = 2\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$

End of document