Matrix Derivatives, Single Entry Matrix and derivatives of $X, X^T X, \det X, \ln\det X$ and $\ln\det X^T X$

Andersen Ang

Mathématique et de Recherche opérationnelle
Faculté polytechnique de Mons
UMONS
Mons, Belgium

email: manshun.ang@umons.ac.be
homepage: *angms.science*

July 14, 2017

Overview

Derivative of matrix w.r.t. scalar

For a matrix $Y \in \mathbb{R}^{m \times n}$,

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

Derivative of sclar w.r.t. matrix

For a matrix $X \in \mathbb{R}^{m\times n}$, $\dfrac{\partial y}{\partial X}$ is

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

Single Entry matrix

Consider the derivative of $\frac{\partial X}{\partial X_{ij}}$, where X_{ij} is the i^{th}-row j^{th}-column element of matrix X, which is a scalar.
For example $(i, j) = (1, 1)$:

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

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

Matrix with all zeros except a one in the position (i, j) is called Single Entry matrix, denoted as J^{ij}

Single Entry matrix as column and row selector

Right-multiplication with J^{ij} = put the i^{th} column to the j^{th} column.
For example

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

Left-multiplication with J^{ij} = put the j^{th} row to i^{th} row.
For example

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

Reference : the Matrix Cookbook equations (444) and (445)

Matrix derivative using single entry matrix

Using single entry matrix we have

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

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

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

It is important to note that these are derivatives of matrix w.r.t. scalar X_{ij}, not matrix. We are not considering $\frac{\partial X}{\partial X}$ here.

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

Illustration of $\frac{\partial X^T X}{\partial X_{ij}} = X^T J^{ij} + J^{ji} X$

Let $X = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$, then $X^T X = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$

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

$$\begin{aligned} \therefore \frac{\partial X^T X}{\partial a_3} &= \begin{pmatrix} 2a_3 & b_3 & c_3 \\ b_3 & 0 & 0 \\ c_3 & c_3 & 0 \end{pmatrix} = \begin{pmatrix} a_3 & 0 & 0 \\ b_3 & 0 & 0 \\ c_3 & 0 & 0 \end{pmatrix} + \begin{pmatrix} a_3 & b_3 & c_3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= X^T J^{31} + J^{13} X \end{aligned}$$

Single Entry matrix with trace

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

$$\mathsf{Tr}(AJ^{ij}) = \mathsf{Tr}(J^{ji}A) = (A^T)_{ij} \tag{6}$$

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

$$\mathsf{Tr}(AJ^{ij}B) = (A^T B^T)_{ij} \tag{7}$$

$$\mathsf{Tr}(AJ^{ji}B) = (BA)_{ij} \tag{8}$$

Illustration of (6):

$$X = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}, \text{ then } \mathsf{Tr}(XJ^{23}) = \mathsf{Tr}\begin{pmatrix} 0 & 0 & a_2 \\ 0 & 0 & b_2 \\ 0 & 0 & c_2 \end{pmatrix} = c_2 = (X^T)_{23}$$

Reference : Matrix Cookbook equations (450-452)

Application of J^{ij} in deriving matrix derivatives

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

$$\frac{\partial \det X}{\partial x} = \det X \cdot \mathsf{Tr}\left(X^{-1}\frac{\partial X}{\partial x}\right)$$

Note that $\det X$, x and $\det X \cdot \mathsf{Tr}\left(X^{-1}\frac{\partial X}{\partial x}\right)$ are all scalars.

To find $\dfrac{\partial \det X}{\partial X}$, use (1) we get

$$\frac{\partial \det X}{\partial X} = \begin{bmatrix} \frac{\partial \det X}{\partial X_{11}} & \cdots & \frac{\partial \det X}{\partial X_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \det X}{\partial X_{n1}} & \cdots & \frac{\partial \det X}{\partial X_{nn}} \end{bmatrix}$$

Thus we have

$$\frac{\partial \det X}{\partial X_{ij}} = \det X \cdot \mathsf{Tr}\left(X^{-1}\frac{\partial X}{\partial X_{ij}}\right)$$

Reference : Matrix Cookbook equations (46)

Derivative of $\det X$ w.r.t. matrix X

Theorem (Matrix Cookbook equation 49) $\dfrac{\partial \det X}{\partial X} = (\det X)X^{-T}$

Proof. Take out the common factor $\det X$ from the last page we get

$$\frac{\partial \det X}{\partial X} = \det X \begin{bmatrix} \mathsf{Tr}\left(X^{-1}\dfrac{\partial X}{\partial X_{11}}\right) & \dots & \mathsf{Tr}\left(X^{-1}\dfrac{\partial X}{\partial X_{1n}}\right) \\ \vdots & \ddots & \vdots \\ \mathsf{Tr}\left(X^{-1}\dfrac{\partial X}{\partial X_{n1}}\right) & \dots & \mathsf{Tr}\left(X^{-1}\dfrac{\partial X}{\partial X_{nn}}\right) \end{bmatrix}$$

By $\frac{\partial X}{\partial X_{ij}} = J^{ij}$ we have

$$\frac{\partial \det X}{\partial X} = \det X \begin{bmatrix} \mathsf{Tr}\left(X^{-1}J^{11}\right) & \dots & \mathsf{Tr}\left(X^{-1}J^{1n}\right) \\ \vdots & \ddots & \vdots \\ \mathsf{Tr}\left(X^{-1}J^{n1}\right) & \dots & \mathsf{Tr}\left(X^{-1}J^{nn}\right) \end{bmatrix}$$

By $\mathrm{Tr}(AJ^{ij}) = (A^T)_{ij}$ we have $\mathrm{Tr}\left(X^{-1}J^{ij}\right) = \left[X^{-1}\right]_{ji}$ and

$$\frac{\partial \det X}{\partial X} = \det X \begin{bmatrix} [X^{-1}]_{11} & \dots & [X^{-1}]_{n1} \\ \vdots & \ddots & \vdots \\ [X^{-1}]_{1n} & \dots & [X^{-1}]_{nn} \end{bmatrix}$$

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

$$\begin{bmatrix} [X^{-1}]_{11} & \dots & [X^{-1}]_{n1} \\ \vdots & \ddots & \vdots \\ [X^{-1}]_{1n} & \dots & [X^{-1}]_{nn} \end{bmatrix} = \frac{1}{\det X} \begin{bmatrix} C_{11} & \dots & C_{1n} \\ \vdots & \ddots & \vdots \\ C_{n1} & \dots & C_{nn} \end{bmatrix}^T = X^{-T}$$

Hence we proved

$$\frac{\partial \det X}{\partial X} = \det X \cdot X^{-T}$$

□

Derivative of log det X w.r.t. matrix X

Theorem (Matrix cookbook equation 57). For positive definite matrix X,
$$\frac{\partial \log \det X}{\partial X} = X^{-T}$$

Proof. Use the previous result and chain rule

$$\frac{\partial \log \det X}{\partial X} = \frac{\partial \log \det X}{\partial \det X} \frac{\partial \det X}{\partial X}$$

Note that $\frac{\partial \log \det X}{\partial X}$ and $\frac{\partial \det X}{\partial X}$ are matrices and $\frac{\partial \log \det X}{\partial \det X}$ is scalar. So the expression is "matrix = scalar × matrix". Furthermore,

- we proved $\frac{\partial \det X}{\partial X} = (\det X) X^{-T}$
- $\det X$ is scalar so $\frac{\partial \log \det X}{\partial \det X}$ is just simply $\frac{1}{\det X}$.

Combines the two we get

$$\frac{\partial \log \det X}{\partial X} = X^{-T} \tag{9}$$

□

Chain Rule with Frobenius inner product

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

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

Important to note that the expression $\frac{\partial g(U)}{\partial x} = \frac{\partial g(U)}{\partial U}\frac{\partial U}{\partial x}$ is **wrong** :
1. $g(U)$ and x are scalar $\implies$ $\frac{\partial g(U)}{\partial x}$ is scalar
2. U is matrix $\implies$ $\frac{\partial g(U)}{\partial U}$, $\frac{\partial U}{\partial x}$ and the product $\frac{\partial g(U)}{\partial U}\frac{\partial U}{\partial x}$ are matrices.
Matrix $\neq$ scalar (unless the matrix is of size 1-by-1).
It is the trace operator turns the matrix into a scalar to fit the equality.

As $\mathrm{Tr}(A^T B) = \|AB\|_F$, equation (10) tells that for the derivative of a scalar-valued function w.r.t. 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 log det $X^T X$ w.r.t. matrix X

Let $g(X) = \log\det f(X)$ and $f(X) = X^T X$ (so $g(X) = \log\det X^T X$)

Consider the derivative w.r.t. x. As g and x are scalar so $\frac{\partial g}{\partial x}$ should be a scalar. It is wrong to write

$$\frac{\partial g(X)}{\partial x} = \frac{\partial \log\det X^T X}{\partial \det X^T X} \frac{\partial \det X^T X}{\partial X^T X} \frac{\partial \det X^T X}{\partial x}$$

as $\frac{\partial \det X^T X}{\partial X^T X}$ and $\frac{\partial \det X^T X}{\partial x}$ are matrices.

The correct way is to apply (10):

$$\begin{aligned} \frac{\partial \log\det X^T X}{\partial x} &= \frac{\partial \log\det X^T X}{\partial \det X^T X} \mathsf{Tr}\Big[\Big(\frac{\partial \det X^T X}{\partial X^T X}\Big)^T \frac{\partial \det X^T X}{\partial x}\Big] \\ &= \frac{1}{\det X^T X} \mathsf{Tr}\Big[\det(X^T X)\Big(X^T X\Big)^{-1} \frac{\partial \det X^T X}{\partial x}\Big] \\ &= \mathsf{Tr}\Big[(X^T X)^{-1} \frac{\partial \det X^T X}{\partial x}\Big] \end{aligned}$$

Derivative of log det $X^T X$ w.r.t. matrix X ... 2

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

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

By (2), we finally have

$$\frac{\partial \log\det X^T X}{\partial X} = 2X(X^T X)^{-1} = 2(X^\dagger)^T$$

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

Last page - summary

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

End of document