Matrix Derivatives

김성민

서울대학교
통계학과, 베이즈통계 연구실

2022. 09. 06

CONTENTS

CONTENTS

Material

- Tracy, D., and Jinadasa, K. Patterned matrix derivatives.
  The Canadian Journal of Statistics/La Revue Canadienne de Statistique (1988), 411–418 (3)
- Deemer, W. L., and Olkin, I. The jacobians of certain matrix transformations useful in multivariate analysis: Based on lectures of pl hsu at the university of north carolina, 1947.
  Biometrika 38, 3/4 (1951), 345–367 (1)
- Magnus, J. R., and Neudecker, H. The commutation matrix: some properties and applications.
  The Annals of Statistics 7, 2 (1979), 381–394 (2)

CONTENTS

Derivatives of matrix

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}, \quad Y = f(X)$$

Then, we can evaluate the derivatives as the follows:

$$\frac{\partial Y}{\partial X} = \frac{\partial vecY}{\partial vecX}$$

Example

$$X = \begin{pmatrix} s & t \\ t & s^2 \end{pmatrix}, \quad Y = f(X) = \begin{pmatrix} st \\ s^2 t \end{pmatrix}$$

- How to obtain matrix derivatives?

Patterned matrix

Definition

If at least one of the following statements about a matrix is true, it is said to be **patterned**:

- Some elements are constant.
- Some elements are function of other elements.

Definition

- **vecX** is the vectorization of matrix **X**.
- **vecpX** is the column vector of all distinct variables obtained from the elements of **vecX**.

$$X = \begin{pmatrix} x_1 & x_2 & 8 & x_4 \\ x_2 & x_1x_3 & x_3^2 & x_1x_3 \end{pmatrix}$$

- **vecX** $= [x_1, x_2, x_2, x_1x_3, 8, x_3^2, x_4, x_1x_3]^T$
- **vecpX** $= [x_1, x_2, x_3, x_4]^T$

CONTENTS

Setting

- Consider the matrix-valued function $Y = f(X)$ of a matrix X. $(Y \in \mathbb{R}^{p,q}, \quad X \in \mathbb{R}^{m,n})$
- Let X be patterned matrix and $|vecpX| = k$.
- Define J from $\mathbb{R}^k$ onto $D \subset \mathbb{R}^{m,n}$. ($D$ is collection of $m \times n$ matrix with the same pattern with X)
- Consider $\tilde{f}(X)$ which is extension of $f(X)$ to the larger domain $\mathbb{R}^{m,n}$.

Derivatives of Patterned matrix

- Consider $g = f \circ J$, i.e., $g(x) = \tilde{f}(J(x)) = f(X)$, since $J(x) \in D$

$$\frac{\partial g}{\partial x} = \frac{\partial \tilde{f}}{\partial J(x)} \frac{\partial J(x)}{\partial x}$$

$$\begin{aligned} \Longrightarrow [\frac{\partial g}{\partial x}] &= [\frac{\partial \tilde{f}}{\partial J(x)}][\frac{\partial J(x)}{\partial x}] \\ &= [\frac{\partial \tilde{f}}{\partial X}][\frac{\partial J(x)}{\partial x}] \\ &= \frac{\partial vecY}{\partial vecX} \frac{\partial vecJ(x)}{\partial x} \end{aligned}$$

Derivatives of Patterned matrix

Definiton

Let $Y = f(X)$ be a matrix valued function of the matrix X. Then,

$$\frac{\partial vecY}{\partial vecpX} = \frac{\partial vecY}{\partial vecX}\frac{\partial vecX}{\partial vecpX}$$

- In calculating $\frac{\partial vecY}{\partial vecX}$, we consider all elements of X are independent variable.

Derivative of Example

$$X = \begin{pmatrix} s & t \\ t & s^2 \end{pmatrix}, \quad Y = f(X) = \begin{pmatrix} st \\ s^2 t \end{pmatrix}$$

- $k = 2.$
- $D = \left\{ A : A = \begin{pmatrix} s & t \\ t & s^2 \end{pmatrix} : s, t \in \mathbb{R} \right\}$
- $x = vecX = [s, t]^T$
- $J(x) = J([s, t]^T) = X$
- $\tilde{f}(\begin{pmatrix} s & u \\ t & v \end{pmatrix}) = \begin{pmatrix} su \\ tv \end{pmatrix}$

Derivative of Example

$$
\begin{aligned}
\frac{\partial vecY}{\partial vecpX} &= \frac{\partial vecY}{\partial vecX}\frac{\partial vecX}{\partial vecpX} \\
&= \begin{pmatrix} u & 0 & s & 0 \\ 0 & v & 0 & t \end{pmatrix}\Bigg|_{u=t, v=s^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 2s & 0 \end{pmatrix} \\
&= \begin{pmatrix} t & 0 & s & 0 \\ 0 & s^2 & 0 & t \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 2s & 0 \end{pmatrix} \\
&= \begin{pmatrix} t & s \\ 2ts & s^2 \end{pmatrix}
\end{aligned}
$$

CONTENTS

Setting

- Consider $g = f \circ J_1$, i.e., $g(x) = \tilde{f}(J_1(x)) = f(X)$, since $J_1(x) \in D$
- Define J_2 from $P \subset \mathbb{R}^{p,q}$ onto $\mathbb{R}^k$. (P is collection of $p \times q$ matrix with the same pattern with Y).
- Consider $\tilde{J}_2(Y)$ which is extension of $J_2(Y)$ to the larger domain $\mathbb{R}^{p,q}$.

Jacobians of Patterned Matrix Transformations

- Consider $g = J_2 \circ f \circ J_1$, i.e., $g(x) = J_2(\tilde{f}(J(x)))$.

$$
\begin{aligned}
[\frac{\partial g}{\partial x}] &= [\frac{\partial J_2}{\partial \tilde{f}}][\frac{\partial \tilde{f}}{\partial J_1(x)}][\frac{\partial J_1(x)}{\partial x}] \\
&= [\frac{\partial J_2}{\partial Y}][\frac{\partial \tilde{f}}{\partial X}][\frac{\partial J_1(x)}{\partial x}] \\
&= \left[\frac{\partial vecJ_2}{\partial vecY}\right]_{\{Y \in P\}} \left[\frac{\partial vecY}{\partial vecX}\right]_{\{X \in D\}} \frac{\partial vecJ_1(x)}{\partial x}
\end{aligned}
$$

Jacobians of Patterned Matrix Transformations

Definiton

Let $Y = f(X)$ be a matrix valued function of the matrix X. Then,

$$\frac{\partial vecpY}{\partial vecpX} = \frac{\partial vecpY}{\partial vecY}\frac{\partial vecY}{\partial vecX}\frac{\partial vecX}{\partial vecpX}$$

- In calculating $\frac{\partial vecY}{\partial vecX}$, we consider all elements of X are independent variable.

CONTENTS

Kronecker product

Definition

$A \otimes B$ is the operator from $X \in \mathbb{R}^{m,n}$ to $Y \in \mathbb{R}^{m,n}$ where $Y = BXA^T$

- We write down as follows:

$$(A \otimes B)X = BXA^T$$
$$\text{OR} \quad (A \otimes B)vecX = vec(BXA^T)$$

- By using "Kronecker Product", we can instantly write down the Jacobian.

$$dvecY = (A \otimes B)dvecX$$

Lemma I

Lemma

Let X, Y, A be $p \times p$ matrices. If we assume that X is a symmetric matrix, Jacobian determinant of the transformation $Y = AXA^T$ transformation is $|A|^{p+1}$.

Proof.

Assume that A is diagonalizable, with $Au_i = \lambda_i u_i$. Let $M_{ij} = u_i u_j^T + u_j u_i^T$ $(i \leq j)$, then M_{ij} form a basis for $p \times p$ symmetric matrix. We also know that

$$AM_{ij}A^T = \lambda_i \lambda_j M_{ij}$$
$$\text{OR} \quad (A \otimes A) vecM_{ij} = \lambda_i \lambda_j vecM_{ij}$$

So that the determinant is $\prod_{i \leq j} \lambda_i \lambda_j = |A|^{p+1}$ $\square$

Lemma II

Lemma

Let X be a $p \times p$ matrix. Then, Jacobian determinant of the transformation $Y = X^{-1}$ transformation is $|X|^{-2p}$

Proof.

Since $XY = I$, we can obtain $(dX)Y + X(dY) = 0$.

$$\begin{aligned} \Rightarrow \ dY &= -X^{-1}(dX)X^{-1} \\ &= -(X^{-T} \otimes X^{-1})dX \\ |J| &= |X|^{-2p} \end{aligned}$$

If X is symmetric matrix, then the Jacobian determinant is $|X|^{-p-1}$ by Lemma I. $\square$

Lemma III

Lemma

Let X *be a* $p \times p$ *matrix. For general* $X \neq X^T$*,*

$$\frac{\mathrm{d}|X|}{\mathrm{d}X}\left(=\frac{\partial |X|}{\partial vecpX}\right) = |X|X^{-T}.$$

For $X = X^T$*,*

$$\frac{\mathrm{d}|X|}{\mathrm{d}X} = 2|X|X^{-T} - diag(|X|X^{-T}).$$

Lemma III

Proof.

$$
\begin{aligned}
[\frac{\mathrm{d}}{\mathrm{d}X}|X|]_{ij} &= \frac{\mathrm{d}}{\mathrm{d}X_{ij}} \sum_{k=1}^{p} (-1)^{k+j} X_{kj} X^{kj} \quad \text{(where } (-1)^{k+j} X^{kj} \text{ is cofactor of } X_{kj}) \\
&= (-1)^{i+j} X^{ij} \\
&= |X|[X^{-1}]_{ji}.
\end{aligned}
$$

For $X = X^T$,

$$
\begin{aligned}
[\frac{\mathrm{d}}{\mathrm{d}X}|X|]_{ij} &= \frac{\mathrm{d}}{\mathrm{d}X_{ij}} \sum_{k=1}^{p} (-1)^{k+j} X_{kj} X^{kj} \\
&= (-1)^{i+j} X^{ij} + (-1)^{j+i} X^{ji} I(i \neq j) \\
&= (-1)^{i+j} X^{ij} (2 - I(i = j)) \\
&= |X|[2X^{-1} - diag(X^{-1})]_{ij}
\end{aligned}
$$

Jeffreys prior I

$$X \sim N(0, \Lambda^{-1})$$

- $L \equiv \log N(X|0, \Lambda^{-1}) = \frac{1}{2}\log|\Lambda| - \frac{1}{2}X^T\Lambda X + const.$
- $A = \frac{dL}{d\Lambda} = \Lambda^{-1} - \frac{1}{2}diag(\Lambda^{-1}) - \frac{1}{2}XX^T$

$$\begin{aligned}
\frac{\partial A}{\partial \Lambda^{-1}} &= \frac{\partial A}{\partial \Lambda}\frac{\partial \Lambda}{\partial \Lambda^{-1}} \\
&= \left| I_{p(p+1)/2} - \frac{1}{2}J \right| |\Lambda|^{-(p+1)} \\
&= \frac{1}{2^p}|\Lambda|^{-(p+1)}
\end{aligned}$$

$$\therefore \quad \pi(\Lambda) \propto |\Lambda|^{-\frac{p+1}{2}}$$

Jeffreys prior II

- Let I be a information matrix of Λ.

$$
\begin{aligned}
I &= \mathbb{E}\left[\left(\frac{\partial L}{\partial vecp(\Lambda)}\right)^T\left(\frac{\partial L}{\partial vecp(\Lambda)}\right)\right] \\
&= \mathbb{E}\left[G^T\left(\frac{\partial L}{\partial vec(\Lambda)}\right)^T\left(\frac{\partial L}{\partial vec(\Lambda)}\right)^T G\right] \quad \text{where} \quad G = \frac{\partial vec(\Lambda)}{\partial vecp(\Lambda)} \\
&= G^T Var(\frac{\partial L}{\partial vec(\Lambda)})G \\
&= G^T Var(\frac{1}{2}\Lambda^{-1} - \frac{1}{2}XX^T)G \\
&= \frac{1}{4}G^T(I + K_p)(\Lambda^{-1} \otimes \Lambda^{-1})G \\
&= \frac{1}{2}G^T(\Lambda^{-1} \otimes \Lambda^{-1})G
\end{aligned}
$$

References I

[1] Deemer, W. L., and Olkin, I. The jacobians of certain matrix transformations useful in multivariate analysis: Based on lectures of pl hsu at the university of north carolina, 1947. *Biometrika 38*, 3/4 (1951), 345–367.

[2] Magnus, J. R., and Neudecker, H. The commutation matrix: some properties and applications. *The Annals of Statistics 7*, 2 (1979), 381–394.

[3] Tracy, D., and Jinadasa, K. Patterned matrix derivatives. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* (1988), 411–418.