L12 Commutation matrices

1. Comments on $\mathrm{vec}(AXB) = (B' \otimes A)\,\mathrm{vec}(X)$

   (1) Isomorphic spaces
   As spaces with inner products, $R^{m\times n}$ and R^{mn} are isomorphic since $\mathrm{vec}(\cdot)$ is a 1-1 mapping from $R^{m\times n}$ to R^{mn} that preserves the inner products. This mapping is called an isomorphism. As vectors in a linear space with inner products we can treat $A \in R^{m\times n}$ and $\mathrm{vec}(A) \in R^{mn}$ "the same thing".

   (2) A linear transformation from $R^{p\times q}$ to $R^{m\times n}$
   $Y = f(X) = AXB$ is a linear transformation from $X \in R^{p\times q}$ to $Y \in R^{m\times n}$ since $f(\alpha X_1 + \beta X_2) = \alpha f(X_1) + \beta f(X_2)$.

   (3) A linear transformation from R^{pq} to R^{mn}
   Thus $\mathrm{vec}(Y) = \mathrm{vec}(AXB)$ must be a linear transformation from $\mathrm{vec}(X) \in R^{pq}$ to $\mathrm{vec}(Y) \in R^{mn}$. But for column vector linear transformation, it has general form $u = Av$. So $\mathrm{vec}(Y) = \mathrm{vec}(AXB) = M\mathrm{vec}(X)$. This M is $M = B' \otimes A$.

2. Commutation matrices

   (1) Relations
   In exploring the relations of $A \otimes B$ and $B \otimes A$; and $\mathrm{vec}(A)$ and $\mathrm{vec}(A')$ for $A \in R^{m\times n}$ and $B \in R^{p\times q}$, it will be shown that It will be shown that

   $$\mathrm{vec}(A') = K_{mn}\mathrm{vec}(A) \text{ and } K_{pm}\,(A \otimes B)\,K_{nq} = B \otimes A.$$

   Here K_{mn} is called a commutation matrix.

   (2) Definition of commutation matrix K_{mn}
   With $I_m = (e_{1m}, ..., e_{mm})$ and $I_n = (e_{1n}, ..., e_{nn})$ define

   $$K_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n} e_{im} \otimes e'_{jn} \otimes e'_{im} \otimes e_{jn} \in R^{mn\times mn}.$$

   Then $K_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n}(e_{im}e'_{jn}) \otimes (e_{im}e'_{jn})'$.

   Ex1: For column vectors x and y, $y' \otimes x = x \otimes y' = xy'$ and $\mathrm{vec}(xy') = y \otimes x$.
   Thus $K_{m,n}$ takes on many forms.
   $K_{m,n} = \sum_{i=1}^{m}\sum_{j=1}^{n} e_{im} \otimes e'_{jn} \otimes e'_{im} \otimes e_{jn} = \sum_{i=1}^{m}\sum_{j=1}^{n}(e_{im}e'_{jn}) \otimes (e_{im}e'_{in})'$

   $$\begin{aligned} K_{m,n} &= \textstyle\sum_{i=1}^{m}\sum_{j=1}^{n}(e_{im} \otimes e_{jn}) \otimes (e_{jn} \otimes e_{im})' = \sum_{i=1}^{m}\sum_{j=1}^{n}(e_{im} \otimes e_{jn})(e_{jn} \otimes e_{im})' \\ &= \textstyle\sum_{i=1}^{m}\sum_{j=1}^{n}\mathrm{vec}(e_{jn}e'_{im})[\mathrm{vec}(e_{im}e'_{jn})]'. \end{aligned}$$

   One can stick to one form.

   (3) Property I
   Transpose of a commutation matrix is a commutation matrix $K'_{mn} = K_{nm}$.

   $$\begin{aligned} \textbf{Proof } K'_{mn} &= \Big[\textstyle\sum_{i=1}^{m}\sum_{j=1}^{n} e_{im} \otimes e'_{jn} \otimes e'_{im} \otimes e_{jn}\Big]' \\ &= \textstyle\sum_{i=1}^{m}\sum_{j=1}^{n} e'_{im} \otimes e_{jn} \otimes e_{im} \otimes e'_{jn} = \sum_{j=1}^{n}\sum_{i=1}^{m} e_{jn} \otimes e'_{im} \otimes e'_{jn} \otimes e_{im} \\ &= K_{nm} \end{aligned}$$

   Ex2: $K_{m1} = \sum_{i=1}^{m} e_{im} \otimes 1' \otimes e'_{im} \otimes 1 = \sum_{i=1}^{m} e_{im}e'_{im} = I_m$.
   $K_{1n} = K'_{n1} = I'_n = I_n$.

(4) Property II
Commutation matrix is an orthogonal matrix $\quad K_{mn}^{-1} = K_{nm}'$.

Proof $\quad K_{mn}K_{nm}$

$$\begin{aligned}
&= \left[\sum_i^m \sum_j^n (e_{im} \otimes e_{jn}' \otimes e_{im}' \otimes e_{jn})\right] \left[\sum_p^n \sum_q^m (e_{pn} \otimes e_{qm}' \otimes e_{pn}' \otimes e_{qm})\right] \\
&= \sum_i^m \sum_j^n \sum_p^n \sum_q^m (e_{im} \otimes e_{jn}' \otimes e_{im}' \otimes e_{jn})(e_{qm}' \otimes e_{pn} \otimes e_{qm} \otimes e_{pn}') \\
&= \sum_i^m \sum_j^n \sum_p^n \sum_q^m (e_{im}e_{qm}') \otimes (e_{jn}'e_{pn}) \otimes (e_{im}'e_{qm}) \otimes (e_{jn}e_{pn}') \\
&= \sum_i^m \sum_j^n (e_{im}e_{im}') \otimes (e_{jn}e_{jn}') = \sum_i^m (e_{im}e_{im}') \otimes \sum_j^n (e_{jn}e_{jn}') \\
&= I_m \otimes I_n = I_{mn}
\end{aligned}$$

So $K_{mn}^{-1} = K_{nm} = K_{mn}'$. K_{mn} is orthogonal.

3. Application of commutation matrices

(1) For $A \in R^{m\times n}$, $K_{mn}\,\mathrm{vec}(A) = \mathrm{vec}(A')$.

Pf: For $A = (a_{ij})_{m\times n}$,

$$\begin{aligned}
\mathrm{vec}(A') &= \mathrm{vec}\left[\sum_{i=1}^m \sum_{j=1}^n (e_{im}'Ae_{jn})(e_{jn}e_{im}')\right] \\
&= \mathrm{vec}\left[\sum_{i=1}^m \sum_{j=1}^n (e_{jn}e_{im}')A(e_{jn}e_{im}')\right] \\
&= \sum_{i=1}^m \sum_{j=1}^n \mathrm{vec}\left[(e_{jn}e_{im}')A(e_{jn}e_{im}')\right] \\
&= \sum_{i=1}^m \sum_{j=1}^n \left[(e_{jn}e_{im}')' \otimes (e_{jn}e_{im}')\right] \mathrm{vec}(A) \\
&= \left(\sum_{i=1}^m \sum_{j=1}^n e_{im} \otimes e_{jn}' \otimes e_{jn} \otimes e_{im}'\right) \mathrm{vec}(A) \\
&= \left(\sum_{i=1}^m \sum_{j=1}^n e_{im} \otimes e_{jn}' \otimes e_{im}' \otimes e_{jn}\right) \mathrm{vec}(A) = K_{mn}\,\mathrm{vec}(A).
\end{aligned}$$

(2) For $A \in R^{m\times n}$ and $B \in R^{p\times q}$, $K_{pm}\,(A \otimes B)\,K_{nq} = B \otimes A$.

Pf: With $A \in R^{m\times n}$ and $B \in^{p\times q}$, for all $X \in R^{q\times n}$,

$$\begin{aligned}
K_{pm}(A \otimes B)\,\mathrm{vec}(X) &= K_{pm}\mathrm{vec}(BXA') = \mathrm{vec}(AX'B') \\
&= (B \otimes A)\,\mathrm{vec}(X') = (B \otimes A)\,K_{qn}\mathrm{vec}(X).
\end{aligned}$$

So $K_{pm}(A \otimes B) = (B \otimes A)\,K_{qn}$. Thus $K_{pm}(A \otimes B)K_{nq} = B \otimes A$.

Ex3: With $A \in R^{m\times n}$ and $y \in R^p$, $K_{pm}(A \otimes y)K_{n1} = y \otimes A$, i.e.,
$K_{pm}(A \otimes y) = y \otimes A$.

L13 Matrices that store all derivatives

1. One function

   (1) Definition
   For one function, argument can be allowed to be a matrix: $y = f(X)$ where $X = (x_{ij})_{p\times q}$.
   Define $\frac{\partial y}{\partial X} = \left(y'_{x_{ij}}\right)_{p\times q}$.

   (2) Formula I: $\frac{\partial x^T Ax}{\partial x} = (A^T + A)x$ and $\frac{\partial x^T Ax}{\partial x^T} = x^T(A^T + A)$.
   With symmetric A, $\frac{\partial x^T Ax}{\partial x} = 2Ax$ and $\frac{\partial x^T Ax}{\partial x^T} = x^T 2A$.

   Ex1: $y = f(x) = x^T \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} x = 2x_1^2 - 2x_1x_2 + 4x_2^2$

   By definition: $\frac{\partial y}{\partial x} = \begin{pmatrix} y'_{x_1} \\ y'_{x_2} \end{pmatrix} = \begin{pmatrix} 4x_1 - 2x_2 \\ -2x_1 + 8x_2 \end{pmatrix}$.

   By formula: $\frac{\partial y}{\partial x} = 2 \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4x_1 - 2x_2 \\ -2x_1 + 8x_2 \end{pmatrix}$

   (3) Formula II: $\frac{\partial \alpha^T X\beta}{\partial X} = \alpha\beta^T \in R^{p\times q}$.

   Ex2: $y = (a_1,\, a_2) \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$

   $x_{11}a_1b_1 + x_{12}a_1b_2 + x_{13}a_1b_3 + x_{21}a_2b_1 + x_{22}a_2b_2 + x_{23}a_2b_3$.

   By definition: $\frac{\partial y}{\partial X} = \begin{pmatrix} y'_{x_11} & y'_{x_{12}} & y'_{x_{13}} \\ y'_{x_{21}} & y'_{x_{22}} & y'_{x_{23}} \end{pmatrix} = \begin{pmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \end{pmatrix}$.

   By formula: $\frac{\partial y}{\partial X} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} (b_1,\, b_2,\, b_3) = \begin{pmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \end{pmatrix}$

   (4) Formula III: With $X = (x_{ij})_{p\times p}$ $\frac{\partial\,\mathrm{tr}(X)}{\partial X} = I_p$.
   Proof: $\frac{\partial\,\mathrm{tr}(X)}{\partial X} = \frac{\partial\, x_{11}+\cdots+x_{pp}}{\partial X} = I_p$.

   (5) Formula IV: $\frac{\partial |X|}{\partial X} = |X|(X')^{-1}$.
   Proof: Let $C = (c_{ij})_{p\times p}$ be the cofactor matrix for $X = (x_{ij})_{p\times p}$.
   From $|X| = x_{i1}c_{i1} + \cdots + x_{ip}c_{ip}$, $\frac{\partial |X|}{\partial X} = \left(|X|'_{x_{ij}}\right)_{p\times p} = (c_{ij})_{p\times p} = C$.
   But $XC' = |X|I_p \Longrightarrow C' = |X|X^{-1} \Longrightarrow C = |X|(X')^{-1}$.
   Hence $\frac{\partial |X|}{\partial X} = |X|(X')^{-1}$.

2. A vector of functions

   (1) Definition
   For a vector of m functions, the argument could be a vector of n components.

   Define $\frac{\partial y}{\partial x^T} = \begin{pmatrix} (y_1)'_{x_1} & \cdots & (y_1)'_{x_p} \\ \vdots & \ddots & \vdots \\ (y_m)'_{x_1} & \cdots & (y_m)'_{x_p} \end{pmatrix}$ and $\frac{\partial y^T}{\partial x} = \begin{pmatrix} (y_1)'_{x_1} & \cdots & (y_m)'_{x_1} \\ \vdots & \ddots & \vdots \\ (y_1)'_{x_p} & \cdots & (y_m)'_{x_p} \end{pmatrix}$

   $\left(\frac{\partial y}{\partial x^T}\right)^T = \frac{\partial y^T}{\partial x}$.

   (2) Formula V: $\frac{\partial Ax}{\partial x^T} = A$ and $\frac{\partial (Ax)^T}{\partial x} = A^T$.

Proof: $y = Ax = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1p}x_p \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mp}x_p \end{pmatrix}$.

$\frac{\partial Ax}{\partial x^T} = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{pmatrix} = A$.

3. A matrix of functions

(1) Definition

For a matrix of functions with one argument, $Y = (y_{st}(x))_{m\times n}$ define $\frac{dY}{dx} = \left(\frac{d\,y_{st}}{d\,x}\right)_{m\times n}$

Ex3: For $Y = \begin{pmatrix} 2x & -4x^2 \end{pmatrix}$, $\frac{d\,Y}{d\,x} = (2,\ -8x)$.

(2) For a matrix of functions with a matrix of arguments

$\frac{\partial\,\mathrm{vec}(Y)}{\partial\,[\mathrm{vec}(X)]^T}$ and/or $\frac{\partial\,[\mathrm{vec}(Y)]^T}{\partial\,\mathrm{vec}(X)}$ store of partial derivatives.

Ex4: For $Y = AXB$, $\frac{\partial\,\mathrm{vec}(AXB)}{\partial\,[\mathrm{vec}(X)]^T} = \frac{\partial\,(B^T\otimes A)\mathrm{vec}(X)}{\partial\,[\mathrm{vec}(X)]^T} = B^T \otimes A$.

(3) For $Y \in R^{m\times n}$ and $X \in R^{p\times q}$,

mn matrices $\frac{\partial\,y_{st}}{\partial\,X} \in R^{p\times q}$, $s = 1, ..., m; t = 1, ..., n$ store all partial derivatives

Ex5: For $Y = AXB$, $\frac{\partial\,(AXB)_{st}}{\partial\,X} = A^T E_{m\times n}(s,t)B^T$ where $E_{m\times n}(s,t) \in R^{m\times n}$ with elements at (s,t) is 1 and 0 elsewhere.

$\frac{\partial\,(AXB)_{st}}{\partial\,X} = \frac{\partial\,e_{sm}^T AXBe_{tn}}{\partial\,X} = A^T e_{sm} e_{tn}^T B^T = A^T E_{m\times n}(s,t)B^T$.

(4) For $Y \in R^{m\times n}$ and $X \in R^{p\times q}$,

pq matrices $\frac{\partial\,Y}{\partial\,x_{ij}} \in R^{m\times n}$, $i = 1, .., p$; $j = 1, ..., q$ store all partial derivatives

Ex6: $\frac{\partial\,AXB}{\partial\,x_{ij}} = AE_{p\times q}(i,j)B$.

$A = (A_1, ..., A_p)$, $A_i = Ae_{ip}$; $B = \begin{pmatrix} B_{(1)}^T \\ \vdots \\ B_{(q)}^T \end{pmatrix}$ $B_{(j)}^T = e_{jq}^T B$

$\frac{\partial\,AXB}{\partial\,x_{ij}} = \frac{\partial\,\sum_{ij} x_{ij} A_i B_{(j)}^T}{\partial\,x_{ij}} = A_i B_{(j)}^T = Ae_{ip}e_{jq}^T B = AE_{p\times q}(i,j)B$.

Summary of formulas

$\frac{\partial\,x^T Ax}{\partial\,x} = (A^T + A)x$; $\quad \frac{\partial\,Ax}{\partial\,x^T} = A$; $\quad \frac{\partial\,\alpha^T X\beta}{\partial\,X} = \alpha\beta^T$

$\frac{\partial\,\mathrm{tr}(X)}{\partial\,X} = I \quad \frac{\partial\,|X|}{\partial\,X} = |X|(X')^{-1}$

$\frac{\partial\,[\mathrm{vec}(AXB)]}{\partial\,[\mathrm{vec}(X)]^T} = B^T \otimes A$; $\quad \frac{\partial\,(AXB)_{st}}{\partial\,X} = A^T E_{m\times n}(s,t)B^T$; $\quad \frac{\partial\,AXB}{\partial\,x_{ij}} = AE_{p\times q}(i,j)B$

Ex7: x, $\beta \in R^p$. Find $\frac{\partial\,x^T\beta}{\partial\,x}$.

$$\frac{\partial\,x^T\beta}{\partial\,x} = \left(\frac{\partial\,\beta^T x}{\partial\,x^T}\right)^T = \left(\beta^T\right)^T = \beta.$$

$$\frac{\partial\,x^T\beta}{\partial\,x} = \frac{\partial\,\beta^T x 1}{\partial\,x} = \beta\,1 = \beta$$