Matrix Derivatives with Chain Rule and Rules for Simple, Hadamard, and Kronecker Products (1)

Mubing Zhou, Rongpeng Li, Wenchao Zhang

SUSTC

May 6, 2014

Notations and Definitions

- $\overline{X}$: $1 \times nm$ row vector of all elements of an $n \times m$ matrix X, taken row by row.
- $D_{\overline{X}}$: $nm \times nm$ diagonal matrix with diagonal elements $d_{ii} = x_i$, the corresponding element of the row vector $\overline{X}$.
- $X * Y$: Hadamard product of two $n \times m$ matrices X and Y, i.e $X * Y = [x_{ij} y_{ij}]$.
- $X \otimes Y$: Kronecker product of two matrices, i.e., $X \otimes Y = [x_{ij} Y]$

Definition of 'derivative': Let Z be a $p \times q$ matrix and Y be a $m \times n$ matrix whose elements are differentiable functions of all elements in Z. Define $\frac{\partial Y}{\partial Z}$ to be a $pq \times mn$ matrix whose ith row is the mn-vector $\frac{\partial \overline{Y}}{\partial z_i}$, where z_i is the i-th element in $\overline{Z}$. Obviously from the definition, $\frac{\partial Y}{\partial Z} = \frac{\partial \overline{Y}}{\partial \overline{Z}}$.

Examples

Suppose $Z = (z_1, z_2)$, $Y = \begin{pmatrix} y_1 & y_2 \\ y_3 & y_4 \end{pmatrix}$ and $X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$. Then

$$\begin{aligned}
\overline{Y} &= (y_1, y_2, y_3, y_4) \\
D_{\overline{Y}} &= diag\{y_1, y_2, y_3, y_4\} \\
X * Y &= \begin{pmatrix} x_1 y_1 & x_2 y_2 \\ x_3 y_3 & x_4 y_4 \end{pmatrix} \\
Z \otimes Y &= (z_1 Y, z_2 Y) \\
&= \begin{pmatrix} z_1 y_1 & z_1 y_2 & z_2 y_1 & z_2 y_2 \\ z_1 y_3 & z_1 y_4 & z_3 y_3 & z_2 y_4 \end{pmatrix}
\end{aligned}$$

Examples (cont'd)

$$\frac{\partial Y}{\partial Z} = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_2}{\partial z_1} & \frac{\partial y_3}{\partial z_1} & \frac{\partial y_4}{\partial z_1} \\ \frac{\partial y_1}{\partial z_2} & \frac{\partial y_2}{\partial z_2} & \frac{\partial y_3}{\partial z_2} & \frac{\partial y_4}{\partial z_2} \end{pmatrix}$$

Chain Rule

Theorem

Suppose each element of a **vector** $Y = [Y_1, Y_2]$ *is a differential function of all elements of a* **vector** Z*, where* Y_1 *contains all the mathematical variables which are independent to each other in* Y *and* Y_2 *are functions of elements in* Y_1*. Furthermore, suppose that each element of a* **vector** X *is a differentiable function of all elements of* Y_1*. Then*

$$\frac{\partial X}{\partial Z} = \frac{\partial Y_1}{\partial Z}\frac{\partial X}{\partial Y_1} \tag{1}$$

Proof of Chain Rule

Suppose X is $1 \times r$, Y_1 is $1 \times s$ and Z is $1 \times t$. Then we have

$$\frac{\partial X}{\partial Z} = \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \cdots & \frac{\partial x_r}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_t} & \cdots & \frac{\partial x_r}{\partial z_t} \end{pmatrix}$$

and

$$\frac{\partial Y_1}{\partial Z}\frac{\partial X}{\partial Y_1} = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_s}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_t} & \cdots & \frac{\partial y_s}{\partial z_t} \end{pmatrix} \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_r}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_s} & \cdots & \frac{\partial x_r}{\partial y_s} \end{pmatrix}$$

The (i,j)th element in $\frac{\partial X}{\partial Z}$ is $\frac{\partial x_i}{\partial z_j}$ while the corresponding element in $\frac{\partial Y_1}{\partial Z}\frac{\partial X}{\partial Y_1}$ is $\sum_{l=1}^{s} \frac{\partial y_l}{\partial z_i}\frac{\partial x_j}{\partial y_l}$. By the chain rule in scalar calculus, we verify Eq.1.

Warning

Generally speaking, the following equation is NOT true!

$$\frac{\partial X}{\partial Z} = \frac{\partial Y}{\partial Z}\frac{\partial X}{\partial Y}$$

A counterexample: $X = [v_1, v_1, v_2, v_2]$, $Y = [v_1, v_2, v_1, v_2]$ and $Z = [v_1, v_2, v_2, v_1]$. Now

$$\frac{\partial X}{\partial Z} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

But $\frac{\partial Y}{\partial Z}\frac{\partial X}{\partial Y} = \begin{pmatrix} 2 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 2 & 2 & 0 & 0 \end{pmatrix}$

Product Rule

Theorem

Suppose each element of a $p \times r$ matrix Y and an $r \times q$ matrix X is a differentiable function of all elements of a $m \times n$ matrix Z. Then

$$\frac{\partial (YX)}{\partial Z} = \frac{\partial Y}{\partial Z}(I^p \otimes X) + \frac{\partial X}{\partial Z}(Y' \otimes I^q) \tag{2}$$

Proof of Product Rule

$\frac{\partial(YX)}{\partial Z}$ is a $mn \times pq$ matrix. Its (i,j)th element corresponds to the i-th element in $\overline{Z}$, say z_β, and j-th element in $\overline{XY}$, say $(XY)_{(s,t)}$. $(XY)_{(s,t)}$ is the product of the s-th row of Y and t-th column of X, or $(XY)_{(s,t)} = \sum_{l=1}^{r} y_{sl}x_{lt}$. So

$$\frac{\partial(YX)}{\partial Z} = [\frac{\partial \sum_{l=1}^{r} y_{sl}x_{lt}}{\partial z_\beta}] = \sum_{l=1}^{r} \frac{\partial y_{sl}}{\partial z_\beta}x_{lt} + \frac{\partial x_{lt}}{\partial z_\beta}y_{sl}$$

Proof of Product Rule (cont'd)

Now consider $\frac{\partial Y}{\partial Z}(I^p \otimes X)$. $\frac{\partial Y}{\partial Z}$ is a $mn \times pr$ matrix while $I^p \otimes X$ is a $pr \times pq$ matrix. The (i,j)th element of $\frac{\partial Y}{\partial Z}(I^p \otimes X)$ is the product of i-th row of $\frac{\partial Y}{\partial Z}$ and j-th column of $I^p \otimes X$. The i-th row of corresponds to the i-th element of $\overline{Z}$, which is z_β. Recall that $I^p \otimes X$ is in the form of $\begin{pmatrix} X & & & \\ & X & & \\ & & \ddots & \\ & & & X \end{pmatrix}$ So the j-th column of $I^p \otimes X$ corresponds to the t-column of X. And the t-th column of X should be multiplied by the s-th row of Y. So the (i,j)th element of $\frac{\partial Y}{\partial Z}(I^p \otimes X)$ is $\sum_{l=1}^{k} \frac{\partial y_{sl}}{\partial z_\beta} x_{lt}$. Similarly, we can prove that the (i,j)th element of $\frac{\partial X}{\partial Z}(Y' \otimes I^q)$ is $\sum_{l=1}^{k} \frac{\partial x_{lt}}{\partial z_\beta} y_{sl}$. Add these two parts up is just the (i,j)th element of $\frac{\partial (YX)}{\partial Z}$.

Properties of Hadamard Product Cont'd

The Hadamard product is commutative, associative and distributive over addition. That is

$$A * B = B * A \tag{3}$$
$$A * (B * C) = (A * B) * C \tag{4}$$
$$A * (B + C) = A * B + A * C \tag{5}$$

The identity matrix under Hadamard multiplication of two m-by-n matrices is m-by-n matrix where all elements are equal to 1, which is different from simple identity matrix I

A matrix has an inverse under Hadamard multiplication if and only if none of the elements are equal to zero.

In programming language like MATLAB, Hadamard product is done by using $.*$ where $*$ stands for the normal multiplication.

Properties of Hadamard Product

For square A and B, the row-sums of their Hadamard product are the diagonal elements of AB^T

$$\sum_j (A * B)_{i,j} = (AB^T)_{i,j} \tag{6}$$

The Hadamard product is a principal submatrix of the Kronecker product. A principal submatrix is a square submatrix where the distinguished rows and columns are the same.

Schur Product Theorem

Theorem

The Hadamard product of two positive-semidefinite matrices is positive-semidefinite. This is known as the Schur product theorem

Proof.

Proof is done using eigendecomposition.
Let $M = \sum u_i m_i m_i^T$ and $N = \sum v_i n_i n_i^T$. Then
$M * N = \sum_{i,j} u_i v_j (m_i m_i^T) * (n_j n_j^T) = \sum_{i,j} u_i v_j (m_i * n_j)(m_i * n_j)^T$. Each $(m_i * n_j)(m_i * n_j)^T$ is positive and $u_i v_j > 0$, thus the sum giving $M * N$ is also positive. □

Product Rule for Hadamard Product

Theorem

Suppose Y and X are $m - n$ matrices and each of whose elements is a differentiable function of all elements of a $p - n$ matrix Z. Then

$$\frac{\partial (Y * X)}{\partial Z} = \frac{\partial Y}{\partial Z}(D_{\hat{x}}) + \frac{\partial X}{\partial Z}(D_{\hat{y}}). \tag{7}$$

Proof.

$$\frac{\partial Y}{\partial Z}(D_{\hat{x}}) + \frac{\partial X}{\partial Z}(D_{\hat{y}}) = [\sum_{ij}(\frac{\partial y_{ij}}{\partial z_{st}} x_{ij}) + \sum_{jk}(y_{jk}\frac{\partial x_{jk}}{\partial z_{st}})] \tag{8}$$

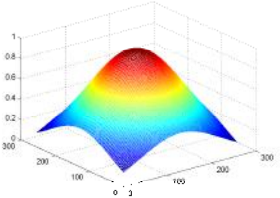
By the definition of D_x,

$$RHS = \frac{\partial (Y * X)}{\partial Z}. \tag{9}$$

□

Physical Meaning of Hadamard Product

A very intuitive way to understand Hadamard product is to view it as a amplification of image processing. Here is a demonstration of Hadamard product that reduces the brightness of a picture. The matrix of original picture Hadamard-multiplies another matrix of exponential decay from the center.



Hadamard Matrix

The Hadamard matrix H is a square with elements 1 or -1. It has several interesting properties.

$$HH^T = nI_n \tag{10}$$

A Hadamard matrix has maximal determinant among matrices with entries of absolute value less than or equal to 1. If M is a complex matrix of order n, whose entries are bounded by $M_{ij} \leq 1$, then $|det(M)| \leq n^{n/2}$.
The Hadamard matrix can be constructed by Sylvester's construction.

Sylvester's construction

If H is a Hadamard matrix, then the matrix below is also a Hadamard matrix. $\begin{pmatrix} H & H \\ H & -H \end{pmatrix}$ For example, $H_1 = [1]$, $H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.
$H_{2^k} = \begin{pmatrix} H_{2^{k-1}} & H_{2^{k-1}} \\ H_{2^{k-1}} & H_{2^{k-1}} \end{pmatrix} = H_2 \otimes H_{2^{k-1}}$ where $\otimes$ denotes the Kronecker product.