Matrix Algebra Marathon

J2200071 Ryuto Saito

May 18, 2024

1 Derivatives of Function of Matrices

1.1 Definition

Definition 1.1.1. *We denote the derivative of a function* $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ *by*

$$\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) \coloneqq \begin{bmatrix} \nabla_{X_{1,1}}(f(\boldsymbol{X})) & \cdots & \nabla_{X_{1,n}}(f(\boldsymbol{X})) \\ \vdots & \ddots & \vdots \\ \nabla_{X_{m,1}}(f(\boldsymbol{X})) & \cdots & \nabla_{X_{m,n}}(f(\boldsymbol{X})) \end{bmatrix}.$$

1.2 Exercise

Exercise 2.14.2.

Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{m\times n}$. Show that

$$\nabla_{\boldsymbol{X}}(tr(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})) = \boldsymbol{A}. \tag{1}$$

Proof. Let $\boldsymbol{X}, \boldsymbol{A} \in \mathbb{R}^{m\times n}$. Recall the definition of inner-product of matrices and

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}).$$

$$\begin{aligned}
\nabla_{\boldsymbol{X}}(tr(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})) &= \begin{bmatrix} \nabla_{X_{1,1}} tr(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}) & \cdots & \nabla_{X_{1,n}} tr(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}) \\ \vdots & \ddots & \vdots \\ \nabla_{X_{m,1}} tr(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}) & \cdots & \nabla_{X_{m,n}} tr(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}) \end{bmatrix} \\
&= \begin{bmatrix} \nabla_{X_{1,1}} \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}X_{i,j}\right) & \cdots & \nabla_{X_{1,n}} \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}X_{i,j}\right) \\ \vdots & \ddots & \vdots \\ \nabla_{X_{m,1}} \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}X_{i,j}\right) & \cdots & \nabla_{X_{m,n}} \left(\sum_{i=1}^m \sum_{j=1}^n A_{i,j}X_{i,j}\right) \end{bmatrix} \\
&= \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix} \\
&= \boldsymbol{A}
\end{aligned}$$

$\square$

Exercise 2.14.4.

Let $\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$. Show that

$$\nabla_{\boldsymbol{X}}(tr(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{B}))) = \boldsymbol{A}. \tag{2}$$

Proof. Let $\boldsymbol{X}, \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$.

$$\begin{aligned}
&\nabla_{\boldsymbol{X}}(tr(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{B}))) \\
&= \begin{bmatrix} \nabla_{X_{1,1}}(tr(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{B}))) & \cdots & \nabla_{X_{1,n}}(tr(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{B}))) \\ \vdots & \ddots & \vdots \\ \nabla_{X_{m,1}}(tr(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{B}))) & \cdots & \nabla_{X_{m,n}}(tr(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}+\boldsymbol{B}))) \end{bmatrix} \\
&= \begin{bmatrix} \nabla_{X_{1,1}}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{i,j}(X_{i,j}+B_{i,j})\right) & \cdots & \nabla_{X_{1,n}}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{i,j}(X_{i,j}+B_{i,j})\right) \\ \vdots & \ddots & \vdots \\ \nabla_{X_{m,1}}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{i,j}(X_{i,j}+B_{i,j})\right) & \cdots & \nabla_{X_{m,n}}\left(\sum_{i=1}^{m}\sum_{j=1}^{n} A_{i,j}(X_{i,j}+B_{i,j})\right) \end{bmatrix} \\
&= \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix} \\
&= \boldsymbol{A}
\end{aligned}$$

$\square$

2 Derivatives of Matrix-Valued Functions

2.1 Definition

Definition 2.1.1. *Consider a matrix-valued function $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m \times n}$ expressed as*

$$\boldsymbol{F}(x) \coloneqq \begin{bmatrix} F_{1,1}(x) & \cdots & F_{1,n}(x) \\ \vdots & \ddots & \vdots \\ F_{m,1}(x) & \cdots & F_{m,n}(x) \end{bmatrix}.$$

The derivative of the function is denoted by

$$\nabla_x \boldsymbol{F}(x) \coloneqq \begin{bmatrix} \nabla_x F_{1,1}(x) & \cdots & \nabla_x F_{1,n}(x) \\ \vdots & \ddots & \vdots \\ \nabla_x F_{m,1}(x) & \cdots & \nabla_x F_{m,n}(x) \end{bmatrix}.$$

2.2 Exercise

Exercise 2.15.3.

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times n}$. Show that

$$\nabla_x \langle \boldsymbol{A}, \boldsymbol{F}(x)\rangle = \langle \boldsymbol{A}, \nabla_x \boldsymbol{F}(x)\rangle. \tag{3}$$

Proof. Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times n}$.

$$\begin{aligned}
\nabla_x \langle \boldsymbol{A}, \boldsymbol{F}(x)\rangle &= \nabla_x \left(\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} F_{i,j}(x) \right) \\
&= \left(\sum_{i=1}^{m} \sum_{j=1}^{n} A_{i,j} \nabla_x F_{i,j}(x) \right) \\
&= \langle \boldsymbol{A}, \nabla_x \boldsymbol{F}(x)\rangle
\end{aligned}$$

□

Exercise 2.15.5.

Let $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times k}$ and $\boldsymbol{G} : \mathbb{R} \to \mathbb{R}^{k\times n}$. Show that

$$\nabla_x (\boldsymbol{F}(x)\boldsymbol{G}(x)) = (\nabla_x \boldsymbol{F}(x))\boldsymbol{G}(x) + \boldsymbol{F}(x)\nabla_x \boldsymbol{G}(x). \tag{4}$$

Proof. Let $\boldsymbol{F} : \mathbb{R} \to \mathbb{R}^{m\times k}$ and $\boldsymbol{G} : \mathbb{R} \to \mathbb{R}^{k\times n}$. For all $i \in \mathbb{N}_m$ and $j \in \mathbb{N}_n$,

$$\begin{aligned}
(\nabla_x (\boldsymbol{F}(x)\boldsymbol{G}(x)))_{i,j} &= \nabla_x (\boldsymbol{F}(x)\boldsymbol{G}(x))_{i,j} \\
&= \nabla_x \left(\sum_{l=1}^{k} F_{i,l}(x) G_{l,j}(x) \right) \\
&= \sum_{l=1}^{k} ((\nabla_x F_{i,l}(x)) G_{l,j}(x) + F_{i,l}(x)(\nabla_x G_{l,j}(x))) \\
&= \sum_{l=1}^{k} (\nabla_x F_{i,l}(x)) G_{l,j}(x) + \sum_{l=1}^{k} F_{i,l}(x) \nabla_x G_{l,j}(x) \\
&= \sum_{l=1}^{k} (\nabla_x \boldsymbol{F}(x))_{i,l} G_{l,j}(x) + \sum_{l=1}^{k} F_{i,l}(x) (\nabla_x \boldsymbol{G}(x))_{l,j} \\
&= ((\nabla_x \boldsymbol{F}(x))\boldsymbol{G}(x))_{i,j} + (\boldsymbol{F}(x)\nabla_x \boldsymbol{G}(x))_{i,j} \\
&= ((\nabla_x \boldsymbol{F}(x))\boldsymbol{G}(x) + \boldsymbol{F}(x)\nabla_x \boldsymbol{G}(x))_{i,j}.
\end{aligned}$$

Therefore, (4) holds. □

3 Diagonal Matrices

3.1 Definition

Definition 3.1.1. *A diagonal matrix is a square matrix whose all the off-diagonal entries are zero. Such a matrix with diagonal entries* $\boldsymbol{d} \coloneqq [d_1, \ldots, d_n]^{\mathrm{T}} \in \mathbb{R}^n$ *is denoted by*

$$diag(\boldsymbol{d}) \coloneqq \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

3.2 Exercise

Exercise 2.16.3.

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Show that

$$diag(\boldsymbol{x})diag(\boldsymbol{y}) = diag(\boldsymbol{y})diag(\boldsymbol{x}). \tag{5}$$

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.

$$\begin{aligned} diag(\boldsymbol{x})diag(\boldsymbol{y}) &= \begin{bmatrix} x_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & x_n \end{bmatrix} \begin{bmatrix} y_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & y_n \end{bmatrix} \\ &= \begin{bmatrix} x_1y_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & x_ny_n \end{bmatrix} \\ &= \begin{bmatrix} y_1x_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & y_nx_n \end{bmatrix} \\ &= \begin{bmatrix} y_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & y_n \end{bmatrix} \begin{bmatrix} x_1 & & \boldsymbol{O} \\ & \ddots & \\ \boldsymbol{O} & & x_n \end{bmatrix} \\ &= diag(\boldsymbol{y})diag(\boldsymbol{x}) \end{aligned}$$

$\square$

4 Orthonormal Matrices

4.1 Definition

Definition 4.1.1. *An $m \times n$ matrix $\boldsymbol{P}$ is said to be orthonormal if the matrix satisfies $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{I}_n$. Symbol $\mathbb{O}^{m\times n}$ is used to denote the set of $m\times n$ orthonormal matrices.*

4.2 Exercise

Exercise 2.17.2.

Let $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n] \in \mathbb{O}^{m\times n}$. Show that $\forall i, \forall j \in \mathbb{N}_n$,

$$\langle \boldsymbol{p}_i, \boldsymbol{p}_j \rangle = \delta_{i,j} \tag{6}$$

where $\delta_{i,j}$ is the Kronecker delta.

Proof. Let $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_n] \in \mathbb{O}^{m\times n}$ and $i, j \in \mathbb{N}_n$.

$$\begin{aligned}
(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})_{i,j} &= \sum_{k=1}^{m} (\boldsymbol{P}^{\mathrm{T}})_{i,k} P_{k,j} \\
&= \sum_{k=1}^{m} \boldsymbol{P}_{k,i} P_{k,j} \\
&= \langle \boldsymbol{p}_i, \boldsymbol{p}_j \rangle
\end{aligned}$$

Furthermore,

$$\begin{aligned}
(\boldsymbol{I}_n)_{i,j} &= \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \\
&= \delta_{i,j}
\end{aligned}$$

Thus, since $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{I}_n$,

$$\langle \boldsymbol{p}_i, \boldsymbol{p}_j \rangle = (\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})_{i,j} = (\boldsymbol{I}_n)_{i,j} = \delta_{i,j}.$$

Then, we get (6). $\square$

Exercise 2.17.4.

Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Show that

$$\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}. \tag{7}$$

Proof. Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Since $\boldsymbol{P} \in \mathbb{O}^{n\times n}$,

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{I}_n.$$

Recall the definition of inverse of square matrices. Then, we get

$$\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}.$$

Thus, (7) holds. □

Exercise 2.17.6.

Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Show that

$$det(\boldsymbol{P}) \in \{\pm 1\}. \tag{8}$$

Proof. Let $\boldsymbol{P} \in \mathbb{O}^{n\times n}$. Recall that $det(\boldsymbol{AB}) = det(\boldsymbol{A})det(\boldsymbol{B})$ and $det(\boldsymbol{A}^{\mathrm{T}}) = det(\boldsymbol{A})$ for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$. Then,

$$det(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}) = det(\boldsymbol{P}^{\mathrm{T}})det(\boldsymbol{P}) = det(\boldsymbol{P})det(\boldsymbol{P}) = (det(\boldsymbol{P}))^2.$$

Recall that

$$det(\boldsymbol{I}_n) = 1.$$

Then, from the definition of orthonormal matrices,

$$(det(\boldsymbol{P}))^2 = det(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}) = det(\boldsymbol{I}_n) = 1.$$

By solving this equation, we have $det(\boldsymbol{P}) = \pm 1$. Hence, $det(\boldsymbol{P}) \in \{\pm 1\}$. □

Exercise 2.17.8.

Let $\boldsymbol{P} \in \mathbb{O}^{m\times k}$ and $\boldsymbol{Q} \in \mathbb{O}^{k\times n}$ where $m \geq k \geq n$. Show that $\boldsymbol{PQ} \in \mathbb{O}^{m\times n}$.

Proof. Let $\boldsymbol{P} \in \mathbb{O}^{m\times k}$ and $\boldsymbol{Q} \in \mathbb{O}^{k\times n}$ where $m \geq k \geq n$. Since $\boldsymbol{P} \in \mathbb{O}^{m\times k}$,

$$\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{I}_k.$$

Similarly,

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}_n.$$

Then,

$$\begin{aligned}(\boldsymbol{PQ})^{\mathrm{T}}\boldsymbol{PQ} &= \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PQ} \\ &= \boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P})\boldsymbol{Q} \\ &= \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{I}_k\boldsymbol{Q} \\ &= \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} \\ &= \boldsymbol{I}_n.\end{aligned}$$

Hence, $\boldsymbol{PQ} \in \mathbb{O}^{m\times n}$. □