# 矩阵理论与方法

# 内容提要 CONTENTS

- ☐ **课程信息**

- ☐ **课程介绍**

- ☐ **矩阵理论与方法**

  **第3章 矩阵分析及其应用**

# 其他微分

函数对矩阵的导数

函数矩阵对矩阵的导数

# 其它微分概念

函数对矩阵的导数(包括向量)

定义:设 $X=(\xi_{ij})_{m\times n}$, $mn$ 元函数

$$f(X) = f(\xi_{11}, \xi_{12}, \cdots, \xi_{1n}, \cdots, \xi_{m\times n})$$

定义 $f(X)$ 对矩阵 $X$ 的导数为

$$\frac{df}{dX} = \left(\frac{\partial f}{\partial \xi_{ij}}\right)_{m\times n} = \begin{bmatrix} \dfrac{\partial f}{\partial \xi_{11}} & \cdots & \dfrac{\partial f}{\partial \xi_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial f}{\partial \xi_{m1}} & \cdots & \dfrac{\partial f}{\partial \xi_{mn}} \end{bmatrix}$$

$$\text{例}11: x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}: \quad f(x) = f(\xi_1, \xi_2, \cdots, \xi_n) \qquad \frac{df}{dx} = \begin{bmatrix} \dfrac{\partial f}{\partial \xi_1} \\ \vdots \\ \dfrac{\partial f}{\partial \xi_n} \end{bmatrix}$$

例11: $x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$: $f(x) = f(\xi_1, \xi_2, \cdots, \xi_n)$ $\qquad \dfrac{df}{dx} = \begin{bmatrix} \dfrac{\partial f}{\partial \xi_1} \\ \vdots \\ \dfrac{\partial f}{\partial \xi_n} \end{bmatrix}$

例12: $A = (a_{ij})_{n \times n}$, $x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$: $f(x) = x^T A x,$ 求 $\dfrac{df}{dx}$

例11: $x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$: $f(x) = f(\xi_1, \xi_2, \cdots, \xi_n)$   $\dfrac{df}{dx} = \begin{bmatrix} \dfrac{\partial f}{\partial \xi_1} \\ \vdots \\ \dfrac{\partial f}{\partial \xi_n} \end{bmatrix}$

例12: $A = (a_{ij})_{n \times n}$, $x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$: $f(x) = x^T Ax$, 求 $\dfrac{df}{dx}$

$$f(x) = \xi_1 \sum_{j=1}^{n} a_{1j}\xi_j + \cdots + \xi_k \sum_{j=1}^{n} a_{kj}\xi_j + \cdots + \xi_n \sum_{j=1}^{n} a_{nj}\xi_j$$

$$\frac{\partial f}{\partial \xi_k} = \xi_1 a_{1k} + \cdots + \xi_{k-1}a_{k-1,k} + \left( \sum_{j=1}^{n} a_{kj}\xi_j + \xi_k a_{kk} \right)$$

$$+ \xi_{k+1}a_{k+1,k} + \cdots + \xi_n a_{nk} = \sum_{j=1}^{n} a_{kj}\xi_j + \sum_{i=1}^{n} a_{ik}\xi_i$$

例12: $A = \left( a_{ij} \right)_{n \times n}, \ x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} : \ f(x) = x^T A x, \ 求 \ \dfrac{df}{dx}$

$\therefore \dfrac{df}{dx} = \left( A + A^T \right) x$

如果 $A = A^T$ ， 有 $\dfrac{df}{dx} = 2Ax$

# 矩阵对矩阵的导数

定义:设 $X = \left(\xi_{ij}\right)_{m \times n}$, $f_{kl}(X) = f_{kl}\left(\xi_{11}, \xi_{12}, \cdots, \xi_{1n}, \cdots, \xi_{m \times n}\right)$

$$F = \begin{bmatrix} f_{11} & \cdots & f_{1s} \\ \vdots & & \vdots \\ f_{r1} & \cdots & f_{rs} \end{bmatrix}, \qquad \frac{\partial F}{\partial \xi_{ij}} = \begin{bmatrix} \dfrac{\partial f_{11}}{\partial \xi_{ij}} & \cdots & \dfrac{\partial f_{1s}}{\partial \xi_{ij}} \\ \vdots & & \vdots \\ \dfrac{\partial f_{r1}}{\partial \xi_{ij}} & \cdots & \dfrac{\partial f_{rs}}{\partial \xi_{ij}} \end{bmatrix},$$

定义 $\dfrac{dF}{dX} = \begin{bmatrix} \dfrac{\partial F}{\partial \xi_{11}} & \cdots & \dfrac{\partial F}{\partial \xi_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial F}{\partial \xi_{m1}} & \cdots & \dfrac{\partial F}{\partial \xi_{mn}} \end{bmatrix}$

例15: $x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$, $F(x) = [f_1(x), f_2(x), \cdots, f_l(x)]$

$$\frac{dF}{dx} =$$

例15: $x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$, $F(x) = \begin{bmatrix} f_1(x), f_2(x), \cdots, f_l(x) \end{bmatrix}$

$$\frac{dF}{dx} = \begin{bmatrix} \dfrac{\partial f_1}{\partial \xi_1} & \cdots & \dfrac{\partial f_l}{\partial \xi_1} \\ \vdots & & \vdots \\ \dfrac{\partial f_1}{\partial \xi_n} & \cdots & \dfrac{\partial f_l}{\partial \xi_n} \end{bmatrix}$$

例$15: x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$, $F(x) = \left[ f_1(x), f_2(x), \cdots, f_l(x) \right]$

$$\frac{dF}{dx} = \begin{bmatrix} \dfrac{\partial f_1}{\partial \xi_1} & \cdots & \dfrac{\partial f_l}{\partial \xi_1} \\ \vdots & & \vdots \\ \dfrac{\partial f_1}{\partial \xi_n} & \cdots & \dfrac{\partial f_l}{\partial \xi_n} \end{bmatrix}$$

例$16:$ $A = \left( a_{ij} \right)_{n \times n}$, $\quad x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$ $\quad Ax = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j}\xi_j \\ \vdots \\ \sum\limits_{j=1}^{n} a_{nj}\xi_j \end{bmatrix}$

$$\frac{d(Ax)}{dx^T} =$$

例$15: x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix}$, $F(x) = \begin{bmatrix} f_1(x), f_2(x), \cdots, f_l(x) \end{bmatrix}$

$$\frac{dF}{dx} = \begin{bmatrix} \dfrac{\partial f_1}{\partial \xi_1} & \cdots & \dfrac{\partial f_l}{\partial \xi_1} \\ \vdots & & \vdots \\ \dfrac{\partial f_1}{\partial \xi_n} & \cdots & \dfrac{\partial f_l}{\partial \xi_n} \end{bmatrix}$$

例$16: \quad A = \left( a_{ij} \right)_{n \times n}, \qquad x = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} \qquad Ax = \begin{bmatrix} \displaystyle\sum_{j=1}^{n} a_{1j} \xi_j \\ \vdots \\ \displaystyle\sum_{j=1}^{n} a_{nj} \xi_j \end{bmatrix}$

$$\frac{d(Ax)}{dx^T} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = A$$

**Types of Matrix Derivatives**

| Types | Scalar | Vector | Matrix |
|---|---|---|---|
| Scalar | $\dfrac{\partial y}{\partial x}$ | $\dfrac{\partial \mathbf{y}}{\partial x}$ | $\dfrac{\partial \mathbf{Y}}{\partial x}$ |
| Vector | $\dfrac{\partial y}{\partial \mathbf{x}}$ | $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$ | |
| Matrix | $\dfrac{\partial y}{\partial \mathbf{X}}$ | | |

## Vector-by-scalar  [ edit ]

The derivative of a vector $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$ , by a scalar $x$ is written (in numerator layout notation) as

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

## Scalar-by-vector [ edit ]

The derivative of a scalar $y$ by a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, is written (in denominator layout notation) as $\dfrac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}$.

## Vector-by-vector [ edit ]

Each of the previous two cases can be considered as an application of the derivative of a vector with respect to a vector, using a vector of involving vectors in a corresponding way.

The derivative of a vector function (a vector whose components are functions) $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$, with respect to an input vector, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$,

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

## Matrix-by-scalar   [ edit ]

The derivative of a matrix function $\mathbf{Y}$ by a scalar $x$ is known as the **tangent matrix** and is given (in numerator layout notation) by

$$\frac{\partial \mathbf{Y}}{\partial x} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{12}}{\partial x} & \cdots & \frac{\partial y_{1n}}{\partial x} \\ \frac{\partial y_{21}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{m1}}{\partial x} & \frac{\partial y_{m2}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix}.$$

## Scalar-by-matrix   [ edit ]

The derivative of a scalar $y$ function of a $p \times q$ matrix $\mathbf{X}$ of independent variables, with respect to the matrix $\mathbf{X}$, is given (in numerator layout notation) by

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

**Result of differentiating various kinds of aggregates with other kinds of aggregates**

| | Scalar $y$ | | Vector y (size $m$) | | Matrix Y (size $m \times n$) | |
| --- | --- | --- | --- | --- | --- | --- |
| | Notation | Type | Notation | Type | Notation | Type |
| Scalar $x$ | $\dfrac{\partial y}{\partial x}$ | scalar | $\dfrac{\partial \mathbf{y}}{\partial x}$ | (numerator layout) size-$m$ column vector<br>(denominator layout) size-$m$ row vector | $\dfrac{\partial \mathbf{Y}}{\partial x}$ | (numerator layout) $m \times n$ matrix |
| Vector x (size $n$) | $\dfrac{\partial y}{\partial \mathbf{x}}$ | (numerator layout) size-$n$ row vector<br>(denominator layout) size-$n$ column vector | $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$ | (numerator layout) $m \times n$ matrix<br>(denominator layout) $n \times m$ matrix | $\dfrac{\partial \mathbf{Y}}{\partial \mathbf{x}}$ | |
| Matrix X (size $p \times q$) | $\dfrac{\partial y}{\partial \mathbf{X}}$ | (numerator layout) $q \times p$ matrix<br>(denominator layout) $p \times q$ matrix | $\dfrac{\partial \mathbf{y}}{\partial \mathbf{X}}$ | | $\dfrac{\partial \mathbf{Y}}{\partial \mathbf{X}}$ | |

分子布局 $\quad \dfrac{\partial y}{\partial x} = \dfrac{dy}{dx^T}$

**Numerator-layout notation** [ edit ]

Using numerator-layout notation, we have:[1]

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \cdots & \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{21}} & \cdots & \frac{\partial y}{\partial x_{p1}} \\ \frac{\partial y}{\partial x_{12}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{p2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{1q}} & \frac{\partial y}{\partial x_{2q}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

分母布局 $\quad \dfrac{\partial y}{\partial x} = \dfrac{dy^T}{dx}$

**Denominator-layout notation** [ edit ]

Using denominator-layout notation, we have:[3]

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}.$$

$$\frac{\partial y}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1q}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2q}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{p1}} & \frac{\partial y}{\partial x_{p2}} & \cdots & \frac{\partial y}{\partial x_{pq}} \end{bmatrix}.$$

## Identities: vector-by-vector $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$

| Condition | Expression | Numerator layout, i.e. by y and x$^T$ | Denominator layout, i.e. by y$^T$ and x |
|---|---|---|---|
| **a** is not a function of **x** | $\dfrac{\partial \mathbf{a}}{\partial \mathbf{x}} =$ | **0** | |
| | $\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}} =$ | **I** | |
| **A** is not a function of **x** | $\dfrac{\partial \mathbf{Ax}}{\partial \mathbf{x}} =$ | $\mathbf{A}$ | $\mathbf{A}^\top$ |
| **A** is not a function of **x** | $\dfrac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$ | $\mathbf{A}^\top$ | $\mathbf{A}$ |
| *a* is not a function of **x**, **u** = **u(x)** | $\dfrac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$ | $a\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ | |
| *a* = *a*(**x**), **u** = **u(x)** | $\dfrac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$ | $a\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\dfrac{\partial a}{\partial \mathbf{x}}$ | $a\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}} + \dfrac{\partial a}{\partial \mathbf{x}}\mathbf{u}^\top$ |
| **A** is not a function of **x**, **u** = **u(x)** | $\dfrac{\partial \mathbf{Au}}{\partial \mathbf{x}} =$ | $\mathbf{A}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ | $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$ |
| **u** = **u(x)**, **v** = **v(x)** | $\dfrac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$ | $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}} + \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ | |
| **u** = **u(x)** | $\dfrac{\partial \mathbf{g(u)}}{\partial \mathbf{x}} =$ | $\dfrac{\partial \mathbf{g(u)}}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ | $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\dfrac{\partial \mathbf{g(u)}}{\partial \mathbf{u}}$ |
| **u** = **u(x)** | $\dfrac{\partial \mathbf{f(g(u))}}{\partial \mathbf{x}} =$ | $\dfrac{\partial \mathbf{f(g)}}{\partial \mathbf{g}}\dfrac{\partial \mathbf{g(u)}}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$ | $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\dfrac{\partial \mathbf{g(u)}}{\partial \mathbf{u}}\dfrac{\partial \mathbf{f(g)}}{\partial \mathbf{g}}$ |

Identities: scalar-by-vector $\dfrac{\partial y}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} y$

| Condition | Expression | Numerator layout, i.e. by $\mathbf{x}^{\mathsf{T}}$; result is row vector | Denominator layout, i.e. by $\mathbf{x}$; result is column vector |
|---|---|---|---|
| $a$ is not a function of $\mathbf{x}$ | $\dfrac{\partial a}{\partial \mathbf{x}} =$ | $\mathbf{0}^{\top}$ [4] | $\mathbf{0}$ [4] |
| $a$ is not a function of $\mathbf{x}$, $u = u(\mathbf{x})$ | $\dfrac{\partial au}{\partial \mathbf{x}} =$ | $a\dfrac{\partial u}{\partial \mathbf{x}}$ | |
| $u = u(\mathbf{x}),\ v = v(\mathbf{x})$ | $\dfrac{\partial(u + v)}{\partial \mathbf{x}} =$ | $\dfrac{\partial u}{\partial \mathbf{x}} + \dfrac{\partial v}{\partial \mathbf{x}}$ | |
| $u = u(\mathbf{x}),\ v = v(\mathbf{x})$ | $\dfrac{\partial uv}{\partial \mathbf{x}} =$ | $u\dfrac{\partial v}{\partial \mathbf{x}} + v\dfrac{\partial u}{\partial \mathbf{x}}$ | |
| $u = u(\mathbf{x})$ | $\dfrac{\partial g(u)}{\partial \mathbf{x}} =$ | $\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{x}}$ | |
| $u = u(\mathbf{x})$ | $\dfrac{\partial f(g(u))}{\partial \mathbf{x}} =$ | $\dfrac{\partial f(g)}{\partial g}\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{x}}$ | |
| $\mathbf{u} = \mathbf{u}(\mathbf{x}),\ \mathbf{v} = \mathbf{v}(\mathbf{x})$ | $\dfrac{\partial(\mathbf{u} \cdot \mathbf{v})}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{u}^{\top}\mathbf{v}}{\partial \mathbf{x}} =$ | $\mathbf{u}^{\top}\dfrac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^{\top}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$<br><br>• assumes numerator layout of $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ | $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v} + \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{u}$<br><br>• assumes denominator layout of $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ |
| $\mathbf{u} = \mathbf{u}(\mathbf{x}),\ \mathbf{v} = \mathbf{v}(\mathbf{x})$, $\mathbf{A}$ is not a function of $\mathbf{x}$ | $\dfrac{\partial(\mathbf{u} \cdot \mathbf{A}\mathbf{v})}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{u}^{\top}\mathbf{A}\mathbf{v}}{\partial \mathbf{x}} =$ | $\mathbf{u}^{\top}\mathbf{A}\dfrac{\partial \mathbf{v}}{\partial \mathbf{x}} + \mathbf{v}^{\top}\mathbf{A}^{\top}\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}$<br><br>• assumes numerator layout of $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ | $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}\mathbf{v} + \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A}^{\top}\mathbf{u}$<br><br>• assumes denominator layout of $\dfrac{\partial \mathbf{u}}{\partial \mathbf{x}}, \dfrac{\partial \mathbf{v}}{\partial \mathbf{x}}$ |
| | $\dfrac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^{T}} =$ | | $\mathbf{H}$, the Hessian matrix[5] |

| | | | |
|---|---|---|---|
| **a** is not a function of **x** | $\dfrac{\partial(\mathbf{a}\cdot\mathbf{x})}{\partial\mathbf{x}}=\dfrac{\partial(\mathbf{x}\cdot\mathbf{a})}{\partial\mathbf{x}}=$ <br><br> $\dfrac{\partial\mathbf{a}^\top\mathbf{x}}{\partial\mathbf{x}}=\dfrac{\partial\mathbf{x}^\top\mathbf{a}}{\partial\mathbf{x}}=$ | $\mathbf{a}^\top$ | $\mathbf{a}$ |
| **A** is not a function of **x** <br> **b** is not a function of **x** | $\dfrac{\partial\mathbf{b}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}}=$ | $\mathbf{b}^\top\mathbf{A}$ | $\mathbf{A}^\top\mathbf{b}$ |
| **A** is not a function of **x** | $\dfrac{\partial\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}}=$ | $\mathbf{x}^\top(\mathbf{A}+\mathbf{A}^\top)$ | $(\mathbf{A}+\mathbf{A}^\top)\mathbf{x}$ |
| **A** is not a function of **x** <br> **A** is symmetric | $\dfrac{\partial\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}}=$ | $2\mathbf{x}^\top\mathbf{A}$ | $2\mathbf{A}\mathbf{x}$ |
| **A** is not a function of **x** | $\dfrac{\partial^2\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}^2}=$ | $\mathbf{A}+\mathbf{A}^\top$ | |
| **A** is not a function of **x** <br> **A** is symmetric | $\dfrac{\partial^2\mathbf{x}^\top\mathbf{A}\mathbf{x}}{\partial\mathbf{x}^2}=$ | $2\mathbf{A}$ | |
| | $\dfrac{\partial(\mathbf{x}\cdot\mathbf{x})}{\partial\mathbf{x}}=\dfrac{\partial\mathbf{x}^\top\mathbf{x}}{\partial\mathbf{x}}=$ | $2\mathbf{x}^\top$ | $2\mathbf{x}$ |
| **a** is not a function of **x**, <br> **u** = **u**(**x**) | $\dfrac{\partial(\mathbf{a}\cdot\mathbf{u})}{\partial\mathbf{x}}=\dfrac{\partial\mathbf{a}^\top\mathbf{u}}{\partial\mathbf{x}}=$ | $\mathbf{a}^\top\dfrac{\partial\mathbf{u}}{\partial\mathbf{x}}$ <br><br> • assumes numerator layout of $\dfrac{\partial\mathbf{u}}{\partial\mathbf{x}}$ | $\dfrac{\partial\mathbf{u}}{\partial\mathbf{x}}\mathbf{a}$ <br><br> • assumes denominator layout of $\dfrac{\partial\mathbf{u}}{\partial\mathbf{x}}$ |
| **a**, **b** are not functions of **x** | $\dfrac{\partial\,\mathbf{a}^\top\mathbf{x}\mathbf{x}^\top\mathbf{b}}{\partial\,\mathbf{x}}=$ | $\mathbf{x}^\top(\mathbf{a}\mathbf{b}^\top+\mathbf{b}\mathbf{a}^\top)$ | $(\mathbf{a}\mathbf{b}^\top+\mathbf{b}\mathbf{a}^\top)\mathbf{x}$ |
| **A**, **b**, **C**, **D**, **e** are not functions of **x** | $\dfrac{\partial\,(\mathbf{A}\mathbf{x}+\mathbf{b})^\top\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e})}{\partial\,\mathbf{x}}=$ | $(\mathbf{D}\mathbf{x}+\mathbf{e})^\top\mathbf{C}^\top\mathbf{A}+(\mathbf{A}\mathbf{x}+\mathbf{b})^\top\mathbf{C}\mathbf{D}$ | $\mathbf{D}^\top\mathbf{C}^\top(\mathbf{A}\mathbf{x}+\mathbf{b})+\mathbf{A}^\top\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{e})$ |
| **a** is not a function of **x** | $\dfrac{\partial\,\|\mathbf{x}-\mathbf{a}\|}{\partial\,\mathbf{x}}=$ | $\dfrac{(\mathbf{x}-\mathbf{a})^\top}{\|\mathbf{x}-\mathbf{a}\|}$ | $\dfrac{\mathbf{x}-\mathbf{a}}{\|\mathbf{x}-\mathbf{a}\|}$ |

| Condition | Expression | Numerator layout, i.e. by y, result is column vector | Denominator layout, i.e. by $\mathbf{y}^\mathsf{T}$, result is row vector |
|---|---|---|---|
| **a** is not a function of $x$ | $\dfrac{\partial \mathbf{a}}{\partial x} =$ | $\mathbf{0}$[4] | |
| $a$ is not a function of $x$, $\mathbf{u} = \mathbf{u}(x)$ | $\dfrac{\partial a\mathbf{u}}{\partial x} =$ | $a\dfrac{\partial \mathbf{u}}{\partial x}$ | |
| **A** is not a function of $x$, $\mathbf{u} = \mathbf{u}(x)$ | $\dfrac{\partial \mathbf{A}\mathbf{u}}{\partial x} =$ | $\mathbf{A}\dfrac{\partial \mathbf{u}}{\partial x}$ | $\dfrac{\partial \mathbf{u}}{\partial x}\mathbf{A}^\mathsf{T}$ |
| $\mathbf{u} = \mathbf{u}(x)$ | $\dfrac{\partial \mathbf{u}^\mathsf{T}}{\partial x} =$ | $\left(\dfrac{\partial \mathbf{u}}{\partial x}\right)^\mathsf{T}$ | |
| $\mathbf{u} = \mathbf{u}(x)$, $\mathbf{v} = \mathbf{v}(x)$ | $\dfrac{\partial (\mathbf{u} + \mathbf{v})}{\partial x} =$ | $\dfrac{\partial \mathbf{u}}{\partial x} + \dfrac{\partial \mathbf{v}}{\partial x}$ | |
| $\mathbf{u} = \mathbf{u}(x)$, $\mathbf{v} = \mathbf{v}(x)$ | $\dfrac{\partial (\mathbf{u}^\mathsf{T} \times \mathbf{v})}{\partial x} =$ | $\left(\dfrac{\partial \mathbf{u}}{\partial x}\right)^\mathsf{T} \times \mathbf{v} + \mathbf{u}^\mathsf{T} \times \dfrac{\partial \mathbf{v}}{\partial x}$ | $\dfrac{\partial \mathbf{u}}{\partial x} \times \mathbf{v} + \mathbf{u}^\mathsf{T} \times \left(\dfrac{\partial \mathbf{v}}{\partial x}\right)^\mathsf{T}$ |
| $\mathbf{u} = \mathbf{u}(x)$ | $\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial x} =$ | $\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial x}$ <br> Assumes consistent matrix layout; see below. | $\dfrac{\partial \mathbf{u}}{\partial x}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$ |
| $\mathbf{u} = \mathbf{u}(x)$ | $\dfrac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial x} =$ | $\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{u}}{\partial x}$ <br> Assumes consistent matrix layout; see below. | $\dfrac{\partial \mathbf{u}}{\partial x}\dfrac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\dfrac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$ |
| $\mathbf{U} = \mathbf{U}(x)$, $\mathbf{v} = \mathbf{v}(x)$ | $\dfrac{\partial (\mathbf{U} \times \mathbf{v})}{\partial x} =$ | $\dfrac{\partial \mathbf{U}}{\partial x} \times \mathbf{v} + \mathbf{U} \times \dfrac{\partial \mathbf{v}}{\partial x}$ | $\mathbf{v}^\mathsf{T} \times \left(\dfrac{\partial \mathbf{U}}{\partial x}\right) + \dfrac{\partial \mathbf{v}}{\partial x} \times \mathbf{U}^\mathsf{T}$ |

## Identities: scalar-by-matrix $\dfrac{\partial y}{\partial \mathbf{X}}$

| Condition | Expression | Numerator layout, i.e. by $\mathbf{X}^{\mathsf{T}}$ | Denominator layout, i.e. by $\mathbf{X}$ |
|---|---|---|---|
| $a$ is not a function of $\mathbf{X}$ | $\dfrac{\partial a}{\partial \mathbf{X}} =$ | $\mathbf{0}^{\top}$ [6] | $\mathbf{0}$ [6] |
| $a$ is not a function of $\mathbf{X}$, $u = u(\mathbf{X})$ | $\dfrac{\partial au}{\partial \mathbf{X}} =$ | $a\dfrac{\partial u}{\partial \mathbf{X}}$ | |
| $u = u(\mathbf{X})$, $v = v(\mathbf{X})$ | $\dfrac{\partial (u + v)}{\partial \mathbf{X}} =$ | $\dfrac{\partial u}{\partial \mathbf{X}} + \dfrac{\partial v}{\partial \mathbf{X}}$ | |
| $u = u(\mathbf{X})$, $v = v(\mathbf{X})$ | $\dfrac{\partial uv}{\partial \mathbf{X}} =$ | $u\dfrac{\partial v}{\partial \mathbf{X}} + v\dfrac{\partial u}{\partial \mathbf{X}}$ | |
| $u = u(\mathbf{X})$ | $\dfrac{\partial g(u)}{\partial \mathbf{X}} =$ | $\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{X}}$ | |
| $u = u(\mathbf{X})$ | $\dfrac{\partial f(g(u))}{\partial \mathbf{X}} =$ | $\dfrac{\partial f(g)}{\partial g}\dfrac{\partial g(u)}{\partial u}\dfrac{\partial u}{\partial \mathbf{X}}$ | |
| $\mathbf{U} = \mathbf{U}(\mathbf{X})$ | [5] $\dfrac{\partial g(\mathbf{U})}{\partial X_{ij}} =$ | $\mathrm{tr}\left( \dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\dfrac{\partial \mathbf{U}}{\partial X_{ij}} \right)$ | $\mathrm{tr}\left( \left(\dfrac{\partial g(\mathbf{U})}{\partial \mathbf{U}}\right)^{\top} \dfrac{\partial \mathbf{U}}{\partial X_{ij}} \right)$ |
| | | Both forms assume *numerator* layout for $\dfrac{\partial \mathbf{U}}{\partial X_{ij}}$, i.e. mixed layout if denominator layout for $\mathbf{X}$ is being used. | |
| $a$ and $b$ are not functions of $\mathbf{X}$ | $\dfrac{\partial a^{\top}\mathbf{X}b}{\partial \mathbf{X}} =$ | $ba^{\top}$ | $ab^{\top}$ |
| $a$ and $b$ are not functions of $\mathbf{X}$ | $\dfrac{\partial a^{\top}\mathbf{X}^{\top}b}{\partial \mathbf{X}} =$ | $ab^{\top}$ | $ba^{\top}$ |
| $a$, $b$ and $\mathbf{C}$ are not functions of $\mathbf{X}$ | $\dfrac{\partial (\mathbf{X}a + b)^{\top}\mathbf{C}(\mathbf{X}a + b)}{\partial \mathbf{X}} =$ | $((\mathbf{C} + \mathbf{C}^{\top})(\mathbf{X}a + b)a^{\top})^{\top}$ | $(\mathbf{C} + \mathbf{C}^{\top})(\mathbf{X}a + b)a^{\top}$ |
| $a$, $b$ and $\mathbf{C}$ are not functions of $\mathbf{X}$ | $\dfrac{\partial (\mathbf{X}a)^{\top}\mathbf{C}(\mathbf{X}b)}{\partial \mathbf{X}} =$ | $(\mathbf{C}\mathbf{X}ba^{\top} + \mathbf{C}^{\top}\mathbf{X}ab^{\top})^{\top}$ | $\mathbf{C}\mathbf{X}ba^{\top} + \mathbf{C}^{\top}\mathbf{X}ab^{\top}$ |

Identities: matrix-by-scalar $\dfrac{\partial \mathbf{Y}}{\partial x}$

| Condition | Expression | Numerator layout, i.e. by Y |
|---|---|---|
| $\mathbf{U} = \mathbf{U}(x)$ | $\dfrac{\partial a\mathbf{U}}{\partial x} =$ | $a\dfrac{\partial \mathbf{U}}{\partial x}$ |
| **A**, **B** are not functions of $x$ <br> $\mathbf{U} = \mathbf{U}(x)$ | $\dfrac{\partial \mathbf{AUB}}{\partial x} =$ | $\mathbf{A}\dfrac{\partial \mathbf{U}}{\partial x}\mathbf{B}$ |
| $\mathbf{U} = \mathbf{U}(x),\ \mathbf{V} = \mathbf{V}(x)$ | $\dfrac{\partial (\mathbf{U} + \mathbf{V})}{\partial x} =$ | $\dfrac{\partial \mathbf{U}}{\partial x} + \dfrac{\partial \mathbf{V}}{\partial x}$ |
| $\mathbf{U} = \mathbf{U}(x),\ \mathbf{V} = \mathbf{V}(x)$ | $\dfrac{\partial (\mathbf{UV})}{\partial x} =$ | $\mathbf{U}\dfrac{\partial \mathbf{V}}{\partial x} + \dfrac{\partial \mathbf{U}}{\partial x}\mathbf{V}$ |
| $\mathbf{U} = \mathbf{U}(x),\ \mathbf{V} = \mathbf{V}(x)$ | $\dfrac{\partial (\mathbf{U} \otimes \mathbf{V})}{\partial x} =$ | $\mathbf{U} \otimes \dfrac{\partial \mathbf{V}}{\partial x} + \dfrac{\partial \mathbf{U}}{\partial x} \otimes \mathbf{V}$ |
| $\mathbf{U} = \mathbf{U}(x),\ \mathbf{V} = \mathbf{V}(x)$ | $\dfrac{\partial (\mathbf{U} \circ \mathbf{V})}{\partial x} =$ | $\mathbf{U} \circ \dfrac{\partial \mathbf{V}}{\partial x} + \dfrac{\partial \mathbf{U}}{\partial x} \circ \mathbf{V}$ |
| $\mathbf{U} = \mathbf{U}(x)$ | $\dfrac{\partial \mathbf{U}^{-1}}{\partial x} =$ | $-\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}$ |
| $\mathbf{U} = \mathbf{U}(x,y)$ | $\dfrac{\partial^2 \mathbf{U}^{-1}}{\partial x \partial y} =$ | $\mathbf{U}^{-1}\left(\dfrac{\partial \mathbf{U}}{\partial x}\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial y} - \dfrac{\partial^2 \mathbf{U}}{\partial x \partial y} + \dfrac{\partial \mathbf{U}}{\partial y}\mathbf{U}^{-1}\dfrac{\partial \mathbf{U}}{\partial x}\right)\mathbf{U}^{-1}$ |
| any matrix function defined <br> s its derivative, and **g′(X)** is | $\dfrac{\partial \mathbf{g}(x\mathbf{A})}{\partial x} =$ | $\mathbf{Ag}'(x\mathbf{A}) = \mathbf{g}'(x\mathbf{A})\mathbf{A}$ |
| **A** is not a function of $x$ | $\dfrac{\partial e^{x\mathbf{A}}}{\partial x} =$ | $\mathbf{A}e^{x\mathbf{A}} = e^{x\mathbf{A}}\mathbf{A}$ |

# 下课，谢谢大家！