

矩阵求导

圆眼睛的阿凡提哥哥

2020 年 12 月 1 日

标量、向量、矩阵间的求导共有 9 种可能：

$\partial \text{标量} / \partial \text{标量}$	$\partial \text{标量} / \partial \text{向量}$	$\partial \text{标量} / \partial \text{矩阵}$
$\partial \text{向量} / \partial \text{标量}$	$\partial \text{向量} / \partial \text{向量}$	$\partial \text{向量} / \partial \text{矩阵}$
$\partial \text{矩阵} / \partial \text{标量}$	$\partial \text{矩阵} / \partial \text{向量}$	$\partial \text{矩阵} / \partial \text{矩阵}$

表 1: 9 种求导情形

$\partial \text{标量} / \partial \text{标量}$ 就是我们熟悉的单变量微积分， $\partial \text{向量} / \partial \text{矩阵}$ 、 $\partial \text{矩阵} / \partial \text{向量}$ 、 $\partial \text{矩阵} / \partial \text{矩阵}$ 会涉及高阶张量，处理更为麻烦，因此本文只考虑剩下的 5 种情形。

设 $\mathbf{u} \in \mathbb{R}^l$ ， $\mathbf{U} \in \mathbb{R}^{m \times n}$ ，则向量、矩阵对标量求导的定义为

$$\frac{\partial \mathbf{u}}{\partial x} \triangleq \begin{bmatrix} \frac{\partial u_1}{\partial x} \\ \frac{\partial u_2}{\partial x} \\ \vdots \\ \frac{\partial u_l}{\partial x} \end{bmatrix}, \quad \frac{\partial \mathbf{U}}{\partial x} \triangleq \begin{bmatrix} \frac{\partial u_{11}}{\partial x} & \frac{\partial u_{12}}{\partial x} & \cdots & \frac{\partial u_{1n}}{\partial x} \\ \frac{\partial u_{21}}{\partial x} & \frac{\partial u_{22}}{\partial x} & \cdots & \frac{\partial u_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{m1}}{\partial x} & \frac{\partial u_{m2}}{\partial x} & \cdots & \frac{\partial u_{mn}}{\partial x} \end{bmatrix}$$

设 $\mathbf{x} \in \mathbb{R}^l$ ， $\mathbf{X} \in \mathbb{R}^{m \times n}$ ，则标量对向量、矩阵求导的定义为

$$\frac{\partial u}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial u}{\partial x_1} & \frac{\partial u}{\partial x_2} & \cdots & \frac{\partial u}{\partial x_l} \end{bmatrix}, \quad \frac{\partial u}{\partial \mathbf{X}} \triangleq \begin{bmatrix} \frac{\partial u}{\partial x_{11}} & \frac{\partial u}{\partial x_{21}} & \cdots & \frac{\partial u}{\partial x_{m1}} \\ \frac{\partial u}{\partial x_{12}} & \frac{\partial u}{\partial x_{22}} & \cdots & \frac{\partial u}{\partial x_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u}{\partial x_{1n}} & \frac{\partial u}{\partial x_{2n}} & \cdots & \frac{\partial u}{\partial x_{mn}} \end{bmatrix}$$

即向量、矩阵对标量求导的结果与分子尺寸相同，标量对向量、矩阵求导的结果与分母的转置尺寸相同。向量对向量求导的定义为 Jacobian 矩阵：

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \triangleq \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \cdots & \frac{\partial u_1}{\partial x_l} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \cdots & \frac{\partial u_2}{\partial x_l} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_l}{\partial x_1} & \frac{\partial u_l}{\partial x_2} & \cdots & \frac{\partial u_l}{\partial x_l} \end{bmatrix}$$

即行数与分子尺寸相同、列数与分母尺寸相同。

以上即为分子布局,其好处是链式法则跟单变量微积分中的顺序一样,坏处是计算标量值函数 $f(\mathbf{x})$ 关于向量变量 \mathbf{x} 的梯度时要多做做一个转置: $\nabla f = (\frac{\partial f}{\partial \mathbf{x}})^\top$, 因为我们更习惯梯度是列向量。分母布局的结果均是分子布局的转置,好处就是算梯度时不用做转置,坏处就是链式法则的顺序要完全反过来。

1 基本结果

以下结果根据定义和单变量微积分的求导法则都是显然的。

单变量微积分中常量的导数为零

$$\frac{\partial a}{\partial x} = 0$$

类似的这里有

$$\frac{\partial \mathbf{a}}{\partial x} = \mathbf{0}, \quad \frac{\partial a}{\partial \mathbf{x}} = \mathbf{0}^\top, \quad \frac{\partial \mathbf{a}}{\partial \mathbf{x}} = \mathbf{0}, \quad \frac{\partial \mathbf{A}}{\partial x} = \mathbf{0}, \quad \frac{\partial a}{\partial \mathbf{X}} = \mathbf{0}^\top$$

单变量微积分中常数标量乘的求导法则为

$$\frac{\partial au}{\partial x} = a \frac{\partial u}{\partial x}$$

类似的这里有

$$\frac{\partial a\mathbf{u}}{\partial x} = a \frac{\partial \mathbf{u}}{\partial x}, \quad \frac{\partial au}{\partial \mathbf{x}} = a \frac{\partial u}{\partial \mathbf{x}}, \quad \frac{\partial a\mathbf{u}}{\partial \mathbf{x}} = a \frac{\partial \mathbf{u}}{\partial \mathbf{x}}, \quad \frac{\partial a\mathbf{U}}{\partial x} = a \frac{\partial \mathbf{U}}{\partial x}, \quad \frac{\partial au}{\partial \mathbf{X}} = a \frac{\partial u}{\partial \mathbf{X}}$$

单变量微积分中加法的求导法则为

$$\frac{\partial(u+v)}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}$$

类似的这里有

$$\begin{aligned} \frac{\partial(\mathbf{u} + \mathbf{v})}{\partial x} &= \frac{\partial \mathbf{u}}{\partial x} + \frac{\partial \mathbf{v}}{\partial x}, & \frac{\partial(u+v)}{\partial \mathbf{x}} &= \frac{\partial u}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}, & \frac{\partial(\mathbf{u} + \mathbf{v})}{\partial \mathbf{x}} &= \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \\ \frac{\partial(\mathbf{U} + \mathbf{V})}{\partial x} &= \frac{\partial \mathbf{U}}{\partial x} + \frac{\partial \mathbf{V}}{\partial x}, & \frac{\partial(u+v)}{\partial \mathbf{X}} &= \frac{\partial u}{\partial \mathbf{X}} + \frac{\partial v}{\partial \mathbf{X}} \end{aligned}$$

单变量微积分中乘法的求导法则为

$$\frac{\partial uv}{\partial x} = \frac{\partial u}{\partial x} v + u \frac{\partial v}{\partial x}$$

类似的这里有

$$\begin{aligned} \frac{\partial \mathbf{u}\mathbf{v}}{\partial x} &= \frac{\partial \mathbf{u}}{\partial x} \mathbf{v} + \mathbf{u} \frac{\partial \mathbf{v}}{\partial x}, & \frac{\partial uv}{\partial \mathbf{x}} &= \frac{\partial u}{\partial \mathbf{x}} v + u \frac{\partial v}{\partial \mathbf{x}} \\ \frac{\partial \mathbf{U}\mathbf{V}}{\partial x} &= \frac{\partial \mathbf{U}}{\partial x} \mathbf{V} + \mathbf{U} \frac{\partial \mathbf{V}}{\partial x}, & \frac{\partial uv}{\partial \mathbf{X}} &= \frac{\partial u}{\partial \mathbf{X}} v + u \frac{\partial v}{\partial \mathbf{X}} \end{aligned}$$

其中第二行是因为

$$\left[\frac{\partial \mathbf{U}\mathbf{V}}{\partial x} \right]_{ij} = \frac{\partial \sum_k u_{ik} v_{kj}}{\partial x} = \sum_k \frac{\partial u_{ik}}{\partial x} v_{kj} + \sum_k u_{ik} \frac{\partial v_{kj}}{\partial x} = \left[\frac{\partial \mathbf{U}}{\partial x} \mathbf{V} \right]_{ij} + \left[\mathbf{U} \frac{\partial \mathbf{V}}{\partial x} \right]_{ij}$$

$$\begin{aligned}
&\Rightarrow \frac{\partial \mathbf{UV}}{\partial x} = \frac{\partial \mathbf{U}}{\partial x} \mathbf{V} + \mathbf{U} \frac{\partial \mathbf{V}}{\partial x} \\
\left[\frac{\partial uv}{\partial \mathbf{X}} \right]_{ij} &= \frac{\partial uv}{\partial x_{ji}} = \frac{\partial u}{\partial x_{ji}} v + u \frac{\partial v}{\partial x_{ji}} = \left[\frac{\partial u}{\partial \mathbf{X}} \right]_{ij} v + u \left[\frac{\partial v}{\partial \mathbf{X}} \right]_{ij} \\
&\Rightarrow \frac{\partial uv}{\partial \mathbf{X}} = \frac{\partial u}{\partial \mathbf{X}} v + u \frac{\partial v}{\partial \mathbf{X}}
\end{aligned}$$

第一行可看作第二行的特例。 $\partial \mathbf{uv} / \partial \mathbf{x}$ 有两种可能，一是 \mathbf{uv} 为标量，即两者的内积，后文会讲；二是 \mathbf{uv} 为矩阵，我们不考虑 ∂ 矩阵 / ∂ 向量 这种情形。

单变量微积分中有 $\partial x / \partial x = 1$ ，类似的这里有

$$\frac{\partial x_i}{\partial \mathbf{x}} = \mathbf{e}_i^\top, \quad \frac{\partial \mathbf{x}}{\partial x_i} = \mathbf{e}_i, \quad \frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{I}, \quad \frac{\partial x_{ij}}{\partial \mathbf{X}} = \mathbf{E}_{ji}, \quad \frac{\partial \mathbf{X}}{\partial x_{ij}} = \mathbf{E}_{ij}$$

其中 \mathbf{E}_{ij} 是 (i, j) 处为 1 其余为 0 的矩阵。

单变量微积分中的链式法则为

$$\frac{\partial g(u)}{\partial x} = \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial x}$$

类似的设 $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} = \mathbf{u}(\mathbf{x}) \in \mathbb{R}^m$, $\mathbf{g} : \mathbb{R}^m \mapsto \mathbb{R}^l$, 则

$$\underbrace{\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}}}_{l \times n} = \underbrace{\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}}_{l \times m} \underbrace{\frac{\partial \mathbf{u}}{\partial \mathbf{x}}}_{m \times n}$$

这是因为

$$\begin{aligned}
\left[\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} \right]_{ij} &= \frac{\partial [\mathbf{g}(\mathbf{u})]_i}{\partial x_j} = \sum_{k \in [m]} \frac{\partial [\mathbf{g}(\mathbf{u})]_i}{\partial u_k} \frac{\partial u_k}{\partial x_j} = \frac{\partial [\mathbf{g}(\mathbf{u})]_i}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial x_j} = \left[\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \right]_{i,:} \left[\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right]_{:,j} = \left[\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right]_{i,j} \\
&\Rightarrow \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}}
\end{aligned}$$

注意若 $n = m = l = 1$, 就退化成了单变量的链式法则。

设 $u = u(\mathbf{X})$, $g : \mathbb{R} \mapsto \mathbb{R}$, 则

$$\frac{\partial g(u)}{\partial \mathbf{X}} = \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{X}}$$

这是因为

$$\left[\frac{\partial g(u)}{\partial \mathbf{X}} \right]_{ij} = \frac{\partial g(u)}{\partial x_{ji}} = \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial x_{ji}} = \frac{\partial g(u)}{\partial u} \left[\frac{\partial u}{\partial \mathbf{X}} \right]_{ij} \Rightarrow \frac{\partial g(u)}{\partial \mathbf{X}} = \frac{\partial g(u)}{\partial u} \frac{\partial u}{\partial \mathbf{X}}$$

设 $\mathbf{U} = \mathbf{U}(x) \in \mathbb{R}^{m \times n}$, $g : \mathbb{R}^{m \times n} \mapsto \mathbb{R}$, 则

$$\frac{\partial g(\mathbf{U})}{\partial x} = \sum_{p,q} \frac{\partial g(\mathbf{U})}{\partial u_{pq}} \frac{\partial u_{pq}}{\partial x} = \text{tr} \left(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial x} \right)$$

其中第二个等号是因为

$$\text{tr}(\mathbf{A}^\top \mathbf{B}) = \sum_q \sum_p a_{pq} b_{pq} = \sum_{p,q} a_{pq} b_{pq}$$

2 向量对标量求导

矩阵和向量的乘积是向量，若 \mathbf{A} 与 \mathbf{x} 无关，易知有

$$\begin{aligned}\left[\frac{\partial \mathbf{A}\mathbf{u}}{\partial x}\right]_i &= \frac{\partial [\mathbf{A}\mathbf{u}]_i}{\partial x} = \frac{\partial \sum_k a_{ik} u_k}{\partial x} = \sum_k a_{ik} \frac{\partial u_k}{\partial x} = \left[\mathbf{A} \frac{\partial \mathbf{u}}{\partial x}\right]_i \Rightarrow \frac{\partial \mathbf{A}\mathbf{u}}{\partial x} = \mathbf{A} \frac{\partial \mathbf{u}}{\partial x} \\ \left[\frac{\partial \mathbf{u}^\top \mathbf{A}}{\partial x}\right]_i &= \frac{\partial [\mathbf{u}^\top \mathbf{A}]_i}{\partial x} = \frac{\partial [\mathbf{A}^\top \mathbf{u}]_i}{\partial x} = \left[\mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial x}\right]_i \Rightarrow \frac{\partial \mathbf{u}^\top \mathbf{A}}{\partial x} = \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial x}\end{aligned}$$

向量的外积也是向量，记 $\mathbf{u} = [u_1(x); u_2(x); u_3(x)]$ ， $\mathbf{v} = [v_1(x); v_2(x); v_3(x)]$ ，则

$$\mathbf{u}^\top \times \mathbf{v} = \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix}$$

于是

$$\frac{\partial (\mathbf{u}^\top \times \mathbf{v})}{\partial x} = \begin{bmatrix} \frac{\partial u_2}{\partial x} v_3 - \frac{\partial u_3}{\partial x} v_2 + u_2 \frac{\partial v_3}{\partial x} - u_3 \frac{\partial v_2}{\partial x} \\ \frac{\partial u_3}{\partial x} v_1 - \frac{\partial u_1}{\partial x} v_3 + u_3 \frac{\partial v_1}{\partial x} - u_1 \frac{\partial v_3}{\partial x} \\ \frac{\partial u_1}{\partial x} v_2 - \frac{\partial u_2}{\partial x} v_1 + u_1 \frac{\partial v_2}{\partial x} - u_2 \frac{\partial v_1}{\partial x} \end{bmatrix} = \left(\frac{\partial \mathbf{u}}{\partial x}\right)^\top \times \mathbf{v} + \mathbf{u}^\top \times \frac{\partial \mathbf{v}}{\partial x}$$

3 标量对向量求导

二次型是标量，设 \mathbf{A} 与 \mathbf{x} 无关，易知有

$$\begin{aligned}\left[\frac{\partial \mathbf{u}^\top \mathbf{A}\mathbf{v}}{\partial x}\right]_i &= \frac{\partial \mathbf{u}^\top \mathbf{A}\mathbf{v}}{\partial x_i} = \frac{\partial \sum_j \sum_k u_j a_{jk} v_k}{\partial x_i} = \sum_j \sum_k u_j a_{jk} \frac{\partial v_k}{\partial x_i} + \sum_j \sum_k \frac{\partial u_j}{\partial x_i} a_{jk} v_k \\ &= \mathbf{u}^\top \mathbf{A} \frac{\partial \mathbf{v}}{\partial x_i} + \mathbf{v}^\top \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial x_i} = \left[\mathbf{u}^\top \mathbf{A} \frac{\partial \mathbf{v}}{\partial x}\right]_i + \left[\mathbf{v}^\top \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial x}\right]_i \\ &\Rightarrow \frac{\partial \mathbf{u}^\top \mathbf{A}\mathbf{v}}{\partial x} = \mathbf{u}^\top \mathbf{A} \frac{\partial \mathbf{v}}{\partial x} + \mathbf{v}^\top \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial x}\end{aligned}$$

特别的，

- 取 $\mathbf{A} = \mathbf{I}$ ，则

$$\frac{\partial \mathbf{u}^\top \mathbf{v}}{\partial x} = \mathbf{u}^\top \frac{\partial \mathbf{v}}{\partial x} + \mathbf{v}^\top \frac{\partial \mathbf{u}}{\partial x}$$

进一步若 $\mathbf{u} = \mathbf{a}$ 与 \mathbf{x} 无关，则

$$\frac{\partial \mathbf{a}^\top \mathbf{v}}{\partial x} = \mathbf{a}^\top \frac{\partial \mathbf{v}}{\partial x}, \quad \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial x} = \mathbf{a}^\top \frac{\partial \mathbf{x}}{\partial x} = \mathbf{a}^\top, \quad \frac{\partial \mathbf{b}^\top \mathbf{A}\mathbf{x}}{\partial x} = \mathbf{b}^\top \mathbf{A}$$

- 取 $\mathbf{u} = \mathbf{v} = \mathbf{x}$ ，则

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial x} = \mathbf{x}^\top \mathbf{A} \frac{\partial \mathbf{x}}{\partial x} + \mathbf{x}^\top \mathbf{A}^\top \frac{\partial \mathbf{x}}{\partial x} = \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top) \stackrel{\text{若}\mathbf{A}\text{对称}}{=} 2\mathbf{x}^\top \mathbf{A}$$

进一步若 $\mathbf{A} = \mathbf{I}$ ，则

$$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial x} = \frac{\partial \|\mathbf{x}\|^2}{\partial x} = 2\mathbf{x}^\top$$

- 若 $\mathbf{A} = \mathbf{b}\mathbf{a}^\top$, 则

$$\frac{\partial \mathbf{x}^\top \mathbf{b} \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^\top \mathbf{x} \mathbf{x}^\top \mathbf{b}}{\partial \mathbf{x}} = \mathbf{x}^\top (\mathbf{a} \mathbf{b}^\top + \mathbf{b} \mathbf{a}^\top)$$

- 更一般的有

$$\begin{aligned} \frac{\partial (\mathbf{A}\mathbf{x} + \mathbf{b})^\top \mathbf{C}(\mathbf{D}\mathbf{x} + \mathbf{e})}{\partial \mathbf{x}} &= \frac{\partial (\mathbf{x}^\top \mathbf{A}^\top \mathbf{C} \mathbf{D} \mathbf{x} + \mathbf{b}^\top \mathbf{C} \mathbf{D} \mathbf{x} + \mathbf{x}^\top \mathbf{A}^\top \mathbf{C} \mathbf{e} + \mathbf{b}^\top \mathbf{e})}{\partial \mathbf{x}} \\ &= \mathbf{x}^\top (\mathbf{A}^\top \mathbf{C} \mathbf{D} + \mathbf{D}^\top \mathbf{C}^\top \mathbf{A}) + \mathbf{b}^\top \mathbf{C} \mathbf{D} + \mathbf{e}^\top \mathbf{C}^\top \mathbf{A} \\ &= (\mathbf{D}\mathbf{x} + \mathbf{e})^\top \mathbf{C}^\top \mathbf{A} + (\mathbf{A}\mathbf{x} + \mathbf{b})^\top \mathbf{C} \mathbf{D} \end{aligned}$$

范数也是标量, 若 \mathbf{a} 与 \mathbf{x} 无关, 则

$$\begin{aligned} \left[\frac{\partial \|\mathbf{x} - \mathbf{a}\|}{\partial \mathbf{x}} \right]_i &= \frac{\partial \|\mathbf{x} - \mathbf{a}\|}{\partial x_i} = \frac{\partial \sqrt{\sum_j (x_j - a_j)^2}}{\partial x_i} = \frac{1}{2} \frac{2(x_i - a_i)}{\sqrt{\sum_j (x_j - a_j)^2}} = \frac{x_i - a_i}{\|\mathbf{x} - \mathbf{a}\|} \\ &\Rightarrow \frac{\partial \|\mathbf{x} - \mathbf{a}\|}{\partial \mathbf{x}} = \frac{(\mathbf{x} - \mathbf{a})^\top}{\|\mathbf{x} - \mathbf{a}\|} \end{aligned}$$

4 向量对向量求导

若 \mathbf{A} 与 \mathbf{x} 无关, 易知有

$$\begin{aligned} \left[\frac{\partial \mathbf{A} \mathbf{u}}{\partial \mathbf{x}} \right]_{ij} &= \frac{\partial [\mathbf{A} \mathbf{u}]_i}{\partial x_j} = \frac{\partial \sum_k a_{ik} u_k}{\partial x_j} = \sum_k a_{ik} \frac{\partial u_k}{\partial x_j} = \left[\mathbf{A} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right]_{ij} \Rightarrow \frac{\partial \mathbf{A} \mathbf{u}}{\partial \mathbf{x}} = \mathbf{A} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \\ \left[\frac{\partial \mathbf{u}^\top \mathbf{A}}{\partial \mathbf{x}} \right]_{ij} &= \frac{\partial [\mathbf{u}^\top \mathbf{A}]_i}{\partial x_j} = \frac{\partial [\mathbf{A}^\top \mathbf{u}]_i}{\partial x_j} = \left[\mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right]_{ij} \Rightarrow \frac{\partial \mathbf{u}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}^\top \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \end{aligned}$$

特别的, 若 $\mathbf{u} = \mathbf{x}$, 则

$$\frac{\partial \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}, \quad \frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}^\top \frac{\partial \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$$

若 $v = v(\mathbf{x})$, 则

$$\left[\frac{\partial v \mathbf{u}}{\partial \mathbf{x}} \right]_{ij} = \frac{\partial v u_i}{\partial x_j} = v \frac{\partial u_i}{\partial x_j} + u_i \frac{\partial v}{\partial x_j} = v \left[\frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right]_{ij} + \left[\mathbf{u} \frac{\partial v}{\partial \mathbf{x}} \right]_{ij} \Rightarrow \frac{\partial v \mathbf{u}}{\partial \mathbf{x}} = v \frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u} \frac{\partial v}{\partial \mathbf{x}}$$

注意第一项是标量乘以 Jacobian 矩阵, 第二项是列向量乘以行向量。

5 矩阵对标量求导

若 $u = u(\mathbf{x})$, $\mathbf{V} = \mathbf{V}(\mathbf{x})$, 则

$$\left[\frac{\partial u \mathbf{V}}{\partial \mathbf{x}} \right]_{ij} = \frac{\partial u v_{ij}}{\partial x} = \frac{\partial u}{\partial x} v_{ij} + u \frac{\partial v_{ij}}{\partial x} = \frac{\partial u}{\partial \mathbf{x}} [\mathbf{V}]_{ij} + u \left[\frac{\partial \mathbf{V}}{\partial \mathbf{x}} \right]_{ij} \Rightarrow \frac{\partial u \mathbf{V}}{\partial \mathbf{x}} = \frac{\partial u}{\partial \mathbf{x}} \mathbf{V} + u \frac{\partial \mathbf{V}}{\partial \mathbf{x}}$$

若乘积求导法则中的 \mathbf{U} 或 \mathbf{V} 可继续分解为 x 相关项的乘积, 例如 $\mathbf{V} \leftarrow \mathbf{V}\mathbf{W}$, 则

$$\frac{\partial \mathbf{U}\mathbf{V}\mathbf{W}}{\partial x} = \frac{\partial \mathbf{U}}{\partial x} \mathbf{V}\mathbf{W} + \mathbf{U} \frac{\partial \mathbf{V}\mathbf{W}}{\partial x} = \frac{\partial \mathbf{U}}{\partial x} \mathbf{V}\mathbf{W} + \mathbf{U} \left(\frac{\partial \mathbf{V}}{\partial x} \mathbf{W} + \mathbf{V} \frac{\partial \mathbf{W}}{\partial x} \right) = \frac{\partial \mathbf{U}}{\partial x} \mathbf{V}\mathbf{W} + \mathbf{U} \frac{\partial \mathbf{V}}{\partial x} \mathbf{W} + \mathbf{U}\mathbf{V} \frac{\partial \mathbf{W}}{\partial x} \quad (1)$$

据此可知若 \mathbf{A} 、 \mathbf{B} 与 x 无关, 则

$$\frac{\partial \mathbf{A}\mathbf{U}\mathbf{B}}{\partial x} = \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} \mathbf{B}$$

当 \mathbf{U} 为方阵、 n 为正整数时有

$$\frac{\partial \mathbf{U}^n}{\partial x} = \mathbf{U}^{n-1} \frac{\partial \mathbf{U}}{\partial x} + \mathbf{U}^{n-2} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U} + \cdots + \mathbf{U} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{n-2} + \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{n-1} = \sum_{i \in [n]} \mathbf{U}^{i-1} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{n-i} \quad (2)$$

令乘积求导法则中的 $\mathbf{V} = \mathbf{U}^{-1}$ 可得

$$\mathbf{0} = \frac{\partial \mathbf{I}}{\partial x} = \frac{\partial \mathbf{U}\mathbf{U}^{-1}}{\partial x} = \mathbf{U} \frac{\partial \mathbf{U}^{-1}}{\partial x} + \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} \Rightarrow \frac{\partial \mathbf{U}^{-1}}{\partial x} = -\mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} \quad (3)$$

进一步结合式 (1) 可得 Hessian 矩阵

$$\begin{aligned} \frac{\partial^2 \mathbf{U}^{-1}}{\partial x \partial y} &= \frac{\partial}{\partial y} \left(-\mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} \right) = -\frac{\partial \mathbf{U}^{-1}}{\partial y} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} - \mathbf{U}^{-1} \frac{\partial^2 \mathbf{U}}{\partial x \partial y} \mathbf{U}^{-1} - \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \frac{\partial \mathbf{U}^{-1}}{\partial y} \\ &= \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial y} \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} - \mathbf{U}^{-1} \frac{\partial^2 \mathbf{U}}{\partial x \partial y} \mathbf{U}^{-1} + \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial y} \mathbf{U}^{-1} \\ &= \mathbf{U}^{-1} \left(\frac{\partial \mathbf{U}}{\partial y} \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial x} - \frac{\partial^2 \mathbf{U}}{\partial x \partial y} + \frac{\partial \mathbf{U}}{\partial x} \mathbf{U}^{-1} \frac{\partial \mathbf{U}}{\partial y} \right) \mathbf{U}^{-1} \end{aligned}$$

矩阵除了常规的乘积外, 还有 Kronecker 积和 Hadamard 积。设 $\mathbf{U} \in \mathbb{R}^{m \times n}$, $\mathbf{V} \in \mathbb{R}^{p \times q}$, 则

$$\begin{aligned} \frac{\partial \mathbf{U} \otimes \mathbf{V}}{\partial x} &= \begin{bmatrix} \frac{\partial u_{11}}{\partial x} \mathbf{V} & \frac{\partial u_{12}}{\partial x} \mathbf{V} & \cdots & \frac{\partial u_{1n}}{\partial x} \mathbf{V} \\ \frac{\partial u_{21}}{\partial x} \mathbf{V} & \frac{\partial u_{22}}{\partial x} \mathbf{V} & \cdots & \frac{\partial u_{2n}}{\partial x} \mathbf{V} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{m1}}{\partial x} \mathbf{V} & \frac{\partial u_{m2}}{\partial x} \mathbf{V} & \cdots & \frac{\partial u_{mn}}{\partial x} \mathbf{V} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial u_{11}}{\partial x} \mathbf{V} + u_{11} \frac{\partial \mathbf{V}}{\partial x} & \frac{\partial u_{12}}{\partial x} \mathbf{V} + u_{12} \frac{\partial \mathbf{V}}{\partial x} & \cdots & \frac{\partial u_{1n}}{\partial x} \mathbf{V} + u_{1n} \frac{\partial \mathbf{V}}{\partial x} \\ \frac{\partial u_{21}}{\partial x} \mathbf{V} + u_{21} \frac{\partial \mathbf{V}}{\partial x} & \frac{\partial u_{22}}{\partial x} \mathbf{V} + u_{22} \frac{\partial \mathbf{V}}{\partial x} & \cdots & \frac{\partial u_{2n}}{\partial x} \mathbf{V} + u_{2n} \frac{\partial \mathbf{V}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{m1}}{\partial x} \mathbf{V} + u_{m1} \frac{\partial \mathbf{V}}{\partial x} & \frac{\partial u_{m2}}{\partial x} \mathbf{V} + u_{m2} \frac{\partial \mathbf{V}}{\partial x} & \cdots & \frac{\partial u_{mn}}{\partial x} \mathbf{V} + u_{mn} \frac{\partial \mathbf{V}}{\partial x} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial u_{11}}{\partial x} \mathbf{V} & \frac{\partial u_{12}}{\partial x} \mathbf{V} & \cdots & \frac{\partial u_{1n}}{\partial x} \mathbf{V} \\ \frac{\partial u_{21}}{\partial x} \mathbf{V} & \frac{\partial u_{22}}{\partial x} \mathbf{V} & \cdots & \frac{\partial u_{2n}}{\partial x} \mathbf{V} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{m1}}{\partial x} \mathbf{V} & \frac{\partial u_{m2}}{\partial x} \mathbf{V} & \cdots & \frac{\partial u_{mn}}{\partial x} \mathbf{V} \end{bmatrix} + \begin{bmatrix} u_{11} \frac{\partial \mathbf{V}}{\partial x} & u_{12} \frac{\partial \mathbf{V}}{\partial x} & \cdots & u_{1n} \frac{\partial \mathbf{V}}{\partial x} \\ u_{21} \frac{\partial \mathbf{V}}{\partial x} & u_{22} \frac{\partial \mathbf{V}}{\partial x} & \cdots & u_{2n} \frac{\partial \mathbf{V}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} \frac{\partial \mathbf{V}}{\partial x} & u_{m2} \frac{\partial \mathbf{V}}{\partial x} & \cdots & u_{mn} \frac{\partial \mathbf{V}}{\partial x} \end{bmatrix} \\ &= \frac{\partial \mathbf{U}}{\partial x} \otimes \mathbf{V} + \mathbf{U} \otimes \frac{\partial \mathbf{V}}{\partial x} \end{aligned}$$

设 $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{m \times n}$, 则

$$\begin{aligned} \frac{\partial \mathbf{U} \circ \mathbf{V}}{\partial x} &= \begin{bmatrix} \frac{\partial u_{11} v_{11}}{\partial x} & \frac{\partial u_{12} v_{12}}{\partial x} & \cdots & \frac{\partial u_{1n} v_{1n}}{\partial x} \\ \frac{\partial u_{21} v_{21}}{\partial x} & \frac{\partial u_{22} v_{22}}{\partial x} & \cdots & \frac{\partial u_{2n} v_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{m1} v_{m1}}{\partial x} & \frac{\partial u_{m2} v_{m2}}{\partial x} & \cdots & \frac{\partial u_{mn} v_{mn}}{\partial x} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial u_{11}}{\partial x} v_{11} & \frac{\partial u_{12}}{\partial x} v_{12} & \cdots & \frac{\partial u_{1n}}{\partial x} v_{1n} \\ \frac{\partial u_{21}}{\partial x} v_{21} & \frac{\partial u_{22}}{\partial x} v_{22} & \cdots & \frac{\partial u_{2n}}{\partial x} v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial u_{m1}}{\partial x} v_{m1} & \frac{\partial u_{m2}}{\partial x} v_{m2} & \cdots & \frac{\partial u_{mn}}{\partial x} v_{mn} \end{bmatrix} + \begin{bmatrix} u_{11} \frac{\partial v_{11}}{\partial x} & u_{12} \frac{\partial v_{12}}{\partial x} & \cdots & u_{1n} \frac{\partial v_{1n}}{\partial x} \\ u_{21} \frac{\partial v_{21}}{\partial x} & u_{22} \frac{\partial v_{22}}{\partial x} & \cdots & u_{2n} \frac{\partial v_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} \frac{\partial v_{m1}}{\partial x} & u_{m2} \frac{\partial v_{m2}}{\partial x} & \cdots & u_{mn} \frac{\partial v_{mn}}{\partial x} \end{bmatrix} \\ &= \frac{\partial \mathbf{U}}{\partial x} \circ \mathbf{V} + \mathbf{U} \circ \frac{\partial \mathbf{V}}{\partial x} \end{aligned}$$

设多项式函数 $g(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots$, 则 $g'(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots$, 设 \mathbf{A} 为与 x 无关的方阵, 记

$$\begin{aligned} g(x\mathbf{A}) &= a_0 \mathbf{I} + a_1 x \mathbf{A} + a_2 x^2 \mathbf{A}^2 + a_3 x^3 \mathbf{A}^3 + \cdots \\ g'(x\mathbf{A}) &= a_1 \mathbf{I} + 2a_2 x \mathbf{A} + 3a_3 x^2 \mathbf{A}^2 + \cdots \end{aligned}$$

易知有

$$\begin{aligned} \frac{\partial g(x\mathbf{A})}{\partial x} &= a_1 \mathbf{A} + 2a_2 x \mathbf{A}^2 + 3a_3 x^2 \mathbf{A}^3 + \cdots \\ &= \mathbf{A}(a_1 \mathbf{I} + 2a_2 x \mathbf{A} + 3a_3 x^2 \mathbf{A}^2 + \cdots) = \mathbf{A} g'(x\mathbf{A}) \\ &= (a_1 \mathbf{I} + 2a_2 x \mathbf{A} + 3a_3 x^2 \mathbf{A}^2 + \cdots) \mathbf{A} = g'(x\mathbf{A}) \mathbf{A} \end{aligned}$$

对于 e^x 、 $\ln x$ 、 $\sin x$ 、 $\cos x$, 上式依然适用, 例如

$$\frac{\partial e^{x\mathbf{A}}}{\partial x} = \mathbf{A} e^{x\mathbf{A}} = e^{x\mathbf{A}} \mathbf{A}$$

6 标量对矩阵求导

矩阵常见的标量函数有迹和行列式。

6.1 迹对矩阵求导

若 a 与 \mathbf{X} 无关, $\mathbf{U} = \mathbf{U}(\mathbf{X})$, $\mathbf{V} = \mathbf{V}(\mathbf{X})$, 则以下结论是显然的:

$$\frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}, \quad \frac{\partial \text{tr}(\mathbf{U} + \mathbf{V})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{U})}{\partial \mathbf{X}} + \frac{\partial \text{tr}(\mathbf{V})}{\partial \mathbf{X}}, \quad \frac{\partial \text{tr}(a\mathbf{U})}{\partial \mathbf{X}} = a \frac{\partial \text{tr}(\mathbf{U})}{\partial \mathbf{X}}$$

对于乘积有

$$\left[\frac{\partial \text{tr}(\mathbf{UV})}{\partial \mathbf{X}} \right]_{ij} = \frac{\partial \text{tr}(\mathbf{UV})}{\partial x_{ji}} = \frac{\partial \sum_p \sum_q u_{pq} v_{qp}}{\partial x_{ji}} = \sum_p \sum_q \left(\frac{\partial u_{pq}}{\partial x_{ji}} v_{qp} + u_{pq} \frac{\partial v_{qp}}{\partial x_{ji}} \right)$$

$$= \text{tr} \left(\frac{\partial \mathbf{U}}{\partial x_{ji}} \mathbf{V} \right) + \text{tr} \left(\mathbf{U} \frac{\partial \mathbf{V}}{\partial x_{ji}} \right) = \text{tr} \left(\frac{\partial \mathbf{UV}}{\partial x_{ji}} \right)$$

由此可知迹和求导的顺序可以交换。特别的,

- 取 $\mathbf{U} = \mathbf{A}$ 与 \mathbf{X} 无关, $\mathbf{V} = \mathbf{X}$, 则

$$\left[\frac{\partial \text{tr}(\mathbf{AX})}{\partial \mathbf{X}} \right]_{ij} = \text{tr} \left(\mathbf{A} \frac{\partial \mathbf{X}}{\partial x_{ji}} \right) = \text{tr}(\mathbf{A} \mathbf{E}_{ji}) = a_{ij} \implies \frac{\partial \text{tr}(\mathbf{AX})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{XA})}{\partial \mathbf{X}} = \mathbf{A}$$

进一步若 \mathbf{B} 与 \mathbf{X} 也无关, 则

$$\frac{\partial \text{tr}(\mathbf{AXB})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{BAX})}{\partial \mathbf{X}} = \mathbf{BA}$$

- 取 $\mathbf{U} = \mathbf{A}$ 与 \mathbf{X} 无关, $\mathbf{V} = \mathbf{X}^\top$, 则

$$\frac{\partial \text{tr}(\mathbf{AX}^\top)}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{XA}^\top)}{\partial \mathbf{X}} = \mathbf{A}^\top$$

- 取 $\mathbf{U} = \mathbf{A}$ 与 \mathbf{X} 无关, $\mathbf{V} = \mathbf{XX}^\top$, 则

$$\begin{aligned} \left[\frac{\partial \text{tr}(\mathbf{AXX}^\top)}{\partial \mathbf{X}} \right]_{ij} &= \text{tr} \left(\mathbf{A} \frac{\partial \mathbf{XX}^\top}{\partial x_{ji}} \right) = \text{tr} \left(\mathbf{A} \frac{\partial \mathbf{X}}{\partial x_{ji}} \mathbf{X}^\top \right) + \text{tr} \left(\mathbf{AX} \frac{\partial \mathbf{X}^\top}{\partial x_{ji}} \right) \\ &= \text{tr}(\mathbf{A} \mathbf{E}_{ji} \mathbf{X}^\top) + \text{tr}(\mathbf{AX} \mathbf{E}_{ij}) \\ &= [\mathbf{X}^\top \mathbf{A}]_{ij} + [\mathbf{AX}]_{ji} \end{aligned}$$

从而

$$\frac{\partial \text{tr}(\mathbf{AXX}^\top)}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^\top \mathbf{AX})}{\partial \mathbf{X}} = \mathbf{X}^\top \mathbf{A} + \mathbf{X}^\top \mathbf{A}^\top = \mathbf{X}^\top (\mathbf{A} + \mathbf{A}^\top)$$

- 取 $\mathbf{U} = \mathbf{A}$ 与 \mathbf{X} 无关, $\mathbf{V} = \mathbf{X}^{-1}$, 结合式 (3) 可得

$$\begin{aligned} \left[\frac{\partial \text{tr}(\mathbf{AX}^{-1})}{\partial \mathbf{X}} \right]_{ij} &= \text{tr} \left(\mathbf{A} \frac{\partial \mathbf{X}^{-1}}{\partial x_{ji}} \right) = \text{tr} \left(-\mathbf{AX}^{-1} \frac{\partial \mathbf{X}}{\partial x_{ji}} \mathbf{X}^{-1} \right) = -\text{tr}(\mathbf{X}^{-1} \mathbf{AX}^{-1} \mathbf{E}_{ji}) = -[\mathbf{X}^{-1} \mathbf{AX}^{-1}]_{ij} \\ &\implies \frac{\partial \text{tr}(\mathbf{AX}^{-1})}{\partial \mathbf{X}} = \frac{\partial \text{tr}(\mathbf{X}^{-1} \mathbf{A})}{\partial \mathbf{X}} = -\mathbf{X}^{-1} \mathbf{AX}^{-1} \end{aligned}$$

- 取 $\mathbf{U} = \mathbf{AXB}$, $\mathbf{V} = \mathbf{X}^\top \mathbf{C}$, 其中 \mathbf{A} 、 \mathbf{B} 、 \mathbf{C} 与 \mathbf{X} 无关, 则

$$\begin{aligned} \left[\frac{\partial \text{tr}(\mathbf{AXBX}^\top \mathbf{C})}{\partial \mathbf{X}} \right]_{ij} &= \text{tr} \left(\frac{\partial \mathbf{AXB}}{\partial x_{ji}} \mathbf{X}^\top \mathbf{C} \right) + \text{tr} \left(\mathbf{AXB} \frac{\partial \mathbf{X}^\top \mathbf{C}}{\partial x_{ji}} \right) \\ &= \text{tr}(\mathbf{A} \mathbf{E}_{ji} \mathbf{BX}^\top \mathbf{C}) + \text{tr}(\mathbf{AXB} \mathbf{E}_{ij} \mathbf{C}) \\ &= [\mathbf{BX}^\top \mathbf{CA}]_{ij} + [\mathbf{CAXB}]_{ji} \\ &\implies \frac{\partial \text{tr}(\mathbf{AXBX}^\top \mathbf{C})}{\partial \mathbf{X}} = \mathbf{BX}^\top \mathbf{CA} + \mathbf{B}^\top \mathbf{X}^\top \mathbf{A}^\top \mathbf{C}^\top \end{aligned}$$

- 取 $\mathbf{U} = \mathbf{A}$ 与 \mathbf{X} 无关, $\mathbf{V} = \mathbf{X}^n$, 其中 n 是正整数, 结合式 (2) 可得

$$\left[\frac{\partial \text{tr}(\mathbf{AX}^n)}{\partial \mathbf{X}} \right]_{ij} = \text{tr} \left(\mathbf{A} \frac{\partial \mathbf{X}^n}{\partial x_{ji}} \right) = \text{tr} \left(\mathbf{A} \sum_{k \in [n]} \mathbf{X}^{k-1} \frac{\partial \mathbf{X}}{\partial x_{ji}} \mathbf{X}^{n-k} \right) = \sum_{k \in [n]} \text{tr} \left(\mathbf{AX}^{k-1} \frac{\partial \mathbf{X}}{\partial x_{ji}} \mathbf{X}^{n-k} \right)$$

$$\begin{aligned}
&= \sum_{k \in [n]} \text{tr}(\mathbf{X}^{n-k} \mathbf{A} \mathbf{X}^{k-1} \mathbf{E}_{ji}) = \sum_{k \in [n]} [\mathbf{X}^{n-k} \mathbf{A} \mathbf{X}^{k-1}]_{ij} \\
&\Rightarrow \frac{\partial \text{tr}(\mathbf{A} \mathbf{X}^n)}{\partial \mathbf{X}} = \sum_{k \in [n]} \mathbf{X}^{n-k} \mathbf{A} \mathbf{X}^{k-1}
\end{aligned}$$

进一步若 $\mathbf{A} = \mathbf{I}$, 则

$$\frac{\partial \text{tr}(\mathbf{X}^n)}{\partial \mathbf{X}} = \sum_{k \in [n]} \mathbf{X}^{n-k} \mathbf{X}^{k-1} = \sum_{k \in [n]} \mathbf{X}^{n-1} = n \mathbf{X}^{n-1}$$

不难发现形式上和单变量的求导公式 $\partial x^n / \partial x = n x^{n-1}$ 是一样的。类似的记

$$\begin{aligned}
e^{\mathbf{X}} &= \mathbf{I} + \mathbf{X} + \frac{\mathbf{X}^2}{2!} + \frac{\mathbf{X}^3}{3!} + \cdots \\
\sin \mathbf{X} &= \mathbf{X} - \frac{\mathbf{X}^3}{3!} + \frac{\mathbf{X}^5}{5!} - \cdots \\
\cos \mathbf{X} &= \mathbf{I} - \frac{\mathbf{X}^2}{2!} + \frac{\mathbf{X}^4}{4!} - \frac{\mathbf{X}^6}{6!} + \cdots
\end{aligned}$$

结合式 (2) 可得

$$\begin{aligned}
\frac{\partial \text{tr}(e^{\mathbf{X}})}{\partial \mathbf{X}} &= \frac{\partial}{\partial \mathbf{X}} \text{tr} \left(\mathbf{I} + \mathbf{X} + \frac{\mathbf{X}^2}{2!} + \frac{\mathbf{X}^3}{3!} + \cdots \right) \\
&= \frac{\partial \text{tr}(\mathbf{I})}{\partial \mathbf{X}} + \frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} + \frac{1}{2!} \frac{\partial \text{tr}(\mathbf{X}^2)}{\partial \mathbf{X}} + \frac{1}{3!} \frac{\partial \text{tr}(\mathbf{X}^3)}{\partial \mathbf{X}} + \cdots \\
&= \mathbf{I} + \mathbf{X} + \frac{\mathbf{X}^2}{2!} + \cdots = e^{\mathbf{X}}
\end{aligned}$$

以及

$$\begin{aligned}
\frac{\partial \text{tr}(\sin \mathbf{X})}{\partial \mathbf{X}} &= \frac{\partial}{\partial \mathbf{X}} \text{tr} \left(\mathbf{X} - \frac{\mathbf{X}^3}{3!} + \frac{\mathbf{X}^5}{5!} - \cdots \right) \\
&= \frac{1}{1!} \frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} - \frac{1}{3!} \frac{\partial \text{tr}(\mathbf{X}^3)}{\partial \mathbf{X}} + \frac{1}{5!} \frac{\partial \text{tr}(\mathbf{X}^5)}{\partial \mathbf{X}} - \cdots \\
&= \mathbf{I} - \frac{\mathbf{X}^2}{2!} + \frac{\mathbf{X}^4}{4!} - \cdots = \cos \mathbf{X} \\
\frac{\partial \text{tr}(\cos \mathbf{X})}{\partial \mathbf{X}} &= \frac{\partial}{\partial \mathbf{X}} \text{tr} \left(\mathbf{I} - \frac{\mathbf{X}^2}{2!} + \frac{\mathbf{X}^4}{4!} - \frac{\mathbf{X}^6}{6!} + \cdots \right) \\
&= \frac{\partial \text{tr}(\mathbf{I})}{\partial \mathbf{X}} - \frac{1}{2!} \frac{\partial \text{tr}(\mathbf{X}^2)}{\partial \mathbf{X}} + \frac{1}{4!} \frac{\partial \text{tr}(\mathbf{X}^4)}{\partial \mathbf{X}} - \frac{1}{6!} \frac{\partial \text{tr}(\mathbf{X}^6)}{\partial \mathbf{X}} + \cdots \\
&= -\mathbf{X} + \frac{\mathbf{X}^3}{3!} - \frac{\mathbf{X}^5}{5!} + \cdots = -\sin \mathbf{X}
\end{aligned}$$

均与单变量的求导公式一样。

6.2 行列式对矩阵求导

设 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 、 $\mathbf{A} \in \mathbb{R}^{l \times m}$ 、 $\mathbf{B} \in \mathbb{R}^{n \times l}$, 则 $\mathbf{Y} = \mathbf{A} \mathbf{X} \mathbf{B} \in \mathbb{R}^{l \times l}$, 易知

$$\left[\frac{\partial |\mathbf{A} \mathbf{X} \mathbf{B}|}{\partial \mathbf{X}} \right]_{ij} = \frac{\partial |\mathbf{Y}|}{\partial x_{ji}} = \sum_p \sum_q \frac{\partial |\mathbf{Y}|}{\partial y_{pq}} \frac{\partial y_{pq}}{\partial x_{ji}}$$

注意 $y_{pq} = \sum_s \sum_t a_{ps} x_{st} b_{tq}$, 因此第二项

$$\frac{\partial y_{pq}}{\partial x_{ji}} = \frac{\partial \sum_s \sum_t a_{ps} x_{st} b_{tq}}{\partial x_{ji}} = a_{pj} b_{iq}$$

记 y_{pq} 有一个微小增量 ϵ 后的矩阵为 $\mathbf{Y}(y_{pq} + \epsilon)$, 根据第 j 行 Laplace 展开易知

$$|\mathbf{Y}(y_{pq} + \epsilon)| - |\mathbf{Y}| = \epsilon C_{pq}$$

其中 C_{pq} 是关于 y_{pq} 的代数余子式, 因此第一项

$$\frac{\partial |\mathbf{Y}|}{\partial y_{pq}} = \lim_{\epsilon \rightarrow 0} \frac{|\mathbf{Y}(y_{pq} + \epsilon)| - |\mathbf{Y}|}{\epsilon} = C_{pq}$$

代入可得

$$\begin{aligned} \left[\frac{\partial |\mathbf{AXB}|}{\partial \mathbf{X}} \right]_{ij} &= \sum_p \sum_q b_{iq} C_{pq} a_{pj} = \left[\mathbf{B} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \mathbf{A} \right]_{ij} = [\mathbf{BY}^* \mathbf{A}]_{ij} \\ &\Rightarrow \frac{\partial |\mathbf{AXB}|}{\partial \mathbf{X}} = \mathbf{B}(\mathbf{AXB})^* \mathbf{A} \end{aligned}$$

若 \mathbf{X} 、 \mathbf{A} 、 \mathbf{B} 均为可逆方阵, 则 $\mathbf{Y} = \mathbf{AXB}$ 亦为可逆方阵, 于是

$$\frac{\partial |\mathbf{AXB}|}{\partial \mathbf{X}} = \mathbf{B}(\mathbf{AXB})^* \mathbf{A} = \mathbf{B}|\mathbf{AXB}|(\mathbf{AXB})^{-1} \mathbf{A} = |\mathbf{AXB}| \mathbf{X}^{-1} \quad (4)$$

进一步若 $\mathbf{A} = \mathbf{B} = \mathbf{I}$, 则

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \mathbf{X}^* = |\mathbf{X}| \mathbf{X}^{-1}$$

事实上式 (4) 也可由上式导出

$$\frac{\partial |\mathbf{AXB}|}{\partial \mathbf{X}} = \frac{\partial |\mathbf{A}| |\mathbf{X}| |\mathbf{B}|}{\partial \mathbf{X}} = |\mathbf{A}| \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} |\mathbf{B}| = |\mathbf{A}| |\mathbf{X}| \mathbf{X}^{-1} |\mathbf{B}| = |\mathbf{AXB}| \mathbf{X}^{-1}$$

若 a 与 \mathbf{X} 无关, 则

$$\frac{\partial \ln |a\mathbf{X}|}{\partial \mathbf{X}} = \frac{\partial \ln a^m |\mathbf{X}|}{\partial \mathbf{X}} = \frac{\partial \ln a^m}{\partial \mathbf{X}} + \frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = \frac{1}{|\mathbf{X}|} \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \frac{\mathbf{X}^*}{|\mathbf{X}|} = \mathbf{X}^{-1}$$