矩阵相关知识汇总

1. tr，矩阵的迹，有关性质

① $\mathrm{tr}(\boldsymbol{A}\pm\boldsymbol{B})=\mathrm{tr}(\boldsymbol{A})\pm\mathrm{tr}(\boldsymbol{B})$

② 如c为常数，则$\mathrm{tr}(c\boldsymbol{A})=c\,\mathrm{tr}(\boldsymbol{A})$

③ $\mathrm{tr}(c_1\boldsymbol{A}\pm c_2\boldsymbol{B})=c_1\mathrm{tr}(\boldsymbol{A})\pm c_2\mathrm{tr}(\boldsymbol{B})$

④ $\mathrm{tr}(\boldsymbol{A}^T)=\mathrm{tr}(\boldsymbol{A})$，$\mathrm{tr}(\boldsymbol{A}^*)=[\mathrm{tr}(\boldsymbol{A})]^*$ (*表示复数共轭)，

$\mathrm{tr}(\boldsymbol{A}^H)=[\mathrm{tr}(\boldsymbol{A})]^*$ (H 表示共轭转置)

⑤ $\boldsymbol{A}_{m\times n}$，$\boldsymbol{B}_{n\times m}$ 则$\mathrm{tr}(\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{BA})$

推广：$\mathrm{tr}(\boldsymbol{ABC})=\mathrm{tr}(\boldsymbol{CAB})=\mathrm{tr}(\boldsymbol{BCA})$

⑥ $\boldsymbol{A}$和$\boldsymbol{B}$为$m\times m$矩阵，并且$\boldsymbol{B}$非奇异，则

$$\mathrm{tr}(\boldsymbol{BAB}^{-1})=\mathrm{tr}(\boldsymbol{B}^{-1}\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{A})$$

⑦ 若$\boldsymbol{A}$为$m\times n$矩阵，则$\mathrm{tr}(\boldsymbol{A}^H\boldsymbol{A})=0\Leftrightarrow\boldsymbol{A}=\boldsymbol{0}$ (零矩阵)

⑧ $\boldsymbol{x}^H\boldsymbol{Ax}=\mathrm{tr}(\boldsymbol{Axx}^H)$和$\boldsymbol{y}^H\boldsymbol{x}=\mathrm{tr}(\boldsymbol{xy}^H)$

⑨ $\mathrm{tr}\begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{C} & \boldsymbol{D}\end{bmatrix}=\mathrm{tr}(\boldsymbol{A})+\mathrm{tr}(\boldsymbol{D})$

⑩ $\mathrm{tr}(\boldsymbol{A})=\lambda_1+\lambda_2+\cdots+\lambda_n$

⑪ $\mathrm{tr}(\boldsymbol{A}^k)=\sum_{i=1}^{n}\lambda_i^k$

2. rank，矩阵的秩，性质

① $\mathrm{rank}(\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A}^T)$

② 若$\boldsymbol{P},\boldsymbol{Q}$非奇异，则$\mathrm{rank}(\boldsymbol{PA})=\mathrm{rank}(\boldsymbol{AQ})=\mathrm{rank}(\boldsymbol{PAQ})=\mathrm{rank}(\boldsymbol{A})$

③ $\mathrm{rank}\begin{pmatrix}\boldsymbol{A} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{B}\end{pmatrix}=\mathrm{rank}(\boldsymbol{A})+\mathrm{rank}(\boldsymbol{B})$

④ $\operatorname{rank}\begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{B} \end{pmatrix} \geq \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B})$

⑤ $\boldsymbol{A}$为$m \times n$，则 $\operatorname{rank}(\boldsymbol{A}) \leq \min(m,n)$

⑥ $\boldsymbol{A}$为$m \times k$，$\boldsymbol{B}$为$k \times n$，则

$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) - k \leq \operatorname{rank}(\boldsymbol{AB}) \leq \min\{\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B})\}$

⑦ $\boldsymbol{A,B}$为$m \times n$，则$\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})$

⑧ $\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^T) = \operatorname{rank}(\boldsymbol{A}^T\boldsymbol{A})=\operatorname{rank}(\boldsymbol{A})$

⑨ 设$\boldsymbol{A}$为n阶方阵，$\boldsymbol{A}^*$为$\boldsymbol{A}$的伴随矩阵，则

$$\operatorname{rank}(\boldsymbol{A}^*) = \begin{cases} n, & \operatorname{rank}(\boldsymbol{A})=n \\ 1, & \operatorname{rank}(\boldsymbol{A})=n-1 \\ 0, & \operatorname{rank}(\boldsymbol{A})<n-1 \end{cases}$$

3. 矩阵及向量范数

1) 向量范数

设$\|\boldsymbol{x}\|$为以向量$\boldsymbol{x}$为自变量的非负实值函数，且满足：

①非负性：当$\boldsymbol{x} \neq \boldsymbol{0}$，$\|\boldsymbol{x}\| > 0$；当$\boldsymbol{x} = \boldsymbol{0}$，$\|\boldsymbol{x}\| = 0$

②齐次性：对$\forall k \in \mathbf{P}$ ($\mathbf{P}$为数域)，$\|k\boldsymbol{x}\| = |k|\|\boldsymbol{x}\|$

③三角不等式：$\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbf{R}^n \text{ or } \mathbf{C}^n$，$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

则称$\|\boldsymbol{x}\|$为向量$\boldsymbol{x}$的范数.

2) 常见的向量范数

$$\|\boldsymbol{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

$\|\boldsymbol{x}\|_2 = (\sum_{i=1}^{n} |x_i|^2)^{1/2}$，$\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^H\boldsymbol{x}$(复数域)，$\boldsymbol{x}^T\boldsymbol{x}$(实数域)

$$\|\boldsymbol{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

$$\|\boldsymbol{x}\|_p=(\sum_{i=1}^{n}|x_i|^p)^{1/p},\ 1\le p<+\infty$$

3) 矩阵范数

定义和向量范数相似

4) 常见的矩阵范数

① Frobenius 范数：$\|\boldsymbol{A}\|_F=(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2)^{1/2}=\left[\mathrm{tr}(\boldsymbol{A}^H\boldsymbol{A})\right]^{1/2}$ (**C**域)

$$\|\boldsymbol{A}\|_F=(\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}^2)^{1/2}=\left[\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A})\right]^{1/2}\ (\mathbf{R}\text{域})$$

② $\|\boldsymbol{A}\|_1=\max\limits_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|$，其中$\boldsymbol{A}$为$m\times n$，列和

③ $\|\boldsymbol{A}\|_\infty=\max\limits_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|$，行和

④ $\|\boldsymbol{A}\|_2=\left[\lambda_{\max}(\boldsymbol{A}^H\boldsymbol{A})\right]^{1/2}$

⑤ F 范数性质：$\boldsymbol{A}\in\mathbf{R}^{m\times n}$，$\boldsymbol{P}$和$\boldsymbol{Q}$分别为$m$阶和$n$阶正交矩阵

$$\|\boldsymbol{PAQ}\|_F=\|\boldsymbol{A}\|_F$$

⑥ 2 范数性质：$\|\boldsymbol{A}\|_2=\max\limits_{\substack{\|\boldsymbol{x}\|_2=1\\ \|\boldsymbol{y}\|_2=1}}|\boldsymbol{y}^H\boldsymbol{A}\boldsymbol{x}|$

$$\|\boldsymbol{A}^H\|_2=\|\boldsymbol{A}^T\|_2=\|\boldsymbol{A}\|_2$$

$$\|\boldsymbol{A}^H\boldsymbol{A}\|_2=\|\boldsymbol{A}\|_2^2$$

$$\|\boldsymbol{A}\|_2^2\le\|\boldsymbol{A}\|_1\|\boldsymbol{A}\|_\infty$$

设$\boldsymbol{A}$为$m\times n$，m阶正交阵$\boldsymbol{P}$和n阶正交阵$\boldsymbol{Q}$，$\|\boldsymbol{PAQ}\|_2=\|\boldsymbol{A}\|_2$

△ 4. 导数或梯度

1) 对标量的导数

设$m\times n$阶矩阵$\boldsymbol{A}(t)$的每一个元素均是标量t的函数，则

$$\frac{d\boldsymbol{A}(t)}{dt}=\begin{bmatrix}\frac{da_{11}}{dt} & \frac{da_{12}}{dt} & \cdots & \frac{da_{1n}}{dt}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{da_{m1}}{dt} & \frac{da_{m2}}{dt} & \cdots & \frac{da_{mn}}{dt}\end{bmatrix}$$

设 $\boldsymbol{x}(t)$ 向量的每一个元素均是标量 t 的函数，则

$$\frac{d\boldsymbol{x}(t)}{dt}=\begin{bmatrix}\frac{dx_1}{dt}\\ \vdots\\ \frac{dx_n}{dt}\end{bmatrix}$$

2) 对向量变量的导数

(1) 数量函数对向量的导数

设 $f(\boldsymbol{x})$ 是以向量 $\boldsymbol{x}=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$ 为自变量的数量函数，则

$$\frac{df(\boldsymbol{x})}{d\boldsymbol{x}}=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{bmatrix}=\nabla_x f$$

(2) 向量函数对向量的导数(梯度)

m 维行向量函数 $\boldsymbol{f}^T(\boldsymbol{x})=\left[f_1(\boldsymbol{x}),f_2(\boldsymbol{x}),...,f_m(\boldsymbol{x})\right]$，对于 n 维列量 $\boldsymbol{x}$ 的梯度为一 $n\times m$ 矩阵，定义为：

$$\frac{\partial \boldsymbol{f}^T(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \nabla_{\boldsymbol{x}} \boldsymbol{f}$$

相似地，m 维列向量函数 $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ 对 n 维行向量 $\boldsymbol{x}^T$ 的

梯度为：

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}^T} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}_{m\times n} = \nabla_{\boldsymbol{x}^T} \boldsymbol{y}$$

总之一句话：行对列，列对行，行不能对行，列不能对列。

(3) 矩阵函数对向量的导数

设矩阵函数 $\boldsymbol{A}(\boldsymbol{x})$，即 $\boldsymbol{A}$ 的每一个元素 a_{ij} 均为向量 $\boldsymbol{x}$ 的函数，

即

$$\boldsymbol{A}(\boldsymbol{x}) = \begin{bmatrix} a_{11}(\boldsymbol{x}) & a_{12}(\boldsymbol{x}) & \cdots & a_{1p}(\boldsymbol{x}) \\ a_{21}(\boldsymbol{x}) & a_{22}(\boldsymbol{x}) & \cdots & a_{2p}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}(\boldsymbol{x}) & a_{m2}(\boldsymbol{x}) & \cdots & a_{mp}(\boldsymbol{x}) \end{bmatrix}_{m\times p}$$

其中 $\boldsymbol{x} = (x_1, x_2, ..., x_n)^T$ 为列向量，则

$$\frac{dA(\boldsymbol{x})}{d\boldsymbol{x}}=\begin{bmatrix}\frac{\partial A(\boldsymbol{x})}{\partial x_1}\\ \frac{\partial A(\boldsymbol{x})}{\partial x_2}\\ \vdots\\ \frac{\partial A(\boldsymbol{x})}{\partial x_n}\end{bmatrix}_{mn\times p}\quad ,\quad \frac{dA(\boldsymbol{x})}{d\boldsymbol{x}^T}=\begin{bmatrix}\frac{\partial A(\boldsymbol{x})}{\partial x_1} & \frac{\partial A(\boldsymbol{x})}{\partial x_2} & \cdots & \frac{\partial A(\boldsymbol{x})}{\partial x_n}\end{bmatrix}_{m\times np}$$

3) 实值函数对矩阵的梯度

设 $f=f(\boldsymbol{A})$ 是以矩阵 $\boldsymbol{A}\in\mathbf{R}^{m\times n}$ 为自变量的实值函数，则

$$\frac{df(\boldsymbol{A})}{d\boldsymbol{A}}=\begin{bmatrix}\frac{df}{da_{11}} & \frac{df}{da_{12}} & \cdots & \frac{df}{da_{1n}}\\ \frac{df}{da_{21}} & \frac{df}{da_{22}} & \cdots & \frac{df}{da_{2n}}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{df}{da_{m1}} & \frac{df}{da_{m2}} & \cdots & \frac{df}{da_{mn}}\end{bmatrix}$$

以上为定义，下面开始讲例子。

$$\begin{pmatrix}\vartriangle\ \frac{\partial\boldsymbol{x}}{\partial\boldsymbol{x}^T}=\boldsymbol{I} & \frac{\partial\boldsymbol{x}^T}{\partial\boldsymbol{x}}=\boldsymbol{I} & \text{其中}\boldsymbol{x}\text{为向量}\\ \vartriangle\ \frac{\partial\mathrm{tr}(\boldsymbol{A}^T)}{\partial\boldsymbol{A}}=\boldsymbol{I} & \frac{\partial\mathrm{tr}(\boldsymbol{A})}{\partial\boldsymbol{A}^T}=\boldsymbol{I} & \text{其中}\boldsymbol{A}\text{为矩阵}\end{pmatrix}$$

例1. $\boldsymbol{A}$ 和 $\boldsymbol{y}$ 均与 $\boldsymbol{x}$ 无关，则

$$\frac{\partial\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}}{\partial\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{y}\qquad \frac{\partial\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x}}{\partial\boldsymbol{x}}=\boldsymbol{A}^T\boldsymbol{y}$$

($\boldsymbol{A}$ 为矩阵， $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 均为向量，下同)

例2. 若 $\boldsymbol{A}$ 与 $\boldsymbol{x}$ 无关，则 $\frac{\partial\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^T\boldsymbol{x}$

例3. $\frac{\partial\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}}{\partial\boldsymbol{A}}=\boldsymbol{x}\boldsymbol{y}^T$

例4. $\dfrac{\partial \mathrm{tr}(\boldsymbol{W})}{\partial \boldsymbol{W}} = \dfrac{\partial \mathrm{tr}(\boldsymbol{W}^T)}{\partial \boldsymbol{W}} = \boldsymbol{I}$

$$\frac{\partial \mathrm{tr}(\boldsymbol{W}^{-1})}{\partial \boldsymbol{W}} = -(\boldsymbol{W}^{-2})^T$$

$$\frac{\partial \mathrm{tr}(\boldsymbol{x}\boldsymbol{y}^T)}{\partial \boldsymbol{x}} = \boldsymbol{y} = \frac{\partial \mathrm{tr}(\boldsymbol{y}\boldsymbol{x}^T)}{\partial \boldsymbol{x}}$$

$$\frac{\partial \mathrm{tr}(\boldsymbol{W}\boldsymbol{A})}{\partial \boldsymbol{W}} = \boldsymbol{A}^T = \frac{\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{W})}{\partial \boldsymbol{W}}$$

$$\frac{\partial \mathrm{tr}(\boldsymbol{W}\boldsymbol{W}^T)}{\partial \boldsymbol{W}} = 2\boldsymbol{W}$$

若 $\boldsymbol{A}$ 与 $\boldsymbol{W}$ 无关，则 $\dfrac{\partial \mathrm{tr}(\boldsymbol{W}^T\boldsymbol{A}\boldsymbol{W})}{\partial \boldsymbol{W}} = (\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{W}$

△ $$\frac{\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{W}^{-1})}{\partial \boldsymbol{W}} = -(\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W}^{-1})^T$$

$$\frac{\partial \mathrm{tr}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}^T\boldsymbol{B})}{\partial \boldsymbol{W}} = \boldsymbol{B}\boldsymbol{W}\boldsymbol{A} + \boldsymbol{B}^T\boldsymbol{W}\boldsymbol{A}^T$$

例5. $\boldsymbol{y}$ 和 $\boldsymbol{z}$ 均为向量 $\boldsymbol{x}$ 的向量函数，即 $\boldsymbol{y}(\boldsymbol{x}) = \left[y_1(\boldsymbol{x}),\ldots,y_n(\boldsymbol{x})\right]$，$\boldsymbol{z}(\boldsymbol{x}) = \left[z_1(\boldsymbol{x}),\ldots,z_n(\boldsymbol{x})\right]$，则有：

$$\frac{d(\boldsymbol{y}^T\boldsymbol{z})}{d\boldsymbol{x}} = \frac{d\boldsymbol{y}^T}{d\boldsymbol{x}}\boldsymbol{z} + \frac{d\boldsymbol{z}^T}{d\boldsymbol{x}}\boldsymbol{y}$$

$$\frac{d(\boldsymbol{y}^T\boldsymbol{z})}{d\boldsymbol{x}^T} = \boldsymbol{y}^T\frac{d\boldsymbol{z}}{d\boldsymbol{x}^T} + \boldsymbol{z}^T\frac{d\boldsymbol{y}}{d\boldsymbol{x}^T}$$

4. 复合函数微分法

① 设 $f = f(\boldsymbol{y})$ 是列向量 $\boldsymbol{y}$ 的标量函数，而 $\boldsymbol{y} = \boldsymbol{y}(t) = \left[y_1(t),\ldots,y_n(t)\right]^T$，则 f 对标量 t 的导数为：

$$\frac{df}{dt} = \frac{df}{d\boldsymbol{y}^T}\frac{d\boldsymbol{y}}{dt} = \frac{d\boldsymbol{y}^T}{dt}\frac{df}{d\boldsymbol{y}}$$

② 设 $\boldsymbol{g}=\begin{bmatrix} g_1(\boldsymbol{y}) \\ g_2(\boldsymbol{y}) \\ \vdots \\ g_m(\boldsymbol{y}) \end{bmatrix}$， $\boldsymbol{y}=\begin{bmatrix} y_1(\boldsymbol{x}) \\ y_2(\boldsymbol{x}) \\ \vdots \\ y_s(\boldsymbol{x}) \end{bmatrix}$， $\boldsymbol{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$， $\boldsymbol{g},\boldsymbol{y},\boldsymbol{x}$ 均为向量，则

$$\left.\begin{aligned} \frac{d\boldsymbol{g}^T}{d\boldsymbol{x}}_{n\times m} &= \frac{d\boldsymbol{y}^T}{d\boldsymbol{x}}\frac{d\boldsymbol{g}^T}{d\boldsymbol{y}} \\ \frac{d\boldsymbol{g}}{d\boldsymbol{x}^T}_{m\times n} &= \frac{d\boldsymbol{g}}{d\boldsymbol{y}^T}\frac{d\boldsymbol{y}}{d\boldsymbol{x}^T} \end{aligned}\right\}\text{链式法则. 记忆要点：能乘维数对就对}$$

5. 矩阵微分

设向量函数 $\boldsymbol{f}(\boldsymbol{c})$ 在 $\boldsymbol{c}$ 可微分，并且向量函数 $\boldsymbol{g}(\boldsymbol{b})$ 在 $\boldsymbol{b}=\boldsymbol{f}(\boldsymbol{c})$ 可微分，则复合向量函数 $\boldsymbol{h}(\boldsymbol{x})=\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))$ 在 $\boldsymbol{c}$ 可微分，且

$$D\boldsymbol{h}(\boldsymbol{c})=(D\left[\boldsymbol{g}(\boldsymbol{b})\right])(D\left[\boldsymbol{f}(\boldsymbol{c})\right])$$

例子：

$$d(\mathrm{tr}\boldsymbol{U})=\mathrm{tr}(d\boldsymbol{U}) \quad \triangle$$
$$d(\boldsymbol{U}\boldsymbol{V})=d(\boldsymbol{U})\boldsymbol{V}+\boldsymbol{U}d\boldsymbol{V} \quad \triangle$$
$$d(\boldsymbol{U}^T)=(d\boldsymbol{U})^T$$
$$d(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})=\boldsymbol{A}(d\boldsymbol{X})\boldsymbol{B}$$
$$d(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^T)=(d\boldsymbol{X})\boldsymbol{A}\boldsymbol{X}^T+\boldsymbol{X}\boldsymbol{A}(d\boldsymbol{X})^T$$
$$d(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})=(d\boldsymbol{X})^T\boldsymbol{A}\boldsymbol{X}+\boldsymbol{X}^T\boldsymbol{A}d\boldsymbol{X}$$
$$d\left|\boldsymbol{X}\right|=\left|\boldsymbol{X}\right|\mathrm{tr}(\boldsymbol{X}^{-1}d\boldsymbol{X})$$
$$d(\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{X}))=2\mathrm{tr}(\boldsymbol{X}^Td\boldsymbol{X}) \quad \triangle$$
$$d(\boldsymbol{X}^{-1})=-\boldsymbol{X}^{-1}(d\boldsymbol{X})\boldsymbol{X}^{-1}$$
$$d\log(\boldsymbol{X})=\boldsymbol{X}^{-1}d\boldsymbol{X}$$
$$d\log(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})=(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})^{-1}\left[(d\boldsymbol{X})^T\boldsymbol{A}\boldsymbol{X}+\boldsymbol{X}^T\boldsymbol{A}d\boldsymbol{X}\right]$$

根据微分求标量函数的梯度：

两个重要公式：

$$df(\boldsymbol{x}) = \boldsymbol{A}\,\mathrm{d}\,\boldsymbol{x} \Leftrightarrow \nabla f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^T$$

$$df(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{A}\,\mathrm{d}\,\boldsymbol{x}) \Leftrightarrow \nabla f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^T$$

例1. $\dfrac{\partial \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = ?$

$$\begin{aligned} d\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}) &= \mathrm{tr}\left[d(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})\right] \\ &= \mathrm{tr}\left[(d\boldsymbol{X})^T\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^T\boldsymbol{A}d\boldsymbol{X}\right] \\ &= \mathrm{tr}\left[\boldsymbol{X}^T\boldsymbol{A}^T d\boldsymbol{X}\right] + \mathrm{tr}\left[\boldsymbol{X}^T\boldsymbol{A}d\boldsymbol{X}\right] \\ &= \mathrm{tr}\left[(\boldsymbol{X}^T\boldsymbol{A}^T + \boldsymbol{X}^T\boldsymbol{A})d\boldsymbol{X}\right] \end{aligned}$$

因此： $\dfrac{\partial \mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = (\boldsymbol{A}^T + \boldsymbol{A})\boldsymbol{X}$

例2. $\dfrac{\partial |\boldsymbol{X}|^2}{\partial \boldsymbol{X}} = ?$

$$d(|\boldsymbol{X}|^2) = 2|\boldsymbol{X}|d(|\boldsymbol{X}|) = 2|\boldsymbol{X}|^2\,\mathrm{tr}(\boldsymbol{X}^{-1}d\boldsymbol{X})$$

因此： $\dfrac{\partial |\boldsymbol{X}|^2}{\partial \boldsymbol{X}} = 2|\boldsymbol{X}|^2\boldsymbol{X}^{-T}$

例3. $\dfrac{\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}} = ?$

$$\begin{aligned} d\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1}) &= \mathrm{tr}\left[d(\boldsymbol{A}\boldsymbol{X}^{-1})\right] \\ &= \mathrm{tr}\left[\boldsymbol{A}d\boldsymbol{X}^{-1}\right] \\ &= -\mathrm{tr}\left[\boldsymbol{A}\boldsymbol{X}^{-1}(d\boldsymbol{X})\boldsymbol{X}^{-1}\right] \\ &= -\mathrm{tr}\left[\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1}d\boldsymbol{X}\right] \end{aligned}$$

因此： $\dfrac{\partial \mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})}{\partial \boldsymbol{X}} = -\boldsymbol{X}^{-T}\boldsymbol{A}^T\boldsymbol{X}^{-T}$

例4. $\dfrac{\partial |\boldsymbol{X}\boldsymbol{X}^T|}{\partial \boldsymbol{X}} = ?$

$$
\begin{aligned}
d\left|\boldsymbol{X}\boldsymbol{X}^T\right| &= \left|\boldsymbol{X}\boldsymbol{X}^T\right| \operatorname{tr}\left[(\boldsymbol{X}\boldsymbol{X}^T)^{-1} d(\boldsymbol{X}\boldsymbol{X}^T)\right] \\
&= \left|\boldsymbol{X}\boldsymbol{X}^T\right| \operatorname{tr}\left[(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\left[(d\boldsymbol{X})\boldsymbol{X}^T + \boldsymbol{X}(d\boldsymbol{X})^T\right]\right] \\
&= \left|\boldsymbol{X}\boldsymbol{X}^T\right| \operatorname{tr}\left[(\boldsymbol{X}\boldsymbol{X}^T)^{-1}(d\boldsymbol{X})\boldsymbol{X}^T + (\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}d(\boldsymbol{X})^T\right] \\
&= \left|\boldsymbol{X}\boldsymbol{X}^T\right| \operatorname{tr}\left[\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}d\boldsymbol{X} + (d\boldsymbol{X})\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-T}\right] \\
&= \left|\boldsymbol{X}\boldsymbol{X}^T\right| \operatorname{tr}\left[\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}d\boldsymbol{X} + \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}d\boldsymbol{X}\right] \\
&= 2\left|\boldsymbol{X}\boldsymbol{X}^T\right| \operatorname{tr}\left[\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T)^{-1}d\boldsymbol{X}\right]
\end{aligned}
$$

因此：$\dfrac{\partial\left|\boldsymbol{X}\boldsymbol{X}^T\right|}{\partial\boldsymbol{X}} = 2\left|\boldsymbol{X}\boldsymbol{X}^T\right|(\boldsymbol{X}\boldsymbol{X}^T)^{-1}\boldsymbol{X}$