矩阵论

2024年秋学期

第六讲

2024年9月25日

第3章 矩阵微分

矩阵微分

矩阵微分是矩阵分析和多变量微积分中的一个重要工具，它在理论研究和实际应用中都有广泛的用途。通过矩阵微分，可以高效地处理和解决涉及多维数据的问题。

矩阵微分的主要用途：

1. 优化问题：在机器学习、统计学和工程优化等领域，经常需要最小化或最大化某个多变量函数（通常表达为矩阵形式）。矩阵微分提供了一种寻找这些函数最优点（极值点）的方法。例如，通过求解梯度等于零的点，可以找到函数的局部最小值或最大值。

2. 机器学习：在机器学习中，矩阵微分用于计算损失函数的梯度，这是许多优化算法（如梯度下降法）的核心步骤。此外，在神经网络的反向传播算法中，矩阵微分被用来高效地计算权重的更新。

函数对变量（元素、向量或者矩阵）的导数/微分

变元、函数和映射

$\boldsymbol{x}=\left[x_1,\cdots,x_m\right]^{\mathrm{T}}\in\mathbb{R}^m$ 为实向量变元

$\boldsymbol{X}=\left[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\right]\in\mathbb{R}^{m\times n}$ 为实矩阵变元

$f(\boldsymbol{x})\in\mathbb{R}$ 为实值标量函数，其变元为 $m\times 1$ 实值向量 $\boldsymbol{x}$， $f:\mathbb{R}^m\rightarrow\mathbb{R}$ ；

$f(\boldsymbol{X})\in\mathbb{R}$ 为实值标量函数，其变元为 $m\times n$ 实值矩阵 $\boldsymbol{X}$， $f:\mathbb{R}^{m\times n}\rightarrow\mathbb{R}$；

$\boldsymbol{f}(\boldsymbol{x})\in\mathbb{R}^p$ 为p维实列向量函数，其变元为 $m\times 1$ 实值向量 $\boldsymbol{x}$ ，$\boldsymbol{f}:\mathbb{R}^m\rightarrow\mathbb{R}^p$;

$\boldsymbol{f}(\boldsymbol{X})\in\mathbb{R}^p$ 为p维实列向量函数，其变元为 $m\times n$ 实值矩阵 $\boldsymbol{X}$,$\boldsymbol{f}:\mathbb{R}^{m\times n}\rightarrow\mathbb{R}^p$;

$\boldsymbol{F}(\boldsymbol{x})\in\mathbb{R}^{p\times q}$ 为$p\times q$实矩阵函数，其变元为 $m\times 1$ 实值向量 $\boldsymbol{x}$ ，$F:\mathbb{R}^m\rightarrow\mathbb{R}^{p\times q}$;

$\boldsymbol{F}(\boldsymbol{X})\in\mathbb{R}^{p\times q}$ 为$p\times q$实矩阵函数，其变元为 $m\times n$实值向量 $\boldsymbol{X}$, $F:\mathbb{R}^{m\times n}\rightarrow\mathbb{R}^{p\times q}$;

变元、函数和映射

表 3.1.1　实值函数的分类

函数类型	向量变元 $\boldsymbol{x} \in \mathbb{R}^m$	矩阵变元 $\boldsymbol{X} \in \mathbb{R}^{m\times n}$
标量函数 $f \in \mathbb{R}$	$f(\boldsymbol{x})$ $f:\ \mathbb{R}^m \to \mathbb{R}$	$f(\boldsymbol{X})$ $f:\ \mathbb{R}^{m\times n} \to \mathbb{R}$
向量函数 $\boldsymbol{f} \in \mathbb{R}^p$	$\boldsymbol{f}(\boldsymbol{x})$ $\boldsymbol{f}:\ \mathbb{R}^m \to \mathbb{R}^p$	$\boldsymbol{f}(\boldsymbol{X})$ $\boldsymbol{f}:\ \mathbb{R}^{m\times n} \to \mathbb{R}^p$
矩阵函数 $\boldsymbol{F} \in \mathbb{R}^{p\times q}$	$\boldsymbol{F}(\boldsymbol{x})$ $\boldsymbol{F}:\ \mathbb{R}^m \to \mathbb{R}^{p\times q}$	$\boldsymbol{F}(\boldsymbol{X})$ $\boldsymbol{F}:\ \mathbb{R}^{m\times n} \to \mathbb{R}^{p\times q}$

Jacobian矩阵

行向量偏导算子

$$\mathrm{D}_{\boldsymbol{x}} \overset{\mathrm{def}}{=} \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left[\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_m} \right]$$

$$\mathrm{D}_{\boldsymbol{x}} f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_m} \right]$$

当实值标量函数的变元为实值矩阵时，存在两种定义

1) $f(\boldsymbol{X})$ **关于矩阵变元** $\boldsymbol{X}$ **的Jacobian矩阵**

$$\boldsymbol{X} \in \mathbb{R}^{m \times n} \quad \mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{n \times m}$$

Jacobian矩阵

2）行偏导向量

$$\mathrm{D}_{\mathrm{vec}\boldsymbol{X}} f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \, \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})} = \left[\frac{\partial f(\boldsymbol{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}} \right]$$

$$\mathrm{D}_{\mathrm{vec}\boldsymbol{X}} f(\boldsymbol{X}) = \mathrm{rvec}\left(\mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X})\right) = \left(\mathrm{vec}\left(\mathrm{D}_{\boldsymbol{X}}^{\mathrm{T}} f(\boldsymbol{X})\right)\right)^{\mathrm{T}}$$

Jacobian矩阵与梯度矩阵

函数为矩阵，变元为矩阵

Jacobian 矩阵

$$\mathrm{D}_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X}) \overset{\text{def}}{=} \frac{\partial \mathrm{vec}(\boldsymbol{F}(\boldsymbol{X}))}{\partial (\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \in \mathbb{R}^{pq\times mn}$$

其具体表达式为

$$\mathrm{D}_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X}) = \begin{bmatrix} \frac{\partial f_{11}}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \\ \vdots \\ \frac{\partial f_{p1}}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \\ \vdots \\ \frac{\partial f_{1q}}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \\ \vdots \\ \frac{\partial f_{pq}}{\partial(\mathrm{vec}\boldsymbol{X})^{\mathrm{T}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \cdots & \frac{\partial f_{11}}{\partial x_{m1}} & \cdots & \frac{\partial f_{11}}{\partial x_{1n}} & \cdots & \frac{\partial f_{11}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{p1}}{\partial x_{11}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m1}} & \cdots & \frac{\partial f_{p1}}{\partial x_{1n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{1q}}{\partial x_{11}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m1}} & \cdots & \frac{\partial f_{1q}}{\partial x_{1n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{pq}}{\partial x_{11}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m1}} & \cdots & \frac{\partial f_{pq}}{\partial x_{1n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}} \end{bmatrix}$$

梯度矩阵

$m \times 1$ 列向量偏导算子即梯度算子记作 $\nabla_{\boldsymbol{x}}$，定义为

$$\nabla_{\boldsymbol{x}} \overset{\text{def}}{=} \frac{\partial}{\partial \boldsymbol{x}} = \left[\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_m} \right]^{\mathrm{T}}$$

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) \overset{\text{def}}{=} \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_m} \right]^{\mathrm{T}} = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}$$

矩阵变元 X 的梯度算子为

$$\nabla_{\mathrm{vec}\boldsymbol{X}} = \frac{\partial}{\partial \mathrm{vec}\boldsymbol{X}} = \left[\frac{\partial}{\partial x_{11}}, \cdots, \frac{\partial}{\partial x_{m1}}, \cdots, \frac{\partial}{\partial x_{1n}}, \cdots, \frac{\partial}{\partial x_{mn}} \right]^{\mathrm{T}}$$

矩阵变元 X 的梯度向量为

$$\nabla_{\mathrm{vec}\,\boldsymbol{X}} f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial\, \mathrm{vec}\, \boldsymbol{X}} = \left[\frac{\partial f(\boldsymbol{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}} \right]^{\mathrm{T}}$$

梯度矩阵

也可以直接定义梯度矩阵

$$\nabla_{X} f(\boldsymbol{X})=\begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial x_{mn}} \end{bmatrix}=\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}$$

对于实值矩阵函数 $\boldsymbol{F}(\boldsymbol{X})\in\mathbb{R}^{p\times q}$ **（其中矩阵变元** $\boldsymbol{X}\in\mathbb{R}^{m\times n}$ **），梯度矩阵定义为**

$$\nabla_{\boldsymbol{X}} \boldsymbol{F}(\boldsymbol{X})=\frac{\partial \operatorname{vec}^{\mathrm{T}} \boldsymbol{F}(\boldsymbol{X})}{\partial \operatorname{vec} \boldsymbol{X}}=\left(\frac{\partial \operatorname{vec} \boldsymbol{F}(\boldsymbol{X})}{\partial \operatorname{vec}^{\mathrm{T}} \boldsymbol{X}}\right)^{\mathrm{T}}$$

$$\nabla_{\boldsymbol{X}} \boldsymbol{F}(\boldsymbol{X})=\left(\mathrm{D}_{\boldsymbol{X}} \boldsymbol{F}(\boldsymbol{X})\right)^{\mathrm{T}}$$

矩阵函数的梯度矩阵是其Jacobian矩阵的转置。

梯度矩阵

(1）若 $f(\boldsymbol{X})=c$ **为常数，其中，** $\boldsymbol{X}$ **为** $m\times n$ **矩阵，则梯度** $\dfrac{\partial c}{\partial \boldsymbol{X}}=\boldsymbol{O}_{m\times n}$

(2）线性法则　若 $f(\boldsymbol{X})$ **和** $g(\boldsymbol{X})$ **分别是矩阵** $\boldsymbol{X}$ **的实值函数，** c_1 **和** c_2 **为实常数，则**

$$\frac{\partial\left[c_1 f(\boldsymbol{X})+c_2 g(\boldsymbol{X})\right]}{\partial \boldsymbol{X}}=c_1\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}+c_2\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}}$$

(3）乘积法则　若 $f(\boldsymbol{X})$**、**$g(\boldsymbol{X})$ **和** $h(\boldsymbol{X})$ **都是矩阵** $\boldsymbol{X}$ **的实值函数，则**

$$\frac{\partial[f(\boldsymbol{X})g(\boldsymbol{X})]}{\partial \boldsymbol{X}}=g(\boldsymbol{X})\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}+f(\boldsymbol{X})\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}}$$

$$\frac{\partial[f(\boldsymbol{X})g(\boldsymbol{X})h(\boldsymbol{X})]}{\partial \boldsymbol{X}}=g(\boldsymbol{X})h(\boldsymbol{X})\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}+f(\boldsymbol{X})h(\boldsymbol{X})\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}}+f(\boldsymbol{X})g(\boldsymbol{X})\frac{\partial h(\boldsymbol{X})}{\partial \boldsymbol{X}}$$

梯度矩阵

(4）商法则　若 $g(\boldsymbol{X}) \neq 0$**，则**

$$\frac{\partial [f(\boldsymbol{X})/g(\boldsymbol{X})]}{\partial \boldsymbol{X}} = \frac{1}{g^2(\boldsymbol{X})}\left[g(\boldsymbol{X})\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} - f(\boldsymbol{X})\frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}}\right]$$

(5）链式法则

$$\frac{\partial g(f(\boldsymbol{X}))}{\partial \boldsymbol{X}} = \frac{\mathrm{d}g(y)}{\mathrm{d}y}\frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}}$$

独立性与基本假设

矩阵求导的独立性主要体现在对矩阵元素的处理上。在对矩阵函数进行求导时，通常假设矩阵中的每个元素都是相互独立的。这种假设简化了求导过程，因为它允许单独对每个元素进行求导，而不用担心其他元素的影响。这一点在计算梯度或雅可比矩阵时尤为重要。

假设实值函数的向量变元 $x=\left[x_i\right]_{i=1}^{m}\in\mathbb{R}^m$ **或者矩阵变元** $\left[x_{ij}\right]_{i=1,j=1}^{m,n}\in\mathbb{R}^{m\times n}$

本身无任何特殊结构，即向量或矩阵变元的元素之间是各自独立的。

上述独立性基本假设可以用数学公式表示成

$$\frac{\partial x_i}{\partial x_j}=\delta_{ij}=\begin{cases}1, & i=j\\ 0, & \text{其他}\end{cases}$$

以及

$$\frac{\partial x_{kl}}{\partial x_{ij}}=\delta_{ki}\delta_{lj}=\begin{cases}1, & k=i \text{ 且 } l=j\\ 0, & \text{其他}\end{cases}$$

独立性与基本假设

示例 求实值函数 $f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 的 Jacobian 矩阵。

由于

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\sum_{k=1}^{n}\sum_{l=1}^{n}a_{kl}x_k x_l,$$

故利用独立性基本假设可求出行偏导向量 $\dfrac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}^{\mathrm{T}}}$ 的第 i 个分量为

$$\left[\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}^{\mathrm{T}}}\right]_i=\frac{\partial}{\partial x_i}\sum_{k=1}^{n}\sum_{l=1}^{n}a_{kl}x_k x_l=\sum_{k=1}^{n}x_k a_{ki}+\sum_{l=1}^{n}x_l a_{il}$$

示例 令 $\boldsymbol{F}(\boldsymbol{X})=\boldsymbol{X}\in\mathbb{R}^{m\times n}$，则直接计算偏导得

$$\frac{\partial f_{kl}}{\partial x_{ij}}=\frac{\partial x_{kl}}{\partial x_{ij}}=\delta_{lj}\delta_{ki}$$

于是得 Jacobian 矩阵

$$\mathrm{D}_{\boldsymbol{X}}\boldsymbol{X}=\boldsymbol{I}_n\otimes\boldsymbol{I}_m=\boldsymbol{I}_{mn}\in\mathbb{R}^{mn\times mn} \longleftarrow \mathrm{D}_{\boldsymbol{X}}\boldsymbol{F}(\boldsymbol{X})\overset{\text{def}}{=}\frac{\partial\,\mathrm{vec}(\boldsymbol{F}(\boldsymbol{X}))}{\partial(\mathrm{vec}\,\boldsymbol{X})^{\mathrm{T}}}$$

一阶实矩阵微分：Jacobian矩阵辨识

矩阵微分用符号 d$\mathbf{X}$ 表示

标量函数 $\mathrm{tr}(\mathbf{U})$的微分

$$\mathrm{d}\left[\mathrm{tr}(\mathbf{U})\right]=\mathrm{d}\left(\sum_{i=1}^{n}u_{ii}\right)=\sum_{i=1}^{n}\mathrm{d}u_{ii}=\mathrm{tr}\left(\mathrm{d}\mathbf{U}\right)$$

矩阵的迹的微分等于矩阵微分的迹

矩阵乘积UV的微分矩阵

$$\mathrm{d}\left(\mathbf{UV}\right)=(\mathrm{d}\,\mathbf{U})\mathbf{V}+\mathbf{U}(\mathrm{d}\,\mathbf{V})$$

一阶实矩阵微分：Jacobian矩阵辨识

标量函数 $f(\boldsymbol{x})$，变元向量 $\boldsymbol{x}=[x_1,x_2,\cdots x_m]^{\mathrm{T}}\in R^m$

微分法则的向量形式

$$\mathrm{d}f(\boldsymbol{x})=\frac{\partial f(\boldsymbol{x})}{\partial x_1}\mathrm{d}x_1+\frac{\partial f(\boldsymbol{x})}{\partial x_2}\mathrm{d}x_2+\cdots+\frac{\partial f(\boldsymbol{x})}{\partial x_m}\mathrm{d}x_m$$

$$=\left[\frac{\partial f(\boldsymbol{x})}{\partial x_1},\cdots,\frac{\partial f(\boldsymbol{x})}{\partial x_m}\right]\begin{bmatrix}\mathrm{d}x_1\\ \vdots\\ \mathrm{d}x_m\end{bmatrix}$$

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

标量函数对向量的导数

标量函数 $f(\boldsymbol{X})$，变元矩阵 $\boldsymbol{X}\in\boldsymbol{R}^{m\times n}$

***A*: Jacobian矩阵**

$$\mathrm{d}f(\boldsymbol{X})=tr(\boldsymbol{A}d\boldsymbol{X})$$

一阶实矩阵微分：Jacobian矩阵辨识

标量函数的Jacobian矩阵辨识

$$\mathrm{d}f(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{x}) \Longleftrightarrow \mathrm{D}_{\boldsymbol{x}} f(\boldsymbol{x}) = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{A}$$

$$\mathrm{d}f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{X}) \Longleftrightarrow \mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}} = \boldsymbol{A}$$

$$\mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}} = \boldsymbol{A} \Longleftrightarrow \nabla_{\boldsymbol{X}} f(\boldsymbol{X}) = \boldsymbol{A}^{\mathrm{T}}$$

习题1：利用矩阵微分证明二次型函数$f(\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ 变元向量 $\boldsymbol{x}$ 的梯度向量

$$\nabla_{\boldsymbol{x}}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\right)=\left(\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{A}\right)\boldsymbol{x}$$

习题2：利用Jacobian矩阵辨识，求包含逆矩阵的迹函数 $\mathrm{tr}\left(\boldsymbol{A}\boldsymbol{X}^{-1}\right)$的梯度矩阵

一阶实矩阵微分：Jacobian矩阵辨识

示例 已知 $\frac{\partial f}{\partial \boldsymbol{Y}}$ 和 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$，求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $$\mathrm{d}f=\mathbf{tr}\left[\left(\frac{\partial f}{\partial \boldsymbol{Y}}\right)^{\mathrm{T}}\mathrm{d}\boldsymbol{Y}\right]=\mathbf{tr}\left[\left(\frac{\partial f}{\partial \boldsymbol{Y}}\right)^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\mathbf{X}\boldsymbol{B}\right]=\mathbf{tr}\left[\boldsymbol{B}\left(\frac{\partial f}{\partial \boldsymbol{Y}}\right)^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\mathbf{X}\right]$$

$$=\mathbf{tr}\left[\left(\boldsymbol{A}^{\mathrm{T}}\frac{\partial f}{\partial \boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}}\right)^{\mathrm{T}}\mathrm{d}\mathbf{X}\right],$$

迹的性质：交换顺序

得 $\frac{\partial f}{\partial \boldsymbol{X}}=\boldsymbol{A}^{\mathrm{T}}\frac{\partial f}{\partial \boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}}$。

一阶实矩阵微分：Jacobian矩阵辨识

示例 已知 $f = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{b}$，求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $\mathrm{d}f = \mathrm{d}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{b} + \boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\boldsymbol{b} + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\mathrm{d}\boldsymbol{b} = \mathbf{tr}\left[\boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\boldsymbol{b}\right]$

0

$= \mathbf{tr}\left[\boldsymbol{b}\boldsymbol{a}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right] = \mathbf{tr}\left[\left(\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}\right)^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]$

得 $\frac{\partial f}{\partial \boldsymbol{X}} = \boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$。

笔记 上述两个例子反复应用了 $\mathbf{tr}(\boldsymbol{A}\boldsymbol{B}) = \mathbf{tr}(\boldsymbol{B}\boldsymbol{A})$（$\boldsymbol{A}\boldsymbol{B}$ 和 $\boldsymbol{B}\boldsymbol{A}$ 特征值相同）。

一阶实矩阵微分：Jacobian矩阵辨识

示例 已知 $f=\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{X}\right]$，求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $\mathrm{d}f=\mathbf{tr}\left[\mathrm{d}\left(\boldsymbol{X}^{\mathrm{T}}\right)\boldsymbol{S}\boldsymbol{X}\right]+\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\mathrm{d}\boldsymbol{X}\right]=\mathbf{tr}\left[(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{S}\boldsymbol{X}\right]+\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\mathrm{d}\boldsymbol{X}\right]$
$=\mathbf{tr}\left[(\boldsymbol{S}\boldsymbol{X})^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]+\mathbf{tr}\left[\boldsymbol{X}^{\mathrm{T}}\boldsymbol{S}\mathrm{d}\boldsymbol{X}\right]=\mathbf{tr}\left[2(\boldsymbol{S}\boldsymbol{X})^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]$，得 $\frac{\partial f}{\partial \boldsymbol{X}}=2\boldsymbol{S}\boldsymbol{X}$。

示例 已知 $f=\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|_2^2$（实数域），求 $\frac{\partial f}{\partial \boldsymbol{X}}$?

解 $\mathrm{d}f=\mathrm{d}\left[(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\right]=\mathrm{d}(\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})+(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{x}$
$=(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{x}+(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{x}=\mathbf{tr}\left[2\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}\right)^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right]$，
得 $\frac{\partial f}{\partial \boldsymbol{X}}=2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}-2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$。

Hessian矩阵

定义 假设有一实值函数 $f(x_1,\cdots,x_n)$，如果 f 的所有二阶偏导数都存在并在定义域内连续，那么函数 f 的**海森矩阵**（Hessian matrix）为

$$\boldsymbol{H}[f(\boldsymbol{x})]=\frac{\partial^2 f(\boldsymbol{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathrm{T}}}=\frac{\partial}{\partial\mathbf{x}}\left[\frac{\partial f(\boldsymbol{x})}{\partial\mathbf{x}^{\mathrm{T}}}\right]\in\mathbb{R}^{m\times m}$$

显然海森矩阵是一个对称矩阵

$$\boldsymbol{H}[f(\boldsymbol{x})]=\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x})=\nabla_{\boldsymbol{x}}\left(\mathrm{D}_{\boldsymbol{x}} f(\boldsymbol{x})\right)$$

$$[\boldsymbol{H}f(\boldsymbol{x})]_{i,j}=\left[\frac{\partial^2 f(\boldsymbol{x})}{\partial\mathbf{x}\partial\mathbf{x}^T}\right]_{i,j}=\frac{\partial}{\partial x_i}\left[\frac{\partial f(\boldsymbol{x})}{\partial x_j}\right]$$

$$\boldsymbol{H}[f(\boldsymbol{x})]=\frac{\partial^2 f(\boldsymbol{x})}{\partial\mathbf{x}\partial\mathbf{x}^T}=\begin{bmatrix}\frac{\partial^2 f}{\partial x_1\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_m}\\ \vdots & \ddots & \vdots\\ \frac{\partial^2 f}{\partial x_m\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_m\partial x_m}\end{bmatrix}\in\mathbb{R}^{m\times m}$$

示例 $f=x^2-y^2$，$\boldsymbol{H}_f=\begin{bmatrix}2 & 0\\ 0 & -2\end{bmatrix}$。

第三章习题

见学在浙大
作业版块
Homework2

10.13 交