矩阵微分

主 讲 人：马 丽 艳
办 公 室：计 1013
Email: liyanma@t.shu.edu.cn
计算机工程与科学学院

2020-7-1

主要内容

- **Jacobian矩阵与梯度分析**
- **一阶实矩阵微分与Jacobian矩阵辨识**

矩阵微分

量与量之间的对应关系(自变量←→ 应变量):

$$\{标量,向量,矩阵\} \longleftrightarrow \{标量,向量,矩阵\}$$

（1）实函数矩阵对标量变元的导数

（2）实矩阵函数对矩阵变元的导数

（3）梯度矩阵、Jacobian矩阵与Hessian矩阵

（4）实值标量函数的矩阵微分及计算

矩阵微分

实值函数的分类

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

矩阵微分

复值函数的分类

函数类型\变量类型	标量变元$z, z^* \in \mathbb{C}$	向量变元$z, z^* \in \mathbb{C}^m$	矩阵变元$Z, Z^* \in \mathbb{C}^{m\times n}$
标量函数$f \in \mathbb{C}$	$f(z, z^*)$ $f: \mathbb{C} \times \mathbb{C} \to \mathbb{C}$	$f(z, z^*)$ $f: \mathbb{C}^m \times \mathbb{C}^m \to \mathbb{C}$	$f(z, z^*)$ $f: \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \to \mathbb{C}$
向量函数$f \in \mathbb{C}^p$	$f(z, z^*)$ $f: \mathbb{C} \times \mathbb{C} \to \mathbb{C}^p$	$f(z, z^*)$ $f: \mathbb{C}^m \times \mathbb{C}^m \to \mathbb{C}^p$	$f(z, z^*)$ $f: \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \to \mathbb{C}^p$
矩阵函数$F \in \mathbb{C}^{p\times q}$	$F(z, z^*)$ $F: \mathbb{C} \times \mathbb{C} \to \mathbb{C}^{p\times q}$	$F(z, z^*)$ $F: \mathbb{C}^m \times \mathbb{C}^m \to \mathbb{C}^{p\times q}$	$F(z, z^*)$ $F: \mathbb{C}^{m\times n} \times \mathbb{C}^{m\times n} \to \mathbb{C}^{p\times q}$

数列极限

- **数列极限的定义**

"$\varepsilon-N$" 定义：设 $\{a_n\}$ 为数列， a 为定数.

若 $\forall\varepsilon>0, \exists N$, 当 $n>N$ 时，有

$$|a_n-a|<\varepsilon,$$

称数列 $\{a_n\}$ 收敛于 a.

函数极限

1^0 “ $\varepsilon-\delta$ ” 定义.

设函数 $f(x)$ 在点 x_0 的某空心邻域内有定义，A 为定数. 若 $\forall\varepsilon>0,\exists\delta>0$, 当 $0<|x-x_0|<\delta$ 时，有

$$|f(x)-A|<\varepsilon.$$

则称函数 $f(x)$ 当 $x\to x_0$ 时以 A 为极限.

2^0 $x\to+\infty$ 的极限定义.

设函数 $f(x)$ 在 $[a,+\infty)$ 上有定义，A 为定数.

$\forall\varepsilon>0,\exists M\geq a$, 当 $x>M$ 时，有

$$|f(x)-A|<\varepsilon.$$

则称函数 $f(x)$ 当 $x\to+\infty$ 时以 A 为极限.

一阶导数

设函数 $f(x)$ 在 x_0 点某邻域内有定义，若极限

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

存在，则称 $f(x)$ 在 x_0 点可导，并称极限值为函数 $f(x)$ 在 x_0 的导数，记 $f'(x_0), f'(x)|_{x=x_0}$.

若极限不存在，则称 $f(x)$ 在 x_0 点不可导.

函数可导的充要条件是函数左、右导数存在且相等.

一阶微分

若存在常数 A 使得 $f(x_0+\Delta x)-f(x_0)=A\Delta x+o(\Delta x)$,
则称 $f(x)$ 在 x_0 点可微，并成 $A\Delta x$ 为 $f(x)$ 在 x_0 点的微分,
记 $df(x)|_{x=x_0}=A\Delta x=Adx$.

(i) 当 x 为自变量时 $dx=\Delta x$. 由导数定义可以证明 $A=f'(x_0)$.

(ii) 可微的定义就是 $f(x)=f(x_0)+f'(x_0)(x-x_0)+o(x-x_0)$,
即 $f(x)$ 等于线性函数 (直线) 加上 x 与 x_0 距离的高阶无穷小量.

(iii) 对一元函数而言，可导与可微是等价的. 但描述的几何意义不同，
导数描述的是变化率，微分其实就是在局部可 “以直代曲”.

(iv) 一阶微分的定义其实就是 Taylor 公式的特殊情况， Taylor 公式
是用多项式局部逼近函数，而微分是用线性函数 (一次多项式)
局部逼近函数.

高阶微分

高阶微分定义： n 阶微分是 $n-1$ 阶微分的微分，记 $d^n f(x)$, 即

$$d^n f(x) = d(d^{n-1} f(x)) = d(f^{(n-1)} dx^{n-1}) = f^{(n)}(x) dx^n.$$

泰勒定理

(1) 设函数 $f(x)$ 在 x_0 点存在直至 n 阶导数，则

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2$$
$$+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+o((x-x_0)^n).$$

(2) 若函数 $f(x)$ 在 $[a,b]$ 上存在直至 n 阶的连续导函数，在 (a,b) 内存在 $(n+1)$ 阶导函数，则对 $\forall x,x_0\in[a,b]$ 至少存在 $\xi\in(a,b)$ 使得

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2$$
$$+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n+\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{(n+1)}.$$

偏导数

(1) 偏导数定义： $f'_x(x_0,y_0)=\lim\limits_{\Delta x\to 0}\dfrac{f(x_0+\Delta x,y_0)-f(x_0,y_0)}{\Delta x}$.

函数 f 对一个自变量求偏导数，是先把其它自变量看作常数，从而变成一元函数的求导问题.

(2) 偏导数的几何意义： $f'_x(x_0,y_0)$ 就是一元函数 $f(x,y_0)$ 在 $x=x_0$ 的导数，就是曲线

$$\mathrm{L}:\begin{cases} x=x, \\ y=y_0, \\ z=f(x,y), \end{cases}$$

在 (x_0,y_0,z_0)(其中 $z_0=f(x_0,y_0)$) 处的切线 T_x 对于 x 轴的斜率，即 T_x 与 x 轴正向所成倾角 α 的正切 $\tan\alpha$.

全微分

(1) 全微分定义：若 $\Delta z = f(x_0+\Delta x, y_0+\Delta y)-f(x_0,y_0) = A\Delta x+B\Delta y+o(\sqrt{\Delta x^2+\Delta y^2})$，称 f 在 (x_0,y_0) 点可微，并称关于 $\Delta x, \Delta y$ 的线性函数 $A\Delta x+B\Delta y$ 为 f 在 (x_0,y_0) 点的全微分，记 $df(x_0,y_0) = A\Delta x+B\Delta y$.

(2) 全微分的几何意义：如果函数 f 在 (x_0,y_0) 点可微，则曲面 $z=f(x,y)$ 在 (x_0,y_0,z_0)(其中 $z_0=f(x_0,y_0)$) 处存在切平面：

$$z-z_0 = f'_x(x_0,y_0)(x-x_0)+f'_y(x_0,y_0)(y-y_0)$$

当点 (x_0,y_0) 变为 $(x_0+\Delta x, y_0+\Delta y)$ 时，Δz 是曲面的增量，而全微分是切平面的增量，函数在某点可微就是在该点附近可以用切平面近似代替曲面.

方向导数

$z=f(x,y)$ 的偏导数 f'_x, f'_y 是函数 $z=f(x,y)$ 沿两个特殊方向的变化率，要考虑函数沿其它特定方向的变化率.

(1) 方向导数的定义：设函数 $z=f(x,y)$ 在点 $P_0(x_0,y_0)$ 的某邻域内有定义， l 是通过 P_0 的任意一条有向直线，其正向与 x,y 轴的正向间的夹角分别为 α,β，再设 $P(x_0+\Delta x, y_0+\Delta y)$ 是 l 上任意一点，记 $\rho=\sqrt{\Delta x^2+\Delta y^2}$，则 $\Delta x=\rho\cos\alpha, \Delta y=\rho\cos\beta$，若

$$\lim_{\rho\to 0}\frac{f(x_0+\rho\cos\alpha, y_0+\rho\cos\beta)-f(x_0,y_0)}{\rho}$$

存在，则称此极限为 $f(x,y)$ 在 P_0 点沿方向 l 的方向导数，记 $\frac{\partial f}{\partial l}|_{p_0}$.

当 l 为 x 轴正方向时，方向导数恰好为 $\frac{\partial f}{\partial l}|_{p_0}=\frac{\partial f}{\partial x}|_{p_0}$;

当 l 为 x 轴负方向时，方向导数恰好为 $\frac{\partial f}{\partial l}|_{p_0}=-\frac{\partial f}{\partial x}|_{p_0}$.

梯度

(1) 梯度定义：设 $f(x,y,z)$ 在点 $P_0(x_0,y_0,z_0)$ 存在对所有自变量的偏导数，则称向量 $(f'_x(P_0),f'_y(P_0),f'_z(P_0))$ 为函数在 f 在 P_0 的梯度，记为

$$\mathrm{grad}\,f(P_0)=(f'_x(P_0),f'_y(P_0),f'_z(P_0)).$$

梯度就是将数量场映射成向量场的算子.

梯度

(2) 梯度的几何意义：设 $l=(\cos\alpha,\cos\beta,\cos\gamma)$, 则方向导数公式又可写成

$$\frac{\partial f}{\partial l}(P_0)=\operatorname{grad} f(P_0)\cdot l=|\operatorname{grad} f(P_0)|\cos\theta,$$

其中 θ 为梯度向量与 l 的夹角.

当 f 在 P_0 点可微时， $\operatorname{grad} f(P_0)$ 就是 f 的值增长最快的方向, 且沿这一方向的变化率就是梯度的模 $|\operatorname{grad} f(P_0)|$;
$-\operatorname{grad} f(P_0)$ 是方向导数取得最小值 $-|\operatorname{grad} f(P_0)|$ 的方向.

函数矩阵导数

Definition 32. 以实变量 t 的函数为元素的矩阵

$$\boldsymbol{A}(t)=\begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & a_{22}(t) & \cdots & a_{2n}(t) \\ \vdots & \vdots & & \vdots \\ a_{m1}(t) & a_{m2}(t) & \cdots & a_{mn}(t) \end{bmatrix}$$

称为函数矩阵, 其中 $a_{ij}(t)$ 都是定义在 $[a,b]$ 上的实函数.

函数矩阵导数

函数矩阵 $\boldsymbol{A}(t)$ 在 $[a,b]$ 上有界、有极限、连续、可微、可积等概念，可用其 $m\times n$ 个元素 $a_{ij}(t)$ 同时在 $[a,b]$ 上有界、有极限、连续、可微、可积来定义. 例如

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\left[\frac{\mathrm{d}}{\mathrm{d}t}a_{ij}(t)\right]_{m\times n};$$

$$\int \boldsymbol{A}(t)\,\mathrm{d}t=\left[\int a_{ij}(t)\,\mathrm{d}t\right]_{m\times n};$$

$$\int_a^b \boldsymbol{A}(t)\,\mathrm{d}t=\left[\int_a^b a_{ij}(t)\,\mathrm{d}t\right]_{m\times n}.$$

函数矩阵导数

Definition 33. 如果所有的元素 $a_{ij}(t)$ 在 $t \to t_0$ 时, 极限存在, 记为常数 a_{ij}, 即

$$\lim_{t \to t_0} a_{ij}(t) = a_{ij},$$

则称矩阵 $\boldsymbol{A}(t)$ 在 $t \to t_0$ 时, 极限存在, 且极限值为 $\boldsymbol{A}$ (常量矩阵), 即

$$\lim_{t \to t_0} \boldsymbol{A}(t) = \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

函数矩阵导数

Definition 34. 如果所有的元素 $a_{ij}(t)$ 在 $t = t_0$ 连续, 即

$$\lim_{t \to t_0} a_{ij}(t) = a_{ij}(t_0),$$

则称矩阵 $\boldsymbol{A}(t)$ 在 $t = t_0$ 连续, 且记为

$$\lim_{t \to t_0} \boldsymbol{A}(t) = \boldsymbol{A}(t_0) = \begin{bmatrix} a_{11}(t_0) & a_{12}(t_0) & \cdots & a_{1n}(t_0) \\ a_{21}(t_0) & a_{22}(t_0) & \cdots & a_{2n}(t_0) \\ \vdots & \vdots & & \vdots \\ a_{m1}(t_0) & a_{m2}(t_0) & \cdots & a_{mn}(t_0) \end{bmatrix}.$$

函数矩阵导数

容易验证下列等式是成立的.

设 $\lim_{t\to t_0}\boldsymbol{A}(t)=\boldsymbol{A},\ \lim_{t\to t_0}\boldsymbol{B}(t)=\boldsymbol{B}$.

1. 若 $\boldsymbol{A}(t),\ \boldsymbol{B}(t)$ 都是 $m\times n$ 阶矩阵, 则

$$\lim_{t\to t_0}[\boldsymbol{A}(t)+\boldsymbol{B}(t)]=\boldsymbol{A}+\boldsymbol{B}=\lim_{t\to t_0}\boldsymbol{A}(t)+\lim_{t\to t_0}\boldsymbol{B}(t).$$

2. 设 k 为常数, 则

$$\lim_{t\to t_0}[k\boldsymbol{A}(t)]=k\boldsymbol{A}.$$

3. 若 $\boldsymbol{A}(t),\ \boldsymbol{B}(t)$ 分别是 $m\times n$ 阶及 $n\times r$ 阶矩阵, 则

$$\lim_{t\to t_0}[\boldsymbol{A}(t)\boldsymbol{B}(t)]=\boldsymbol{A}\boldsymbol{B}=\lim_{t\to t_0}\boldsymbol{A}(t)\lim_{t\to t_0}\boldsymbol{B}(t).$$

函数矩阵导数

Definition 3.4

若函数矩阵 $\boldsymbol{A}(t)$ 中所有的元素 $a_{ij}(t)$ 在 t_0 处 (或在区间 (a,b) 上) 可微, 则称函数矩阵 $\boldsymbol{A}(t)$ 在 t_0 处 (或在区间 (a,b) 上)可微. 并记为

$$\begin{aligned}
\boldsymbol{A}'(t_0) &= \left.\frac{\mathrm{d}\boldsymbol{A}(t)}{\mathrm{d}t}\right|_{t=t_0} \\
&= \lim_{\Delta t\to 0}\frac{\boldsymbol{A}(t_0+\Delta t)-\boldsymbol{A}(t_0)}{\Delta t} \\
&= \begin{bmatrix} a'_{11}(t_0) & a'_{12}(t_0) & \cdots & a'_{1n}(t_0) \\ a'_{21}(t_0) & a'_{22}(t_0) & \cdots & a'_{2n}(t_0) \\ \vdots & \vdots & & \vdots \\ a'_{m1}(t_0) & a'_{m2}(t_0) & \cdots & a'_{mn}(t_0) \end{bmatrix}.
\end{aligned}$$

函数矩阵导数

Example 3.5

已知 $\boldsymbol{A}(t) = \begin{bmatrix} \sin t & 2t^3 \\ 2\sqrt{t} & \mathrm{e}^{2t} \end{bmatrix}$，求 $\dfrac{\mathrm{d}\boldsymbol{A}(t)}{\mathrm{d}t}$.

解:

$$\frac{\mathrm{d}\boldsymbol{A}(t)}{\mathrm{d}t} = \begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}t}(\sin t) & \frac{\mathrm{d}}{\mathrm{d}t}(2t^3) \\ \frac{\mathrm{d}}{\mathrm{d}t}(2\sqrt{t}) & \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{2t}) \end{bmatrix} = \begin{bmatrix} \cos t & 6t^2 \\ \frac{1}{\sqrt{t}} & 2\mathrm{e}^{2t} \end{bmatrix}. \quad \square$$

函数矩阵导数

关于函数矩阵, 有下面的求导法则:

1. 若 $\boldsymbol{A}(t)$, $\boldsymbol{B}(t)$ 都是 $m\times n$ 阶矩阵可微矩阵, 则

$$\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{A}(t)+\boldsymbol{B}(t)]=\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)+\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t).$$

2. 若 $\boldsymbol{A}(t)$, $\boldsymbol{B}(t)$ 分别是 $m\times n$ 阶及 $n\times r$ 阶矩阵, 则

$$\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{A}(t)\boldsymbol{B}(t)]=\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right]\boldsymbol{B}(t)+\boldsymbol{A}(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)\right].$$

3. 若 $\boldsymbol{A}(u)$ 可微, 且 $u=f(t)$ 关于 t 可微, 则

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(f(t))=f'(t)\frac{\mathrm{d}}{\mathrm{d}u}\boldsymbol{A}(u).$$

4. 若 $\boldsymbol{A}(t)$ 与 $\boldsymbol{A}^{-1}(t)$ 都可微, 则

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{A}^{-1}(t)\right)=-\boldsymbol{A}^{-1}(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right]\boldsymbol{A}^{-1}(t).$$

函数矩阵导数

证: (4) 注意到 $\boldsymbol{A}(t)\boldsymbol{A}^{-1}(t)=\boldsymbol{I}$, 两端对 t 求导, 得

$$\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right]\boldsymbol{A}^{-1}(t)+\boldsymbol{A}(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{A}^{-1}(t)\right)\right]=\boldsymbol{O},$$

即

$$\boldsymbol{A}(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{A}^{-1}(t)\right)\right]=-\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right]\boldsymbol{A}^{-1}(t)$$

两边左乘以 $\boldsymbol{A}^{-1}(t)$, 得

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{A}^{-1}(t)\right)=-\boldsymbol{A}^{-1}(t)\left[\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right]\boldsymbol{A}^{-1}(t). \quad \square$$

☞ 注意矩阵乘法不满足交换律. 例如

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^2(t)=\frac{\mathrm{d}}{\mathrm{d}t}[\boldsymbol{A}(t)\boldsymbol{A}(t)]=\boldsymbol{A}'(t)\cdot\boldsymbol{A}(t)+\boldsymbol{A}(t)\cdot\boldsymbol{A}'(t)$$

$$\neq 2\boldsymbol{A}(t)\cdot\boldsymbol{A}'(t).$$

函数矩阵导数

Definition 3.6

若 $\boldsymbol{A}(t)$ 中所有的元素 $a_{ij}(t)$ 在区间 $[a, b]$ 上可积, 则称函数矩阵 $\boldsymbol{A}(t)$ 在区间 $[a, b]$ 上可积, 并规定

$$\int_a^b \boldsymbol{A}(t)\,\mathrm{d}t = \Big[\int_a^b a_{ij}(t)\,\mathrm{d}t\Big]_{m\times n}.$$

Theorem 3.7

1. 若 $\boldsymbol{A}(t)$ 在区间 $[a, b]$ 上连续, 则对任一 $t \in (a, b)$, $\int_a^t \boldsymbol{A}(\tau)\,\mathrm{d}\tau$ 可微, 且

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big[\int_a^t \boldsymbol{A}(\tau)\,\mathrm{d}\tau\Big] = \boldsymbol{A}(t).$$

2. 若 $\boldsymbol{A}(t)$ 在区间 $[a, b]$ 上可微, 则

$$\int_a^t \Big[\frac{\mathrm{d}}{\mathrm{d}s}\boldsymbol{A}(s)\Big]\mathrm{d}s = \boldsymbol{A}(t) - \boldsymbol{A}(a), \qquad t \in [a, b].$$

函数矩阵导数

例如, 对于如下函数的幂级数展开式

$$(1-z)^{-1}=\sum_{m=0}^{\infty} z^m \qquad (R=1),$$

$$\ln(1+z)=\sum_{m=0}^{\infty} \frac{(-1)^m}{m+1} z^{m+1} \qquad (R=1).$$

相应地有矩阵函数

$$(\boldsymbol{I}-\boldsymbol{A})^{-1}=\sum_{m=0}^{\infty} \boldsymbol{A}^m \qquad (\rho(\boldsymbol{A})<1),$$

$$\ln(\boldsymbol{I}+\boldsymbol{A})=\sum_{m=0}^{\infty} \frac{(-1)^m}{m+1} \boldsymbol{A}^{m+1} \qquad (\rho(\boldsymbol{A})<1).$$

以及

- 矩阵指数函数: $\mathrm{e}^{\boldsymbol{A}}=\boldsymbol{I}+\boldsymbol{A}+\frac{\boldsymbol{A}^2}{2!}+\cdots+\frac{\boldsymbol{A}^k}{k!}+\cdots.$
- 矩阵正弦函数: $\sin \boldsymbol{A}=\boldsymbol{A}-\frac{\boldsymbol{A}^3}{3!}+\cdots+(-1)^{k-1}\frac{\boldsymbol{A}^{2k-1}}{(2k-1)!}+\cdots,$
- 矩阵余弦函数: $\cos \boldsymbol{A}=\boldsymbol{I}-\frac{\boldsymbol{A}^2}{2!}+\cdots+(-1)^k\frac{\boldsymbol{A}^{2k}}{(2k)!}+\cdots.$

函数矩阵导数

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(a(t)A(t)\right)=\frac{\mathrm{d}a(t)}{\mathrm{d}t}A(t)+a(t)\frac{d}{\mathrm{d}t}A(t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{tA}=Ae^{tA}=\ e^{tA}A$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\cos(tA)=-Asin(tA)=-\sin(tA)\,A$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sin(tA)=Acos(tA)=\cos(tA)\,A$$

◆ 高阶导数

$$\frac{\mathrm{d}^k A(t)}{\mathrm{d}t^k}=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}^{k-1}A(t)}{\mathrm{d}t^{k-1}}\right)=\left(\frac{\mathrm{d}^k a_{ij}(t)}{\mathrm{d}\ ^k}\right)_{m\times n}$$

函数矩阵导数

应用：矩阵微分方程的解

定理1：满足初始条件 $\mathbf{x}(t)|_{t=t_0}=\mathbf{x}(t_0)$ 的一阶线性常系数齐次微分方程组

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t}=\mathbf{A}\mathbf{x}(t)$$

有且仅有唯一解 $\mathbf{x}(t)=\mathrm{e}^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0)$

函数矩阵导数

应用：矩阵微分方程的解

定理2： 一阶线性常系数非齐次微分方程组

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{A}\mathbf{x}(t) + \mathbf{b}(t)$$

的通解为$\mathbf{x}(t) = \mathrm{e}^{tA}\left(\mathbf{c} + \int_{t_0}^{t} \mathrm{e}^{-sA}\,\mathbf{b}(s)\mathrm{d}s\right)$

其中c为任意常数向量。

函数矩阵导数

应用：矩阵微分方程的解

定理3： n阶常系数齐次线性微分方程

$$\begin{cases} x^{(n)}(t)+a_1x^{(n-1)}(t)+a_2x^{(n-2)}(t)+\cdots+a_nx(t)=0 \\ x^{(i)}(t)\Big|_{t=t_0}=x^{(i)}(t_0), \quad i=0,1,\cdots,n-1 \end{cases}$$

的解为：

$$x(t)=[1 \quad 0 \quad \cdots \quad 0]\mathrm{e}^{(t-t_0)A}\begin{bmatrix} x(t_0) \\ x'(t_0) \\ \vdots \\ x^{(n-1)}(t_0) \end{bmatrix}$$

其中

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix}$$

函数矩阵导数

应用：矩阵微分方程的解

定理4： n阶常系数非齐次线性微分方程

$$x^{(n)}(t)+a_1x^{(n-1)}(t)+a_2x^{(n-2)}(t)+\cdots+a_nx(t)=f(t)$$

的通解为

$$x(t)=[10\cdots 0]\left(\mathrm{e}^{tA}c+\int_{t_0}^{t}\mathrm{e}^{A(t-s)}bf(s)\mathrm{d}s\right)$$

其中c为任意常数向量；$b=[0\quad 0\quad \cdots\quad 1]^{\mathrm{T}}$；而$A$同定理3.

标量函数关于矩阵导数

在场论中，我们对数量函数 $f(x,y,z)$ 定义梯度为

$$\mathbf{grad}\, f = \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right),$$

这可以理解为数量函数 $f(x,y,z)$ 对向量 (x,y,z) 的导数.

注意梯度是向量，有时也记为

$$\mathbf{grad}\, f = \frac{\partial f}{\partial x}\boldsymbol{i} + \frac{\partial f}{\partial y}\boldsymbol{j} + \frac{\partial f}{\partial z}\boldsymbol{k}.$$

下面我们将这一概念推广到一般情形.

标量函数关于矩阵导数

函数 u 的梯度定义为

$$\nabla u = \frac{\partial u}{\partial x}\boldsymbol{i} + \frac{\partial u}{\partial y}\boldsymbol{j} + \frac{\partial u}{\partial z}\boldsymbol{k}$$

$\boldsymbol{E}$ 的散度定义为

$$\nabla \cdot \boldsymbol{E} = (\frac{\partial}{\partial x}\boldsymbol{i} + \frac{\partial}{\partial y}\boldsymbol{j} + \frac{\partial}{\partial z}\boldsymbol{k}) \cdot (E_x\boldsymbol{i} + E_y\boldsymbol{j} + E_z\boldsymbol{k})$$

$$= \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z}$$

From：黄正华

标量函数关于矩阵导数

Definition 3.8

设 $\boldsymbol{X}=[x_{ij}]\in\mathbb{R}^{m\times n}$ 为变量矩阵, $f(\boldsymbol{X})$ 为矩阵 $\boldsymbol{X}$ 的数量函数, 即看成是 $m\times n$ 元函数, 即

$$f(\boldsymbol{X})=f\left(x_{11},x_{12},\cdots,x_{1n},x_{21},x_{22},\cdots,x_{2n},\cdots,x_{m1},x_{m2},\cdots,x_{mn}\right).$$

则规定数量函数 $f(\boldsymbol{X})$ 对于矩阵 $\boldsymbol{X}$ 的导数为

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}}=\left[\frac{\partial f}{\partial x_{ij}}\right]_{m\times n}=\begin{bmatrix}\frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{bmatrix}.$$

From：黄正华

标量函数关于矩阵导数

特别地，若变量矩阵为 m 维向量 $\boldsymbol{X} = [x_i]_{m\times 1}$，这时数量函数为 $f(x_1, x_2, \cdots, x_n)$，则

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{X}} = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_m} \end{bmatrix}.$$

此即梯度.

From：黄正华

标量函数关于矩阵导数

◆列向量偏导和行向量偏导($\mathbf{x} \in \mathbb{R}^{m\times1}$)

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

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

链式法则：A为矩阵，$y = f(A)$和$g(y)$分别是实值函数，则

$$\frac{\partial g(f(A))}{\partial A} = \frac{dg(y)}{dy}\frac{\partial f(A)}{\partial A}$$

标量函数关于矩阵导数

（1）若$\mathbf{X} \in \mathbb{R}^{m\times n}$且$f(\mathbf{X}) = c$为常数，则$\frac{\mathrm{d}c}{\mathrm{d}X} = O_{m\times n}$.

（2）若c_1, c_2为实常数，则

$$\frac{\mathrm{d}(c_1 f(\mathbf{X}) + c_2 g(\mathbf{X}))}{\mathrm{d}\mathbf{X}} = c_1 \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}} + c_2 \frac{\mathrm{d}g(\mathbf{X})}{\mathrm{d}\mathbf{X}}.$$

（3）

$$\frac{\mathrm{d}f(\mathbf{X})g(\mathbf{X})}{\mathrm{d}\mathbf{X}} = g(\mathbf{X}) \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}} + f(\mathbf{X}) \frac{\mathrm{d}g(\mathbf{X})}{\mathrm{d}\mathbf{X}}.$$

（4）若$g(\mathbf{X}) \neq 0$，则

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

标量函数关于矩阵导数

（5）$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b}}{\mathrm{d}\mathbf{X}} = \mathbf{a}\mathbf{b}^{\mathrm{T}} \quad (\mathbf{X} \in \mathbb{R}^{m\times n})$

（6）$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\mathbf{b}}{\mathrm{d}\mathbf{X}} = \boldsymbol{b}\boldsymbol{a}^{\mathrm{T}} \quad (\mathbf{X} \in \mathbb{R}^{m\times n})$

（7）若$\mathbf{X} \in \mathbb{R}^{n\times n}$非奇异，则

$$\frac{\mathrm{d}\mathbf{a}^{\mathrm{T}}\mathbf{X}^{-1}\mathbf{b}}{\mathrm{d}\mathbf{X}} = -\mathbf{X}^{-\mathrm{T}}\mathbf{a}\mathbf{b}^{\mathrm{T}}\mathbf{X}^{-\mathrm{T}}$$

（8）$\frac{\mathrm{d}a^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b}}{\mathrm{d}\mathbf{X}} = \mathbf{X}\left(\mathbf{a}\mathbf{b}^{\mathrm{T}} + \mathbf{b}\mathbf{a}^{\mathrm{T}}\right)$

（9）$\frac{\mathrm{d}\exp(\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b})}{\mathrm{d}\mathbf{X}} = \mathbf{a}\mathbf{b}^{\mathrm{T}}\exp\left(\mathbf{a}^{\mathrm{T}}\mathbf{X}\mathbf{b}\right)$

标量函数关于矩阵导数

（10） $$\frac{d\mathbf{a}^T\mathbf{x}}{d\mathbf{x}} = \frac{d\mathbf{x}^T\mathbf{a}}{d\mathbf{x}} = \mathbf{a}$$

（11） $$\frac{d\mathbf{x}^T\mathbf{A}\mathbf{b}}{d\mathbf{x}} = \mathbf{A}\mathbf{b}, \quad \frac{d\mathbf{b}^T\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \mathbf{A}^T\mathbf{b}$$

（12） $$\frac{d\mathbf{x}^T\mathbf{A}\mathbf{x}}{d\mathbf{x}} = \left(\mathbf{A} + \mathbf{A}^T\right)\mathbf{x}$$

标量函数关于矩阵导数

```
clear
a=rand(1,4)';
b=rand(1,4)';
X=rand(4,4);
delta=0.01;
Z=zeros(4,4);
for ii=1:4
    for jj=1:4
        XX=X;
        XX(ii,jj)=XX(ii,jj)+delta;
        f=a'*XX*b-a'*X*b;
        Z(ii,jj)=f/delta;
    end
end
disp(Z);   disp(a*b');
```

```
>> Z
Z =
    0.5152    0.0795    0.2269    0.4456
    0.5728    0.0884    0.2523    0.4954
    0.0803    0.0124    0.0354    0.0694
    0.5776    0.0891    0.2544    0.4995

>> a*b'
ans =
    0.5152    0.0795    0.2269    0.4456
    0.5728    0.0884    0.2523    0.4954
    0.0803    0.0124    0.0354    0.0694
    0.5776    0.0891    0.2544    0.4995
```

标量函数关于矩阵导数

```
clear
a=rand(1,4)';  b=rand(1,4)';
X=rand(4,4);  Z=zeros(4,4);
X1=inv(X);
delta=0.001;
for ii=1:4
    for jj=1:4
        XX=X;
        XX(ii,jj)=XX(ii,jj)+delta;
        X2=inv(XX);
        f=a'*X2*b-a'*X1*b;
        Z(ii,jj)=f/delta;
    end
end
```

```
>> Z
Z =

  -2.1607    1.5003    1.3472    0.4124
   1.7311   -1.1853   -1.0604   -0.3269
   2.6782   -1.8321   -1.6455   -0.5061
  -1.2717    0.8770    0.7862    0.2407

>> -X1'*a*b'*X1'
ans =
  -2.1752    1.4957    1.3419    0.4122
   1.7249   -1.1860   -1.0641   -0.3269
   2.6710   -1.8365   -1.6478   -0.5062
  -1.2729    0.8753    0.7853    0.2412
```

矩阵迹的微分

$$d(tr(X)) = d(\sum_i x_{ii}) = \sum_i dx_{ii} = tr(dX)$$

若$A = \dfrac{\partial f(x)}{\partial x^T}$，则一阶微分$df(x) = \dfrac{\partial f(x)}{\partial x^T} dx = tr(Adx)$

$$df(x) = tr(Adx) \Leftrightarrow \frac{\partial f(x)}{\partial x^T} A$$

矩阵迹的微分

(13) $$\frac{\mathrm{dtr}(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \mathbf{I}$$

(14) $$\frac{\mathrm{dtr}(\mathbf{X}^{-1})}{\mathrm{d}\mathbf{X}} = -(\mathbf{X}^{-2})^{\mathrm{T}}$$

(15) $$\frac{\mathrm{dtr}(\mathbf{AX})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{dtr}(\mathbf{XA})}{\mathrm{d}\mathbf{X}} = \mathbf{A}^{\mathrm{T}}$$

(16)

$$\frac{\mathrm{dtr}(\mathbf{AX}^{\mathrm{T}})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{dtr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})}{\mathrm{d}\mathbf{X}} = \mathbf{A}, \qquad \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AXB}) = \mathbf{A}^T\mathbf{B}^T$$

$$\frac{\mathrm{dtr}(\mathbf{ax}^{\mathrm{T}})}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{dtr}(\mathbf{xa}^{\mathrm{T}})}{\mathrm{d}\mathbf{x}} = \mathbf{a}. \qquad \frac{\partial}{\partial \mathbf{X}}\mathrm{Tr}(\mathbf{AX}^T\mathbf{B}) = \mathbf{BA}$$

矩阵迹的微分

$$d(tr(X)) = d(\sum_i x_{ii}) = \sum_i dx_{ii} = tr(dX)$$

$$(17)\ \frac{\mathrm{d}\,\mathrm{tr}(\mathbf{X}\mathbf{X}^{\mathrm{T}})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{d}\,\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})}{\mathrm{d}\mathbf{X}} = 2\mathbf{X}$$

$$\begin{aligned}\mathrm{d}\,\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) &= \mathrm{tr}\left(\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})\right) = \mathrm{tr}\left((\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)\\ &= \mathrm{tr}\left((\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{X}\right) + \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})\\ &= \mathrm{tr}\left(2\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)\end{aligned}$$

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

矩阵迹的微分

(18) $\frac{\mathrm{dtr}(\mathbf{AX}^{-1})}{\mathrm{d}\mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{AX}^{-1})^{\mathrm{T}}$

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

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

矩阵迹的微分

(19) $$\frac{\mathrm{d}\,\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X})}{\mathrm{d}\mathbf{X}} = (\mathbf{A}+\mathbf{A}^{\mathrm{T}})\mathbf{X}$$

$$\frac{\mathrm{d}\,\mathrm{tr}(\mathbf{A}\mathbf{X}^{\mathrm{T}})}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{d}\,\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})}{\mathrm{d}\mathbf{X}} = \mathbf{A},$$

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

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

矩阵行列式的微分

(20) $\dfrac{\mathrm{d}|\mathbf{X}|}{\mathrm{d}X} = |\mathbf{X}|\mathbf{X}^{-\mathrm{T}}$

$$A^* = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

令 c_{ij} 是 x_{ij} 的代数余子式， $|\boldsymbol{X}| = \sum_{i=1}^{n} c_{ij} x_{ij}$

有 $\dfrac{\partial|\boldsymbol{X}|}{\partial x_{ij}} = c_{ij}$,

$$\mathrm{d}|\boldsymbol{X}| = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\mathrm{d}x_{ij} = \mathrm{tr}(\boldsymbol{X}^{\#}\mathrm{d}\boldsymbol{X})$$

式中，$\boldsymbol{X}^{\#}$ 是矩阵 $\boldsymbol{X}$ 的伴随矩阵。由于 $\boldsymbol{X}^{-1} = \dfrac{\boldsymbol{X}^{\#}}{|\boldsymbol{X}|}$，故立即有

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

$$\frac{\partial|\boldsymbol{X}|}{\partial\boldsymbol{X}} = |\boldsymbol{X}|(\boldsymbol{X}^{-1})^{\mathrm{T}} = |\boldsymbol{X}|\boldsymbol{X}^{-\mathrm{T}}$$

矩阵行列式的微分

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

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

更一般地，$|\boldsymbol{X}^k|$ 的矩阵微分为

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

由此得梯度矩阵

$$\frac{\partial|\boldsymbol{X}^k|}{\partial \boldsymbol{X}} = k|\boldsymbol{X}^k|\boldsymbol{X}^{-\mathrm{T}}$$

矩阵行列式的微分

(21) $$\frac{\mathrm{d}|\mathbf{X}^{-1}|}{\mathrm{d}\mathbf{X}} = -|\mathbf{X}|^{-1}\mathbf{X}^{-\mathrm{T}}$$

(22) $$\frac{\mathrm{d}\log|\mathbf{X}|}{\mathrm{d}\mathbf{X}} = \frac{1}{|\mathbf{X}|}\frac{\mathrm{d}|\mathbf{X}|}{\mathrm{d}\mathbf{X}}$$

(23) $$\frac{\mathrm{d}|\mathbf{X}^{\mathrm{T}}\mathbf{X}|}{\mathrm{d}\mathbf{X}} = 2|\mathbf{X}^{\mathrm{T}}\mathbf{X}|\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \quad (\mathrm{rank}(\mathbf{X}) = n)$$

$$\frac{\mathrm{d}|\mathbf{X}\mathbf{X}^{\mathrm{T}}|}{\mathrm{d}\mathbf{X}} = 2|\mathbf{X}\mathbf{X}^{\mathrm{T}}|(\mathbf{X}\mathbf{X}^{\mathrm{T}})^{-1}\mathbf{X} \quad (\mathrm{rank}(\mathbf{X}) = m)$$

矩阵行列式的微分

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

即得梯度矩阵 $\dfrac{\partial|\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}|}{\partial\boldsymbol{X}} = 2|\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}|(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}$

类似地， $\mathrm{d}|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}| = \operatorname{tr}\left(2|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}|(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)$

矩阵行列式的微分

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

$$
\frac{\partial|\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}|}{\partial\boldsymbol{X}} = |\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}|\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{B}^{\mathrm{T}}
$$

矩阵行列式的微分

(24) $$\frac{\mathrm{d}|\mathbf{XAX}^{\mathrm{T}}|}{\mathrm{d}\mathbf{X}} = |\mathbf{XAX}^{\mathrm{T}}| \times \left[\left(\mathbf{XAX}^{\mathrm{T}}\right)^{-\mathrm{T}}\mathbf{XA}^{\mathrm{T}} + \left(\mathbf{XAX}^{\mathrm{T}}\right)^{-1}\mathbf{XA}\right]$$

$$\begin{aligned}
&\mathrm{d}|\boldsymbol{XAX}^{\mathrm{T}}| \\
&= |\boldsymbol{XAX}^{\mathrm{T}}|\mathrm{tr}\left((\boldsymbol{XAX}^{\mathrm{T}})^{-1}\mathrm{d}(\boldsymbol{XAX}^{\mathrm{T}})\right) \\
&= |\boldsymbol{XAX}^{\mathrm{T}}|\mathrm{tr}\left((\boldsymbol{XAX}^{\mathrm{T}})^{-1}\left[(\mathrm{d}\boldsymbol{X})\boldsymbol{AX}^{\mathrm{T}} + \boldsymbol{XA}(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\right]\right) \\
&= |\boldsymbol{XAX}^{\mathrm{T}}|\left[\mathrm{tr}\left((\boldsymbol{XAX}^{\mathrm{T}})^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{AX}^{\mathrm{T}}\right) + \mathrm{tr}\left((\boldsymbol{XAX}^{\mathrm{T}})^{-1}\boldsymbol{XA}(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\right)\right] \\
&= |\boldsymbol{XAX}^{\mathrm{T}}|\left[\mathrm{tr}\left((\boldsymbol{XAX}^{\mathrm{T}})^{-1}(\mathrm{d}\boldsymbol{X})\boldsymbol{AX}^{\mathrm{T}}\right) + \mathrm{tr}\left((\boldsymbol{XA})^{\mathrm{T}}(\boldsymbol{XAX}^{\mathrm{T}})^{-\mathrm{T}}\mathrm{d}\boldsymbol{X}\right)\right] \\
&= |\boldsymbol{XAX}^{\mathrm{T}}|\left[\mathrm{tr}\left(\boldsymbol{AX}^{\mathrm{T}}(\boldsymbol{XAX}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{X}\right) + \mathrm{tr}\left((\boldsymbol{XA})^{\mathrm{T}}(\boldsymbol{XA}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}\mathrm{d}\boldsymbol{X}\right)\right] \\
&= |\boldsymbol{XAX}^{\mathrm{T}}|\,\mathrm{tr}\left([\boldsymbol{AX}^{\mathrm{T}}(\boldsymbol{XAX}^{\mathrm{T}})^{-1} + (\boldsymbol{XA})^{\mathrm{T}}(\boldsymbol{XA}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}})^{-1}]\mathrm{d}\boldsymbol{X}\right)
\end{aligned}$$

矩阵行列式的微分

(25) $$\frac{\mathrm{d}|\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}|}{\mathrm{d}\mathbf{X}} = |\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}| \times \left[\mathbf{A}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}\right)^{-1} + \mathbf{A}^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}\right)^{-\mathrm{T}}\right]$$

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

$$\begin{aligned}
\frac{\partial|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|}{\partial\boldsymbol{X}} &= |\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|[\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X})^{-1}] \\
&= 2|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}|\boldsymbol{A}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})^{-1}, \quad \text{若 } \boldsymbol{A} \text{ 为对称矩阵}
\end{aligned}$$

矩阵对矩阵的微分

Definition 3.9

设 $\boldsymbol{A}=[a_{ij}]_{m\times n}$, $\boldsymbol{X}=[x_{kl}]_{p\times q}$, 且 $\boldsymbol{A}$ 中的各元素 a_{ij} 是矩阵 $\boldsymbol{X}$ 中各元素 x_{kl} 的可微函数, 则矩阵 $\boldsymbol{A}$ 对矩阵 $\boldsymbol{X}$ 的导数定义为

$$\frac{\mathrm{d}\boldsymbol{A}}{\mathrm{d}\boldsymbol{X}}=\begin{bmatrix} \dfrac{\partial \boldsymbol{A}}{\partial x_{11}} & \dfrac{\partial \boldsymbol{A}}{\partial x_{12}} & \cdots & \dfrac{\partial \boldsymbol{A}}{\partial x_{1q}} \\ \dfrac{\partial \boldsymbol{A}}{\partial x_{21}} & \dfrac{\partial \boldsymbol{A}}{\partial x_{22}} & \cdots & \dfrac{\partial \boldsymbol{A}}{\partial x_{2q}} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial \boldsymbol{A}}{\partial x_{p1}} & \dfrac{\partial \boldsymbol{A}}{\partial x_{p2}} & \cdots & \dfrac{\partial \boldsymbol{A}}{\partial x_{pq}} \end{bmatrix},$$

矩阵对矩阵的微分

其中

$$\frac{\partial \boldsymbol{A}}{\partial x_{kl}} = \begin{bmatrix} \frac{\partial a_{11}}{\partial x_{kl}} & \frac{\partial a_{12}}{\partial x_{kl}} & \cdots & \frac{\partial a_{1n}}{\partial x_{kl}} \\ \frac{\partial a_{21}}{\partial x_{kl}} & \frac{\partial a_{22}}{\partial x_{kl}} & \cdots & \frac{\partial a_{2n}}{\partial x_{kl}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial a_{m1}}{\partial x_{kl}} & \frac{\partial a_{m2}}{\partial x_{kl}} & \cdots & \frac{\partial a_{mn}}{\partial x_{kl}} \end{bmatrix}, \quad k = 1, 2, \cdots, p;\ l = 1, 2, \cdots, q.$$

矩阵对矩阵的微分

Example 3.10

设 $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$, 求向量 $\boldsymbol{X}^{\mathrm{T}} = (x_1, x_2, \cdots, x_n)$ 对向量 $\boldsymbol{X}$ 的导数.

解:

$$\frac{\mathrm{d}\boldsymbol{X}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{X}} = \begin{bmatrix} \dfrac{\partial \boldsymbol{X}^{\mathrm{T}}}{\partial x_1} \\ \dfrac{\partial \boldsymbol{X}^{\mathrm{T}}}{\partial x_2} \\ \vdots \\ \dfrac{\partial \boldsymbol{X}^{\mathrm{T}}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \boldsymbol{I}_n. \quad \square$$

矩阵对矩阵的微分

Example 3.10

设 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$, 求向量 $\boldsymbol{X}^{\mathrm{T}}=(x_1,x_2,\cdots,x_n)$ 对向量 $\boldsymbol{X}$ 的导数.

解:

$$\frac{\mathrm{d}\boldsymbol{X}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{X}}=\begin{bmatrix}\dfrac{\partial \boldsymbol{X}^{\mathrm{T}}}{\partial x_1}\\ \dfrac{\partial \boldsymbol{X}^{\mathrm{T}}}{\partial x_2}\\ \vdots\\ \dfrac{\partial \boldsymbol{X}^{\mathrm{T}}}{\partial x_n}\end{bmatrix}=\begin{bmatrix}1&0&0&\cdots&0\\0&1&0&\cdots&0\\ \vdots&\vdots&\vdots&&\vdots\\0&0&0&\cdots&1\end{bmatrix}=\boldsymbol{I}_n. \qquad \square$$

梯度矩阵、Jacobian矩阵与Hessian矩阵

定义：梯度向量和梯度矩阵

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

$$\nabla_{\mathbf{X}} f(\mathbf{X}) \overset{\text{def}}{=} \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{\mathbf{m}\times n}$$

$$\mathrm{vec}(\nabla_{\mathbf{X}} f(\mathbf{X})) = \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{dvec}(\mathbf{X})}$$

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

梯度矩阵、Jacobian矩阵与Hessian矩阵

定义：协梯度向量和Jacobian矩阵（协梯度矩阵）：

$$\mathrm{D}_x f(\mathbf{x}) \stackrel{\text{def}}{=} \frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial f(\mathbf{x})}{\partial x_m}\right]$$

$$D_X f(\mathbf{X}) \stackrel{\text{def}}{=} \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}_{n\times m}$$

命题1： 给定是指标量函数f(x),其中$\mathbf{X} \in \mathbb{R}^{m\times n}$，则

$$\mathrm{rvec}(\mathrm{D}_X f(\mathbf{X})) = \mathrm{D}_{\mathrm{vec}(X)} f(\mathbf{X})$$

或

$$\mathrm{D}_{\mathbf{X}} f(\mathbf{X}) = \mathrm{unrvec}\left(\mathrm{D}_{\mathrm{vec}(\mathbf{X})} f(\mathbf{X})\right)$$

命题2： $\nabla_{\mathbf{X}} f(\mathbf{X}) = \mathrm{D}_{\mathbf{X}}^{\mathrm{T}} \mathbf{f}(\mathbf{X})$

梯度矩阵、Jacobian矩阵与Hessian矩阵

命题3： 给定f(x)，其中$f(\mathbf{X})$。若已求出$\mathrm{D}_{\mathrm{vec}(x)}f(\mathbf{X})$，则

$$\nabla_X f(\mathbf{X}) = \mathrm{unvec}\left(\mathrm{D}^{\mathrm{T}}_{\mathrm{vec}(\mathbf{X})} f(\mathbf{X})\right) \qquad (1)$$

换言之，若

$$\mathrm{D}_{\mathrm{vec}(\mathbf{X})} f(\mathbf{X}) = [d_1, d_2, \cdots, d_{mn}]$$

则

$$[\nabla_{\mathbf{X}} f(\mathbf{X})]_{i,j} = d_{i+(j-1)n} \quad \begin{cases} i = 1, \cdots, m \\ j = 1, \cdots n \end{cases} \qquad (2)$$

梯度矩阵、Jacobian矩阵与Hessian矩阵

定义：Jacobian矩阵或协梯度矩阵：

$$D_x f(x) = \frac{\mathrm{d}\mathbf{f}(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \begin{bmatrix} \frac{\mathrm{d}f_1(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} \\ \vdots \\ \frac{\mathrm{d}f_p(\mathbf{x})}{\mathrm{d}\mathbf{x}^{\mathrm{T}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_p(\mathbf{x})}{\partial x_m} \end{bmatrix}_{p\times m}$$

◆ 实值矩阵函数情形

$$\mathbf{F}(\mathbf{X}) = [f_{kl}]_{k=1,l=1}^{p,q} \in \mathbb{R}^{p\times q} \quad (\text{其中}\mathbf{X} \in \mathbb{R}^{m\times n})$$

$$\mathbf{f}(\mathbf{X}) \overset{\mathrm{def}}{=} \mathrm{vec}(\mathbf{F}(\mathbf{X})) \in \mathbb{R}^{pq\times 1}$$

$$= \left[f_{11}(\mathbf{X}), \cdots, f_{p1}(\mathbf{X}), \cdots, f_{1q}(\mathbf{X}), \cdots, f_{pq}(\mathbf{X})\right]^{\mathrm{T}}$$

梯度矩阵、Jacobian矩阵与Hessian矩阵

矩阵函数F(X)的行向量偏导

$$\mathrm{D}_{\mathrm{vec}(\mathbf{X})}\mathbf{F}(\mathbf{X}) \overset{\mathrm{def}}{=} \frac{\mathrm{d}\mathbf{f}(\mathbf{X})}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{X})} = \frac{\mathrm{dvec}(\mathbf{F}(\mathbf{X}))}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{X})} \in \mathbb{R}^{pq\times mn}$$

$$= \left[\frac{\mathrm{d}f_{11}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}, \cdots, \frac{\mathrm{d}f_{p1}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}, \cdots, \frac{\mathrm{d}f_{1q}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}, \cdots, \frac{\mathrm{d}f_{pq}}{\mathrm{dvec}^{\mathrm{T}}(\mathbf{x})}\right]^{\mathrm{T}}$$

$$= \begin{bmatrix} \frac{\partial f_{11}}{\partial x_{11}} & \cdots & \frac{\partial f_{11}}{\partial x_{m1}} & \cdots & \frac{\partial f_{11}}{\partial x_{1n}} & \cdots & \frac{\partial f_{11}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{p1}}{\partial x_{11}} & \cdots & \frac{\partial f_{p1}}{\partial x_{m1}} & \cdots & \frac{\partial f_{p1}}{\partial x_{1n}} & \cdots & \frac{\partial f_{p1}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{1q}}{\partial x_{11}} & \cdots & \frac{\partial f_{1q}}{\partial x_{m1}} & \cdots & \frac{\partial f_{1q}}{\partial x_{1n}} & \cdots & \frac{\partial f_{1q}}{\partial x_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_{pq}}{\partial x_{11}} & \cdots & \frac{\partial f_{pq}}{\partial x_{m1}} & \cdots & \frac{\partial f_{pq}}{\partial x_{1n}} & \cdots & \frac{\partial f_{pq}}{\partial x_{mn}} \end{bmatrix}_{pq\times mn}$$

梯度矩阵、Jacobian矩阵与Hessian矩阵

定义：标量函数的Hessian矩阵

$$\frac{\mathrm{d}^2 f(\mathbf{x})}{\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{x}^{\mathrm{T}}} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}^{\mathrm{T}}}\left[\frac{\mathrm{d}f(\mathbf{x})}{\mathrm{d}\mathbf{x}}\right] = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_m} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_m} \end{bmatrix}_{m\times m}$$

◆ 记$\nabla_{\mathbf{x}}^2 f(\mathbf{x}) = \mathrm{D}_{\mathbf{x}}(\nabla_{\mathbf{x}} f(\mathbf{x})) = \nabla_{\mathbf{x}^{\mathrm{T}}}(\nabla_{\mathbf{x}} f(\mathbf{x}))$

◆ 实值标量函数f(x)的Hessian矩阵

$$\frac{\partial^2 f(\mathbf{X})}{\partial \mathbf{X} \partial \mathbf{X}^{\mathrm{T}}} = \frac{\partial}{\partial \mathbf{X}^{\mathrm{T}}}\left[\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}}\right]$$

或记作$\nabla_X^2 f(\mathbf{X}) = \mathrm{D}_{\mathbf{X}}(\nabla_{\mathbf{X}} f(\mathbf{X})) = \nabla_{\mathbf{X}^T}(\nabla_{\mathbf{X}} f(\mathbf{X}))$

实值标量函数的矩阵微分及计算

A.实值阵微分

◆ 全微分

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

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

$$= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{T}}}\mathrm{d}\mathbf{x}$$

实值标量函数的矩阵微分及计算

$$\mathrm{d}f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_1}\mathrm{d}\mathbf{x}_1 + \cdots + \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_n}\mathrm{d}\mathbf{x}_n$$

$$= \left[\frac{\partial f(\mathbf{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{m1}}\right]\begin{bmatrix}\mathrm{d}x_{11}\\ \vdots \\ \mathrm{d}x_{m1}\end{bmatrix} + \cdots + \left[\frac{\partial f(\mathbf{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{mn}}\right]\begin{bmatrix}\mathrm{d}x_{1n}\\ \vdots \\ \mathrm{d}x_{mn}\end{bmatrix}$$

$$= \left[\frac{\partial f(\mathbf{X})}{\partial x_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{m1}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial x_{mn}}\right]\begin{bmatrix}\mathrm{d}x_{11}\\ \vdots \\ \mathrm{d}x_{m1}\\ \vdots \\ \mathrm{d}x_{1n}\\ \vdots \\ \mathrm{d}x_{mn}\end{bmatrix}$$

$$= \mathrm{rvec}(\mathbf{A})\mathrm{vec}(\mathrm{d}\mathbf{X})$$

实值标量函数的矩阵微分及计算

其中

$$\mathbf{A} = \mathrm{D}_{\mathbf{X}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix}$$

且

$$\mathrm{d}\mathbf{X} = \begin{bmatrix} \mathrm{d}x_{11} & \cdots & \mathrm{d}x_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{d}x_{m1} & \cdots & \mathrm{d}x_{mn} \end{bmatrix}$$

进一步有

$$\mathrm{d}f(\mathbf{X}) = \left(\mathrm{vec}\left(\mathbf{A}^{\mathrm{T}}\right)\right)^{\mathrm{T}} \mathrm{vec}(\mathrm{d}\mathbf{X})$$

则

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

实值标量函数的矩阵微分及计算

命题4： 一阶偏导矩阵A是唯一确定的。即，若存在A_1和A_2满足

$$\mathrm{d}f(\mathbf{X}) = \mathbf{A}_i \mathrm{d}\mathbf{X}, \quad i = 1,2$$

则$\mathbf{A}_1 = \mathbf{A}_2$。

命题5： 若实值标量函数f(X)在X可微分，则

$$\mathrm{d}f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}\mathrm{d}\mathbf{X}) \quad \Leftrightarrow \quad \nabla_{\mathbf{X}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^{\mathrm{T}}$$

实值标量函数的矩阵微分及计算

B.实矩阵微分计算

1.一般规则

（1）$\mathrm{d}\left(\mathbf{X}^{\mathrm{T}}\right)=(\mathrm{d}\mathbf{X})^{\mathrm{T}}$

（2）$\mathrm{d}(\alpha\mathbf{X}+\beta\mathbf{Y})=\alpha\mathrm{d}\mathbf{X}+\beta\mathrm{d}\mathbf{Y}$

例1.考虑标量函数$\mathrm{tr}(\mathbf{U})$的微分，得

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

即有$\mathrm{d}(\mathrm{tr}\mathbf{U})=\mathrm{tr}(\mathrm{d}\mathbf{U})$。

实值标量函数的矩阵微分及计算

常用计算公式

（1） $\mathrm{d}A = 0$

（2) $\mathrm{d}(\alpha X) = \alpha \mathrm{d}\mathrm{X}$

（3) $\mathrm{d}\left(X^{\mathrm{T}}\right) = (\mathrm{d}X)^{\mathrm{T}})$

（4) $\mathrm{d}(\mathbf{U} \pm \mathbf{V}) = \mathrm{d}\mathbf{U} \pm \mathrm{d}\mathbf{V}$

（5) $\mathrm{d}(\mathbf{AXB}) = \mathbf{A}(\mathrm{d}\mathbf{X})\mathbf{B}$

实值标量函数的矩阵微分及计算

常用计算公式

（6）

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

$$\mathrm{d}(\mathbf{UVW}) = (\mathrm{d}\mathbf{U})\mathbf{VW} + \mathbf{U}(\mathrm{d}\mathbf{V})\mathbf{W} + \mathbf{UV}(\mathrm{d}\mathbf{W})$$

特别地，若A为常数矩阵，则

$$\mathrm{d}\left(\mathbf{XAX}^{\mathrm{T}}\right) = (\mathrm{d}\mathbf{X})\mathbf{AX}^{\mathrm{T}} + \mathbf{XA}(\mathrm{d}\mathbf{X})^{\mathrm{T}}$$

和 $$\mathrm{d}\left(\mathbf{X}^{\mathrm{T}}\mathbf{AX}\right) = (\mathrm{d}\mathbf{X})^{\mathrm{T}}\mathbf{AX} + \mathbf{X}^{\mathrm{T}}\mathbf{A}\mathrm{d}\mathbf{X}$$

(7） $\mathrm{d}(\mathbf{U} \otimes \mathbf{V}) = (\mathrm{d}\mathbf{U}) \otimes \mathbf{V} + \mathbf{U} \otimes \mathrm{d}\mathbf{V}$

(8) $\mathrm{d}(\mathbf{U} \odot \mathbf{V}) = (\mathrm{d}\mathbf{U}) \odot \mathbf{V} + \mathbf{U} \odot \mathrm{d}\mathbf{V}$

(9) $\mathrm{d}(\mathrm{vec}(\mathbf{X})) = \mathrm{vec}(\mathrm{d}\mathbf{X})$

实值标量函数的矩阵微分及计算

(10) $\mathrm{d}\log\mathbf{X} = \mathbf{X}^{-1}\mathrm{d}\mathbf{X}, \quad \mathrm{d}\log(\mathbf{F}(\mathbf{X})) = (\mathbf{F}(\mathbf{X}))^{-1}\mathrm{d}(\mathbf{F}(\mathbf{X}))$

(11) $\mathrm{d}|\mathbf{X}| = |\mathbf{X}|\mathrm{tr}(\mathbf{X}^{-1}\mathrm{d}\mathbf{X}), \quad \mathrm{d}|\mathbf{F}(\mathbf{X})| = |\mathbf{U}|\mathrm{tr}(\mathbf{U}^{-1}\mathrm{d}\mathbf{X})$

(12） $\mathrm{d}(\mathrm{tr}(\mathbf{X})) = \mathrm{tr}(\mathrm{d}\mathbf{X}), \quad \mathrm{d}(\mathrm{tr}(\mathbf{F}(\mathbf{X}))) = \mathrm{tr}(\mathrm{d}(\mathbf{F}(\mathbf{X})))$

(13） $\mathrm{d}(\mathbf{X}^{-1}) = -\mathbf{X}^{-1}(\mathrm{d}\mathbf{X})\mathbf{X}^{-1}$

(14） $$\mathrm{d}(\mathbf{X}^{\dagger}) = -\mathbf{X}^{\dagger}(\mathrm{d}\mathbf{X})\mathbf{X}^{\dagger} + \mathbf{X}^{\dagger}(\mathbf{X}^{\dagger})^{\mathrm{T}}(\mathrm{d}\mathbf{X}^{\mathrm{T}})(\mathbf{I} - \mathbf{X}\mathbf{X}^{\dagger}) + (\mathbf{I} - \mathbf{X}^{\dagger}\mathbf{X})(\mathrm{d}\mathbf{X}^{\mathrm{T}})(\mathbf{X}^{\dagger})^{\mathrm{T}}\mathbf{X}^{\dagger}$$

$$\mathrm{d}(\mathbf{X}^{\dagger}\mathbf{X}) = \mathbf{X}^{\dagger}(\mathrm{d}\mathbf{X})(\mathbf{I} - \mathbf{X}^{\dagger}\mathbf{X}) + \left(\mathbf{X}^{\dagger}(\mathrm{d}\mathbf{X})(\mathbf{I} - \mathbf{X}^{\dagger}\mathbf{X})\right)^{\mathrm{T}}$$

$$\mathrm{d}(\mathbf{X}\mathbf{X}^{\dagger}) = (\mathbf{I} - \mathbf{X}\mathbf{X}^{\dagger})(\mathrm{d}\mathbf{X})\mathbf{X}^{\dagger} + \left((\mathbf{I} - \mathbf{X}\mathbf{X}^{\dagger})(\mathrm{d}\mathbf{X})\mathbf{X}^{\dagger}\right)^{\mathrm{T}}$$

实值标量函数的矩阵微分及计算

C.利用矩阵微分计算梯度矩阵

$$\mathrm{d}f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}\mathrm{d}\mathbf{X}) \quad \Leftrightarrow \quad \nabla_{\mathbf{X}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^{\mathrm{T}}$$

1.一般标量函数的梯度矩阵

2.迹函数的梯度矩阵

3.行列式的梯度矩阵

实值标量函数的矩阵微分及计算

D.二阶实微分矩阵与实Hessian矩阵

令x, $\mathbf{x}$,$\mathbf{X}$分别代表函数的实标量变元、m×1实向量变元和$m \times n$实矩阵变元， 而$f(\cdot)$，$\mathbf{f}(\cdot)$，$\mathbf{F}(\cdot)$分别表示实标量函数、$p\times 1$实向量函数和$p\times q$实矩阵函数。

表3 二阶辨识表

实函数	二阶实微分矩阵	实Hessian矩阵H	H的维数
$f(x)$ $f(\mathbf{x})$ $f(\mathbf{X})$	$\mathrm{d}^2[f(x)] = \beta(\mathrm{d}x)^2$ $\mathrm{d}^2[f(x)] = (\mathrm{d}x)^{\mathrm{T}} B \mathrm{d}x$ $\mathrm{d}^2[f(X)] = (\mathrm{dvec}(X))^{\mathrm{T}} B \mathrm{d}(\mathrm{vec}(X))$	$\mathbf{H}\vert f(x)] = \beta$ $\mathbf{H}[f(\mathbf{x})] = \frac{1}{2}(\mathbf{B}+\mathbf{B}^{\mathrm{T}})$ $\mathbf{H}[f(\mathbf{X})] = \frac{1}{2}(\mathbf{B}+\mathbf{B}^{\mathrm{T}})$	1×1 $m\times m$ $mn\times mn$
$f(x)$ $f(\mathbf{x})$ $f(\mathbf{X})$	$\mathrm{d}^2[f(x)] = b(\mathrm{d}x)^2$ $\mathrm{d}^2[f(x)] = (I_m \otimes \mathrm{d}x)^{\mathrm{T}} B \mathrm{d}x$ $\mathrm{d}^2[f(X)] = (I_m \otimes \mathrm{dvec}(X))^{\mathrm{T}} B \mathrm{d}(\mathrm{vec}(X))$	$\mathbf{H}[\mathbf{f}(x)] = \mathbf{b}$ $\mathbf{H}[\mathbf{f}(\mathbf{x})] = \frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$ $\mathbf{H}[\mathbf{f}(\mathbf{X})] = \frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$	$p\times 1$ $pm\times m$ $pmn\times mn$
$\mathbf{F}(x)$ $\mathbf{F}(\mathbf{x})$ $\mathbf{F}(\mathbf{X})$	$\mathrm{d}^2[\mathbf{F}(x)] = \mathbf{B}(\mathrm{d}x)^2$ $\mathrm{d}^2[\mathrm{vec}(\mathbf{F})] = \left(\mathbf{I}_{mp} \otimes \mathrm{d}\mathbf{x}\right)^{\mathrm{T}} \mathbf{B}\mathrm{d}\mathbf{x}$ $\mathrm{d}^2[\mathrm{vec}(\mathbf{F})]$ $= \left(\mathbf{I}_{mp} \otimes \mathrm{dvec}(\mathbf{X})\right)^{\mathrm{T}} \mathbf{B}\mathrm{d}(\mathrm{vec}(\mathbf{X}))$	$\mathbf{H}[\mathbf{F}(x)] = \mathrm{vec}(\mathbf{B})$ $\mathbf{H}[\mathbf{F}(\mathbf{x})] = \frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$ $\mathbf{H}\left[\mathbf{F}'(\mathbf{x})\right] = \frac{1}{2}\left[\mathbf{B}+(\mathbf{B}')_v\right]$	$pq\times 1$ $pmq\times m$ $pmqn\times mn$

实值标量函数的矩阵微分及计算

表中，对于实向量函数$\boldsymbol{f}$，

$$\mathbf{B}=\begin{bmatrix}\mathbf{B}_1\\ \mathbf{B}_2\\ \vdots\\ \mathbf{B}_p\end{bmatrix},\quad (\mathbf{B}')_v=\begin{bmatrix}\mathbf{B}_1^{\mathrm{T}}\\ \mathbf{B}_2^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_p^{\mathrm{T}}\end{bmatrix}$$

而对于实矩阵函数$\mathbf{F}$，

$$\mathbf{B}=\begin{bmatrix}\mathbf{B}_{11}\\ \vdots\\ \mathbf{B}_{p1}\\ \vdots\\ \mathbf{B}_{1q}\\ \vdots\\ \mathbf{B}_{pq}\end{bmatrix}\quad (\mathbf{B}')_v=\begin{bmatrix}\mathbf{B}_{11}^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_{p1}^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_{1q}^{\mathrm{T}}\\ \vdots\\ \mathbf{B}_{pq}^{\mathrm{T}}\end{bmatrix}$$

梯度数值计算

```
clear all
x=rand(5,1)
y=diff(x)

x =
    0.1576
    0.9706
    0.9572
    0.4854
    0.8003
y =
    0.8130
   -0.0134
   -0.4718
    0.3149
```

```
%DIFF Difference and approximate derivative.
%   DIFF(X), for a vector X, is [X(2)-X(1)  X(3)-X(2) ... X(n)-X(n-1)].
%   DIFF(X), for a matrix X, is the matrix of row differences,
%      [X(2:n,:) - X(1:n-1,:)].
%   DIFF(X), for an N-D array X, is the difference along the first
%      non-singleton dimension of X.
%   DIFF(X,N) is the N-th order difference along the first non-singleton     dimension (denote it by DIM). If N >= size(X,DIM), DIFF takes     successive differences along the next non-singleton dimension.
%   DIFF(X,N,DIM) is the Nth difference function along dimension DIM.
%      If N >= size(X,DIM), DIFF returns an empty array.
```

梯度数值计算

```
>> X = [3 7 5
        0 9 2]
X =
     3     7     5
     0     9     2

>> diff(X,1,1)
ans =
    -3     2    -3

>> diff(X,2,1)
ans =
   空矩阵: 0×3
```

```
>> diff(X,1,2)
ans =
     4    -2
     9    -7

>> diff(X,2,2)
ans =
    -6
   -16
```

梯度数值计算

$y = f(x)$, but we only know the values of f at a finite set of points,

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

Forward Difference

$$f'(x_i) = y_i' \approx \frac{y_{i+1} - y_i}{x_{i+1} - x_i}.$$

Backward Difference

$$f'(x_i) \approx \frac{y_i - y_{i-1}}{x_i - x_{i-1}}.$$

If the points are evenly spaced, i.e. $x_{i+1} - x_i = x_i - x_{i-1} = h$,

Central Difference

$$f'(x_i) = y_i' \approx \frac{y_{i+1} - y_{i-1}}{2h}.$$

梯度数值计算

Suppose $u = u(x, y)$, $\quad u_{i,j} = u(x_i, y_j)$

The central difference formulas

$$u_x(x_i, y_j) \approx \frac{1}{2h}\left(u_{i+1,j} - u_{i-1,j}\right) \quad \text{and}$$

$$u_y(x_i, y_j) \approx \frac{1}{2k}\left(u_{i,j+1} - u_{i,j-1}\right).$$

The second partial derivatives are

$$u_{xx}(x_i, y_j) \approx \frac{1}{h^2}\left(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\right) \quad \text{and}$$

$$u_{yy}(x_i, y_j) \approx \frac{1}{k^2}\left(u_{i,j+1} - 2u_{i,j} + u_{i,j-1}\right),$$

and the mixed partial derivative is

$$u_{xy}(x_i, y_j) \approx \frac{1}{4hk}\left(u_{i+1,j+1} - u_{i+1,j-1} - u_{i-1,j+1} + u_{i-1,j-1}\right).$$

边界处理

1、补零

2、对称性边界(Symmetric Boundary Conditions)

3、周期性边界(Periodic Boundary Conditions)

卷积

C = conv(A, B, SHAPE) returns a subsection of the convolution with size

specified by SHAPE:

'full' - (default) returns the full convolution,

'same' - returns the central part of the convolution
that is the same size as A.

'valid' - returns only those parts of the convolution
that are computed without the zero-padded edges.
LENGTH(C)is MAX(LENGTH(A)-MAX(0,LENGTH(B)-1),0).

上海大学 SHANGHAI UNIVERSITY

卷积

```
>> A=rand(5,1)
A =
    0.1419
    0.4218
    0.9157
    0.7922
    0.9595
>> B=rand(3,1)
B =
    0.6557
    0.0357
    0.8491
>> conv(A,B)
ans =
    0.0930
    0.2816
    0.7360
    0.9103
    1.4350
    0.7070
    0.8147
```

```
>> conv(A,B,'full')
ans =
    0.0930
    0.2816
    0.7360
    0.9103
    1.4350
    0.7070
    0.8147

>> conv(A,B,'same')
ans =
    0.2816
    0.7360
    0.9103
    1.4350
    0.7070
```

```
>> conv(A,B,'valid')

ans =

    0.7360
    0.9103
    1.4350
```

谢　谢！