



研究生课程教学课件

矩阵分析与应用

讲授：张贤达

清华大学自动化系

Email: zxd-dau@tsinghua.edu.cn

办公室: FIT大楼3-117

电话: 010-62794875



[访问清华主页](#)

[标题页](#)



第 1 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)

第3章 矩阵微分

- 3.1 Jacobian矩阵、梯度矩阵与Hessian矩阵
- 3.2 实值函数的矩阵微分
- 3.3 实矩阵微分计算
- 3.4 梯度与共轭梯度
- 3.5 复矩阵微分
- 3.6 复矩阵微分计算



访问清华主页

标题页



第 2 页 共 72 页

返回

全屏显示

关闭

退出

3.1 Jacobian矩阵、梯度矩阵与Hessian矩阵

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 $f \in \mathbb{R}$	$f(\boldsymbol{x})$ $f: \mathbb{R}^m \rightarrow \mathbb{R}$	$f(\boldsymbol{X})$ $f: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$
向量函数 $\boldsymbol{f} \in \mathbb{R}^p$	$\boldsymbol{f}(\boldsymbol{x})$ $\boldsymbol{f}: \mathbb{R}^m \rightarrow \mathbb{R}^p$	$\boldsymbol{f}(\boldsymbol{X})$ $\boldsymbol{f}: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$
矩阵函数 $\boldsymbol{F} \in \mathbb{R}^{p \times q}$	$\boldsymbol{F}(\boldsymbol{x})$ $\boldsymbol{F}: \mathbb{R}^m \rightarrow \mathbb{R}^{p \times q}$	$\boldsymbol{F}(\boldsymbol{X})$ $\boldsymbol{F}: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{p \times q}$

实值向量偏导 $\left\{ \begin{array}{l} \text{实值行向量偏导} \\ \text{实值列向量偏导} \end{array} \right.$



访问清华主页

标题页

◀ ▶

◀ ▶

第 3 页 共 72 页

返回

全屏显示

关闭

退出

1. 实值标量函数的Jacobian矩阵与梯度矩阵

$1 \times m$ 行向量偏导算子

$$D_x \stackrel{\text{def}}{=} \frac{\partial}{\partial \mathbf{x}^T} \triangleq \left[\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_m} \right]$$

实值标量函数 $f(\mathbf{x})$ 的行向量偏导

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

若变元为实值矩阵 $\mathbf{X} \in \mathbb{R}^{m \times n}$ ，则需要预先将矩阵变元进行列向量化，然后转置，变成行向量之后，再定义行向量偏导：

$$\begin{aligned} D_{\text{vec}^T(\mathbf{X})} f(\mathbf{X}) &= \frac{\partial f(\mathbf{X})}{\partial \text{vec}^T(\mathbf{X})} \\ &= \left[\frac{\partial f(\mathbf{X})}{\partial X_{11}}, \dots, \frac{\partial f(\mathbf{X})}{\partial X_{m1}}, \dots, \frac{\partial f(\mathbf{X})}{\partial X_{1n}}, \dots, \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \right] \end{aligned}$$

式中 $\text{vec}^T(\mathbf{X}) = [\text{vec}(\mathbf{X})]^T$ 。



访问清华主页

标题页

◀ ▶

◀ ▶

第 4 页 共 72 页

返回

全屏显示

关闭

退出



如直接以矩阵形式定义实值标量函数 $f(\mathbf{X})$ 在 \mathbf{X} 处的偏导

$$\mathbf{D}_{\mathbf{X}}f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}^T} = \left[\frac{\partial f(\mathbf{X})}{\partial x_{ji}} \right]_{j=1, i=1}^{n, m}$$

则有

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

$\mathbf{D}_{\text{vec}^T(\mathbf{X})}f(\mathbf{X})$ 和 $\mathbf{D}_{\mathbf{X}}f(\mathbf{X})$ 分别称为标量函数 $f(\mathbf{X})$ 在 \mathbf{X} 的行向量偏导和Jacobian矩阵。

[访问清华主页](#)

[标题页](#)

[«](#) [»](#)

[◀](#) [▶](#)

第 5 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)



比较 $D_{\text{vec}^T(\mathbf{X})}f(\mathbf{X})$ 和 $D_{\mathbf{X}}f(\mathbf{X})$ 知，行向量偏导与Jacobian矩阵之间存在下述关系。

命题1 给定一实值标量函数 $f(\mathbf{X})$ ，其中 $\mathbf{X} \in \mathbb{R}^{m \times n}$ ，则

$$\text{rvec}(D_{\mathbf{X}}f(\mathbf{X})) = D_{\text{vec}^T(\mathbf{X})}f(\mathbf{X})$$

或

$$D_{\mathbf{X}}f(\mathbf{X}) = \text{unrvec}(D_{\text{vec}^T(\mathbf{X})}f(\mathbf{X}))$$

式中 $\text{unrvec}(\cdot)$ 是行向量的矩阵化函数。

命题1的意义：Jacobian矩阵往往是感兴趣的待求矩阵，而行向量偏导则容易运算，为中间运算工具。

[访问清华主页](#)

[标题页](#)

[«](#) [»](#)

[◀](#) [▶](#)

第 6 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)

$m \times 1$ 列向量偏导算子定义为

$$\nabla_x \stackrel{\text{def}}{=} \frac{\partial}{\partial \mathbf{x}} = \left[\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_m} \right]^T$$

习惯称之为梯度算子。因此，实值标量函数 $f(\mathbf{x})$ 的梯度向量为 $m \times 1$ 列向量，定义为

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

将矩阵变元 \mathbf{X} 列向量化后，即可直接定义关于矩阵变元 \mathbf{X} 的梯度算子为

$$\begin{aligned} \nabla_{\text{vec}(\mathbf{X})} &= \frac{\partial}{\partial \text{vec}(\mathbf{X})} \\ &= \left[\frac{\partial}{\partial X_{11}}, \cdots, \frac{\partial}{\partial X_{m1}}, \cdots, \frac{\partial}{\partial X_{1n}}, \cdots, \frac{\partial}{\partial X_{mn}} \right]^T \end{aligned}$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 7 页 共 72 页

返回

全屏显示

关闭

退出



于是，实值标量函数 $f(\mathbf{X})$ 关于矩阵变元 \mathbf{X} 的梯度矩阵的列向量形式：

$$\begin{aligned}\text{vec}(\nabla_{\mathbf{X}} f(\mathbf{X})) &= \frac{\partial f(\mathbf{X})}{\partial \text{vec}(\mathbf{X})} \\ &= \left[\frac{\partial f(\mathbf{X})}{\partial X_{11}}, \dots, \frac{\partial f(\mathbf{X})}{\partial X_{m1}}, \dots, \frac{\partial f(\mathbf{X})}{\partial X_{1n}}, \dots, \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \right]^T\end{aligned}$$

将这一列向量结果矩阵化，易知

$$\nabla_{\mathbf{X}} f(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{11}} & \dots & \frac{\partial f(\mathbf{X})}{\partial X_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial X_{m1}} & \dots & \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \end{bmatrix} = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} \quad (2)$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 8 页 共 72 页

返回

全屏显示

关闭

退出

比较式(2)和式(1)，立即有

$$\nabla_X f(\mathbf{X}) = D_X^T f(\mathbf{X})$$

即是说，实值标量函数 $f(\mathbf{X})$ 的梯度矩阵等于Jacobian 矩阵的转置。

Jacobian矩阵：在流形计算、几何物理、微分几何以及矩阵微分等中，当定义一个标量函数关于变元向量的偏导数时，行向量偏导向量和Jacobian 矩阵是“最自然的”选择。

梯度矩阵：在最优化和许多工程问题中，采用列向量形式定义的偏导(梯度向量和梯度矩阵)是一种比行向量偏导和Jacobian矩阵更自然的选择。



访问清华主页

标题页

◀ ▶

◀ ▶

第 9 页 共 72 页

返回

全屏显示

关闭

退出



由于两者之间的转换关系，行向量形式的偏导向量是列向量形式的梯度向量的协变形式(covariant form of the gradient vector)，故又简称为协梯度向量(cogradient vector)。类似地，Jacobian矩阵有时也被称为梯度矩阵的协变形式或简称协(同)梯度矩阵。协梯度是一协变算子(covariant operator)，它本身虽然不是梯度，但却是梯度的紧密伙伴(转置后即变身为梯度)。

有鉴于此，Jacobian算子 $\frac{\partial}{\partial \mathbf{x}^T}$ 和 $\frac{\partial}{\partial \mathbf{X}^T}$ 又称(行)偏导算子、梯度算子的协变形式或协梯度算子(cogradient operator)。

访问清华主页

标题页

◀ ▶

◀ ▶

第 10 页 共 72 页

返回

全屏显示

关闭

退出



由 $\nabla_{\mathbf{X}} f(\mathbf{X}) = \mathbf{D}_{\mathbf{X}}^T f(\mathbf{X})$ 及命题1立即有以下结果。

命题2 给定一实值标量函数 $f(\mathbf{X})$ ，其中 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 。若已求出行向量偏导 $\mathbf{D}_{\text{vec}^T(\mathbf{X})} f(\mathbf{X})$ ，则

$$\nabla_{\mathbf{X}} f(\mathbf{X}) = \underline{\text{unvec}} \left(\mathbf{D}_{\text{vec}^T(\mathbf{X})}^T f(\mathbf{X}) \right) \quad (3)$$

即是说，实值标量函数 $f(\mathbf{X})$ 的梯度矩阵由行向量偏导的转置(列向量形式)的矩阵化结果决定。

若 $\mathbf{D}_{\text{vec}^T(\mathbf{X})} f(\mathbf{X}) = [d_1, \dots, d_{mn}]$ ，则梯度矩阵第 (i, j) 个元素

$$[\nabla_{\mathbf{X}} f(\mathbf{X})]_{i,j} = d_{i+(j-1)n} \quad \begin{cases} i = 1, \dots, m \\ j = 1, \dots, n \end{cases} \quad (4)$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 11 页 共 72 页

返回

全屏显示

关闭

退出

梯度方向的负方向称为变元 x 的梯度流(gradient flow), 记作

$$\dot{x} = -\nabla_x f(x)$$

从梯度向量的定义式可以看出:

(1)一个以向量为变元的实值标量函数的梯度为一列向量。

(2)梯度向量的每个分量给出了标量函数在该分量方向上的变化率。

重要性质: 梯度向量指出了当变元增大时实值标量函数 $f(x)$ 的最大增大率。相反, 梯度的负值(简称负梯度)则指出了当变元增大时函数 $f(x)$ 的最大减小率。这是梯度下降法的基础。



访问清华主页

标题页

« »

◀ ▶

第 12 页 共 72 页

返回

全屏显示

关闭

退出

2. 实值向量函数的协梯度矩阵

对 $p \times 1$ 实值向量函数 $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \dots, f_p(\mathbf{x})]^\mathrm{T}$ 的元素 $f_i(\mathbf{x}), i = 1, \dots, p$ 使用实值标量函数的行向量偏导公式，可以直接定义实值向量函数的偏导如下：

$$\mathbf{D}_x \mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}^\mathrm{T}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}^\mathrm{T}} \\ \vdots \\ \frac{\partial f_p(\mathbf{x})}{\partial \mathbf{x}^\mathrm{T}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_1(\mathbf{x})}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p(\mathbf{x})}{\partial x_1} & \dots & \frac{\partial f_p(\mathbf{x})}{\partial x_m} \end{bmatrix} \in \mathbb{R}^{p \times m}$$

并称之为向量函数 $\mathbf{f}(\mathbf{x})$ 在 \mathbf{x} 处的 Jacobian 矩阵或协梯度矩阵，其第 (i, j) 个元素定义为向量函数 $\mathbf{f}(\mathbf{x})$ 的第 i 个分量 $f_i(\mathbf{x})$ 相对于向量变元 \mathbf{x} 的第 j 个元素的偏导，即

$$[\mathbf{D}_x \mathbf{f}(\mathbf{x})]_{ij} = \frac{\partial f_i(\mathbf{x})}{\partial x_j}$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 13 页 共 72 页

返回

全屏显示

关闭

退出

3. 实值矩阵函数的协梯度矩阵

现在考虑实值矩阵函数 $F(\mathbf{X}) = [F_{kl}]_{k=1,l=1}^{p,q} \in \mathbb{R}^{p \times q}$ 的情况，其中，矩阵变元 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 。

为了利用向量函数的行向量偏导和Jacobian矩阵的定义，需要预先通过列向量化，将 $p \times q$ 矩阵函数转换成 $pq \times 1$ 列向量：

$$\begin{aligned} \mathbf{f}(\text{vec} \mathbf{X}) &\triangleq \text{vec}(\mathbf{F}(\mathbf{X})) \in \mathbb{R}^{pq} \\ &= [F_{11}(\mathbf{X}), \dots, F_{p1}(\mathbf{X}), \dots, F_{1q}(\mathbf{X}), \dots, F_{pq}(\mathbf{X})]^T \end{aligned}$$

于是，矩阵函数 $F(\mathbf{X})$ 的行向量偏导定义为

$$D_{\text{vec}^T(\mathbf{X})} \mathbf{F}(\mathbf{X}) \triangleq \frac{\partial \mathbf{f}(\text{vec} \mathbf{X})}{\partial \text{vec}^T(\mathbf{X})} = \frac{\partial \text{vec}(\mathbf{F}(\mathbf{X}))}{\partial \text{vec}^T(\mathbf{X})} \in \mathbb{R}^{pq \times mn}$$

其具体表达式为



访问清华主页

标题页

◀

▶

◀

▶

第 14 页 共 72 页

返回

全屏显示

关闭

退出



$$\mathbf{D}_{\text{vec}^T(\mathbf{X})} \mathbf{F}(\mathbf{X})$$

$$= \left[\frac{\partial F_{11}}{\partial \text{vec}^T(\mathbf{X})}, \dots, \frac{\partial F_{p1}}{\partial \text{vec}^T(\mathbf{X})}, \dots, \frac{\partial F_{1q}}{\partial \text{vec}^T(\mathbf{X})}, \dots, \frac{\partial F_{pq}}{\partial \text{vec}^T(\mathbf{X})} \right]^T$$

$$= \begin{bmatrix} \frac{\partial F_{11}}{\partial X_{11}} & \dots & \frac{\partial F_{11}}{\partial X_{m1}} & \dots & \frac{\partial F_{11}}{\partial X_{1n}} & \dots & \frac{\partial F_{11}}{\partial X_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{p1}}{\partial X_{11}} & \dots & \frac{\partial F_{p1}}{\partial X_{m1}} & \dots & \frac{\partial F_{p1}}{\partial X_{1n}} & \dots & \frac{\partial F_{p1}}{\partial X_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{1q}}{\partial X_{11}} & \dots & \frac{\partial F_{1q}}{\partial X_{m1}} & \dots & \frac{\partial F_{1q}}{\partial X_{1n}} & \dots & \frac{\partial F_{1q}}{\partial X_{mn}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{pq}}{\partial X_{11}} & \dots & \frac{\partial F_{pq}}{\partial X_{m1}} & \dots & \frac{\partial F_{pq}}{\partial X_{1n}} & \dots & \frac{\partial F_{pq}}{\partial X_{mn}} \end{bmatrix}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 15 页 共 72 页

返回

全屏显示

关闭

退出



4. 实值标量函数的Hessian矩阵

实值函数 $f(\boldsymbol{x})$ 相对于 $m \times 1$ 实向量 \boldsymbol{x} 的二阶偏导是一个由 m^2 个二阶偏导组成的矩阵(称为Hessian矩阵), 定义为

$$\frac{\partial^2 f(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^T} = \frac{\partial}{\partial \boldsymbol{x}^T} \left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right]$$

或记作

$$\nabla_x^2 f(\boldsymbol{x}) = D_x(\nabla_x f(\boldsymbol{x})) = D_x \boldsymbol{g}(\boldsymbol{x})$$

即实值标量函数 $f(\boldsymbol{x})$ 的Hessian矩阵是梯度向量函数 $\boldsymbol{g}(\boldsymbol{x}) = \nabla_x f(\boldsymbol{x})$ 的协梯度矩阵(Jacobian 矩阵)。

访问清华主页

标题页

◀ ▶

◀ ▶

第 16 页 共 72 页

返回

全屏显示

关闭

退出



实值函数 $f(\mathbf{X})$ 相对于 $m \times n$ 实矩阵 \mathbf{X} 的二阶偏导是一个由 mn 个二阶偏导组成的矩阵(称为Hessian矩阵), 定义为

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

或记作

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

即实值标量函数 $f(\mathbf{X})$ 的Hessian矩阵是梯度矩阵函数 $\mathbf{G}(\mathbf{X}) = \nabla_{\mathbf{X}} f(\mathbf{X})$ 的协梯度矩阵(Jacobian 矩阵)。

访问清华主页

标题页

◀ ▶

◀ ▶

第 17 页 共 72 页

返回

全屏显示

关闭

退出



求Hessian矩阵的二步法:

1. 求实值函数 $f(x)$ [或 $f(X)$]关于向量变元 x [或矩阵变元 X]的偏导数, 得到实值函数的梯度向量 $\frac{\partial f(x)}{\partial x}$ 或梯度矩阵 $\frac{\partial f(X)}{\partial X}$ 。
2. 再求梯度向量 $\frac{\partial f(x)}{\partial x}$ 相对于 $1 \times n$ 行向量 x^T 的偏导数(Jacobian矩阵, 协梯度矩阵), 或者求梯度矩阵 $\frac{\partial f(X)}{\partial X}$ 相对于 $n \times m$ 转置矩阵 X^T 的协梯度矩阵, 得到Hessian矩阵。

访问清华主页

标题页

◀ ▶

◀ ▶

第 18 页 共 72 页

返回

全屏显示

关闭

退出

3.2 实值标量函数的矩阵微分

在多元函数的微积分中，称多元函数 $f(x_1, \dots, x_m)$ 在点 (x_1, \dots, x_m) 可微分，若 $f(x_1, \dots, x_m)$ 的全改变量可以写为

$$\begin{aligned}\Delta f(x_1, \dots, x_m) &= f(x_1 + \Delta x_1, \dots, x_m + \Delta x_m) - f(x_1, \dots, x_m) \\ &= A_1 \Delta x_1 + \dots + A_m \Delta x_m + O(\Delta x_1, \dots, \Delta x_m)\end{aligned}$$

式中， A_1, \dots, A_m 分别与 $\Delta x_1, \dots, \Delta x_m$ 无关，而 $O(\Delta x_1, \dots, \Delta x_m)$ 表示偏改变量 $\Delta x_1, \dots, \Delta x_m$ 的二阶及高阶项。这时，函数 $f(x_1, \dots, x_m)$ 的偏导数 $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_m}$ 一定存在，并且

$$\frac{\partial f}{\partial x_1} = A_1, \quad \dots, \quad \frac{\partial f}{\partial x_m} = A_m$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 19 页 共 72 页

返回

全屏显示

关闭

退出



全改变量 $\Delta f(x_1, \cdots, x_m)$ 的线性主部

$$A_1 \Delta x_1 + \cdots + A_m \Delta x_m = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_m} dx_m$$

称为多变量函数 $f(x_1, \cdots, x_m)$ 的全微分，记为

$$df(x_1, \cdots, x_m) = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_m} dx_m \quad (5)$$

多变量函数 $f(x_1, \cdots, x_m)$ 在点 (x_1, \cdots, x_m) 可微分的充分条件是：若 $f(x_1, \cdots, x_m)$ 在点 (x_1, \cdots, x_m) 的偏导数 $\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_m}$ 均存在而且连续，则多变量函数在该点是可微分的。

访问清华主页

标题页

◀ ▶

◀ ▶

第20页共72页

返回

全屏显示

关闭

退出



1. 实矩阵微分

考虑实值标量函数 $f(\mathbf{x})$ ，其变元 $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathbb{R}^m$ 。将变元向量的元素 x_1, \dots, x_m 视为 m 个变量，则利用式(5)，可以直接引出以向量为变元的实值标量函数 $f(\mathbf{x})$ 的全微分表达式：

$$\begin{aligned} df(\mathbf{x}) &= \frac{\partial f(\mathbf{x})}{\partial x_1} dx_1 + \dots + \frac{\partial f(\mathbf{x})}{\partial x_m} dx_m \\ &= \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_m} \right] \begin{bmatrix} dx_1 \\ \vdots \\ dx_m \end{bmatrix} \end{aligned}$$

或简记为 $df(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T} d\mathbf{x}$ ，其中 $\frac{df}{d\mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T}$

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T} = \left[\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_m} \right]$$
$$d\mathbf{x} = [dx_1, \dots, dx_m]^T$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 21 页 共 72 页

返回

全屏显示

关闭

退出



考查实值标量函数 $f(\mathbf{X})$ ，其变元为 $m \times n$ 实矩阵 $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ 。记 $\mathbf{x}_j = [x_{1j}, \cdots, x_{mj}]^T, j = 1, \cdots, n$ ，则由实值标量函数 $f(\mathbf{x})$ 的全微分公式易知，实值矩阵作变元的实值标量函数 $f(\mathbf{X})$ 的全微分为

$$\begin{aligned} \mathrm{d}f(\mathbf{X}) &= \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_1} \mathrm{d}\mathbf{x}_1 + \cdots + \frac{\partial f(\mathbf{X})}{\partial \mathbf{x}_n} \mathrm{d}\mathbf{x}_n \\ &= \left[\frac{\partial f(\mathbf{X})}{\partial X_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial X_{m1}} \right] \begin{bmatrix} \mathrm{d}X_{11} \\ \vdots \\ \mathrm{d}X_{m1} \end{bmatrix} + \cdots + \left[\frac{\partial f(\mathbf{X})}{\partial X_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \right] \begin{bmatrix} \mathrm{d}X_{1n} \\ \vdots \\ \mathrm{d}X_{mn} \end{bmatrix} \\ &= \left[\frac{\partial f(\mathbf{X})}{\partial X_{11}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial X_{m1}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial X_{1n}}, \cdots, \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \right] \begin{bmatrix} \mathrm{d}X_{11} \\ \vdots \\ \mathrm{d}X_{m1} \\ \vdots \\ \mathrm{d}X_{1n} \\ \vdots \\ \mathrm{d}X_{mn} \end{bmatrix} \end{aligned}$$

[访问清华主页](#)[标题页](#)[◀](#) [▶](#)[◀](#) [▶](#)

第 22 页 共 72 页

[返回](#)[全屏显示](#)[关闭](#)[退出](#)

借助向量化函数，上式可以简写为

$$df(\mathbf{X}) = \text{rvec}(\mathbf{A})\text{vec}(d\mathbf{X}) \quad (6)$$

式中 $\text{rvec}(\mathbf{A})$ 是Jacobian矩阵的行向量化，并且

$$d\mathbf{X} = \begin{bmatrix} dX_{11} & \cdots & dX_{1n} \\ \vdots & \ddots & \vdots \\ dX_{m1} & \cdots & dX_{mn} \end{bmatrix} \quad (7)$$

以及

$$\mathbf{A} = D_{\mathbf{x}}f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}^T} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial X_{m1}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \end{bmatrix} \quad (8)$$

是实标量函数 $f(\mathbf{X})$ 在实矩阵点 \mathbf{X} 的偏导矩阵，即Jacobian矩阵。



访问清华主页

标题页

◀ ▶

◀ ▶

第 23 页 共 72 页

返回

全屏显示

关闭

退出

应用行向量化与列向量化的关系 $\text{rvec}(\mathbf{A}) = (\text{vec}(\mathbf{A}^T))^T$ ，式(6)可以改写为 $df(\mathbf{X}) = (\text{vec}(\mathbf{A}^T))^T \text{vec}(d\mathbf{X})$ 。然后，在向量化算子 vec 与迹函数之间的关系式 $\text{tr}(\mathbf{B}^T \mathbf{C}) = (\text{vec}(\mathbf{B}))^T \text{vec}(\mathbf{C})$ 中，令 $\mathbf{B} = \mathbf{A}^T$ 和 $\mathbf{C} = d\mathbf{X}$ ，又可将 $df(\mathbf{X}) = (\text{vec}(\mathbf{A}^T))^T \text{vec}(d\mathbf{X})$ 最终表示为

$$\underline{df(\mathbf{X}) = \text{tr}(\mathbf{A} d\mathbf{X})} \quad (9)$$

从上述讨论，可以引出如下定义：

- (1) 称实值标量函数 $f(\mathbf{X})$ 在实矩阵点 \mathbf{X} 是可微分的(differentiable);
- (2) 称线性主部的实矩阵 \mathbf{A} 为实值标量函数 $f(\mathbf{X})$ 相对于变元矩阵 $\mathbf{X} \in \mathbb{R}^{m \times n}$ 的一阶偏导矩阵即Jacobian矩阵。



访问清华主页

标题页

◀ ▶

◀ ▶

第 24 页 共 72 页

返回

全屏显示

关闭

退出



一阶偏导矩阵 A 是唯一确定的，即：若存在 A_1 和 A_2 满足 $df(\mathbf{X}) = A_i d\mathbf{X}, i = 1, 2$ ，则 $A_1 = A_2$ 。

重要的是，实值标量函数 $f(\mathbf{X})$ 相对于 $m \times n$ 矩阵变元 \mathbf{X} 的Jacobian矩阵和梯度矩阵之间存在以下关系：

$$\nabla_{\mathbf{X}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial X_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial X_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial X_{mn}} \end{bmatrix} = \mathbf{A}^T \quad (10)$$

综合式(9)和式(10)立即得到下面的重要结果。

[访问清华主页](#)

[标题页](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

第 25 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)



命题3 若矩阵的实值标量函数 $f(\mathbf{X})$ 在 $m \times n$ 矩阵点 \mathbf{X} 可微分, 则

$$df(\mathbf{X}) = \text{tr}(\mathbf{A}d\mathbf{X}) \Leftrightarrow \nabla_{\mathbf{x}} f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^T$$

命题3启示了利用矩阵微分求实值目标函数 $f(\mathbf{X})$ 的梯度矩阵 $\nabla f(\mathbf{X})$ 的有效方法:

- (1) 求实值目标函数 $f(\mathbf{X})$ 相对于变元矩阵 \mathbf{X} 的矩阵微分 $df(\mathbf{X})$, 并将其表示成规范形式 $df(\mathbf{X}) = \text{tr}(\mathbf{A}d\mathbf{X})$;
- (2) 实值目标函数 $f(\mathbf{X})$ 相对于 $m \times n$ 变元矩阵 \mathbf{X} 的梯度矩阵由 $\nabla f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^T$ 直接给出。

访问清华主页

标题页

◀ ▶

◀ ▶

第26页共72页

返回

全屏显示

关闭

退出



2. 实Hessian矩阵的二阶辨识表

令 $x, \boldsymbol{x}, \boldsymbol{X}$ 分别代表函数的实标量变元、 $m \times 1$ 实向量变元和 $m \times n$ 实矩阵变元，而 $f(\cdot), \boldsymbol{f}(\cdot), \boldsymbol{F}(\cdot)$ 则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。

Hessian矩阵可以通过下面的二阶辨识表求出。

访问清华主页

标题页

◀ ▶

◀ ▶

第27页共72页

返回

全屏显示

关闭

退出



表3 二阶辨识表

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

访问清华主页

标题页

◀ ▶

◀ ▶

第28页共72页

返回

全屏显示

关闭

退出

表中，对于实向量函数 f ,

$$B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_p \end{bmatrix}, \quad (B')_v = \begin{bmatrix} B_1^T \\ B_2^T \\ \vdots \\ B_p^T \end{bmatrix}$$

而对于实矩阵函数 F ,

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



[访问清华主页](#)

[标题页](#)

[◀◀](#) [▶▶](#)

[◀](#) [▶](#)

第 29 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)

3.3 实矩阵微分计算

1. 实矩阵微分的常用计算公式

通过考查 $(d\mathbf{X})_{ij}$ ，容易证明实矩阵微分具有以下两个基本性质。

转置： 矩阵转置的微分等于矩阵微分的转置，
即有 $d(\mathbf{X}^T) = (d\mathbf{X})^T$ 。

线性： $d(\alpha\mathbf{X} + \beta\mathbf{Y}) = \alpha d\mathbf{X} + \beta d\mathbf{Y}$ 。

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

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

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



访问清华主页

标题页

◀ ▶

◀ ▶

第30页共72页

返回

全屏显示

关闭

退出

例2 考虑矩阵乘积 UV 的微分矩阵，有

$$\begin{aligned} [d(UV)]_{ij} &= d([UV]_{ij}) \\ &= d\left(\sum_k u_{ik}v_{kj}\right) \\ &= \sum_k d(u_{ik}v_{kj}) \\ &= \sum_k [(du_{ik})v_{kj} + u_{ik}dv_{kj}] \\ &= \sum_k (du_{ik})v_{kj} + \sum_k u_{ik}dv_{kj} \\ &= [(dU)V]_{ij} + [UdV]_{ij} \end{aligned}$$

从而得 $d(UV) = (dU)V + UdV$ 。



访问清华主页

标题页

◀ ▶

◀ ▶

第31页共72页

返回

全屏显示

关闭

退出



矩阵微分的常用计算公式

- (1) 常数矩阵的微分矩阵为零矩阵，即 $d\mathbf{A} = \mathbf{O}$ 。
- (2) 常数 α 与矩阵 \mathbf{X} 的乘积的微分矩阵 $d(\alpha\mathbf{X}) = \alpha d\mathbf{X}$ 。
- (3) 矩阵转置的微分矩阵等于原矩阵的微分矩阵的转置，即 $d(\mathbf{X}^T) = (d\mathbf{X})^T$ 。
- (4) 两个矩阵函数的和(差)的微分矩阵为 $d(\mathbf{U} \pm \mathbf{V}) = d\mathbf{U} \pm d\mathbf{V}$ 。

访问清华主页

标题页

◀ ▶

◀ ▶

第 32 页 共 72 页

返回

全屏显示

关闭

退出

(5) 常数矩阵与矩阵乘积的微分矩阵为 $\underline{d(AXB) = A(dX)B}$ 。

(6) 矩阵函数 $U = F(X), V = G(X), W = H(X)$ 乘积的微分矩阵为

$$d(UV) = (dU)V + U(dV)$$

$$\underline{d(UVW) = (dU)VW + U(dV)W + UV(dW)}$$

特别地，若 A 为常数矩阵，则

$$\underline{d(AXX^T) = (dX)AX^T + XA(dX)^T}$$

和

$$\underline{d(X^TAX) = (dX)^TAX + X^TAdX}$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 33 页 共 72 页

返回

全屏显示

关闭

退出



(7) 矩阵函数的Kronecker积的微分矩阵为

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

(8) 矩阵函数的Hadamard积的微分矩阵为

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

(9) 向量化函数 $\text{vec}(\mathbf{X})$ 的微分矩阵等于 \mathbf{X} 的微分矩阵的向量化函数，即

$$d(\text{vec}(\mathbf{X})) = \text{vec}(d\mathbf{X})$$

访问清华主页

标题页

◀ ▶

◀ ▶

第34页共72页

返回

全屏显示

关闭

退出

(10) 矩阵 \mathbf{X} 的迹的矩阵微分 $d(\text{tr}(\mathbf{X}))$ 等于矩阵微分 $d\mathbf{X}$ 的迹 $\text{tr}(d\mathbf{X})$, 即有

$$d(\text{tr}(\mathbf{X})) = \text{tr}(d\mathbf{X})$$

特别地, 矩阵函数 $F(\mathbf{X})$ 的迹的矩阵微分为

$$d(\text{tr}(\mathbf{F}(\mathbf{X}))) = \text{tr}(d(\mathbf{F}(\mathbf{X})))$$

例如 $d(\text{tr}(\mathbf{X}^T \mathbf{X})) = 2\text{tr}(\mathbf{X}^T d\mathbf{X})$ 。

(11) 行列式的微分为

$$d|\mathbf{X}| = |\mathbf{X}| \text{tr}(\mathbf{X}^{-1} d\mathbf{X})$$

特别地, 矩阵函数的行列式的微分为

$$d|\mathbf{F}(\mathbf{X})| = |\mathbf{U}| \text{tr}(\mathbf{U}^{-1} d\mathbf{X})$$



访问清华主页

标题页

◀

▶

◀

▶

第 35 页 共 72 页

返回

全屏显示

关闭

退出



(12) 矩阵对数的微分矩阵为

$$d \log \mathbf{X} = \mathbf{X}^{-1} d\mathbf{X}$$

特别地，矩阵函数的对数的微分矩阵为

$$d \log(\mathbf{F}(\mathbf{X})) = (\mathbf{F}(\mathbf{X}))^{-1} d(\mathbf{F}(\mathbf{X}))$$

(13) 逆矩阵的微分矩阵为

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

访问清华主页

标题页

◀ ▶

◀ ▶

第 36 页 共 72 页

返回

全屏显示

关闭

退出



(14) Moore-Penrose逆矩阵的微分矩阵为

$$\begin{aligned} d(\mathbf{X}^\dagger) = & -\mathbf{X}^\dagger(d\mathbf{X})\mathbf{X}^\dagger + \mathbf{X}^\dagger(\mathbf{X}^\dagger)^T(d\mathbf{X}^T)(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger) \\ & + (\mathbf{I} - \mathbf{X}^\dagger\mathbf{X})(d\mathbf{X}^T)(\mathbf{X}^\dagger)^T\mathbf{X}^\dagger \end{aligned}$$

$$\begin{aligned} d(\mathbf{X}^\dagger\mathbf{X}) &= \mathbf{X}^\dagger(d\mathbf{X})(\mathbf{I} - \mathbf{X}^\dagger\mathbf{X}) + (\mathbf{X}^\dagger(d\mathbf{X})(\mathbf{I} - \mathbf{X}^\dagger\mathbf{X}))^T \\ d(\mathbf{X}\mathbf{X}^\dagger) &= (\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger)(d\mathbf{X})\mathbf{X}^\dagger + ((\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger)(d\mathbf{X})\mathbf{X}^\dagger)^T \end{aligned}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 37 页 共 72 页

返回

全屏显示

关闭

退出



迹函数的梯度矩阵

对于 $\text{tr}(\mathbf{X}^T \mathbf{X})$, 注意到 $\text{tr}(\mathbf{A}^T \mathbf{B}) = \text{tr}(\mathbf{B}^T \mathbf{A})$,
有

$$\begin{aligned} d \text{tr}(\mathbf{X}^T \mathbf{X}) &= \text{tr}(d(\mathbf{X}^T \mathbf{X})) \\ &= \text{tr}((d\mathbf{X})^T \mathbf{X} + \mathbf{X}^T d\mathbf{X}) \\ &= \text{tr}((d\mathbf{X})^T \mathbf{X}) + \text{tr}(\mathbf{X}^T d\mathbf{X}) \\ &= \text{tr}(2\mathbf{X}^T d\mathbf{X}) \end{aligned}$$

故由命题3直接得 $\text{tr}(\mathbf{X}^T \mathbf{X})$ 关于 \mathbf{X} 的梯度矩阵为

$$\frac{\partial \text{tr}(\mathbf{X}^T \mathbf{X})}{\partial \mathbf{X}} = (2\mathbf{X}^T)^T = 2\mathbf{X}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 38 页 共 72 页

返回

全屏显示

关闭

退出



考虑三个矩阵乘积的迹函数 $\text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X})$ ，其微分

$$\begin{aligned} d \text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) &= \text{tr}(d(\mathbf{X}^T \mathbf{A} \mathbf{X})) \\ &= \text{tr}((d\mathbf{X})^T \mathbf{A} \mathbf{X} + \mathbf{X}^T \mathbf{A} d\mathbf{X}) \\ &= \text{tr}((d\mathbf{X})^T \mathbf{A} \mathbf{X}) + \text{tr}(\mathbf{X}^T \mathbf{A} d\mathbf{X}) \\ &= \text{tr}((\mathbf{A} \mathbf{X})^T d\mathbf{X}) + \text{tr}(\mathbf{X}^T \mathbf{A} d\mathbf{X}) \\ &= \text{tr}(\mathbf{X}^T (\mathbf{A}^T + \mathbf{A}) d\mathbf{X}) \end{aligned}$$

从而得梯度矩阵

$$\frac{\partial \text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = [\mathbf{X}^T (\mathbf{A}^T + \mathbf{A})]^T = (\mathbf{A} + \mathbf{A}^T) \mathbf{X}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 39 页 共 72 页

返回

全屏显示

关闭

退出



再看一个包含了逆矩阵的迹函数 $\text{tr}(\mathbf{A}\mathbf{X}^{-1})$ 。
计算得

$$\begin{aligned} d \text{tr}(\mathbf{A}\mathbf{X}^{-1}) &= \text{tr} [d(\mathbf{A}\mathbf{X}^{-1})] \\ &= \text{tr} [\mathbf{A}d\mathbf{X}^{-1}] \\ &= -\text{tr} [\mathbf{A}\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}] \\ &= -\text{tr} (\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}d\mathbf{X}) \end{aligned}$$

由此得梯度矩阵

$$\frac{\partial \text{tr}(\mathbf{A}\mathbf{X}^{-1})}{\partial \mathbf{X}} = -(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})^T$$

课堂练习：求四个矩阵乘积的迹函数
 $\text{tr}(\mathbf{X}\mathbf{A}\mathbf{X}\mathbf{B})$ 的微分矩阵和梯度矩阵。

[访问清华主页](#)

[标题页](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

第 40 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)



以上举例可以总结出应用命题3的要点如下:

(1) 实值标量函数 $f(\mathbf{X})$ 总可以写成迹函数的形式, 因为 $f(\mathbf{X}) = \text{tr}(f(\mathbf{X}))$;

(2) 无论 $d\mathbf{X}$ 出现在迹函数内的任何位置, 总可以通过迹函数的性质 $\text{tr}[\mathbf{A}(d\mathbf{X})\mathbf{B}] = \text{tr}(\mathbf{B}\mathbf{A}d\mathbf{X})$, 将 $d\mathbf{X}$ 写到迹函数变量的最右端, 从而得到迹函数微分矩阵的规范形式。

(3) 对于 $(d\mathbf{X})^T$, 总可以通过迹函数的性质

$$\text{tr}[\mathbf{A}(d\mathbf{X})^T\mathbf{B}] = \text{tr}[(\mathbf{A}(d\mathbf{X})^T\mathbf{B})^T] = \text{tr}(\mathbf{A}^T\mathbf{B}^Td\mathbf{X})$$

写成迹函数微分矩阵的规范形式。

访问清华主页

标题页

◀ ▶

◀ ▶

第 41 页 共 72 页

返回

全屏显示

关闭

退出

表4 几种迹函数的微分矩阵与梯度矩阵的对应关系

迹函数 $f(\mathbf{X})$	微分矩阵 $\mathrm{d}f(\mathbf{X})$	梯度矩阵 $\partial f(\mathbf{X})/\partial \mathbf{X}$
$\mathrm{tr}(\mathbf{X})$	$\mathrm{tr}(\mathbf{I} \mathrm{d}\mathbf{X})$	\mathbf{I}
$\mathrm{tr}(\mathbf{X}^{-1})$	$-\mathrm{tr}(\mathbf{X}^{-2} \mathrm{d}\mathbf{X})$	$(\mathbf{X}^{-2})^{\mathrm{T}}$
$\mathrm{tr}(\mathbf{A}\mathbf{X})$	$\mathrm{tr}(\mathbf{A} \mathrm{d}\mathbf{X})$	\mathbf{A}^{T}
$\mathrm{tr}(\mathbf{X}^2)$	$2\mathrm{tr}(\mathbf{X} \mathrm{d}\mathbf{X})$	$2\mathbf{X}^{\mathrm{T}}$
$\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{X})$	$2\mathrm{tr}(\mathbf{X}^{\mathrm{T}} \mathrm{d}\mathbf{X})$	$2\mathbf{X}$
$\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X})$	$\mathrm{tr}[\mathbf{X}^{\mathrm{T}}(\mathbf{A} + \mathbf{A}^{\mathrm{T}}) \mathrm{d}\mathbf{X}]$	$(\mathbf{A} + \mathbf{A}^{\mathrm{T}})\mathbf{X}$
$\mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}})$	$\mathrm{tr}[(\mathbf{A} + \mathbf{A}^{\mathrm{T}})\mathbf{X}^{\mathrm{T}} \mathrm{d}\mathbf{X}]$	$\mathbf{X}(\mathbf{A} + \mathbf{A}^{\mathrm{T}})$
$\mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X})$	$\mathrm{tr}[(\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{A}) \mathrm{d}\mathbf{X}]$	$\mathbf{X}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}$
$\mathrm{tr}(\mathbf{A}\mathbf{X}^{-1})$	$-\mathrm{tr}(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1} \mathrm{d}\mathbf{X})$	$-(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})^{\mathrm{T}}$
$\mathrm{tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})$	$-\mathrm{tr}(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1} \mathrm{d}\mathbf{X})$	$-(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^{\mathrm{T}}$
$\mathrm{tr}[(\mathbf{X} + \mathbf{A})^{-1}]$	$-\mathrm{tr}[(\mathbf{X} + \mathbf{A})^{-2} \mathrm{d}\mathbf{X}]$	$-[(\mathbf{X} + \mathbf{A})^{-2}]^{\mathrm{T}}$
$\mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}\mathbf{B})$	$\mathrm{tr}[(\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{B}\mathbf{X}\mathbf{A}) \mathrm{d}\mathbf{X}]$	$(\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{B}\mathbf{X}\mathbf{A})^{\mathrm{T}}$
$\mathrm{tr}(\mathbf{X}\mathbf{A}\mathbf{X}^{\mathrm{T}}\mathbf{B})$	$\mathrm{tr}[(\mathbf{A}\mathbf{X}^{\mathrm{T}}\mathbf{B} + \mathbf{A}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}) \mathrm{d}\mathbf{X}]$	$\mathbf{B}^{\mathrm{T}}\mathbf{X}\mathbf{A}^{\mathrm{T}} + \mathbf{B}\mathbf{X}\mathbf{A}$
$\mathrm{tr}(\mathbf{A}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{B})$	$\mathrm{tr}[\mathbf{X}^{\mathrm{T}}(\mathbf{B}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}) \mathrm{d}\mathbf{X}]$	$(\mathbf{B}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}})\mathbf{X}$
$\mathrm{tr}(\mathbf{A}\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{B})$	$\mathrm{tr}[(\mathbf{B}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}})\mathbf{X}^{\mathrm{T}} \mathrm{d}\mathbf{X}]$	$\mathbf{X}(\mathbf{B}\mathbf{A} + \mathbf{A}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}})$



访问清华主页

标题页

◀

▶

◀

▶

第 42 页 共 72 页

返回

全屏显示

关闭

退出

3.4 梯度与共轭梯度

梯度：目标函数相对于复向量或者复矩阵本身的梯度

共轭梯度：目标函数相对于复共轭向量或者复共轭矩阵的梯度

表5 复值函数的分类

函数类型	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 $f \in \mathbb{C}$	$f(z, z^*)$ $f : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}$	$f(z, z^*)$ $f : \mathbb{C}^m \times \mathbb{C}^m \rightarrow \mathbb{C}$	$f(Z, Z^*)$ $f : \mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n} \rightarrow \mathbb{C}$
向量函数 $f \in \mathbb{C}^p$	$f(z, z^*)$ $f : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}^p$	$f(z, z^*)$ $f : \mathbb{C}^m \times \mathbb{C}^m \rightarrow \mathbb{C}^p$	$f(Z, Z^*)$ $f : \mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^p$
矩阵函数 $F \in \mathbb{C}^{p \times q}$	$F(z, z^*)$ $F : \mathbb{C} \times \mathbb{C} \rightarrow \mathbb{C}^{p \times q}$	$F(z, z^*)$ $F : \mathbb{C}^m \times \mathbb{C}^m \rightarrow \mathbb{C}^{p \times q}$	$F(Z, Z^*)$ $F : \mathbb{C}^{m \times n} \times \mathbb{C}^{m \times n} \rightarrow \mathbb{C}^{p \times q}$



访问清华主页

标题页

◀

▶

◀

▶

第 43 页 共 72 页

返回

全屏显示

关闭

退出

令复变函数 $f(z)$ 可以用实部 $u(x, y)$ 和虚部 $v(x, y)$ 写作

$$f(z) = u(x, y) + \mathrm{j}v(x, y)$$

式中 $z = x + \mathrm{j}y$ ，并且 $u(x, y)$ 和 $v(x, y)$ 分别是实值函数。

关于全纯函数，以下四个叙述等价：

1. 复变函数 $f(z)$ 是全纯函数(即复解析函数);
2. 复变函数的导数 $f'(z)$ 存在，并且连续;
3. 复变函数 $f(z)$ 满足Cauchy-Riemann方程:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{和} \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}$$

4. 复变函数 $f(z)$ 的所有导数存在，并且具有一个收敛的幂级数。



访问清华主页

标题页

◀ ▶

◀ ▶

第 44 页 共 72 页

返回

全屏显示

关闭

退出



Cauchy-Riemann条件的一个直接结果是：函数 $f(z) = u(x, y) + jv(x, y)$ 为全纯函数，仅当实变函数 $u(x, y)$ 和 $v(x, y)$ 同时满足Laplace方程：

$$\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0$$

和

$$\frac{\partial^2 v(x, y)}{\partial x^2} + \frac{\partial^2 v(x, y)}{\partial y^2} = 0$$

满足Laplace方程

$$\frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} = 0$$

的实变函数 $f(x, y)$ 称为调和函数(harmonic function)。

[访问清华主页](#)

[标题页](#)

[«](#) [»](#)

[◀](#) [▶](#)

第 45 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)



一个复变函数 $f(z) = u(x, y) + jv(x, y)$ 只要其中任何一个实变函数 $u(x, y)$ 或者 $v(x, y)$ 不满足Cauchy-Riemann条件或者Laplace条件, 那么它就不是一个全纯函数。

虽然幂函数 z^n 、指数函数 e^z 、对数函数 $\ln z$ 、正弦函数 $\sin z$ 和余弦函数 $\cos z$ 等许多函数都是全纯函数, 即全复平面上的解析函数。但是, 一些实际中经常遇到的常用函数却不是全纯函数:

1. 复变函数 $f(z) = z^* = x - jy = u(x, y) + jv(x, y)$ 中的实变函数 $u(x, y) = x$ 和 $v(x, y) = -y$ 显然不满足Cauchy-Riemann条件 $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ 。

2. 复变函数 $f(z) = \operatorname{Re}(z) = x$ 和 $f(z) = \operatorname{Im}(z) = y$ 都不满足Cauchy-Riemann条件。

访问清华主页

标题页

◀ ▶

◀ ▶

第 46 页 共 72 页

返回

全屏显示

关闭

退出

3.任何一个非常数的实值复变函数 $f(z) \in \mathbb{R}$ 都不满足Cauchy-Riemann条件 $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ 和 $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$, 因为 $f(z) = u(x, y) + jv(x, y)$ 中的实变函数 $v(x, y) = 0$ 。特别地, 实值函数 $f(z) = |z| = \sqrt{x^2 + y^2}$ 是不可微分的, 而 $f(z) = |z|^2 = x^2 + y^2 = u(x, y) + jv(x, y)$ 中的实变函数 $u(x, y) = x^2 + y^2$ 不是调和函数, 因为它不满足Laplace条件 $\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0$ 。

问题: 对于一个给定的复变函数, 是否能够保证它是全纯(即复解析)函数, 从而求出它关于 z 或者 z^* 的偏导呢?

为了回答这个问题, 有必要回忆复变函数论中对复数 z 和共轭复数 z^* 的偏导数的定义:



访问清华主页

标题页

◀ ▶

◀ ▶

第 47 页 共 72 页

返回

全屏显示

关闭

退出

$$\frac{\partial}{\partial z} = \frac{1}{2} \left(\frac{\partial}{\partial x} - j \frac{\partial}{\partial y} \right), \quad \frac{\partial}{\partial z^*} = \frac{1}{2} \left(\frac{\partial}{\partial x} + j \frac{\partial}{\partial y} \right)$$

关于复变量 $z = x + jy$ 的偏导，有一个实部和虚部的独立性基本假设：

$$\frac{\partial x}{\partial y} = 0 \quad \text{和} \quad \frac{\partial y}{\partial x} = 0$$

由实际偏导的定义及上述独立性假设，容易求出

$$\begin{aligned} \frac{\partial z}{\partial z^*} &= \frac{\partial x}{\partial z^*} + j \frac{\partial y}{\partial z^*} = \frac{1}{2} \left(\frac{\partial x}{\partial x} + j \frac{\partial x}{\partial y} \right) + j \frac{1}{2} \left(\frac{\partial y}{\partial x} + j \frac{\partial y}{\partial y} \right) \\ &= \frac{1}{2}(1 + 0) + j \frac{1}{2}(0 + j) \\ \frac{\partial z^*}{\partial z} &= \frac{\partial x}{\partial z} - j \frac{\partial y}{\partial z} = \frac{1}{2} \left(\frac{\partial x}{\partial x} - j \frac{\partial x}{\partial y} \right) - j \frac{1}{2} \left(\frac{\partial y}{\partial x} - j \frac{\partial y}{\partial y} \right) \\ &= \frac{1}{2}(1 - 0) - j \frac{1}{2}(0 - j) \end{aligned}$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 48 页 共 72 页

返回

全屏显示

关闭

退出

即有

$$\frac{\partial z}{\partial z^*} = 0 \quad \text{和} \quad \frac{\partial z^*}{\partial z} = 0$$

复变量理论的一个基本结果：在复变量的实部与虚部相互独立的基本假设下，求偏导数时，复变量 z 和其复数共轭变量 z^* 可以当作两个相互独立的变量处理，即任何一个相对于另一个都可认为是常数。因此，上述非全纯函数如何微分的问题便有了解决方法：在标准的复变函数框架内，一个复变函数 $f(z)$ (其中 $z = x + jy$) 使用实(极)坐标 $r \triangleq (x, y)^T$ 表示为 $f(r) = f(x, y)$ ，而在复导数的框架内，则使用共轭坐标 $c \triangleq (z, z^*)^T$ 替代实坐标 $r = (x, y)^T$ ，将复变函数 $f(z)$ 写成 $f(c) = f(z, z^*)$ 的形式。



访问清华主页

标题页

◀ ▶

◀ ▶

第 49 页 共 72 页

返回

全屏显示

关闭

退出



如果将一个非全纯函数 $f(z)$ 写成 $f(z, z^*)$ 的形式, 则对于固定的 z^* , 复变函数 $f(z, z^*)$ 在 $z = x + jy$ 全平面是解析的, 并且对于固定的 z 值, 复变函数 $f(z, z^*)$ 在 $z^* = x - jy$ 全平面上也是解析的。换句话说, 一个复变函数写成 $f(z, z^*)$ 之后, 则其偏导

$$\nabla_z f(z, z^*) = \left. \frac{\partial f(z, z^*)}{\partial z} \right|_{z^*=\text{常数}}$$

和共轭偏导

$$\nabla_{z^*} f(z, z^*) = \left. \frac{\partial f(z, z^*)}{\partial z^*} \right|_{z=\text{常数}}$$

一定存在, 并且连续。这一结果告诉我们, 求复变函数 $f(z, z^*)$ 关于 z 的偏导 $\frac{\partial f(z, z^*)}{\partial z}$ 或者关于 z^* 的偏导 $\frac{\partial f(z, z^*)}{\partial z^*}$, 只需要将 z^* 或者 z 分别视为常数即可。

访问清华主页

标题页

◀ ▶

◀ ▶

第 50 页 共 72 页

返回

全屏显示

关闭

退出

复变函数偏导的常用公式与法则:

1. 复变函数共轭 $f^*(z, z^*)$ 关于 z 变量共轭 z^* 的偏导等于原复变函数 $f(z, z^*)$ 关于 z 变量的偏导的共轭, 即

$$\frac{\partial f^*(z, z^*)}{\partial z^*} = \left(\frac{\partial f(z, z^*)}{\partial z} \right)^*$$

2. 复变函数共轭 $f^*(z, z^*)$ 关于 z 变量的偏导等于原复变函数 $f(z, z^*)$ 关于共轭变量 z^* 的偏导的共轭, 即

$$\frac{\partial f^*(z, z^*)}{\partial z} = \left(\frac{\partial f(z, z^*)}{\partial z^*} \right)^*$$

3. 复微分法则:

$$df(z, z^*) = \frac{\partial f(z, z^*)}{\partial z} dz + \frac{\partial f(z, z^*)}{\partial z^*} dz^*$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 51 页 共 72 页

返回

全屏显示

关闭

退出



4.链式法则:

$$\frac{\partial h(g(z, z^*))}{\partial z} = \frac{\partial h(g(z, z^*))}{\partial g(z, z^*)} \frac{\partial g(z, z^*)}{\partial z} + \frac{\partial h(g(z, z^*))}{\partial g^*(z, z^*)} \frac{\partial g^*(z, z^*)}{\partial z}$$
$$\frac{\partial h(g(z, z^*))}{\partial z^*} = \frac{\partial h(g(z, z^*))}{\partial g(z, z^*)} \frac{\partial g(z, z^*)}{\partial z^*} + \frac{\partial h(g(z, z^*))}{\partial g^*(z, z^*)} \frac{\partial g^*(z, z^*)}{\partial z^*}$$

令 $\mathbf{z} = \mathbf{x} + \mathrm{j}\mathbf{y} = [z_1, \dots, z_m]^T \in \mathbb{C}^m$, 其中 $\mathbf{x} = [x_1, \dots, x_m]^T \in \mathbb{R}^m$, $\mathbf{y} = [y_1, \dots, y_m]^T \in \mathbb{R}^m$, 即 $z_i = x_i + \mathrm{j}y_i, i = 1, \dots, m$, 并且实部 x_i 与虚部 y_i 是相互独立的变元。

关于复变元列向量 $\mathbf{z} \in \mathbb{C}^m$ 的协梯度算子(cogradient operator)和共轭协梯度算子(conjugate cogradient operator)均采用行向量形式, 分别定义为

$$\mathbf{D}_{\mathbf{z}} = \frac{\partial}{\partial \mathbf{z}^T} \triangleq \left[\frac{\partial}{\partial z_1}, \dots, \frac{\partial}{\partial z_m} \right] \text{ 和 } \mathbf{D}_{\mathbf{z}^*} = \frac{\partial}{\partial \mathbf{z}^H} \triangleq \left[\frac{\partial}{\partial z_1^*}, \dots, \frac{\partial}{\partial z_m^*} \right]$$

[访问清华主页](#)[标题页](#)[◀◀](#) [▶▶](#)[◀](#) [▶](#)

第 52 页 共 72 页

[返回](#)[全屏显示](#)[关闭](#)[退出](#)

对行向量 $\mathbf{z}^T = [z_1, \dots, z_m]$ 的每个元素运用复变函数的偏导算子

$$D_z = \frac{\partial}{\partial z_i} = \frac{1}{2} \left(\frac{\partial}{\partial x_i} - j \frac{\partial}{\partial y_i} \right) \quad \text{和} \quad D_{z^*} = \frac{\partial}{\partial z_i^*} = \frac{1}{2} \left(\frac{\partial}{\partial x_i} + j \frac{\partial}{\partial y_i} \right)$$

便得到用复变向量 \mathbf{z} 的实部 \mathbf{x} 与虚部 \mathbf{y} 表示的协梯度算子

$$D_z = \frac{\partial}{\partial \mathbf{z}^T} = \frac{1}{2} \left(\frac{\partial}{\partial \mathbf{x}^T} - j \frac{\partial}{\partial \mathbf{y}^T} \right)$$

和共轭协梯度算子

$$D_{z^*} = \frac{\partial}{\partial \mathbf{z}^H} = \frac{1}{2} \left(\frac{\partial}{\partial \mathbf{x}^T} + j \frac{\partial}{\partial \mathbf{y}^T} \right)$$

类似地，梯度算子(gradient operator)和共轭梯度算子(conjugate gradient operator)采用列向量形式，分别定义为

$$\nabla_z = \frac{\partial}{\partial \mathbf{z}} \triangleq \left[\frac{\partial}{\partial z_1}, \dots, \frac{\partial}{\partial z_m} \right]^T, \quad \nabla_{z^*} = \frac{\partial}{\partial \mathbf{z}^*} \triangleq \left[\frac{\partial}{\partial z_1^*}, \dots, \frac{\partial}{\partial z_m^*} \right]^T$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 53 页 共 72 页

返回

全屏显示

关闭

退出

对复列向量 $\mathbf{z} = [z_1, \dots, z_m]^T$ 的每个元素运用复变函数的偏导算子，立即得到用复变向量 \mathbf{z} 的实部 \mathbf{x} 与虚部 \mathbf{y} 表示的梯度算子

$$\nabla_{\mathbf{z}} = \frac{\partial}{\partial \mathbf{z}} = \frac{1}{2} \left(\frac{\partial}{\partial \mathbf{x}} - j \frac{\partial}{\partial \mathbf{y}} \right)$$

和共轭梯度算子

$$\nabla_{\mathbf{z}^*} = \frac{\partial}{\partial \mathbf{z}^*} = \frac{1}{2} \left(\frac{\partial}{\partial \mathbf{x}} + j \frac{\partial}{\partial \mathbf{y}} \right)$$

利用梯度算子和共轭梯度算子的定义公式，不难求出

$$\frac{\partial \mathbf{z}^T}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{z}} + j \frac{\partial \mathbf{y}^T}{\partial \mathbf{z}} = \frac{1}{2} \left(\frac{\partial \mathbf{x}^T}{\partial \mathbf{x}} - j \frac{\partial \mathbf{x}^T}{\partial \mathbf{y}} \right) + j \frac{1}{2} \left(\frac{\partial \mathbf{y}^T}{\partial \mathbf{x}} - j \frac{\partial \mathbf{y}^T}{\partial \mathbf{y}} \right)$$

$$= \mathbf{I}_{m \times m}$$

$$\frac{\partial \mathbf{z}^T}{\partial \mathbf{z}^*} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{z}^*} + j \frac{\partial \mathbf{y}^T}{\partial \mathbf{z}^*} = \frac{1}{2} \left(\frac{\partial \mathbf{x}^T}{\partial \mathbf{x}} + j \frac{\partial \mathbf{x}^T}{\partial \mathbf{y}} \right) + j \frac{1}{2} \left(\frac{\partial \mathbf{y}^T}{\partial \mathbf{x}} + j \frac{\partial \mathbf{y}^T}{\partial \mathbf{y}} \right)$$

$$= \mathbf{O}_{m \times m}$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 54 页 共 72 页

返回

全屏显示

关闭

退出



将上述结果以及它们的共轭、转置和复共轭转置一并写出，便有下列重要结果：

$$\begin{aligned}\frac{\partial \mathbf{z}^T}{\partial \mathbf{z}} &= \mathbf{I}, & \frac{\partial \mathbf{z}^H}{\partial \mathbf{z}^*} &= \mathbf{I}, & \frac{\partial \mathbf{z}}{\partial \mathbf{z}^T} &= \mathbf{I}, & \frac{\partial \mathbf{z}^*}{\partial \mathbf{z}^H} &= \mathbf{I} \\ \frac{\partial \mathbf{z}^T}{\partial \mathbf{z}^*} &= \mathbf{O}, & \frac{\partial \mathbf{z}^H}{\partial \mathbf{z}} &= \mathbf{O}, & \frac{\partial \mathbf{z}}{\partial \mathbf{z}^H} &= \mathbf{O}, & \frac{\partial \mathbf{z}^*}{\partial \mathbf{z}^T} &= \mathbf{O}\end{aligned}$$

从这些结果可以总结出协梯度算子和梯度算子的下列应用法则：

1. 无论使用协梯度算子 $\frac{\partial}{\partial \mathbf{z}^T}$ 还是梯度算子 $\frac{\partial}{\partial \mathbf{z}}$ ，复共轭变元向量 \mathbf{z}^* 均可视为一常数向量；
2. 无论使用共轭协梯度算子 $\frac{\partial}{\partial \mathbf{z}^H}$ 还是共轭梯度算子 $\frac{\partial}{\partial \mathbf{z}^*}$ ，向量 \mathbf{z} 都可当作一常数向量处理。

访问清华主页

标题页

◀ ▶

◀ ▶

第 55 页 共 72 页

返回

全屏显示

关闭

退出



不妨称上述法则为复偏导算子的独立法则：当使用复偏导算子(协梯度算子、共轭协梯度算子、梯度算子和共轭梯度算子)时，复变元向量 z 和 z^* 可以当作两个相互独立的变元向量处理，即其中一个向量作为变元时，另一个向量便可视为常数向量。

给定一复变函数 $f(z, z^*) \in \mathbb{R}$ ，其中 $z \in \mathbb{C}^m$ 。

协梯度向量、共轭协梯度向量的定义公式：

$$\begin{aligned} D_z f(z, z^*) &= \left. \frac{\partial f(z, z^*)}{\partial z^T} \right|_{z^* = \text{常数向量}} \\ D_{z^*} f(z, z^*) &= \left. \frac{\partial f(z, z^*)}{\partial z^H} \right|_{z = \text{常数向量}} \end{aligned}$$

梯度向量、共轭梯度向量的定义公式：

$$\begin{aligned} \nabla_z f(z, z^*) &= \left. \frac{\partial f(z, z^*)}{\partial z} \right|_{z^* = \text{常数向量}} \\ \nabla_{z^*} f(z, z^*) &= \left. \frac{\partial f(z, z^*)}{\partial z^*} \right|_{z = \text{常数向量}} \end{aligned}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 56 页 共 72 页

返回

全屏显示

关闭

退出



实值标量函数 $f(\mathbf{Z}, \mathbf{Z}^*)$ 的梯度矩阵和共轭梯度矩阵分别定义为

$$\nabla_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*) = \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}}$$
$$\nabla_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*) = \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*}$$

梯度矩阵和共轭梯度矩阵的计算公式：

$$\nabla_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*) = \left. \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} \right|_{\mathbf{Z}^* = \text{常数矩阵}}$$
$$\nabla_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*) = \left. \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} \right|_{\mathbf{Z} = \text{常数矩阵}}$$

Hessian矩阵(共轭梯度的协梯度)

$$\nabla_{\mathbf{Z}^T} [\nabla_{\mathbf{Z}^*} f(\mathbf{z}, \mathbf{Z}^*)] = \frac{\partial}{\partial \mathbf{Z}^T} \left[\frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^*} \right]$$

[访问清华主页](#)

[标题页](#)

[«](#) [»](#)

[◀](#) [▶](#)

第 57 页 共 72 页

[返回](#)

[全屏显示](#)

[关闭](#)

[退出](#)



访问清华主页

标题页

◀ ▶

◀ ▶

第 58 页 共 72 页

返回

全屏显示

关闭

退出

下面是梯度矩阵的运算法则。

1. 若 $f(\mathbf{Z}, \mathbf{Z}^*) = c$ 为常数，则梯度矩阵和共轭梯度矩阵均等于零矩阵，即 $\frac{\partial c}{\partial \mathbf{Z}} = \mathbf{O}$ 和 $\frac{\partial c}{\partial \mathbf{Z}^*} = \mathbf{O}$ 。

2. 线性法则：若 $f(\mathbf{Z}, \mathbf{Z}^*)$ 和 $g(\mathbf{Z}, \mathbf{Z}^*)$ 分别是复变函数，而 c_1 和 c_2 为复常数，则

$$\frac{\partial [c_1 f(\mathbf{Z}, \mathbf{Z}^*) + c_2 g(\mathbf{Z}, \mathbf{Z}^*)]}{\partial \mathbf{Z}^*} = c_1 \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} + c_2 \frac{\partial g(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*}$$

3. 乘积法则：

$$\frac{\partial f(\mathbf{Z}, \mathbf{Z}^*) g(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} = g(\mathbf{Z}, \mathbf{Z}^*) \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} + f(\mathbf{Z}, \mathbf{Z}^*) \frac{\partial g(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*}$$

4. 商法则：若 $g(\mathbf{Z}, \mathbf{Z}^*) \neq 0$ ，则

$$\begin{aligned} & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*) / g(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} \\ &= \frac{1}{g^2(\mathbf{Z}, \mathbf{Z}^*)} \left[g(\mathbf{Z}, \mathbf{Z}^*) \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} - f(\mathbf{Z}, \mathbf{Z}^*) \frac{\partial g(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} \right] \end{aligned}$$



5. 复合函数 $H(\mathbf{Z}, \mathbf{Z}^*) = G(F(\mathbf{Z}, \mathbf{Z}^*))$ 的梯度矩阵和共轭梯度矩阵的链式法则:

$$\begin{aligned}\frac{\partial H(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} &= \frac{\partial G(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} + \\ &\quad \frac{\partial G(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} \\ \frac{\partial H(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} &= \frac{\partial G(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} + \\ &\quad \frac{\partial G(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*}\end{aligned}$$

特别地, 若 $h(\mathbf{Z}, \mathbf{Z}^*) = g(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))$, 则

$$\begin{aligned}\frac{\partial h(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} &= \frac{\partial g(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} + \\ &\quad \frac{\partial g(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} \\ \frac{\partial h(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} &= \frac{\partial g(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} + \\ &\quad \frac{\partial g(F(\mathbf{Z}, \mathbf{Z}^*), F^*(\mathbf{Z}, \mathbf{Z}^*))}{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)} \cdot \frac{\partial F^*(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*}\end{aligned}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 59 页 共 72 页

返回

全屏显示

关闭

退出

3.5 复矩阵微分



单变元复变函数的复微分法则

$$df(z, z^*) = \frac{\partial f(z, z^*)}{\partial z} dz + \frac{\partial f(z, z^*)}{\partial z^*} dz^*$$

多变元复变函数 $f(\cdot) = f((z_1, z_1^*), \dots, (z_m, z_m^*))$ 的复微分法则:

$$df(\cdot) = \frac{\partial f(\cdot)}{\partial z_1} dz_1 + \dots + \frac{\partial f(\cdot)}{\partial z_m} dz_m + \frac{\partial f(\cdot)}{\partial z_1^*} dz_1^* + \dots + \frac{\partial f(\cdot)}{\partial z_m^*} dz_m^*$$

这一复微分法则是复矩阵微分的基础。

将 $m \times 1$ 复变元向量 $z = [z_1, \dots, z_m]^T$ 的每个元素视为多元复变函数的复变量, 则由多元复变函数的复微分法则, 有

访问清华主页

标题页

◀ ▶

◀ ▶

第 60 页 共 72 页

返回

全屏显示

关闭

退出

$$\begin{aligned}
 df(\mathbf{z}, \mathbf{z}^*) &= \left[\frac{f(\mathbf{z}, \mathbf{z}^*)}{\partial z_1}, \dots, \frac{f(\mathbf{z}, \mathbf{z}^*)}{\partial z_m} \right] \begin{bmatrix} dz_1 \\ \vdots \\ dz_m \end{bmatrix} \\
 &\quad + \left[\frac{f(\mathbf{z}, \mathbf{z}^*)}{\partial z_1^*}, \dots, \frac{f(\mathbf{z}, \mathbf{z}^*)}{\partial z_m^*} \right] \begin{bmatrix} dz_1^* \\ \vdots \\ dz_m^* \end{bmatrix} \\
 &= \frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^T} d\mathbf{z} + \frac{\partial f(\mathbf{z}, \mathbf{z}^*)}{\partial \mathbf{z}^H} d\mathbf{z}^*
 \end{aligned}$$

或简记作

$$df(\mathbf{z}, \mathbf{z}^*) = D_{\mathbf{z}} f(\mathbf{z}, \mathbf{z}^*) d\mathbf{z} + D_{\mathbf{z}^*} f(\mathbf{z}, \mathbf{z}^*) d\mathbf{z}^*$$

式中

$$d\mathbf{z} = [dz_1, \dots, dz_m]^T$$

$$d\mathbf{z}^* = [dz_1^*, \dots, dz_m^*]^T$$



访问清华主页

标题页

◀

▶

◀

▶

第 61 页 共 72 页

返回

全屏显示

关闭

退出



$$D_z f(z, z^*) = \frac{\partial f(z, z^*)}{\partial \mathbf{z}^T} = \left[\frac{\partial f(z, z^*)}{\partial z_1}, \dots, \frac{\partial f(z, z^*)}{\partial z_m} \right] \in \mathbb{C}^{1 \times m}$$
$$D_{z^*} f(z, z^*) = \frac{\partial f(z, z^*)}{\partial \mathbf{z}^H} = \left[\frac{\partial f(z, z^*)}{\partial z_1^*}, \dots, \frac{\partial f(z, z^*)}{\partial z_m^*} \right] \in \mathbb{C}^{1 \times m}$$

分别是复变函数 $f(z, z^*)$ 的协梯度向量和共轭协梯度向量。于是，复变函数 $f(z, z^*)$ 的梯度向量和共轭梯度向量分别为

$$\nabla_z f(z, z^*) = \frac{\partial f(z, z^*)}{\partial \mathbf{z}} = (D_z f(z, z^*))^T$$
$$\nabla_{z^*} f(z, z^*) = \frac{\partial f(z, z^*)}{\partial \mathbf{z}^*} = (D_{z^*} f(z, z^*))^T$$

[访问清华主页](#)[标题页](#)[◀◀](#) [▶▶](#)[◀](#) [▶](#)

第 62 页 共 72 页

[返回](#)[全屏显示](#)[关闭](#)[退出](#)

考虑以矩阵作变元的复变函数 $f(\mathbf{Z}, \mathbf{Z}^*)$ ，其中 $\mathbf{Z} \in \mathbb{C}^{m \times n}$ 。将矩阵变元 \mathbf{Z} 和 \mathbf{Z}^* 分别向量化，则

$$df(\mathbf{Z}, \mathbf{Z}^*) = \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^T(\mathbf{Z})} d(\text{vec} \mathbf{Z}) + \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^H(\mathbf{Z})} d(\text{vec} \mathbf{Z}^*)$$

或写作

$$df(\mathbf{Z}, \mathbf{Z}^*) = \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^T(\mathbf{Z})} \text{vec}(d\mathbf{Z}) + \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^H(\mathbf{Z})} \text{vec}(d\mathbf{Z}^*) \quad (11)$$

式中

$$\begin{aligned} \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^T(\mathbf{Z})} &= \left[\frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{11}}, \dots, \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{m1}}, \dots, \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{1n}}, \dots, \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{mn}} \right] \\ \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^H(\mathbf{Z})} &= \left[\frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{11}^*}, \dots, \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{m1}^*}, \dots, \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{1n}^*}, \dots, \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{mn}^*} \right] \end{aligned}$$

另一方面，复变函数 $f(\mathbf{Z}, \mathbf{Z}^*)$ 的Jacobian矩阵和共轭Jacobian矩阵分别定义为



访问清华主页

标题页

◀ ▶

◀ ▶

第 63 页 共 72 页

返回

全屏显示

关闭

退出



访问清华主页

标题页

◀ ▶

◀ ▶

第 64 页 共 72 页

返回

全屏显示

关闭

退出

$$D_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*) \triangleq \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^T} = \begin{bmatrix} \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{11}} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{1n}} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{mn}} \end{bmatrix}$$

$$D_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*) \triangleq \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^H} = \begin{bmatrix} \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{11}^*} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{m1}^*} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{1n}^*} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{mn}^*} \end{bmatrix}$$

类似地，复变函数 $f(\mathbf{Z}, \mathbf{Z}^*)$ 的梯度矩阵和共轭梯度矩阵分别定义为

$$\nabla_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*) \triangleq \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}} = \begin{bmatrix} \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{11}} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{m1}} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{mn}} \end{bmatrix}$$

$$\nabla_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*) \triangleq \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \mathbf{Z}^*} = \begin{bmatrix} \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{11}^*} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{1n}^*} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{m1}^*} & \dots & \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial Z_{mn}^*} \end{bmatrix}$$

若令

$$\mathbf{A} = \mathbf{D}_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*) \quad \text{和} \quad \mathbf{B} = \mathbf{D}_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*)$$

则

$$\begin{aligned} \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^T(\mathbf{Z})} &= \text{rvec}(\mathbf{D}_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*)) = \text{rvec}(\mathbf{A}) = \text{vec}^T(\mathbf{A}^T) \\ \frac{\partial f(\mathbf{Z}, \mathbf{Z}^*)}{\partial \text{vec}^H(\mathbf{Z})} &= \text{rvec}(\mathbf{D}_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*)) = \text{rvec}(\mathbf{B}) = \text{vec}^T(\mathbf{B}^T) \end{aligned}$$

于是, 式(11)可以写成

$$\mathrm{d}f(\mathbf{Z}, \mathbf{Z}^*) = \text{vec}^T(\mathbf{A}^T) \text{vec}(\mathrm{d}\mathbf{Z}) + \text{vec}^T(\mathbf{B}^T) \text{vec}(\mathrm{d}\mathbf{Z}^*) \quad (12)$$

利用 $\text{tr}(\mathbf{C}^T \mathbf{D}) = \text{vec}^T(\mathbf{C}) \text{vec}(\mathbf{D})$, 式(12)可以用迹函数形式表示为

$$\mathrm{d}f(\mathbf{Z}, \mathbf{Z}^*) = \text{tr}(\mathbf{A} \mathrm{d}\mathbf{Z} + \mathbf{B} \mathrm{d}\mathbf{Z}^*) \quad (13)$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 65 页 共 72 页

返回

全屏显示

关闭

退出



命题 给定一复变函数 $f(\mathbf{Z}, \mathbf{Z}^*) \in \mathbb{C}$, 其中 $\mathbf{Z} \in \mathbb{C}^{m \times n}$, 则

$$df(\mathbf{Z}, \mathbf{Z}^*) = \text{tr}(\mathbf{A}d\mathbf{Z} + \mathbf{B}d\mathbf{Z}^*) \Leftrightarrow \begin{cases} \nabla_{\mathbf{Z}} f(\mathbf{Z}, \mathbf{Z}^*) = \mathbf{A}^T \\ \nabla_{\mathbf{Z}^*} f(\mathbf{Z}, \mathbf{Z}^*) = \mathbf{B}^T \end{cases}$$

即复变函数的梯度矩阵和共轭梯度矩阵分别由矩阵 \mathbf{A} 和 \mathbf{B} 的转置唯一辨识。

注意：在矩阵微分结果 $df(\mathbf{X}, \mathbf{X}^*)$ 中，必须将 $d\mathbf{X}$ 和 $d\mathbf{X}^*$ 分别写在有关项的最右边，并称之为复矩阵微分的标准形式。

访问清华主页

标题页

◀ ▶

◀ ▶

第 66 页 共 72 页

返回

全屏显示

关闭

退出

3.6 复矩阵微分计算

迹函数 $\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^* \mathbf{B})$ 的复矩阵微分

$$\begin{aligned} d[\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^* \mathbf{B})] &= \text{tr}(d\mathbf{X} \cdot \mathbf{A} \mathbf{X}^* \mathbf{B}) + \text{tr}(\mathbf{X} \mathbf{A} d\mathbf{X}^* \mathbf{B}) \\ &= \text{tr}(\mathbf{A} \mathbf{X}^* \mathbf{B} d\mathbf{X}) + \text{tr}(\mathbf{B} \mathbf{X} \mathbf{A} d\mathbf{X}^*) \end{aligned}$$

由此得迹函数 $\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^* \mathbf{B})$ 的梯度矩阵和共轭梯度矩阵分别为

$$\begin{aligned} \nabla_{\mathbf{X}} \text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^* \mathbf{B}) &= (\mathbf{A} \mathbf{X}^* \mathbf{B})^T = \mathbf{B}^T \mathbf{X}^H \mathbf{A}^T \\ \nabla_{\mathbf{X}^*} \text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^* \mathbf{B}) &= (\mathbf{B} \mathbf{X} \mathbf{A})^T = \mathbf{A}^T \mathbf{X}^T \mathbf{B}^T \end{aligned}$$



访问清华主页

标题页

◀ ▶

◀ ▶

第 67 页 共 72 页

返回

全屏显示

关闭

退出



类似地，迹函数 $\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B})$ 的复矩阵微分为

$$\begin{aligned} \text{d}[\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B})] &= \text{tr}(\mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B} \text{d}\mathbf{X}) + \text{tr}(\mathbf{X} \mathbf{A} \text{d}\mathbf{X}^{\text{H}} \cdot \mathbf{B}) \\ &= \text{tr}(\mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B} \text{d}\mathbf{X}) + \text{tr}[(\mathbf{B} \mathbf{X} \mathbf{A})^{\text{T}} \text{d}\mathbf{X}^*] \end{aligned}$$

故迹函数 $\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B})$ 的梯度矩阵和共轭梯度矩阵分别为

$$\nabla_{\mathbf{X}} \text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B}) = (\mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B})^{\text{T}} = \mathbf{B}^{\text{T}} \mathbf{X}^* \mathbf{A}^{\text{T}}$$

$$\nabla_{\mathbf{X}^*} \text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^{\text{H}} \mathbf{B}) = [(\mathbf{B} \mathbf{X} \mathbf{A})^{\text{T}}]^{\text{T}} = \mathbf{B} \mathbf{X} \mathbf{A}$$

访问清华主页

标题页

◀ ▶

◀ ▶

第 68 页 共 72 页

返回

全屏显示

关闭

退出

行列式 $|\mathbf{X}\mathbf{X}^*|$ 和 $|\mathbf{X}\mathbf{X}^H|$ 的复矩阵微分分别为

$$\begin{aligned} d|\mathbf{X}\mathbf{X}^*| &= |\mathbf{X}\mathbf{X}^*| \text{tr}[(\mathbf{X}\mathbf{X}^*)^{-1} d(\mathbf{X}\mathbf{X}^*)] \\ &= |\mathbf{X}\mathbf{X}^*| \text{tr}[(\mathbf{X}\mathbf{X}^*)^{-1} d\mathbf{X} \cdot \mathbf{X}^*] \\ &\quad + |\mathbf{X}\mathbf{X}^*| \text{tr}[(\mathbf{X}\mathbf{X}^*)^{-1} \mathbf{X} d\mathbf{X}^*] \\ &= |\mathbf{X}\mathbf{X}^*| \text{tr}[\mathbf{X}^* (\mathbf{X}\mathbf{X}^*)^{-1} d\mathbf{X}] \\ &\quad + |\mathbf{X}\mathbf{X}^*| \text{tr}[(\mathbf{X}\mathbf{X}^*)^{-1} \mathbf{X} d\mathbf{X}^*] \end{aligned}$$

和

$$\begin{aligned} d|\mathbf{X}\mathbf{X}^H| &= |\mathbf{X}\mathbf{X}^H| \text{tr}[(\mathbf{X}\mathbf{X}^H)^{-1} d(\mathbf{X}\mathbf{X}^H)] \\ &= |\mathbf{X}\mathbf{X}^H| \text{tr}[(\mathbf{X}\mathbf{X}^H)^{-1} d\mathbf{X} \cdot \mathbf{X}^H] \\ &\quad + |\mathbf{X}\mathbf{X}^H| \text{tr}[(\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X} d\mathbf{X}^H] \\ &= |\mathbf{X}\mathbf{X}^H| \text{tr}[\mathbf{X}^H (\mathbf{X}\mathbf{X}^H)^{-1} d\mathbf{X}] \\ &\quad + |\mathbf{X}\mathbf{X}^H| \text{tr}\{[(\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X}]^T d\mathbf{X}^*\} \end{aligned}$$



访问清华主页

标题页

◀

▶

◀

▶

第 69 页 共 72 页

返回

全屏显示

关闭

退出



几种迹函数的梯度矩阵与共轭梯度矩阵

$f(X, X^*)$	梯度矩阵 $\partial f(X, X^*)/\partial X$	共轭梯度矩阵 $\partial f(X, X^*)/\partial X^*$
$\text{tr}(AX)$	A^T	O
$\text{tr}(AX^H)$	O	A
$\text{tr}(XAX^TB)$	$B^TAX^T + BXA$	O
$\text{tr}(XAX^*B)$	$B^TX^HA^T$	$A^TX^TB^T$
$\text{tr}(X^*AXB)$	$A^TX^HB^T$	$B^TX^TA^T$
$\text{tr}(XAX^HB)$	$B^TX^*A^T$	BXA
$\text{tr}(X^HAXB)$	$A^TX^*B^T$	AXB

访问清华主页

标题页

◀ ▶

◀ ▶

第 70 页 共 72 页

返回

全屏显示

关闭

退出



几种行列式函数的梯度矩阵与共轭梯度矩阵

$f(X, X^*)$	梯度矩阵 $\partial f(X, X^*)/\partial X$	共轭梯度矩阵 $\partial f(X, X^*)/\partial X^*$
$ X $	$ X X^{-T}$	O
$ XX^T $	$2 XX^T (XX^T)^{-1}X$	O
$ X^TX $	$2 X^TX X(X^TX)^{-1}$	O
$ XX^* $	$ XX^* (X^HX^T)^{-1}X^H$	$ XX^* X^T(X^HX^T)^{-1}$
$ X^*X $	$ X^*X X^H(X^TX^H)^{-1}$	$ X^*X (X^TX^H)^{-1}X^T$
$ XX^H $	$ XX^H (X^*X^T)^{-1}X^*$	$ XX^H (XX^H)^{-1}X$
$ X^HX $	$ X^HX X^*(X^TX^*)^{-1}$	$ X^HX X(X^HX)^{-1}$

访问清华主页

标题页

◀ ▶

◀ ▶

第 71 页 共 72 页

返回

全屏显示

关闭

退出



一句话概括：矩阵微分是一种重要的数学工具，在流形计算、几何物理、微分几何、经济计量(econometrics)和最优化中有着广泛的应用！

习题1：求迹函数 $\text{tr}[(\mathbf{A} + \mathbf{X}^T \mathbf{C} \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{B} \mathbf{X})]$ 的微分矩阵和梯度矩阵。

习题2：计算复矩阵微分

$$d[\text{tr}(\mathbf{X} \mathbf{A} \mathbf{X}^H \mathbf{B})] \quad \text{和} \quad d|\mathbf{X} \mathbf{X}^H|$$

附加题：求复变函数 $f(\mathbf{X}, \mathbf{X}^*) = |\mathbf{X} \mathbf{A} \mathbf{X}^H \mathbf{B}|$ 的梯度矩阵 $\nabla_{\mathbf{X}} f(\mathbf{X}, \mathbf{X}^*) = \frac{\partial f(\mathbf{X}, \mathbf{X}^*)}{\partial \mathbf{X}}$ 和共轭梯度矩阵 $\nabla_{\mathbf{X}^*} f(\mathbf{X}, \mathbf{X}^*) = \frac{\partial f(\mathbf{X}, \mathbf{X}^*)}{\partial \mathbf{X}^*}$ 。

访问清华主页

标题页

◀ ▶

◀ ▶

第 72 页 共 72 页

返回

全屏显示

关闭

退出