清華大學

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

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

$$\text{实值向量偏导}\begin{cases}\text{实值行向量偏导}\\ \text{实值列向量偏导}\end{cases}$$

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

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

实值标量函数$f(\boldsymbol{x})$的行向量偏导

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

若变元为实值矩阵$\boldsymbol{X}\in\mathbb{R}^{m\times n}$，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：

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

式中$\mathrm{vec}^{\mathrm{T}}(\boldsymbol{X}) = [\mathrm{vec}(\boldsymbol{X})]^{\mathrm{T}}$。

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

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

则有

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

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

比较$\mathrm{D}_{\mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})} f(\boldsymbol{X})$和$\mathrm{D}_{\boldsymbol{X}} f(\boldsymbol{X})$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$f(\boldsymbol{X})$，其中$\boldsymbol{X} \in \mathbb{R}^{m\times n}$，则

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

或

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

式中$\mathrm{unrvec}(\cdot)$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。

$m \times 1$列向量偏导算子定义为

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

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

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

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

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

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$f(\boldsymbol{X})$关于矩阵变元$\boldsymbol{X}$的梯度矩阵的列向量形式：

$$\operatorname{vec}(\nabla_{\boldsymbol{X}} f(\boldsymbol{X})) = \frac{\partial f(\boldsymbol{X})}{\partial \operatorname{vec}(\boldsymbol{X})}$$

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

将这一列向量结果矩阵化，易知

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

比较式(2)和式(1)，立即有

$$\nabla_X f(\boldsymbol{X}) = \mathrm{D}_X^{\mathrm{T}} f(\boldsymbol{X})$$

即是说，实值标量函数$f(\boldsymbol{X})$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。

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$\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}$和$\frac{\partial}{\partial \boldsymbol{X}^{\mathrm{T}}}$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$\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) = \mathrm{D}_{\boldsymbol{X}}^{\mathrm{T}} f(\boldsymbol{X})$及命题1立即有以下结果。

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

$$\nabla_{\boldsymbol{X}} f(\boldsymbol{X}) = \mathrm{unvec}\left(\mathrm{D}_{\mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})}^{\mathrm{T}} f(\boldsymbol{X})\right) \tag{3}$$

即是说，实值标量函数$f(\boldsymbol{X})$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$\mathrm{D}_{\mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})} f(\boldsymbol{X}) = [d_1, \cdots, d_{mn}]$，则梯度矩阵第$(i,j)$个元素

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

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 $\boldsymbol{x}$ 的梯度流(gradient flow)，记作

$$\dot{\boldsymbol{x}} = -\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$$

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 $f(\boldsymbol{x})$ 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 $f(\boldsymbol{x})$ 的最大减小率。这是梯度下降法的基础。

2. 实值向量函数的协梯度矩阵

对$p\times 1$实值向量函数$\boldsymbol{f}(\boldsymbol{x})=[f_1(\boldsymbol{x}),\cdots,f_p(\boldsymbol{x})]^{\mathrm{T}}$的元素$f_i(\boldsymbol{x}), i=1,\cdots,p$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：

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

并称之为向量函数$\boldsymbol{f}(\boldsymbol{x})$在$\boldsymbol{x}$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$\boldsymbol{f}(\boldsymbol{x})$的第$i$个分量$f_i(\boldsymbol{x})$相对于向量变元$\boldsymbol{x}$的第$j$个元素的偏导，即

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

3. 实值矩阵函数的协梯度矩阵

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

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：

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

于是，矩阵函数$\boldsymbol{F}(\boldsymbol{X})$的行向量偏导定义为

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

其具体表达式为

$$
\begin{aligned}
&\mathrm{D}_{\mathrm{vec}^{\mathrm{T}}(\boldsymbol{x})}\boldsymbol{F}(\boldsymbol{X}) \\
&= \left[\frac{\partial F_{11}}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})}, \cdots, \frac{\partial F_{p1}}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})}, \cdots, \frac{\partial F_{1q}}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})}, \cdots, \frac{\partial F_{pq}}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{X})}\right]^{\mathrm{T}} \\
&= \begin{bmatrix}
\frac{\partial F_{11}}{\partial X_{11}} & \cdots & \frac{\partial F_{11}}{\partial X_{m1}} & \cdots & \frac{\partial F_{11}}{\partial X_{1n}} & \cdots & \frac{\partial F_{11}}{\partial X_{mn}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial F_{p1}}{\partial X_{11}} & \cdots & \frac{\partial F_{p1}}{\partial X_{m1}} & \cdots & \frac{\partial F_{p1}}{\partial X_{1n}} & \cdots & \frac{\partial F_{p1}}{\partial X_{mn}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial F_{1q}}{\partial X_{11}} & \cdots & \frac{\partial F_{1q}}{\partial X_{m1}} & \cdots & \frac{\partial F_{1q}}{\partial X_{1n}} & \cdots & \frac{\partial F_{1q}}{\partial X_{mn}} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial F_{pq}}{\partial X_{11}} & \cdots & \frac{\partial F_{pq}}{\partial X_{m1}} & \cdots & \frac{\partial F_{pq}}{\partial X_{1n}} & \cdots & \frac{\partial F_{pq}}{\partial X_{mn}}
\end{bmatrix}
\end{aligned}
$$

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

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

或记作

$$\nabla_{\boldsymbol{x}}^2 f(\boldsymbol{x}) = \mathrm{D}_{\boldsymbol{x}}(\nabla_{\boldsymbol{x}} f(\boldsymbol{x})) = \mathrm{D}_{\boldsymbol{x}}\boldsymbol{g}(\boldsymbol{x})$$

即实值标量函数$f(\boldsymbol{x})$的Hessian矩阵是梯度向量函数$\boldsymbol{g}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} f(\boldsymbol{x})$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$f(\boldsymbol{X})$相对于$m \times n$实矩阵$\boldsymbol{X}$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

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

或记作

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

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

求Hessian矩阵的二步法：

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

3.2 实值标量函数的矩阵微分

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

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

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

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

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

$$A_1\Delta x_1+\cdots+A_m\Delta x_m=\frac{\partial f}{\partial x_1}\mathrm{d}x_1+\cdots+\frac{\partial f}{\partial x_m}\mathrm{d}x_m$$

称为多变量函数$f(x_1,\cdots,x_m)$的全微分，记为

$$\mathrm{d}f(x_1,\cdots,x_m)=\frac{\partial f}{\partial x_1}\mathrm{d}x_1+\cdots+\frac{\partial f}{\partial x_m}\mathrm{d}x_m \qquad (5)$$

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。

1.实矩阵微分

考虑实值标量函数$f(\boldsymbol{x})$，其变元$\boldsymbol{x} = [x_1, \cdots, x_m]^{\mathrm{T}} \in \mathbb{R}^m$。将变元向量的元素$x_1, \cdots, x_m$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$f(\boldsymbol{x})$的全微分表达式：

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

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

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

$$
\mathrm{d}\boldsymbol{x} = [\mathrm{d}x_1, \cdots, \mathrm{d}x_m]^{\mathrm{T}}
$$

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

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

借助向量化函数，上式可以简写为

$$\mathrm{d}f(\boldsymbol{X}) = \mathrm{rvec}(\boldsymbol{A})\mathrm{vec}(\mathrm{d}\boldsymbol{X}) \tag{6}$$

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

$$\mathrm{d}\boldsymbol{X} = \begin{bmatrix} \mathrm{d}X_{11} & \cdots & \mathrm{d}X_{1n} \\ \vdots & \ddots & \vdots \\ \mathrm{d}X_{m1} & \cdots & \mathrm{d}X_{mn} \end{bmatrix} \tag{7}$$

以及

$$\boldsymbol{A} = \mathrm{D}_x f(\boldsymbol{X}) = \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{X})}{\partial X_{11}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial X_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{X})}{\partial X_{1n}} & \cdots & \frac{\partial f(\boldsymbol{X})}{\partial X_{mn}} \end{bmatrix} \tag{8}$$

是实标量函数$f(\boldsymbol{X})$在实矩阵点$\boldsymbol{X}$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$\mathrm{rvec}(\boldsymbol{A}) = (\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}}))^{\mathrm{T}}$，式(6)可以改写为$\mathrm{d}f(\boldsymbol{X}) = (\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}}))^{\mathrm{T}}\mathrm{vec}(\mathrm{d}\boldsymbol{X})$。然后，在向量化算子vec与迹函数之间的关系式$\mathrm{tr}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C}) = (\mathrm{vec}(\boldsymbol{B}))^{\mathrm{T}}\mathrm{vec}(\boldsymbol{C})$中，令$\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}$和$\boldsymbol{C} = \mathrm{d}\boldsymbol{X}$，又可将$\mathrm{d}f(\boldsymbol{X}) = (\mathrm{vec}(\boldsymbol{A}^{\mathrm{T}}))^{\mathrm{T}}\mathrm{vec}(\mathrm{d}\boldsymbol{X})$最终表示为

$$\mathrm{d}f(\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{X}) \tag{9}$$

从上述讨论，可以引出如下定义：

(1) 称实值标量函数$f(\boldsymbol{X})$在实矩阵点$\boldsymbol{X}$是可微分的(differentiable)；

(2) 称线性主部的实矩阵$\boldsymbol{A}$为实值标量函数$f(\boldsymbol{X})$相对于变元矩阵$\boldsymbol{X} \in \mathbb{R}^{m\times n}$的一阶偏导矩阵即Jacobian矩阵。

一阶偏导矩阵A是唯一确定的，即：若存在$\boldsymbol{A}_1$和$\boldsymbol{A}_2$满足$\mathrm{d}f(\boldsymbol{X}) = \boldsymbol{A}_i \mathrm{d}\boldsymbol{X}, i = 1, 2$，则$\boldsymbol{A}_1 = \boldsymbol{A}_2$。

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

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

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$f(\boldsymbol{X})$在$m\times n$矩阵点$\boldsymbol{X}$ 可微分，则

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

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

(1) 求实值目标函数$f(\boldsymbol{X})$相对于变元矩阵$\boldsymbol{X}$的矩阵微分$\mathrm{d}f(\boldsymbol{X})$，并将其表示成规范形式$\mathrm{d}f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\mathrm{d}\boldsymbol{X})$；

(2) 实值目标函数$f(\boldsymbol{X})$相对于$m\times n$变元矩阵$\boldsymbol{X}$的梯度矩阵由$\nabla f(\boldsymbol{X})=\frac{\partial f(\boldsymbol{x})}{\partial\boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}$直接给出。

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$\boldsymbol{H}$	$\boldsymbol{H}$的维数
$f(x)$	$\mathrm{d}^2[f(x)] = \beta(\mathrm{d}x)^2$	$\boldsymbol{H}[f(x)] = \beta$	1×1
$f(\boldsymbol{x})$	$\mathrm{d}^2[f(\boldsymbol{x})] = (\mathrm{d}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}\mathrm{d}\boldsymbol{x}$	$\boldsymbol{H}[f(\boldsymbol{x})] = \frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}})$	$m \times m$
$f(\boldsymbol{X})$	$\mathrm{d}^2[f(\boldsymbol{X})] = (\mathrm{d}\,\mathrm{vec}(\boldsymbol{X}))^{\mathrm{T}}\boldsymbol{B}\mathrm{d}(\mathrm{vec}(\boldsymbol{X}))$	$\boldsymbol{H}[f(\boldsymbol{X})] = \frac{1}{2}(\boldsymbol{B} + \boldsymbol{B}^{\mathrm{T}})$	$mn \times mn$
$\boldsymbol{f}(x)$	$\mathrm{d}^2[\boldsymbol{f}(x)] = \boldsymbol{b}(\mathrm{d}x)^2$	$\boldsymbol{H}[\boldsymbol{f}(x)] = \boldsymbol{b}$	$p \times 1$
$\boldsymbol{f}(\boldsymbol{x})$	$\mathrm{d}^2[\boldsymbol{f}(\boldsymbol{x})] = (\boldsymbol{I}_m \otimes \mathrm{d}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}\mathrm{d}\boldsymbol{x}$	$\boldsymbol{H}[\boldsymbol{f}(\boldsymbol{x})] = \frac{1}{2}[\boldsymbol{B} + (\boldsymbol{B}')_v]$	$pm \times m$
$\boldsymbol{f}(\boldsymbol{X})$	$\mathrm{d}^2[\boldsymbol{f}(\boldsymbol{X})] = (\boldsymbol{I}_m \otimes \mathrm{d}\,\mathrm{vec}(\boldsymbol{X}))^{\mathrm{T}}\boldsymbol{B}\mathrm{d}(\mathrm{vec}(\boldsymbol{X}))$	$\boldsymbol{H}[\boldsymbol{f}(\boldsymbol{X})] = \frac{1}{2}[\boldsymbol{B} + (\boldsymbol{B}')_v]$	$pmn \times mn$
$\boldsymbol{F}(x)$	$\mathrm{d}^2[\boldsymbol{F}(x)] = \boldsymbol{B}(\mathrm{d}x)^2$	$\boldsymbol{H}[\boldsymbol{F}(x)] = \mathrm{vec}(\boldsymbol{B})$	$pq \times 1$
$\boldsymbol{F}(\boldsymbol{x})$	$\mathrm{d}^2[\mathrm{vec}(\boldsymbol{F})] = (\boldsymbol{I}_{mp} \otimes \mathrm{d}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{B}\mathrm{d}\boldsymbol{x}$	$\boldsymbol{H}[\boldsymbol{F}(\boldsymbol{x})] = \frac{1}{2}[\boldsymbol{B} + (\boldsymbol{B}')_v]$	$pmq \times m$
$\boldsymbol{F}(\boldsymbol{X})$	$\mathrm{d}^2[\mathrm{vec}(\boldsymbol{F})] = (\boldsymbol{I}_{mp} \otimes \mathrm{d}\,\mathrm{vec}(\boldsymbol{X}))^{\mathrm{T}}\boldsymbol{B}\mathrm{d}(\mathrm{vec}(\boldsymbol{X}))$	$\boldsymbol{H}[\boldsymbol{F}(\boldsymbol{x})] = \frac{1}{2}[\boldsymbol{B} + (\boldsymbol{B}')_v]$	$pmqn \times mn$

表中，对于实向量函数f，

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

而对于实矩阵函数F，

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

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$(\mathrm{d}\boldsymbol{X})_{ij}$，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$\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}) = (\mathrm{d}\boldsymbol{X})^{\mathrm{T}}$。

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

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

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

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

例2　考虑矩阵乘积UV的微分矩阵，有

$$
\begin{aligned}
[\mathrm{d}(\boldsymbol{U}\boldsymbol{V})]_{ij} &= \mathrm{d}\left([\boldsymbol{U}\boldsymbol{V}]_{ij}\right) \\
&= \mathrm{d}\left(\sum_k u_{ik}v_{kj}\right) \\
&= \sum_k \mathrm{d}(u_{ik}v_{kj}) \\
&= \sum_k \left[(\mathrm{d}u_{ik})v_{kj} + u_{ik}\mathrm{d}v_{kj}\right] \\
&= \sum_k (\mathrm{d}u_{ik})v_{kj} + \sum_k u_{ik}\mathrm{d}v_{kj} \\
&= [(\mathrm{d}\boldsymbol{U})\boldsymbol{V}]_{ij} + [\boldsymbol{U}\mathrm{d}\boldsymbol{V}]_{ij}
\end{aligned}
$$

从而得$\mathrm{d}(\boldsymbol{U}\boldsymbol{V}) = (\mathrm{d}\boldsymbol{U})\boldsymbol{V} + \boldsymbol{U}\mathrm{d}\boldsymbol{V}$。

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$\mathrm{d}\boldsymbol{A}=\boldsymbol{O}$。

(2) 常数α与矩阵$\boldsymbol{X}$的乘积的微分矩阵$\mathrm{d}(\alpha\boldsymbol{X})=\alpha\mathrm{d}\boldsymbol{X}$。

(3) 矩阵转置的微分矩阵等于原矩阵的微分矩阵的转置，即$\mathrm{d}(\boldsymbol{X}^{\mathrm{T}})=(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}$。

(4) 两个矩阵函数的和(差)的微分矩阵为$\mathrm{d}(\boldsymbol{U}\pm\boldsymbol{V})=\mathrm{d}\boldsymbol{U}\pm\mathrm{d}\boldsymbol{V}$。

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

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

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

$$\mathrm{d}(\boldsymbol{U}\boldsymbol{V}\boldsymbol{W}) = (\mathrm{d}\boldsymbol{U})\boldsymbol{V}\boldsymbol{W} + \boldsymbol{U}(\mathrm{d}\boldsymbol{V})\boldsymbol{W} + \boldsymbol{U}\boldsymbol{V}(\mathrm{d}\boldsymbol{W})$$

特别地，若$\boldsymbol{A}$为常数矩阵，则

$$\mathrm{d}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}) = (\mathrm{d}\boldsymbol{X})\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{A}(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}$$

和

$$\mathrm{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}) = (\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\mathrm{d}\boldsymbol{X}$$

(7) 矩阵函数的Kronecker积的微分矩阵为

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

(8) 矩阵函数的Hadamard积的微分矩阵为

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

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

$$\mathrm{d}(\mathrm{vec}(\boldsymbol{X})) = \mathrm{vec}(\mathrm{d}\boldsymbol{X})$$

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

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

特别地，矩阵函数$\boldsymbol{F}(\boldsymbol{X})$的迹的矩阵微分为

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

例如 $\mathrm{d}(\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})) = 2\mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})$。

(11) 行列式的微分为

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

特别地，矩阵函数的行列式的微分为

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

(12) 矩阵对数的微分矩阵为

$$\mathrm{d}\log \boldsymbol{X} = \boldsymbol{X}^{-1}\mathrm{d}\boldsymbol{X}$$

特别地，矩阵函数的对数的微分矩阵为

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

(13) 逆矩阵的微分矩阵为

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

(14) Moore-Penrose逆矩阵的微分矩阵为

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

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

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

迹函数的梯度矩阵

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

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

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

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

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

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

从而得梯度矩阵

$$
\frac{\partial \mathrm{tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X})}{\partial \boldsymbol{X}} = \left[\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A})\right]^{\mathrm{T}} = (\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\boldsymbol{X}
$$

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$\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})$。计算得

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

由此得梯度矩阵

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

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

访问清华主页
标题页
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$f(\boldsymbol{X})$总可以写成迹函数的形式，因为$f(\boldsymbol{X}) = \mathrm{tr}(f(\boldsymbol{X}))$；

(2) 无论$\mathrm{d}\boldsymbol{X}$出现在迹函数内的任何位置，总可以通过迹函数的性质$\mathrm{tr}[\boldsymbol{A}(\mathrm{d}\boldsymbol{X})\boldsymbol{B}] = \mathrm{tr}(\boldsymbol{B}\boldsymbol{A}\mathrm{d}\boldsymbol{X})$，将$\mathrm{d}\boldsymbol{X}$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$(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}$，总可以通过迹函数的性质

$$\mathrm{tr}[\boldsymbol{A}(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{B}] = \mathrm{tr}[(\boldsymbol{A}(\mathrm{d}\boldsymbol{X})^{\mathrm{T}}\boldsymbol{B})^{\mathrm{T}}] = \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\mathrm{d}\boldsymbol{X})$$

写成迹函数微分矩阵的规范形式。

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

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

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：

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

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。

Cauchy-Riemann条件的一个直接结果是：函数$f(z) = u(x, y) + \mathrm{j}v(x, y)$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$f(z) = u(x,y) + \mathrm{j}v(x,y)$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$f(z) = z^* = x - \mathrm{j}y = u(x,y) + \mathrm{j}v(x,y)$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$f(z) = \mathrm{Re}(z) = x$和$f(z) = \mathrm{Im}(z) = y$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$f(z) = u(x,y) + \mathrm{j}v(x,y)$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$f(z) = |z|^2 = x^2 + y^2 = u(x,y) + \mathrm{j}v(x,y)$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：

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

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

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

由实际偏导的定义及上述独立性假设，容易求出

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

$$=\frac{1}{2}(1+0)+\mathrm{j}\frac{1}{2}(0+\mathrm{j})$$

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

$$=\frac{1}{2}(1-0)-\mathrm{j}\frac{1}{2}(0-\mathrm{j})$$

即有

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

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$z = x + \mathrm{j}y$) 使用实(极)坐标$r \triangleq (x, y)^{\mathrm{T}}$表示为$f(r) = f(x, y)$，而在复导数的框架内，则使用共轭坐标$c \triangleq (z, z^*)^{\mathrm{T}}$替代实坐标$r = (x, y)^{\mathrm{T}}$，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$z=x+\mathrm{j}y$全平面是解析的，并且对于固定的z值，复变函数$f(z,z^*)$在$z^*=x-\mathrm{j}y$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:

$$\mathrm{d}f(z,z^*) = \frac{\partial f(z,z^*)}{\partial z}\mathrm{d}z + \frac{\partial f(z,z^*)}{\partial z^*}\mathrm{d}z^*$$

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$\boldsymbol{z}=\boldsymbol{x}+\mathrm{j}\boldsymbol{y}=[z_1,\cdots,z_m]^{\mathrm{T}}\in\mathbb{C}^m$，其中$\boldsymbol{x}=[x_1,\cdots,x_m]^{\mathrm{T}}\in\mathbb{R}^m,\boldsymbol{y}=[y_1,\cdots,y_m]^{\mathrm{T}}\in\mathbb{R}^m$，即$z_i=x_i+\mathrm{j}y_i,i=1,\cdots,m$，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$\boldsymbol{z}\in\mathbb{C}^m$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

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

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$\boldsymbol{z}^{\mathrm{T}} = [z_1, \cdots, z_m]$的每个元素运用复变函数的偏导算子

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

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

$$\mathrm{D}_{\boldsymbol{z}} = \frac{\partial}{\partial \boldsymbol{z}^{\mathrm{T}}} = \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}} - \mathrm{j}\frac{\partial}{\partial \boldsymbol{y}^{\mathrm{T}}}\right)$$

和共轭协梯度算子

$$\mathrm{D}_{\boldsymbol{z}^*} = \frac{\partial}{\partial \boldsymbol{z}^{\mathrm{H}}} = \frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}} + \mathrm{j}\frac{\partial}{\partial \boldsymbol{y}^{\mathrm{T}}}\right)$$

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

$$\nabla_{\boldsymbol{z}} = \frac{\partial}{\partial \boldsymbol{z}} \triangleq \left[\frac{\partial}{\partial z_1}, \cdots, \frac{\partial}{\partial z_m}\right]^{\mathrm{T}}, \quad \nabla_{\boldsymbol{z}^*} = \frac{\partial}{\partial \boldsymbol{z}^*} \triangleq \left[\frac{\partial}{\partial z_1^*}, \cdots, \frac{\partial}{\partial z_m^*}\right]^{\mathrm{T}}$$

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$\boldsymbol{z} = [z_1, \cdots, z_m]^{\mathrm{T}}$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$\boldsymbol{z}$的实部$\boldsymbol{x}$与虚部$\boldsymbol{y}$表示的梯度算子

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

和共轭梯度算子

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

利用梯度算子和共轭梯度算子的定义公式，不难求出

$$\begin{aligned}\frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{z}} &= \frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{z}} + \mathrm{j}\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{z}} = \frac{1}{2}\left(\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{x}} - \mathrm{j}\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{y}}\right) + \mathrm{j}\frac{1}{2}\left(\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}} - \mathrm{j}\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{y}}\right)\\ &= \boldsymbol{I}_{m\times m}\\ \frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{z}^*} &= \frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{z}^*} + \mathrm{j}\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{z}^*} = \frac{1}{2}\left(\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{x}} + \mathrm{j}\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{y}}\right) + \mathrm{j}\frac{1}{2}\left(\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{x}} + \mathrm{j}\frac{\partial \boldsymbol{y}^{\mathrm{T}}}{\partial \boldsymbol{y}}\right)\\ &= \boldsymbol{O}_{m\times m}\end{aligned}$$

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：

$$\frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{z}} = \boldsymbol{I}, \quad \frac{\partial \boldsymbol{z}^{\mathrm{H}}}{\partial \boldsymbol{z}^{*}} = \boldsymbol{I}, \quad \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{z}^{\mathrm{T}}} = \boldsymbol{I}, \quad \frac{\partial \boldsymbol{z}^{*}}{\partial \boldsymbol{z}^{\mathrm{H}}} = \boldsymbol{I}$$

$$\frac{\partial \boldsymbol{z}^{\mathrm{T}}}{\partial \boldsymbol{z}^{*}} = \boldsymbol{O}, \quad \frac{\partial \boldsymbol{z}^{\mathrm{H}}}{\partial \boldsymbol{z}} = \boldsymbol{O}, \quad \frac{\partial \boldsymbol{z}}{\partial \boldsymbol{z}^{\mathrm{H}}} = \boldsymbol{O}, \quad \frac{\partial \boldsymbol{z}^{*}}{\partial \boldsymbol{z}^{\mathrm{T}}} = \boldsymbol{O}$$

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$\frac{\partial}{\partial \boldsymbol{z}^{\mathrm{T}}}$还是梯度算子$\frac{\partial}{\partial \boldsymbol{z}}$，复共轭变元向量$\boldsymbol{z}^{*}$均可视为一常数向量；

2.无论使用共轭协梯度算子$\frac{\partial}{\partial \boldsymbol{z}^{\mathrm{H}}}$还是共轭梯度算子$\frac{\partial}{\partial \boldsymbol{z}^{*}}$，向量$\boldsymbol{z}$ 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$\boldsymbol{z}$和$\boldsymbol{z}^*$ 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$f(\boldsymbol{z},\boldsymbol{z}^*) \in \mathbb{R}$，其中$\boldsymbol{z} \in \mathbb{C}^m$。

协梯度向量、共轭协梯度向量的定义公式：

$$\mathrm{D}_{\boldsymbol{z}} f(\boldsymbol{z},\boldsymbol{z}^*) = \left.\frac{\partial f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial \boldsymbol{z}^{\mathrm{T}}}\right|_{\boldsymbol{z}^*=\text{常数向量}}$$

$$\mathrm{D}_{\boldsymbol{z}^*} f(\boldsymbol{z},\boldsymbol{z}^*) = \left.\frac{\partial f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial \boldsymbol{z}^{\mathrm{H}}}\right|_{\boldsymbol{z}=\text{常数向量}}$$

梯度向量、共轭梯度向量的定义公式：

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

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

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$f(\boldsymbol{Z},\boldsymbol{Z}^*)$的梯度矩阵和共轭梯度矩阵分别定义为

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

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

梯度矩阵和共轭梯度矩阵的计算公式：

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

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

Hessian矩阵(共轭梯度的协梯度)

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

下面是梯度矩阵的运算法则。

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

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

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

3.乘积法则：

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

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

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

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

$$\frac{\partial H(Z,Z^*)}{\partial Z}=\frac{\partial G(F(Z,Z^*),F^*(Z,Z^*))}{\partial F(Z,Z^*)}\cdot\frac{\partial F(Z,Z^*)}{\partial Z}+\frac{\partial G(F(Z,Z^*),F^*(Z,Z^*))}{\partial F^*(Z,Z^*)}\cdot\frac{\partial F^*(Z,Z^*)}{\partial Z}$$

$$\frac{\partial H(Z,Z^*)}{\partial Z^*}=\frac{\partial G(F(Z,Z^*),F^*(Z,Z^*))}{\partial F(Z,Z^*)}\cdot\frac{\partial F(Z,Z^*)}{\partial Z^*}+\frac{\partial G(F(Z,Z^*),F^*(Z,Z^*))}{\partial F^*(Z,Z^*)}\cdot\frac{\partial F^*(Z,Z^*)}{\partial Z^*}$$

特别地，若$h(Z,Z^*)=g(F(Z,Z^*),F^*(Z,Z^*))$，则

$$\frac{\partial h(Z,Z^*)}{\partial Z}=\frac{\partial g(F(Z,Z^*),F^*(Z,Z^*))}{\partial F(Z,Z^*)}\cdot\frac{\partial F(Z,Z^*)}{\partial Z}+\frac{\partial g(F(Z,Z^*),F^*(Z,Z^*))}{\partial F^*(Z,Z^*)}\cdot\frac{\partial F^*(Z,Z^*)}{\partial Z}$$

$$\frac{\partial h(Z,Z^*)}{\partial Z^*}=\frac{\partial g(F(Z,Z^*),F^*(Z,Z^*))}{\partial F(Z,Z^*)}\cdot\frac{\partial F(Z,Z^*)}{\partial Z^*}+\frac{\partial g(F(Z,Z^*),F^*(Z,Z^*))}{\partial F^*(Z,Z^*)}\cdot\frac{\partial F^*(Z,Z^*)}{\partial Z^*}$$

3.5 复矩阵微分

单变元复变函数的复微分法则

$$\mathrm{d}f(z,z^*)=\frac{\partial f(z,z^*)}{\partial z}\mathrm{d}z+\frac{\partial f(z,z^*)}{\partial z^*}\mathrm{d}z^*$$

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

$$\mathrm{d}f(\cdot)=\frac{\partial f(\cdot)}{\partial z_1}\mathrm{d}z_1+\cdots+\frac{\partial f(\cdot)}{\partial z_m}\mathrm{d}z_m+\frac{\partial f(\cdot)}{\partial z_1^*}\mathrm{d}z_1^*+\cdots+\frac{\partial f(\cdot)}{\partial z_m^*}\mathrm{d}z_m^*$$

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$\boldsymbol{z}=[z_1,\cdots,z_m]^{\mathrm{T}}$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

$$\mathrm{d}f(\boldsymbol{z},\boldsymbol{z}^*)=\left[\frac{f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial z_1},\cdots,\frac{f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial z_m}\right]\begin{bmatrix}\mathrm{d}z_1\\ \vdots\\ \mathrm{d}z_m\end{bmatrix}$$

$$+\left[\frac{f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial z_1^*},\cdots,\frac{f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial z_m^*}\right]\begin{bmatrix}\mathrm{d}z_1^*\\ \vdots\\ \mathrm{d}z_m^*\end{bmatrix}$$

$$=\frac{\partial f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial \boldsymbol{z}^{\mathrm{T}}}\mathrm{d}\boldsymbol{z}+\frac{\partial f(\boldsymbol{z},\boldsymbol{z}^*)}{\partial \boldsymbol{z}^{\mathrm{H}}}\mathrm{d}\boldsymbol{z}^*$$

或简记作

$$\mathrm{d}f(\boldsymbol{z},\boldsymbol{z}^*)=\mathrm{D}_{\boldsymbol{z}}f(\boldsymbol{z},\boldsymbol{z}^*)\mathrm{d}\boldsymbol{z}+\mathrm{D}_{\boldsymbol{z}^*}f(\boldsymbol{z},\boldsymbol{z}^*)\mathrm{d}\boldsymbol{z}^*$$

式中

$$\mathrm{d}\boldsymbol{z}=[\mathrm{d}z_1,\cdots,\mathrm{d}z_m]^{\mathrm{T}}$$

$$\mathrm{d}\boldsymbol{z}^*=[\mathrm{d}z_1^*,\cdots,\mathrm{d}z_m^*]^{\mathrm{T}}$$

$$\mathrm{D}_{\boldsymbol{z}} f(\boldsymbol{z}, \boldsymbol{z}^*) = \frac{\partial f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial \boldsymbol{z}^{\mathrm{T}}} = \left[\frac{f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial z_1}, \cdots, \frac{f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial z_m} \right] \in \mathbb{C}^{1\times m}$$

$$\mathrm{D}_{\boldsymbol{z}^*} f(\boldsymbol{z}, \boldsymbol{z}^*) = \frac{\partial f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial \boldsymbol{z}^{\mathrm{H}}} = \left[\frac{f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial z_1^*}, \cdots, \frac{f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial z_m^*} \right] \in \mathbb{C}^{1\times m}$$

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

$$\nabla_{\boldsymbol{z}} f(\boldsymbol{z}, \boldsymbol{z}^*) = \frac{\partial f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial \boldsymbol{z}} = (\mathrm{D}_{\boldsymbol{z}} f(\boldsymbol{z}, \boldsymbol{z}^*))^{\mathrm{T}}$$

$$\nabla_{\boldsymbol{z}^*} f(\boldsymbol{z}, \boldsymbol{z}^*) = \frac{\partial f(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial \boldsymbol{z}^*} = (\mathrm{D}_{\boldsymbol{z}^*} f(\boldsymbol{z}, \boldsymbol{z}^*))^{\mathrm{T}}$$

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

$$\mathrm{d}f(\boldsymbol{Z},\boldsymbol{Z}^*)=\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{Z})}\mathrm{d}(\mathrm{vec}\boldsymbol{Z})+\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{H}}(\boldsymbol{Z})}\mathrm{d}(\mathrm{vec}\boldsymbol{Z}^*)$$

或写作

$$\mathrm{d}f(\boldsymbol{Z},\boldsymbol{Z}^*)=\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{Z})}\mathrm{vec}(\mathrm{d}\boldsymbol{Z})+\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{H}}(\boldsymbol{Z})}\mathrm{vec}(\mathrm{d}\boldsymbol{Z}^*) \tag{11}$$

式中

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

$$\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{H}}(\boldsymbol{Z})}=\left[\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{11}^*},\cdots,\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{m1}^*},\cdots,\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{1n}^*},\cdots,\frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{mn}^*}\right]$$

另一方面，复变函数$f(\boldsymbol{Z},\boldsymbol{Z}^*)$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

$$\mathrm{D}_{Z} f(\boldsymbol{Z},\boldsymbol{Z}^*) \triangleq \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \boldsymbol{Z}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{11}} & \cdots & \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{1n}} & \cdots & \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z_{mn}} \end{bmatrix}$$

$$\mathrm{D}_{Z^*} f(\boldsymbol{Z},\boldsymbol{Z}^*) \triangleq \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial \boldsymbol{Z}^{\mathrm{H}}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z^*_{11}} & \cdots & \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z^*_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z^*_{1n}} & \cdots & \frac{\partial f(\boldsymbol{Z},\boldsymbol{Z}^*)}{\partial Z^*_{mn}} \end{bmatrix}$$

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

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

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

若令

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

则

$$\frac{\partial f(\boldsymbol{Z}, \boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{Z})} = \mathrm{rvec}(\mathrm{D}_{\boldsymbol{Z}} f(\boldsymbol{Z}, \boldsymbol{Z}^*)) = \mathrm{rvec}(\boldsymbol{A}) = \mathrm{vec}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})$$

$$\frac{\partial f(\boldsymbol{Z}, \boldsymbol{Z}^*)}{\partial \mathrm{vec}^{\mathrm{H}}(\boldsymbol{Z})} = \mathrm{rvec}(\mathrm{D}_{\boldsymbol{Z}^*} f(\boldsymbol{Z}, \boldsymbol{Z}^*)) = \mathrm{rvec}(\boldsymbol{B}) = \mathrm{vec}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}})$$

于是，式(11)可以写成

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

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

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

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

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

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

注意：在矩阵微分结果$\mathrm{d}f(\boldsymbol{X},\boldsymbol{X}^*)$中，必须将$\mathrm{d}\boldsymbol{X}$和$\mathrm{d}\boldsymbol{X}^*$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。

3.6 复矩阵微分计算

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

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

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

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

$$\nabla_{\boldsymbol{X}^*}\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^*\boldsymbol{B}) = (\boldsymbol{B}\boldsymbol{X}\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}$$

类似地，迹函数$\mathrm{tr}(\boldsymbol{XAX}^{\mathrm{H}}\boldsymbol{B})$的复矩阵微分为

$$\begin{aligned}\mathrm{d}[\mathrm{tr}(\boldsymbol{XAX}^{\mathrm{H}}\boldsymbol{B})] &= \mathrm{tr}(\boldsymbol{AX}^{\mathrm{H}}\boldsymbol{B}\mathrm{d}\boldsymbol{X}) + \mathrm{tr}(\boldsymbol{XA}\mathrm{d}\boldsymbol{X}^{\mathrm{H}}\cdot\boldsymbol{B}) \\ &= \mathrm{tr}(\boldsymbol{AX}^{\mathrm{H}}\boldsymbol{B}\mathrm{d}\boldsymbol{X}) + \mathrm{tr}[(\boldsymbol{BXA})^{\mathrm{T}}\mathrm{d}\boldsymbol{X}^*]\end{aligned}$$

故迹函数$\mathrm{tr}(\boldsymbol{XAX}^{\mathrm{H}}\boldsymbol{B})$的梯度矩阵和共轭梯度矩阵分别为

$$\nabla_{\boldsymbol{X}}\mathrm{tr}(\boldsymbol{XAX}^{\mathrm{H}}\boldsymbol{B}) = (\boldsymbol{AX}^{\mathrm{H}}\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^*\boldsymbol{A}^{\mathrm{T}}$$

$$\nabla_{\boldsymbol{X}^*}\mathrm{tr}(\boldsymbol{XAX}^{\mathrm{H}}\boldsymbol{B}) = [(\boldsymbol{BXA})^{\mathrm{T}}]^{\mathrm{T}} = \boldsymbol{BXA}$$

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$|\boldsymbol{X}\boldsymbol{X}^*|$ 和$|\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}|$的复矩阵微分分别为

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

和

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

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

$f(\boldsymbol{X},\boldsymbol{X}^*)$	梯度矩阵$\partial f(\boldsymbol{X},\boldsymbol{X}^*)/\partial \boldsymbol{X}$	共轭梯度矩阵$\partial f(\boldsymbol{X},\boldsymbol{X}^*)/\partial \boldsymbol{X}^*$
$\mathrm{tr}(\boldsymbol{A}\boldsymbol{X})$	$\boldsymbol{A}^{\mathrm{T}}$	$\boldsymbol{O}$
$\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{\mathrm{H}})$	$\boldsymbol{O}$	$\boldsymbol{A}$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B})$	$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{X}^{\mathrm{T}}+\boldsymbol{B}\boldsymbol{X}\boldsymbol{A}$	$\boldsymbol{O}$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^*\boldsymbol{B})$	$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{T}}$	$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{X}^*\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})$	$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{T}}$	$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$
$\mathrm{tr}(\boldsymbol{X}\boldsymbol{A}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{B})$	$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{X}^*\boldsymbol{A}^{\mathrm{T}}$	$\boldsymbol{B}\boldsymbol{X}\boldsymbol{A}$
$\mathrm{tr}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})$	$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}^*\boldsymbol{B}^{\mathrm{T}}$	$\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}$

几种行列式函数的梯度矩阵与共轭梯度矩阵

$f(\boldsymbol{X},\boldsymbol{X}^*)$	梯度矩阵$\partial f(\boldsymbol{X},\boldsymbol{X}^*)/\partial \boldsymbol{X}$	共轭梯度矩阵$\partial f(\boldsymbol{X},\boldsymbol{X}^*)/\partial \boldsymbol{X}^*$
$\lvert\boldsymbol{X}\rvert$	$\lvert\boldsymbol{X}\rvert\boldsymbol{X}^{-\mathrm{T}}$	$\boldsymbol{O}$
$\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rvert$	$2\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\rvert(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}$	$\boldsymbol{O}$
$\lvert\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\rvert$	$2\lvert\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\rvert\boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}$	$\boldsymbol{O}$
$\lvert\boldsymbol{X}\boldsymbol{X}^*\rvert$	$\lvert\boldsymbol{X}\boldsymbol{X}^*\rvert(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^{\mathrm{H}}$	$\lvert\boldsymbol{X}\boldsymbol{X}^*\rvert\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}^{\mathrm{T}})^{-1}$
$\lvert\boldsymbol{X}^*\boldsymbol{X}\rvert$	$\lvert\boldsymbol{X}^*\boldsymbol{X}\rvert\boldsymbol{X}^{\mathrm{H}}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{H}})^{-1}$	$\lvert\boldsymbol{X}^*\boldsymbol{X}\rvert(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{H}})^{-1}\boldsymbol{X}^{\mathrm{T}}$
$\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\rvert$	$\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\rvert(\boldsymbol{X}^*\boldsymbol{X}^{\mathrm{T}})^{-1}\boldsymbol{X}^*$	$\lvert\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}}\rvert(\boldsymbol{X}\boldsymbol{X}^{\mathrm{H}})^{-1}\boldsymbol{X}$
$\lvert\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\rvert$	$\lvert\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\rvert\boldsymbol{X}^*(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}^*)^{-1}$	$\lvert\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}\rvert\boldsymbol{X}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})^{-1}$

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$\mathrm{tr}[(\boldsymbol{A}+\boldsymbol{X}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{X})^{-1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{X})]$的微分矩阵和梯度矩阵。

习题2：计算复矩阵微分

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

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