实验物理中的统计方法

补充：向量微分间接

杨振伟

yangzw@pku.edu.cn

北京大学物理学院

要点

- 向量微分定义
- 标量函数对向量的梯度
- 向量函数对向量的梯度
- 向量微分常用公式
- 举例

提纲

- 很多实际问题涉及到向量微分
  - 可按分量展开解决
  - 但直接用向量微分会方便很多
- 函数对向量的微分形式与普通微分类似
- 向量微分的结果是向量或者矩阵，不论被求导的函数本身是标量函数还是矢量函数(或者矩阵)
- 本讲义以实函数为例

行向量和列向量的定义

- 行向量

$$\boldsymbol{x}^T \equiv [x_1, x_2, \ldots, x_n] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T$$

- 列向量

$$\boldsymbol{x} \equiv [x_1, x_2, \ldots, x_n]^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

向量微分的定义(1)

➢ 对列向量和行向量的微分（梯度）定义

对列向量的梯度：

相对于$n\times 1$向量（列向量）$\boldsymbol{x}$ 的梯度算子记作$\nabla_{\boldsymbol{x}}$，定义为

$$\nabla_{\boldsymbol{x}} \equiv \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n}\right]^T = \frac{\partial}{\partial \boldsymbol{x}}$$

对行向量的梯度：

相对于$1\times n$向量（行向量）$\boldsymbol{x}^T$ 的梯度算子记作$\nabla_{\boldsymbol{x}^T}$，定义为

$$\nabla_{\boldsymbol{x}^T} \equiv \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n}\right] = \frac{\partial}{\partial \boldsymbol{x}^T}$$

标量函数对向量的梯度

所以，假设标量函数$f(\boldsymbol{x})$以n维列向量$\boldsymbol{x}$为变元，其相对于$\boldsymbol{x}$的梯度是n维列向量：

$$\nabla_{\boldsymbol{x}} f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]^T = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}$$

而函数$f(\boldsymbol{x})$相对于行向量$\boldsymbol{x}^T$的梯度是n维行向量：

$$\nabla_{\boldsymbol{x}^T} f(\boldsymbol{x}) = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right] = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^T}$$

1）梯度的分量 $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$给出了$f(\boldsymbol{x})$在第$i$个方向上的变化率

2）标量函数对列向量的微分结果为列向量
标量函数对行向量的微分结果为行向量

举例: 标量函数$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x}\left(= \sum_{i=1}^{n} x_i^2\right)$

求$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}$和 $\frac{\partial f\left(\boldsymbol{x}^T\right)}{\partial \boldsymbol{x}^T}$。

解：根据对列向量和行向量微分的定义，

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]^T$$
$$= [2x_1, 2x_2, \cdots, 2x_n]^T = 2\boldsymbol{x} \text{ (列向量)}$$

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}^T} = \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \cdots, \frac{\partial f(\boldsymbol{x})}{\partial x_n}\right]$$
$$= [2x_1, 2x_2, \cdots, 2x_n] = 2\boldsymbol{x}^T \text{ (行向量)}$$

推广到矩阵函数(1)

m维行向量函数$\boldsymbol{f}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_m(\boldsymbol{x})]$

相对于n维列向量$\boldsymbol{x}$的梯度是一个$n\times m$矩阵：

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

推广到矩阵函数(2)

m维列向量函数$\boldsymbol{y}(\boldsymbol{x}) = [y_1(\boldsymbol{x}), y_2(\boldsymbol{x}), \cdots, y_m(\boldsymbol{x})]^T$

相对于n维行向量$\boldsymbol{x}^T$的梯度是一个$m\times n$矩阵：

$$\frac{\partial \boldsymbol{y}(\boldsymbol{x})}{\partial \boldsymbol{x}^T} = \begin{bmatrix} \frac{\partial y_1(\boldsymbol{x})}{\partial x_1}, & \frac{\partial y_1(\boldsymbol{x})}{\partial x_2}, & \cdots, & \frac{\partial y_1(\boldsymbol{x})}{\partial x_n} \\ \frac{\partial y_2(\boldsymbol{x})}{\partial x_1}, & \frac{\partial y_2(\boldsymbol{x})}{\partial x_2}, & \cdots, & \frac{\partial y_2(\boldsymbol{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m(\boldsymbol{x})}{\partial x_1}, & \frac{\partial y_m(\boldsymbol{x})}{\partial x_2}, & \cdots, & \frac{\partial y_m(\boldsymbol{x})}{\partial x_n} \end{bmatrix} = \nabla_{\boldsymbol{x}^T} \boldsymbol{y}(\boldsymbol{x})$$

这正是雅克比矩阵

再举例(1)：行向量函数$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}^T$,求$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}}$

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^T}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1}, \frac{\partial x_2}{\partial x_1}, \cdots, \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2}, \frac{\partial x_2}{\partial x_2}, \cdots, \frac{\partial x_n}{\partial x_2} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ \frac{\partial x_1}{\partial x_n}, \frac{\partial x_2}{\partial x_n}, \cdots, \frac{\partial x_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1,0,\cdots,0 \\ 0,1,\cdots,0 \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ 0,0,\cdots,1 \end{bmatrix} = \boldsymbol{I}$$

再举例(2)：列向量函数$\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x}$,求$\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}^T}$

$$\frac{\partial \boldsymbol{g}(\boldsymbol{x})}{\partial \boldsymbol{x}^T} = \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^T} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1}, \frac{\partial x_2}{\partial x_1}, \cdots, \frac{\partial x_n}{\partial x_1} \\ \frac{\partial x_1}{\partial x_2}, \frac{\partial x_2}{\partial x_2}, \cdots, \frac{\partial x_n}{\partial x_2} \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ \frac{\partial x_1}{\partial x_n}, \frac{\partial x_2}{\partial x_n}, \cdots, \frac{\partial x_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1,0, \cdots, 0 \\ 0,1, \cdots, 0 \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ 0,0, \cdots, 1 \end{bmatrix} = \boldsymbol{I}$$

一些常用公式 (1)

$$(1)\ \frac{\partial \boldsymbol{x}^T}{\partial \boldsymbol{x}} = \boldsymbol{I}, \qquad \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^T} = \boldsymbol{I}$$

若$\boldsymbol{A}$为n维方矩阵，$\boldsymbol{y}$为n维列向量，且$\boldsymbol{A}$和$\boldsymbol{y}$都与$\boldsymbol{x}$无关：

$$(2)\ \frac{\partial \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}^T} = \boldsymbol{A}, \qquad \frac{\partial \boldsymbol{x}^T\boldsymbol{A}}{\partial \boldsymbol{x}} = \boldsymbol{A}$$

$$(3)\ \frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{y}, \qquad \frac{\partial \boldsymbol{y}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{y}\ (!!!)$$

$$(4)\ \frac{\partial \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{x}$$

一些常用公式(2)

形式上，普通微分的一些法则都适用于向量微分，如

$$(1)\ f(\boldsymbol{x}) = c \quad \Longrightarrow \quad \frac{\partial c}{\partial \boldsymbol{x}} = 0$$

$$(2)\ \frac{\partial[c_1 f(\boldsymbol{x}) + c_2 g(\boldsymbol{x})]}{\partial \boldsymbol{x}} = c_1 \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} + c_2 \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}$$

$$(3)\ \frac{\partial[f(\boldsymbol{x}) g(\boldsymbol{x})]}{\partial \boldsymbol{x}} = \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} g(\boldsymbol{x}) + f(\boldsymbol{x}) \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}$$

$$(4)\ \frac{\partial}{\partial \boldsymbol{x}}\left[\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\right] = \frac{1}{g^2(\boldsymbol{x})}\left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} g(\boldsymbol{x}) - f(\boldsymbol{x}) \frac{\partial g(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]$$

$$(5)\ \frac{\partial f\big(\boldsymbol{y}(\boldsymbol{x})\big)}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{y}^T(\boldsymbol{x})}{\partial \boldsymbol{x}} \frac{\partial f(\boldsymbol{y})}{\partial \boldsymbol{y}}$$

例: 多维随机变量的线性变换的概率密度

假设n维随机变量$\boldsymbol{x} = (x_1, \cdots, x_n)^T$的联合概率密度为$f(\boldsymbol{x})$，变量$\boldsymbol{y} = (y_1, \cdots, y_n)$是$\boldsymbol{x}$的线性变换，$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$，即

$$y_i = \sum_{j=1}^{n} A_{ij} x_j \,.$$

假设逆变换$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$ 存在。

(a) 证明$\boldsymbol{y}$的联合概率密度为$g(\boldsymbol{y}) = f(\boldsymbol{A}^{-1}\boldsymbol{x})|\det(\boldsymbol{A}^{-1})|$。
(b) 当$\boldsymbol{A}$是正交矩阵时，即$\boldsymbol{A}^{-1} = \boldsymbol{A}^T$，求$g(\boldsymbol{y})$

解：(a) 根据*Statistical Data Analysis* (G. Cowan)的式(1.37)，在逆变换$\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{y}$ 存在的情况下，$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$的联合概率密度为

$$g(\boldsymbol{y})=f(\boldsymbol{x})|J|=f(\boldsymbol{A}^{-1}\boldsymbol{y})|J|$$

其中J是Jacobian行列式，即

$$J=\det\left(\frac{\partial\boldsymbol{x}}{\partial\boldsymbol{y}^T}\right)=\det\left(\frac{\partial\boldsymbol{A}^{-1}\boldsymbol{y}}{\partial\boldsymbol{y}^T}\right)=\det(\boldsymbol{A}^{-1})$$

所以，$g(\boldsymbol{y})=f(\boldsymbol{A}^{-1}\boldsymbol{y})|J|=f(\boldsymbol{A}^{-1}\boldsymbol{y})|\det(\boldsymbol{A}^{-1})|$。

(b) 当$\boldsymbol{A}$是正交矩阵时，即$\boldsymbol{A}^{-1}=\boldsymbol{A}^T$。

由$\det(\boldsymbol{A}\boldsymbol{B})=\det(\boldsymbol{A})\det(\boldsymbol{B})$和$\det(\boldsymbol{A}^T)=\det(\boldsymbol{A})$，可得

$$1=\det(\boldsymbol{I})=\det(\boldsymbol{A}^{-1}\boldsymbol{A})=\det(\boldsymbol{A}^T\boldsymbol{A})=\det(\boldsymbol{A}^T)\det(\boldsymbol{A})=(\det(\boldsymbol{A}^T))^2$$

所以，$|\det(\boldsymbol{A}^{-1})|=|\det(\boldsymbol{A}^T)|=1$，即

$$g(\boldsymbol{y})=f(\boldsymbol{A}^{-1}\boldsymbol{y})$$

这比用分量形式简洁很多。

参考资料

《矩阵分析与应用》(2004)，张贤达，清华大学出版社（第5.1.2节)