粒子物理与核物理实验中的数据分析

杨振伟
清华大学

补充讲义：数学基础：

——向量微分

本讲要点

- 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

简介

- 本课程很多问题涉及到向量微分。
  处理这些问题，可按分量展开解决，
  但直接进行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，差别在于微分后的结果是向量或者矩阵，不论被微分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、列向量定义

行向量：

$$x^T = [x_1, x_2, \ldots, x_n]$$

列向量：

$$x = [x_1, x_2, \ldots, x_n]^T$$

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)$\mathbf{x}$的梯度算子记作$\nabla_{\mathbf{x}}$，定义为

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

对行向量的梯度：

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

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

标量函数对向量的梯度

所以，假设标量函数$f(\mathbf{x})$以n维列向量$\mathbf{x}$为变元，

其相对于$\mathbf{x}$的梯度是n维列向量，

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

而$f(\mathbf{x})$相对于行向量$\mathbf{x}^T$的梯度是n维行向量，

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

1) 梯度的分量$\dfrac{\partial f(\mathbf{x})}{\partial x_i}$给出了$f(\mathbf{x})$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$f(\mathbf{x})=\mathbf{x}^T\mathbf{x}\ \left(=\sum_{i=1}^{n}x_i^2\right)$，求$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}$和$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^T}$

解：根据对列向量和行向量微分的定义，

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

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

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

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

推广到矩阵函数(1)

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

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

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

推广到矩阵函数(2)

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

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

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

注：这正是***Jacobi***矩阵

再举例(1)

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

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

再举例(2)

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

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

一些常用公式(1)

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

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

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

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

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

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

$$(1)\; f(\mathbf{x}) = c \Rightarrow \frac{\partial c}{\partial \mathbf{x}} = 0$$

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

$$(3)\; \frac{\partial f(\mathbf{x}) g(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} g(\mathbf{x}) + f(\mathbf{x}) \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}}$$

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

$$(5)\; \frac{\partial f(\mathbf{y}(\mathbf{x}))}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}^T(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}}$$

例子(Exercise1.8)

假设n维随机变量$\boldsymbol{x}=(x_1,\ldots,x_n)^T$的联合p.d.f.为$f(\boldsymbol{x})$，变量$\boldsymbol{y}=(y_1,\ldots,y_n)^T$是$\boldsymbol{x}$的线性变换，$\boldsymbol{y}=A\boldsymbol{x}$，即

$$y_i=\sum_{j=1}^{n}A_{ij}x_j。$$

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

(a)证明$\boldsymbol{y}$的联合p.d.f.为$g(\boldsymbol{y})=f(A^{-1}\boldsymbol{y})\left|\det(A^{-1})\right|$。

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

续上

解:(a)根据公式(1.37), $g(a_1,\ldots,a_n)=f(x_1,\ldots,x_n)\left|J\right|$,

很容易得到本题所求为

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

其中J为Jacobian矩阵的行列式，即

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

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

(b)如果A是正交矩阵，则$A^{-1}=A^T$。

由$\det(AB)=\det(A)\det(B)$，以及$\det(A^T)=\det(A)$得到

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

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

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

续上

如果用分量方式，由于Jacobian行列式$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \cdots & \cdots & \cdots & \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$

则要计算出$\frac{\partial x_i}{\partial y_j}$，证明$\frac{\partial x_i}{\partial y_j} = (A^{-1})_{ij}$

由$x_i = \sum_{k=1}^{n} (A^{-1})_{ik} y_k$,显然可得

$$\frac{\partial x_i}{\partial y_j} = \sum_{k=1}^{n} (A^{-1})_{ik} \frac{\partial y_k}{\partial y_j} = \sum_{k=1}^{n} (A^{-1})_{ik} \frac{\partial y_k}{\partial y_j} = \sum_{k=1}^{n} (A^{-1})_{ik} \delta_{kj} = (A^{-1})_{ij}。$$

得证。但不如直接向量微分简洁。

参考资料

1. 《矩阵分析与应用》(2004)，张贤达，清华大学出版社（第5.1.2节）