

基本概念

标量 | 只有大小，没有方向，可用实数表示的量。

实值函数 | 函数 $y = f(x), y \in \mathbb{R}, x \in \mathbb{R} \cup \mathbb{C}$

梯度 | 根据自变量和应变量的不同可以分为：

✓ 自变量为实向量的标量函数关于向量的梯度。

$$f \in \mathbb{R}, \nabla_x f = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]^T = \frac{\partial f}{\partial x}.$$

✓ 自变量为实向量的向量函数关于向量的梯度。

$$f \in \mathbb{R}^{1 \times n}, \nabla_x f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \frac{\partial f}{\partial x}.$$

✓ 自变量为实向量的标量函数关于矩阵的梯度。

$$f \in \mathbb{R}, \nabla_X f = \begin{bmatrix} \frac{\partial f_1}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \dots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \dots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \dots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix} = \frac{\partial f}{\partial X}.$$

Jacobian矩阵 | 若函数 $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ，有

$$x = [x_1, x_2, \dots, x_n]^T, f(x) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix}.$$

其Jacobian矩阵 $J(x)$ 可以写为

$$f \in \mathbb{R}^{1 \times n}, J(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \nabla_x^T f.$$

✓ Jacobian矩阵表现了向量函数的最佳线性逼近。

Hessian矩阵 | 若 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 为二次可导函数， $x = [x_1, \dots, x_n]^T$ ， f 的Hessian矩阵为

$$H(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} = \nabla_x^T f.$$

✓ Hessian矩阵使用函数的二阶信息，常用于Newton法解决大规模的优化问题。

实值函数有关向量的梯度

函数 f 关于行向量 x^T 的梯度为

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

函数 f 关于列向量 x 的梯度计算公式为

✓ 若 f 为常数，那么对应的梯度为 $\mathbf{0}$ 。

✓ 如果 A 和 y 与 x 无关，那么

$$\frac{\partial x^T A y}{\partial x} = \frac{\partial x^T}{\partial x} A y = A y.$$

加法法则 | 若 $f(x), g(x)$ 均为 x 的实值函数

$$\frac{\partial [pf(x) + qg(x)]}{\partial x} = p \frac{\partial f(x)}{\partial x} + q \frac{\partial g(x)}{\partial x}.$$

乘法法则 | 若 $f(x), g(x)$ 均为 x 的实值函数

$$\frac{\partial f(x)g(x)}{\partial x} = f(x) \frac{\partial g(x)}{\partial x} + g(x) \frac{\partial f(x)}{\partial x}.$$

除法法则 | 若 $f(x)$ 为向量 x 的向量值函数

$$\frac{\partial g(f(x))}{\partial x} = \frac{\partial f^T(x)}{\partial x} \frac{\partial g(f)}{\partial f}.$$

链式法则 | 若 $g(x)$ 为 x 的向量值函数

$$\frac{\partial f(g(x))/g(x)}{\partial x} = \frac{\partial g^T(x)}{\partial x} \frac{\partial f(g)}{\partial g}.$$

常见类型的实值函数的梯度为

$$\begin{aligned} \nabla(x^T x) &= \frac{\partial x^T x}{\partial x} = 2x^T \\ \nabla(a^T x) &= \frac{\partial a^T x}{\partial x} = a, \nabla(x^T A) = \frac{\partial x^T A}{\partial x} = A \\ \nabla(x^T A x) &= \frac{\partial x^T A x}{\partial x} = (A + A^T)x \end{aligned}$$

矩阵迹、行列式的梯度矩阵

矩阵迹的性质 | 二次型函数的迹与它本身相等。

$$f(x) = x^T A x = \text{Tr}(x^T A x) = \text{Tr}(x x^T A).$$

✓ 有关矩阵的迹，常见的梯度计算公式：

$$\begin{aligned} \nabla(\text{Tr}(X)) &= \frac{\partial \text{Tr}(X)}{\partial X} = I \\ \nabla(\text{Tr}(X^{-1})) &= \frac{\partial \text{Tr}(X^{-1})}{\partial X} = -(X^{-1})^T \\ \nabla(\text{Tr}(X^T X)) &= \frac{\partial \text{Tr}(X^T X)}{\partial X} = (2X^T)^T = 2X \\ \nabla(\text{Tr}(X A)) &= \frac{\partial \text{Tr}(X A)}{\partial X} = \frac{\partial \text{Tr}(A X)}{\partial X} = A^T \end{aligned}$$

✓ 有关矩阵的行列式，常见的梯度计算公式：

$$\begin{aligned} \nabla(\det(X)) &= \frac{\partial \det(X)}{\partial X} = \det(X) X^{-T} \\ \nabla(\det(X^{-1})) &= \frac{\partial \text{Tr}(X^{-1})}{\partial X} = -(\det(X))^{-1} (X^{-1})^T \\ \nabla(\det(X X^T)) &= \frac{\partial \det(X X^T)}{\partial X} = 2(\det(X X^T))^2 X^{-T} \\ \nabla(\det(\log X)) &= \frac{1}{\det(X)} \frac{\partial \det(X)}{\partial X} = 2X^{-1} - \text{diag}(X^{-1}) \end{aligned}$$

实值函数的梯度矩阵

实值函数的梯度函数 | 实值函数有关矩阵的梯度。

✓ 若 $X \in \mathbb{R}^{m \times n}, f(X) = c, \nabla_X f(X) = \mathbf{0}_{m \times n}$

加法法则 | 若 $f(X), g(X)$ 均为矩阵 X 的实值函数

$$\frac{\partial [pf(X) + qg(X)]}{\partial X} = p \frac{\partial f(X)}{\partial X} + q \frac{\partial g(X)}{\partial X}.$$

乘法法则 | 若 $f(X), g(X)$ 均为矩阵 X 的实值函数

$$\frac{\partial f(X)g(X)}{\partial X} = f(X) \frac{\partial g(X)}{\partial X} + g(X) \frac{\partial f(X)}{\partial X}.$$

除法法则 | 若 $f(X), g(X)$ 为 X 的函数， $g(X) \neq 0$

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

链式法则 | 若 $g(X)$ 是自变量为矩阵 X 的实值函数， $f(y)$ 是自变量为标量 y 的实值函数

$$\frac{\partial f(g(X))}{\partial X} = \frac{\partial f(y)}{\partial y} \frac{\partial g(X)}{\partial X}.$$

常见类型的实值函数的梯度矩阵计算公式

$$\begin{aligned} \nabla_X(a^T X y) &= \frac{\partial a^T X y}{\partial X} = a y^T \\ \nabla_X(a^T X X^T y) &= \frac{\partial a^T X y}{\partial X} = (a y^T + y x^T) X \\ \nabla_X(e^{a^T X y}) &= \frac{\partial e^{a^T X y}}{\partial X} = a y^T e^{a^T X y} \end{aligned}$$

标量函数的微分

✓ 标量函数 $f(x)$ 的导数 $f'(x)$ 定义为

$$f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta}$$

即 $f(x + \Delta x) = f(x) + \Delta x f'(x) + R, \lim_{\Delta x \rightarrow 0} \frac{R}{\Delta x} = 0$ 。

上式称为**泰勒公式的一阶展开式**。

✓ $f(x)$ 在 x 点的**一阶微分**为 $df(x) = \Delta x f'(x)$ 。

矩阵的微分

矩阵微分 | 实函数微分对矩阵函数的推广情况。

如果 $x, \Delta x$ 为 $n \times 1$ 的向量， $\exists A(x) \in \mathbb{R}^{m \times n}$ ，使得

$$f(x + \Delta x) = f(x) + A(x) \Delta x + R.$$

其中 $\lim_{\Delta x \rightarrow 0} \frac{R}{\|\Delta x\|_2} = 0$ ，那么函数 $f(X)$ 在向量 x 处的一阶微分向量为

$$df(x) = A(x) \Delta x.$$

$A(x)$ 称为向量函数 $f(x)$ 的一阶**导数矩阵**；如果向量函数 $f(x)$ 在 c 处可微， $u \in \mathbb{R}^{n \times 1}$

$$df(c) = [D(f(x))] \mu, D(f(x)) \in \mathbb{R}^{m \times n}.$$

D_{ij} 表示 $f(x)$ 第 i 个元素关于 c 的第 j 个元素的偏导； $D(f(x))$ 实质是 $f(x)$ 在 c 处的Jacobian矩阵。

梯度矩阵 | $f(x)$ 在 c 处的梯度矩阵为 $(D(f(x)))^T$ 。

求解 $m \times 1$ 的向量函数 $f(c)$ 的梯度矩阵 $\nabla f(X)$ ：

✓ 求解向量函数微分 $df(c)$ ，获得Jacobian矩阵；

✓ 将Jacobian矩阵转置，得到梯度矩阵 $\nabla f(c)$ 。

常数矩阵微分 | $dC = 0$ 。

常数与矩阵的乘积 | $dcX = cdX$ 。

矩阵加和的微分 | $d(X + Y) = dX + dY$ 。

矩阵乘积的微分 | $d(XY) = (dX)Y + X(dY)$ 。

常用的矩阵微分计算公式

$$\begin{aligned} d(X^T) &= (dX)^T \\ d(\text{Tr}(X)) &= \text{Tr}(dX) \\ d(X^{-1}) &= -X^{-1}(dX)X^{-1} \\ d(\ln(X)) &= X^{-1}dX \\ d(\det(X)) &= \det(X) \text{Tr}(X^{-1}dX) \end{aligned}$$

二阶微分矩阵 | 矩阵函数 $f(X)$ 的二阶微分为

$$d^2 f(X) = \begin{cases} \text{Tr}(U(dX)^T V(dX)) & \text{或者} \\ \text{Tr}(U(dX)V(dX)) \end{cases}.$$

与之对应的Hessian矩阵可写为

$$H(f(X)) = \begin{cases} \frac{1}{2}(U^T \otimes V + U \otimes V^T) & \text{或者} \\ \frac{1}{2}K_{nm}(U^T \otimes V + V^T \otimes U) \end{cases}.$$

其中， K_{nm} 为交换矩阵。