1.2预备知识

1.2.1 向量范数与矩阵范数

1.2.1　向量范数与矩阵范数

$$\mathrm{x}:=(\mathrm{x}_1,\mathrm{x}_2,\cdots,x_n)^T=\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix},\langle x,y\rangle:=x^Ty=\sum_{i=1}^{n}x_iy_i.$$

定义：称映射$\|\cdot\|:R^n\to R$为R^n上的范数，当且仅当：

(1)$\forall x\in R^n,\|x\|\geq 0$且$\|\mathrm{x}\|=0$当且仅当$\mathrm{x}=0$;

(2) $\forall x\in R^n,a\in R,\|ax\|=|a|\,\|\mathrm{x}\|.$

(3)$\forall x,y\in R^n,\|x+y\|\leq\|x\|+\|y\|.$

常用范数

l_1-范数：$\|x\|_1 := \sum_{i=1}^{n} |x_i|$;

l_2-范数： $\|\mathrm{x}\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}$;

l_∞-范数： $\|\mathrm{x}\|_\infty = \max\{|\mathrm{x}_i|: \mathrm{i} \in \{1,2,\cdots,\mathrm{n}\}\}$;

l_p-范数： $\|\mathrm{x}\|_\mathrm{p} = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}, p \in [1,\infty)$;

椭球范数：$\|x\|_A := \sqrt{x^T A x}, \forall x \in R^n$ (A为正定矩阵)

范数间的关系：$\|\mathrm{x}\|_\infty \le \|\mathrm{x}\|_2 \le \|x\|_1 \le n\|\mathrm{x}\|_\infty$;

命题： 假设$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$是定义在R^n上的任意两种范数，那么总存在两个正数ξ_1，ξ_2使得对任意$x \in R^n$都有

$$\xi_1 \|x\|_\alpha \le \|x\|_\beta \le \xi_2 \|x\|_\alpha$$

定义1.2.2（矩阵范数）称映射$\|\cdot\|: R^{n\times n} \to R$上的范数，当且仅当它具有一下性质：

(1)对$\forall A \in R^{n\times n}, \|A\| \ge 0,$ 且$\|A\| = 0$当且仅当$A = 0$;

(2)$\forall A \in R^{n\times n}, \forall \alpha \in R, \|\alpha A\| = |\alpha| \|A\|;$

(3)$\forall A, B \in R^{n\times n}, \|A + B\| \le \|A\| + \|B\|.$

常用矩阵范数

诱导范数：$\|A\| := \max_{x \in R^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}$

列范数：$\|A\|_1 = \max\{\sum_{i=1}^{n} |a_{ij}| \,|\, j \in \{1, 2, \cdots, n\}\};$

行范数：$\|A\|_\infty = \max\{\sum_{j=1}^{n} |a_{ij}| \,|\, i \in \{1, 2, \cdots, n\}\};$

谱范数：$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)}$($\lambda_{\max}$表示矩阵的最大特征值);

Frobenius 范数：$\|A\|_F = \sqrt{Tr(A^T A)}$($Tr(\cdot)$表示矩阵的迹)

相容性：$\|AB\| \le \|A\|\|B\|; \|Ax\| \le \|A\|\|x\|$

几个常用不等式

$Cauchy-Schwarz$不等式：$\forall x,y\in R^n, |x^Ty|\le \|x\|_2\|y\|_2$

且等号成立当且仅当x和y线性相关；

广义$Cauchy-Schwarz$不等式：设$A\in R^{n\times n}$正定，则对

$\forall x,y\in R^n, |x^Ty|\le \|x\|_A\|y\|_{A^{-1}}$；

$Young$不等式：设$p,q>1$且$\frac{1}{p}+\frac{1}{q}=1$，如果$a,b\in R$，则

$|ab|\le \frac{|a|^p}{p}+\frac{|b|^q}{q}$(``="成立当且仅当$|a|^p=|b|^q$)；

$\ddot{H}older$不等式：$\forall x,y\in R^n$，

$$\left|x^Ty\right|\le \|x\|_p\|y\|_q=\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}\left(\sum_{i=1}^n |y|^q\right)^{\frac{1}{q}}$$

其中$p,q>1$且$\frac{1}{p}+\frac{1}{q}=1$；

$Minkowski$不等式：$\forall x,y\in R^n, p\in[1,\infty)$

$$\|x+y\|^p\le \|x\|^p+\|y\|^p=\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}+\left(\sum_{i=1}^n |y|^q\right)^{\frac{1}{q}}.$$

1.2.2 函数的可微性

函数的可微性

1、 连续、连续可微、二次连续可微：

(1) f在F上连续（$f \in C$)：

给定函数$f:\mathrm{F} \subseteq R^n$,如果$f$在每一点$x \in \mathrm{F}$连续；

(2) f在F上连续可微（$f \in C^1$)：

若F为开集且在每一点$x \in \mathrm{F}$处，每一个偏导数$\dfrac{\partial f(x)}{\partial x_i}(i \in \{1,2,\cdots,n\})$

存在且连续；

(3) f在F上二次连续可微（$f \in C^2$)：

若在每一点$x \in \mathrm{F}$处，每一个二阶偏导数$\dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}(i,j \in \{1,2,\cdots,n\})$

存在且连续。

2 梯度、海塞阵(Hesse)：

（1） 梯度：

设$f:F\subseteq R^n\to R$是一阶连续可微的，则f在x处的一阶偏导数（f在x处的梯度）

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

(2) 海塞阵(Hesse)：

设f是二阶连续可微的，则f在x处的二阶导数（f在x处的Hesse阵）

$$\nabla^2 f(x)=\begin{pmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\partial x_n}\\ \frac{\partial^2 f}{\partial x_2\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\partial x_n}\\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n\partial x_1} & \frac{\partial^2 f}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{pmatrix}=\left(\frac{\partial^2 f}{\partial x_i\partial x_j}\right)_{n\times n}$$

注：二次函数$f(x)=\frac{1}{2}x^T Ax+b^T x+c$,其中$A\in R^{n\times n},b\in R^n,c\in R$,则

$$\nabla f(x)=Ax+b,\quad \nabla^2 f(x)=A$$

(3) 多变量向量值函数的Jacobi阵:

设多变量向量值函数 $F:\mathsf{F} \subseteq R^n \to R^m$在$x \in \mathsf{F}$ 处连续可微,

则F在$x \in \mathsf{F}$ 处的一阶导数为

$$F'(x) = \begin{pmatrix} \dfrac{\partial F_1(x)}{\partial x_1} & \dfrac{\partial F_1(x)}{\partial x_2} & \cdots & \dfrac{\partial F_1(x)}{\partial x_n} \\ \dfrac{\partial F_2(x)}{\partial x_1} & \dfrac{\partial F_2(x)}{\partial x_2} & \cdots & \dfrac{\partial F_2(x)}{\partial x_n} \\ \vdots & & & \\ \dfrac{\partial F_m(x)}{\partial x_1} & \dfrac{\partial F_m(x)}{\partial x_2} & \cdots & \dfrac{\partial F_m(x)}{\partial x_n} \end{pmatrix} = \left(\dfrac{\partial F_i(x)}{\partial x_j} \right)_{m\times n} \in R^{m\times n}$$

称为F在x处的Jacobi阵.

例：设多变量向量值函数$F(x) = Ax$,其中$A \in R^{m\times n}$,则Jacobi阵为

$$F'(x) = A.$$

例　求函数$f(x)=x_1^2+2x_2^2-2x_1x_2-4x_1+3$的梯度与Hesse阵。

解：（法1）$\dfrac{\partial f(x)}{\partial x_1}=2x_1-2x_2-4;\dfrac{\partial f(x)}{\partial x_2}=4x_2-2x_1;$

所以$\nabla f(x)=(2x_1-2x_2-4,4x_2-2x_1)^T.$

$$\frac{\partial^2 f(x)}{\partial x_1^2}=2;\frac{\partial^2 f(x)}{\partial x_1\partial x_2}=-2;\frac{\partial^2 f(x)}{\partial x_2\partial x_1}=-2;\frac{\partial^2 f(x)}{\partial x_2^2}=4;$$

所以$\partial^2 f(x)=\begin{pmatrix}2 & -2\\ -2 & 4\end{pmatrix}.$

（法2）$f(x)=\dfrac{1}{2}x^T\begin{pmatrix}2 & -2\\ -2 & 4\end{pmatrix}x+\begin{pmatrix}-4\\ 0\end{pmatrix}^T x+3;$

$$\nabla f(x)=Ax+b=\begin{pmatrix}2 & -2\\ -2 & 4\end{pmatrix}x+\begin{pmatrix}-4\\ 0\end{pmatrix}=\begin{pmatrix}2x_1-2x_2-4\\ -2x_1+4x_2\end{pmatrix}.$$

$$\partial^2 f(x)=A=\begin{pmatrix}2 & -2\\ -2 & 4\end{pmatrix}.$$

例 求函数$f(x)=(x_2-x_1^2)^2+(1+x_1)^2$的梯度与Hesse阵。

解：$\dfrac{\partial f(x)}{\partial x_1}=2(x_2-x_1^2)(-2x_1)+2(1+x_1);\dfrac{\partial f(x)}{\partial x_2}=2(x_2-x_1^2);$

所以$\nabla f(x)=\begin{pmatrix}2(x_2-x_1^2)(-2x_1)+2(1+x_1)\\2(x_2-x_1^2)\end{pmatrix};$

$$\frac{\partial^2 f(x)}{\partial x_1^2}=12x_1^2-4x_2+2;\frac{\partial^2 f(x)}{\partial x_1\partial x_2}=-4x_1;\frac{\partial^2 f(x)}{\partial x_2^2}=2;$$

$$\frac{\partial^2 f}{\partial x_2^2}=\begin{pmatrix}12x_1^2-4x_2+2 & -4x_1\\-4x_1 & 2\end{pmatrix}.$$

例 求向量值函数$F(x)=(e^{x_1x_2}+3\sin x_2,1+x_2^2\cos x_1)^T$的Jacobi阵。

解：$F'(x)=\begin{pmatrix}\dfrac{\partial F_1(x)}{\partial x_1} & \dfrac{\partial F_1(x)}{\partial x_2}\\ \dfrac{\partial F_2(x)}{\partial x_1} & \dfrac{\partial F_2(x)}{\partial x_2}\end{pmatrix}=\begin{pmatrix}x_2e^{x_1x_2} & x_1e^{x_1x_2}+3\cos x_2\\-x_2^2\sin x_1 & 2x_2\cos x_1\end{pmatrix}.$

3 多变量实值函数的 中值定理、泰勒公式

定理1.2.1 设$f: \mathrm{F} \subseteq R^n \to R$，且$x, x^* \in \mathrm{F}$，如果$f$是一阶连续可微的，则

(1) 存在$\alpha \in$（0,1）使得：$f(x) = f(x^*) + \nabla f(\xi)^T (x - x^*)$

其中$\xi = x^* + \alpha(x - x^*)$

（2）f在x^*处有一阶$Taylor$公式：

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + o(\|x - x^*\|)$$

如果f在F上是二阶连续可微的，则

（3）存在$\alpha \in$（0,1）使得：

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla f^2(\xi)(x - x^*)$$

其中$\xi = x^* + \alpha(x - x^*)$

（4）f在x^*处有二阶$Taylor$公式：

$$f(x) = f(x^*) + \nabla f(x^*)^T (x - x^*) + \frac{1}{2}(x - x^*)^T \nabla f^2(x^*)(x - x^*) + o(\|x - x^*\|^2)$$

证明： 引入函数$\phi(t)=f(x^*+t(x-x^*))$,则$\phi(0)=f(x^*),\phi(1)=f(x)$.

由$f(x)$二阶连续可微知：$\phi(t)$二阶连续可微且

$$\phi'(t)=\sum_{i=1}^{n}\frac{\partial f(x^*+t(x-x^*))}{\partial x_i}(x_i-x_i^*)=\nabla f(x^*+t(x-x^*))^T(x-x^*),$$

$$\phi''(t)=\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\partial^2 f(x^*+t(x-x^*))}{\partial x_i\partial x_j}(x_j-x_j^*)\right)(x_i-x_i^*)$$

$$=(x-x^*)^T\nabla^2 f(x^*+t(x-x^*))(x-x^*).$$

从而$\phi'(0)=\nabla f(x^*)^T(x-x^*);\phi''(0)=(x-x^*)^T\nabla^2 f(x^*)(x-x^*)$.

利用一元函数的Taylor展开式得：

$$\phi(1)=\phi(0)+\phi'(\alpha);\quad \phi(1)=\phi(0)+\phi'(0)+\frac{1}{2}\phi''(\beta),\text{其中}\alpha,\beta\in(0,1).$$

因此(1),(3)得证。

记$y := \frac{x-x^*}{\left\| x-x^* \right\|}, \gamma = \left\| x-x^* \right\|$.令$g(\gamma) := f(x^* + \gamma y)$则

$g(\gamma) = f(x); g(0) = f(x^*); g^{'}(0)\gamma = \nabla f(x^*)^T (x-x^*);$

$g^{''}(0)\gamma^2 = (x-x^*)\nabla^2 f(x^*)(x-x^*).$

由一元函数的Taylor展开公式得：

$g(\gamma) = g(0) + g'(0)\gamma + o(\gamma),$

$g(\gamma) = g(0) + g'(0)\gamma + \frac{1}{2} g''(0)\gamma^2 + o(\gamma^2).$

因此(2)(4)得证。