梅加强——微分中值定理和拟微分中值定理

张浩然

2024 年 11 月 2 日

凸域就是开的凸集.

题目 1. 设 $f: D \to \mathbb{R}$ 在区域 $D \subset \mathbb{R}^n$ 中处处可微. 如果 f 的梯度场是常值的，证明 f 是线性函数.

解答.

设 f 的梯度场为 $(u_1, u_2, \cdots, u_n) = u^T$, $\nabla\left(f(x) - u^T \cdot x\right) = u^T - u^T = 0$

任取 $x, y \in D$, 根据微分中值定理：

$$\begin{aligned} f(x) - u^T \cdot x - f(y) + u^T \cdot y &= \nabla\left(f(\xi) - u^T \cdot \xi\right) \cdot (x - y) \\ &= 0 \end{aligned}$$

$f(x) - u^T \cdot x = f(y) - u^T \cdot y$.

存在 $b \in \mathbb{R}$, 使得 $f(x) = u^T \cdot x + b$.

题目 1 的注记.

此处 $b \neq 0$ 时，我们也说是线性函数，与高代要求可加性和齐次性不同，就像我们认为一元函数 $y = 2x + 1$ 也是线性函数一样.

题目 2. 设 $f:\mathbb{R}^n\to\mathbb{R}^n$ 处处可微. 当 $x\in\mathbb{R}^n$ 时，矩阵 $Jf(x)(Jf)^T(x)$ 的最大特征值记为 $\Lambda(x)$.

若 $\Lambda=\sup_{x\in\mathbb{R}^n}\Lambda(x)<+\infty$, 证明：

$$\|f(x)-f(y)\|\leqslant\sqrt{\Lambda}\|x-y\|$$

解答.

这是一个线性代数问题，熟知结论：

$$\left|\lambda E_{n\times n}-Jf(x)(Jf)^T(x)\right|=\left|\lambda E_{n\times n}-(Jf)^T(x)Jf(x)\right|$$

$Jf(x)(Jf)^T(x)$ 和 $(Jf)^T(x)Jf(x)$ 具有相同特征值，最大特征值都为 $\Lambda(x)$.

根据 2-范数的拟微分中值定理：

$$\|f(x)-f(y)\|_2\leqslant\|Jf(\xi)\|_2\cdot\|x-y\|_2$$

范数是 2-范数,那么 $\|Jf(\xi)\|_2=\sqrt{\max\lambda((Jf)^T(\xi)Jf(\xi))}$, 则 $\|Jf(\xi)\|_2=\sqrt{\Lambda(\xi)}$

$$\begin{aligned}\|f(x)-f(y)\|_2&\leqslant\|Jf(\xi)\|_2\cdot\|x-y\|_2\\&\leqslant\sqrt{\Lambda(\xi)}\cdot\|x-y\|_2\\&\leqslant\sqrt{\Lambda}\cdot\|x-y\|_2\end{aligned}$$

对于向量，2-范数和 Frobenius 范数相等，即 $\|x\|=\|x\|_2,\forall x\in\mathbb{R}^n$

$$\|f(x)-f(y)\|\leqslant\sqrt{\Lambda}\cdot\|x-y\|$$

题目 2 的注记.

这题也有点奇怪，此处范数应该只能是 2-范数，不然我不会做……

我们来证明 2-范数的拟微分中值定理：$D\subset\mathbb{R}^n$ 是凸域, $f:D\to R^m$ 可微，任取 $x,y\in D$, 存在 $\xi=\theta x+(1-\theta)y,\theta\in(0,1)$, 使得：

$$\|f(x)-f(y)\|_2\leqslant\|Jf(\xi)\|_2\cdot\|x-y\|_2$$

证明：任取 $x, y \in D$, $D \subset \mathbb{R}^n$ 是凸域，不妨设 $f(x) \neq f(y)$, 否则平凡. 那么我们取 $\sigma(t) = tx + (1-t)y, \forall t \in [0,1]$.

设函数 $\varphi(t) = \langle f(x) - f(y), f(\sigma(t)) \rangle$, 这是 $\mathbb{R}^m$ 上的标准内积.

根据一元的微分中值定理：存在 $\theta \in (0,1)$:

$$\begin{aligned}
&\varphi(1) - \varphi(0) = \varphi'(\theta) \\
&\langle f(x) - f(y), f(x) - f(y) \rangle = \langle f(x) - f(y), Jf(\theta x + (1-\theta)y)(x-y) \rangle
\end{aligned}$$

我们利用 $\|x\| = \|x\|_2, \forall x \in \mathbb{R}^m$,

$$\begin{aligned}
&\langle f(x) - f(y), f(x) - f(y) \rangle = \langle f(x) - f(y), Jf(\theta x + (1-\theta)y)(x-y) \rangle \\
&\|f(x) - f(y)\|_2^2 = \langle f(x) - f(y), Jf(\theta x + (1-\theta)y)(x-y) \rangle \\
&\|f(x) - f(y)\|_2^2 \leqslant \|f(x) - f(y)\|_2 \cdot \|Jf(\theta x + (1-\theta)y)(x-y)\|_2 \\
&\|f(x) - f(y)\|_2 \leqslant \|Jf(\theta x + (1-\theta)y)(x-y)\|_2 \\
&\|f(x) - f(y)\|_2 \leqslant \|Jf(\theta x + (1-\theta)y)\|_2 \cdot \|(x-y)\|_2
\end{aligned}$$

这里用到 $\|Ax\|_2 \leqslant \|A\|_2 \|x\|_2$, 于是定理成立.

我们叙述 2-范数, 对于向量 $x \in \mathbb{R}^n$, 向量的 2-范数定义如下：

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

根据 2-范数诱导的矩阵范数称为矩阵 2-范数，对于 $m \times n$ 的实系数矩阵，定义如下：

$$\|A\|_2 = \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sup_{\|x\|_2 = 1} \|Ax\|_2$$

于是，$\|A\|_2^2 = \sup_{\|x\|=1} \|Ax\|_2^2 = \sup_{\|x\|=1} (Ax)^T Ax = \sup_{\|x\|=1} x^T A^T Ax$.

$A^T A$ 是实的对称矩阵，那么可以对角化，而且 $\|Ax\|_2^2 = x^T A^T Ax \geqslant 0$, $A^T A$ 半正定，那么其有 n 个非负特征值，排列如下 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$.

那么存在正交矩阵 V，使得：

$$A^TA = V\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}V^T$$

于是，

$$A^TAV = V\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

设 V 的列向量为 $b_1, b_2, \cdots, b_n$, 它们都是单位向量 (根据正交矩阵性质) 它们为 A^TA 的特征向量，对应特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$

任取 $i \neq j, b_i^T b_j = 0$(用到正交矩阵列向量正交) . 于是，$b_1, b_2, \cdots, b_n$ 两两正交，构成 $\mathbb{R}^n$ 上的一组标准正交基.

于是设 $x = k_1u_1 + k_2u_2 + \cdots + k_nu_n$, 则

$$\begin{aligned} x^TA^TAx &= (k_1u_1 + +k_2u_2 + \cdots + k_nu_n)^T(k_1\lambda_1u_1 + k_2\lambda_2u_2 + \cdots + k_n\lambda_nu_n) \\ &= k_1^2\lambda_1 + \cdots + k_n^2\lambda_n \\ &\leqslant \lambda_1(k_1^2 + k_2^2 + \cdots + k_n^2) \\ &= \lambda_1 \\ &= \max\lambda(A^TA) \end{aligned}$$

这是因为在 $\|x\|_2 = 1, k_1^2 + k_2^2 + \cdots + k_n^2 = 1$ 条件下, 只需要令 $x = u_1$ 就可以取到等号，于是：

$$\|A\|_2 = \sup_{\|x\|_2\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \sup_{\|x\|_2=1}\|Ax\|_2 = \sqrt{\max\lambda(A^TA)}$$

题目 3. 设 $D \subset \mathbb{R}^n$ 是开集，$f \in C^1(D)$. 如果 $\overline{B_r(x^0)} \subset D$, 则任给 $\varepsilon > 0$, 存在 $\delta > 0$, 使得当 $x, y \in \overline{B_r(x^0)}$ 且 $\|x - y\| < \delta$ 时，

$$\|f(y) - f(x) - Jf(x)(y - x)\| \leqslant \varepsilon\|x - y\|$$

解答. 我们做最一般的证明，f 是向量值函数时，我们用不了微分中值定理，但是可以用拟微分中值定理：

$$g(z) = f(z) - Jf(x)z$$

满足在 $\overline{B_r(x^0)}$ 上处处可微，对任意的 $x, y \in \overline{B_r(x^0)}$ 那么存在 $\theta \in (0, 1)$:

$$\|g(y) - g(x)\| \leqslant \|Jg(\theta x + (1 - \theta)y)\|\|x - y\|$$

也就是：

$$\|f(y) - f(x) - Jf(x)(y - x)\| \leqslant \|Jf(\theta x + (1 - \theta)y) - Jf(x)\|\|x - y\|$$

根据 $f \in C^1(D)$, 任取 $\varepsilon > 0$, 存在 $\delta > 0$, 使对于 $x, y \in \overline{B_r(x^0)}$ 且 $\|x-y\| < \delta$，则 $\|Jf(y) - Jf(x)\| \leqslant \varepsilon$ ，而又有：

$$\|\theta x + (1 - \theta)y - x\| = \|(1 - \theta)(x - y)\| \leqslant \|x - y\| < \delta$$

使得：

$$\|Jf(\theta x + (1 - \theta)y) - Jf(x)\| \leqslant \varepsilon$$

进而，

$$\begin{aligned}\|f(y) - f(x) - Jf(x)(y - x)\| &\leqslant \|Jf(\theta x + (1 - \theta)y) - Jf(x)\|\|x - y\| \\ &\leqslant \varepsilon\|x - y\|\end{aligned}$$

题目 3 的注记.

我们终于证明了结论……

至于为什么能在 $\overline{B_r(x^0)}$ 上用拟微分中值定理，这时只需要认为它包含于一个处处可微的凸域 $D'(D'$ 包含于 $D)$ 中就行了，根据欧式空间的性质，这样的凸域是存在的.

题目 4. 设 $D \subset \mathbb{R}^n$ 为凸域，$f \in C^{1,1}(D)$, 即 f 处处可微，且存在常数 L, 使得

$$\|Jf(x) - Jf(y)\| \leqslant L\|x-y\|, \forall x, y \in D.$$

证明:

$$\|f(y) - f(x) - Jf(x)(y-x)\| \leqslant \frac{1}{2}L\|x-y\|^2, \forall x, y \in D.$$

解答.

我们还是证明最一般的情况：f 是 $C^{1,1}(D)$ 的向量值函数.

先推广定义,对一元向量值函数 $G : [a,b] \to \mathbb{R}^m$, 各分量为 $g_1, g_2, \cdots, g_m$, 其积分为:

$$\int_a^b G(t)dt = \begin{pmatrix} \int_a^b g_1(t)\mathrm{d}t \\ \int_a^b g_1(t)\mathrm{d}t \\ \vdots \\ \int_a^b g_1(t)\mathrm{d}t \end{pmatrix}$$

G 黎曼可积当且仅当各分量黎曼可积.

对于多元向量值函数 $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$,$f_1, f_2, \cdots .f_m$ 为分量, 都是多元函数,如果可微,那么 $f_i(x+t(y-x))$ 在 $t \in [0,1]$ 连续, 如果 $f_i'(x+t(y-x))$ 在 $t \in [0,1]$ 黎曼可积, 那么

$$f_i(y) - f_i(x) = \int_0^1 f_i'(x + t(y-x))(y-x)\mathrm{d}t$$

于是我们可以写出：

$$f(y)-f(x)=\int_0^1 Jf(x+t(y-x))(y-x)\mathrm{d}t$$

引理：如果一元向量值函数 $G\,[a.b]\to\mathbb{R}^m$ 连续，那么

$$\left\|\int_a^b G(t)\mathrm{d}t\right\|\leqslant\int_a^b\|G(t)\|\mathrm{d}t$$

对于本题：

$$\begin{aligned}\|f(y)-f(x)-Jf(x)(y-x)\| &= \left\|\int_0^1 Jf(x+t(y-x))(y-x)\mathrm{d}t-Jf(x)(y-x)\right\|\\ &= \left\|\int_0^1 Jf(x+t(y-x))(y-x)\mathrm{d}t-\int_0^1 Jf(x)(y-x)\mathrm{d}t\right\|\\ &= \left\|\int_0^1 [Jf(x+t(y-x))-Jf(x)](y-x)\mathrm{d}t\right\|\\ &\leqslant \int_0^1 \|[Jf(x+t(y-x))-Jf(x)](y-x)\|\mathrm{d}t\\ &\leqslant \int_0^1 \|Jf(x+t(y-x))-Jf(x)\|\cdot\|(y-x)\|\mathrm{d}t\\ &\leqslant \int_0^1 L\|t(y-x)\|\cdot\|(y-x)\|\mathrm{d}t\\ &= L\|(y-x)\|^2\int_0^1 t\mathrm{d}t\\ &= \frac{1}{2}L\|x-y\|^2\end{aligned}$$

题目 4 的注记.

引理：如果一元向量值函数 $G\,[a.b]\to\mathbb{R}^m$ 连续，那么

$$\left\|\int_a^b G(t)\mathrm{d}t\right\|\leqslant\int_a^b\|G(t)\|\mathrm{d}t$$

引理的证明: 范数是一个连续函数，那么 $\|G(t)\|$ 黎曼可积，任取 $\varepsilon>0$, 存在分划 $a=t_0<t_1<\cdots<t_{p-1}<t_p=b$, 使得

$$\left\|\int_a^b G(t)\mathrm{d}t-\sum_1^p G(t_i)(t_i-t_{i-1})\right\|<\varepsilon$$

以及

$$\left|\int_a^b \|G(t)\|\mathrm{d}t - \sum_1^p \|G(t_i)\|(t_i - t_{i-1})\right| < \varepsilon$$

于是，

$$\begin{aligned}\left\|\int_a^b G(t)\mathrm{d}t\right\| &\leqslant \left\|\sum_1^p G(t_i)(t_i - t_{i-1})\right\| + \varepsilon \\ &\leqslant \sum_1^p \|G(t_i)\|(t_i - t_{i-1}) + \varepsilon \\ &\leqslant \int_a^b \|G(t)\|\mathrm{d}t + 2\varepsilon\end{aligned}$$

根据 ε 的任意性，

$$\left\|\int_a^b G(t)\mathrm{d}t\right\| \leqslant \int_a^b \|G(t)\|\mathrm{d}t$$