2022秋季 学院路校区(凸优化与近似)
最优化方法　　第 1 次作业
提交日期：2022 年9 月27日周二课前

2022 年 9 月 18 日

1. 指出下面哪些集合是凸集. 是凸集的说明理由，不是凸集的也需要说明理由. 以下记题目所给集合为S.

   (1) $\{\boldsymbol{x} \in \mathbb{R}^2 : x_1 + i^2 x_2 \le 1, i = 1, \cdots, 10\}$

   (2) $\{\boldsymbol{x} \in \mathbb{R}^2 : x_1^2 + 2i x_2 x_2 + i^2 x_2^2 \le 1, i = 1, \cdots, 10\}$

   (3) $\{\boldsymbol{x} \in \mathbb{R}^2 : x_1^2 + i x_1 x_2 + i^2 x_2^2 \le 1, i = 1, \cdots, 10\}$

   (4) $\{\boldsymbol{x} \in \mathbb{R}^2 : x_1^2 + 5x_1x_2 + 4x_2^2 \le 1\}$

   (5) $\{\boldsymbol{x} \in \mathbb{R}^{10} : x_1^2 + 2x_2^2 + 3x_3^2 + \cdots + 10x_{10}^2 \le 2004x_1 - 2003x_2 + 2002x_3 - \cdots + 1996x_9 - 1995x_{10}\}$

   (6) $\{\boldsymbol{x} \in \mathbb{R}^2 : \exp\{x_1\} \le x_2\}$

   (7) $\{\boldsymbol{x} \in \mathbb{R}^2 : \exp\{x_1\} \ge x_2\}$

   (8) $\{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 = 1\}$

   (9) $\{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \le 1\}$

   (10) $\{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \ge 1\}$

   (11) $\{\boldsymbol{x} \in \mathbb{R}^n : \max_{i=1,\cdots,n} x_i \le 1\}$

   (12) $\{\boldsymbol{x} \in \mathbb{R}^n : \max_{i=1,\cdots,n} x_i \ge 1\}$

   (13) $\{\boldsymbol{x} \in \mathbb{R}^n : \max_{i=1,\cdots,n} x_i = 1\}$

   (14) $\{\boldsymbol{x} \in \mathbb{R}^n : \min_{i=1,\cdots,n} x_i \le 1\}$

   (15) $\{\boldsymbol{x} \in \mathbb{R}^n : \min_{i=1,\cdots,n} x_i \ge 1\}$

   (16) $\{\boldsymbol{x} \in \mathbb{R}^n : \min_{i=1,\cdots,n} x_i = 1\}$

2. (理解函数凸等价于它的上图凸. 这样，可以利用凸集研究凸函数的性质) 设 f 是定义在凸集 $C \subseteq \mathbb{R}^n$ 上的函数, 定义集合

$$[f, C] = \{(\boldsymbol{x}, r) \in \mathbb{R}^n \times \mathbb{R} : \boldsymbol{x} \in C, f(\boldsymbol{x}) \leq r\},$$

称该集合为 f 在 C 上的**上图**(epigraph). 证明 $f(\boldsymbol{x})$ 是凸函数当且仅当 $[f, C]$ 是凸集.

3. (Jensen不等式是研究中分析涉及凸函数不等式的重要工具) 设 C 是 $\mathbb{R}^n$ 中的凸集，$f(\cdot) : C \to \mathbb{R}$. 证明 f 是 C 上的凸函数当且仅当对任一整数 $k \geq 2$, $\boldsymbol{x}_i \in C, \theta_i \geq 0, i = 1, \cdots, k, \sum_{i=1}^{k} \theta_i = 1$ 蕴含着 $f\left(\sum_{i=1}^{k} \theta_i \boldsymbol{x}_i\right) \leq \sum_{i=1}^{k} \theta_i f(\boldsymbol{x}_i)$.

4. (复习并熟悉多元函数梯度和海塞矩阵的概念，特别注意(d)，其是非线性最小二乘问题的目标函数，所得结果是设计求解非线性最小二乘问题算法的基础) 确定下列 n 元函数的梯度向量和 Hessian 阵：

   (a) $\boldsymbol{a}^T \boldsymbol{x}$: $\boldsymbol{a}$ 是已知向量；

   (b) $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}$: $\boldsymbol{A}$ 是非对称已知矩阵；

   (c) $\frac{1}{2} \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} - \boldsymbol{b}^T \boldsymbol{x}$: $\boldsymbol{A}$ 是已知的对称矩阵，$\boldsymbol{b}$ 是已知向量；

   (d) $\boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{r}(\boldsymbol{x})$: $\boldsymbol{r}(\boldsymbol{x}) = (r_1(\boldsymbol{x}), \cdots, r_m(\boldsymbol{x}))^T$ 是依赖于 $\boldsymbol{x}$ 的 m 维向量，记 $\nabla \boldsymbol{r}^T$ 为 $\boldsymbol{A}^T$，它一般不是常量.

5. (理解梯度方向的性质) 假设在点 $\boldsymbol{x}'$ 有 $\boldsymbol{g}' \neq \boldsymbol{0}$，证明在所有单位向量 $\boldsymbol{p}^T \boldsymbol{p} = 1$ 中，函数沿方向向量 $\boldsymbol{p} = \boldsymbol{g}' / \|\boldsymbol{g}'\|_2$ 的斜率最大. 称该方向是函数的**最速上升**(steepest ascent)方向.

6. (应用无约束优化的最优性条件分析线性最小二乘问题) 考虑问题 $\min\limits_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$, 其中 $\boldsymbol{A}$ 是 $m \times n$ 矩阵, $\boldsymbol{b}$ 是 m 维向量，$m > n$.

   (a) 给出问题的几何解释.

   (b) 写出最优性的必要条件. 它也是一个充分条件吗？

   (c) 最优解唯一吗？理由是什么？

   (d) 你能给出最优解的一种闭合形式(解析式)吗？在满足题目所给信息下，允许规定任何你所需的假设.

   (e) 求解该问题, 其中

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

7. 利用凸函数的一阶刻画, 证明：

   (a) $(1-a)^t \le e^{-at},\ \forall a \in (0,1], \forall t > 0.$

   (b) $\ln(x+1) \le x,\ \forall x > -1.$

8. (当函数不可微时，可以考虑次梯度和次微分的概念. 利用这些概念刻画最优解，进而设计求解问题的算法.) 设 f 是定义在 $\mathbb{R}^n$ 上的函数. 如果向量 $\boldsymbol{p} \in \mathbb{R}^n$ 满足

$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \boldsymbol{p}^T(\boldsymbol{y}-\boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^n,$$

   称 $\boldsymbol{p}$ 是 f 在 $\boldsymbol{x}$ 处的**次梯度**(subgradient)； f 在 $\boldsymbol{x}$ 处所有次梯度的集合记为 $\partial f(\boldsymbol{x})$, 称为 f 在 $\boldsymbol{x}$ 处的**次微分**(subdifferential).

   (a) 证明：如果 f 是可微凸函数，那么 $\partial f(\boldsymbol{x}) = \{\nabla f(\boldsymbol{x})\}$.

   (b) 设 $f(x) = |x|$ ，求 $\partial f(x)$.

   (c) 设$f(x) = \max\{0, x\}$. 求 $\partial f(x)$.

9. (选做题) 请依次完成以下问题：

   (a) 设$\{f_i\}_{i\in I}$是$\mathbb{R}^n$上的凸函数族(可能不可数无穷多). 证明

$$f(x) := \sup_{i\in I} f_i(x)$$

   在$\mathbb{R}^n$上是凸函数. (提示：可用定义，或者第2 题的结论和凸集的交还是凸集证明)

   (b) 设$x \in \mathbb{R}^n$使得$\mathcal{A}(x) := \{i \in I : f(x) = f_i(x)\} \ne \emptyset$, 此即$f$的定义中，取到上确界的函数指标集. 证明

$$\left\{ \sum_{i\in I_k} \theta_i g_i : \forall k \in \mathbb{Z}_{++}, g_i \in \partial f_i(x), \theta_i \ge 0, \forall i \in I_k \subseteq \mathcal{A}(x), |I_k| = k, \sum_{i\in I_k} \theta_i = 1 \right\} \subseteq \partial f(x).$$

   并分别对函数$|x| = \max\{x, -x\}$和$\max\{0, x\}$应用该结论.

   (c) 证明凸函数的和是凸函数；凸函数的非负倍数是凸函数.

   (d) (Nesterov构造的世界最差函数) 已知$t < n$. 对于某$\gamma, \alpha \in \mathbb{R}_{++}$，考虑函数[1]

$$f(x) = \gamma \max_{1\le i\le t} x[i] + \tfrac{\alpha}{2}\|x\|^2.$$

   这里$\alpha > 0, \gamma > 0$是待定参数.

   - 由(a)和(c)说明函数f是凸函数.

[1]为了避免与迭代指标混淆，本节用$x[i]$ 表示x的第i个分量.

- 由(b)给出$\partial f(x)$的计算公式.

$$\partial f(x) = \alpha x + \gamma \mathrm{conv}\Big\{e_i \colon i \in \operatorname*{argmax}_{1\le j\le t} x[j]\Big\}.$$

- 已知$x_* \in \mathbb{R}^n$ 满足

$$x_*[i] = \begin{cases} -\frac{\gamma}{\alpha t} & \text{如果} 1 \le i \le t \\ 0 & \text{否则}. \end{cases}$$

  对于$t = 2$，请给出$\partial f(x_*)$的显式表示. 请问$0 \in \partial f(x_*)$吗？为什么？给出函数的极小点和极小值.