

最优化算法

孙天阳

中国科学技术大学数学科学学院

`tysun@mail.ustc.edu.cn`

2025 年 3 月 4 日

目录

目录	1
1 凸集	2
2 凸函数	3
3 最优化问题解的存在性	4
4 无约束可微问题的最优性理论	5
5 对偶理论	6
6 共轭函数	7
7 第一次作业	8

1 凸集

2 凸函数

在最优化领域, 经常会涉及对某个函数其中的一个变量取 \inf 或 \sup 的操作, 这导致函数的取值可能为无穷, 于是有了下面的定义

定义 2.1. 令 $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间, 则映射 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 称为广义实值函数.

定义 2.2. 给定广义实值函数 f 和非空集合 \mathcal{X} . 如果存在 $x \in \mathcal{X}$ 使得 $f(x) < +\infty$ 且对任意的 $x \in \mathcal{X}$ 有 $f(x) > -\infty$, 那么称函数 f 关于集合 \mathcal{X} 是适当的.

对于最优化问题 $\min f(x)$, 适当函数可以帮助我们去掉一些不感兴趣的函数, 从而在一个比较合理的函数类中考虑最优化问题. 规定适当函数的定义域为 $\text{dom } f = \{x | f(x) < +\infty\}$.

定义 2.3. 设 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, 称 $C_\alpha = \{x | f(x) \leq \alpha\}$ 为 f 的 α -下水平集.

定义 2.4. 设 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, 称

$$\text{epi } f = \{(x, t) \in \mathbb{R}^{n+1} | t \geq f(x)\}$$

为 f 的上方图. 若 $\text{epi } f$ 为闭集, 则称 f 为闭函数.

定义 2.5. 设 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, 若对任意的 $x \in \mathbb{R}^n$ 有 $\liminf_{y \rightarrow x} f(y) \geq f(x)$, 则称 $f(x)$ 为下半连续函数.

命题 2.1. 设 $f: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ 是适当函数, 则以下命题等价

- (1) $f(x)$ 是下半连续的
- (2) $f(x)$ 是闭函数
- (3) $f(x)$ 的任意 α -下水平集都是闭集

3 最优化问题解的存在性

考虑优化问题

$$\min f(x), \quad x \in \mathcal{X}$$

其中 $\mathcal{X} \subset \mathbb{R}^n$ 为可行域. 首先要考虑的是最优解的存在性, 其次是唯一性, 然后考虑如何求出最优解. 我们知道定义在紧集上的连续函数一定存在最大值点和最小值点, 但在许多实际问题中, 定义域可能不是紧的, 目标函数也不一定连续, 因此我们对此命题进行推广

定理 3.1. 设 \mathcal{X} 是闭集, 函数 $f: \mathcal{X} \rightarrow (-\infty, +\infty]$ 适当且闭, 且以下条件中任意一个成立,

- (1) $\text{dom } f := \{x \in \mathcal{X} : f(x) < +\infty\}$ 是有界的.
- (2) 存在一个常数 $\bar{\gamma}$ 使得下水平集 $C_{\bar{\gamma}} := \{x \in \mathcal{X} : f(x) \leq \bar{\gamma}\}$ 是非空且有界的.
- (3) 对于任何满足 $\|x^k\| \rightarrow +\infty$ 的点列 $\{x^k\} \subset \mathcal{X}$, 都有 $\lim_{k \rightarrow \infty} f(x^k) = +\infty$.

则函数 f 的最小值点集 $\{x \in \mathcal{X} | f(x) \leq f(y), \forall y \in \mathcal{X}\}$ 非空且紧.

证明.

- (2) 设 $t = \inf_{x \in \mathcal{X}} f(x) = -\infty$, 存在点列 $\{x^k\}_{k=1}^{\infty} \subset C_{\bar{\gamma}}$ 使得 $\lim_{k \rightarrow \infty} f(x^k) = -\infty$, 因为 $C_{\bar{\gamma}}$ 有界, 所以 $\{x^k\}_{k=1}^{\infty}$ 必有收敛子列, 不妨仍记为 $\{x^k\}_{k=1}^{\infty}$, 设它的序列极限是 x^* , 因为 \mathcal{X} 是闭集, 所以 $x^* \in \mathcal{X}$. 易知 $(x^k, f(x^k))$ 收敛于 (x^*, t) , 因为 $\text{epi } f$ 是闭集, 所以 $(x^*, t) \in \text{epi } f$, 所以 $f(x^*) \leq t = -\infty$, 这与 f 是适当的矛盾, 故 t 是有限值. 考察集合 $f^{-1}(t)$, 因为闭集的原像是闭集, 而 $f^{-1}(t) \subset C_{\bar{\gamma}}$ 是有界集, 所以是紧集.

- (1) 由 f 的适当性与条件 (1) 显然能推出条件 (2).
- (3) 假设某个下水平集无界, 由条件 (3) 显然能导出矛盾.

□

4 无约束可微问题的最优性理论

5 对偶理论

6 共轭函数

函数 f 的共轭函数定义为

$$f^*(y) = \sup_x (y^T x - f(x)).$$

我们来理解一下这个函数, 给定 y , 决定了关于 x 的以 y 为斜率的线性函数 $y^T x$, 比函数 $f(x)$ 高的最大距离, 就是 $f^*(y)$. 因此一个具有明显几何意义的结论是

$$f(x) \geq y^T x - f^*(y)$$

即函数 $f(x)$ 全部在直线 $y^T x - f^*(y)$ 的上方, 且这件事是刚好成立的. 所以或许可以将 $f^*(y)$ 诠释为 $f(x)$ 上以 y 为斜率的位置的切线的负截距. 所以变量 y 应该被理解为函数 $f(x)$ 的斜率. 回忆 Legendre 变换,

$$f^*(y) = y^T x - f(x),$$

其中将 x 通过隐式方程 $y = \nabla f(x)$ 视作关于 y 的函数.

7 第一次作业

问题 1 (判断集合是否是凸集)

1. 考虑这样点的集合, 这些点离给定点 x_0 比离给定集合 S 中的任何点都更近, 即集合 $\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\}$, where $S \subseteq \mathbb{R}^n$.
2. 记 $n \times n$ 的对称矩阵集合为 \mathbb{S}^n , 集合 $\{X \in \mathbb{S}^n \mid \lambda_{\min}(X) \geq 1\}$.

证明.

1.

$$\{x \mid \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for all } y \in S\} = \bigcap_{y \in S} \{x \mid \|x - x_0\|_2 \leq \|x - y\|_2\}$$

而凸集的任意交是凸集, 所以问题中集合是凸集.

2. 设 $X_1, X_2 \in \{X \in \mathbb{S}^n \mid \lambda_{\min}(X) \geq 1\}$, 设 $\theta \in [0, 1]$, 则

$$\lambda_{\min}(\theta X_1 + (1 - \theta)X_2) = \min_{\|x\|=1} x^T(\theta X_1 + (1 - \theta)X_2)x \geq \theta \lambda_{\min}X_1 + (1 - \theta)\lambda_{\min}X_2 \geq 1.$$

□

问题 2 (判断是否是凸函数)

1. 函数 $f(x) = \sum_{i=1}^r |x|_{[i]}$ 在 \mathbb{R}^n 上定义, 其中向量 $|x|$ 的分量满足 $|x|_i = |x_i|$ (即, $|x|$ 是 x 的每个分量的绝对值), 而 $|x|_{[i]}$ 是 $|x|$ 中第 i 大的分量. 换句话说, $|x|_{[1]} \geq |x|_{[2]} \geq \dots \geq |x|_{[n]}$ 是 x 的分量的绝对值按非增序排序.
2. 若 f, g 都是凸函数, 并且都非递减, 而且 f, g 函数值都是正的. 那么他们的乘积函数 $h = fg$ 是否为凸函数?

证明.

1. 考虑两个向量 x, y 满足 $x_i \leq y_i$, 则 $x_{[i]} \leq y_{[i]}$, 下证明之. 对指标进行轮换, 不妨设 y 已经降序排列. 设 $x_{[i]} = x_j$, 分类讨论 $j < i$ 和 $j \geq i$

- 当 $j \geq i$ 时 $x_{[i]} = x_j \leq y_j \leq y_i = y_{[i]}$.
- 当 $j < i$ 时, 考虑 $x_{[1]}, \dots, x_{[i-1]}$, 其中一定有一项 x_k 满足 $k \geq i$. 则

$$x_{[i]} = x_j < x_k \leq y_k \leq y_i = y_{[i]}.$$

$$\theta f(x) + (1 - \theta)f(y) = \sum_{i=1}^r \theta |x|_{[i]} + (1 - \theta)|y|_{[i]} \geq \sum_{i=1}^r |\theta x + (1 - \theta)y|_{[i]} = f(\theta x + (1 - \theta)y)$$

中间不等号成立的原因是

$$|\theta x + (1 - \theta)y|_j \leq \theta |x|_j + (1 - \theta)|y|_j$$

2. 按照定义验证 h 是凸函数.

$$\begin{aligned}
 h(\theta x + (1 - \theta)y) &= f(\theta x + (1 - \theta)y)g(\theta x + (1 - \theta)y) \\
 &\leq (\theta f(x) + (1 - \theta)f(y))(\theta g(x) + (1 - \theta)g(y)) \\
 &= \theta f(x)g(x) + (1 - \theta)f(y)g(y) + \theta(1 - \theta)(f(y) - f(x))(g(x) - g(y)) \\
 &\leq \theta f(x)g(x) + (1 - \theta)f(y)g(y) \\
 &= \theta h(x) + (1 - \theta)h(y)
 \end{aligned}$$

□

问题 3

对于最大分量函数 $f(x) = \max_{i=1, \dots, n} x_i, x \in \mathbb{R}^n$, 证明其共轭函数为

$$f^*(y) = \begin{cases} 0 & y_i \geq 0, \sum_i y_i = 1 \\ +\infty & \text{otherwise} \end{cases}.$$

证明. 当 $y_i \geq 0, \sum_i y_i = 1$ 时, $y^T x \leq f(x)$, 且等号在 $x = 0$ 处可以取到, 所以 $f^*(y) = 0$.

当存在 $y_i < 0$ 时, 构造 $x_i < 0$ 其余 $x_j = 0$, 则 $y^T x > 0$ 且可趋于无穷但 $f(x) = 0$.

当 $y_i \geq 0$ 且 $\sum_i y_i > 1$ 时, 构造 $x = (t, \dots, t)$ 其中 $t > 0$, 所以 $f^*(y) = +\infty$.

当 $y_i \geq 0$ 且 $\sum_i y_i < 1$ 时, 构造 $x = (t, \dots, t)$ 其中 $t < 0$, 所以 $f^*(y) = +\infty$.

□

问题 4 对于分式线性问题

$$\min f_0(x) = \frac{c^T x + d}{e^T x + f}, \quad \text{s.t. } Gx \leq h, Ax = b$$

其中 $\text{dom } f_0(x) = \{x \mid e^T x + f > 0\}$. 证明该问题等价于一个线性规划问题

$$\min c^T y + dz, \quad \text{s.t. } Gy \leq hz, Ay = bz, e^T y + fz = 1, z \geq 0.$$

证明. 在原问题中作变量代换 $x = y/z$, 其中要求 $z > 0$, 这样一来目标函数和约束变成

$$\min \frac{c^T y + dz}{e^T y + fz}, \quad \text{s.t. } Gy \leq hz, Ay = bz, z > 0$$

定义域为 $e^T y + fz > 0$. 因为分子分母同时乘大于零的数并不影响目标函数的值, 所以可以仅仅考虑 $e^T y + fz = 1$ 的情形. 即将问题修改为

$$\min c^T y + dz, \quad \text{s.t. } Gy \leq hz, Ay = bz, e^T y + fz = 1, z > 0.$$

如何证明最小值不在 $z = 0$ 时取到?

□