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 2 月 28 日

目录

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 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ 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 $\operatorname{dom} f = \{x | f(x) < +\infty\}$.

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

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

$$\operatorname{epi} f = \left\{(x,t) \in \mathbb{R}^{n+1} | t \geqslant f(x)\right\}$$

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

定义 2.5. 设 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, 若对任意的 $x \in \mathbb{R}^n$ 有 $\liminf\limits_{y \to x} f(y) \geqslant f(x)$, 则称 $f(x)$ 为下半连续函数.

命题 2.1. 设 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ 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 $f\colon \mathcal{X} \to (-\infty, +\infty]$ 适当且闭, 且以下条件中任意一个成立,

(1) $\operatorname{dom} f := \{x \in \mathcal{X}\colon f(x) < +\infty\}$ 是有界的.

(2) 存在一个常数 $\bar{\gamma}$ 使得下水平集 $C_{\bar{\gamma}} := \{x \in \mathcal{X}\colon f(x) \leqslant \bar{\gamma}\}$ 是非空且有界的.

(3) 对于任何满足 $\|x^k\| \to +\infty$ 的点列 $\left\{x^k\right\} \subset \mathcal{X}$, 都有 $\lim\limits_{k\to\infty} f(x^k) = +\infty$.

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

证明.

(2) 设 $t = \inf\limits_{x\in\mathcal{X}} f(x) = -\infty$, 存在点列 $\left\{x^k\right\}_{k=1}^{\infty} \subset C_{\bar{\gamma}}$ 使得 $\lim\limits_{k\to\infty} f(x^k) = -\infty$, 因为 $C_{\bar{\gamma}}$ 有界, 所以 $\left\{x^k\right\}_{k=1}^{\infty}$ 必有收敛子列, 不妨仍记为 $\left\{x^k\right\}_{k=1}^{\infty}$, 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 $\operatorname{epi} f$ 是闭集, 所以 $(x^*, t) \in \operatorname{epi} f$, 所以 $f(x^*) \leqslant t = -\infty$, 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.

$\square$

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

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

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

□

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，$|x|_{[1]} \geq |x|_{[2]} \geq \ldots \geq |x|_{[n]}$ 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 $x_i \leqslant y_i$, 则 $x_{[i]} \leqslant y_{[i]}$, 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 $j \geqslant i$
   - 当 $j \geqslant i$ 时 $x_{[i]} = x_j \leqslant y_j \leqslant y_i = y_{[i]}$.
   - 当 $j < i$ 时, 考虑 $x_{[1]}, \cdots, x_{[i-1]}$, 其中一定有一项 x_k 满足 $k \geqslant i$. 则
     $$x_{[i]} = x_j < x_k \leqslant y_k \leqslant y_i = y_{[i]}.$$

   $$\theta f(x) + (1-\theta)f(y) = \sum_{i=1}^{r} \theta|x|_{[i]} + (1-\theta)|y|_{[i]} \geqslant \sum_{i=1}^{r} |\theta x + (1-\theta)y|_{[i]} = f(\theta x + (1-\theta)y)$$
   中间不等号成立的原因是
   $$|\theta x + (1-\theta)y|_j \leqslant \theta|x|_j + (1-\theta)|y|_j$$
2.

□