

# 一种基于回溯策略的多目标加速邻近梯度算法\*

黄成志, 陈健, 唐莉萍

(重庆师范大学 重庆国家应用数学中心, 重庆 401331)

**摘要:**为了解决复合多目标优化问题的目标函数中光滑函数梯度 Lipschitz 常数未知的问题,在加速邻近梯度算法的基础上,提出了一种新的回溯策略。该策略通过构建与 Lipschitz 常数的估计序列满足某种等式关系的更新规则,允许估计序列以非递增方式进行更新。并在适当条件下,证明了该算法生成序列的所有聚点都是弱 Pareto 有效解。此外,利用效用函数得到该算法的次线性收敛率为  $O(1/k^2)$ 。数值实验表明,与未采用回溯策略的加速邻近梯度算法相比,新提出的算法在运行时间、迭代次数和函数评估次数方面均表现出明显优势。

**关键词:**多目标优化;加速邻近梯度算法;回溯策略;收敛速率

**中图分类号:**O221.6

**文献标志码:**A

**文章编号:**1672-6693(2025)05-0008-12

本文研究如下无约束复合多目标优化问题(UCMOP):

$$\min_{x \in \mathbf{R}^n} F(x) = (F_1(x), \dots, F_m(x))^T, \quad (1)$$

其中:  $F_i(x) = f_i(x) + g_i(x)$  ( $i=1, 2, \dots, m$ ), 且  $f_i: \mathbf{R}^n \rightarrow \mathbf{R}$  是可微的凸函数,  $g_i: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{+\infty\}$  是下半连续、真凸函数。

上述问题在机器学习和图像恢复等领域<sup>[1-3]</sup>中具有广泛应用,同时在多目标优化理论中也具有重要的理论价值<sup>[4-6]</sup>。例如,可利用罚函数方法将约束多目标优化问题转化为无约束的复合多目标优化问题进行求解<sup>[4]</sup>。当问题(1)中的目标函数为光滑函数与非光滑函数之和时, Tanabe 等人<sup>[6]</sup>首次提出了求解该类问题的有效方法——多目标邻近梯度法,迭代格式为:

$$d_k = \operatorname{argmin}_d \max_{i=1,2,\dots,m} \left\{ \langle \nabla f_i(x_k), d \rangle + g_i(x_k + d) - g_i(x_k) + \frac{l}{2} \|d\|^2 \right\},$$
$$x_{k+1} = x_k + \gamma_k d_k.$$

其中:  $k$  是迭代次数,  $d_k$  是迭代方向,  $l > 0$  是预先给定的常数,  $\gamma_k$  是步长。此方法融合了多目标最速下降算法<sup>[7]</sup>和多目标邻近点算法<sup>[8]</sup>的核心思想,对光滑函数部分  $(f_1, \dots, f_m)^T$  采用多目标最速下降算法进行梯度下降,而对非光滑函数部分  $(g_1, \dots, g_m)^T$  则通过多目标邻近点算法实现隐式梯度下降,进而通过求解一个极小极大子问题来获取更新方向,并设计适当的步长策略进行迭代,直到找到 Pareto 稳定点。

Tanabe 等人<sup>[9]</sup>通过引入效用函数,对多目标邻近梯度算法进行了深入研究,分析了该算法在不同情形下的收敛速率:非凸问题为  $O(\sqrt{1/k})$ 、凸问题为  $O(1/k)$  以及强凸问题为  $O(r^k)$ , 其中  $r \in (0, 1)$ 。为了进一步提升计算效率, Zhao 等人<sup>[10]</sup>提出了一种改进的 Armijo 线搜索策略,并证明了在非凸情况下,该算法所产生的序列的每个聚点(若存在)都是 Pareto 稳定点;在满足凸性条件的情况下,该算法所产生的序列收敛到问题(1)的弱 Pareto 最优解,并且达到了与文献[9]所提算法相同的收敛速率。此外, Chen 等人<sup>[11]</sup>提出了一种基于 Bregman 距离的多目标优化邻近梯度算法,并证明无论采用常数步长还是回溯步长,该算法生成的序列均能收敛至 Pareto 稳定

\* 收稿日期:2025-01-23 网络出版时间:2025-05-15T14:14

资助项目:国家自然科学基金——重大项目(No. 11991024), 重点项目(No. 12431010), 面上项目(No. 12171060); 重庆市自然科学基金——面上项目(No. NCAMC2022-MSXM01), 创新发展联合基金重点项目(No. CSTB2024NSCQ-LZX0140); 重庆市教育委员会科学技术研究计划——重大项目(No. KJZD-M202300504), 青年项目(No. KJQN202400520); 重庆师范大学博望学者青年拔尖人才项目; 重庆师范大学启动基金项目(No. 22XLB006)

第一作者简介:黄成志,男,研究方向为多目标优化, E-mail: huangczmaths@163.com; 通信作者简介:陈健,男,博士, E-mail: chenjian\_math@163.com

网络出版地址: <https://link.cnki.net/urlid/50.1165.N.20250507.1635.003>

点,且该算法具有与文献[9]所提算法相同的收敛速率。Chen 等人<sup>[12]</sup>则从多目标优化问题的个性特征(即目标函数之间的不平衡性)出发,提出了一种基于 Barzilai-Borwein 准则的邻近梯度多目标算法(简记为 BBPGMO),证明了 BBPGMO 拥有比文献[9]所提算法更快的收敛速率。

为了进一步提升多目标邻近梯度法的收敛速率,Tanabe 等人<sup>[13]</sup>将加速思想引入到多目标邻近梯度法框架中,提出了一种多目标加速邻近梯度算法(简记为 mFISTA),在凸性情形下证明了该算法的次线性收敛率为  $O(1/k^2)$ 。然而,与单目标优化问题的情况类似,在 mFISTA 的迭代中目标函数值可能会增加,子问题的不精确计算也可能导致算法发散。为了解决该问题,Nishimura 等人<sup>[14]</sup>通过对目标函数值施加单调性,提出了多目标单调加速近端梯度法,在凸性情形下证明了它与 mFISTA 具有相同的收敛率  $O(1/k^2)$ 。此外,Zhang 等人<sup>[15]</sup>基于 Tanabe 等人<sup>[13]</sup>提出的多目标邻近梯度法,结合 Nestorv 的加速技巧,提出了一种收敛速率快于  $O(1/k^2)$  的多目标加速邻近梯度算法。

尽管在 mFISTA 方面取得了一些重要进展<sup>[13-15]</sup>,但仍然存在一些不足:

1) 光滑函数梯度的真实 Lipschitz 常数往往难以获取,而 mFISTA 的子问题中需要预先给定光滑函数梯度的真实 Lipschitz 常数,这就使得子问题中常数的选取变得困难。选择的常数过小,可能会导致算法不收敛;选择的常数过大,又会导致算法步长过小,降低效率。

2) 估计 Lipschitz 常数的回溯策略可能导致算法求解效率低,甚至可能无法保证加速算法的收敛性。一般采用回溯策略来估计 Lipschitz 常数时,估计值呈单调递增趋势,使得在算法后期子问题的计算难度增加,明显降低了算法的求解速率;此外,多目标优化问题与单目标优化问题不同,由于无法保证  $\min_{i=1,2,\dots,m} (F_i(\mathbf{x}_k) - F_i(\mathbf{z}))$  的非负性,所以加速算法难以收敛。式中  $\{\mathbf{x}_k\}$  为加速算法产生的序列。

注意到,在单目标优化问题的加速算法中,Scheinberg 等人<sup>[16]</sup>提出了一种基于回溯策略的加速邻近梯度算法。该策略允许 Lipschitz 常数的估计值  $L_k$  在迭代过程中局部调整,即沿着迭代局部增加或减少;并且估计值  $L_k$  在相邻 2 步间存在某种等式关系,这种关系可以用于解决基于回溯策略的多目标加速邻近梯度算法在收敛性分析中面临的难题。受文献[16]的启发,本文针对目标函数中光滑函数梯度的 Lipschitz 常数未知的复合多目标优化问题,提出一种基于回溯策略的 mFISTA。该策略局部估计光滑函数梯度的 Lipschitz 常数,并允许估计值局部增加或减少;通过构建与文献[16]中类似的关于估计序列的某种等式关系,来获得算法的收敛性。此外,利用效用函数给出算法在凸性条件下的次线性收敛率为  $O(1/k^2)$ ,并通过数值实验将本文算法与 mFISTA<sup>[13]</sup> 进行比较。

## 1 预备知识

在  $n$  维欧式空间  $\mathbf{R}^n$  中,非负象限定义为:

$$\mathbf{R}_+^n := \{\mathbf{v} \in \mathbf{R}^n \mid v_i \geq 0, i=1, 2, \dots, n\} \subseteq \mathbf{R}^n,$$

并记  $\text{int } \mathbf{R}_+^n := \{\mathbf{v} \in \mathbf{R}^n \mid v_i > 0, i=1, 2, \dots, n\}$ 。对任意  $\mathbf{v}^1, \mathbf{v}^2 \in \mathbf{R}^n$ , 定义序关系为:

$$\mathbf{v}^1 \leq \mathbf{v}^2 \Leftrightarrow \mathbf{v}^2 - \mathbf{v}^1 \in \mathbf{R}_+^n,$$

$$\mathbf{v}^1 < \mathbf{v}^2 \Leftrightarrow \mathbf{v}^2 - \mathbf{v}^1 \in \text{int } \mathbf{R}_+^n,$$

$$\mathbf{v}^1 \not\leq \mathbf{v}^2 \Leftrightarrow \mathbf{v}^2 - \mathbf{v}^1 \notin \mathbf{R}_+^n.$$

$n$  维欧式空间  $\mathbf{R}^n$  上的单纯形定义为  $\Delta^n := \{\lambda \in \mathbf{R}_+^n \mid \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$ 。记  $\mathbf{e}_n = (1, \dots, 1)^\top \in \mathbf{R}^n$ 。

**定义 1** 设  $f: \mathbf{R}^n \rightarrow \mathbf{R}$  是一个连续可微函数。称函数  $f$  是  $L$ -光滑的,如果它的梯度是 Lipschitz 连续的,即存在 1 个常数  $L > 0$ ,使得对所有  $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ ,有以下不等式成立:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|,$$

其中:  $\nabla f(\mathbf{x})$  表示  $f$  在  $\mathbf{x}$  的梯度,  $\|\cdot\|$  是 Euclid 范数。

根据下降引理<sup>[17]</sup>,当函数  $f$  是  $L$ -光滑函数时,有如下重要不等式关系:

$$f(\mathbf{p}) - f(\mathbf{q}) \leq \langle \nabla f(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle + \frac{L}{2} \|\mathbf{p} - \mathbf{q}\|^2, \forall \mathbf{p}, \mathbf{q} \in \mathbf{R}^n.$$

因此,当问题(1)的目标函数中的  $f_i (i=1, 2, \dots, m)$  是  $L_i$ -光滑时,对任意的  $i \in \{1, 2, \dots, m\}$ ,有:

$$f_i(\mathbf{p}) - f_i(\mathbf{q}) \leq \langle \nabla f_i(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle + \frac{L_i}{2} \|\mathbf{p} - \mathbf{q}\|^2, \forall \mathbf{p}, \mathbf{q} \in \mathbf{R}^n.$$

若取  $L(f) = \max_{i=1,2,\dots,m} L_i$ , 进而有:

$$f_i(\mathbf{p}) - f_i(\mathbf{q}) \leq \langle \nabla f_i(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle + \frac{L(f)}{2} \|\mathbf{p} - \mathbf{q}\|^2, \forall \mathbf{p}, \mathbf{q} \in \mathbf{R}^n;$$

且有:

$$\begin{aligned} F_i(\mathbf{p}) - F_i(\mathbf{r}) &= f_i(\mathbf{p}) - f_i(\mathbf{q}) + g_i(\mathbf{p}) + f_i(\mathbf{q}) - F_i(\mathbf{r}) \leq \\ &\langle \nabla f_i(\mathbf{q}), \mathbf{p} - \mathbf{q} \rangle + g_i(\mathbf{p}) + f_i(\mathbf{q}) - F_i(\mathbf{r}) + \frac{L(f)}{2} \|\mathbf{p} - \mathbf{q}\|^2. \end{aligned} \quad (2)$$

下面介绍多目标优化问题(1)的(弱)Pareto 最优解的概念。

**定义 2** 如果不存在  $\mathbf{x} \in \mathbf{R}^n$  使得  $F(\mathbf{x}) < F(\mathbf{x}^*)$ , 则称  $\mathbf{x}^* \in \mathbf{R}^n$  是问题(1)的弱 Pareto 最优解, 记所有弱 Pareto 最优解构成的集合为  $X^*$ 。

**定义 3** 如果不存在  $\mathbf{x} \in \mathbf{R}^n$  使得  $F(\mathbf{x}) \leq F(\mathbf{x}^*)$  且  $F(\mathbf{x}) \neq F(\mathbf{x}^*)$ , 则称  $\mathbf{x}^* \in \mathbf{R}^n$  是问题(1)的 Pareto 最优解, 称所有 Pareto 最优解对应的目标函数值的集合为 Pareto 前沿。

定义效用函数  $u_0: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$  为:

$$u_0(\mathbf{x}) := \sup_{\mathbf{z} \in \mathbf{R}^n} \min_{i=1,2,\dots,m} [F_i(\mathbf{x}) - F_i(\mathbf{z})].$$

利用效用函数  $u_0$ , 可以刻画问题(1)的弱 Pareto 最优解。

**引理 1**<sup>[18]</sup> 对于所有  $\mathbf{x} \in \mathbf{R}^n$ , 有  $u_0(\mathbf{x}) \geq 0$ ; 此外,  $\mathbf{x}$  是问题(1)的弱 Pareto 最优解当且仅当  $u_0(\mathbf{x}) = 0$ 。

## 2 一种基于回溯策略的 mFISTA

本节基于文献[13]中的子问题框架并参考文献[16], 针对复合优化问题(1)建立  $t_k$  与  $L_k$  之间的等式关系, 由此提出一种基于回溯策略的加速邻近梯度算法。同时, 结合子问题的特殊形式, 给出  $t_k$  的重要性质, 以便于后续收敛性分析。本文均假设问题(1)的目标函数中的  $f_i (i=1, 2, \dots, m)$  是  $L_i$ -光滑的, 且记  $L(f) = \max_{i=1,2,\dots,m} L_i$ 。

首先, 算法的子问题构造为如下形式:

$$\min_{\mathbf{z} \in \mathbf{R}^n} \varphi_{L(f)}(\mathbf{z}; \mathbf{x}, \mathbf{y}), \quad (3)$$

其中:  $\mathbf{x} \in \text{dom } F, \mathbf{y} \in \mathbf{R}^n$ , 且有:

$$\varphi_{L(f)}(\mathbf{z}; \mathbf{x}, \mathbf{y}) := \max_{i=1,2,\dots,m} [\langle \nabla f_i(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + g_i(\mathbf{z}) + f_i(\mathbf{y}) - F_i(\mathbf{x})] + \frac{L(f)}{2} \|\mathbf{z} - \mathbf{y}\|^2. \quad (4)$$

由于  $g_i (i=1, 2, \dots, m)$  是凸函数, 所以函数  $\mathbf{z} \mapsto \varphi_{L(f)}(\mathbf{z}; \mathbf{x}, \mathbf{y})$  是强凸的。由此可得, 子问题(1)具有唯一的最优解  $p_{L(f)}(\mathbf{x}, \mathbf{y})$ , 且最优值为  $\theta_{L(f)}(\mathbf{x}, \mathbf{y})$ , 即:

$$p_{L(f)}(\mathbf{x}, \mathbf{y}) := \operatorname{argmin}_{\mathbf{z} \in \mathbf{R}^n} \varphi_{L(f)}(\mathbf{z}; \mathbf{x}, \mathbf{y}), \theta_{L(f)}(\mathbf{x}, \mathbf{y}) := \min_{\mathbf{z} \in \mathbf{R}^n} \varphi_{L(f)}(\mathbf{z}; \mathbf{x}, \mathbf{y}). \quad (5)$$

此外, 根据子问题(1)的最优性条件: 对于所有  $\mathbf{x} \in \text{dom } F$  和所有  $\mathbf{y} \in \mathbf{R}^n$ , 存在  $\eta(\mathbf{x}, \mathbf{y}) \in \partial g(p_{L(f)}(\mathbf{x}, \mathbf{y}))$  和一个 Lagrange 乘子  $\lambda(\mathbf{x}, \mathbf{y}) \in \mathbf{R}^m$ , 使得:

$$\sum_{i=1}^m \lambda_i(\mathbf{x}, \mathbf{y}) [\nabla f_i(\mathbf{y}) + \eta_i(\mathbf{x}, \mathbf{y})] = -L(f) [p_{L(f)}(\mathbf{x}, \mathbf{y}) - \mathbf{y}], \quad (6)$$

并且对所有  $j \notin \mathcal{I}(\mathbf{x}, \mathbf{y}), \mathcal{I}(\mathbf{x}, \mathbf{y}) := \operatorname{argmin}_{i=1,2,\dots,m} [\langle \nabla f_i(\mathbf{x}), p_{L(f)}(\mathbf{x}, \mathbf{y}) - \mathbf{y} \rangle + g_i(p_{L(f)}(\mathbf{x}, \mathbf{y})) + f_i(\mathbf{y}) - F_i(\mathbf{x})]$ , 有:

$$\lambda(\mathbf{x}, \mathbf{y}) \in \Delta^m, \lambda_j(\mathbf{x}, \mathbf{y}) = 0, \quad (7)$$

其中:  $\Delta^m$  表示单纯形。

注意到, 通过在式(4)中取  $\mathbf{z} = \mathbf{y}$ , 对于所有  $\mathbf{x} \in \text{dom } F$  和所有  $\mathbf{y} \in \mathbf{R}^n$ , 有:

$$\theta_{L(f)}(\mathbf{x}, \mathbf{y}) \leq \varphi_{L(f)}(\mathbf{y}; \mathbf{x}, \mathbf{y}) = \max_{i=1,2,\dots,m} \{F_i(\mathbf{y}) - F_i(\mathbf{x})\}.$$

此外, 从式(2)中取  $\mathbf{p} = \mathbf{z}, \mathbf{q} = \mathbf{y}, \mathbf{r} = \mathbf{x}$ , 对所有  $\mathbf{x} \in \text{dom } F$  和  $\mathbf{y} \in \mathbf{R}^n$  可得:

$$\theta_{L(f)}(\mathbf{x}, \mathbf{y}) \geq \max_{i=1,2,\dots,m} \{F_i(p_{L(f)}(\mathbf{x}, \mathbf{y})) - F_i(\mathbf{x})\}.$$

借助文献[13]的性质 4.1, 可通过  $p_{L(f)}$  和  $\theta_{L(f)}$  来刻画弱 Pareto 最优性。

**引理 2**<sup>[13]</sup> 若  $p_{L(f)}(\mathbf{x}, \mathbf{y})$  和  $\theta_{L(f)}(\mathbf{x}, \mathbf{y})$  如式(5)所定义, 则以下 3 个条件等价:

- 1)  $\mathbf{y} \in \mathbf{R}^n$  是问题(1)的弱 Pareto 最优解;
- 2) 存在某个  $\mathbf{x} \in \mathbf{R}^n$ , 使得  $p_{L(f)}(\mathbf{x}, \mathbf{y}) = \mathbf{y}$ ;
- 3) 存在某个  $\mathbf{x} \in \mathbf{R}^n$ , 使得  $\theta_{L(f)}(\mathbf{x}, \mathbf{y}) = \max_{i=1,2,\dots,m} [F_i(\mathbf{y}) - F_i(\mathbf{x})]$ 。

引理 2 表明可以使用“ $L(f) \|p_{L(f)}(\mathbf{x}, \mathbf{y}) - \mathbf{y}\|_\infty < \epsilon$  对某些  $\epsilon > 0$  成立”作为算法停止条件。此外, 与文献[16]类似, 本文采用满足  $\beta L_k \leq L(f)$  的序列  $\{L_k\}$  来寻找合适的  $L(f)$  估计值。

下面给出该算法的具体内容。

**算法 1** 基于回溯策略的多目标加速邻近梯度算法(mFISTA\_backtracking)

步骤 0, 设置初始参数: 初始点  $\mathbf{x}_{-1} = \mathbf{x}_0 = \mathbf{y}_0 \in \mathbf{R}^n$ , Lipschitz 初始估计值  $L_0 > 0$ , Lipschitz 估计值更新参数  $\beta > 1$ , FISTA 序列初始值  $t_{-1} = 0, t_0 = 1$ , Lipschitz 常数比值参数  $\omega_0 = 1$ , 误差界限  $\epsilon > 0$ ,  $m$  维单位向量  $\mathbf{e}_m$ , 最大更新次数  $K$ 。令  $k := 0$ 。

步骤 1, 令  $L_k = L_{k-1}/\beta$ 。

步骤 2, 计算梯度  $\nabla f_i(\mathbf{y}_k), i=1, 2, \dots, m$ , 以及  $\hat{\mathbf{x}}_k = p_{L_k}(\mathbf{x}_k, \mathbf{y}_k)$ 。

步骤 3, 若  $F(\hat{\mathbf{x}}_k) \preceq \langle JF(\mathbf{y}_k), \hat{\mathbf{x}}_k - \mathbf{y}_k \rangle + F(\mathbf{y}_k) + \frac{L_k}{2} \|\hat{\mathbf{x}}_k - \mathbf{y}_k\|^2 \mathbf{e}_m$ , 则:

$$\begin{aligned} L_k &= \beta L_k, \\ \omega_{k-1} &= \beta \omega_{k-1}, \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4\omega_k t_k^2}}{2}, \\ \theta_{k+1} &= \frac{t_k - 1}{t_{k+1}}, \\ \mathbf{y}_{k+1} &= \mathbf{x}_k + \theta_{k+1}(\mathbf{x}_k - \mathbf{x}_{k-1}), \end{aligned} \quad (8)$$

其中: 对于所有  $k \geq 1$ ,  $\omega_k = L_{k+1}/L_k$ , 且  $L_k$  是每次迭代中  $L_f$  的估计值, 返回步骤 2; 否则  $\mathbf{x}_{k+1} = \hat{\mathbf{x}}_k$ 。

步骤 4, 若  $L_k \|p_{L_k}(\mathbf{x}_k, \mathbf{y}_k) - \mathbf{y}_k\|_\infty < \epsilon$ , 则输出弱 Pareto 最优解  $\mathbf{x}_k$ 。否则, 令  $k := k + 1$ , 返回步骤 1。

**假设 1** 设  $X^*$  是弱 Pareto 最优解的集合, 且  $\mathcal{L}_F(c) := \{\mathbf{x} \in \mathbf{R}^n \mid F(\mathbf{x}) \leq c\}$ , 则对于任意  $\mathbf{x} \in \mathcal{L}_F(F(\mathbf{x}_0))$ ,  $k \geq 0$ , 存在  $\mathbf{x}^* \in X^*$ , 使得  $F(\mathbf{x}^*) \leq F(\mathbf{x})$ , 且有下式成立:

$$C := \sup_{F^* \in F(X^* \cap \mathcal{L}_F(F(\mathbf{x}_0)))} \inf_{\mathbf{z} \in F^{-1}(F^*)} \|\mathbf{x}_0 - \mathbf{z}\|^2 < \infty.$$

现在给出序列  $\{t_k\}$  的性质, 便于后续收敛性证明。

**定理 1** 设  $\{t_k\}$  由式(8)定义。那么, 对于所有  $k \geq 1$ , 以下性质成立:

- 1)  $\frac{1}{2} + \sqrt{\omega_k} t_k \leq t_{k+1} \leq (1 + \sqrt{\omega_k}) t_k$ ;
- 2)  $L_{k+1}^{-1} t_{k+1} (t_{k+1} - 1) = L_k^{-1} t_k^2$ ;
- 3) 对于任意  $k \in \mathbf{N}$ , 有  $L_k^{-1} t_k^2 \geq \frac{(k+1)^2}{4\beta L(f)}$ 。

**证明** 首先, 由不等式关系:

$$\frac{1}{2} + \sqrt{(L_{k+1}/L_k)} t_k \leq \frac{1 + \sqrt{1 + 4(L_{k+1}/L_k) t_k^2}}{2} \leq 1 + \sqrt{(L_{k+1}/L_k)} t_k,$$

可得结论 1) 成立。

其次, 代入  $t_k$  的定义可得如下等式:

$$\begin{aligned} L_{k+1}^{-1} t_{k+1} (t_{k+1} - 1) &= L_{k+1}^{-1} \left( \frac{2 + 2\sqrt{1 + 4(L_{k+1}/L_k) t_k^2} + 4(L_{k+1}/L_k) t_k^2 - 2 - 2\sqrt{1 + 4(L_{k+1}/L_k) t_k^2}}{4} \right) = \\ &= L_{k+1}^{-1} \frac{4(L_{k+1}/L_k) t_k^2}{4} = L_k^{-1} t_k^2. \end{aligned}$$

于是结论 2) 成立。

最后, 由结论 1) 可得  $t_{k+1} \geq \frac{1}{2} + \sqrt{L_{k+1}/L_k} t_k$ 。在不等式两边同时乘  $\sqrt{1/L_{k+1}}$  得:

$$\sqrt{1/L_{k+1}} t_{k+1} \geq \frac{\sqrt{1/L_{k+1}}}{2} + \sqrt{1/L_k} t_k。$$

反复使用该不等式, 有  $\sqrt{1/L_{k+1}} t_{k+1} \geq \sum_{i=1}^{k+1} \frac{\sqrt{1/L_i}}{2}$ 。又由  $\beta L_i \leq L(f)$  可得  $L_k^{-1} t_k^2 \geq \frac{(k+1)^2}{4\beta L(f)}$ 。证毕

### 3 收敛性分析

本节主要在假设 1 下, 证明算法的收敛速率为  $O(1/k^2)$ 。为此, 首先介绍在文献[13]中用于证明收敛速率的辅助序列  $\sigma_k(\mathbf{z})$  和  $\rho_k(\mathbf{z})$ : 对于  $k \geq 0$ , 序列  $\sigma_k: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{-\infty\}$  和  $\rho_k: \mathbf{R}^n \rightarrow \mathbf{R}$  的定义如下:

$$\sigma_k(\mathbf{z}) := \min_{i=1,2,\dots,m} [F_i(\mathbf{x}_k) - F_i(\mathbf{z})], \rho_k(\mathbf{z}) := t_{k-1} \mathbf{x}_k - (t_{k-1} - 1) \mathbf{x}_{k-1} - \mathbf{z}。 \quad (9)$$

为了便于后续的证明, 给出序列  $\sigma_k(\mathbf{z})$  的一个重要性质。

**性质 1** 若  $\sigma_k(\mathbf{z})$  如式(9)定义, 则有:

$$\sigma_k(\mathbf{z}) - \sigma_{k+1}(\mathbf{z}) \geq -\frac{L_k}{2} [2\langle \mathbf{y}_k - \mathbf{x}_{k+1}, \mathbf{y}_k - \mathbf{x}_k \rangle + \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2], \quad (10)$$

以及

$$\sigma_{k+1}(\mathbf{z}) \leq \frac{L_k}{2} [2\langle \mathbf{y}_k - \mathbf{x}_{k+1}, \mathbf{y}_k - \mathbf{z} \rangle - \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2]。 \quad (11)$$

**证明** 首先, 由式(6)和(7)可得, 存在  $\eta(\mathbf{x}_k, \mathbf{y}_k) \in \partial g(\mathbf{x}_{k+1})$  和 Lagrange 乘子  $\lambda_i(\mathbf{x}_k, \mathbf{y}_k) \in \mathbf{R}^m$  满足子问题(3)的 KKT 条件, 于是有:

$$\sigma_{k+1}(\mathbf{z}) = \min_{i=1,2,\dots,m} [F_i(\mathbf{x}_{k+1}) - F_i(\mathbf{z})] \leq \sum_{i=1}^m \lambda_i(\mathbf{x}_k, \mathbf{y}_k) [F_i(\mathbf{x}_{k+1}) - F_i(\mathbf{z})]。$$

随后由下降引理可进一步将上式推广:

$$\sigma_{k+1}(\mathbf{z}) \leq \sum_{i=1}^m \lambda_i(\mathbf{x}_k, \mathbf{y}_k) [\langle \nabla f_i(\mathbf{y}_k), \mathbf{x}_{k+1} - \mathbf{y}_k \rangle + g_i(\mathbf{x}_{k+1}) + f_i(\mathbf{y}_k) - F_i(\mathbf{z})] + \frac{L_k}{2} \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2,$$

因此, 由  $f_i$  和  $g_i$  的凸性可得:

$$\sigma_{k+1}(\mathbf{z}) \leq \sum_{i=1}^m \lambda_i(\mathbf{x}_k, \mathbf{y}_k) [\langle \nabla f_i(\mathbf{y}_k) + \eta_i(\mathbf{x}_k, \mathbf{y}_k), \mathbf{x}_{k+1} - \mathbf{z} \rangle] + \frac{L_k}{2} \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2。$$

由式(6)和  $\mathbf{x}_{k+1} = p_{L_k}(\mathbf{x}_k, \mathbf{y}_k)$  可得:

$$\sigma_{k+1}(\mathbf{z}) \leq \frac{L_k}{2} [2\langle \mathbf{y}_k - \mathbf{x}_{k+1}, \mathbf{y}_k - \mathbf{z} \rangle - \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2]。$$

另一方面, 由  $\sigma_k(\mathbf{z})$  的定义可得:

$$\sigma_k(\mathbf{z}) - \sigma_{k+1}(\mathbf{z}) \geq -\max_{i=1,2,\dots,m} [F_i(\mathbf{x}_{k+1}) - F_i(\mathbf{x}_k)]。$$

根据下降引理有:

$$\begin{aligned} & \sigma_k(\mathbf{z}) - \sigma_{k+1}(\mathbf{z}) \geq \\ & -\sum_{i=1}^m \lambda_i(\mathbf{x}_k, \mathbf{y}_k) [\langle \nabla f_i(\mathbf{y}_k), \mathbf{x}_{k+1} - \mathbf{y}_k \rangle + g_i(\mathbf{x}_{k+1}) + f_i(\mathbf{y}_k) - F_i(\mathbf{x}_k)] - \frac{L_k}{2} \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2。 \end{aligned}$$

又由如下等式:

$$\|\mathbf{b} - \mathbf{a}\|^2 + 2\langle \mathbf{b} - \mathbf{a}, \mathbf{a} - \mathbf{c} \rangle = \|\mathbf{b} - \mathbf{c}\|^2 - \|\mathbf{a} - \mathbf{c}\|^2, \forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{R}^n, \quad (12)$$

和  $f_i$  和  $g_i$  的凸性, 可以得到:

$$\sigma_k(\mathbf{z}) - \sigma_{k+1}(\mathbf{z}) \geq \frac{L_k}{2} [2\langle \mathbf{x}_{k+1} - \mathbf{y}_k, \mathbf{y}_k - \mathbf{x}_k \rangle - \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2]。 \quad \text{证毕}$$

**引理 3**<sup>[13]</sup> 算法 1 生成的序列  $\{\mathbf{x}_k\}$  满足对于所有  $i=1, 2, \dots, m$  和  $k \geq 0$ , 有:

$$F_i(\mathbf{x}_k) \leq F_i(\mathbf{x}_0)。$$

接下来,利用性质 1 的不等式来证明算法的收敛速率。

**定理 2** 假设  $\{\mathbf{x}_k\}$  和  $\{\mathbf{y}_k\}$  是算法 1 生成的序列,对于任何  $\mathbf{z} \in \mathbf{R}^n$ , 都有:

$$u_0(\mathbf{x}_k) \leq \frac{4\beta L(f)C}{(k+1)^2}.$$

**证明** 首先,将式(10)乘  $(t_{k+1}-1)$ ,将式(11)乘  $t_{k+1}$ ,然后将两者相加,得到:

$$2L_k^{-1}t_k(t_k-1)\sigma_k(\mathbf{z}) - 2L_k^{-1}t_k^2\sigma_{k+1}(\mathbf{z}) \geq 2t_k \langle \mathbf{x}_{k+1} - \mathbf{y}_k, t_k \mathbf{y}_k - (t_k-1)\mathbf{x}_k - \mathbf{z} \rangle + t_k^2 \|\mathbf{x}_{k+1} - \mathbf{y}_k\|^2.$$

结合式(12)及定理 1 的结论 2),得到:

$$2L_{k-1}^{-1}t_{k-1}^2\sigma_k(\mathbf{z}) - 2L_k^{-1}t_k^2\sigma_{k+1}(\mathbf{z}) \geq \|\rho_{k+1}(\mathbf{z})\|^2 - \|\rho_k(\mathbf{z})\|^2.$$

反复使用该不等式并累加所得,有:

$$2L_k^{-1}t_k^2\sigma_{k+1}(\mathbf{z}) + \|\rho_{k+1}(\mathbf{z})\|^2 \leq 2L_0^{-1}t_0^2\sigma_1(\mathbf{z}) + \|\rho_1(\mathbf{z})\|^2 \leq \|\mathbf{x}_0 - \mathbf{z}\|^2.$$

又结合定理 1 的结论 3)可得:

$$\sigma_{k+1}(\mathbf{z}) \leq \frac{4\beta L(f)}{(k+1)^2} \|\mathbf{x}_0 - \mathbf{z}\|^2.$$

与文献[9]中定理 5.2 的论证类似可得:

$$u_0(\mathbf{x}_k) \leq \frac{4\beta L(f)C}{(k+1)^2}.$$

故算法的收敛速率为  $O(1/k^2)$ 。

证毕

根据上述收敛性分析可以得出以下结论。

**推论 1** 若假设 1 成立,则由算法 1 所得的序列  $\{\mathbf{x}_k\}$  的每个聚点都是弱 Pareto 最优解。若水平集  $\mathcal{L}_F(F(\mathbf{x}^0))$  有界,则  $\{\mathbf{x}_k\}$  必定存在聚点,且所有聚点均为弱 Pareto 最优解。此外,若每个  $F_i$  都是严格凸的,则聚点将是 Pareto 最优解。

## 4 数值实验

本节将展示算法在不同问题上的数值结果。所有数值实验均使用 Python 3.10 实现,并在一台配置为 12 代 Intel(R) Core(TM) i7-12700H CPU @ 2.70 GHz 和 16 GB 内存的个人电脑上执行。对于算法,设置  $L_0=1$ ,  $\beta=2$  以及误差界限  $\epsilon=10^{-3}$ ,以确保算法在有限的迭代次数后终止。此外,最大迭代次数被设置为 1 000 次。

为了便于比较,表 1 列出了所有测试问题的详细信息,用 L1 表示  $g(\mathbf{x}) = \frac{1}{n} \|\mathbf{x}\|_1$  的情况,用 0 表示  $g(\mathbf{x}) = 0$  的情况。表 1 的第 2 列和第 3 列分别给出  $f$  和  $g$  的形式,第 4 列和第 5 列分别给出变量和目标函数的维度,第 6 列和第 7 列分别给出变量  $\mathbf{x}$  的下界和上界,第 8 列给出测试问题的来源。

表 1 测试问题的详细信息

Tab. 1 Test problems

序号	$f$	$g$	$n$	$m$	下界	上界	信息来源
1	BK1	L1	2	2	(-5, -5)	(10, 10)	文献[19]
2	Deb	L1	2	2	(0.1, 0.1)	(1, 1)	文献[19]
3	DD1	L1	5	2	(-20, ..., -20)	(20, ..., 20)	文献[20]
4	FF1	L1	2	2	(-1, -1)	(1, 1)	文献[19]
5	FF1	0	2	2	(-1, -1)	(1, 1)	文献[19]
6	Hil1	L1	2	2	(0, 0)	(1, 1)	文献[21]
7	Hil1	0	2	2	(0, 0)	(1, 1)	文献[21]
8	Imb2	L1	2	2	(-2, -2)	(2, 2)	文献[22]
9	Imb2	0	2	2	(-2, -2)	(2, 2)	文献[22]
10	JOS1	L1	2	2	(-5, -5)	(5, 5)	文献[23]
11	MHHM1	0	1	3	0	1	文献[19]

续表 1

序号	$f$	$g$	$n$	$m$	下界	上界	信息来源
12	MHHM2	0	2	3	(0,0)	(1,1)	文献[19]
13	PNR	L1	2	2	(-2,-2)	(2,2)	文献[24]
14	SP1	L1	2	2	(2,-2)	(3,3)	文献[19]
15	VFM1	0	2	3	(-2,-2)	(2,2)	文献[19]
16	VU1	L1	2	2	(-3,-3)	(3,3)	文献[19]
17	WIT1	L1	2	2	(-2,-2)	(2,2)	文献[25]
18	WIT2	L1	2	2	(-2,-2)	(2,2)	文献[25]
19	WIT3	L1	2	2	(-2,-2)	(2,2)	文献[25]
20	WIT4	L1	2	2	(-2,-2)	(2,2)	文献[25]
21	WIT5	L1	2	2	(-2,-2)	(2,2)	文献[25]
22	WIT6	L1	2	2	(-2,-2)	(2,2)	文献[25]

接下来,将本文算法 1 记为 mFISTA\_backtracking,文献[13]的算法记为 mFISTA,进行性能比较。对于每个问题,使用相同的初始点进行 200 次实验,以评估不同算法的表现。初始点在指定的下限和上限范围内随机选择。经过 200 次实验后,得到每个问题的平均运行时间、平均迭代次数以及平均函数评估次数(表 2)。

表 2 测试问题的运行记录结果

Tab. 2 The results of test problems

问题序号	mFISTA_backtracking			mFISTA		
	平均运行时间/s	平均迭代次数	平均函数评估次数	平均运行时间/s	平均迭代次数	平均函数评估次数
1	<b>0.03</b>	<b>20.20</b>	<b>41.54</b>	0.91	516.98	723.91
2	<b>0.02</b>	<b>9.88</b>	<b>26.21</b>	0.16	43.72	152.45
3	<b>0.05</b>	<b>36.38</b>	<b>60.68</b>	0.55	962.53	1 015.19
4	<b>0.30</b>	<b>182.02</b>	<b>360.12</b>	0.45	860.86	892.27
5	<b>0.01</b>	<b>9.17</b>	<b>16.51</b>	0.01	13.33	13.48
6	<b>0.53</b>	<b>381.64</b>	<b>525.94</b>	1.15	745.61	793.16
7	<b>0.78</b>	<b>682.80</b>	<b>708.15</b>	1.05	712.35	725.64
8	<b>0.11</b>	<b>65.70</b>	<b>185.05</b>	0.33	308.99	390.55
9	<b>0.47</b>	<b>149.83</b>	<b>353.61</b>	0.58	770.31	785.12
10	<b>0.03</b>	<b>19.21</b>	<b>38.22</b>	1.11	133.55	646.36
11	<b>0.01</b>	<b>5.62</b>	<b>5.62</b>	0.01	8.93	8.93
12	<b>0.01</b>	<b>6.47</b>	<b>6.47</b>	0.01	11.61	11.61
13	<b>0.02</b>	<b>8.06</b>	<b>22.45</b>	0.78	457.51	529.57
14	<b>0.02</b>	<b>9.80</b>	<b>21.59</b>	0.81	410.57	565.00
15	<b>0.01</b>	<b>6.45</b>	<b>6.45</b>	0.14	131.02	131.02
16	<b>0.07</b>	<b>47.74</b>	<b>97.19</b>	2.74	746.55	992.39
17	<b>0.30</b>	<b>228.22</b>	<b>463.94</b>	0.36	477.78	486.43
18	<b>0.15</b>	<b>112.04</b>	<b>231.06</b>	0.23	304.01	314.27
19	<b>0.11</b>	<b>86.39</b>	<b>176.88</b>	0.16	205.30	213.71
20	<b>0.04</b>	<b>27.92</b>	<b>57.90</b>	0.08	103.84	108.51
21	<b>0.03</b>	<b>23.89</b>	<b>49.17</b>	0.04	56.57	59.32
22	<b>0.01</b>	<b>5.53</b>	<b>12.82</b>	0.01	12.80	12.80

注:加粗的数据表示对每个问题运行 200 次所得平均值的最优值。

从表 2 可以看出,mFISTA\_backtracking 在运行时间、迭代次数和函数评估次数上均表现出明显的优势。这表明,采用本文提出的回溯策略能够有效解决 Lipschitz 常数未知情况下过大估计量导致的较小步长问题,从而有效提高计算效率。

2 种算法在不同测试问题中获得的 Pareto 前沿如图 1~图 15 所示。

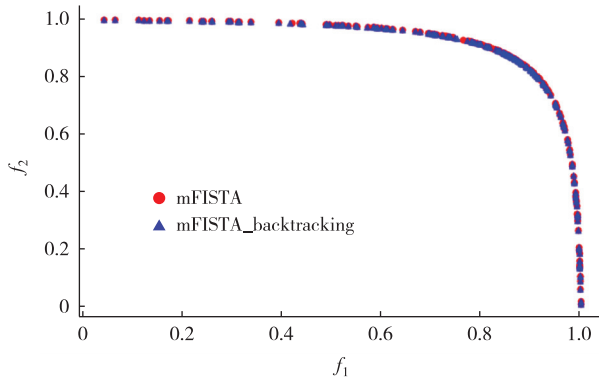


图 1 2 种算法获得的 FF1&0 问题的 Pareto 前沿对比  
Fig. 1 The comparison of Pareto fronts of the FF1&0 problem obtained by two algorithms

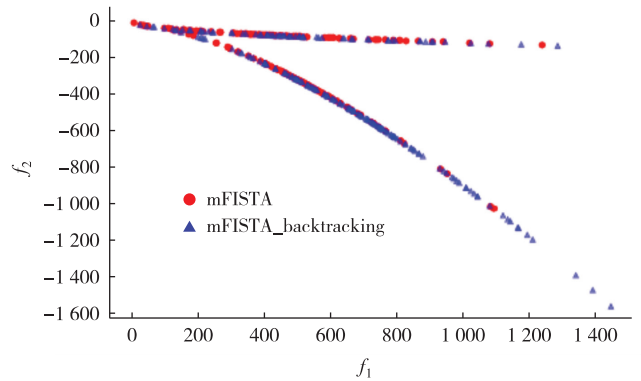


图 2 2 种算法获得的 DD1&0 问题的 Pareto 前沿对比  
Fig. 2 The comparison of Pareto fronts of the Hill&0 problem obtained by two algorithms

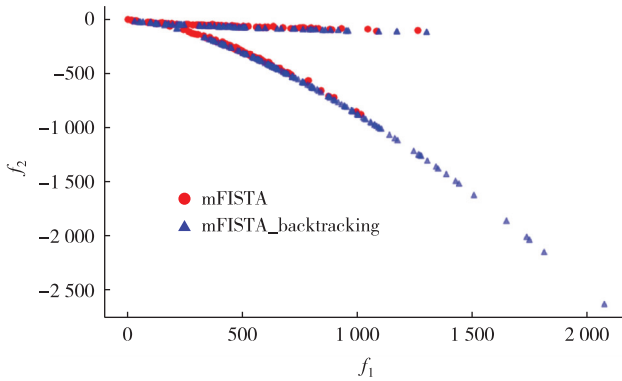


图 3 2 种算法获得的 DD1&L1 问题的 Pareto 前沿对比  
Fig. 3 The comparison of Pareto fronts of the DD1&L1 problem obtained by two algorithms

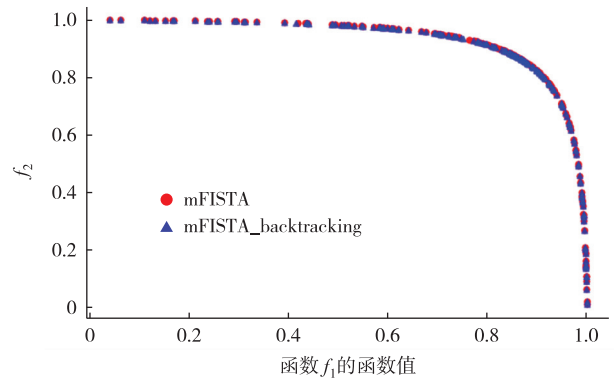


图 4 2 种算法获得的 FF1&L1 问题的 Pareto 前沿对比  
Fig. 4 The comparison of Pareto fronts of the FF1&L1 problem obtained by two algorithms

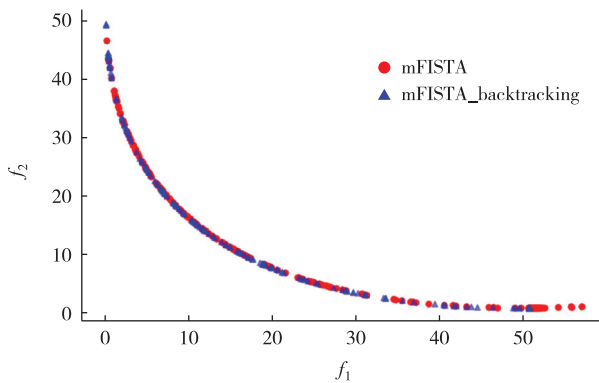


图 5 2 种算法获得的 BK1&L1 问题的 Pareto 前沿对比  
Fig. 5 The comparison of Pareto fronts of the BK1&L1 problem obtained by two algorithms

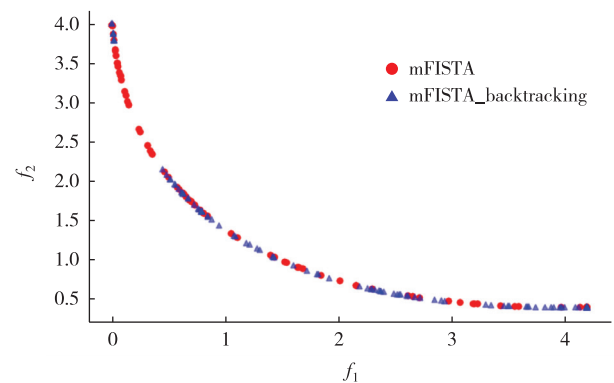


图 6 2 种算法获得的 JOS1&L1 问题的 Pareto 前沿对比  
Fig. 6 The comparison of Pareto fronts of the JOS1&L1 problem obtained by two algorithms

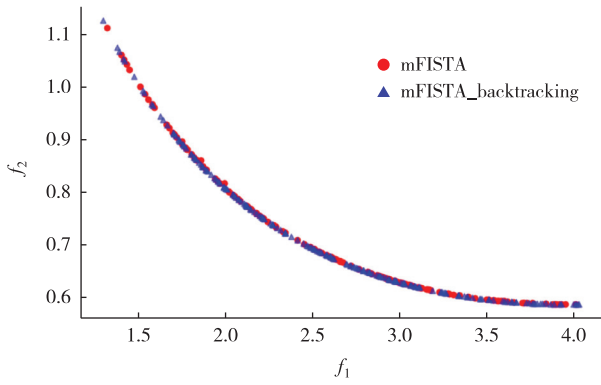


图 7 2 种算法获得的 SP1&L1 问题的 Pareto 前沿对比  
 Fig. 7 The comparison of Pareto fronts of the SP1&L1 problem obtained by two algorithms

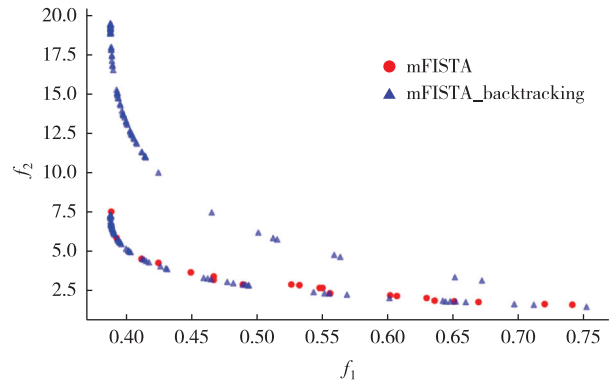


图 8 2 种算法获得的 VU1&L1 问题的 Pareto 前沿对比  
 Fig. 8 The comparison of Pareto fronts of the VU1&L1 problem obtained by two algorithms

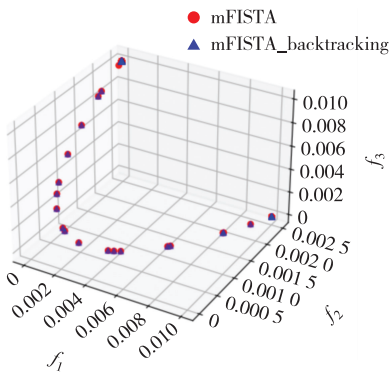


图 9 2 种算法获得的 MHHM1&0 问题的 Pareto 前沿对比  
 Fig. 9 The comparison of Pareto fronts of the MHHM1&0 problem obtained by two algorithms

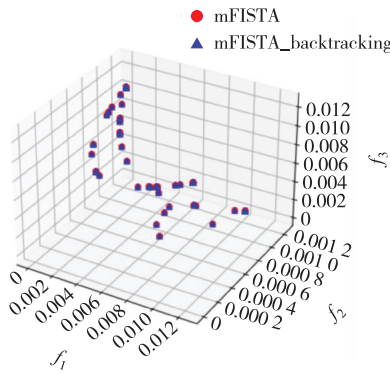


图 10 2 种算法获得的 MHHM2&0 问题的 Pareto 前沿对比  
 Fig. 10 The comparison of Pareto fronts of the MHHM2&0 problem obtained by two algorithms

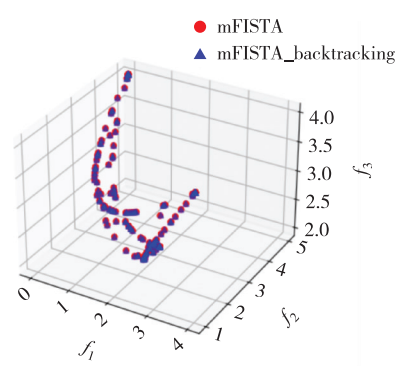


图 11 2 种算法获得的 VFM1&0 问题的 Pareto 前沿对比  
 Fig. 11 The comparison of Pareto fronts of the VFM1&0 problem obtained by two algorithms

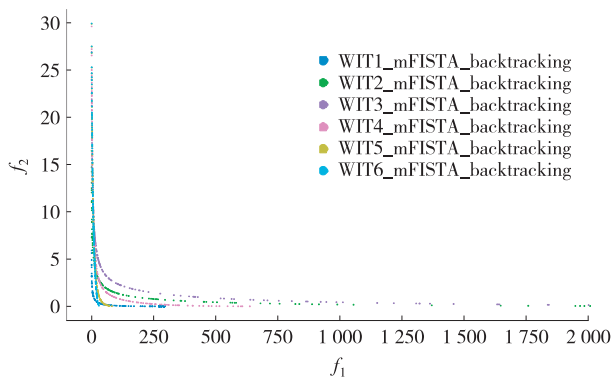


图 12 mFISTA\_backtracking 获得的 WIT1-WIT6 问题的 Pareto 前沿  
 Fig. 12 Pareto fronts of the WIT1-WIT6 problem obtained by mFISTA\_backtracking

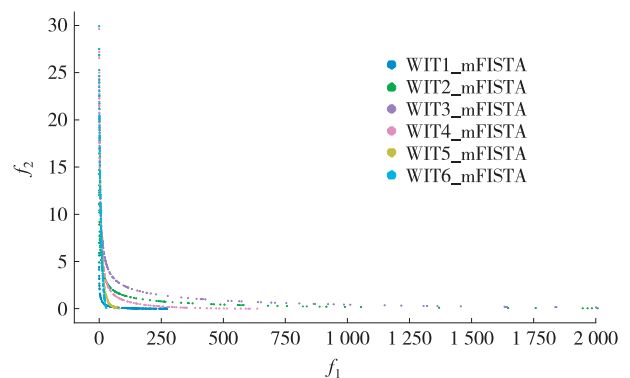


图 13 mFISTA 获得的 WIT1-WIT6 问题的 Pareto 前沿  
 Fig. 13 Pareto fronts of the WIT1-WIT6 problem obtained by mFISTA

为了评估所获得的 Pareto 前沿的效率,提供关于平均迭代次数、平均 CPU 运行时间和平均函数评估次数的性能指标图(图 16~图 18)。

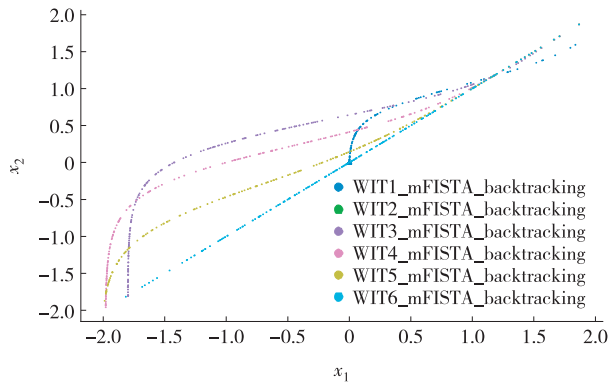


图 14 mFISTA\_backtracking 获得的 WIT1-WIT6 问题的弱 Pareto 最优解集

Fig. 14 Weak Pareto optimal solutions of the WIT1-WIT6 problem obtained by mFISTA\_backtracking

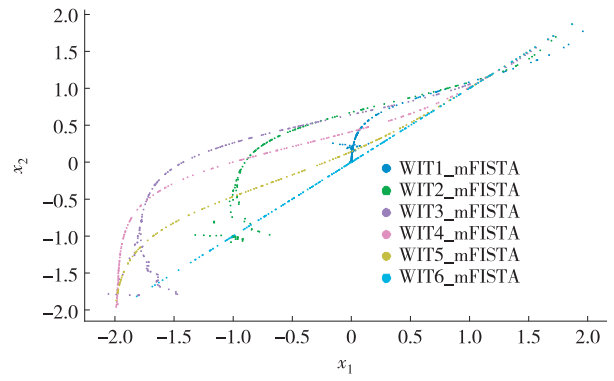


图 15 mFISTA 获得的 WIT1-WIT6 问题的弱 Pareto 最优解集

Fig. 15 Weak Pareto optimal solutions of the WIT1-WIT6 problem obtained by mFISTA

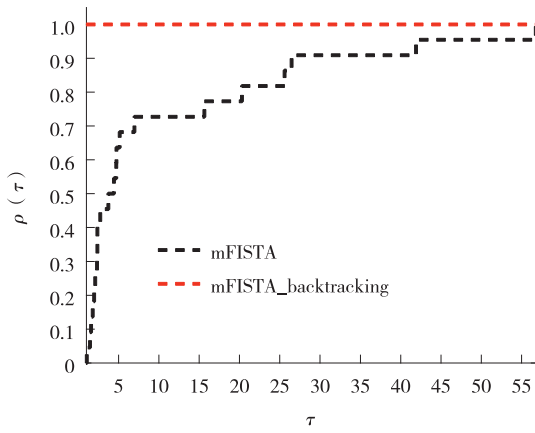


图 16 平均迭代次数性能指标

Fig. 16 Performance Profile: average number of iteration

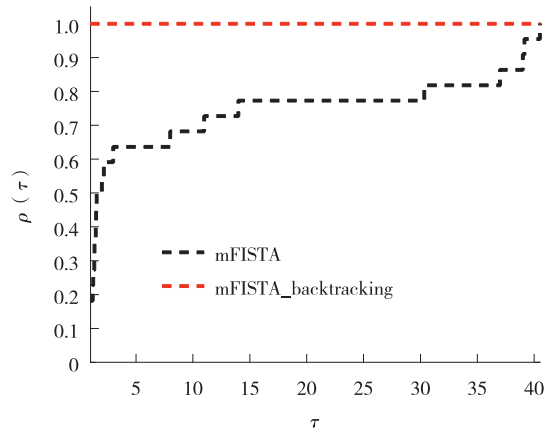


图 17 平均 CPU 运行时间性能指标

Fig. 17 Performance Profile: average time

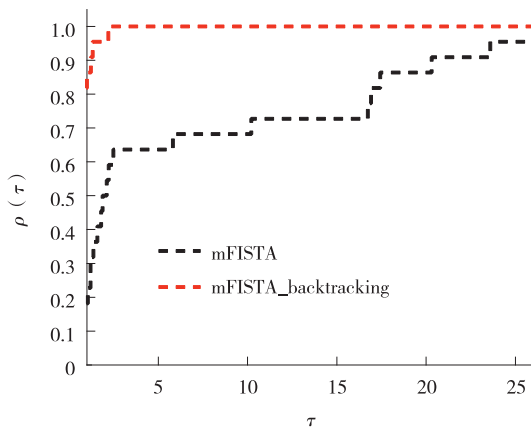


图 18 平均函数评估次数性能指标

Fig. 18 Performance Profile: average number of function evaluations

### 5 结论

本文提出一种新的回溯策略,以解决 mFISTA 无法在目标函数光滑部分梯度的 Lipschitz 常数  $L(f)$  未知情形下给出合适估计值的问题。具体而言,该回溯策略通过  $L_k$  估计  $L(f)$  并建立与  $t_k$  的等式关系,克服了由

$L(f)$ 未知带来的难题。此外,该回溯策略允许  $L_k$  可以在下一次迭代前适当缩小,从而避免产生较小的迭代步长,有效提高了计算效率。未来,将进一步探索这一新策略在更广泛应用场景中的表现,例如用于求解非光滑多目标优化问题。

#### 参考文献:

- [1] BINDER M, MOOSBAUER J, THOMAS J, et al. Multi-objective hyperparameter tuning and feature selection using filter ensembles[C]//Proceedings of the 2020 Genetic and Evolutionary Computation Conference. New York: ACM, 2020: 471-479.
- [2] GONG M G, ZHANG M Y, YUAN Y. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(1): 544-557.
- [3] STADLER W. Multicriteria optimization in engineering and in the sciences[M]. Boston: Springer, 1988.
- [4] HUANG X X, YANG X Q. Nonlinear Lagrangian for multiobjective optimization and applications to duality and exact penalization[J]. SIAM Journal on Optimization, 2002, 13(3): 675-692.
- [5] FUKUDA E H, DRUMMOND L M G. A survey on multiobjective descent methods[J]. Pesquisa Operacional, 2014, 34(3): 585-620.
- [6] TANABE H, FUKUDA E H, YAMASHITA N. Proximal gradient methods for multiobjective optimization and their applications[J]. Computational Optimization and Applications, 2019, 72(2): 339-361.
- [7] FLIEGE J, SVAITER B F. Steepest descent methods for multicriteria optimization[J]. Mathematical Methods of Operations Research, 2000, 51(3): 479-494.
- [8] BONNEL H, IUSEM A N, SVAITER B F. Proximal methods in vector optimization[J]. SIAM Journal on Optimization, 2005, 15(4): 953-970.
- [9] TANABE H, FUKUDA E H, YAMASHITA N. Convergence rates analysis of a multiobjective proximal gradient method[J]. Optimization Letters, 2023, 17(2): 333-350.
- [10] ZHAO X P, GHOSH D, QIN X L, et al. On the convergence analysis of a proximal gradient method for multiobjective optimization[J]. TOP, 2025, 33(1): 102-132.
- [11] CHEN K M, FUKUDA E H, YAMASHITA N. A proximal gradient method with Bregman distance in multi-objective optimization[J]. Pacific Journal of Optimization, 2024, 20(4): 809-826.
- [12] CHEN J, TANG L P, YANG X M. Barzilai-Borwein proximal gradient methods for multiobjective composite optimization problems with improved linear convergence[EB/OL]. (2023-09-13)[2025-01-23]. <https://arxiv.org/pdf/2309.06929>.
- [13] TANABE H, FUKUDA E H, YAMASHITA N. An accelerated proximal gradient method for multiobjective optimization[J]. Computational Optimization and Applications, 2023, 86(2): 421-455.
- [14] NISHIMURA Y, FUKUDA E H, YAMASHITA N. Monotonicity for multiobjective accelerated proximal gradient methods[J]. Journal of the Operations Research Society of Japan, 2024, 67(1): 1-17.
- [15] ZHANG J, YANG X M. The convergence rate of the accelerated proximal gradient algorithm for multiobjective optimization is faster than  $O(1/k^2)$ [EB/OL]. (2023-12-12)[2025-01-23]. <https://arxiv.org/pdf/2312.06913>.
- [16] SCHEINBERG K, GOLDFARB D, BAI X. Fast first-order methods for composite convex optimization with backtracking[J]. Foundations of Computational Mathematics, 2014, 14(3): 389-417.
- [17] BERTSEKAS D P. Nonlinear Programming[M]. Belmont: Athena Scientific, 1999.
- [18] TANABE H, FUKUDA E H, YAMASHITA N. New merit functions for multiobjective optimization and their properties[J]. Optimization, 2024, 73(13): 3821-3858.
- [19] HUBAND S, HINGSTON P, BARONE L, et al. A review of multiobjective test problems and a scalable test problem toolkit[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(5): 477-506.
- [20] DAS I, DENNIS J E. Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems[J]. SIAM Journal on Optimization, 1998, 8(3): 631-657.
- [21] HILLERMEIER C. Generalized Homotopy approach to multiobjective optimization[J]. Journal of Optimization Theory and Applications, 2001, 110(3): 557-583.
- [22] CHEN J, TANG L P, YANG X M. A Barzilai-Borwein descent method for multiobjective optimization problems[J]. European Journal of Operational Research, 2023, 311(1): 196-209.

- [23] JIN Y C, OLHOFFER M, SENDHOFF B. Dynamic Weighted Aggregation for evolutionary multi-objective optimization: why does it work and how?[C]//Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation, New York: ACM, 2001:1042-1049.
- [24] PREUSS M, NAUJOKS B, RUDOLPH G. Pareto set and EMOA behavior for simple multimodal multiobjective functions [C]//RUNARSSON T P, BEYER H G, BURKE E, et al. Parallel Problem Solving from Nature-PPSN IX. Berlin: Springer, 2006:513-522.
- [25] WITTING K. Numerical algorithms for the treatment of parametric multiobjective optimization problems and applications[D]. Paderborn: Universitätsbibliothek, 2012.

## Operations Research and Cybernetics

### Accelerated Proximal Gradient Method Based on Backtracking Strategy for Multiobjective Optimization

HUANG Chengzhi, CHEN Jian, TANG Liping

(National Center for Applied Mathematics in Chongqing, Chongqing Normal University, Chongqing 401331, China)

**Abstract:** To solve the problem generated by unknown Lipschitz constants for the gradient of the smooth part of the objective function, it proposes a new backtracking strategy for composite multi-objective optimization problems based on the accelerated proximal gradient algorithm. This strategy constructs an update rule for an estimated sequence of Lipschitz constants that satisfies a specific equality relationship, allowing the sequence to be updated in a non-monotonic way. Under appropriate conditions, it is proven that all cluster points of the sequence generated by the algorithm are weakly Pareto optimal solutions. Furthermore, the sublinear convergence rate  $O(1/k^2)$  of the algorithm is established using merit functions. Numerical experiments demonstrate that, compared to the accelerated proximal gradient algorithm without the backtracking strategy, the proposed algorithm exhibits significant advantages regarding runtime, iteration count, and function evaluation count.

**Keywords:** multiobjective optimization; accelerated proximal gradient algorithm; backtracking strategy; convergence rate

(责任编辑 黄颖)