

基于近似一阶信息的改进的加速水平束方法

李艳妮

上海建桥学院职业技术学院，上海

收稿日期：2024年3月19日；录用日期：2024年4月15日；发布日期：2024年4月22日

摘要

本文提出了一个近似一阶信息的改进的加速水平束方法，该方法结合多步加速策略，引入了三个迭代点序列进行求解，通过引入非欧氏距离代替传统的欧式距离，从而可以充分利用可行集的几何集合，与Lan的方法相比，本文提出的方法仅需求解一个子问题，从而可以减少算法的计算量。最后对所提出的算法进行收敛性分析以及计算相应的迭代复杂度。

关键词

近似一阶信息，加速水平束方法，全局收敛性

An Modified Accelerated Level Bundle Method Based on Approximate First Order Information

Yanni LI

Vocational College, Shanghai Jianqiao University, Shanghai

Received: Mar. 19th, 2024; accepted: Apr. 15th, 2024; published: Apr. 22nd, 2024

Abstract

In this paper, we propose an improved acceleration level bundle method with approximate first-order information, which combines the multistep acceleration strategy and introduces three iterative point sequences to solve, and can make full use of the geometric set of feasible sets by introducing non-Euclidean distances instead of traditional Euclidean distances. Finally, Global convergence of the algorithm is proved and the iterative complexity is analyzed.

Keywords

Approximate First Order Information, Accelerated Level Bundle Method, Global Convergence

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本文考虑以下凸优化问题：

$$f^* := \min_{x \in X} f(x) \quad (1.1)$$

其中 $X \subseteq R^n$ 是非空紧凸集，函数 $f: R^n \rightarrow R$ 是一个闭凸函数。假设上述优化问题的解集 X^* 非空，则对于 $\forall x \in X$ ，目标函数 f 满足如下不等式：

$$f(y) - f(x) - \langle f'(x), y - x \rangle \leq \frac{M}{1+\rho} \|y - x\|^{1+\rho}, \forall x, y \in X, \quad (1.2)$$

其中目标函数 f 在此点的函数值为 $f(x)$ ，次梯度为 $f'(x) \in \partial f(x)$ ，且常数 $M > 0$ ， $\rho \in [0, 1]$ ， $\|\cdot\|$ 是定义在 R^n 上的范数， $\langle \cdot, \cdot \rangle$ 表示相应的内积。当 $\rho = 0$ 时，函数 $f(x)$ 是非光滑函数；当 $\rho = 1$ 时，函数 $f(x)$ 是光滑函数；当 $\rho \in (0, 1)$ 时，函数 $f(x)$ 是弱光滑函数。文献[1] [2] [3] [4]证明了对任意的 $\varepsilon > 0$ ，任何的一阶算法为找到(1.1)的一个 ε -解，即可以找到一个点 p 使得 $f(p) - f^* \leq \varepsilon$ ，算法所需要计算 f 的一阶信息即次梯度的次数无法小于 $O\left(1/\varepsilon^{\frac{2}{1+3\rho}}\right)$ 。

本文主要讨论水平束类算法，该类算法利用当前所有迭代点处的函数值和次梯度产生线性化函数，通过对这些线性化函数取最大，产生一系列近似凸多面体切平面模型，再极小化割平面模型来逼近原目标函数，并利用切平面模型来产生一系列关于最优值 f^* 的上界和下界，该类方法主要是通过迭代不断压缩上下界的间隙来找到原问题的一个近似解。然而，在很多实际问题中，函数在某点的精确的一阶信息(函数值和次梯度)可能比较难甚至是无法计算。因此，设计有效的算法利用近似的一阶信息求解目标函数并使其具有理论上最优的迭代复杂度的近似算法具有重要的理论意义和实际应用价值。目前，关于近似水平束类算法已经有了很多的研究，梁玲[5]提出了非光滑优化基于非精确的加速水平束方法，陈韵梅和张维[6]提出了基于近似一阶信息的加速水平束方法，近似解达到了最佳精度以及相应的迭代复杂度。本文主要基于 Lan [7]引入的多步加速策略和采用的非欧式距离函数结合陈韵梅[6]提出的基于近似一阶信息的加速水平束方法，提出了基于近似一阶信息的改进的加速水平束方法，简记为“IMAPL”。该方法采用非欧几里得距离代替水平束方法中的欧氏距离，从而可以充分利用可行集的几何结构，大大提高了计算速度。文献[5]“IAPL”非精确数据的加速邻近水平束方法有效性依赖于两个子问题的求解：第一个是为了计算原问题的最优值的下界需要求解一个线性规划问题；第二个是为了更新迭代点需要去求解一个二次规划问题。本文对近似一阶信息的加速邻近水平束方法进行改进，去掉了第一个子问题的求解，第二个子问题用来定义新的迭代点和更新下界，并且可以根据可行集具体的几何结构选择合适的邻近函数。

2. 算法设计

本文采用非精确的一阶信息,即在每一次迭代中,对 $\forall x \in X$, 产生函数 f 满足以下条件的近似一阶信息:

$$\begin{cases} f_\delta \in [f(x) - \delta, f(x)], \\ g_\delta \in \partial_\delta f(x). \end{cases} \quad (2.1)$$

其中, f_δ 和 g_δ 分别叫做函数 f 在点 x 的近似函数值和近似次梯度, $\partial_\delta f(x)$ 的定义如下:

$$\partial_\delta f(x) = \{g \in R^n : f(y) \geq f_\delta + \langle g, y - x \rangle - \delta, \forall y \in X\}. \quad (2.2)$$

根据(2.1)和 f 的凸性可以得到

$$f(\cdot) \geq f_\delta(x) + \langle g_\delta(x), \cdot - x \rangle - \delta, \quad f_\delta(x) \in [f(x) - \delta, f(x)]. \quad (2.3)$$

假设目标函数(1.2)的近似一阶信息满足下面条件,即对 $\forall x \in X$, 函数 f 在这点的近似一阶信息 $(f_\delta(x), g_\delta(x))$ 满足如下不等式:

$$f(y) - f_\delta(x) - \langle g_\delta(x), y - x \rangle \leq \frac{L}{2} \|y - x\|^2 + \delta, \quad \forall x, y \in X, \quad (2.4)$$

为了方便算法的描述和定理的证明的需要,下面给出一些基本知识。

定义邻近函数:

$$\varphi(x_0, x) = \omega(x) - [\omega(x_0) + \langle \omega'(x_0), x - x_0 \rangle], \quad \forall x \in X, \quad (2.5)$$

其中, 函数 $\omega: R^n \rightarrow R$ 是集合 X 上系数为 σ_ω 强凸函数, $x_0 = \arg \min_{x \in X} \varphi(x_0, x)$ 为函数 $\varphi(x_0, x)$ 的邻近中心。

记集合 X 对应于函数 $\omega(x)$ 的大小为:

$$D_{\omega, X} := \{\omega(x) - [\omega(z) + \langle \nabla \omega(z), x - z \rangle], \forall x, z \in X\}, \quad (2.6)$$

记 $L_f(l) := \{x \in X : f(x) \leq l\}$ 为函数 f 在 X 上的水平集,假如存在一个紧凸集 X' , 满足 $L_f(l) \subseteq X' \subseteq X$, 则称集合 X' 为水平集 $L_f(l)$ 的一个定位器。

定义间隙为:

$$\Delta_k := \bar{f}_k - \underline{f}_k. \quad (2.7)$$

根据 Δ_k 的定义易知 Δ_k 是非负的,通过不断压缩间隙 Δ_k , 寻找原问题(1.1)的近似最优解。该算法的主要思想是利用目标函数的近似一阶信息来产生线性化,进而产生最优值 f^* 的上下界,然后不断压缩上下界的间隙来得到给定精度的最优解,算法由外迭代和内迭代子程序构成,每调用一次内迭代子程序,最优值上下界之间的间隙都会随着减少一个常数倍,下面给出算法的具体步骤:

算法

步骤 0 (参数选取) 给定精度 $\varepsilon > 0$, 选取参数 $\lambda, \theta \in (0, 1)$, 初始模型精度 δ_0 。

步骤 1 (初始化) 给定初始点 $p_0 \in X$, 令 $p_1 \in \arg \min_{x \in X} h_{\delta_0}(p_0, x)$, $\text{ub}_1 := \min\{f_{\delta_0}(p_0), f_{\delta_0}(p_1)\} + \delta_0$, $\text{lb}_1 := h_{\delta_0}(p_0, p_1)$, 令 $s := 1$ 。

步骤 2 (终止准则) 如果 $\text{ub}_s - \text{lb}_s \leq \varepsilon$, 则终止算法, 输出近似解 p_s 。

步骤 3 (调用子程序) 令 $(p_{s+1}, \text{lb}_{s+1}) = \varsigma_{\text{IMAPL}}(p_s, \text{lb}_s, \lambda, \theta, \delta_s)$, δ_s 为当前模型精度, $\text{ub}_{s+1} := f_{\delta_s}(p_{s+1}) + \delta_s$ 。

步骤 4 (循环) 令 $s := s + 1$, 返回步骤 2。

注 1: (1) λ 是在内循环时计算水平参数的一个参数, 每当 s 增加 1, 算法 IMAPL 进入下一个阶段, 该算法的每一次外迭代主要判断最优值的上下界间隙与给定精度的大小, 若间隙大于给定精度, 则调用子程序来压缩间隙, 否则, 终止算法, 输出近似最优解。

下面给出算法的内迭代子程序的具体过程, 选定一个搜索点 p 和最优值 f^* 的下界 lb , 输出新的搜索点 p^+ 和新的下界 lb^+ 满足 $f(p^+) - lb^+ \leq q[f(p) - lb]$, 其中 $q \in (0, 1)$ 且依赖于算法外迭代的输入参数 λ, θ 。

算法内迭代子程序: $(p^+, lb^+) = \varsigma_{\text{IMAPL}}(p, lb, \lambda, \theta, \delta)$ 。

步骤 0 (初始化) 令 $x_0^u = p$, $\bar{f}_0 := f(x_0^u) = ub$, $\underline{f}_0 := lb$, $l = \lambda \underline{f}_0 + (1 - \lambda) \bar{f}_0$ 。选取 $x_0 \in X$, $X'_0 = X$, 考虑由(2.5)定义的邻近函数 $\varphi(x_0, x)$, 令 $k := 1$ 。

步骤 1 (更新割平面模型) 令

$$x_k^l := (1 - \alpha_k) x_{k-1}^u + \alpha_k x_{k-1} \quad (2.8)$$

$$h_\delta(x_k^l, x) := f_\delta(x_k^l) + \langle g_\delta(x_k^l), x - x_k^l \rangle \quad (2.9)$$

$$\underline{X}_k := \{x \in X'_{k-1} : h_\delta(x_k^l, x) \leq l\} \quad (2.10)$$

步骤 2 (更新迭代点和下界) 令

$$x_k := \arg \min_{x \in \underline{X}_k} \varphi(x_0, x) \quad (2.11)$$

若 $\underline{X}_k = \emptyset$, 则终止程序, 输出 $p^+ = x_{k-1}^u$, $lb^+ = l$ 。

步骤 3 (更新上界) 令

$$\tilde{x}_k^u := (1 - \alpha_k) x_{k-1}^u + \alpha_k x_{k-1} \quad (2.12)$$

$$x_k^u = \begin{cases} \tilde{x}_k^u & \text{如果 } f_\delta(\tilde{x}_k^u) + \delta < \bar{f}_{k-1}, \\ x_{k-1}^u & \text{否则,} \end{cases} \quad (2.13)$$

令 $\bar{f}_k := f_\delta(x_k^u) + \delta$, 如果 $\bar{f}_k \leq l + \theta(\bar{f}_0 - l)$, 则终止程序, 并输出 $p^+ = x_k^u$, $lb^+ = lb$ 。

步骤 4 (更新定位器) 选取任意多面体集 X'_k 满足 $\underline{X}_k \subseteq X'_k \subseteq \bar{X}_k$, 其中

$$\bar{X}_k := \{x \in X, \langle \nabla \varphi(x_0, x_k), x - x_k \rangle \geq 0\} \quad (2.14)$$

步骤 5 (循环) 令 $k := k + 1$, 返回步骤 1。

注 2: 步骤 0 中初始上下界来自于算法的外循环, 水平值是初始上下界的凸组合, 是固定不变的; 步骤 1 中为了保证算法有限次终止, 算法中步长 $\{\alpha_k\}$ 的选取需要满足一定的条件; 该算法在步骤 2 和步骤 3 分别有终止出口, 当上界有显著下降或者下界有显著上升, 则终止内循环; 步骤 4 中定位器的更新是可以任意选取集合 X'_k 满足 $\underline{X}_k \subseteq X'_k \subseteq \bar{X}_k$, 我们选取 X'_k 在这两个集合中间, 能够控制(2.10)中约束的数目, 减少迭代次数。

引理 2.1 对内迭代子程序 $\varsigma_{\text{IMAPL}}(p, lb, \lambda, \theta, \delta)$ 由下列结论成立:

(a) $\{X'_k\}_{k \geq 0}$ 是水平集 $L_f(l)$ 的定位器, 即 $L_f(l) \subseteq X'_k$, $\forall k \geq 0$ 成立。

(b) 对任意的 $k \geq 1$, 有 $\bar{f}_0 \geq \bar{f}_1 \geq \dots \geq \bar{f}_k \geq f^*$ 。

(c) 如果 $L_f(l) \neq \emptyset$, 则对任意的 $k \geq 1$, 可以找到 X'_k 使得 $L_f(l) \subseteq \underline{X}_k \subseteq X'_k \subseteq \bar{X}_k$ 成立。

(d) 如果 $\underline{X}_k \neq \emptyset$, 则子问题(2.11)有唯一解。如果程序在步骤 2 终止, 则水平值 l 为最优值的一个下界, 即 $l \leq f^*$ 成立。

证明: (a) 用数学归纳法:

1) 当 $k=0$ 时, 显然有 $L_f(l) \subseteq X = X'_0$ 成立;

2) 假设 $k-1$ 时, 有 $L_f(l) \subseteq X'_{k-1}$ 成立, 往证 $L_f(l) \subseteq X'_k$;

3) 由 h_δ 的定义(2.9)和 f 的凸性可知, $h_\delta(x_k^l, x) \leq f(x) \leq l$ 成立, 结合 ii) 对 $\forall x \in L_f(l)$ 有 $x \in X'_{k-1}$ 且 $h_\delta(x_k^l, x) \leq f(x) \leq l$ 成立, 根据 $\underline{X}_k$ 的定义(2.10)可得 $L_f(l) \subseteq \underline{X}_k$, 又因为 $\underline{X}_k \subseteq X'_k$ 可知 $L_f(l) \subseteq X'_k$, 综上所述, $\{X'_k\}_{k \geq 0}$ 是水平集 $L_f(l)$ 的定位器。

(b) 由内迭代子程序的步骤 3 可以知道, $\bar{f}_k := \min\{\bar{f}_{k-1}, f_\delta(\tilde{x}_k'') + \delta\} \leq \bar{f}_{k-1}$, 故对任意的 $k \geq 1$ 时, 有 $f^* \leq \bar{f}_k \leq \bar{f}_{k-1} \leq \dots \leq \bar{f}_1 \leq \bar{f}_0$ 。

(c) 由(a)的证明可知, $L_f(l) \subseteq \underline{X}_k$ 成立, 根据子程序的步骤 4 可知 $X'_k \neq \emptyset$, 否则子程序在步骤 2 终止, 下证 $\underline{X}_k \subseteq \bar{X}_k$ 成立。

根据子问题(2.11)的一阶最优性条件可知: $0 \in \nabla \varphi(x_0, x_k) + N_{\underline{X}_k}(x_k)$, 即 $-\nabla \varphi(x_0, x_k) \in N_{\underline{X}_k}(x_k)$, 由法锥的定义有 $\langle \nabla \varphi(x_0, x_k), x - x_k \rangle \geq 0, \forall x \in \underline{X}_k$ 。再根据 $\bar{X}_k$ 的定义(2.14)可知, 对 $\forall x \in \underline{X}_k$ 有 $x \in \bar{X}_k$, 所以 $\underline{X}_k \subseteq \bar{X}_k$, 即存在 X'_k 满足 $L_f(l) \subseteq \underline{X}_k \subseteq X'_k \subseteq \bar{X}_k, \forall k \geq 1$ 。

(d) 根据 $\underline{X}_k$ 的定义(2.10)和 X 的紧凸集可知 $\underline{X}_k$ 是紧凸集, 又因为 $\varphi(x_0, x)$ 是强凸函数, 如果 $\underline{X}_k \neq \emptyset$, 则子问题(2.11)存在唯一解, 如果程序在步骤 2 终止, 那么意味着 $\underline{X}_k = \emptyset$, 并且根据(a)可知 $L_f(l) \subseteq \underline{X}_k$, 所以 $L_f(l) = \emptyset$, 于是对 $\forall x \in X$, 有 $l \leq f(x)$ 成立, 从而 $l \leq f^* \leq f(x), \forall x \in X$, 即 l 是 f^* 的一个下界。□

为了保证内迭代子程序 ς_{IMAPL} 能够终止且达到最优的迭代复杂度, 需要选取合适的步长 $\{\alpha_k\}$, Lan 在文献[6]中给出了序列 $\{\alpha_k\}$ 的一个广义的选取规则。Chen 等[7]提出了一种更为简洁的选取方法, 即序列 $\{\alpha_k\}$ 需要满足如下条件:

$$\alpha_1 = 1, 0 < \alpha_k \leq 1, \alpha_k < \frac{C_1}{k}, \frac{1 - \alpha_{k+1}}{\alpha_{k+1}^2} \leq \frac{1}{\alpha_k^2} \text{ 且 } \alpha_k^2 \sum_{i=1}^k \frac{1}{\alpha_i^2} \leq C_2 k, \forall k \geq 1, \quad (2.15)$$

其中, $C_1, C_2 > 0$ 且为常数, 下面提供两个具体的 $\{\alpha_k\}$ 的选取的例子, 可类似文献[7]来验证这两个例子满足(2.15)。

引理 2.2 [8] (1) 若选取 $\alpha_k = 2/(k+1), k=1, 2, \dots$, 取 $C_1 = 2, C_2 = 2$, 条件(2.15)满足。(2) 若选取 $\{\alpha_k\}$ 满足以下递归关系:

$\alpha_1 = 1, \alpha_{k+1}^2 = (1 - \alpha_{k+1})\alpha_k^2, \forall k \geq 1$, 则当 $C_1 = 2, C_2 = 2$, 条件(2.15)满足。

引理 2.3 [7] 当程序 ς_{IMAPL} 终止时, 有 $f(p^+) - lb^+ \leq q(f(p) - lb)$ 。其中,

$$q \equiv q(\lambda, \theta) = \max\{\lambda, 1 - \lambda(1 - \theta)\} \quad (2.16)$$

注意到 $f(p)$ 和 $f(p^+)$ 分别表示子程序 ς_{IMAPL} 输入和输出最优值 f^* 的上界, 引理 2.3 说明了每当程序 ς_{IMAPL} 终止时, 最优值 f^* 的上界和下界的差值总是减少了一个常数倍。

3. 收敛性与复杂度分析

本节将分析 IMAPL 算法的全局收敛性以及证明该算法终止时总的迭代复杂度。

定理 3.1 在 ς_{IMAPL} 中, 如果步长 $\{\alpha_k\}$ 满足条件(2.15), 那么对 $\forall K \geq 1$, 若 ς_{IMAPL} 在第 K 次迭代没有终止, 则下面结论成立:

$$\bar{f}_K - l \leq \frac{C_1 L D_{\omega, X}}{K^2 \sigma_\omega} + 2C_2 K \delta. \quad (3.1)$$

证明: 假设内迭代子程序 ζ_{IMAPL} 在第 $K(>0)$ 次迭代没有终止, 由(2.10), (2.11)可知 $x_k \in \underline{X}_k \subseteq X_{k-1} \subseteq \bar{X}_{k-1}, \forall 2 \leq k \leq K$, 同时由 $x_1 \in X'_0$, $x_0 = \arg \min_{x \in X'_0} \varphi(x_0, x)$, 有

$$\varphi(x_0, x) \geq \frac{\sigma_\omega}{2} \|x - x_0\|^2. \quad (3.2)$$

另外, 由 $\bar{X}_k$ 的定义有 $\langle \nabla \varphi(x_0, x_k), x - x_k \rangle \geq 0, \forall x \in \bar{X}_k$, 又因为 $X'_k \subseteq \bar{X}_k$, 显然 $\langle \nabla \varphi(x_0, x_k), x - x_k \rangle \geq 0, \forall x \in X'_k$ 成立, 根据子问题(2.11)可知 $x_{k+1} \in \underline{X}_{k+1} \subseteq X'_k$, 所以 $\langle \nabla \varphi(x_0, x_k), x_{k+1} - x_k \rangle \geq 0$, 所以结合 $\varphi(x_0, x)$ 的强凸性和子问题的最优性条件可以得到

$$\begin{aligned} \frac{\sigma_\omega}{2} \|x_k - x_{k-1}\|^2 &\leq \varphi(x_0, x_k) - \varphi(x_0, x_{k-1}) - \langle \nabla \varphi(x_0, x_{k-1}), x_k - x_{k-1} \rangle \\ &\leq \varphi(x_0, x_k) - \varphi(x_0, x_{k-1}). \end{aligned} \quad (3.3)$$

所以,

$$\begin{aligned} \frac{\sigma_\omega}{2} \sum_{i=1}^k \|x_i - x_{i-1}\|^2 &\leq \varphi(x_0, x_k) - \varphi(x_0, x_0) \\ &\leq \varphi(x_0, x_k) \\ &= \max_{y \in X} \varphi(x_0, y) \\ &\leq \max_{x, y \in X} \varphi(x, y) \\ &= D_{\omega, X}. \end{aligned} \quad (3.4)$$

由 $\bar{f}_k$ 的定义及(2.4)和(2.12)即知对 $\forall k \geq 1$, 有

$$\begin{aligned} \bar{f}_k &= f_\delta(x_k^u) + \delta = f_\delta(\tilde{x}_k^u) + \delta \leq f(\tilde{x}_k^u) + \delta \\ &\leq f_\delta(x_k^l) + \langle g_\delta(x_k^l), \tilde{x}_k^u - x_k^l \rangle + \frac{L}{2} \|\tilde{x}_k^u - x_k^l\|^2 + 2\delta \quad \text{由(2.4)} \\ &= (1 - \alpha_k) \left[f_\delta(x_k^l) + \langle g_\delta(x_k^l), x_{k-1}^u - x_k^l \rangle \right] + \alpha_k \left[f_\delta(x_k^l) + \langle g_\delta(x_k^l), x_k - x_k^l \rangle \right] \\ &\quad + \frac{L\alpha_k^2}{2} \|x_k - x_{k-1}\|^2 + 2\delta \quad \text{由(2.12)} \\ &\leq (1 - \alpha_k) f(x_{k-1}^u) + \alpha_k l + \frac{L\alpha_k^2}{2} \|x_k - x_{k-1}\|^2 + 2\delta \quad \text{由(2.3)和(2.10)} \\ &\leq (1 - \alpha_k) \left[f_\delta(x_{k-1}^u) + \delta \right] + \alpha_k l + \frac{L\alpha_k^2}{2} \|x_k - x_{k-1}\|^2 + 2\delta \quad \text{由(2.1)} \\ &= (1 - \alpha_k) \bar{f}_{k-1} + \alpha_k l + \frac{L\alpha_k^2}{2} \|x_k - x_{k-1}\|^2 + 2\delta. \quad \text{由 } \bar{f}_k \text{ 的定义} \end{aligned}$$

对上式两边都减去 l 可得

$$\bar{f}_k - l \leq (1 - \alpha_k) (\bar{f}_{k-1} - l) + \frac{L\alpha_k^2}{2} \|x_k - x_{k-1}\|^2 + 2\delta \quad (3.5)$$

对式(3.5)两边都除以 α_k^2 可得

$$\frac{\bar{f}_k - l}{\alpha_k^2} \leq \frac{1 - \alpha_k}{\alpha_k^2} (\bar{f}_{k-1} - l) + \frac{L}{2} \|x_k - x_{k-1}\|^2 + \frac{2\delta}{\alpha_k^2} \quad (3.6)$$

当 $k=1$ 时, $\alpha_1=1$, 上式(3.6)为

$$\frac{\bar{f}_1-l}{\alpha_1^2} \leq \frac{L}{2} \|x_1-x_0\|^2 + \frac{2\delta}{\alpha_1^2} \quad (3.7)$$

对 $\forall k > 1$, 由(2.15)知 $\frac{1-\alpha_{k+1}}{\alpha_{k+1}^2} \leq \frac{1}{\alpha_k^2}$, 上式(3.6)为

$$\frac{\bar{f}_k-l}{\alpha_k^2} \leq \frac{1}{\alpha_{k-1}^2} (\bar{f}_{k-1}-l) + \frac{L}{2} \|x_k-x_{k-1}\|^2 + \frac{2\delta}{\alpha_k^2} \quad (3.8)$$

对(3.8)从 1 到 K 求和再与(3.7)求和, 有

$$\frac{\bar{f}_K-l}{\alpha_K^2} \leq \frac{L}{2} \sum_{i=1}^K \|x_i-x_{i-1}\|^2 + 2\delta \sum_{i=1}^K \frac{1}{\alpha_i^2}$$

所以, $\bar{f}_K-l \leq \frac{L\alpha_K^2}{2} \sum_{i=1}^K \|x_i-x_{i-1}\|^2 + 2\delta\alpha_K^2 \sum_{i=1}^K \frac{1}{\alpha_i^2}$, 结合(2.15)和(3.4)可得

$$\bar{f}_K-l \leq \frac{C_1 LD_{\omega,X}}{K^2 \sigma_\omega} + 2C_2 K \delta$$

□

由上述定理, 我们已经得到了 $\bar{f}_K-l$ 上界的一个估计, 下面的定理将使用该估计来进一步计算子程序 ς_{IMAPL} 满足步骤 3 中终止条件所需要的迭代次数。由于加速一阶算法无法避免受到误差累积效应的影响, 在 L 、 R 和 δ 固定的情况下, $\bar{f}_K-l$ 并不能无限递减排至零, 而是存在着一个关于这些参数的一个最小值, 以下定理说明了, 在 δ 满足一定的条件下, 内迭代子程序 ς_{IMAPL} 的迭代次数有一个上界。

定理 3.2 在 ς_{IMAPL} 中, 如果步长 $\{\alpha_k\}$ 满足条件(2.15), 且 δ 满足如下条件

$$\delta \leq \delta_\Delta := \frac{\left(\frac{1}{3}\lambda\theta\Delta\right)^{\frac{3}{2}} \sqrt{\sigma_\omega}}{\sqrt{LD_{\omega,X}} C_1 C_2} \quad (3.9)$$

其中, 根据间隙的定义(2.7), Δ 表示每次调用 ς_{IMAPL} 时输入的上界和下界的差值, 即 $\Delta = ub - lb$, 则内迭代子程序的迭代次数不超过

$$K(\Delta) := \frac{C_1}{\sqrt{\frac{\lambda\theta}{3}}} \cdot \sqrt{\frac{LD_{\omega,X}}{\Delta\sigma_\omega}} + 1 \quad (3.10)$$

证明: 由(3.1)结合 $\delta \leq \delta_\Delta$ 可得,

$$\bar{f}_K-l \leq \frac{C_1 LD_{\omega,X}}{K^2 \sigma_\omega} + 2C_2 K \delta \leq \frac{C_1 LD_{\omega,X}}{K^2 \sigma_\omega} + 2C_2 K \delta_\Delta \quad (3.11)$$

不等式右边达到最小时

$$K = \left(\frac{C_1^2 LD_{\omega,X}}{C_2 \delta_\Delta \sigma_\omega} \right)^{\frac{1}{3}} \quad (3.12)$$

将式(3.12)代入式(3.11)可得

$$\bar{f}_k - l \leq \frac{C_1^2 LD_{\omega, X}}{\left(\frac{C_1^2 LD_{\omega, X}}{C_2 \delta_\Delta \sigma_\omega}\right)^{\frac{2}{3}}} + 2C_2 \left(\frac{C_1^2 LD_{\omega, X}}{C_2 \delta_\Delta \sigma_\omega}\right)^{\frac{1}{3}} \delta_\Delta$$

$$\text{解得, } \delta_\Delta \leq \frac{\left(\frac{1}{3} \lambda \theta \Delta\right)^{\frac{3}{2}} \sqrt{\sigma_\omega}}{\sqrt{LD_{\omega, X}} C_1 C_2}.$$

因此(3.11)右边达到最小时,

$$K^* := \left(\frac{C_1^2 LD_{\omega, X}}{C_2 \delta_\Delta \sigma_\omega}\right)^{\frac{1}{3}} = \frac{C_1}{\sqrt{\frac{\lambda \theta}{3}}} \cdot \sqrt{\frac{LD_{\omega, X}}{\Delta \sigma_\omega}}$$

因此有

$$\bar{f}_{K^*} - l \leq (3C_1 C_2 \delta_\Delta)^{\frac{2}{3}} \left(\frac{LD_{\omega, X}}{\sigma_\omega}\right)^{\frac{1}{3}} = \lambda \theta \Delta \quad (3.13)$$

观察内迭代子程序步骤 3 中终止条件 $\bar{f}_k \leq l + \theta(\bar{f}_0 - l)$ 和水平值 l 的定义可知,

$$\bar{f}_k - l \leq \theta(\bar{f}_0 - l) = \theta \lambda (\bar{f}_0 - f_0) = \theta \lambda (ub - lb) = \theta \lambda \Delta \quad (3.14)$$

即步骤 3 的终止条件等价于(3.14), 因此, 由(3.13)可知, 至少在第 K^* 次迭代后内循环终止。所以, 内循环或者在满足步骤 2 中终止条件时提前终止或者最多在第 K^* 次迭代后满足步骤 3 中终止条件而终止, 即内循环的迭代次数的上界 $K(\Delta) = K^* + 1$ 。□

以下定理讨论算法的收敛性和迭代复杂度。

定理 3.3 对于任意给定的 $\varepsilon > 0$, 如果算法中步长 $\{\alpha_k\}$ 选取满足条件(2.15), 且对于任意的 $s \geq 1$ 均有 $\delta_s \leq \delta_{\Delta_s}$, 其中 δ_{Δ_s} 在(3.9)中定义, 则算法将收敛到(1.1)的一个 ε 解, 且以下结论成立:

1) 调用内循环的次数不超过

$$S(\varepsilon) := \left\lceil \max \left\{ 0, \log_{\frac{1}{q}} \frac{D_{\omega, X} L + 2\delta_0 \sigma_\omega}{\varepsilon \sigma_\omega} \right\} \right\rceil \quad (3.15)$$

2) IMAPL 算法总的迭代次数不超过

$$S(\varepsilon) + \frac{\sqrt{3} C_1 \sqrt{LD_{\omega, X}}}{\sqrt{\lambda \theta \sigma_\omega \varepsilon}} \cdot \frac{1}{1 - \sqrt{q}} \quad (3.16)$$

证明: (1) 观察算法 IMAPL 的步骤 1, 并结合 $f_{\delta_0}(p_1) \leq f(p_1)$ 和(2.4)容易得到

$$\begin{aligned} \Delta_1 &= ub_1 - lb_1 \quad \text{由(2.7)} \\ &\leq f_{\delta_0}(p_1) + \delta_0 - h_{\delta_0}(p_0, p_1) \quad \text{由 } ub_1, lb_1 \text{ 的定义} \\ &\leq \frac{L}{2} \|p_0 - p_1\|^2 + 2\delta_0 \quad \text{由(2.4)} \\ &\leq \frac{LD_{\omega, X}}{\sigma_\omega} + 2\delta_0 \quad \text{由(3.4)} \end{aligned}$$

同时, 根据引理 3.2 可知, $\Delta_{s+1} \leq q\Delta_s$, $\forall s \geq 1$, 根据递推关系可以得到

$$\Delta_{s+1} \leq q\Delta_s \leq q^2\Delta_{s-1} \leq \cdots \leq q^s\Delta_1, \quad \forall s \geq 1 \quad (3.17)$$

容易看都 $\{\Delta_s\}_{s=1}^\infty$ 几何级数递减并必将最终小于 ε , 因此可以假设在第 $S(\varepsilon)$ 次调用 ς_{IMAPL} 后首次得到一个 ε 解, 即算法在 $s := S(\varepsilon) + 1$ 时终止, 则

$$\Delta_{S(\varepsilon)+1} \leq \varepsilon < \Delta_{S(\varepsilon)} \quad (3.18)$$

结合 $\Delta_1 \leq \frac{LD_{\omega,X}}{\sigma_\omega} + 2\delta_0$ 和(3.17)可以得到,

$$\varepsilon < q^{S(\varepsilon)-1}\Delta_1 \leq q^{S(\varepsilon)-1} \left(\frac{LD_{\omega,X}}{\sigma_\omega} + 2\delta_0 \right) \quad (3.19)$$

因此,

$$\left(\frac{1}{q} \right)^{S(\varepsilon)-1} \leq \frac{D_{\omega,X}L + 2\delta_0\sigma_\omega}{\sigma_\omega\varepsilon}$$

即,

$$S(\varepsilon) = \left\lceil \max \left\{ 0, \log_{\frac{1}{q}} \frac{D_{\omega,X}L + 2\delta_0\sigma_\omega}{\sigma_\omega\varepsilon} \right\} \right\rceil$$

(2) 假设 IMAPL 算法已经调用了 $S(\varepsilon)$ 次子程序, 对于 $1 \leq s \leq S(\varepsilon)$, 根据(3.17)、(3.18)可得

$$\varepsilon < \Delta_{S(\varepsilon)} \leq q\Delta_{S(\varepsilon)-1} \leq q^2\Delta_{S(\varepsilon)-2} \leq \cdots \leq q^{S(\varepsilon)-s}\Delta_s$$

即

$$\Delta_s \geq \varepsilon q^{s-S(\varepsilon)}, \quad s = 1, 2, \dots, S(\varepsilon)$$

则

$$\sum_{s=1}^{S(\varepsilon)} \Delta_s \leq \sum_{s=1}^{S(\varepsilon)} \varepsilon \frac{1}{q^{\frac{1}{2}}} q^{\frac{S(\varepsilon)-s}{2}} = \sum_{t=0}^{S(\varepsilon)-1} \varepsilon \frac{1}{q^{\frac{1}{2}}} q^{\frac{t}{2}} \leq \frac{1}{\sqrt{\varepsilon}(1-\sqrt{q})}$$

结合引理 2.1 和(3.10), 将调用子程序的所有迭代次数相加, 得到算法 IMAPL 的总迭代次数不超过

$$\begin{aligned} \sum_{s=1}^{S(\varepsilon)} K_s &\leq \sum_{s=1}^{S(\varepsilon)} K(\Delta_s) \leq S(\varepsilon) + \sum_{s=1}^{S(\varepsilon)} \frac{C_1}{\sqrt{\lambda\theta}} \cdot \sqrt{\frac{LD_{\omega,X}}{\Delta_s\sigma_\omega}} \\ &\leq S(\varepsilon) + \frac{\sqrt{3}C_1\sqrt{LD_{\omega,X}}}{\sqrt{\lambda\theta\sigma_\omega\varepsilon}} \cdot \frac{1}{1-\sqrt{q}}. \end{aligned}$$

其中, K_s 表示算法第 s 次调用子程序时的迭代次数。 □

4. 结束语

本文提出了一个近似一阶信息的改进的加速水平束方法, 该方法结合多步加速策略, 引入了三个迭代点序列进行求解, 并且通过引入邻近函数代替传统的欧式距离, 进而充分利用可行集的几何集合, 加快算法的收敛速度, 并且与文献[5] [6]相比, 减少了一个子问题的求解, 从而可以减少算法的计算量。最

后证明了算法的全局收敛性并分析了迭代复杂度。

参考文献

- [1] Nemirovski, A., Yudin, D.B. and Dawson, E.R. (1983) Problem Complexity and Method Efficiency in Optimization. A Wiley-Interscience Publication, New York.
- [2] Nemirovski, A. and Nesterov, Y. (1985) Optimal Methods of Smooth Convex Minimization. *Zh Vychisl Mat Mat Fiz*, **25**, 356-369. [https://doi.org/10.1016/0041-5553\(85\)90100-4](https://doi.org/10.1016/0041-5553(85)90100-4)
- [3] Nesterov, Y. (1988) On an Approach to the Construction of Optimal Methods of Minimization of Smooth Convex Functions. *Ekonomika Mat Metody*, **24**, 509-517.
- [4] Devolder, O., Glineur, F. and Nesterov, Y. (2014) First-Order Methods of Smooth Convex Optimization with Inexact Oracle. *Mathematical Programming*, **146**, 37-75. <https://doi.org/10.1007/s10107-013-0677-5>
- [5] 梁玲. 非光滑优化基于非精确数据的加速水平束方法[D]: [硕士学位论文]. 南宁: 广西大学, 2018.
- [6] Lan, G.H. (2013) Bundle-Level Type Methods Uniformly Optimal for Smooth and Nonsmooth Convex Optimization. *Mathematical Programming*, **149**, 1-45. <https://doi.org/10.1007/s10107-013-0737-x>
- [7] Chen, Y.M., Lan, G.H., Ouyang, Y.Y., *et al.* (2014) Fast Bundle Level Type Methods for Unconstrained and Ball-Constrained Convex Optimization. ArXiv:1412.2128
<https://doi.org/10.21236/ADA612792>
- [8] 陈韵梅, 张维. 基于近似一阶信息的加速的 Bundle Level 算法[J]. 中国科学, 2017, 10(47): 1119-1142.