

A Modified Level Bundle Method with Inexact Data for Nonsmooth Constrained Optimization

Yanni Li, Haiyan Zheng*, Chunming Tang

College of Mathematics and Information Science, Guangxi University, Nanning Guangxi
Email: ynli@st.gxu.edu.cn, zhenghaiyan166@126.com, cmtang@gxu.edu.cn

Received: Aug. 13th, 2019; accepted: Aug. 28th, 2019; published: Sep. 4th, 2019

Abstract

This paper presents a modified level bundle method with inexact data for nonsmooth constrained optimization. In the method, the inexact data and the approximate improvement function are introduced. Moreover, in the projection subproblem, the Bregman distance is used to replace the classical Euclidean distance, in order that the geometric structure of the feasible set can be taken into account, which can reduce the computation of the algorithm. Global convergence of the algorithm is proved and the iterative complexity is analyzed.

Keywords

Level Bundle Method, Inexact Data, Bregman Distance, Global Convergence, Complexity Analysis

非光滑约束优化基于非精确数据的改进水平束方法

李艳妮, 郑海艳*, 唐春明

广西大学数学与信息科学学院, 广西 南宁
Email: ynli@st.gxu.edu.cn, zhenghaiyan166@126.com, cmtang@gxu.edu.cn

收稿日期: 2019年8月13日; 录用日期: 2019年8月28日; 发布日期: 2019年9月4日

摘要

本文提出了一个求解非光滑约束优化问题基于非精确数据的改进水平束方法。该方法引入了非精确数据*通讯作者。

及相应的近似改进函数。此外,通过在投影子问题中引入Bregman距离以代替传统的欧氏距离,从而可以充分利用可行集的几何集合,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。

关键词

水平束方法, 非精确数据, Bregman距离, 全局收敛, 复杂度分析

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

1. 引言

本文研究求解以下非光滑约束凸优化问题:

$$\begin{aligned} f^* &:= \min f(x) \\ \text{s.t. } &c(x) \leq 0, \\ &x \in X, \end{aligned} \quad (1)$$

其中, f, c 均为 $R^n \rightarrow R$ 上的非光滑凸函数, $X \subseteq R^n$ 为非空紧凸集。

束方法是求解一般非光滑优化问题最有效的方法之一。求解约束问题(1)的束方法可细分为罚函数法[1], 滤子法[2], 改进函数法[3]等。罚函数法的基本思想是在目标函数中增加一个惩罚项, 从而构造出新的目标函数, 通过求罚函数的最优解得到原问题的最优解。但是在具体的算法实现中, 罚参数选取较困难, 如果选取不当, 则导致算法效率较差, 为避免罚参数的选取问题, Fletcher 和 Leyffer [4]提出了滤子法来求解约束优化问题, 其基本思想是当目标函数或者约束函数值有一个减少时, 迭代点将被滤子接收。但滤子法的缺点是算法结构复杂, 理论分析难度较大。为此, 本文将基于改进函数法思想研究相应的水平束方法。考虑如下改进函数:

$$h(x; f^*) = \max \{f(x) - f^*, c(x)\}. \quad (2)$$

从而求解问题(1)转化为在 X 上极小化改进函数(2)。但是在实际问题中最优值 f^* 通常是不知道的, Lemaréchal 等[3]提出了用最优值 f^* 的下界 f_{low}^k 近似替代 f^* , 并通过算法的不断迭代, 使得产生下界序列逼近最优解。Ackooij 和 De Oliveira [5]基于改进函数(2)提出了一种限制存储信息的水平束方法, 确保每一步迭代计算量较小且不影响算法的收敛性。

文献[5]中的投影子问题采用了传统的欧氏距离, 为了充分利用可行集的几何结构, 本文将利用 Bregman 距离代替传统的欧氏距离。此前, Kiwiel [6]在邻近束方法中引入 Bregman 距离代替二次项, 提出基于 Bregman 距离函数的邻近束方法。Ben-Tal 和 Nemirovski [7]提出了非欧氏限制存储的水平束方法, 可以根据可行集的几何结构选取合适的邻近函数。

另一方面, 在很多实际问题中, 往往很难精确计算函数值和次梯度, 或者计算成本较高。因此, 设计基于非精确数据的有效算法具有重要的理论意义和实用价值。最近, 梁玲[8]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[9]提出了基于近似一阶信息的加速水平束方法, 并分析了算法的一致最优迭代复杂度。

本文针对非光滑约束优化问题(1), 对文[5]的工作进行改进和推广, 提出了基于非欧氏距离的非精确约束水平束方法。该方法能够有效处理非精确数据, 并且通过引入 Bregman 距离代替传统的欧氏距离,

在计算时能充分利用可行集的几何结构, 选取合适的邻近函数, 提升计算效率。该方法利用多面体模型近似原问题的目标函数和约束函数, 并引入非精确的改进函数作为近似最优性判别函数。最后证明了算法的全局收敛性并分析了迭代复杂度。

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 $\forall x \in X$, 分别产生函数 f, c 满足以下条件的近似函数值和次梯度:

$$\begin{cases} f(x) \geq f_x \geq f(x) - \varepsilon_f^x, \tilde{g}_f^x \in \partial_{\varepsilon_f^x} f(x), \\ c(x) \geq c_x \geq c(x) - \varepsilon_c^x, \tilde{g}_c^x \in \partial_{\varepsilon_c^x} c(x), \end{cases} \quad (3)$$

其中, f_x, c_x, \tilde{g}_f^x 和 \tilde{g}_c^x 分别表示函数 f, c 在点 x 的近似函数值和近似次梯度, 误差 $\varepsilon_f^x, \varepsilon_c^x \geq 0$, δ -次微分 $\partial_\delta f(x)$ 定义如下:

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

根据(3)可得

$$f(\cdot) \geq f(x) + \langle \tilde{g}_f^x, \cdot - x \rangle - \varepsilon_f^x, \quad c(\cdot) \geq c(x) + \langle \tilde{g}_c^x, \cdot - x \rangle - \varepsilon_c^x.$$

假设每一个估计的误差都是有界的, 即存在常数 $\eta_f, \eta_c \geq 0$ 使得

$$\varepsilon_f^x \leq \eta_f, \quad \varepsilon_c^x \leq \eta_c, \quad \forall x \in X.$$

为简便, 分别用 $\tilde{g}_f^k, \tilde{g}_c^k$ 表示函数 f, c 在 x^k 处的近似次梯度, $\varepsilon_f^k, \varepsilon_c^k$ 表示相应的误差。定义邻近函数:

$$\varphi(x; \hat{x}) := \omega(x) - \omega(\hat{x}) - \langle \nabla \omega(\hat{x}), x - \hat{x} \rangle,$$

其中, 函数 $\omega: \mathbb{R}^n \rightarrow \mathbb{R}$ 是集合 X 上系数为 $\sigma_\omega > 0$ 可微强凸函数, 即

$$\omega(y) \geq \omega(x) + \langle \nabla \omega(x), y - x \rangle + \frac{\sigma_\omega}{2} \|y - x\|^2, \quad \forall x, y \in X.$$

易知, $\varphi(x; \hat{x}) \geq 0$, $\varphi(\hat{x}; \hat{x}) = 0$ 及

$$\langle \nabla \omega(x) - \nabla \omega(z), x - z \rangle \geq \sigma_\omega \|x - z\|^2, \quad \forall x, z \in X.$$

在实际计算中, 可根据 X 的特殊结构选择适当的 $\omega(x)$, 以提高计算效率。

记集合 X 对应于函数 $\omega(x)$ 的直径为:

$$D_{\omega, X}^2 := \max \{ \omega(x) - [\omega(z) + \langle \nabla \omega(z), x - z \rangle], \forall x, z \in X \}. \quad (4)$$

因此, 可以得到

$$\|x - z\|^2 \leq \frac{2}{\sigma_\omega} D_{\omega, X}^2 =: \Omega_{\omega, X}, \quad \forall x, z \in X. \quad (5)$$

根据函数的非精确信息, 定义近似线性化函数:

$$\begin{cases} l_f^k(\cdot) := f_{x^k} + \langle \tilde{g}_f^k, \cdot - x^k \rangle, \\ l_c^k(\cdot) := c_{x^k} + \langle \tilde{g}_c^k, \cdot - x^k \rangle. \end{cases}$$

由函数的凸性可得 $l_f^k(\cdot) \leq f(\cdot)$, $l_c^k(\cdot) \leq c(\cdot)$, 并且在点 x^k 处有 $l_f^k(x^k) = f_{x^k}$, $l_c^k(x^k) = c_{x^k}$ 成立。定

义函数 $f(x), c(x)$ 在第 k 次迭代的多面体模型:

$$\hat{f}^k(x) := \max_{j \in J_f^k} l_f^j(x), \quad \hat{c}^k(x) := \max_{j \in J_c^k} l_c^j(x),$$

其中, $J_f^k, J_c^k \subseteq \{1, \dots, k\}$ 分别是线性化函数 $l_f^j(x), l_c^j(x)$ 对应的某些迭代指标构成的集合。根据多面体模型的定义可以得到:

$$\hat{f}^k(x) \leq f(x), \quad \hat{c}^k(x) \leq c(x), \quad \forall k. \quad (6)$$

设 $f_{\text{low}}^k (\leq f^*)$ 是问题最优值的一个下界, 定义非精确改进函数如下:

$$\tilde{h}(x; f_{\text{low}}^k) = \max\{f_x - f_{\text{low}}^k, c_x\}. \quad (7)$$

用以上函数作为近似最优性判别函数, 并采用如下方式记录改进函数当前的最小值和相应的迭代点:

$$h_{\text{rec}}^k := \begin{cases} \tilde{h}(x^0; f_{\text{low}}^0), & k = 0, \\ \min \left\{ \min_{j \in J_f^k \cap J_c^k} \tilde{h}(x^j; f_{\text{low}}^k), h_{\text{rec}}^{k-1} \right\}, & k > 0, \end{cases} \quad (8)$$

$$x_{\text{rec}}^k \in \{x^j\}_{j \leq k} \quad \text{s.t.} \quad \tilde{h}(x_{\text{rec}}^k; f_{\text{low}}^k) = h_{\text{rec}}^k, \quad (9)$$

由(8)和(9)易知 $\{h_{\text{rec}}^k\}$ 是单调不增的序列。

设 f_{lev}^k 为当前水平值, 采用上述邻近函数代替欧氏距离, 本文算法每次迭代求解如下子问题:

$$x^{k+1} := \arg \min_{x \in X^k} \varphi(x; \hat{x}^k), \quad (10)$$

其中, \hat{x}^k 为当前稳定中心, 水平集 X^k 定义如下:

$$X^k := \{x \in X : \hat{f}^k(x) \leq f_{\text{lev}}^k, \hat{c}^k(x) \leq 0\}. \quad (11)$$

以下引理启发了下界 f_{low}^k 的一个更新规则: 当 X^k 是空集时, 可用当前的水平值 f_{lev}^k 来更新下界, 即令 $f_{\text{low}}^k = f_{\text{lev}}^k$ 。

引理 1 [5]: 如果水平集 X^k 是空集, 则水平值 f_{lev}^k 是问题(1)的最优值的下界。

下面将给出子问题(10)的最优解的一些重要性质, 其证明是文献[5]命题 1 的推广。

命题 1: 假设约束规格 $X^k \cap \text{ri}X \neq \emptyset$ 成立, 则 x^{k+1} 为子问题(10)最优解的充要条件是 $x^{k+1} \in X$, $\hat{f}^k(x^{k+1}) \leq f_{\text{lev}}^k$, $\hat{c}^k(x^{k+1}) \leq 0$, 并且存在向量 $s^k \in N_X(x^{k+1})$, $\hat{g}_f^k \in \partial \hat{f}^k(x^{k+1})$, $\hat{g}_c^k \in \partial \hat{c}^k(x^{k+1})$ 和 $\mu_f^k, \mu_c^k \geq 0$ 使得

$$\nabla \varphi(x^{k+1}; \hat{x}^k) + s^k + \mu_f^k \hat{g}_f^k + \mu_c^k \hat{g}_c^k = 0, \quad \mu_f^k (\hat{f}^k(x^{k+1}) - f_{\text{lev}}^k) = 0, \quad \mu_c^k \hat{c}^k(x^{k+1}) = 0. \quad (12)$$

此外, 聚集线性化

$$\bar{f}^{a_k}(x) := \hat{f}^k(x^{k+1}) + \langle \hat{g}_f^k, x - x^{k+1} \rangle \text{ 满足 } \bar{f}^{a_k}(x) \leq \hat{f}^k(x) \leq f(x), \quad \forall x \in X, \quad (13)$$

$$\bar{c}^{a_k}(x) := \hat{c}^k(x^{k+1}) + \langle \hat{g}_c^k, x - x^{k+1} \rangle \text{ 满足 } \bar{c}^{a_k}(x) \leq \hat{c}^k(x) \leq c(x), \quad \forall x \in X. \quad (14)$$

同时, 有

$$\arg \min_{x \in X^k} \varphi(x; \hat{x}^k) = \arg \min_{x \in X^{a_k}} \varphi(x; \hat{x}^k) \quad (15)$$

成立。其中, 聚集水平集 X^{a_k} 定义为 $X^{a_k} := \{x \in X : \bar{f}^{a_k}(x) \leq f_{\text{lev}}^k, \bar{c}^{a_k}(x) \leq 0\}$ 。

下面给出本文算法的具体步骤:

算法 1:

步骤 0 (初始化): 选取初始点 $x^0 \in X$, 参数 $\gamma \in (0,1)$, 终止参数 $\delta_{\text{Tol}} \geq 0$, 束的最大容量 $\text{nb} \geq 2$. 选取初始下界 $f_{\text{low}}^0 \leq f^*$, 计算初始近似函数值和近似次梯度 (f_{x^0}, \tilde{g}_f^0) , (c_{x^0}, \tilde{g}_c^0) . 令 $\hat{x}^0 = x^0$, $k=0$, $l=0$, $k(l)=0$, $J_f^0 = \{0\}$, $J_c^0 = \{0\}$.

步骤 1 (更新记录值): 分别通过(8)和(9)更新 h_{rec}^k 和 x_{rec}^k .

步骤 2 (终止测试): 如果 $h_{\text{rec}}^k \leq \delta_{\text{Tol}}$, 则算法终止并输出 x_{rec}^k .

步骤 3 (下降测试): 如果 $h_{\text{rec}}^k \leq (1-\gamma)h_{\text{rec}}^{k(l)}$, 则令 $l:=l+1$, $k(l):=k$, 并选取 $\hat{x}^k \in \{x^j : j \in J_f^k \cap J_c^k\}$.

步骤 4 (更新水平集): 令 $f_{\text{lev}}^k := f_{\text{low}}^k + \gamma h_{\text{rec}}^k$, 更新水平集 X^k .

步骤 5 (可行性检测): 如果 X^k 为非空集, 则转到步骤 6; 否则, 转步骤 7.

步骤 6 (子问题求解): 求解(10)产生新迭代点 x^{k+1} , 并计算相应的近似函数值和近似次梯度 $(f_{x^{k+1}}, \tilde{g}_f^{k+1})$, $(c_{x^{k+1}}, \tilde{g}_c^{k+1})$. 令 $f_{\text{low}}^{k+1} := f_{\text{low}}^k$, $\hat{x}^{k+1} := \hat{x}^k$.

步骤 7 (更新下界): 令 $l:=l+1$, $k(l):=k$, $f_{\text{low}}^k := f_{\text{lev}}^k$, 选择 $\hat{x}^k \in \{x^j : j \in J_f^k \cap J_c^k\}$, 返回步骤 1.

步骤 8 (束管理): 如果 $|J_f^k| < \text{nb}$, 则 $J_f^{k+1} = J_f^k \cup \{k+1\}$; 否则选择指标集 $I_f \subseteq J_f^k$, $|I_f| \geq 2$, 更新束 $J_f^{k+1} = J_f^k \setminus I_f \cup \{k+1, a_k\}$. 如果 $|J_c^k| < \text{nb}$, 则 $J_c^{k+1} = J_c^k \cup \{k+1\}$; 否则选择指标集 $I_c \subseteq J_c^k$, $|I_c| \geq 2$, 更新束 $J_c^{k+1} = J_c^k \setminus I_c \cup \{k+1, a_k\}$.

步骤 9 (循环): 令 $k:=k+1$, 返回步骤 1.

注 1: 令 K^l 为第 l 个循环的指标集, 由算法可知对任意的 $k \in K^l$, 稳定中心 \hat{x}^k 和下界 f_{low}^k 不变. 因此, 对每一个固定的 l , 序列 $\{f_{\text{lev}}^j\}_{j \in K^l}$ 是非增的.

下面我们分别讨论误差固定不变和随着迭代趋近于零两种情况, 即

情形 I: $\varepsilon_f^k \equiv \varepsilon_f \geq 0, \varepsilon_c^k \equiv \varepsilon_c \geq 0, \forall k$;

情形 II: $\varepsilon_f^k \rightarrow 0, \varepsilon_c^k \rightarrow 0, k \rightarrow \infty$.

引理 2: 假设 $\lim_{k \rightarrow \infty} h_{\text{rec}}^k \leq 0$, 则

a) 对于情形 I, 序列 $\{x_{\text{rec}}^k\}$ 的任意聚点都是问题(1)的 ε 最优解, 其中 $\varepsilon := \max\{\varepsilon_f, \varepsilon_c\}$;

b) 对于情形 II, 序列 $\{x_{\text{rec}}^k\}$ 的任意聚点都是问题(1)的最优解.

特别地, 如果对某个 k , 有 $h_{\text{rec}}^k \leq 0$, 则对于情形 I, x_{rec}^k 是问题(1)的一个 ε 最优解; 对于情形 II, x_{rec}^k 是问题(1)的一个 ε^{j_k} 最优解, 其中 $\varepsilon^{j_k} := \max\{\varepsilon_f^{j_k}, \varepsilon_c^{j_k}\}$, $j_k (\leq k)$ 是使得 $h_{\text{rec}}^k = \tilde{h}(x_{\text{rec}}^k; f_{\text{low}}^{j_k})$ 成立的某个指标.

证明: 由 h_{rec}^k 的定义(8)和(9)以及 X 有界, 可以得到序列 $\{h_{\text{rec}}^k\}$ 单调有界, 因此必有极限. 故不失一般性, 可设序列 $\{f_{x_{\text{rec}}^k}\}$, $\{c_{x_{\text{rec}}^k}\}$ 有极限. 由假设 $\lim_{k \rightarrow \infty} h_{\text{rec}}^k \leq 0$, 可得

$$0 \geq \lim_k (f_{x_{\text{rec}}^k} - f_{\text{low}}^{j_k}) \geq \lim_k (f(x_{\text{rec}}^k) - \varepsilon_f^{j_k} - f_{\text{low}}^{j_k}) \geq \lim_k (f(x_{\text{rec}}^k) - \varepsilon_f^{j_k} - f^*),$$

$$0 \geq \lim_k c_{x_{\text{rec}}^k} \geq \lim_k (c(x_{\text{rec}}^k) - \varepsilon_c^{j_k}).$$

设 \bar{x} 是序列 $\{x_{\text{rec}}^k\}$ 的一个聚点, 对于情形 I, 由以上两个不等式可以得到 $f(\bar{x}) \leq f^* + \varepsilon_f$ 和 $c(\bar{x}) \leq \varepsilon_c$, 因此 \bar{x} 是问题(1)的 ε 最优解. 对于情形 II, 再次利用以上两个不等式可得 $f(\bar{x}) \leq f^*$ 和 $c(\bar{x}) \leq 0$, 故 \bar{x} 是问题(1)的最优解.

特别地, 如果对某个 k , $h_{\text{rec}}^k \leq 0$ 成立, 类似以上分析可知, 对于情形 I 有: $f(x_{\text{rec}}^k) \leq f^* + \varepsilon_f$ 和 $c(x_{\text{rec}}^k) \leq \varepsilon_c$, 故 x_{rec}^k 是问题(1)的 ε 最优解; 对于情形 II, 有 $f(x_{\text{rec}}^k) \leq f^* + \varepsilon_f^{j_k}$ 和 $c(x_{\text{rec}}^k) \leq \varepsilon_c^{j_k}$, 从而 x_{rec}^k 是问题(1)的一个 ε^{j_k} 最优解. \square

3. 收敛性与复杂度分析

本节将证明算法 1 的全局收敛性并分析其计算复杂度。记 Λ 为近似次梯度的上界，即 $\|\tilde{g}_f^k\| \leq \Lambda$ ， $\|\tilde{g}_c^k\| \leq \Lambda, \forall k$ 。由算法步骤 2 和引理 2 可知，若 $h_{\text{rec}}^k \leq 0$ ，则算法终止，并且 x_{rec}^k 是问题(1)的近似最优解，因此在下面分析中将假设 $h_{\text{rec}}^k > 0, \forall k$ 。下面引理给出相邻两个迭代点之间距离的下界，其证明类似于文[5]的引理 6。

引理 3: 算法 1 产生的迭代点满足下列关系:

$$\|x^{k+1} - x^k\| \geq \frac{1-\gamma}{\Lambda} h_{\text{rec}}^k, \quad k > k(l),$$

$$\|x^{k+1} - \hat{x}^k\| \geq \frac{1-\gamma}{\Lambda} h_{\text{rec}}^k, \quad k = k(l).$$

证明: 对于任意的 k 和 $j \in J_f^k \cap J_c^k$ ，根据子问题 $x^{k+1} := \arg \min_{x \in X^k} \varphi(x; \hat{x}^k)$ 可以得到 $x^{k+1} \in X^k$ 。由 X^k 的定义(11)结合(3)式，可得

$$f_{x^j} + \langle \tilde{g}_f^j, x^{k+1} - x^j \rangle \leq f_{\text{lev}}^k,$$

$$c_{x^j} + \langle \tilde{g}_c^j, x^{k+1} - x^j \rangle \leq 0.$$

即

$$f_{x^j} - f_{\text{lev}}^k \leq \langle \tilde{g}_f^j, x^{k+1} - x^j \rangle,$$

$$c_{x^j} \leq \langle \tilde{g}_c^j, x^{k+1} - x^j \rangle.$$

结合 Cauchy-Schwarz 不等式可以得到

$$f_{x^j} - f_{\text{lev}}^k \leq \langle \tilde{g}_f^j, x^{k+1} - x^j \rangle \leq \|\tilde{g}_f^j\| \|x^{k+1} - x^j\| \leq \Lambda \|x^{k+1} - x^j\|.$$

类似的，可以得到

$$c_{x^j} \leq \langle \tilde{g}_c^j, x^{k+1} - x^j \rangle \leq \|\tilde{g}_c^j\| \|x^{k+1} - x^j\| \leq \Lambda \|x^{k+1} - x^j\|.$$

再结合 $f_{\text{lev}}^k = f_{\text{low}}^k + \gamma h_{\text{rec}}^k$ 及 $h_{\text{rec}}^k > 0$ ，可得

$$\begin{aligned} \Lambda \|x^{k+1} - x^j\| &\geq \max \{ f_{x^j} - f_{\text{low}}^k - \gamma h_{\text{rec}}^k, c_{x^j} \} \\ &\geq \max \{ f_{x^j} - f_{\text{low}}^k - \gamma h_{\text{rec}}^k, c_{x^j} - \gamma h_{\text{rec}}^k \} \\ &= -\gamma h_{\text{rec}}^k + \max \{ f_{x^j} - f_{\text{low}}^k, c_{x^j} \} \\ &= -\gamma h_{\text{rec}}^k + \tilde{h}(x^j; f_{\text{low}}^k) \quad \text{由(7)} \\ &= (1-\gamma) h_{\text{rec}}^k. \quad \text{由(8)} \end{aligned}$$

当 $k > k(l)$ ，则束管理确保 $k \in J_f^k \cap J_c^k$ 。故令 $j = k$ 使得 $\|x^{k+1} - x^k\| \geq \frac{1-\gamma}{\Lambda} h_{\text{rec}}^k$ 成立，当 $k = k(l)$ 时，选取 \hat{x}^k 为稳定中心，故存在 $j \in J_f^k \cap J_c^k$ 使得 $x^j = \hat{x}^k$ ，即 $\|x^{k+1} - \hat{x}^k\| \geq \frac{1-\gamma}{\Lambda} h_{\text{rec}}^k$ 成立。

□

下面将证明每一个循环 K^l 中的迭代次数是有限的。

引理 4: 对于任意的 $l \geq 0$ ，在第 l 个循环中的迭代指标 $k (k \in K^l)$ 满足:

$$k - k(l) + 1 \leq \Omega_{\omega, X} \left(\frac{\Lambda}{(1-\gamma)h_{\text{rec}}^k} \right)^2 + 1.$$

证明: 对任意的 $k > k(l)$, 由子问题(10)知 $x^k = \arg \min_{x \in X^{k-1}} \varphi(x; \hat{x}^{k-1})$, 故由一阶最优性条件可得:

$$\langle \nabla \varphi(x^k; \hat{x}^{k-1}), x - x^k \rangle \geq 0, \quad \forall x \in X^{k-1}. \quad (16)$$

i) 假如在第 $k-1$ 步到第 k 步没有束压缩机制, 那么根据 $\hat{f}^k(x)$ 的定义(6)可知 $\hat{f}^k(x) \geq \hat{f}^{k-1}(x)$, $\hat{c}^k(x) \geq \hat{c}^{k-1}(x), \forall x \in R^n$. 根据注 1 可得 $f_{\text{lev}}^k \leq f_{\text{lev}}^{k-1}$, 从而 $X^k \subseteq X^{k-1}$. 因为 $k \in K^l$, 所以有 X^k 非空, 并且 $x^{k+1} \in X^k$. 又因为在每个 l 循环中稳定中心不变, 即: $\hat{x}^{k-1} = \hat{x}^k$, 从而可以得到 $\langle \nabla \varphi(x^k; \hat{x}^k), x^{k+1} - x^k \rangle \geq 0$.

ii) 若在第 $k-1$ 步到第 k 步有束压缩, 则聚集指标 $a_k \in J_f^k, a_k \in J_c^k$, 故 $\hat{f}^k(x) \geq \bar{f}^{a_k}(x)$, $\hat{c}^k(x) \geq \bar{c}^{a_k}(x)$, $\forall x \in R^n$, 从而 $X^k \subseteq X^{a_k}$. 由(15)可知 $x^k = \arg \min_{x \in X^{a_k}} \{ \varphi(x; \hat{x}^{k-1}) \}$, 根据一阶最优性条件有:

$$\langle \nabla \varphi(x^k; \hat{x}^{k-1}), x - x^k \rangle \geq 0, \quad \forall x \in X^{a_k}. \quad (17)$$

类似地, 可利用稳定中心在同一个 l 循环中不变, 即 $\hat{x}^{k-1} = \hat{x}^k$, 并且 $x^{k+1} \in X^{a_k}$, 结合(17)得到 $\langle \nabla \varphi(x^k; \hat{x}^k), x^{k+1} - x^k \rangle \geq 0$. 又因为 $k \in K^l$, $h_{\text{rec}}^k > \delta_{\text{Tot}}$, 算法不终止且 $\hat{x}^{k-1} = \hat{x}^k$, 于是

$$\langle \nabla \varphi(x^k; \hat{x}^{k(l)}), x^{k+1} - x^k \rangle \geq 0.$$

由 $\varphi(x; \hat{x}^{k(l)})$ 是强凸函数, 有

$$\frac{\sigma}{2} \|x^{k+1} - x^k\|^2 \leq \varphi(x^{k+1}; \hat{x}^{k(l)}) - \varphi(x^k; \hat{x}^{k(l)}) - \langle \nabla \varphi(x^k; \hat{x}^{k(l)}), x^{k+1} - x^k \rangle \leq \varphi(x^{k+1}; \hat{x}^{k(l)}) - \varphi(x^k; \hat{x}^{k(l)}).$$

从而有

$$\frac{\sigma}{2} \|x^{k+1} - x^k\|^2 \leq \varphi(x^{k+1}; \hat{x}^{k(l)}) - \varphi(x^k; \hat{x}^{k(l)}).$$

对上式从 $k(l)+1$ 到 k 求和, 得到:

$$\frac{\sigma}{2} \sum_{j=k(l)+1}^k \|x^{j+1} - x^j\|^2 \leq \varphi(x^{k+1}; \hat{x}^{k(l)}) - \varphi(x^{k(l)+1}; \hat{x}^{k(l)}). \quad (18)$$

对不等式(18)左边缩小, 结合引理 3 可得

$$\frac{\sigma}{2} \sum_{j=k(l)+1}^k \|x^{j+1} - x^j\|^2 \geq \frac{\sigma}{2} \sum_{j=k(l)+1}^k \left(\frac{1-\gamma}{\Lambda} h_{\text{rec}}^j \right)^2 \geq \frac{\sigma}{2} \sum_{j=k(l)+1}^k \left(\frac{1-\gamma}{\Lambda} h_{\text{rec}}^k \right)^2.$$

对不等式(18)右边放大, 由 $\varphi(x; \hat{x})$ 的性质以及(4)可得

$$\begin{aligned} & \varphi(x^{k+1}; \hat{x}^{k(l)}) - \varphi(x^{k(l)+1}; \hat{x}^{k(l)}) \\ & \leq \varphi(x^{k+1}; \hat{x}^{k(l)}) \leq \max_{x \in X} \{ \varphi(x; \hat{x}^{k(l)}) \} \leq \max_{x, y \in X} \{ \varphi(x; y) \} = D_{\omega, X}^2. \end{aligned}$$

于是

$$D_{\omega, X}^2 \geq \frac{\sigma}{2} \sum_{j=k(l)+1}^k \left(\frac{1-\gamma}{\Lambda} h_{\text{rec}}^k \right)^2 = \frac{\sigma}{2} (k - k(l)) \left(\frac{1-\gamma}{\Lambda} h_{\text{rec}}^k \right)^2,$$

从而结合(5)有

$$k - k(l) \leq \frac{2D_{\omega,X}^2}{\sigma} \left(\frac{\Lambda}{(1-\gamma)h_{\text{rec}}^k} \right)^2 = \Omega_{\omega,X} \left(\frac{\Lambda}{(1-\gamma)h_{\text{rec}}^k} \right)^2.$$

□

以下定理给出了算法 1 的全局收敛性。

定理 1: 假设 $\delta_{\text{Tol}} = 0$ 且算法不终止, 则 $\lim_k h_{\text{rec}}^k \leq 0$, 并且

a) 如果 $\varepsilon_f^k \equiv \varepsilon_f \geq 0$, $\varepsilon_c^k \equiv \varepsilon_c \geq 0$, $\forall k$, 则序列 $\{x_{\text{rec}}^k\}$ 的任意聚点是问题(1)的 ε 最优解, 其中 $\varepsilon := \max\{\varepsilon_f, \varepsilon_c\}$ 。

b) 如果 $\varepsilon_f^k \rightarrow 0$, $\varepsilon_c^k \rightarrow 0$, $k \rightarrow \infty$, 则序列 $\{x_{\text{rec}}^k\}$ 的任意聚点是问题(1)的最优解。

证明: 首先, 类似于文献[5]中的定理 4, 可以证明 $\lim_k h_{\text{rec}}^k \leq 0$ 。其次, 根据引理 2 易知定理成立。

□

下面给出算法 1 的计算复杂度。

定理 2: 设 $-\infty < f_{\text{low}}^0 \leq f^*$, $\delta_{\text{Tol}} > 0$, 且不考虑步骤 3, 则算法 1 执行的迭代次数的上界为:

$$\left(1 + \frac{f^* - f_{\text{low}}^0}{\gamma\delta_{\text{Tol}}} \right) \left(\Omega_{\omega,X} \left(\frac{\Lambda}{(1-\gamma)\delta_{\text{Tol}}} \right)^2 + 1 \right).$$

证明: 根据算法 1 可知, 当 X^k 为空集时, 最优值 f^* 的下界 f_{low}^k 增加了 $\gamma h_{\text{rec}}^k (> \gamma\delta_{\text{Tol}})$ 。即

$$f^* \geq f_{\text{low}}^0 + \gamma h_{\text{rec}}^1 + \cdots + \gamma h_{\text{rec}}^k \geq f_{\text{low}}^0 + \gamma\delta_{\text{Tol}} + \cdots + \gamma\delta_{\text{Tol}}.$$

因为 f_{low}^0 是有限的, 所以 X^k 出现空集的次数也是有限的, 令 N 为出现空集的次数, 即 N 是有限的。因此, $N \cdot \gamma\delta_{\text{Tol}} \leq f^* - f_{\text{low}}^0$, 从而有:

$$N \leq \frac{f^* - f_{\text{low}}^0}{\gamma\delta_{\text{Tol}}}.$$

若算法在第 k 次迭代不终止, 有 $h_{\text{rec}}^k > \delta_{\text{Tol}}$, 根据引理 4 可知算法中每个 K^l 至多有 $\Omega_{\omega,X} \left(\frac{\Lambda}{(1-\gamma)h_{\text{rec}}^k} \right)^2 + 1$

次迭代。那么有:

$$\Omega_{\omega,X} \left(\frac{\Lambda}{(1-\gamma)h_{\text{rec}}^k} \right)^2 + 1 \leq \Omega_{\omega,X} \left(\frac{\Lambda}{(1-\gamma)\delta_{\text{Tol}}} \right)^2 + 1.$$

令 $k_{\delta_{\text{Tol}}}$ 是使得 $h_{\text{rec}}^k > \delta_{\text{Tol}}$ 成立的最大指标集, 那么

$$k_{\delta_{\text{Tol}}} \leq \left(1 + \frac{f^* - f_{\text{low}}^0}{\gamma\delta_{\text{Tol}}} \right) \left(\Omega_{\omega,X} \left(\frac{\Lambda}{(1-\gamma)\delta_{\text{Tol}}} \right)^2 + 1 \right).$$

□

4. 结束语

本文提出了一个基于非精确数据的带有非欧氏距离的约束水平束方法, 该方法结合非精确信息并引入 Bregman 距离代替传统的欧氏距离, 充分利用可行集的几何集合, 加快算法的收敛速度, 减少计算量。最后证明了算法的全局收敛性并分析了迭代复杂度。

基金项目

获国家自然科学基金项目(11761013, 71861002); 广西自然科学基金项目(2018GXNSFFA281007; 2017GXNSFBA198238)资助。

参考文献

- [1] Kiwiel, K.C. (1985) An Exact Penalty Function Algorithm for Non-Smooth Convex Constrained Minimization Problems. *IMA Journal of Numerical Analysis*, **5**, 111-119. <https://doi.org/10.1093/imanum/5.1.111>
- [2] Karas, E., Ribeiro, A., Sagastizábal, C., et al. (2009) A Bundle-Filter Method for Nonsmooth Convex Constrained Optimization. *Mathematical Programming*, **116**, 297-320. <https://doi.org/10.1007/s10107-007-0123-7>
- [3] Lemarechal, C., Nemirovskii, A. and Nesterov, Y. (1995) New Variants of Bundle Methods. *Mathematical Programming*, **69**, 111-147. <https://doi.org/10.1007/BF01585555>
- [4] Fletcher, R. and Leyffer, S. (2002) Nonlinear Programming without a Penalty Function. *Mathematical Programming*, **91**, 239-269. <https://doi.org/10.1007/s101070100244>
- [5] Van Ackooij, W. and De Oliveira, W. (2014) Level Bundle Methods for Constrained Convex Optimization with Various Oracles. *Computational Optimization and Applications*, **57**, 555-597. <https://doi.org/10.1007/s10589-013-9610-3>
- [6] Kiwiel, K.C. (1999) A Bundle Bregman Proximal Method for Convex Nondifferentiable Minimization. *Mathematical Programming*, **85**, 241-258. <https://doi.org/10.1007/s101070050056>
- [7] Ben-Tal, A. and Nemirovski, A. (2005) Non-Euclidean Restricted Memory Level Method for Large-Scale Convex Optimization. *Mathematical Programming*, **102**, 407-456. <https://doi.org/10.1007/s10107-004-0553-4>
- [8] 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.
- [9] 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.