

非凸非光滑优化问题的惯性 Bregman ADMM 的收敛性分析*

魏 娇, 罗洪林

(重庆师范大学 数学科学学院, 重庆 401331)

摘要:【目的】针对具有可分结构的非凸非光滑优化问题,提出一种内置惯性 Bregman 交替方向乘子法。【方法】为了加快算法的收敛速度,在 Bregman 交替方向乘子法的框架下,对子问题中的 Bregman 度量内置惯性项。【结果】在生成的点列有界的条件下,利用 Kurdyka-Lojasiewicz 性质,证明了算法的渐进收敛性。【结论】数值实验结果表明了该算法的有效性。
关键词:具有可分结构的非凸非光滑优化;内置惯性 Bregman 交替方向乘子法;Kurdyka-Lojasiewicz 性质;渐进收敛性
中图分类号: O221.2 **文献标志码:** A **文章编号:** 1672-6693(2022)05-0014-10

20 世纪 70 年代初, Gabay 等人^[1]提出了交替方向乘子法(Alternating direction method of multipliers, ADMM)。ADMM 是一类重要的分布式算法,它将问题的求解分解成规模更小的一系列子问题,通过子问题的有效求解生成点列逐步逼近原问题的解。ADMM 广泛应用于机器学习^[2]、大规模电网^[3]、信号处理^[4]、经济调度^[5]等领域。针对凸可分的目标函数,ADMM 在较弱的条件下收敛^[6],收敛率为 $o(1/t)$ 。

本文考虑如下形式的非凸非光滑优化问题:

$$\begin{aligned} \min & f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{z}), \\ \text{s. t. } & \mathbf{Ax} + \mathbf{By} + \mathbf{Cz} = \mathbf{0}. \end{aligned} \tag{1}$$

其中: $\mathbf{A} \in \mathbf{R}^{m \times n_1}$, $\mathbf{B} \in \mathbf{R}^{m \times n_2}$, $\mathbf{C} \in \mathbf{R}^{m \times n_3}$ 为给定的系数矩阵, $f: \mathbf{R}^{n_1} \rightarrow \mathbf{R} \cup \{\infty\}$ 与 $g: \mathbf{R}^{n_2} \rightarrow \mathbf{R} \cup \{\infty\}$ 为两个下半连续函数, $h: \mathbf{R}^{n_3} \rightarrow \mathbf{R} \cup \{\infty\}$ 是连续可微函数。问题(1)的增广拉格朗日函数定义为:

$$L_\alpha(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda}) = f(\mathbf{x}) + g(\mathbf{y}) + h(\mathbf{z}) + \langle \boldsymbol{\lambda}, \mathbf{Ax} + \mathbf{By} + \mathbf{Cz} \rangle + \frac{\alpha}{2} \|\mathbf{Ax} + \mathbf{By} + \mathbf{Cz}\|^2.$$

求解问题(1)的经典 ADMM 的第 $k+1$ 步迭代式为:

$$\begin{cases} \mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbf{R}^{n_1}} L_\alpha(\mathbf{x}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) \\ \mathbf{y}^{k+1} = \arg \min_{\mathbf{y} \in \mathbf{R}^{n_2}} L_\alpha(\mathbf{x}^{k+1}, \mathbf{y}, \mathbf{z}^k, \boldsymbol{\lambda}^k) \\ \mathbf{z}^{k+1} = \arg \min_{\mathbf{z} \in \mathbf{R}^{n_3}} L_\alpha(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k) \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha(\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} + \mathbf{Cz}^{k+1}) \end{cases}.$$

即便在凸性假设条件下,该算法也不一定收敛^[7]。当目标函数满足一定的强凸性假设时,该算法具有全局收敛性^[8-9]。2018 年,徐宗本等人^[10]对非凸非光滑优化问题(1)提出了一种 Bregman ADMM,并给出了算法的全局收敛性,算法的第 $k+1$ 步迭代式为:

$$\begin{cases} \mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbf{R}^{n_1}} L_\alpha(\mathbf{x}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \Delta_{\varphi_1}(\mathbf{x}, \mathbf{x}^k) \\ \mathbf{y}^{k+1} = \arg \min_{\mathbf{y} \in \mathbf{R}^{n_2}} L_\alpha(\mathbf{x}^{k+1}, \mathbf{y}, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \Delta_{\varphi_2}(\mathbf{y}, \mathbf{y}^k) \\ \mathbf{z}^{k+1} = \arg \min_{\mathbf{z} \in \mathbf{R}^{n_3}} L_\alpha(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k) + \Delta_{\varphi_3}(\mathbf{z}, \mathbf{z}^k) \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha(\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} + \mathbf{Cz}^{k+1}) \end{cases}, \tag{2}$$

* 收稿日期:2021-10-21 修回日期:2021-11-30 网络出版时间:2022-9-20 10:04

资助项目:国家自然科学基金(No. 11991024; No. 11771064);重庆市高校创新研究群体项目(No. 20A110029)

第一作者简介:魏娇,女,研究方向为最优化理论与算法, E-mail: 1980237356@qq.com;通信作者:罗洪林,男,教授,博士, E-mail: luohonglin@cqnu.edu.cn

网络出版地址: <https://kns.cnki.net/kcms/detail/50.1165.n.20220916.1656.004.html>

其中: Δ_{φ_i} ($i=1,2,3$)为给定的 Bregman 度量。这种算法可以针对不同的目标函数选择不同的 Bregman 度量以提高算法的效率。

2021 年, Xu 等人^[11]考虑了如下形式的非凸非光滑的优化问题:

$$\begin{aligned} \min & f(\mathbf{x}) + g(\mathbf{y}), \\ \text{s. t.} & \mathbf{Ax} + \mathbf{By} = \mathbf{b}. \end{aligned}$$

其中: $\mathbf{A} \in \mathbf{R}^{m \times n_1}$, $\mathbf{B} \in \mathbf{R}^{m \times n_2}$, $\mathbf{b} \in \mathbf{R}^m$, $f: \mathbf{R}^{n_1} \rightarrow \mathbf{R} \cup \{\infty\}$ 是下半连续函数, $g: \mathbf{R}^{n_2} \rightarrow \mathbf{R} \cup \{\infty\}$ 是连续可微函数。将增广拉格朗日函数定义为: $L_\alpha(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) = f(\mathbf{x}) + g(\mathbf{y}) + \langle \boldsymbol{\lambda}, \mathbf{Ax} + \mathbf{By} - \mathbf{b} \rangle + \frac{\alpha}{2} \|\mathbf{Ax} + \mathbf{By} - \mathbf{b}\|^2$, 并引入一种惯性

Bregman ADMM:

$$\begin{cases} \mathbf{x}^{k+1} \in \arg \min_{\mathbf{x} \in \mathbf{R}^{n_1}} L_\alpha(\mathbf{x}, \mathbf{y}^k, \boldsymbol{\lambda}^k) + \Delta_{\varphi_1}(\mathbf{x}, \mathbf{x}^k) + \theta_k(\mathbf{x}, \mathbf{x}^{k-1} - \mathbf{x}^k) \\ \mathbf{x}_{\text{ad}}^{k+1} = \beta \mathbf{Ax}^{k+1} + (1-\beta)(\mathbf{b} - \mathbf{By}^k) \\ \mathbf{y}^{k+1} \in \arg \min_{\mathbf{y} \in \mathbf{R}^{n_2}} g(\mathbf{y}) + \langle \boldsymbol{\lambda}, \mathbf{By} \rangle + \Delta_{\varphi_2}(\mathbf{y}, \mathbf{y}^k) + \frac{\alpha}{2} \|\mathbf{Ax}_{\text{ad}}^{k+1} + \mathbf{By} - \mathbf{b}\|^2 + \rho_k(\mathbf{y}, \mathbf{y}^{k-1} - \mathbf{y}^k) \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha(\mathbf{Ax}_{\text{ad}}^{k+1} + \mathbf{By}^{k+1} - \mathbf{b}) \end{cases} \quad (3)$$

可以发现, 该算法关于变量 \mathbf{x} 的迭代不但引入了 Bregman 度量 Δ_{φ_1} , 还引入了一个惯性项 $\theta_k(\mathbf{x}, \mathbf{x}^{k-1} - \mathbf{x}^k)$, 数值实验表明该惯性项对算法的收敛有加速的效用。但是, 该算法引入了惯性项后, 关于变量 \mathbf{x} 和 \mathbf{y} 的每次迭代点的选取可能不再具有唯一性(也即关于变量 \mathbf{x} 和 \mathbf{y} 的子问题的最优解可能不唯一), 这无疑为算法的具体实现带来了一定的困难。

受上述文献启发, 本文针对问题(1)提出了一种新的内置惯性 Bregman ADMM(简称 I-B-ADMM)。在算法生成的点列满足有界性的条件下, 利用 Kurdyka-Lojasiewicz 性质证明了算法的渐进收敛性, 并利用该算法求解了文献[10]中的数值算例, 数值结果显示了该算法的有效性。

1 预备知识

定义 1^[12] 若函数 $f: \mathbf{R}^n \rightarrow \mathbf{R}$ 是下半连续的, 且在 \mathbf{x}_0 处满足 Kurdyka-Lojasiewicz 性质(简称 K-L 性质), 则存在 $\eta > 0, \delta > 0, \gamma$, 使得对所有的:

$$\mathbf{x} \in O(\mathbf{x}_0, \delta) \cap \{\mathbf{x} \in \mathbf{R}^n : f(\mathbf{x}_0) < f(\mathbf{x}) < f(\mathbf{x}_0) + \eta\},$$

都有 $\gamma'(f(\mathbf{x}) - f(\mathbf{x}_0)) \text{dist}(0, \partial f(\mathbf{x})) \geq 1$, 其中 $\text{dist}(\mathbf{x}_0, \partial f(\mathbf{x})) = \inf\{\|\mathbf{x}_0 - \mathbf{y}\| : \mathbf{y} \in \partial f(\mathbf{x})\}$, γ 表示 $[0, \eta] \rightarrow \mathbf{R}^+$ 上的函数且满足: 1) γ 在 $[0, \eta]$ 上连续; 2) γ 在 $[0, \eta]$ 上光滑凹; 3) $\gamma(0) = 0, \gamma'(t) > 0, \forall t \in (0, \eta)$ 。

性质 1^[13] 设 $\Omega \subseteq \mathbf{R}^n$ 是紧集, 函数 $f: \mathbf{R}^n \rightarrow \mathbf{R}$ 是下半连续的。若函数 f 在集合 Ω 上是常数, 并在 Ω 上任一个点处满足 K-L 性质, 则存在 $\eta > 0, \delta > 0, \gamma$, 使得对 $\forall \mathbf{x}_0 \in \Omega$ 和 $\forall \mathbf{x} \in \{\mathbf{x} \in \mathbf{R}^n : \text{dist}(\mathbf{x}, \Omega) < \delta\} \cap \{\mathbf{x} \in \mathbf{R}^n : f(\mathbf{x}_0) < f(\mathbf{x}) < f(\mathbf{x}_0) + \eta\}$, 都有 $\gamma'(f(\mathbf{x}) - f(\mathbf{x}_0)) \text{dist}(0, \partial f(\mathbf{x})) \geq 1$ 。常见的满足 K-L 性质的函数有强凸函数、实解析函数、次解析函数和半代数函数。

定义 2^[14] 如果集合 C_1 被称为半代数集, 则 $C_1 = \bigcup_{j=1}^r \bigcap_{i=1}^s \{\mathbf{x} \in \mathbf{R}^n : g_{i,j}(\mathbf{x}) = 0, h_{i,j}(\mathbf{x}) < 0\}$, 其中 $g_{i,j}, h_{i,j}$ 是 $\mathbf{R}^n \rightarrow \mathbf{R}$ 上的实多项式函数。如果函数 $p: \mathbf{R}^n \rightarrow \mathbf{R}$ 的图 $G(p) = \{(\mathbf{x}, \mathbf{y}) \in \mathbf{R}^{n+1} : p(\mathbf{x}) = \mathbf{y}\}$ 在 \mathbf{R}^{n+1} 上是半代数集, 则称函数 p 是半代数函数。如常见的函数 $\|\mathbf{x}\|_1, \|\mathbf{x}\|, \|\mathbf{x}\|_\infty, \|\mathbf{Ax} - \mathbf{b}\|, \|\mathbf{Ax} - \mathbf{b}\|_q, \|\mathbf{Ax} - \mathbf{b}\|_\infty$ 等都是半代数函数。

定义 3^[10] 如果集合 C_2 被称为次解析集, 则 $C_2 = \bigcup_{u=1}^r \bigcap_{v=1}^s \{\mathbf{x} \in \mathbf{R}^n : f_{u,v}(\mathbf{x}) = 0, k_{u,v}(\mathbf{x}) < 0\}$, 其中 $f_{u,v}, k_{u,v}$ 是 $\mathbf{R}^n \rightarrow \mathbf{R}$ 上的解析函数。如果函数 $q: \mathbf{R}^n \rightarrow \mathbf{R}$ 的图在 \mathbf{R}^{n+1} 上是半代数集, 则称函数 q 是次解析函数。

值得注意的是, 实解析函数和半代数函数都是次解析函数。两个次解析函数的和不一定是次解析函数, 如果函数 φ_1 和 φ_2 是次解析函数, φ_1 将有界集映射到有界集, 那么 $\varphi_1 + \varphi_2$ 才是次解析函数。实解析函数和半代数函数的和是次解析函数。

定义 4^[10] 设 κ 是 $\mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$ 上的可微凸函数, 则称 $\Delta_\kappa(\mathbf{x}, \mathbf{y}) = \kappa(\mathbf{x}) - \kappa(\mathbf{y}) - \langle \nabla \kappa(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$ 为函数 κ

导出的 Bregman 度量。特别的,若 $\kappa(\mathbf{x}) = \|\mathbf{x}\|^2$, 则 $\Delta_\kappa(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$ 。常见的 Bregman 度量还有 Itakura-Saito 距离 $\sum_{i=1}^n x_i \left(\log \frac{x_i}{y_i}\right) - \sum_{i=1}^n (x_i - y_i)$ 、Kullback-Leibler 散度 $\sum_{i=1}^n x_i \left(\log \frac{x_i}{y_i}\right)$ 及 Mahalanobis 距离 $\|\mathbf{x} - \mathbf{y}\|_Q^2 = \langle Q\mathbf{x}, \mathbf{x} \rangle$, 其中 Q 是对称正定矩阵。

性质 2^[15] 设 κ 是 $\mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$ 上的可微强凸函数, 且强凸系数 $\delta > 0$, 有 $\Delta_\kappa(\mathbf{x}, \mathbf{y}) \geq 0, \Delta_\kappa(\mathbf{x}, \mathbf{y}) = 0$ 当且仅当 $\mathbf{x} = \mathbf{y}$ 时, 有 $\Delta_\kappa(\mathbf{x}, \mathbf{y}) \geq \frac{\delta}{2} \|\mathbf{x} - \mathbf{y}\|^2$ 。

算法 1 (I-B-ADMM)

步骤 1, 输入初始值 $\mathbf{x}^0, \mathbf{x}^1 \in \mathbf{R}^{n_1}; \mathbf{y}^0, \mathbf{y}^1 \in \mathbf{R}^{n_2}; \mathbf{z}^0, \mathbf{z}^1 \in \mathbf{R}^{n_3}; \boldsymbol{\lambda}^0 \in \mathbf{R}^m; \alpha > 0; a, b, c \in (0, 1); \varphi, \tau, \psi$ 都是强凸函数, 误差 $\zeta = 10^{-8}$ 。

步骤 2, 设置 $k = 1$;

步骤 3, $\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbf{R}^{n_1}} L_a(\mathbf{x}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \Delta_\varphi(\mathbf{x}, \hat{\mathbf{x}}^k)$, 其中 $\hat{\mathbf{x}}^k = \mathbf{x}^{k-1} + a(\mathbf{x}^k - \mathbf{x}^{k-1})$;

步骤 4, $\mathbf{y}^{k+1} = \arg \min_{\mathbf{y} \in \mathbf{R}^{n_2}} L_a(\mathbf{x}^{k+1}, \mathbf{y}, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \Delta_\tau(\mathbf{y}, \hat{\mathbf{y}}^k)$, 其中 $\hat{\mathbf{y}}^k = \mathbf{y}^{k-1} + b(\mathbf{y}^k - \mathbf{y}^{k-1})$;

步骤 5, $\mathbf{z}^{k+1} = \arg \min_{\mathbf{z} \in \mathbf{R}^{n_3}} L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}, \boldsymbol{\lambda}^k) + \Delta_\psi(\mathbf{z}, \hat{\mathbf{z}}^k)$, 其中 $\hat{\mathbf{z}}^k = \mathbf{z}^{k-1} + c(\mathbf{z}^k - \mathbf{z}^{k-1})$;

步骤 6, $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} + \mathbf{C}\mathbf{z}^{k+1})$;

步骤 7, 若终止条件 $\frac{\|(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}) - (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)\|}{\|(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k)\| + 1} \geq \xi$, 则令 $k = k + 1$, 返回 3 重新开始, 否则算法终止。

注 1 I-B-ADMM 与文献[10]中的算法作比较, 不同之处是 I-B-ADMM 在 Bregman 度量中内置了惯性项, 通过调整惯性系数 a, b, c 的取值来提高收敛速率。从数值实验看, I-B-ADMM 明显提高了收敛速率, 且直接影响了算法的迭代步数。

注 2 I-B-ADMM 与文献[11]中的算法作比较, 两个算法都引入了惯性项。(3)式中变量 \mathbf{x} 和 \mathbf{y} 每进行一次迭代时, 生成的迭代点有多个, 导致下一次迭代点的选取不唯一。而 I-B-ADMM 的好处就在于变量每进行一次迭代时, 生成的迭代点唯一。

2 收敛性分析

假设 f, g, h 为问题(1)中的函数, 满足: 1) 存在 $\mu_0 > 0$, 使得 $\mathbf{C}\mathbf{C}^T \geq \mu_0 \mathbf{I}$; 2) $\nabla h, \nabla \varphi, \nabla \tau, \nabla \psi$ 是利普希茨连续, 且利普希茨系数分别为 $l_h, l_\varphi, l_\tau, l_\psi$; 3) φ, τ, ψ 均为强凸函数, 强凸系数分别为 $\mu_\varphi, \mu_\tau, \mu_\psi$; 4) $f + g + h$ 是次解析函数; 5) 若点列 \mathbf{p}^{k+1} 是有界的, 则存在 $N \in \mathbf{R}$, 使得 $\|\mathbf{p}^k - \mathbf{p}^{k-1}\| \leq N \|\mathbf{p}^{k+1} - \mathbf{p}^k\|$ 。

引理 1 若由算法生成的序列 $\boldsymbol{\chi}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)$ 有界, 且满足假设条件 1)~3), 则存在 $\sigma_i > 0, i = 0, 1, 2$, 当 $\frac{\mu_\psi}{2} - l_\psi N(1-c) - \frac{3(l_h + l_\varphi)^2}{\alpha\mu_0} - \frac{3l_\varphi^2 c^2}{\alpha\mu_0} - \frac{3l_\psi^2(1-c)^2}{\alpha\mu_0} > 0, \frac{\mu_\varphi}{2} - l_\varphi N(1-a) > 0, \frac{\mu_\tau}{2} - l_\tau N(1-b) > 0$ 时, 有:

$$\sigma_2 (\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2) \leq \hat{L}(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k, \mathbf{z}^{k-1}, \mathbf{z}^{k-2}) - \hat{L}(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}, \mathbf{z}^k, \mathbf{z}^{k-1}),$$

其中: $\hat{L}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda}, \mathbf{u}, \mathbf{v}) = L_a(\mathbf{x}, \mathbf{y}, \mathbf{z}, \boldsymbol{\lambda}) + \frac{\sigma_0}{2} \|\mathbf{z} - \mathbf{u}\|^2 + \frac{\sigma_1}{2} \|\mathbf{u} - \mathbf{v}\|^2$ 。

证明 由 \mathbf{z} 的第 $k+1$ 步迭代有 $\nabla h(\mathbf{z}^{k+1}) + \mathbf{C}^T \boldsymbol{\lambda}^k + \alpha \mathbf{C}^T (\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} + \mathbf{C}\mathbf{z}^{k+1}) + \nabla \psi(\mathbf{z}^{k+1}) - \nabla \psi(\hat{\mathbf{z}}^k) = 0$ 。再结合 $\boldsymbol{\lambda}$ 的第 $k+1$ 步迭代整理可得:

$$\mathbf{C}^T \boldsymbol{\lambda}^{k+1} = -\nabla h(\mathbf{z}^{k+1}) - \nabla \psi(\mathbf{z}^{k+1}) + \nabla \psi(\hat{\mathbf{z}}^k). \quad (4)$$

则有:

$$\begin{aligned} \|\mathbf{C}^T (\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k)\|^2 &= \|(\nabla h(\mathbf{z}^{k+1}) - \nabla h(\mathbf{z}^k)) + (\nabla \psi(\mathbf{z}^{k+1}) - \nabla \psi(\mathbf{z}^k)) + (\nabla \psi(\hat{\mathbf{z}}^{k-1}) - \nabla \psi(\hat{\mathbf{z}}^k))\|^2 \leq \\ &(\|\nabla h(\mathbf{z}^{k+1}) - \nabla h(\mathbf{z}^k)\| + \|\nabla \psi(\mathbf{z}^{k+1}) - \nabla \psi(\mathbf{z}^k)\| + \|\nabla \psi(\hat{\mathbf{z}}^{k-1}) - \nabla \psi(\hat{\mathbf{z}}^k)\|)^2 \leq \\ &(l_h \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + l_\psi \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + l_\psi \|\hat{\mathbf{z}}^{k-1} - \hat{\mathbf{z}}^k\|)^2 = \end{aligned}$$

$$\begin{aligned} & [(l_h + l_\psi) \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + l_\psi \|\mathbf{z}^{k-2} + c(\mathbf{z}^{k-1} - \mathbf{z}^{k-2}) - \mathbf{z}^{k-1} - c(\mathbf{z}^k - \mathbf{z}^{k-1})\|]^2 = \\ & [(l_h + l_\psi) \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + l_\psi \|(1-c)(\mathbf{z}^{k-2} - \mathbf{z}^{k-1}) + c(\mathbf{z}^{k-1} - \mathbf{z}^k)\|]^2 \leq \\ & [(l_h + l_\psi) \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + l_\psi(1-c) \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\| + l_\psi c \|\mathbf{z}^{k-1} - \mathbf{z}^k\|]^2 \leq \\ & 3[(l_h + l_\psi)^2 \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + l_\psi^2(1-c)^2 \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|^2 + l_\psi^2 c^2 \|\mathbf{z}^{k-1} - \mathbf{z}^k\|^2]. \end{aligned}$$

又因为:

$$\|\mathbf{C}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k)\|^2 = \langle \mathbf{C}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k), \mathbf{C}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) \rangle = \langle \mathbf{C}\mathbf{C}^T(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k), \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k \rangle \geq \mu_0 \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2.$$

即:

$$\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \leq \frac{3(l_h + l_\psi)^2}{\mu_0} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2 c^2}{\mu_0} \|\mathbf{z}^{k-1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2(1-c)^2}{\mu_0} \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|^2. \quad (5)$$

由 \mathbf{x}^{k+1} 是极小值点, 于是有:

$$\begin{aligned} L_a(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_a(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) &\leq \Delta\varphi(\mathbf{x}^k, \hat{\mathbf{x}}^k) - \Delta\varphi(\mathbf{x}^{k+1}, \hat{\mathbf{x}}^k) = \varphi(\mathbf{x}^k) - \varphi(\mathbf{x}^{k+1}) + \\ &\langle \nabla\varphi(\hat{\mathbf{x}}^k), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle = -\Delta\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) + \langle \nabla\varphi(\hat{\mathbf{x}}^k) - \nabla\varphi(\mathbf{x}^k), \mathbf{x}^{k+1} - \mathbf{x}^k \rangle \leq \\ &-\frac{\mu_\varphi}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + l_\varphi \|\hat{\mathbf{x}}^k - \mathbf{x}^k\| \|\mathbf{x}^{k+1} - \mathbf{x}^k\| \leq \\ &-\frac{\mu_\varphi}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + l_\varphi N(1-a) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 = -\left(\frac{\mu_\varphi}{2} - l_\varphi N(1-a)\right) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2. \end{aligned}$$

同理可得:

$$\begin{aligned} L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) - L_a(\mathbf{x}^{k+1}, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) &\leq -\left(\frac{\mu_\tau}{2} - l_\tau N(1-b)\right) \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2, \\ L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) - L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^k, \boldsymbol{\lambda}^k) &\leq -\left(\frac{\mu_\psi}{2} - l_\psi N(1-c)\right) \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2. \end{aligned}$$

而 $L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}) - L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^k) = \langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{y}^{k+1} + \mathbf{C}\mathbf{z}^{k+1} \rangle = \frac{1}{\alpha} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$.

于是有:

$$\begin{aligned} L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}) - L_a(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) &\leq -\left(\frac{\mu_\varphi}{2} - l_\varphi N(1-a)\right) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \\ &\left(\frac{\mu_\tau}{2} - l_\tau N(1-b)\right) \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 - \left(\frac{\mu_\psi}{2} - l_\psi N(1-c)\right) \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \frac{1}{\alpha} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \leq \\ &-\left(\frac{\mu_\varphi}{2} - l_\varphi N(1-a)\right) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \left(\frac{\mu_\tau}{2} - l_\tau N(1-b)\right) \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 - \left(\frac{\mu_\psi}{2} - l_\psi N(1-c)\right) \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \\ &\frac{3(l_h + l_\psi)^2}{\alpha\mu_0} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2 c^2}{\alpha\mu_0} \|\mathbf{z}^{k-1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2(1-c)^2}{\alpha\mu_0} \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|^2 = \\ &-\left(\frac{\mu_\varphi}{2} - l_\varphi N(1-a)\right) \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - \left(\frac{\mu_\tau}{2} - l_\tau N(1-b)\right) \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 - \\ &\left(\frac{\mu_\psi}{2} - l_\psi N(1-c) - \frac{3(l_h + l_\psi)^2}{\alpha\mu_0}\right) \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2 c^2}{\alpha\mu_0} \|\mathbf{z}^{k-1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2(1-c)^2}{\alpha\mu_0} \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|^2, \end{aligned}$$

也即:

$$L_a(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1}) + \frac{\sigma_0}{2} \|\mathbf{z}^k - \mathbf{z}^{k+1}\|^2 + \frac{\sigma_1}{2} \|\mathbf{z}^k - \mathbf{z}^{k-1}\|^2 =$$

$$L_a(\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k) + \frac{\sigma_0}{2} \|\mathbf{z}^{k-1} - \mathbf{z}^k\|^2 + \frac{\sigma_1}{2} \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|^2 - \sigma_2 (\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2).$$

其中: $\frac{\sigma_0}{2} = \frac{3l_\psi^2 c^2}{\alpha\mu_0} + \frac{3l_\psi^2(1-c)^2}{\alpha\mu_0}$, $\frac{\sigma_1}{2} = \frac{3l_\psi^2(1-c)^2}{\alpha\mu_0}$, $\sigma_2 = \min\left\{\frac{\mu_\varphi}{2} - l_\varphi N(1-a), \frac{\mu_\tau}{2} - l_\tau N(1-b), \frac{\mu_\psi}{2} - l_\psi N(1-c) - \frac{3(l_h + l_\psi)^2}{\alpha\mu_0} - \frac{3l_\psi^2 c^2}{\alpha\mu_0} - \frac{3l_\psi^2(1-c)^2}{\alpha\mu_0}\right\}$.

证毕

引理 2 若由算法生成的序列 $\boldsymbol{\chi}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)$ 有界, 则 $\sum_{k=0}^{\infty} \|\boldsymbol{\chi}^k - \boldsymbol{\chi}^{k+1}\|^2 < \infty$.

证明 令 $\boldsymbol{\chi}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k, \mathbf{z}^{k-1}, \mathbf{z}^{k-2})$, 则 $\boldsymbol{\omega}^k$ 存在子列 $\boldsymbol{\omega}^{k_j} \rightarrow \boldsymbol{\omega}^*$. 由 f, g 是下半连续函数, 而 \hat{L} 也是下半连续函数, 则 $\liminf_{j \rightarrow \infty} \hat{L}(\boldsymbol{\omega}^{k_j}) \geq \hat{L}(\boldsymbol{\omega}^*)$, 即 $\hat{L}(\boldsymbol{\omega}^{k_j})$ 有下界. 由引理 1 知 $\hat{L}(\boldsymbol{\omega}^k)$ 单调递减, 故 $\hat{L}(\boldsymbol{\omega}^{k_j})$ 收敛, 从而 $\hat{L}(\boldsymbol{\omega}^k)$ 也收敛, 并且对任意的 k 都有 $\hat{L}(\boldsymbol{\omega}^k) \geq \hat{L}(\boldsymbol{\omega}^*)$.

现在固定 $k \in \mathbf{N}^*$, 由引理 1 可得:

$$\sigma_2 \sum_{i=0}^k (\|\mathbf{x}^{i+1} - \mathbf{x}^i\|^2 + \|\mathbf{y}^{i+1} - \mathbf{y}^i\|^2 + \|\mathbf{z}^{i+1} - \mathbf{z}^i\|^2) \leq \sum_{i=0}^k (\hat{L}(\boldsymbol{\omega}^i) - \hat{L}(\boldsymbol{\omega}^{i+1})) = \hat{L}(\boldsymbol{\omega}^0) - \hat{L}(\boldsymbol{\omega}^{k+1}) \leq \hat{L}(\boldsymbol{\omega}^0) - \hat{L}(\boldsymbol{\omega}^*).$$

$$\text{则 } \sum_{k=0}^{\infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 < \infty, \sum_{k=0}^{\infty} \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 < \infty, \sum_{k=0}^{\infty} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 < \infty.$$

由(5)式有 $\sum_{k=0}^{\infty} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 < \infty$, 于是 $\sum_{k=0}^{\infty} \|\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\|^2 < \infty$, 特别的 $\|\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\| \rightarrow 0$. 证毕

设 $\boldsymbol{\chi}^* = (\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*, \boldsymbol{\lambda}^*)$ 是 $\boldsymbol{\chi}^k$ 的聚点, 则序列 $\boldsymbol{\chi}^k$ 存在子列 $\boldsymbol{\chi}^{k_j} \rightarrow \boldsymbol{\chi}^*$. 当 $k \rightarrow \infty$ 时, 有 $\|\boldsymbol{\chi}^{k+1} - \boldsymbol{\chi}^k\| \rightarrow 0$, 则 $\boldsymbol{\chi}^{k_j}$ 和 $\boldsymbol{\chi}^{k_j+1}$ 有相同的极限点 $\boldsymbol{\chi}^*$. 从 I-B-ADMM 中可以得到:

$$\begin{cases} -\mathbf{A}^T \boldsymbol{\lambda}^{k_j} - \alpha \mathbf{A}^T (\mathbf{A} \mathbf{x}^{k_j+1} + \mathbf{A}_2 \mathbf{x}_2^{k_j} + \mathbf{C} \mathbf{z}^{k_j}) - \nabla \varphi(\mathbf{x}^{k_j+1}) + \nabla \varphi(\hat{\mathbf{x}}^{k_j}) \in \partial f(\mathbf{x}^{k_j+1}) \\ -\mathbf{B}^T \boldsymbol{\lambda}^{k_j} - \alpha \mathbf{B}^T (\mathbf{A} \mathbf{x}^{k_j+1} + \mathbf{B} \mathbf{y}^{k_j+1} + \mathbf{C} \mathbf{z}^{k_j}) - \nabla \tau(\mathbf{y}^{k_j+1}) + \nabla \tau(\hat{\mathbf{y}}^{k_j}) \in \partial g(\mathbf{y}^{k_j+1}) \\ \nabla h(\mathbf{z}^{k_j+1}) = -\mathbf{C}^T \boldsymbol{\lambda}^{k_j+1} - \nabla \psi(\mathbf{z}^{k_j+1}) + \nabla \psi(\hat{\mathbf{z}}^{k_j}) \\ \boldsymbol{\lambda}^{k_j+1} = \boldsymbol{\lambda}^{k_j} + \alpha (\mathbf{A} \mathbf{x}^{k_j+1} + \mathbf{B} \mathbf{y}^{k_j+1} + \mathbf{C} \mathbf{z}^{k_j+1}) \end{cases}.$$

令 $j \rightarrow \infty$ 时, 有 $\mathbf{A} \mathbf{x}^* + \mathbf{B} \mathbf{y}^* + \mathbf{C} \mathbf{z}^* = 0, \nabla h(\mathbf{z}^*) = -\mathbf{C}^T \boldsymbol{\lambda}^*, -\mathbf{A}^T \boldsymbol{\lambda}^* \in \partial f(\mathbf{x}^*), -\mathbf{B}^T \boldsymbol{\lambda}^* \in \partial g(\mathbf{y}^*)$.

引理 3 若由算法生成的序列 $\boldsymbol{\chi}^{k+1} = (\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{z}^{k+1}, \boldsymbol{\lambda}^{k+1})$ 有界, 则存在 $\theta > 0$, 对任意的 k 都有:

$$\text{dist}(0, \partial \hat{L}(\boldsymbol{\omega}^{k+1})) \leq \theta (\|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\| + \|\mathbf{x}^{k+1} - \mathbf{x}^k\| + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|).$$

证明 由 \mathbf{x} 的第 $k+1$ 步迭代, 有 $-\mathbf{A}^T \boldsymbol{\lambda}^k - \alpha \mathbf{A}^T (\mathbf{A} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^k + \mathbf{C} \mathbf{z}^k) - \nabla \varphi(\mathbf{x}^{k+1}) + \nabla \varphi(\hat{\mathbf{x}}^k) \in \partial f(\mathbf{x}^{k+1})$, 由 \hat{L} 的定义有:

$$\begin{aligned} \mathbf{A}^T (\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) + \alpha \mathbf{A}^T (\mathbf{B} \mathbf{y}^{k+1} - \mathbf{B} \mathbf{y}^k) + \alpha \mathbf{A}^T (\mathbf{C} \mathbf{z}^{k+1} - \mathbf{C} \mathbf{z}^k) + (\nabla \varphi(\hat{\mathbf{x}}^k) - \nabla \varphi(\mathbf{x}^{k+1})) \in \\ \partial_x \hat{L}(\boldsymbol{\omega}^{k+1}) = \partial f(\mathbf{x}^{k+1}) + \mathbf{A}^T \boldsymbol{\lambda}^{k+1} + \alpha \mathbf{A}^T (\mathbf{A} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^{k+1} + \mathbf{C} \mathbf{z}^{k+1}). \end{aligned}$$

再由 \mathbf{y} 的第 $k+1$ 步迭代, 有 $-\mathbf{B}^T \boldsymbol{\lambda}^k - \alpha \mathbf{B}^T (\mathbf{A} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^{k+1} + \mathbf{C} \mathbf{z}^k) - \nabla \tau(\mathbf{y}^{k+1}) + \nabla \tau(\hat{\mathbf{y}}^k) \in \partial g(\mathbf{y}^{k+1})$.

同理有:

$$\begin{aligned} \mathbf{B}^T (\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) + \alpha \mathbf{B}^T (\mathbf{C} \mathbf{z}^{k+1} - \mathbf{C} \mathbf{z}^k) + (\nabla \tau(\hat{\mathbf{y}}^k) - \nabla \tau(\mathbf{y}^{k+1})) \in \\ \partial_y \hat{L}(\boldsymbol{\omega}^{k+1}) = \partial_y g(\mathbf{y}^{k+1}) + \mathbf{B}^T \boldsymbol{\lambda}^{k+1} + \alpha \mathbf{B}^T (\mathbf{A} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^{k+1} + \mathbf{C} \mathbf{z}^{k+1}). \end{aligned}$$

又有:

$$\begin{aligned} \partial_z \hat{L}(\boldsymbol{\omega}^{k+1}) = \nabla h(\mathbf{z}^{k+1}) + \mathbf{C}^T \boldsymbol{\lambda}^{k+1} + \alpha \mathbf{C}^T (\mathbf{A} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^{k+1} + \mathbf{C} \mathbf{z}^{k+1}) + \sigma_0 (\mathbf{z}^{k+1} - \mathbf{z}^k) = \\ \nabla \psi(\hat{\mathbf{z}}^k) - \nabla \psi(\mathbf{z}^{k+1}) + \mathbf{C}^T (\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) + \sigma_0 (\mathbf{z}^{k+1} - \mathbf{z}^k). \end{aligned}$$

上式可由(4)式得到.

且还有:

$$\partial_x \hat{L}(\boldsymbol{\omega}^{k+1}) = \mathbf{A} \mathbf{x}^{k+1} + \mathbf{B} \mathbf{y}^{k+1} + \mathbf{C} \mathbf{z}^{k+1} = \frac{1}{\alpha} (\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k),$$

$$\partial_u \hat{L}(\boldsymbol{\omega}^{k+1}) = -\sigma_0 (\mathbf{z}^{k+1} - \mathbf{z}^k) + \sigma_1 (\mathbf{z}^k - \mathbf{z}^{k-1}),$$

$$\partial_v \hat{L}(\boldsymbol{\omega}^{k+1}) = -\sigma_1 (\mathbf{z}^k - \mathbf{z}^{k-1}).$$

则存在: $\theta_0 = \max\{\|\mathbf{A}\| + \|\mathbf{B}\| + \|\mathbf{C}\| + \frac{1}{\alpha}, N+1, \alpha \|\mathbf{A}^T \mathbf{B}\| + N+1, \alpha \|\mathbf{A}^T \mathbf{C}\| + \alpha \|\mathbf{B}^T \mathbf{C}\| + N+1 + 2\sigma_0, 2\sigma_1\}$, 有:

$$\text{dist}(0, \partial \hat{L}(\boldsymbol{\omega}^{k+1})) \leq \theta_0 (\|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| + \|\mathbf{x}^{k+1} - \mathbf{x}^k\| + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|).$$

另一方面,由(3)式有:

$$\begin{aligned} \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| &\leq \left(\frac{3(l_h + l_\psi)^2}{\mu_0} \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2 c^2}{\mu_0} \|\mathbf{z}^{k-1} - \mathbf{z}^k\|^2 + \frac{3l_\psi^2 (1-c)^2}{\mu_0} \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|^2 \right)^{\frac{1}{2}} \leq \\ &\frac{\sqrt{3}(l_h + l_\psi)}{\mu_0} \|\mathbf{z}^{k+1} - \mathbf{z}^k\| + \frac{\sqrt{3}l_\psi c}{\mu_0} \|\mathbf{z}^{k-1} - \mathbf{z}^k\| + \frac{\sqrt{3}l_\psi (1-c)}{\mu_0} \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\| \leq \\ &\theta_1 (\|\mathbf{z}^{k+1} - \mathbf{z}^k\| + \|\mathbf{z}^{k-1} - \mathbf{z}^k\| + \|\mathbf{z}^{k-2} - \mathbf{z}^{k-1}\|). \end{aligned} \quad (6)$$

其中: $\theta_1 = \frac{\sqrt{3}(l_h + l_\psi)}{\mu_0}$, 令 $\theta = \theta_0 + \theta_1$, 则结论得证。

证毕

定理 1 若由算法生成的序列 $\boldsymbol{\chi}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)$ 有界且满足假设条件 4), 则 $\sum_{k=0}^{\infty} \|\boldsymbol{\chi}^k - \boldsymbol{\chi}^{k+1}\| < \infty$, 且序列 $\boldsymbol{\chi}^k$ 收敛到问题 1) 的稳定点。

证明 令 \mathbf{J} 为 $\boldsymbol{\omega}^k = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k, \mathbf{z}^{k-1}, \mathbf{z}^{k-2})$ 的聚点, 由引理 2 知 $\boldsymbol{\chi}^k, \boldsymbol{\chi}^{k+1}$ 有相同的聚点, 则 $\boldsymbol{\omega}^k$ 存在一个子列 $\boldsymbol{\omega}^{k_j}$ 使得 $\boldsymbol{\omega}^{k_j} \rightarrow \boldsymbol{\omega}^*$ 且 $\boldsymbol{\omega}^* = (\mathbf{x}^*, \mathbf{y}^*, \mathbf{z}^*, \boldsymbol{\lambda}^*, \mathbf{z}^*, \mathbf{z}^*) \in \mathbf{J}$. 再由 $\hat{L}(\boldsymbol{\omega}^{k_j}) \rightarrow \hat{L}(\boldsymbol{\omega}^*)$, 序列 $\hat{L}(\boldsymbol{\omega}^k)$ 是收敛序列, 有 $\hat{L}(\boldsymbol{\omega}^k) \rightarrow \hat{L}(\boldsymbol{\omega}^*)$. 对 $\hat{L}(\boldsymbol{\omega}^k)$ 考虑如下两种情况:

1) 若存在 $k_0 \in \mathbf{N}^*$ 使得 $\hat{L}(\boldsymbol{\omega}^{k_0}) = \hat{L}(\boldsymbol{\omega}^*)$ 成立. 则由引理 1 可知, 对任意的 $k > k_0$, 有:

$$\sigma_2 (\|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 + \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2) \leq \hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^{k+1}) \leq \hat{L}(\boldsymbol{\omega}^{k_0}) - \hat{L}(\boldsymbol{\omega}^*) = 0.$$

则由上式可知, 除有限项外序列 $\boldsymbol{\chi}^k$ 是常数序列, 于是 $\boldsymbol{\chi}^k$ 是收敛序列。

2) 对任意 $k \in \mathbf{N}^*$ 都有 $\hat{L}(\boldsymbol{\omega}^k) > \hat{L}(\boldsymbol{\omega}^*)$. 由假设 $f + g + h$ 是次解析函数, 则有 \hat{L} 也是次解析函数并且满足 K-L 性质. 由 \mathbf{J} 的定义知 $\lim_{k \rightarrow \infty} \text{dist}(\mathbf{J}, \boldsymbol{\omega}^k) = 0$, 再加上 $\hat{L}(\boldsymbol{\omega}^k) \rightarrow \hat{L}(\boldsymbol{\omega}^*)$, 则存在 $\beta > 0, \epsilon > 0, k_1 \in \mathbf{N}^*$ 使得对任意 $k > k_1$ 都有 $\text{dist}(\boldsymbol{\omega}^k, \mathbf{J}) < \epsilon$ 且满足 $\hat{L}(\boldsymbol{\omega}^k) < \hat{L}(\boldsymbol{\omega}^*) + \beta$.

于是由性质 1 可知, 存在 $\beta > 0, \epsilon > 0, \gamma \in A_n, A_n$ 是满足 γ 函数条件构成的一个函数集, $\vartheta \in A_n$, 且满足 $\text{dist}(\boldsymbol{\omega}^k, \mathbf{J}) < \epsilon$ 和 $\hat{L}(\boldsymbol{\omega}^*) < \hat{L}(\boldsymbol{\omega}^k) < \hat{L}(\boldsymbol{\omega}^*) + \beta$, 有 $\vartheta'(\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*)) \text{dist}(0, \partial \hat{L}(\boldsymbol{\omega}^k)) \geq 1$. 再由引理 3 得到:

$$\begin{aligned} \frac{1}{\vartheta'(\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*))} &\leq \text{dist}(0, \partial \hat{L}(\boldsymbol{\omega}^k)) \leq \\ &\theta (\|\mathbf{x}^k - \mathbf{x}^{k-1}\| + \|\mathbf{y}^k - \mathbf{y}^{k-1}\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\| + \|\mathbf{z}^{k-2} - \mathbf{z}^{k-3}\|). \end{aligned}$$

再由 ϑ 的凹性, 有:

$$\begin{aligned} \hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^{k+1}) &= (\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*)) - (\hat{L}(\boldsymbol{\omega}^{k+1}) - \hat{L}(\boldsymbol{\omega}^*)) \leq \\ &\frac{\vartheta(\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*)) - \vartheta(\hat{L}(\boldsymbol{\omega}^{k+1}) - \hat{L}(\boldsymbol{\omega}^*))}{\vartheta'(\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*))} \leq \\ &\theta (\|\mathbf{x}^{k-1} - \mathbf{x}^k\| + \|\mathbf{y}^{k-1} - \mathbf{y}^k\| + \|\mathbf{z}^k - \mathbf{z}^{k-1}\| + \\ &\quad \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\| + \|\mathbf{z}^{k-2} - \mathbf{z}^{k-3}\|) \times \\ &[\vartheta(\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*)) - \varphi(\hat{L}(\boldsymbol{\omega}^{k+1}) - \hat{L}(\boldsymbol{\omega}^*))]. \end{aligned}$$

则由引理 1 有:

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2 &\leq \\ \frac{\theta}{\sigma_2} (\|\mathbf{x}^{k-1} - \mathbf{x}^k\| + \|\mathbf{y}^{k-1} - \mathbf{y}^k\| + \|\mathbf{z}^{k-1} - \mathbf{z}^k\| + \|\mathbf{z}^{k-1} - \mathbf{z}^{k-2}\| + \|\mathbf{z}^{k-2} - \mathbf{z}^{k-3}\|) \times \\ &[\vartheta(\hat{L}(\boldsymbol{\omega}^k) - \hat{L}(\boldsymbol{\omega}^*)) - \vartheta(\hat{L}(\boldsymbol{\omega}^{k+1}) - \hat{L}(\boldsymbol{\omega}^*))]. \end{aligned}$$

从而有:

$$4(\|\mathbf{x}^{k+1} - \mathbf{x}^k\| + \|\mathbf{y}^{k+1} - \mathbf{y}^k\| + \|\mathbf{z}^{k+1} - \mathbf{z}^k\|) \leq 4\sqrt{3}(\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + \|\mathbf{y}^{k+1} - \mathbf{y}^k\|^2 + \|\mathbf{z}^{k+1} - \mathbf{z}^k\|^2)^{\frac{1}{2}} \leq$$

$$2(\|x^{k-1} - x^k\| + \|y^{k-1} - y^k\| + \|z^{k-1} - z^k\| + \|z^{k-1} - z^{k-2}\| + \|z^{k-2} - z^{k-3}\|)^{\frac{1}{2}} \times \\ \sqrt{\frac{12\theta}{\sigma_2}} [\vartheta(\hat{L}(\omega^k) - \hat{L}(\omega^*)) - \vartheta(\hat{L}(\omega^{k+1}) - \hat{L}(\omega^*))]^{\frac{1}{2}} \leq \|x^{k-1} - x^k\| + \|y^{k-1} - y^k\| + \|z^{k-1} - z^k\| + \\ \|z^{k-1} - z^{k-2}\| + \|z^{k-2} - z^{k-3}\| + \frac{12\theta}{\sigma_2} [\vartheta(\hat{L}(\omega^k) - \hat{L}(\omega^*)) - \vartheta(\hat{L}(\omega^{k+1}) - \hat{L}(\omega^*))].$$

于是有:

$$4 \sum_{i=k_1}^k (\|y^{i+1} - y^i\| + \|z^{i+1} - z^i\| + \|x^{i+1} - x^i\|) \leq \sum_{i=k_1}^k (\|z^{i-1} - z^i\| + \|z^{i-1} - z^{i-2}\| + \|z^{i-2} - z^{i-3}\| + \\ \|x^{i-1} - x^i\| + \|y^{i-1} - y^i\|) + \frac{12\theta}{\sigma_2} \sum_{i=k_1}^k [\vartheta(\hat{L}(\omega^k) - \hat{L}(\omega^*)) - \vartheta(\hat{L}(\omega^{k+1}) - \hat{L}(\omega^*))].$$

等价于:

$$\sum_{i=k_1}^k \|y^{i+1} - y^i\| + 3 \sum_{i=k_1}^k \|x^{i+1} - x^i\| + 3 \sum_{i=k_1}^k \|z^{i+1} - z^i\| \leq \\ \sum_{i=k_1}^k (\|z^{i-1} - z^i\| - \|z^{i+1} - z^i\|) + \sum_{i=k_1}^k (\|z^{i-1} - z^{i-2}\| - \|z^{i+1} - z^i\|) + \sum_{i=k_1}^k (\|z^{i-2} - z^{i-3}\| - \|z^{i+1} - z^i\|) + \\ \sum_{i=k_1}^k (\|x^{i-1} - x^i\| - \|x^{i+1} - x^i\|) + \sum_{i=k_1}^k (\|y^{i-1} - y^i\| - \|y^{i+1} - y^i\|) + \frac{12\theta}{\sigma_2} \sum_{i=k_1}^k [\vartheta(\hat{L}(\omega^i) - \hat{L}(\omega^*)) - \\ \vartheta(\hat{L}(\omega^{i+1}) - \hat{L}(\omega^*))] \leq 3\|z^{k_1-1} - z^{k_1}\| + 2\|z^{k_1-1} - z^{k_1-2}\| + \|z^{k_1-2} - z^{k_1-3}\| + \|x^{k_1-1} - x^{k_1}\| + \\ \|y^{k_1-1} - y^{k_1}\| + \frac{12\theta}{\sigma_2} [\vartheta(\hat{L}(\omega^{k_1}) - \hat{L}(\omega^*)) - \vartheta(\hat{L}(\omega^{k_1+1}) - \hat{L}(\omega^*))] \leq 3\|z^{k_1-1} - z^{k_1}\| + \\ 2\|z^{k_1-1} - z^{k_1-2}\| + \|z^{k_1-2} - z^{k_1-3}\| + \|x^{k_1-1} - x^{k_1}\| + \|y^{k_1-1} - y^{k_1}\| + \frac{12\theta}{\sigma_2} [\vartheta(\hat{L}(\omega^{k_1}) - \hat{L}(\omega^*))].$$

于是有:

$$\sum_{k=0}^{\infty} (\|y^{k+1} - y^k\| + \|z^{k+1} - z^k\| + \|x^{k+1} - x^k\|) < \infty,$$

再由(6)式可以推导出 $\sum_{k=0}^{\infty} \|\lambda^{k+1} - \lambda^k\| < \infty$, 从而 $\sum_{k=0}^{\infty} \|\chi^{k+1} - \chi^k\| < \infty$. 因此序列 χ^k 是收敛的。 证毕

3 数值实验

针对问题(1)考虑矩阵分解应用^[10]中的 3 块非凸优化问题,基本模型如下:

$$\min_{X,Y,Z} \|X\|_* + \lambda \|Y\|_1 + \frac{\mu}{2} \|Z - M\|_F^2, \\ \text{s. t. } Z = X + Y. \tag{7}$$

其中: M 表示 $m \times n$ 维观测矩阵, $\|X\|_* = \sum_{i=1}^{\min(m,n)} |\sigma_i(X)|^{\frac{1}{2}}$, $\|Y\|_1 = \sum_{i=1}^m \sum_{j=1}^n |Y_{ij}|$, λ 是低秩项 $\|X\|_*$ 和稀疏项 $\|Y\|_1$ 之间的交换参数, μ 是与噪声水平相关的惩罚参数。

问题(7)的增广拉格朗日函数可表示为:

$$L_\alpha(X, Y, Z, \Lambda) = \|X\|_* + \lambda \|Y\|_1 + \frac{\mu}{2} \|Z - M\|_F^2 + \langle \Lambda, Z - X - Y \rangle + \frac{\alpha}{2} \|Z - X - Y\|_F^2,$$

利用 I-B-ADMM, 函数 φ, τ, ψ 分别为 $\varphi(X) = \|X\|_F^2, \tau(Y) = \|Y\|_F^2, \psi(Z) = \|Z\|_F^2$. 则问题(7)可以表示为:

$$\begin{cases} X^{k+1} = \arg \min_X \|X\|_* - \langle \Lambda^k, X \rangle + \frac{\alpha}{2} \|Z^k - X - Y^k\|_F^2 + \frac{\rho}{2} \|X - \hat{X}^k\|_F^2 \\ Y^{k+1} = \arg \min_Y \lambda \|Y\|_1 - \langle \Lambda^k, Y \rangle + \frac{\alpha}{2} \|Z^k - X^{k+1} - Y\|_F^2 + \frac{\rho}{2} \|Y - \hat{Y}^k\|_F^2 \\ Z^{k+1} = \arg \min_Z \frac{\mu}{2} \|Z - M\|_F^2 + \langle \Lambda^k, Z \rangle + \frac{\alpha}{2} \|Z - X^{k+1} - Y^{k+1}\|_F^2 + \frac{\rho}{2} \|Z - \hat{Z}^k\|_F^2 \\ \Lambda^{k+1} = \Lambda^k + \alpha(Z^{k+1} - X^{k+1} - Y^{k+1}) \end{cases}.$$

其中: $\hat{\mathbf{X}}^k = \mathbf{X}^{k-1} + a(\mathbf{X}^k - \mathbf{X}^{k-1})$, $\hat{\mathbf{Y}}^k = \mathbf{Y}^{k-1} + b(\mathbf{Y}^k - \mathbf{Y}^{k-1})$, $\hat{\mathbf{Z}}^k = \mathbf{Z}^{k-1} + c(\mathbf{Z}^k - \mathbf{Z}^{k-1})$ 。

于是上述子问题可以处理成如下形式:

$$\begin{cases} \mathbf{X}^{k+1} = H \left(\frac{\alpha \left(\mathbf{Z}^k - \mathbf{Y}^{k+1} + \frac{\mathbf{\Lambda}^k}{\alpha} \right) + \rho \hat{\mathbf{X}}^k}{\alpha + \rho}, \frac{1}{\alpha + \rho} \right) \\ \mathbf{Y}^{k+1} = S \left(\frac{\alpha \left(\mathbf{Z}^k - \mathbf{X}^{k+1} + \frac{\mathbf{\Lambda}^k}{\alpha} \right) + \rho \hat{\mathbf{Y}}^k}{\alpha + \rho}, \frac{\lambda}{\alpha + \rho} \right) \\ \mathbf{Z}^{k+1} = \frac{\mu \mathbf{M} + \alpha \left(\mathbf{X}^{k+1} + \mathbf{Y}^{k+1} - \frac{\mathbf{\Lambda}^k}{\alpha} \right) + \rho \hat{\mathbf{Z}}^k}{\mu + \alpha + \rho} \\ \mathbf{\Lambda}^{k+1} = \mathbf{\Lambda}^k + \alpha (\mathbf{Z}^{k+1} - \mathbf{X}^{k+1} - \mathbf{Y}^{k+1}) \end{cases}。$$

其中: $H(\mathbf{D}, n)$ 是半 shrinkage 算子^[16], 且 $H(\mathbf{D}, n) = \{h_n(d_{i1}), h_n(d_{i2}), \dots, h_n(d_{in})\}^T, i = 1, 2, \dots, m; h_n(d_{ij}) = \begin{cases} \frac{2d_{ij}}{3} \left(1 + \cos \frac{2}{3}(\pi - \varphi(|d_{ij}|)) \right), & |d_{ij}| > \frac{\sqrt[3]{54}}{4} n^{\frac{2}{3}} \\ 0, & d_{ij} = 0 \end{cases}$, 其中 $\varphi(d_{ij}) = \arccos \left(\frac{n}{8} \left(\frac{|d_{ij}|}{3} \right)^{-\frac{3}{2}} \right)$; $S(\mathbf{E}, u)$ 是软

shrinkage 算子, $S(\mathbf{E}, u) = \{s_u(e_{i1}), s_u(e_{i2}), \dots, s_u(e_{iu})\}^T, i = 1, 2, \dots, v$, 其中 $s_u(e_{ij}) = \begin{cases} e_{ij} - u, & e_{ij} > u \\ 0, & |e_{ij}| < u \\ e_{ij} + u, & e_{ij} < -u \end{cases}$ 。

本文数值实验运行环境为英特尔 Core i5-7200U 2.71 GHz 处理器、4.00 GB 内存和 64 位 Windows 10 操作系统, 运行 Matlab R2018a 验证所提出的算法的可行性。针对问题(5), 定义 $\alpha = 0.03, \rho = 0.03$, 对应的初始矩阵 $\mathbf{X}^0, \mathbf{Y}^0, \mathbf{Z}^0$ 都是维数为 100 的零矩阵。观测矩阵 \mathbf{M} 是维数为 100 的随机矩阵。停止条件为 $\text{RelChg}: \frac{\|(\mathbf{X}^{k+1}, \mathbf{Y}^{k+1}, \mathbf{Z}^{k+1}) - (\mathbf{X}^k, \mathbf{Y}^k, \mathbf{Z}^k)\|_F}{\|(\mathbf{X}^k, \mathbf{Y}^k, \mathbf{Z}^k)\|_F + 1} \leq 10^{-8}$ 。

首先, 分析参数 μ, λ, σ 对 I-B-ADMM 的影响, 固定惯性系数 a, b, c 都为 0.25, 重复实验 100 次(且 100 次实验均是在同一操作环境中进行), 并对主要实验结果进行了记录。表 1 中呈现的数据结果分别为 100 次实验中最好、最差和平均的结果。其中 $\sigma = 0$ 表示无噪音情况, $\sigma = 0.01$ 表示 Gaussian 噪音情况, $k_{\min}, k_{\max}, \bar{k}$ 分别表示 100 次实验中最少的迭代步数、最多的迭代步数及平均迭代步数。

表 1 固定惯性系数试验结果比较

Tab. 1 Comparison of fixed inertia coefficient test results

μ	λ	σ	k_{\min}	k_{\max}	\bar{k}	μ	λ	σ	k_{\min}	k_{\max}	\bar{k}
10 000	0.01	0	86	114	91	100	0.01	0	87	132	92
10 000	0.01	0.01	72	97	79	100	0.01	0.01	80	107	88
10 000	0.1	0	91	100	94	100	0.1	0	94	137	98
10 000	0.1	0.01	77	127	86	100	0.1	0.01	82	117	91

注: 加黑的数据表示该组结果最优

从表 1 中可以发现, 当 μ 和 σ 相同时, λ 越小, I-B-ADMM 需要迭代的步数越少。当 λ 和 σ 相同时, μ 越大, I-B-ADMM 需要的迭代步数越少。当 λ 和 μ 相同时, σ 越大时, I-B-ADMM 需要的迭代步数越少。通过上述分析, 为使 I-B-ADMM 需要的迭代步数更少, 应在数值实验时选择较大的 μ, σ 和较小的 λ 。

基于参数 μ, λ, σ 较好的情况下, 取 $\mu = 10\ 000, \lambda = 0.01, \sigma = 0.01$, 对不同的惯性系数 a, b, c , 比较了 I-B-ADMM 与 B-ADMM。图 1 展示了在 Gaussian 噪音情况下, 考虑惯性系数 a, b, c 相同的情形时, 并分别取 0.25, 0.5, 0.75 和 1 时关于 RelChg 误差与迭代步数的关系。当惯性系数 a, b, c 分别取 0.25, 0.5, 0.75 时, 表示

的是 I-B-ADMM; 当惯性系数 a, b, c 取 1 时, I-B-ADMM 退化到文献[11]中的 B-ADMM。

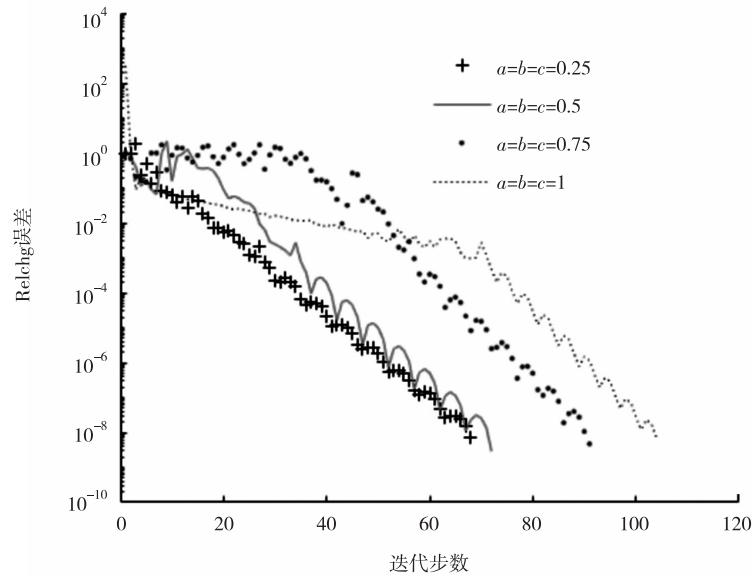


图 1 RelChg 误差与迭代步数

Fig. 1 RelChg error and iteration steps

从图 1 中可以发现随着惯性系数 a, b, c 的增大而需要的迭代步数增大, 而惯性系数越接近 1 它需要的迭代步数最多, 数值实验结果表明 I-B-ADMM 优于文献[11]中的 B-ADMM。

4 结论

针对可分结构的非凸非光滑优化问题, 本文提出了一种新的内置惯性 Bregman 交替方向乘子法。在算法生成的点列满足有界性的条件下, 利用 K-L 性质, 证明了算法的渐进收敛性。本文数值实验结果说明该算法在 Bregman 度量中内置了惯性项, 通过调整惯性系数 a, b, c 的取值能提高收敛速率, 且效果比文献[11]中的 B-ADMM 好。不足之处在于证明收敛性分析时运用了 K-L 性质, 但这是一个比较繁琐的工具, 是否可以摆脱 K-L 性质来证明收敛性分析值得有待研究。

参考文献:

- [1] GABAY D, MERCIER B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation [J]. Computers & Mathematics with Applications, 1976, 2(1): 17-40.
- [2] DHAR S, YI C R, RAMAKRISHNAN N, et al. ADMM based scalable machine learning on Spark[C]//2015 IEEE International Conference on Big Data. Santa Clara: IEEE, 2015: 1174-1182. doi: 10.1109/BigData.2015.7363871.
- [3] HUEBNER N, RINK Y, SURIYAH M, et al. Distributed AC-DC optimal power flow in the European transmission grid with ADMM[EB/OL]. (2019-12-19)[2021-10-21]. <https://arxiv.org/pdf/1912.03942.pdf>.
- [4] SCHIZAS I D, RIBEIRO A, GIANNAKIS G B. Consensus in *Ad Hoc* WSNs with noisy links: part I: distributed estimation of deterministic signals[J]. IEEE Transactions on Signal Processing, 2008, 56(1): 350-364.
- [5] YANG Q, CHEN G, WANG T. ADMM-based distributed algorithm for economic dispatch in power systems with both packet drops and communication delays[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(3): 842-852.
- [6] 何炳生. 凸优化和单调变分不等式收缩算法的统一框架[J]. 中国科学, 2018, 48(2): 255-272.
HE B S. A uniform framework of contraction methods for convex optimization and monotone variational inequality[J]. Science Sinica (Mathematica), 2018, 48(2): 255-272.
- [7] YE Y Y, CHEN C H, HE B S, et al. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent[J]. Mathematical Programming, 2016, 155: 57-79.
- [8] YANG L, PONG T K, CHEN X J. Alternating Direction Method of Multipliers for Nonconvex Background/Foreground Extraction[EB/OL]. (2016-9-29)[2021-10-21]. <https://doi.org/10.48550/arXiv.1506.07029>.

- [9] CAI X J, HAN D R, YUAN X M. On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function[J]. Computational Optimization Applications, 2017, 66: 39-73.
- [10] WANG F H, CAO W F, XU Z B. Convergence of multi-block Bregman ADMM for nonconvex composite problems[J]. Science China (Information Sciences), 2018, 61(12): 153-64.
- [11] XU J W, CHAO M T. An inertial Bregman generalized alternating direction method of multipliers for nonconvex optimization [J]. Journal of Applied Mathematics and Computing, 2022, 68: 1-27.
- [12] KURDYKA K. On gradients of functions definable in o-minimal structures[J]. Annales-Institut Fourier, 1998, 48(3): 769-783.
- [13] BOLTE J, DANILIDIS A, LEY O, et al. Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity[J]. Transactions of the American Mathematical Society, 2010, 362(6): 3319-3363.
- [14] XU Y Y, YIN W T. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion[J]. SIAM Journal on Imaging Sciences, 2015, 6(3): 1758-1789.
- [15] BREGMAN L M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming[J]. USSR Computational Mathematics and Mathematical Physics, 1967, 73(3): 200-217.
- [16] XU Z B, CHANG X Y, XU F M, et al. L-1/2 regularization; a thresholding representation theory and a fast solver[J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(7): 1013-1027.

Operations Research and Cybernetics

Convergence of Build-in Inertial Bregman ADMM for Nonsmooth Nonconvex Optimization

WEI Jiao, LUO Honglin

(School of Mathematical Sciences, Chongqing Normal University, Chongqing 401331, China)

Abstract: [Purposes] A class of linearly constrained nonconvex nonsmooth optimization problems with separable structures is considered. a new Bregman ADMM with build-in inertial force is proposed for solving such nonconvex nonsmooth optimization problems. [Methods] In order to accelerate the convergence speed of the algorithm, in the framework of the Bregman Multiplier lalternating direction method, the built-in inertia term of Bregman metric in the subproblem. [Findings] It proves that the iterative sequence generated by the algorithm converges asymptotically to a critical point of the problem under some mild assumptions. [Conclusions] Some preliminary numerical results on solving a nonconvex matrix decomposition problem with a low-rank term and a sparse term to illustrate the efficiency of the proposed method.

Keywords: separable nonconvex nonsmooth optimization problems; build-in inertial Bregman ADMM; Kurdyka-Lojasiewicz property; asymptotic convergence

(责任编辑 黄颖)