

求解非凸优化问题的带惯性项Majorized Bregman交替方向乘子法

吴展雄*, 陆 莎#, 黄清梅

南宁师范大学数学与统计学院, 广西 南宁

收稿日期: 2025年5月17日; 录用日期: 2025年6月11日; 发布日期: 2025年6月18日

摘 要

对非凸两分块优化问题, 提出一种带惯性的Majorized Bregman交替方向乘子法。该算法在迭代中结合了目标函数的极大化线性技术和Bregman距离来简化子问题的求解, 同时通过引入惯性项加快收敛效果。在适当条件下证明了算法的收敛性质。初步数值实验结果表明该算法是有效的。

关键词

交替方向乘子法, Bregman距离, 惯性项, KL性质

Majorized Bregman Alternating Direction Method of Multipliers with Inertia Terms for Solving Nonconvex Optimisation Problems

Zhanxiong Wu*, Sha Lu#, Qingmei Huang

School of Mathematics and Statistics, Nanning Normal University, Nanning Guangxi

Received: May 17th, 2025; accepted: Jun. 11th, 2025; published: Jun. 18th, 2025

Abstract

A Majorized Bregman alternating direction method of multipliers (Bregman-ADMM) with inertial terms is proposed for nonconvex two-block optimization problems. The algorithm combines a linearization technique for the objective function and the Bregman distance in each iteration to simplify

*第一作者。

#通讯作者。

subproblem solutions, while accelerating the convergence rate through inertial terms. The convergence of the algorithm is established under appropriate conditions. Preliminary numerical experiments demonstrate the effectiveness of the proposed algorithm.

Keywords

Alternating Direction Method of Multipliers, Bregman Distance, Inertial Term, Kurdyka-Łojasiewicz Property

Copyright © 2025 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

本文考虑如下带线性等式约束的两块可分结构非凸优化问题：

$$\begin{aligned} \min & f(x) + g(y) \\ \text{s.t.} & Ax - By = 0, \end{aligned} \quad (1)$$

其中 $f: R^{s_1} \rightarrow R$ 是梯度 Lipschitz 连续可微的凸函数, $g: R^{s_2} \rightarrow R \cup \{+\infty\}$ 是适当下半连续函数, $A \in R^{m \times s_1}, B \in R^{m \times s_2}$ 。问题(1)广泛应用于信号处理[1]、图像处理[2]、机器学习[3]和经济调度[4]等领域,其中 f 通常是一个(二次或逻辑)损失函数, g 通常是一个正则项,如 l_1 范数或 $l_{1/2}$ 拟范数。问题(1)的增广拉格朗日函数定义为

$$L_\beta(x, y, \lambda) = f(x) + g(y) + \langle \lambda, Ax - By \rangle + \frac{\beta}{2} \|Ax - By\|^2,$$

其中 $\lambda \in R^m$ 是拉格朗日乘子, $\beta > 0$ 是一个惩罚参数。

求解问题(1)的经典 ADMM 的第 $k+1$ 步迭代格式为

$$\begin{cases} y^{k+1} = \arg \min_y L_\beta(x^k, y, \lambda^k) \\ x^{k+1} = \arg \min_x L_\beta(x, y^{k+1}, \lambda^k) \\ \lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} - By^{k+1}). \end{cases} \quad (2)$$

尽管 ADMM 在解决凸优化问题方面得到了广泛的应用,但对于非凸优化问题,ADMM 的收敛性分析仍值得研究,目前一些学者已经得到了部分研究成果。如 Deng Zhao 等[5]针对非凸两分块优化问题,在增广拉格朗日函数满足 KL 性质且罚参数大于某阈值的条件下,证明了 ADMM 的收敛性。在此基础上, Wang Y 等[6]进一步拓展了 ADMM 的理论框架,他们放松了文献[5]中对矩阵列满秩的严格限制,并在目标函数和系数矩阵满足一些常见约束条件下,证明了 ADMM 的全局收敛性。

一般将 ADMM 直接应用于非凸优化问题时,其收敛性不能保证,由此学者们对原始 ADMM 方法做了改进研究,如线性化技术、松弛技巧等。为了有利于迭代子问题的求解, Wang F 等[7]针对非凸两分块优化问题(1)引入 Bregman 距离改善子问题目标函数的结构,提出 Bregman ADMM (BADMM)算法。其过程如下

$$\begin{cases} y^{k+1} = \arg \min_y \{L_\beta(x^k, y, \lambda^k) + D_{T_1}(y, y^k)\} \\ x^{k+1} = \arg \min_x \{L_\beta(x, y^{k+1}, \lambda^k) + D_{T_2}(x, x^k)\} \\ \lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} - By^{k+1}) \end{cases}$$

其中 $T_i (i=1,2)$ 是凸可微函数, 其 Bregman 距离 D_{T_i} 定义为:

$$D_{T_i}(x, y) = T_i(x) - T_i(y) - \langle \nabla T_i(y), x - y \rangle, \quad \forall x \in \text{dom} T_i, y \in \text{int dom} T_i (i=1,2).$$

比如当 $T_i(x) = x^T C x$, 其中 C 是半正定矩阵时有 $D_{T_i}(x_1, x_2) = (x_1 - x_2)^T C (x_1 - x_2)$ 。

在此基础上, 陈建华等[8]的 Majorized Bregman ADMM(MBADMM)算法针对非凸两分块优化问题, 对目标函数中的光滑函数项进行了极大化线性处理, 由于目标函数中的 f 是光滑凸函数, 这意味着存在一个半正定矩阵 Σ_f 使得对任意 $x, x' \in R^{s_1}$, 都有: $f(x) \geq f(x') + \langle \nabla f(x'), x - x' \rangle + \frac{1}{2} \|x - x'\|_{\Sigma_f}^2$, 又由于函数 f 的梯度是 Lipschitz 连续的, 所以可以找到一个半正定矩阵 $\hat{\Sigma}_f$, 且 $\hat{\Sigma}_f \succeq \Sigma_f$, 使得对任意的 $x, x' \in R^{s_1}$ 都有

$$f(x) \leq \hat{f}(x, x') := f(x') + \langle \nabla f(x'), x - x' \rangle + \frac{1}{2} \|x - x'\|_{\hat{\Sigma}_f}^2.$$

于是针对问题(1), 文献[8]构建 Majorized 增广拉格朗日函数为

$$\hat{L}_\beta(x, y, \lambda, x') = \hat{f}(x, x') + g(y) + \lambda^T (Ax - By) + \frac{\beta}{2} \|Ax - By\|^2. \quad (3)$$

相应的 MBADMM 算法迭代过程如下:

$$\begin{cases} y^{k+1} = \arg \min_y \{\hat{L}_\beta(x^k, y, \lambda^k, x^k) + D_{T_1}(y, y^k)\} \\ x^{k+1} = \arg \min_x \{\hat{L}_\beta(x, y^{k+1}, \lambda^k, x^k) + D_{T_2}(x, x^k)\} \\ \lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} - By^{k+1}) \end{cases} \quad (4)$$

另一方面, 注意到在迭代算法中, 增加惯性力或惯性项可以在产生新方向时充分利用部分旧方向的信息, 从而对算法起到加速的作用[9]-[11], Xu J 等[12]提出带惯性项的 Bregman Generalized ADMM (IB-GADMM)算法。针对线性约束 $Ax + By = b$ 非凸优化问题, [12]结合惯性步和 Bregman 距离思想给出迭代过程

$$\begin{cases} x^{k+1} \in \arg \min_x \{L_\beta(x, y^k, \lambda^k) + D_{T_2}(x, x^k) + \theta_k \langle x, x^{k-1} - x^k \rangle\} \\ x_{ad}^{k+1} = \alpha Ax^{k+1} + (1 - \alpha)(b - By^k) \\ y^{k+1} \in \arg \min_y \{g(y) - \langle \lambda^k, By^k \rangle + \frac{\beta}{2} \|x_{ad}^{k+1} + By - b\|^2 + D_{T_1}(y, y^k) + \rho_k \langle y, y^{k-1} - y^k \rangle\} \\ \lambda^{k+1} = \lambda^k - \beta(x_{ad}^{k+1} + By^{k+1} - b), \end{cases}$$

并建立了算法的子序列收敛性和全局收敛性。

受上述文献的启发, 对目标函数中的光滑函数项进行极大化线性处理, 可以简化关于该函数子问题的求解, 针对目标函数加入 Bregman 距离可以提升算法在求解非凸优化问题时的性能和收敛性, 引入惯性项也可以缩短收敛所需时间。本文提出了针对非凸优化问题(1)的带惯性项 Majorized-Bregman-ADMM (IMBADMM)算法, 该算法涵盖了 ADMM 和 MBADMM 作为其特殊情形。与文献[13]有所区别的是, 本

文在两个子问题中引入了惯性力[9]-[11]。此外, 效益函数的充分递减特性并不需要假设 A 或 B 是满秩。在 IMBADMM 生成的序列具有有界性且效益函数满足 KL 性质的条件下, 我们证明了 IMBADMM 的子序列收敛性和全局收敛性。最后运用所提出的 IMBADMM 算法对稀疏逻辑回归问题进行了数值实验比较。

2. 预备知识

2.1. 基本概念与结论

对任意 $x, y \in \mathbb{R}^n$, 定义它们的内积为 $\langle x, y \rangle = x^T y$, 定义 x 的 G 范数 $\|x\|_G = \sqrt{x^T G x}$, 其中 $G \succeq 0$ (> 0) 表明 G 是半正定(正定)矩阵。 $\lambda_{\max}(G)$ 表示矩阵 G 的最大特征值。

任意点 y 到集合 S 的距离定义如下

$$\text{dist}(y, S) = \begin{cases} \inf_{x \in S} \|x - y\|, & S \neq \emptyset \\ +\infty, & S = \emptyset. \end{cases}$$

定义 1 ([14], 引理 2.2) 称可微函数 $f(x)$ 是 μ -强凸函数, 若如下不等式成立:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2, \quad \forall x, y \in \text{dom} f.$$

关于 Bregman 距离, 有下述性质。

性质 1 ([15], 命题 2.3) 令 $T: \mathbb{R}^n \rightarrow \mathbb{R}$ 是可微凸函数, $D_T(x, y)$ 是与它相关的 Bregman 距离, 则

- 1) 对任意的 $x, y \in \mathbb{R}^n$, 有: $D_T(x, y) \geq 0$, $D_T(x, x) = 0$;
- 2) 对于固定的 y , $D_T(\cdot, y)$ 是凸的;
- 3) 对于任意 $x, y \in \mathbb{R}^n$, $D_T(x, y) \geq \frac{\mu}{2} \|x - y\|^2$, 其中 μ 是 T 的强凸系数;
- 4) 如果 ∇T 是 Lipschitz 连续的, 且其 Lipschitz 常数为 l_T , 则对任意 $x, y \in \mathbb{R}^n$, 有

$$\nabla T(x, y) \leq \frac{l_T}{2} \|x - y\|^2.$$

定义 2 ([14], 定义 8.5) 定义满足以下条件的凹连续函数 $\varphi: [0, \eta) \rightarrow \mathbb{R}_+$ 的集合为 Φ_η 函数类:

- 1) $\varphi(0) = 0$;
- 2) φ 在 $(0, \eta)$ 内连续可微, 在点 0 处连续;
- 3) 对任意 $s \in (0, \eta)$, 都有 $\varphi'(s) > 0$ 。

根据上面的定义我们可以引入 KL 性质。

定义 3 ([14], 定义 8.6) (KL 性质) 设 $g: \mathbb{R}^{s_2} \rightarrow \mathbb{R} \cup \{+\infty\}$ 为适当下半连续函数。

1) 称函数 g 在给定点 $\bar{y} \in \text{dom} \partial g \stackrel{\text{def}}{=} \{y | \partial g(y) \neq \emptyset\}$ 处具有 KL 性质, 若存在 $\eta \in (0, +\infty]$, $\bar{y}$ 的一个邻域 U 及函数 $\varphi \in \Phi_\eta$, 使得对

$$\forall y \in U \cap [g(\bar{y}) < g < g(\bar{y}) + \eta],$$

有以下不等式成立:

$$\varphi'(g(y) - g(\bar{y})) \text{dist}(0, \partial g(y)) \geq 1.$$

2) 若 g 在 $\text{dom} \partial g$ 上处处都满足 KL 性质, 则称 g 是一个 KL 函数。

性质 2 ([14], 引理 8.4) (一致 KL 性质) 设 Ω 为紧集, $g: \mathbb{R}^{s_2} \rightarrow \mathbb{R} \cup \{+\infty\}$ 为适当下半连续函数, 在 Ω 上为常数且在 Ω 的每个点处都满足 KL 性质, 则存在 $\varepsilon > 0, \eta > 0, \varphi \in \Phi_\eta$, 使得对任意 $\bar{y} \in \Omega$ 和所有如下集

合中的 y :

$$y \in \{y \in R^{S_2} : \text{dist}(y, \Omega) < \varepsilon\} \cap [g(\bar{y}) < g < g(\bar{y}) + \eta],$$

有

$$\varphi'(g(y) - g(\bar{y})) \text{dist}(0, \partial g(y)) \geq 1.$$

性质 3 ([16], 引理 A.2) 设 $A \in R^{m \times s_1}$ 是一个非零矩阵, μ_{A^T} 表示 $A^T A$ 的最小正特征值, P_A 是到 $\text{Im}(A) = \{y | y = Ax, x \in R^{s_1}\}$ 的欧几里得投影, 则有

$$\|P_A(u)\| \leq \frac{1}{\sqrt{\mu_{A^T}}} \|A^T u\|, \forall u \in R^{s_1}.$$

2.2. 主要结论与论文结构

本文主要贡献集中在第 3 部分和第 4 部分, 分别给出新的算法及其收敛性证明和数值实验。

第 3 部分提出求解非凸优化问题的带惯性项 Majorized Bregman ADMM (IMBADMM) 算法, 针对两个子问题, 在每一步迭代中引入惯性力, 在保证算法收敛性的前提下, 有效地提高了算法的运算速度。在较弱的条件下, 证明了算法的全局收敛性。

第 4 部分我们将 IMBADMM 算法应用到稀疏逻辑回归问题中, 并分别与 MBADMM、ADMM、SGD 和 LBADMM 算法进行比较, 数值实验都表明了算法的有效性。

最后, 在第 5 部分给出本文总结。

3. 求解非凸优化问题的带惯性 Majorized Bregman ADMM 算法

对优化问题(1), 我们先给出带惯性项的 Majorized Bregman ADMM 算法, 再讨论该算法的收敛性。

3.1. IMBADMM 算法描述

算法 1: 带惯性项的 Majorized Bregman ADMM(IMBADMM)算法

输入: $A \in R^{m \times s_1}, B \in R^{m \times s_2}, \lambda > 0, \rho_0 > 0, \theta_0 > 0, \rho > 0, \theta > 0, \text{eps} > 0,$

$\text{maxiter} > 0$, 选取合适的凸可微函数 T_1, T_2 用于构造 Bregman 距离;

初始化: $x^{-1} = x^0 = \text{zeros}(n, 1), y^{-1} = y^0 = \text{zeros}(n, 1), \lambda^0 = \text{zeros}(n, 1)$; 令 $k := 0$ 。

步 1: 更新参数, 选取适当 ρ_k 和 θ_k 满足 $0 \leq \rho_k \leq \rho, 0 \leq \theta_k \leq \theta$;

步 2: 计算

$$y^{k+1} \in \arg \min_y \left\{ \hat{L}_\beta(x^k, y, \lambda^k, x^k) + D_{T_1}(y, y^k) + \rho_k \langle y, y^{k-1} - y^k \rangle \right\}; \quad (5a)$$

步 3: 固定其他, 计算

$$x^{k+1} = \arg \min_x \left\{ \hat{L}_\beta(x, y^{k+1}, \lambda^k, x^k) + D_{T_2}(x, x^k) + \theta_k \langle x, x^{k-1} - x^k \rangle \right\}; \quad (5b)$$

步 4: 更新乘子

$$\lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} - By^{k+1}); \quad (5c)$$

步 5: 若 $\frac{\|x^{k+1} - x^k\|}{\max\{\|x^k\|, 1\}} \leq \text{eps}$, 则停止迭代, 令 $y = y^{k+1}, x = x^{k+1}, \lambda = \lambda^{k+1}$, 输出 y, x ,

λ ; 否则令 $k := k + 1$, 转步 1。

其中 T_1, T_2 是两个可微凸函数, $D_{T_1}(y, y^k), D_{T_2}(x, x^k)$ 是 Bregman 距离, $\rho_k \langle y, y^{k-1} - y^k \rangle, \theta_k \langle x, x^{k-1} - x^k \rangle$ 为惯性项。

3.2. 收敛性分析

在分析收敛性之前, 我们需要如下假设。

假设 1 1) $A \neq 0, \text{Im } B \subset \text{Im } A$;

2) $f: R^{s_1} \rightarrow R$ 是连续可微凸函数, ∇f 是 l_f -Lipschitz 连续的, 且 $\|\nabla^2 f(x)\|$ 有界,

$g: R^{s_2} \rightarrow R \cup \{+\infty\}$ 是适当下半连续函数;

3) 构造 Bregman 距离函数 $T_i (i=1, 2)$ 是 μ_i -强凸的, 且 ∇T_i 是 l_i -Lipschitz 连续的, $\mu_1 - 2\rho > 0$ 且 $\mu_2 > 6 \left(\frac{(l_f + l_{T_2} + \lambda_{\max} + \theta)^2 + \theta^2 + (l_{T_2} + \lambda_{\max})^2}{\beta \mu_{A^T}} + \frac{\theta}{3} \right) + \lambda_{\max} + 2l_f$, 其中 μ_{A^T} 是矩阵 $A^T A$ 的最小严格正特征值,

$\lambda_{\max}$ 是矩阵 $\hat{\Sigma}_f$ 的最大特征值。

为了简化表达, 记 $z^k = (x^k, y^k, \lambda^k)$, $\hat{L}_\beta(\cdot)$ 的临界点 $z^* = (x^*, y^*, \lambda^*)$,

$$\Delta x^k = x^k - x^{k-1}, \Delta y^k = y^k - y^{k-1}, \Delta \lambda^k = \lambda^k - \lambda^{k-1},$$

$$\hat{z}^k = (x^k, y^k, \lambda^k, x^{k-1}, y^{k-1}, \lambda^{k-1}, x^{k-2}).$$

引理 1 若假设 1 成立, 令 $\{z^k\}$ 是由 IMBADMM 算法生成的序列, 则对于所有 $k \in N$, 有

$$\|\Delta \lambda^{k+1}\| \leq \frac{1}{\sqrt{\mu_{A^T}}} \left((l_f + l_{T_2} + \lambda_{\max} + \theta_k) \|\Delta x^k\| + (l_{T_2} + \lambda_{\max}) \|\Delta x^{k+1}\| + \theta_{k-1} \|\Delta x^{k-1}\| \right). \quad (6)$$

对 IMBADMM 算法的收敛性分析, 构造效益函数

$$\hat{L}(x, y, \lambda, x', \hat{y}, \hat{x}, \bar{x}) = \hat{L}_\beta(x, y, \lambda, x') + \frac{\rho}{2} \|y - \hat{y}\|^2 + n_1 \|x - \hat{x}\|^2 + n_2 \|\hat{x} - \bar{x}\|^2,$$

$$\hat{L}(\hat{z}^k) = \hat{L}_\beta(x^k, y^k, \lambda^k, x^{k-1}) + \frac{\rho}{2} \|\Delta y^k\|^2 + n_1 \|\Delta x^k\|^2 + n_2 \|\Delta x^{k-1}\|^2, \quad (7)$$

$$\text{其中 } n_1 = \frac{3(l_f + l_{T_2} + \lambda_{\max} + \theta)^2}{\beta \mu_{A^T}} + \frac{\theta + \lambda_{\max} + 2l_f}{2} + \frac{3\theta^2}{\beta \mu_{A^T}}, n_2 = \frac{3\theta^2}{\beta \mu_{A^T}}.$$

以下所述的充分下降性质将在我们的收敛性分析中占据关键地位, 为了分析 $\{\hat{L}(\hat{z}^k)\}$ 的单调性, 我们记

$$\begin{cases} \kappa_1 := \kappa_1(\mu_1, \rho) = \frac{\mu_1 - 2\rho}{2} \end{cases} \quad (8a)$$

$$\begin{cases} \kappa_2 := \kappa_2(\mu_2, \beta, \theta) = \frac{\mu_2 - \lambda_{\max} - 2l_f}{2} - 3 \left(\frac{(l_f + l_{T_2} + \lambda_{\max} + \theta)^2 + \theta^2 + (l_{T_2} + \lambda_{\max})^2}{\beta \mu_{A^T}} + \frac{\theta}{3} \right). \end{cases} \quad (8b)$$

证明: 因为 $\lambda^{k+1} - \lambda^k = \beta(Ax^{k+1} - By^{k+1}) \in \text{Im } A$, 所以 $\Delta \lambda^{k+1} = P_A(\Delta \lambda^{k+1})$, 由性质 3 得 $\|P_A(\Delta \lambda^{k+1})\| \leq \frac{1}{\sqrt{\mu_{A^T}}} \|A^T \Delta \lambda^{k+1}\|$ 。

根据(5b)式的最优性条件得

$$0 = \nabla f(x^k) + \hat{\Sigma}_f(x^{k+1} - x^k) + A^T \lambda^{k+1} + \nabla T_2(x^{k+1}) - \nabla T_2(x^k) + \theta_k(x^{k-1} - x^k), \quad (9a)$$

(9a)式移项可得

$$A^T \lambda^{k+1} = -\nabla f(x^{k+1}) - \hat{\Sigma}_f(x^{k+1} - x^k) - \nabla T_2(x^{k+1}) + \nabla T_2(x^k) - \theta_k(x^{k-1} - x^k) \quad (9b)$$

令 $k = k-1$, 有

$$A^T \lambda^k = -\nabla f(x^k) - \hat{\Sigma}_f(x^k - x^{k-1}) - \nabla T_2(x^k) + \nabla T_2(x^{k-1}) - \theta_k(x^{k-2} - x^{k-1}) \quad (9c)$$

(9b)和(9c)式相减可得

$$\begin{aligned} A^T(\lambda^{k+1} - \lambda^k) &= -(\nabla f(x^k) - \nabla f(x^{k-1})) - \hat{\Sigma}_f[(x^{k+1} - x^k) - (x^k - x^{k-1})] \\ &\quad - [\nabla T_2(x^{k+1}) - \nabla T_2(x^k)] + [\nabla T_2(x^k) - \nabla T_2(x^{k-1})] \\ &\quad - \theta_k(x^{k-1} - x^k) + \theta_{k-1}(x^{k-2} - x^{k-1}) \end{aligned}$$

由于 ∇f 是 l_f -Lipschitz 连续的, ∇T_2 是 l_{T_2} -Lipschitz 连续的, 且 $\lambda_{\max}$ 是矩阵 $\hat{\Sigma}_f$ 的最大特征值, 则

$$\begin{aligned} \|A^T \Delta \lambda^{k+1}\| &\leq l_f \|\Delta x^k\| + \lambda_{\max} \|\Delta x^{k+1} + \Delta x^k\| + l_{T_2} \|\Delta x^{k+1}\| + l_{T_2} \|\Delta x^k\| + \theta_k \|\Delta x^k\| + \theta_{k-1} \|\Delta x^{k-1}\| \\ &= (l_f + l_{T_2} + \lambda_{\max} + \theta_k) \|\Delta x^k\| + (l_{T_2} + \lambda_{\max}) \|\Delta x^{k+1}\| + \theta_{k-1} \|\Delta x^{k-1}\| \end{aligned}$$

上述不等式两边乘 $\frac{1}{\sqrt{\mu_{A^T}}}$ 得

$$\|P_A(\Delta \lambda^{k+1})\| = \|\Delta \lambda^{k+1}\| \leq \frac{1}{\sqrt{\mu_{A^T}}} ((l_f + l_{T_2} + \lambda_{\max} + \theta_k) \|\Delta x^k\| + (l_{T_2} + \lambda_{\max}) \|\Delta x^{k+1}\| + \theta_{k-1} \|\Delta x^{k-1}\|).$$

引理 2 若假设 1 成立, 令 $\{z^k\}$ 是由 IMBADMM 算法生成的序列, 那么以下结论成立:

- 1) 对于所有 $k \in N$, 有 $\hat{L}(\hat{z}^{k+1}) \leq \hat{L}(\hat{z}^k) - \kappa_1 \|\Delta y^{k+1}\|^2 - \kappa_2 \|\Delta x^{k+1}\|^2$;
- 2) 如果 $\{z^k\}$ 是有界的, 则 $\sum_{k=0}^{+\infty} \|z^{k+1} - z^k\|^2$ 。

证明:

1) (充分下降性质)

由于 y^{k+1} 是(5a)式的最优解, 所以

$$\begin{aligned} &\hat{L}_\beta(x^k, y^{k+1}, \lambda^k, x^k) - \hat{L}_\beta(x^k, y^k, \lambda^k, x^k) \\ &\leq -D_{T_1}(y^{k+1}, y^k) + \rho_k \langle y^k - y^{k+1}, y^{k-1} - y^k \rangle \\ &\leq \frac{\rho_k - \mu_1}{2} \|\Delta y^{k+1}\|^2 + \frac{\rho_k}{2} \|\Delta y^k\|^2, \end{aligned} \quad (11)$$

其中第 2 个不等式根据性质 1 的 3) 得出, 由于 x^{k+1} 是(5b)式的最优解, 故有

$$\begin{aligned} &\hat{f}(x^{k+1}, x^k) + \langle \lambda^k, Ax^{k+1} \rangle + \frac{\beta}{2} \|Ax^{k+1} - By^{k+1}\|^2 + D_{T_2}(x^{k+1}, x^k) + \theta_k \langle x^{k+1}, x^{k-1} - x^k \rangle \\ &\leq \hat{f}(x^k, x^k) + \langle \lambda^k, Ax^k \rangle + \frac{\beta}{2} \|Ax^k - By^{k+1}\|^2 + \theta_k \langle x^k, x^{k-1} - x^k \rangle, \end{aligned}$$

再利用 $\|a-b\|^2 - \|a\|^2 = -2\langle a, b \rangle + \|b\|^2$ 和性质 1 的 3), 可推出

$$\begin{aligned} &\hat{f}(x^{k+1}, x^k) - \hat{f}(x^k, x^k) + \langle \lambda^k, A\Delta x^{k+1} \rangle \\ &\leq -\beta \langle Ax^{k+1} - By^{k+1}, A\Delta x^{k+1} \rangle + \frac{\beta}{2} \|A\Delta x^{k+1}\|^2 + \frac{\theta_k - \mu_2}{2} \|\Delta x^{k+1}\|^2 + \frac{\theta_k}{2} \|\Delta x^k\|^2, \end{aligned} \quad (12)$$

根据 Majorized 增广拉格朗日函数 $\hat{L}_\beta(\cdot)$ 的定义得

$$\hat{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k, x^k) - \hat{L}_\beta(x^k, y^{k+1}, \lambda^k, x^k) \stackrel{(12)}{\leq} \frac{\theta_k}{2} \|\Delta x^k\|^2 + \frac{\theta_k - \mu_2}{2} \|\Delta x^{k+1}\|^2, \quad (13)$$

根据 $\hat{L}_\beta(\cdot)$ 的定义及引理 1, 得

$$\begin{aligned} & \hat{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}, x^k) - \hat{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k, x^k) \\ & \leq \frac{3}{\beta \mu_{A^T}} \left((l_f + l_{T_2} + \lambda_{\max} + \theta_k)^2 \|\Delta x^k\|^2 + (l_{T_2} + \lambda_{\max})^2 \|\Delta x^{k+1}\|^2 + \theta_{k-1}^2 \|\Delta x^{k-1}\|^2 \right), \end{aligned} \quad (14)$$

根据 $\hat{L}_\beta(\cdot)$ 的定义及拉格朗日中值定理, 得

$$\hat{L}_\beta(x^k, y^k, \lambda^k, x^k) - \hat{L}_\beta(x^k, y^k, \lambda^k, x^{k-1}) \leq \left(l_f + \frac{\lambda_{\max}}{2} \right) \|\Delta x^k\|^2, \quad (15)$$

由(11)、(13)、(14)、(15)式相加可得

$$\begin{aligned} & \hat{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}, x^k) - \hat{L}_\beta(x^k, y^k, \lambda^k, x^{k-1}) \\ & \leq \frac{\rho - \mu_1}{2} \|\Delta y^{k+1}\|^2 + \frac{\rho}{2} \|\Delta y^k\|^2 + \left(\frac{3(l_{T_2} + \lambda_{\max})^2}{\beta \mu_{A^T}} + \frac{\theta - \mu_2}{2} \right) \|\Delta x^{k+1}\|^2 \\ & \quad + \left(\frac{3(l_f + l_{T_2} + \lambda_{\max} + \theta)^2}{\beta \mu_{A^T}} + \frac{\theta + \lambda_{\max} + 2l_f}{2} \right) \|\Delta x^k\|^2 + \frac{3\theta^2}{\beta \mu_{A^T}} \|\Delta x^{k-1}\|^2, \end{aligned} \quad (16)$$

由算法过程可知 $0 \leq \rho_k \leq \rho$ 和 $0 \leq \theta_k \leq \theta$, 其中最后一个不等式由此得出。

根据(16)式, 得

$$\begin{aligned} & \hat{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}, x^k) + \frac{\rho}{2} \|\Delta y^{k+1}\|^2 + n_1 \|\Delta x^{k+1}\|^2 + n_2 \|\Delta x^k\|^2 \\ & \leq \hat{L}_\beta(x^k, y^k, \lambda^k, x^{k-1}) + \frac{\rho}{2} \|\Delta y^k\|^2 + n_1 \|\Delta x^k\|^2 + n_2 \|\Delta x^{k-1}\|^2 - \kappa_1 \|\Delta y^{k+1}\|^2 - \kappa_2 \|\Delta x^{k+1}\|^2, \end{aligned}$$

$$\text{则 } \hat{L}(\hat{z}^{k+1}) \leq \hat{L}(\hat{z}^k) - \kappa_1 \|\Delta y^{k+1}\|^2 - \kappa_2 \|\Delta x^{k+1}\|^2.$$

2) 由于 $\{z^k\}$ 有界, 则 $\{\hat{z}^k\}$ 有界并且至少有一个聚点, 即存在子序列 $\{\hat{z}^{k_j}\}$, 使得 $\lim_{j \rightarrow \infty} \hat{z}^{k_j} = \hat{z}^*$ 。因为 f 是连续可微函数, g 是下半连续函数, 所以 $\hat{L}(\cdot)$ 为下半连续函数, 因此, $\liminf_{j \rightarrow \infty} \hat{L}(\hat{z}^{k_j}) \geq \hat{L}(\hat{z}^*)$, $\hat{L}(\hat{z}^{k_j})$ 有下界。

根据假设 1 的 3) 知 $\kappa_i > 0, i=1, 2$, 因此 $\{\hat{L}(\hat{z}^k)\}$ 单调非增, 由于单调序列的子列收敛等价于全序列收敛, 即有 $\{\hat{L}(\hat{z}^k)\}$ 收敛且 $\hat{L}(\hat{z}^*) \leq \hat{L}(\hat{z}^k)$ 。由(10)式, 得

$$\kappa_1 \sum_{k=0}^{+\infty} \|\Delta y^{k+1}\|^2 + \kappa_2 \sum_{k=0}^{+\infty} \|\Delta x^{k+1}\|^2 \leq \hat{L}(\hat{z}^0) - \hat{L}(\hat{z}^*) < +\infty.$$

基于 $\kappa_i > 0, i=1, 2$, 可推导出 $\sum_{k=0}^{\infty} \|\Delta y^{k+1}\|^2 < +\infty$ 与 $\sum_{k=0}^{\infty} \|\Delta x^{k+1}\|^2 < +\infty$ 。

进而由(6)式, 可得 $\sum_{k=0}^{\infty} \|\Delta \lambda^{k+1}\|^2 < +\infty$ 。因此 $\sum_{k=0}^{\infty} \|z^{k+1} - z^k\|^2 < +\infty$ 。

注 1 由引理 2 的 1) 可知, 若 $\kappa_i > 0, i=1, 2$, 则序列 $\{\hat{L}(\hat{z}^k)\}$ 是非增的。

下面的引理给出效益函数次梯度的上限估计。

引理 3 若假设 1 成立, 令 $\{z^k\}$ 是由 IMBADMM 算法生成的序列, 当它有界时, 对于每一个 $k \geq 0$, 有

$$\varepsilon^{k+1} = (\varepsilon_1^{k+1}, \varepsilon_2^{k+1}, \varepsilon_3^{k+1}, \varepsilon_4^{k+1}, \varepsilon_5^{k+1}, \varepsilon_6^{k+1}, \varepsilon_7^{k+1})^T \in \partial \hat{L}(\hat{z}^{k+1}),$$

$$\text{其中 } \begin{cases} \varepsilon_1^{k+1} = \nabla f(x^k) + \hat{\Sigma}_f(x^{k+1} - x^k) + A^T \lambda^{k+1} + \beta A^T(Ax^{k+1} - By^{k+1}) + 2n_1 \Delta x^{k+1} \\ \varepsilon_2^{k+1} \in \partial g(y^{k+1}) - B^T \lambda^{k+1} - \beta B^T(Ax^{k+1} - By^{k+1}) + \rho \Delta y^{k+1} \\ \varepsilon_3^{k+1} = \|\Delta \lambda^{k+1}\| / \beta \\ \varepsilon_4^{k+1} = (\nabla^2 f(x^k) - \hat{\Sigma}_f) \Delta x^{k+1} \\ \varepsilon_5^{k+1} = -\rho \Delta y^{k+1} \\ \varepsilon_6^{k+1} = 2n_2 \Delta x^k - 2n_1 \Delta x^{k+1} \\ \varepsilon_7^{k+1} = -2n_2 \Delta x^k, \end{cases} \quad \circ$$

且存在 $\tau > 0$, 使得 $d(0, \partial \hat{L}(\hat{z}^{k+1})) \leq \|\varepsilon^{k+1}\| \leq \tau (\|\Delta y^k\| + \|\Delta y^{k+1}\| + \|\Delta x^{k-1}\| + \|\Delta x^k\| + \|\Delta x^{k+1}\|)$ 。

证明: 根据(7)式效益函数的定义得

$$\begin{aligned} \hat{L}(\hat{z}^{k+1}) &= \hat{f}(x^{k+1}, x^k) + g(y^{k+1}) + \langle \lambda^{k+1}, Ax^{k+1} - By^{k+1} \rangle + \frac{\beta}{2} \|Ax^{k+1} - By^{k+1}\|^2 \\ &\quad + \frac{\rho}{2} \|\Delta y^{k+1}\|^2 + n_1 \|\Delta x^{k+1}\|^2 + n_2 \|\Delta x^k\|^2 \end{aligned} \quad (17)$$

对 $\hat{L}(\cdot)$ 各分量求偏导, ε^{k+1} 可以表示为

$$\varepsilon_1^{k+1} = \nabla f(x^k) + \hat{\Sigma}_f(x^{k+1} - x^k) + A^T \lambda^{k+1} + \beta A^T(Ax^{k+1} - By^{k+1}) + 2n_1 \Delta x^{k+1} \quad (18a)$$

$$\varepsilon_2^{k+1} \in \partial g(y^{k+1}) - B^T \lambda^{k+1} - \beta B^T(Ax^{k+1} - By^{k+1}) + \rho \Delta y^{k+1} \quad (18b)$$

$$\varepsilon_3^{k+1} = Ax^{k+1} - By^{k+1} \quad (18c)$$

$$\varepsilon_4^{k+1} = (\nabla^2 f(x^k) - \hat{\Sigma}_f) \Delta x^{k+1} \quad (18d)$$

$$\varepsilon_5^{k+1} = -\rho \Delta y^{k+1} \quad (18e)$$

$$\varepsilon_6^{k+1} = 2n_2 \Delta x^k - 2n_1 \Delta x^{k+1} \quad (18f)$$

$$\varepsilon_7^{k+1} = -2n_2 \Delta x^k. \quad (18g)$$

根据(5b)式的最优性条件得

$$\nabla f(x^k) + \hat{\Sigma}_f(x^{k+1} - x^k) = -A^T \lambda^k - \beta A^T(Ax^{k+1} - By^{k+1}) + \nabla T_2(x^k) - \nabla T_2(x^{k+1}) - \theta_k(x^{k-1} - x^k)$$

代入(18a)式有

$$\varepsilon_1^{k+1} = A^T \Delta \lambda^{k+1} + \nabla T_2(x^k) - \nabla T_2(x^{k+1}) - \theta_k(x^{k-1} - x^k) + 2n_1 \Delta x^{k+1}.$$

由于 ∇T_2 是 l_{T_2} -Lipschitz 连续的, 则

$$\|\varepsilon_1^{k+1}\| \leq \|A\| \|\Delta \lambda^{k+1}\| + (l_{T_2} + 2n_1) \|\Delta x^{k+1}\| + \theta_k \|\Delta x^k\|. \quad (19)$$

根据(5a)式的最优性条件得

$$\partial g(y^{k+1}) \ni B^T \lambda^k + \beta B^T(Ax^k - By^{k+1}) + \nabla T_1(y^k) - \nabla T_1(y^{k+1}) - \rho_k(y^{k-1} - y^k)$$

代入(18b)式有

$$\varepsilon_2^{k+1} = -B^T \Delta \lambda^{k+1} - \beta B^T A \Delta x^{k+1} + \nabla T_1(y^k) - \nabla T_1(y^{k+1}) + \rho_k \Delta y^k + \rho \Delta y^{k+1}.$$

由于 ∇T_1 是 l_{T_1} -Lipschitz 连续的, 则

$$\|\varepsilon_2^{k+1}\| \leq \|B\| \|\Delta \lambda^{k+1}\| + \beta \|B\| \|A\| \|\Delta x^{k+1}\| + (l_{T_1} + \rho) \|\Delta y^{k+1}\| + \rho_k \|\Delta y^k\|. \quad (20)$$

根据(5c)式得 $Ax^{k+1} - By^{k+1} = \frac{1}{\beta} \Delta \lambda^{k+1}$, 则

$$\|\varepsilon_3^{k+1}\| = \frac{1}{\beta} \|\Delta \lambda^{k+1}\|. \quad (21)$$

因此, $\varepsilon^{k+1} \in \partial \hat{L}(\hat{z}^{k+1})$, 根据(18d~18g)式及(19~21)式, 存在 $\xi_1 > 0$ 使得

$$\|\varepsilon^{k+1}\| \leq \xi_1 (\|\Delta x^k\| + \|\Delta x^{k+1}\| + \|\Delta y^k\| + \|\Delta y^{k+1}\| + \|\Delta \lambda^{k+1}\|).$$

根据引理 1, 存在 $\xi_2 > 0$ 使得 $\|\Delta \lambda^{k+1}\| \leq \xi_2 (\|\Delta x^k\| + \|\Delta x^{k+1}\| + \|\Delta x^{k-1}\|)$ 。

综合上述两个不等式, 当 $\varepsilon^{k+1} \in \partial \hat{L}(\hat{z}^{k+1})$ 时, 可得

$$\begin{aligned} d(0, \partial \hat{L}(\hat{z}^{k+1})) &\leq \|\varepsilon^{k+1}\| \leq \xi_1 (1 + \xi_2) (\|\Delta y^k\| + \|\Delta y^{k+1}\| + \|\Delta x^{k-1}\| + \|\Delta x^k\| + \|\Delta x^{k+1}\|) \\ &= \tau (\|\Delta y^k\| + \|\Delta y^{k+1}\| + \|\Delta x^{k-1}\| + \|\Delta x^k\| + \|\Delta x^{k+1}\|), \end{aligned}$$

其中 $\tau = \xi_1 (1 + \xi_2) > 0$ 。

定理 1 (子序列收敛性) 若假设 1 成立, $\{z^k\}$ 为 IMBADMM 算法生成的序列, 且假设 $\{z^k\}$ 有界。令 Ω 表示序列 $\{z^k\}$ 的聚点集, $\hat{\Omega}$ 表示序列 $\{\hat{z}^k\}$ 的聚点集, 则以下结论成立:

- 1) Ω 是一个非空紧集, 且当 $k \rightarrow \infty$ 时, $d(z^k, \Omega) \rightarrow 0$; Ω 里面的点都是 $L_\beta(\cdot)$ 的临界点; $(x^*, y^*, \lambda^*, x^*, \hat{y}^*, \hat{x}^*, \bar{x}^*) \in \hat{\Omega}$, 当且仅当 $x^* = x^* = \hat{x}^* = \bar{x}^*$, $y^* = \hat{y}^*$ 且 $(x^*, y^*, \lambda^*) \in \Omega$;
- 2) $\hat{L}(\cdot)$ 是有限的且在 $\hat{\Omega}$ 上为常数, $\inf_{k \in N} \hat{L}(\hat{z}^k) = \lim_{k \rightarrow +\infty} \hat{L}(\hat{z}^k)$ 。

接下来的结果显示, 当效益函数是 KL 函数时, 序列 $\{z^k\}$ 将收敛到优化问题(1)的临界点。

证明:

1) 因 $\{z^k\}$ 有界, 故 $\{z^k\}$ 至少存在一个聚点, 其聚点集 Ω 非空且为紧集。

令 $z^* \in \Omega$, 那么存在 $\{z^k\}$ 的一个子序列 $\{z^{k_j}\}$, 使得: $\lim_{j \rightarrow \infty} (z^{k_j}) = z^*$ 。

根据引理 2 的 2), 有 $\lim_{j \rightarrow \infty} \|z^{k_j+1} - z^{k_j}\| = 0$, 进而可得 $\lim_{j \rightarrow \infty} (z^{k_j+1}) = z^*$ 。

由于 $\lambda^{k_j+1} = \lambda^{k_j} + \beta (Ax^{k_j+1} - By^{k_j+1})$ 且 $\beta > 0$, 从而 $Ax^* - By^* = 0$ 。

又由于 y^{k+1} 是(5a)的最优解, 故有

$$\begin{aligned} \hat{L}(x^k, y^{k+1}, \lambda^k, x^k, y^k, x^{k-1}, x^{k-2}) &\leq \hat{L}(x^k, y^*, \lambda^k, x^k, y^k, x^{k-1}, x^{k-2}) \\ &+ D_{T_1}(y^*, y^k) - D_{T_1}(y^{k+1}, y^k) + \frac{\rho}{2} \|y^{k+1} - y^k\|^2 - \frac{\rho}{2} \|y^* - y^k\|^2 - \rho_k \langle y^* - y^{k+1}, \Delta y^k \rangle. \end{aligned} \quad (22)$$

因为效益函数 $\hat{L}(x, y, \lambda, x', \hat{y}, \hat{x}, \bar{x})$ 关于变量 $(x, \lambda, x', \hat{y}, \hat{x}, \bar{x})$ 是连续的, 所以

$$\begin{aligned} \limsup_{j \rightarrow \infty} \hat{L}(\hat{z}^{k_j+1}) &\stackrel{(22)}{\leq} \limsup_{j \rightarrow \infty} \{ \hat{L}(x^{k_j}, y^*, \lambda^{k_j}, x^{k_j}, y^{k_j}, x^{k_j-1}, x^{k_j-2}) + \Delta_{T_1}(y^*, y^{k_j}) - \Delta_{T_1}(y^{k_j+1}, y^{k_j}) \\ &+ \frac{\rho}{2} \|y^{k_j+1} - y^{k_j}\|^2 - \frac{\rho}{2} \|y^* - y^{k_j}\|^2 - \rho_k \langle y^* - y^{k_j+1}, \Delta y^{k_j} \rangle \} = \hat{L}(\hat{z}^*). \end{aligned} \quad (23)$$

由于 f 是连续函数且 g 是下半连续函数, 则 $\hat{L}(\cdot)$ 是下半连续函数。因此,

$$\hat{L}(\hat{z}^*) \leq \liminf_{j \rightarrow \infty} \hat{L}(\hat{z}^{k_j+1}). \quad (24)$$

根据(23)式和(24)式可得

$$\lim_{j \rightarrow \infty} \hat{L}(\hat{z}^{k_j+1}) = \hat{L}(\hat{z}^*). \quad (25)$$

由此可得 $\lim_{k_j \rightarrow \infty} g(y^{k_j+1}) = g(y^*)$ 。利用 ∇f 的连续性和 ∂g 的闭合性质, 结合(5)式的最优性条件得 $B^T \lambda^* \in \partial g(y^*)$, $\nabla f(x^*) = -A^T \lambda^*$, $Ax^* - By^* = 0$ 。

这表明 z^* 是 $L_\beta(\cdot)$ 的临界点。

2) 根据(25)式有 $\liminf_{k \rightarrow \infty} \hat{L}(\hat{z}^{k+1}) = \lim_{j \rightarrow \infty} \hat{L}(\hat{z}^{k_j+1}) = \hat{L}(\hat{z}^*) = f(x^*) + g(y^*) = L_\beta(z^*)$, 又由引理 2, $\hat{L}(\cdot)$ 是单调非增的, 故有 $\inf_{k \in N} \hat{L}(\hat{z}^k) = \lim_{k \rightarrow \infty} \hat{L}(\hat{z}^k) = \hat{L}(\hat{z}^*)$ 。

定理 2 (全局收敛性) 若假设 1 成立, $\{z^k\}$ 是 IMBADMM 算法生成的序列, 且假设 $\{z^k\}$ 有界。设 $\hat{\Omega}$ 表示序列 $\{\hat{z}^k\}$ 的聚点集, 如果 $\hat{L}(\cdot)$ 满足 $\hat{\Omega}$ 上的 KL 性质, 则以下结论成立:

1) 序列 $\{z^k\}$ 具有有限长度, 即: $\sum_{k=0}^{+\infty} \|z^{k+1} - z^k\| < +\infty$ 。

2) 序列 $\{z^k\}$ 收敛到优化问题(1)的临界点。

证明: 根据定理 1 的 2) 得 $\lim_{k \rightarrow \infty} \hat{L}(\hat{z}^k) = \hat{L}(\hat{z}^*)$, 对所有 $\hat{z}^* \in \hat{\Omega}$, 我们考虑以下两种情况。

a) 假设存在 $k_0 \in N$, 满足 $\hat{L}(\hat{z}^{k_0}) = \hat{L}(\hat{z}^*)$ 。由引理 2 的 1) 知

当 $\kappa_1 > 0$ 且 $\kappa_2 > 0$ 时, 不等式 $\kappa_1 \|\Delta y^{k+1}\|^2 + \kappa_2 \|\Delta x^{k+1}\|^2 \leq \hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^{k+1})$ 成立。由于假设了 $\hat{L}(\hat{z}^{k_0}) = \hat{L}(\hat{z}^*)$, 则对于任意 $k > k_0$, 有 $\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^{k+1}) \leq \hat{L}(\hat{z}^{k_0}) - \hat{L}(\hat{z}^*) = 0$ 。结合 $\kappa_1 > 0$ 且 $\kappa_2 > 0$, 可推出对任意 $k > k_0$, 有 $x^{k+1} = x^k$ 且 $y^{k+1} = y^k$ 。

再结合引理 1, 可知对任意 $k > k_0$, $\lambda^{k+1} = \lambda^k$, 综上有 $z^{k+1} = z^k$ 。

b) 假设对于任意 $k > 1$, 均有 $\hat{L}(\hat{z}^k) > \hat{L}(\hat{z}^*)$ 。因为 $d(\hat{z}^k, \hat{\Omega}) \rightarrow 0$, 对于给定 $\varepsilon > 0$, 存在 $k_1 > 0$, 当 $k > k_1$ 时, 有 $d(\hat{z}^k, \hat{\Omega}) < \varepsilon$ 。又因为 $\lim_{k \rightarrow \infty} \hat{L}(\hat{z}^k) = \hat{L}(\hat{z}^*)$, 对于给定 $\eta > 0$, 存在 $k_2 > 0$, 当 $k > k_2$ 时, 有

$\hat{L}(\hat{z}^k) < \hat{L}(\hat{z}^*) + \eta$ 。如果 $k > \tilde{k} = \max\{k_1, k_2\}$, 那么 $d(\hat{z}^k, \hat{\Omega}) < \varepsilon$, $\hat{L}(\hat{z}^*) < \hat{L}(\hat{z}^k) < \hat{L}(\hat{z}^*) + \eta$ 。根据性质 2, 对任意的 $k > \tilde{k}$, 存在可微凹函数 $\varphi \in \Phi_\eta$, 有 $\varphi'(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*))d(0, \partial \hat{L}(\hat{z}^k)) \geq 1$ 。因为: $\varphi'(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*)) > 0$, 所以

$$\frac{1}{\varphi'(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*))} \leq d(0, \partial \hat{L}(\hat{z}^k)). \quad (26)$$

根据 φ 的凹性得

$$\varphi(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*)) - \varphi(\hat{L}(\hat{z}^{k+1}) - \hat{L}(\hat{z}^*)) \geq \varphi'(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*))(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^{k+1}))$$

$$\text{由此可推导出 } 0 \leq \hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^{k+1}) \leq \frac{\varphi(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*)) - \varphi(\hat{L}(\hat{z}^{k+1}) - \hat{L}(\hat{z}^*))}{\varphi'(\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^*))}.$$

为了简化表达, 定义 $\Delta_{p,q} := \varphi(\hat{L}(\hat{z}^p) - \hat{L}(\hat{z}^*)) - \varphi(\hat{L}(\hat{z}^q) - \hat{L}(\hat{z}^*))$, 以及

$$\Gamma_k = \|\Delta y^{k-1}\| + \|\Delta y^k\| + \|\Delta x^k\| + \|\Delta x^{k-1}\| + \|\Delta x^{k-2}\|.$$

由(26)式及引理 3, 对于任意 $k \geq 1$, 有

$$\hat{L}(\hat{z}^k) - \hat{L}(\hat{z}^{k+1}) \stackrel{(26)}{\leq} \Delta_{k, k+1} d(0, \partial \hat{L}(\hat{z}^k)) \leq \tau \Delta_{k, k+1} \Gamma_k.$$

根据引理 2 的 1) 及上述不等式, 可知对任意 $k > \tilde{k}$, 有

$$\kappa_1 \|\Delta y^{k+1}\|^2 + \kappa_2 \|\Delta x^{k+1}\|^2 \leq \tau \Delta_{k, k+1} \Gamma_k.$$

上述分析表明, 存在 $\xi = \min\{\kappa_1, \kappa_2\} > 0$, 使得

$$\|\Delta x^{k+1}\|^2 + \|\Delta y^{k+1}\|^2 \leq \frac{\tau}{\xi} \Delta_{k, k+1} \Gamma_k.$$

利用不等式 $2ab \leq a^2 + b^2$, 有

$$(\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|)^2 \leq 2(\|\Delta x^{k+1}\|^2 + \|\Delta y^{k+1}\|^2) \leq 2 \frac{\tau}{\xi} \Delta_{k, k+1} \Gamma_k.$$

可推出

$$4(\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|) \leq \Gamma_k + \frac{8\tau}{\xi} \Delta_{k, k+1} \leq \Gamma_k + \|\Delta y^{k-2}\| + \frac{8\tau}{\xi} \Delta_{k, k+1}.$$

$$\text{由于 } 2\Gamma_k^{\frac{1}{2}} \left(2\sqrt{\frac{2\tau}{\xi}} \Delta_{k, k+1}^{\frac{1}{2}} \right) \leq \Gamma_k + \frac{8\tau}{\xi} \Delta_{k, k+1},$$

根据上述公式得

$$4 \sum_{k=k+1}^m (\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|) \leq \sum_{k=k+1}^m (\Gamma_k + \|\Delta y^{k-2}\|) + \frac{8\tau}{\xi} \Delta_{\tilde{k}+1, m+1}.$$

两边取极限 $m \rightarrow +\infty$, 得

$$\begin{aligned} \sum_{k=k+1}^{+\infty} (\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|) &\leq \sum_{k=k+1}^{+\infty} (\|\Delta y^{k-1}\| + \|\Delta y^k\| + \|\Delta x^k\| + \|\Delta x^{k-1}\| + \|\Delta x^{k-2}\| + \|\Delta y^{k-2}\|) + \frac{8\tau}{\xi} \Delta_{\tilde{k}+1, m+1} \\ &\leq (3\|\Delta y^{\tilde{k}+1}\| + 2\|\Delta y^{\tilde{k}}\| + \|\Delta y^{\tilde{k}-1}\| + 3\|\Delta x^{\tilde{k}+1}\| + 2\|\Delta x^{\tilde{k}}\| + \|\Delta x^{\tilde{k}-1}\|) + \frac{8\tau}{\xi} \varphi(\hat{L}(\hat{z}^{\tilde{k}+1}) - \hat{L}(\hat{z}^*)) \end{aligned}$$

因此, $\sum_{k=k+1}^{\infty} (\|\Delta x^{k+1}\| + \|\Delta y^{k+1}\|) < +\infty$ 。

根据引理 1 及 $0 \leq \theta_k \leq \theta$, 我们推导出 $\sum_{k=1}^{\infty} (\|\Delta \lambda^{k+1}\|) < +\infty$ 。

进一步观察到 $\|z^{k+1}\| = (\|\Delta x^{k+1}\|^2 + \|\Delta y^{k+1}\|^2 + \|\Delta \lambda^{k+1}\|^2)^{\frac{1}{2}} \leq \|\Delta x^{k+1}\| + \|\Delta y^{k+1}\| + \|\Delta \lambda^{k+1}\|$ 。

由此可得 $\sum_{k=1}^{\infty} \|\Delta z^{k+1}\| < +\infty$, 这表明 $\{z^k\}$ 是一个柯西序列, 从而 $\{z^k\}$ 收敛。

从定理 1, 我们知 $\{z^k\}$ 收敛到 $L_{\beta}(\cdot)$ 的临界点。

4. 数值实验

下面通过数值实验对比 IMBADMM、MBADMM [8]、ADMM [17]、SGD [18] 与 LBADMM [19] 算法

的性能。所有实验均在配备 16 GB 内存、3.20 GHz、AMD Ryzen 7 6800H with Radeon Graphics CPU 的联想电脑上进行，操作系统为 Windows 11，使用 Matlab R2023a 版本进行编码和执行。

针对问题(1)考虑稀疏逻辑回归问题，基本模型如下：

$$\min_x \sum_{i=1}^n \log(1 + \exp(-b_i a_i^T x)) + n\rho \|x\|_q. \quad (27)$$

其中： $a_i \in R^d$ 是样本 $i(i=1,2,\dots,n)$ 的特征向量， $b_i \in \{1,-1\}$ 是相应的二分类， ρ 是一个正则化参数， $\|\cdot\|_q$ 是 l_q 范数，对任意的 $x \in R^d$ 有 $\|x\|_q = \left(\sum_{i=1}^n |x_i|^q\right)^{1/q}$ ， $q \in (0,1]$ 。令 $A = (a_1, a_2, \dots, a_n)^T \in R^{n \times d}$ ， $b = (b_1, b_2, \dots, b_n)^T \in R^n$ ，则 (a_i, b_i) ($i \in \{1,2,\dots,n\}$) 是训练集 (a,b) 的一个样本。该问题有 n 个样本，维数为 d 。

在实际应用中，我们将(27)式转化为：

$$\min_x \sum_{i=1}^n \log(1 + \exp(-b_i a_i^T x)) + n\rho \|y\|_q, \quad \text{s.t. } x - y = 0. \quad (28)$$

于是(28)式即为优化问题(1)的形式，其中 $f(x) = \sum_{i=1}^n \log(1 + \exp(-b_i a_i^T x))$ ， $g(y) = n\rho \|y\|_q$ 。易知： ∇f 是 Lipschitz 连续的，其 Lipschitz 连续常数 $L = \frac{1}{4} \|A\|^2$ 。所以存在一个半正定矩阵 $\hat{\Sigma}_f$ 使得对任意给定的 $x' \in R^d$ ，有：

$$f(x) \leq \hat{f}(x, x') := f(x') + \langle \nabla f(x'), x - x' \rangle + \frac{1}{2} \|x - x'\|_{\hat{\Sigma}_f}^2,$$

令 $C_i^T = -b_i a_i^T$ ，经计算有：

$$\nabla f(x) = \sum_{i=1}^n \frac{C_i e^{C_i x}}{1 + e^{C_i x}}, \quad \nabla^2 f(x) = C_i C_i^T \sum_{i=1}^n \frac{C_i e^{C_i x}}{(1 + e^{C_i x})^2} \leq \frac{1}{4} \sum_{i=1}^n C_i C_i^T, \quad \text{因此，} \hat{\Sigma}_f \text{ 的近似项可以选择为：}$$

$\hat{\Sigma}_f = \frac{1}{4} C C^T$ ，其中 $C = [C_1, \dots, C_n] \in R^{d \times n}$ 。对五种算法最大迭代次数设为 5000，取 $\rho_k = 0.01$ ， $\theta_k = 0.01$ 。

本实验数据来源于 LIBSVM 数据库(<https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/>)提供的 a2000a 数据集进行了数值实验。分别取不同的范数正则项 $q = 1$ 和 $q = 1/2$ ，表 1 展示了 IMBADMM、MBADMM、ADMM、SGD 以及 LBADMM 算法在处理稀疏逻辑回归问题(28)时的数值结果比较。观察表 1 中的数据，在计算效率方面，IMBADMM 仅需 0.07 秒即可完成优化，相比 MBADMM 提速 25~30 倍；在收敛速度方面，IMBADMM 所需迭代次数最少，相比 ADMM 减少约 50%~65% 的迭代次数；特别值得注意的是，在保持计算精度的前提下($q = 1$ 时目标函数值 616.01 vs MBADMM 的 616.02)，IMBADMM 实现了计算效率的显著提升。对于非光滑情况($q = 1/2$)，IMBADMM 同样展现出强大优势，相比 LBADMM 将计算时间从 11.37 秒降至 0.07 秒，提速 162 倍，同时迭代次数从 2326 次减少至 57 次。这些实验结果充分证明了 IMBADMM 算法在处理大规模优化问题的卓越性能。与随机梯度下降(SGD)算法的对比显示：虽然 SGD 在计算时间上略有优势(0.02 s vs 0.07 s)，但 IMBADMM 在优化质量和收敛性方面表现出优势。为了直观分析算法的收敛性能，我们分别绘制了 IMBADMM、MBADMM 和 ADMM 算法在参数取值为 $q = 1$ 和 $q = 1/2$ 时的收敛曲线。其中，图 1 展示了 IMBADMM 算法的误差随迭代次数的演化过程，图 2 则呈现了三种算法的目标函数值随迭代次数的收敛特性。通过图 2 可知，IMBADMM 算法展现出最平滑的收敛轨迹，算法更具有稳定性且收敛速度快。

Table 1. Comparison of numerical results of the five algorithms
表 1. 五个算法的数值结果对比

算法	维度大小	数据来源	迭代次数	计算时间/s	目标函数值
IMBADMM ($q = 1$)	999×60	a2000a	55	0.07	616.01
MBADMM ($q = 1$)	999×60	a2000a	55	1.82	616.02
ADMM ($q = 1$)	999×60	a2000a	119	1.76	613.91
SGD ($q = 1$)	999×60	a2000a	150	0.02	962.71
LBADMM ($q = 1$)	999×60	a2000a	3006	1.18	620.00
IMBADMM ($q = 1/2$)	999×60	a2000a	57	0.07	672.07
MBADMM ($q = 1/2$)	999×60	a2000a	57	2.07	672.07
ADMM ($q = 1/2$)	999×60	a2000a	166	1.96	670.5
SGD ($q = 1/2$)	999×60	a2000a	150	0.02	1395.03
LBADMM ($q = 1/2$)	999×60	a2000a	2326	11.37	692.45

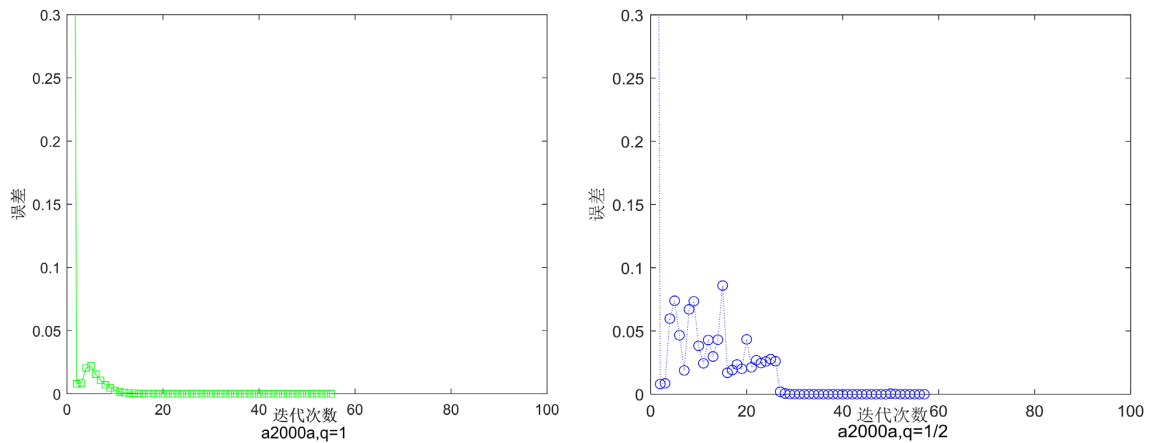


Figure 1. Evolution curve of the error with number of iterations when $q = 1$ and $q = 1/2$
图 1. 当 $q = 1$ 和 $q = 1/2$ 时, 误差随迭代次数的演化曲线

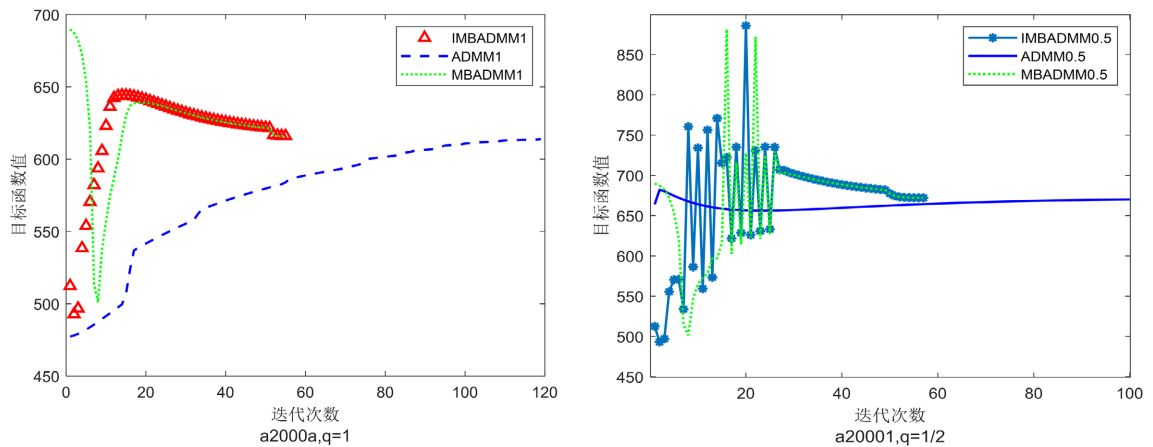


Figure 2. Evolution curve of the objective function value with the number of iterations when $q = 1$ and $q = 1/2$
图 2. 当 $q = 1$ 和 $q = 1/2$ 时, 目标函数值随迭代次数的演化曲线

5. 结论

本文针对非凸两分块优化问题, 提出了一种带惯性项的 Majorized Bregman 交替方向乘子法(简称 IMBADMM), 新算法结合了 Majorized Bregman ADMM 和惯性 Bregman generalized ADMM 的基本思想。在适当的条件下, 我们建立了算法的子序列收敛性和全局收敛性。最后, 初步数值实验结果表明 IMBADMM 算法的有效性和可靠性。

基金项目

广西自然科学基金(2022GXNSFAA035618)。

参考文献

- [1] Schizas, I.D., Ribeiro, A. and Giannakis, G.B. (2007) Consensus in AD Hoc WSNs with Noisy Links—Part I: Distributed Estimation of Deterministic Signals. *IEEE Transactions on Signal Processing*, **56**, 350-364. <https://doi.org/10.1109/TSP.2007.906734>
- [2] Afonso, M.V., Bioucas-Dias, J.M. and Figueiredo, M.A.T. (2010) An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems. *IEEE Transactions on Image Processing*, **20**, 681-695. <https://doi.org/10.1109/TIP.2010.2076294>
- [3] Dhar, S., Yi, C., Ramakrishnan, N. and Shah, M. (2015) ADMM Based Scalable Machine Learning on Spark. 2015 *IEEE International Conference on Big Data (Big Data)*, Santa Clara, 29 October 2015-1 November 2015, 1174-1182. <https://doi.org/10.1109/bigdata.2015.7363871>
- [4] Yang, Q., Chen, G. and Wang, T. (2020) ADMM-Based Distributed Algorithm for Economic Dispatch in Power Systems with Both Packet Drops and Communication Delays. *IEEE/CAA Journal of Automatica Sinica*, **7**, 842-852. <https://doi.org/10.1109/jas.2020.1003156>
- [5] 邓钊, 晁锦涛, 简金宝. 非凸两分块问题乘子交替方向法的收敛性分析[J]. 广西科学, 2016, 23(5): 422-427.
- [6] Wang, Y., Yin, W. and Zeng, J. (2019) Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. *Journal of Scientific Computing*, **78**, 29-63. <https://doi.org/10.1007/s10915-018-0757-z>
- [7] Wang, F., Cao, W. and Xu, Z. (2018) Convergence of Multi-Block Bregman ADMM for Nonconvex Composite Problems. *Science China Information Sciences*, **61**, Article No. 12201. <https://doi.org/10.1007/s11432-017-9367-6>
- [8] 陈建华, 彭建文, 罗洪林. 求解非凸两分块优化问题的 Majorized Bregman 交替方向乘子法[J]. 重庆师范大学学报(自然科学版), 2023, 40(5): 1-10.
- [9] Polyak, B.T. (1964) Some Methods of Speeding up the Convergence of Iteration Methods. *USSR Computational Mathematics and Mathematical Physics*, **4**, 1-17. [https://doi.org/10.1016/0041-5553\(64\)90137-5](https://doi.org/10.1016/0041-5553(64)90137-5)
- [10] Ochs, P., Chen, Y., Brox, T. and Pock, T. (2014) iPiano: Inertial Proximal Algorithm for Nonconvex Optimization. *SIAM Journal on Imaging Sciences*, **7**, 1388-1419. <https://doi.org/10.1137/130942954>
- [11] Boţ, R.I., Csetnek, E.R. and László, S.C. (2016) An Inertial Forward-Backward Algorithm for the Minimization of the Sum of Two Nonconvex Functions. *EURO Journal on Computational Optimization*, **4**, 3-25. <https://doi.org/10.1007/s13675-015-0045-8>
- [12] Xu, J. and Chao, M. (2022) An Inertial Bregman Generalized Alternating Direction Method of Multipliers for Nonconvex Optimization. *Journal of Applied Mathematics and Computing*, **68**, 1-27. <https://doi.org/10.1007/s12190-021-01590-1>
- [13] Chao, M.T., Zhang, Y. and Jian, J.B. (2020) An Inertial Proximal Alternating Direction Method of Multipliers for Nonconvex Optimization. *International Journal of Computer Mathematics*, **98**, 1199-1217. <https://doi.org/10.1080/00207160.2020.1812585>
- [14] 刘浩洋, 卢将, 李勇锋, 等. 最优化: 建模、算法与理论[M]. 北京: 高等教育出版社, 2020.
- [15] Wang, F., Xu, Z. and Xu, H.K. (2014) Convergence of Bregman Alternating Direction Method with Multipliers for Nonconvex Composite Problems. arXiv:1410.8625.
- [16] Gonçalves, M.L.N., Melo, J.G. and Monteiro, R.D.C. (2017) Convergence Rate Bounds for a Proximal ADMM with Over-Relaxation Step Size Parameter for Solving Nonconvex Linearly Constrained Problems. arXiv:1702.01850.
- [17] Boyd, S. (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, **3**, 1-122. <https://doi.org/10.1561/22000000016>

- [18] 李晶晶. 基于梯度的三种优化方法及比较[J]. 统计学与应用, 2024, 13(1): 21-29.
- [19] Chao, M., Deng, Z. and Jian, J. (2020) Convergence of Linear Bregman ADMM for Nonconvex and Nonsmooth Problems with Nonseparable Structure. *Complexity*, **2020**, Article 6237942. <https://doi.org/10.1155/2020/6237942>