

求解非凸非光滑优化问题的一类惯性Moreau包络交替方向乘子法

史梦蝶

河北工业大学理学院, 天津

收稿日期: 2026年2月5日; 录用日期: 2026年2月28日; 发布日期: 2026年3月11日

摘要

针对一类带有线性约束的非凸非光滑优化问题, 本文提出了一类惯性Moreau包络交替方向乘子法 (IMADMM)。该方法通过引入Moreau包络对非光滑项进行光滑近似, 并嵌入惯性步以加速迭代; 另外, 在每次迭代中采用Armijo线搜索确定步长, 在保持算法稳定性的同时提升了收敛速度。在理论分析方面, 在Kurdyka-Łojasiewicz (KL) 框架下证明了算法的全局收敛性。最后, 数值实验以稀疏最小二乘问题为例, 验证了所提算法的有效性和优越性。

关键词

非凸非光滑优化, 交替方向乘子法, Moreau包络, Armijo线搜索, Kurdyka-Łojasiewicz不等式

An Inertial Moreau Envelope Alternating Direction Method of Multipliers for Solving Nonconvex Nonsmooth Optimization Problems

Mengdie Shi

School of Science, Hebei University of Technology, Tianjin

Received: February 5, 2026; accepted: February 28, 2026; published: March 11, 2026

Abstract

This paper proposes an inertial Moreau envelope alternating direction method of multipliers (IMADMM) for a class of nonconvex nonsmooth optimization problems with linear constraints. The

method introduces the Moreau envelope to smoothly approximate the nonsmooth term and embeds an inertial step to accelerate the iteration. Moreover, an Armijo line search is employed at each iteration to determine the step size, enhancing the convergence rate while maintaining algorithmic stability. Theoretically, the global convergence of the algorithm is established under the Kurdyka-Łojasiewicz (KL) framework. Finally, numerical experiments on sparse least squares problems validate the effectiveness and superiority of the proposed method.

Keywords

Nonconvex Nonsmooth Optimization, Alternating Direction Method of Multipliers, Moreau Envelope, Armijo Line Search, Kurdyka-Łojasiewicz Inequality

Copyright © 2026 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

本文考虑如下形式的一类非凸非光滑优化问题:

$$\min_{x,y} f(x) + g(y) \text{ s.t. } Ax + By = b, \tag{1}$$

其中, $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是一个适当下半连续函数, 并且没有凸性假设, $g: \mathbb{R}^m \rightarrow \mathbb{R}$ 是带有梯度 Lipschitz 连续的可微函数, $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{m \times m}, b \in \mathbb{R}^m$ 分别是一些给定的矩阵和向量。这类问题在许多领域都有重要应用, 例如信号和图像处理, 机器学习, 统计学[1]-[3]等。随着现代科技的快速发展, 这类问题的重要性日益凸显。例如, 在机器学习领域, 深度神经网络的训练问题本质上是非凸的, 而正则化技术如 L1 正则化则引入了非光滑性[4]; 在信号处理中, 稀疏编码和压缩感知问题中的稀疏约束同样带来了非凸非光滑优化挑战[5]。

求解问题(1)的一类经典方法是交替方向乘子法(Alternating Direction Method of Multipliers, 简称 ADMM) [6], 可描述为,

$$\begin{cases} x^{k+1} \in \arg \min_x \mathcal{L}_\beta(x, y^k, \lambda^k), \\ y^{k+1} \in \arg \min_y \mathcal{L}_\beta(x^{k+1}, y, \lambda^k), \\ \lambda^{k+1} = \lambda^k + \beta(Ax^{k+1} + By^{k+1} - b). \end{cases}$$

其中,

$$\mathcal{L}_\beta(x, y, \lambda) = f(x) + g(y) + \lambda^T (Ax + By - b) + \frac{\beta}{2} \|Ax + By - b\|^2, \tag{2}$$

为增广拉格朗日函数, $\lambda \in \mathbb{R}^m$ 是拉格朗日乘子向量, $\beta > 0$ 是惩罚参数。

由于 ADMM 这类一阶算法通常会存在收敛速度慢的问题, 从而有一些学者考虑对其进行改进。近年来, 由于惯性技术可以有效地提高算法的数值性能, 其已经被广泛应用于一阶算法中。惯性思想最早起源于文[7]中所提出的重球法,

$$x^{k+1} = x^k - \eta_1 P(x^k) + \eta_2 (x^k - x^{k-1}),$$

其中, $P(x)$ 为希尔伯特空间中某些光滑函数的梯度, $\eta_2 (x^n - x^{n-1})$ 为惯性外推步。这里, $\eta_1 > 0, \eta_2 \in [0, 1)$ 。

当 $\eta_2 = 0$ 时, 重球法则退化为梯度法。Alvarez 在文献[8]中首次提出了惯性邻近点算法(IPPA), 惯性步有效地加快了邻近点算法的收敛速度。最近, 一些学者研究了基于惯性的可分优化分裂方法, 如求解某些可分非凸优化问题和强凸优化问题的惯性前向后向分裂方法[9] [10], 惯性 D-R 算子分裂方法[11], 求解非凸优化问题的惯性邻近梯度法[12]和惯性邻近交替线性化极小化方法[13]等。同时, 惯性技术也被应用到 ADMM 算法中来求解凸优化和非凸优化问题。Chen 等人[14]融合邻近 ADMM 算法与惯性邻近点算法的核心思想, 设计出一种惯性邻近 ADMM 算法, 专门用于求解含线性约束的可分离凸优化问题; 该研究同时指出, 算法涉及的矩阵不必局限于正定类型, 半正定矩阵同样适用, 并且完成了对算法收敛性的严格理论证明。Chao 等人[15]在经典交替方向乘子法中引入了惯性步长与邻近正则项, 为线性约束的非凸优化问题提出了一类惯性邻近算法框架。在此基础上, Xu 和 Chao [16]进一步结合了 Bregman 距离与广义松弛形式, 提出了一个更广泛的惯性 Bregman 广义交替方向乘子法, 并在 Kurdyka-Łojasiewicz 性质下建立了其全局收敛性理论。上述工作表明, 惯性技术对提高算法的计算性能有很好的效果。

在处理非光滑问题时, 一类常用的方法是进行光滑近似, 该方法不仅可以保留问题结构, 也能兼容一阶算法, 其中, 文献[17] [18]提出的 Moreau 包络(Moreau envelope)是一种经典的光滑化技术, 它通过二次项来逼近非光滑函数。该方法将复杂的优化问题转化为可高效求解的形式, 并为处理非光滑优化问题奠定了理论基础。此外, Moreau 包络的光滑性还能缓解在求解某些非凸优化问题时的振荡现象。近年来, 文献[19]针对带线性约束的弱凸优化问题, 提出了 Moreau 包络增广拉格朗日方法(MEAL)。此后, 文献[20]针对一类特定的稀疏优化问题, 进一步提出了 Moreau 包络交替方向乘子法(MADMM), 但该方法缺乏系统完备的理论分析。

本文基于上述文献, 提出了一类惯性 Moreau 包络交替方向乘子法(IMADMM)用于求解问题(1)。首先, 在 x 的更新上, 算法融合了 Moreau 包络光滑化技巧, 通过 Moreau 包络逼近非光滑函数, 将原本复杂的子问题转化为更容易处理的形式, 从而提升了计算效率。该技术不仅提供了理论保证, 还有利于减少振荡并提高数值稳定性。在 y 的更新上, 加入了惯性步以提高算法的收敛速率, 同时采用 Armijo 线搜索确定步长, 该技巧在保证迭代方向充分下降的同时, 能够自适应地对步长进行调整, 从而提高了算法的灵活性, 增强了其在实际应用中的鲁棒性与适用性。在收敛性分析上, 构造了特定的辅助函数, 在 KL 框架下完成了收敛性分析。

本文内容结构如下: 第二节介绍相关预备知识。第三节详细阐述了基本假设以及所提出的 IMADMM 算法框架。第四节分析了算法的收敛性。第五节通过数值实验, 以稀疏最小二乘问题为例, 验证所提算法相较于经典 ADMM 及非惯性 MADMM 方法的有效性。最后, 第六节总结全文并展望未来研究方向。

2. 预备知识

定义 2.1. (次微分) [21] 令 $h: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ 为适当下半连续函数。

(1) 对于每个 $x \in \text{dom}(h)$, h 在 x 处的 Fréchet 次微分记为 $\hat{\partial}h(x)$, 它表示如下所有向量 $u \in \mathbb{R}^d$ 所构成的集合,

$$\hat{\partial}h(x) := \left\{ u \in \mathbb{R}^d : \liminf_{y \neq x, y \rightarrow x} \frac{h(y) - h(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0 \right\}.$$

当 $x \notin \text{dom}(h)$ 时, 令 $\hat{\partial}h(x) = \emptyset$ 。

(2) h 在 $x \in \text{dom}(h)$ 处的极限次微分(简称次微分), 记为 $\partial h(x)$, 其定义为

$$\partial h(x) := \left\{ u \in \mathbb{R}^d : \exists x^k \rightarrow x, h(x^k) \rightarrow h(x), u^k \in \hat{\partial}h(x^k) \rightarrow u \right\}.$$

注 2.1. 定义 2.1 表明 $\hat{\partial}h(x)$ 为闭凸集, 同时 $\partial h(x)$ 也是闭集。如果 $h_1: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ 是适当下半连

续的, 并且 $h_2: \mathbb{R}^d \rightarrow \mathbb{R}$ 是连续可微的, 则对于任意的 $x \in \text{dom}(h_1)$ 有 $\partial(h_1 + h_2)(x) = \partial h_1(x) + \nabla h_2(x)$ 。

定义 2.2. 给定函数 $h: \mathbb{R}^d \rightarrow \mathbb{R} \cup +\infty$, 其 Moreau 包络[22]定义为

$$M_{\gamma, h}(x) = \min_z \left\{ h(z) + \frac{1}{2\gamma} \|x - z\|^2 \right\},$$

其中 $\gamma > 0$ 为惩罚参数。对应的邻近算子定义为

$$\text{Prox}_{\gamma, h}(x) = \arg \min_z \left\{ h(z) + \frac{1}{2\gamma} \|x - z\|^2 \right\}.$$

若 h 是 ρ -弱凸的且 $\gamma \in (0, \rho^{-1})$, 则 $\text{Prox}_{\gamma, h}$ 是单调、单值且 Lipschitz 连续的, 并且 $M_{\gamma, h}$ 可微, 其梯度为

$$\nabla M_{\gamma, h}(x) = \frac{1}{\gamma} (x - \text{Prox}_{\gamma, h}(x)) \in \partial h(\text{Prox}_{\gamma, h}(x)).$$

此外, 还有 $M_{\gamma, h}(\text{Prox}_{\gamma, h}(x)) \leq h(x)$, 这表明 Moreau 包络可作为 h 的一个光滑下近似。更多结果可参考文献[22]。

定义 2.3. [23] 设 $h: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ 是一个适当下半连续函数。若存在 $\eta \in (0, +\infty]$, $\bar{x} \in \text{dom}(\partial h)$ 的一个邻域 U , 以及一个连续凹函数 $\varphi: [0, \eta] \rightarrow \mathbb{R}_+$ 使得

- (1) $\varphi(0) = 0$, 且 φ 在 $(0, \eta)$ 上连续可微, 对任意 $s \in (0, \eta)$ 满足 $\varphi'(s) > 0$;
- (2) 对所有 $x \in U \cap [x \in \mathbb{R}^d : h(\bar{x}) < h(x) < h(\bar{x}) + \eta]$, 都有如下 Kurdyka-Łojasiewicz 不等式成立

$$\varphi'(h(x) - h(\bar{x})) \text{dist}(0, \partial h(x)) \geq 1,$$

则称函数 h 在 $\bar{x} \in \text{dom}(\partial h)$ 处满足 Kurdyka-Łojasiewicz (KŁ) 性质。

注 2.2. (1) 在 $\text{dom}(\partial h)$ 的每一点均满足 KŁ 性质的适当下半连续函数称为 KŁ 函数。

(2) 为方便起见, 将满足定义 2.3 (1) 中的函数类记为 Φ_η 。满足定义 2.3 (1) 和 (2) 中的函数称为 h 的去奇化函数。

(3) 半代数函数提供了非常丰富的一类满足 KŁ 性质的函数, 其去奇化函数可取如下形式的函数

$$\varphi(s) = as^{1-\theta},$$

其中 $a > 0$, $\theta \in [0, 1)$, θ 称为 KŁ 指数[23]。

下述引理给出了重要的一致 KŁ 性质。

引理 2.1. (一致 KŁ 性质) [24] 设 Ω 是一个紧集, 且 $h: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ 是一个适当下半连续函数。假设 h 在 Ω 上为常数, 且在 Ω 的每一点处均满足 KŁ 性质。则存在 $\varepsilon > 0$, $\eta > 0$ 以及 $\varphi \in \Phi_\eta$, 使得对于 $\forall \bar{x} \in \Omega$ 以及所有属于下述交集集中的 x :

$$\{x \in \mathbb{R}^d : \text{dist}(x, \Omega) < \varepsilon\} \cap [h(\bar{x}) < h(x) < h(\bar{x}) + \eta],$$

都有

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

3. 惯性 Moreau 包络 ADMM

在介绍算法之前, 首先给出如下几条基本假设。

假设 3.1. (1) $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 是适当下半连续弱凸函数, 系数为 $\rho > 0$ 。

(2) $g: \mathbb{R}^m \rightarrow \mathbb{R}$ 为连续可微函数, 并且具有全局 Lipschitz 光滑性, 系数为 L_g 。

(3) 矩阵 \mathbf{B} 是列满秩的。

回顾问题(1)的增广拉格朗日函数,

$$\mathcal{L}_\beta(x, y, \lambda) = f(x) + g(y) + \lambda^T (\mathbf{A}x + \mathbf{B}y - b) + \frac{\beta}{2} \|\mathbf{A}x + \mathbf{B}y - b\|^2.$$

定义 $\mathcal{L}_\beta(x, y, \lambda)$ 的 Moreau 包络如下,

$$\phi_\beta(x, y, \lambda) = \min_{z \in \mathbb{R}^n} \left\{ \mathcal{L}_\beta(z, y, \lambda) + \frac{1}{2\gamma} \|z - x\|^2 \right\},$$

其中 $\gamma > 0$ 为邻近参数。下面介绍用于求解问题(1)的惯性 Moreau 包络交替方向乘法。

算法 1

a) 选择参数 $\gamma \in (0, \rho^{-1})$, $\alpha \in (\gamma, 2\gamma)$, $\sigma \in (0, 1)$, $\delta \in (0, 1)$, $\beta > 0$ 。初始化 $(z^0, x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n_2} \times \mathbb{R}^m$ 。

b) 对于每个 $k = 0, 1, 2, \dots$, 通过以下迭代步骤产生 z^{k+1} , x^{k+1} ,

$$z^{k+1} = \arg \min_{z \in \mathbb{R}^n} \left\{ \mathcal{L}_\beta(z, y^k, \lambda^k) + \frac{1}{2\gamma} \|z - x^k\|^2 \right\}, \quad (3)$$

$$x^{k+1} = x^k - \alpha \nabla_x \phi_\beta(x^k, y^k, \lambda^k) = x^k - \frac{\alpha}{\gamma} (x^k - z^{k+1}). \quad (4)$$

c) 产生搜索方向: $d_y^k = \hat{y}^k - y^k$, 其中,

$$\hat{y}^k = \arg \min_{y \in \mathbb{R}^{n_2}} \left\{ \mathcal{L}_\beta(z^{k+1}, y, \lambda^k) \right\}. \quad (5)$$

d) (Armijo 线搜索) 在序列 $\{s = \sigma^i, i = 0, 1, 2, \dots\}$ 中选择使得以下不等式成立的最小的 i , 作为迭代步长 $s_k \in (0, 1)$,

$$\mathcal{L}_\beta(z^{k+1}, y^k, \lambda^k) - \mathcal{L}_\beta(z^{k+1}, y^k + s(\hat{y}^k - y^k), \lambda^k) \geq \delta s \|y^k - \hat{y}^k\|^2. \quad (6)$$

e) 更新

$$y^{k+1} = y^k + s_k (\hat{y}^k - y^k), \quad (7)$$

$$\lambda^{k+1} = \lambda^k + \beta (\mathbf{A}z^{k+1} + \mathbf{B}y^{k+1} - b). \quad (8)$$

f) 令 $k := k + 1$, 返回步骤 b)。

注 3.1. (1) 本文算法依赖于一个核心前提, 函数 f 的邻近算子容易求解, 如 SCAD、MCP、 ℓ_p ($0 < p < 1$) 等常见非凸正则项。对于闭式解不存在的情形, 步骤 **b** 本身是一个光滑强凸子问题, 可借助内层数值优化(牛顿法、坐标下降等)高精度求解。本文的收敛性分析建立在子问题被精确求解的假设之上, 这为算法的理论合理性提供了保障。

(2) 虽然引入 Armijo 线搜索会增加每次迭代的计算成本, 但在非凸非光滑优化中, 合适的步长对保证收敛至关重要。线搜索通过确保充分下降, 显著减少了算法收敛所需的总迭代次数, 从而在实际中往往带来整体计算效率的提升, 在第 5 节的数值实验中也验证了该策略的有效性。

4. 收敛性分析

本节分析了算法的收敛性。假设 $\{\omega^k = (z^k, x^k, y^k, \lambda^k)\}$ 为 IMADMM 算法生成的序列, 记

$$\widetilde{\mathcal{L}}_\beta(z, x, y, \lambda) = \mathcal{L}_\beta(z, y, \lambda) + \frac{1}{2\gamma} \|z - x\|^2. \quad (9)$$

下面的引理说明了 $\widetilde{\mathcal{L}}_\beta$ 具有下降性, 以及所生成序列是平方可和的。

引理 4.1. 设假设 3.1 成立, 且参数按照算法 1 所述设置。设 $\{\omega^k\}$ 为算法 IMADMM 生成的序列。则有以下结论成立

(1) 对所有 $k \in \mathbb{N}$, 有

$$\widetilde{\mathcal{L}}_\beta(\omega^k) - \widetilde{\mathcal{L}}_\beta(\omega^{k+1}) \geq \tau_1 \left(\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2 + \|y^k - \hat{y}^k\|^2 + \|y^{k+1} - y^k\|^2 + \|\hat{y}^{k+1} - \hat{y}^k\|^2 \right), \quad (10)$$

其中,

$$\tau_1 = \min \left\{ \frac{\gamma^{-1} - \rho}{2}, \frac{2\gamma - \alpha}{2\gamma\alpha}, \delta s, \frac{2\beta \|B^T B\|^2}{\sigma_{\min}(B^T B)}, \frac{2(\beta \|B^T B\| + L_g)^2}{\beta \sigma_{\min}(B^T B)} \right\} > 0.$$

(2) 如果序列 $\{\omega^k\}$ 为有界序列, 则有

$$\sum_{k=1}^{+\infty} \|\omega^{k+1} - \omega^k\|^2 < +\infty, \sum_{k=1}^{+\infty} \|y^k - \hat{y}^k\|^2 < +\infty.$$

(3) 当 $k \rightarrow +\infty$ 时, $\|z^k - x^k\| \rightarrow 0$ 。

因此, 序列 $\{\mathcal{L}_\beta(z^k, y^k, \lambda^k)\}$ 是收敛的, 并且

$$\lim_{k \rightarrow +\infty} \mathcal{L}_\beta(x^k, y^k, \lambda^k) = \lim_{k \rightarrow +\infty} \widetilde{\mathcal{L}}_\beta(\omega^k).$$

证明 (1) 由于迭代(3)中的 z^{k+1} 通过最小化强凸函数 $\widetilde{\mathcal{L}}_\beta(z, x^k, y^k, \lambda^k)$ 得到, 其系数至少为 $(\gamma^{-1} - \rho)$, 从而有

$$\widetilde{\mathcal{L}}_\beta(z^k, x^k, y^k, \lambda^k) - \widetilde{\mathcal{L}}_\beta(z^{k+1}, x^k, y^k, \lambda^k) \geq \frac{\gamma^{-1} - \rho}{2} \|z^{k+1} - z^k\|^2. \quad (11)$$

由(4)中 x^{k+1} 的更新可知, 有如下等式

$$2z^{k+1} - x^k - x^{k+1} = \left(\frac{2\gamma}{\alpha} - 1 \right) (x^{k+1} - x^k),$$

成立。则可以得到

$$\begin{aligned} & \widetilde{\mathcal{L}}_\beta(z^{k+1}, x^k, y^k, \lambda^k) - \widetilde{\mathcal{L}}_\beta(z^{k+1}, x^{k+1}, y^k, \lambda^k) \\ &= \frac{1}{2\gamma} \langle x^{k+1} - x^k, 2z^{k+1} - x^k - x^{k+1} \rangle \\ &= \left(\frac{1}{\alpha} - \frac{1}{2\gamma} \right) \|x^{k+1} - x^k\|^2. \end{aligned} \quad (12)$$

算法中线搜索步骤(6)可得

$$\widetilde{\mathcal{L}}_\beta(z^{k+1}, x^{k+1}, y^k, \lambda^k) - \widetilde{\mathcal{L}}_\beta(z^{k+1}, x^{k+1}, y^{k+1}, \lambda^k) \geq \delta s \|y^k - \hat{y}^k\|^2. \quad (13)$$

又由算法中(5)式的一阶最优性条件以及 λ 的更新,

$$\begin{aligned} 0 &= \nabla g(\hat{y}^k) + B^T \lambda^k + \beta B^T (Az^{k+1} + By^k - b) \\ &= \nabla g(\hat{y}^k) + B^T \lambda^{k+1} - \beta B^T B (y^{k+1} - \hat{y}^k), \end{aligned} \quad (14)$$

从而有

$$\mathbf{B}^T \lambda^{k+1} = \beta \mathbf{B}^T \mathbf{B} (y^{k+1} - \hat{y}^k) - \nabla g(\hat{y}^k), \quad (15)$$

进一步可以推出

$$\mathbf{B}^T (\lambda^{k+1} - \lambda^k) = \beta \mathbf{B}^T \mathbf{B} (y^{k+1} - y^k) + \beta \mathbf{B}^T \mathbf{B} (\hat{y}^{k-1} - \hat{y}^k) + \nabla g(\hat{y}^{k-1}) - \nabla g(\hat{y}^k).$$

因此, 可得

$$\begin{aligned} \|\lambda^{k+1} - \lambda^k\| &\leq \frac{1}{\sqrt{\sigma_{\min}(\mathbf{B}^T \mathbf{B})}} \|\mathbf{B}^T (\lambda^{k+1} - \lambda^k)\| \\ &\leq \frac{1}{\sqrt{\sigma_{\min}(\mathbf{B}^T \mathbf{B})}} \left(\beta \|\mathbf{B}^T \mathbf{B}\| \|y^{k+1} - y^k\| + (\beta \|\mathbf{B}^T \mathbf{B}\| + L_g) \|\hat{y}^{k-1} - \hat{y}^k\| \right), \end{aligned} \quad (16)$$

以及

$$\|\lambda^{k+1} - \lambda^k\|^2 \leq \frac{2\beta^2 \|\mathbf{B}^T \mathbf{B}\|^2}{\sigma_{\min}(\mathbf{B}^T \mathbf{B})} \|y^{k+1} - y^k\|^2 + \frac{2(\beta \|\mathbf{B}^T \mathbf{B}\| + L_g)^2}{\sigma_{\min}(\mathbf{B}^T \mathbf{B})} \|\hat{y}^{k-1} - \hat{y}^k\|^2, \quad (17)$$

其中 $\sigma_{\min}(\mathbf{B}^T \mathbf{B})$ 为 $\mathbf{B}^T \mathbf{B}$ 矩阵的最小正特征值。从而根据(17), 可以得到

$$\begin{aligned} &\widetilde{\mathcal{L}}_{\beta}(z^{k+1}, x^{k+1}, y^{k+1}, \lambda^k) - \widetilde{\mathcal{L}}_{\beta}(z^{k+1}, x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ &= \langle \lambda^k - \lambda^{k+1}, \mathbf{A}z^{k+1} + \mathbf{B}y^{k+1} - b \rangle \\ &= -\frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|^2 \\ &\geq \frac{2\beta \|\mathbf{B}^T \mathbf{B}\|^2}{\sigma_{\min}(\mathbf{B}^T \mathbf{B})} \|y^{k+1} - y^k\|^2 + \frac{2(\beta \|\mathbf{B}^T \mathbf{B}\| + L_g)^2}{\beta \sigma_{\min}(\mathbf{B}^T \mathbf{B})} \|\hat{y}^{k-1} - \hat{y}^k\|^2. \end{aligned} \quad (18)$$

结合(11)~(13)以及(18), 即可证得结果

$$\begin{aligned} &\widetilde{\mathcal{L}}_{\beta}(z^k, x^k, y^k, \lambda^k) - \widetilde{\mathcal{L}}_{\beta}(z^{k+1}, x^{k+1}, y^{k+1}, \lambda^{k+1}) \\ &\geq \frac{\gamma^{-1} - \rho}{2} \|z^{k+1} - z^k\|^2 + \left(\frac{1}{\alpha} - \frac{1}{2\gamma} \right) \|x^{k+1} - x^k\|^2 + \delta_S \|y^k - \hat{y}^k\|^2 \\ &\quad + \frac{2\beta \|\mathbf{B}^T \mathbf{B}\|^2}{\sigma_{\min}(\mathbf{B}^T \mathbf{B})} \|y^{k+1} - y^k\|^2 + \frac{2(\beta \|\mathbf{B}^T \mathbf{B}\| + L_g)^2}{\beta \sigma_{\min}(\mathbf{B}^T \mathbf{B})} \|\hat{y}^{k-1} - \hat{y}^k\|^2. \end{aligned}$$

证明完毕。

(2) 由于序列 $\{\omega^k\}$ 有界, 从而它至少存在一个极限点。令 $\omega^* = (z^*, x^*, y^*, \lambda^*)$ 为序列 $\{\omega^k\}$ 的一个极限点, 并令 $\{\omega^{k_j}\}$ 为收敛到该极限点的一个子列, 即 $\omega^{k_j} \rightarrow \omega^*$ 。又由函数 f 是下半连续的且 g 是连续的, 则函数 $\mathcal{L}_{\beta}(\cdot)$ 也是下半连续的, 因此

$$\mathcal{L}_{\beta}(z^*, y^*, \lambda^*) \leq \liminf_{j \rightarrow +\infty} \mathcal{L}_{\beta}(z^{k_j}, y^{k_j}, \lambda^{k_j}).$$

从而 $\mathcal{L}_{\beta}(z^{k_j}, y^{k_j}, \lambda^{k_j})$ 有下界。同时 $\{\widetilde{\mathcal{L}}_{\beta}(\omega^{k_j})\}$ 也具有下界, 再结合 $\{\widetilde{\mathcal{L}}_{\beta}(\omega^k)\}$ 单调非增这一事实, 意味着 $\{\widetilde{\mathcal{L}}_{\beta}(\omega^{k_j})\}$ 收敛。因此, $\{\widetilde{\mathcal{L}}_{\beta}(\omega^k)\}$ 收敛。设 $\mathcal{L}^* = \lim_{k \rightarrow +\infty} \widetilde{\mathcal{L}}_{\beta}(\omega^k)$ 。进一步, 不等式(10)表明

$$\begin{aligned} & \tau_1 \left(\sum_{k=1}^n \|z^{k+1} - z^k\|^2 + \sum_{k=1}^n \|x^{k+1} - x^k\|^2 + \sum_{k=1}^n \|y^k - \hat{y}^k\|^2 + \sum_{k=1}^n \|y^{k+1} - y^k\|^2 + \sum_{k=1}^n \|\hat{y}^{k-1} - \hat{y}^k\|^2 \right) \\ & \leq \widetilde{\mathcal{L}}_\beta(\omega^1) - \widetilde{\mathcal{L}}_\beta(\omega^{n+1}) \leq \widetilde{\mathcal{L}}_\beta(\omega^1) - \mathcal{L}^* \leq +\infty. \end{aligned}$$

由于 $\tau_1 > 0$ ，则有

$$\begin{aligned} \sum_{k=1}^{+\infty} \|z^{k+1} - z^k\|^2 &< +\infty, \sum_{k=1}^{+\infty} \|x^{k+1} - x^k\|^2 < +\infty, \sum_{k=1}^{+\infty} \|y^{k+1} - y^k\|^2 < +\infty, \\ \sum_{k=1}^{+\infty} \|y^k - \hat{y}^k\|^2 &< +\infty, \sum_{k=1}^{+\infty} \|\hat{y}^{k-1} - \hat{y}^k\|^2 < +\infty. \end{aligned}$$

再回顾(8)式可得

$$\sum_{k=1}^{+\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty.$$

综上，可以得到 $\sum_{k=1}^{+\infty} \|\omega^{k+1} - \omega^k\|^2 < +\infty$ 以及 $\sum_{k=1}^{+\infty} \|y^k - \hat{y}^k\|^2 < +\infty$ 。

(3) 由(4)式中 x^{k+1} 的更新可得

$$z^{k+1} - x^{k+1} = \left(1 - \frac{\gamma}{\alpha}\right) (x^k - x^{k+1}). \tag{19}$$

再结合引理 4.1. (2)，显然有 $\|z^k - x^k\| \rightarrow 0$ 。事实上，序列 $\{z^k\}$ 和 $\{x^k\}$ 的极限点是相同的。根据(9)式中 $\widetilde{\mathcal{L}}_\beta$ 的定义可知， $\{\mathcal{L}_\beta(z^k, y^k, \lambda^k)\}$ 是收敛的，并且

$$\lim_{k \rightarrow +\infty} \mathcal{L}_\beta(x^k, y^k, \lambda^k) = \lim_{k \rightarrow +\infty} \widetilde{\mathcal{L}}_\beta(\omega^k).$$

证明完毕。

下面的引理说明了辅助函数 $\widetilde{\mathcal{L}}_\beta$ 次梯度的有界性。

引理 3.2. 设假设 3.1 成立，且参数按照算法 1 所述设置。设 $\{\omega^k\}$ 为由 IMADMM 算法生成的有界序列。定义 $\kappa^{k+1} := (\kappa_z^{k+1}, \kappa_x^{k+1}, \kappa_y^{k+1}, \kappa_\lambda^{k+1})$ ，其中

$$\begin{cases} \kappa_z^{k+1} = \beta A^T (B y^{k+1} - B y^k) + \frac{1}{\gamma} (x^k - x^{k+1}) + A^T (\lambda^{k+1} - \lambda^k), \\ \kappa_x^{k+1} = \left(\frac{1}{\gamma} - \frac{1}{\alpha}\right) (x^{k+1} - x^k), \\ \kappa_y^{k+1} = \nabla g(y^{k+1}) - \nabla g(\hat{y}^k) + B^T (\lambda^{k+1} - \lambda^k) + \beta B^T B (y^{k+1} - \hat{y}^k), \\ \kappa_\lambda^{k+1} = \frac{1}{\beta} (\lambda^{k+1} - \lambda^k), \end{cases}$$

则有

$$\kappa^{k+1} \in \partial \widetilde{\mathcal{L}}_\beta(\omega^{k+1}).$$

并且，存在 $\tau_2 > 0$ 使得

$$\text{dist}\left(0, \partial \widetilde{\mathcal{L}}_\beta(\omega^{k+1})\right) \leq \tau_2 \left(\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| + \|y^{k+1} - \hat{y}^k\| + \|\hat{y}^k - \hat{y}^{k-1}\| \right).$$

证明 由函数 $\widetilde{\mathcal{L}}_\beta(\cdot)$ 的定义，

$$\begin{cases} \partial f(z^{k+1}) + A^T \lambda^{k+1} + \beta A^T (Az^{k+1} + By^{k+1} - b) + \frac{1}{\gamma}(z^{k+1} - x^{k+1}) \in \partial_z \widetilde{\mathcal{L}}_\beta(\omega^{k+1}), \\ -\frac{1}{\gamma}(z^{k+1} - x^{k+1}) = \partial_x \widetilde{\mathcal{L}}_\beta(\omega^{k+1}), \\ \nabla g(y^{k+1}) + B^T \lambda^{k+1} + \beta B^T (Az^{k+1} + By^{k+1} - b) = \partial_y \widetilde{\mathcal{L}}_\beta(\omega^{k+1}), \\ Az^{k+1} + By^{k+1} - b = \partial_\lambda \widetilde{\mathcal{L}}_\beta(\omega^{k+1}). \end{cases}$$

又由(3)式的一阶最优性条件得出

$$0 \in \partial f(z^{k+1}) + A^T \lambda^k + \beta A^T (Az^{k+1} + By^k - b) + \frac{1}{\gamma}(z^{k+1} - x^k), \tag{20}$$

因此有 $\kappa_z^{k+1} \in \partial_z \widetilde{\mathcal{L}}_\beta(\omega^{k+1})$ 。对于变量 x ，结合(19)式可得 $\kappa_x^{k+1} \in \partial_x \widetilde{\mathcal{L}}_\beta(\omega^{k+1})$ 。回顾(14)式，整理后得

$$\beta B^T (Az^{k+1} + By^{k+1} - b) = -\nabla g(\hat{y}^k) - B^T \lambda^k + \beta B^T B(y^{k+1} - \hat{y}^k),$$

将上式代入 $\partial_y \widetilde{\mathcal{L}}_\beta(\omega^{k+1})$ 中，从而有 $\kappa_y^{k+1} \in \partial_y \widetilde{\mathcal{L}}_\beta(\omega^{k+1})$ 。再结合算法中 λ 的更新，显然有 $\kappa_\lambda^{k+1} \in \partial_\lambda \widetilde{\mathcal{L}}_\beta(\omega^{k+1})$ 。进一步，由于序列 $\{\omega^k\}$ 有界并且 ∇g 在 \mathbb{R}^{n_2} 的有界子集上 Lipschitz 连续，

$$\begin{aligned} \|\kappa^{k+1}\| &\leq \|\kappa_z^{k+1}\| + \|\kappa_x^{k+1}\| + \|\kappa_y^{k+1}\| + \|\kappa_\lambda^{k+1}\| \\ &\leq \left(\frac{2}{\gamma} - \frac{1}{\alpha}\right) \|x^{k+1} - x^k\| + \left(\|A\| + \|B\| + \frac{1}{\beta}\right) \|\lambda^{k+1} - \lambda^k\| \\ &\quad + \beta \|A^T B\| \|y^{k+1} - y^k\| + (L_g + \beta \|B^T B\|) \|y^{k+1} - \hat{y}^k\|. \end{aligned}$$

将(16)式代入上式，则有

$$\begin{aligned} \text{dist}(0, \partial \widetilde{\mathcal{L}}_\beta(\omega^{k+1})) &\leq \|\kappa^{k+1}\| \\ &\leq \left(\frac{2}{\gamma} - \frac{1}{\alpha}\right) \|x^{k+1} - x^k\| + r_1 \|y^{k+1} - y^k\| + r_2 \|y^{k+1} - \hat{y}^k\| + r_3 \|\hat{y}^k - \hat{y}^{k-1}\| \\ &\leq \tau_2 (\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| + \|y^{k+1} - \hat{y}^k\| + \|\hat{y}^k - \hat{y}^{k-1}\|), \end{aligned}$$

这里 $\tau_2 = \max\left\{\frac{2}{\gamma} - \frac{1}{\alpha}, r_1, r_2, r_3\right\} > 0$ ，其中，

$$\begin{aligned} r_1 &= \beta \|A^T B\| + \left(\|A\| + \|B\| + \frac{1}{\beta}\right) \frac{\beta \|B^T B\|}{\sqrt{\sigma_{\min}(B^T B)}}, \\ r_2 &= L_g + \beta \|B^T B\|, \\ r_3 &= \left(\|A\| + \|B\| + \frac{1}{\beta}\right) \frac{r_2}{\sqrt{\sigma_{\min}(B^T B)}}. \end{aligned}$$

证明完毕。

记 $v^k = (x^k, y^k, \lambda^k)$ 。令 $v^* = (x^*, y^*, \lambda^*)$ 为序列 $\{v^k\}$ 的一个极限点。由引理 4.1 可知， $\omega^* = (x^*, x^*, y^*, \lambda^*)$ 为 $\{\omega^k\}$ 极限点。分别记 $\{v^k\}$ ， $\{\omega^k\}$ 的极限点集为 $S^*(\bar{v})$ 和 $S^*(\bar{\omega})$ 。

引理 3.3. 设假设 3.1 成立，且参数按照算法 1 所述设置。设 $\{\omega^k\}$ 为 IMADMM 算法生成的有界序列。则有以下结论成立，

- (1) $S^*(\bar{\omega})$ 是非空紧集，并且 $\lim_{k \rightarrow +\infty} \text{dist}(\omega^k, S^*(\bar{\omega})) = 0$ 。

- (2) $S^*(\bar{v}) \subset \text{crit}(\mathcal{L}_\beta)$ 且 $S^*(\bar{\omega}) \subset \text{crit}(\widetilde{\mathcal{L}}_\beta)$ 。
- (3) 函数 $\widetilde{\mathcal{L}}_\beta$ 在 $S^*(\bar{\omega})$ 上为常值。

证明 (1) 由定义显然得证。详细证明见参考文献[24]。

(2) 令 $\omega^* \in S^*(\bar{\omega})$ ，则存在 $\{\omega^k\}$ 的子列 $\{\omega^{k_j}\}$ 收敛到 $\omega^* = (x^*, x^*, y^*, \lambda^*)$ 。由于 $\|\omega^{k+1} - \omega^k\| \rightarrow 0$ ，容易得到 $\{\omega^{k_j+1}\}$ 也收敛到 ω^* 。由迭代(3)式可得，

$$\mathcal{L}_\beta(z^{k+1}, y^k, \lambda^k) + \frac{1}{2\gamma} \|z^{k+1} - x^k\|^2 \leq \mathcal{L}_\beta(x^*, y^k, \lambda^k) + \frac{1}{2\gamma} \|x^* - x^k\|^2.$$

因此，

$$\begin{aligned} f(z^{k+1}) &\leq f(x^*) + (\lambda^k)^\top (Ax^* - Az^{k+1}) + \frac{\beta}{2} \|Ax^* + By^k - b\|^2 \\ &\quad - \frac{\beta}{2} \|Az^{k+1} + By^k - b\|^2 + \frac{1}{2\gamma} \|x^* - x^k\|^2 - \frac{1}{2\gamma} \|z^{k+1} - x^k\|^2 \\ &= f(x^*) + (\lambda^k)^\top (Ax^* - Az^{k+1}) + \frac{\beta}{2} \|Az^{k+1} + By^k - b + (Ax^* - Az^{k+1})\|^2 \\ &\quad - \frac{\beta}{2} \|Az^{k+1} + By^k - b\|^2 + \frac{1}{2\gamma} \|z^{k+1} - x^k + (x^* - z^{k+1})\|^2 - \frac{1}{2\gamma} \|z^{k+1} - x^k\|^2 \\ &\leq f(x^*) + (\lambda^k)^\top A(x^* - z^{k+1}) + \frac{\beta}{2} \|Az^{k+1} + By^{k+1} - b + (By^k - By^{k+1})\|^2 \\ &\quad + \frac{1}{\gamma} \|x^* - z^{k+1}\|^2 + \frac{1}{2\gamma} \|z^{k+1} - x^k\|^2 + \beta \|Ax^* - Az^{k+1}\|^2 \\ &\leq f(x^*) + (\lambda^k)^\top A(x^* - z^{k+1}) + \beta \|Ax^* - Az^{k+1}\|^2 + \frac{1}{\beta} \|\lambda^{k+1} - \lambda^k\|^2 \\ &\quad + \beta \|By^{k+1} - By^k\|^2 + \frac{1}{\gamma} \|x^* - z^{k+1}\|^2 + \frac{\gamma}{2\alpha^2} \|x^{k+1} - x^k\|^2, \end{aligned} \tag{21}$$

其中最后一个不等式是由迭代(4)式以及(8)式得到的。令不等式(21)式中 $k = k_j$ 有，

$$\begin{aligned} \limsup_{j \rightarrow +\infty} f(z^{k_j+1}) &\leq \limsup_{j \rightarrow +\infty} \left(f(x^*) + (\lambda^{k_j})^\top A(x^* - z^{k_j+1}) + \beta \|Ax^* - Az^{k_j+1}\|^2 \right. \\ &\quad \left. + \frac{1}{\beta} \|\lambda^{k_j+1} - \lambda^{k_j}\|^2 + \beta \|By^{k_j+1} - By^{k_j}\|^2 + \frac{1}{\gamma} \|x^* - z^{k_j+1}\|^2 \right. \\ &\quad \left. + \frac{\gamma}{2\alpha^2} \|x^{k_j+1} - x^{k_j}\|^2 \right) \\ &= f(x^*). \end{aligned} \tag{22}$$

由 f 的下半连续性可得，

$$\liminf_{j \rightarrow +\infty} f(z^{k_j+1}) \geq f(x^*). \tag{23}$$

再结合(22)和(23)，

$$\lim_{j \rightarrow +\infty} f(z^{k_j+1}) = f(x^*). \tag{24}$$

回顾(20)式，再结合 ∂f 的闭性，令 $k = k_j$ 在其子序列中取极限有 $0 \in \partial f(x^*) + A^\top \lambda^*$ 。又在引理 4.1 (2)中，当 $k \rightarrow +\infty$ 时有 $\|y^k - \hat{y}^k\| \rightarrow 0$ 。考虑(14)式，再结合 ∇g 的连续性可以得出 $0 = \nabla g(y^*) + B^\top \lambda^*$ 。最后再根据算法中 λ 的更新，沿子序列取极限。综上分析得出结论，

$$\begin{cases} 0 \in \partial f(x^*) + A^T \lambda^*, \\ 0 = \nabla g(y^*) + B^T \lambda^*, \\ Ax^* + By^* - b = 0. \end{cases}$$

因此,

$$(x^*, y^*, \lambda^*) \in \text{crit}(\mathcal{L}_\beta).$$

结合(24)式和函数 g 的连续性, 可得

$$\lim_{j \rightarrow +\infty} \widetilde{\mathcal{L}}_\beta(\omega^{k_j+1}) = \widetilde{\mathcal{L}}_\beta(\omega^*). \tag{25}$$

此外, 由引理 4.1 和引理 4.2 可知, $\kappa^k \in \partial \widetilde{\mathcal{L}}_\beta(\omega^k)$ 且当 $k \rightarrow +\infty$ 时 $\kappa^k \rightarrow 0$ 。类似地, 基于 $\partial \widetilde{\mathcal{L}}_\beta$ 的闭性, 有 $0 \in \partial \widetilde{\mathcal{L}}_\beta(\omega^*)$ 。因此, ω^* 是 $\widetilde{\mathcal{L}}_\beta$ 的一个临界点。

(3) 对于任意点 $\omega^* \in S^*(\bar{\omega})$, 则存在 $\{\omega^k\}$ 的一个子序列 $\{\omega^{k_j}\}$ 收敛到 $\omega^* = (x^*, x^*, y^*, \lambda^*)$ 。结合(25)式和 $\{\widetilde{\mathcal{L}}_\beta\}$ 的单调非增性, 有

$$\lim_{k \rightarrow +\infty} \widetilde{\mathcal{L}}_\beta(\omega^k) = \widetilde{\mathcal{L}}_\beta(\omega^*).$$

因此, $\widetilde{\mathcal{L}}_\beta(\cdot)$ 在 $S^*(\bar{\omega})$ 上是恒定的。证明完毕。

基于上述引理分析, 现给出本文证明的主要结果。

定理 4.1. 设假设 3.1 成立, 且参数按照算法 1 所述设置。设 $\{\omega^k = (z^k, x^k, y^k, \lambda^k)\}$ 为 IMADMM 算法生成的有界序列。假设函数 f 和 g 均为半代数函数, 则序列 $\{\omega^k\}$ 为柯西列。并且序列 $\{(x^k, y^k, \lambda^k)\}$ 收敛到 \mathcal{L}_β 的临界点。

Proof 根据引理 4.3, 对于任意的 $\omega^* \in S^*(\bar{\omega})$ 都有 $\lim_{k \rightarrow +\infty} \widetilde{\mathcal{L}}_\beta(\omega^k) = \widetilde{\mathcal{L}}_\beta(\omega^*)$ 成立。因此, 我们考虑以下两种情况。

情况 1: 如果存在整数 \tilde{k} 使得 $\widetilde{\mathcal{L}}_\beta(\omega^{\tilde{k}}) = \widetilde{\mathcal{L}}_\beta(\omega^*)$, 则对任意 $k > \tilde{k}$, 根据(10)式可得

$$\begin{aligned} & \tau_1 (\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2 + \|y^k - \hat{y}^k\|^2 + \|y^{k+1} - y^k\|^2 + \|\hat{y}^{k-1} - \hat{y}^k\|^2) \\ & \leq \widetilde{\mathcal{L}}_\beta(\omega^k) - \widetilde{\mathcal{L}}_\beta(\omega^{k+1}) \leq \widetilde{\mathcal{L}}_\beta(\omega^{\tilde{k}}) - \widetilde{\mathcal{L}}_\beta(\omega^*) = 0. \end{aligned}$$

因此, 对于任意 $k > \tilde{k}$ 有, $z^{k+1} = z^k$, $x^{k+1} = x^k$ 以及 $y^{k+1} = y^k$ 。证明完毕。

情况 2: 对于所有的 $k \geq 1$, 现假设 $\widetilde{\mathcal{L}}_\beta(\omega^k) > \widetilde{\mathcal{L}}_\beta(\omega^*)$ 成立。由于 $\lim_{k \rightarrow +\infty} \widetilde{\mathcal{L}}_\beta(\omega^k) = \widetilde{\mathcal{L}}_\beta(\omega^*)$, 从而有对于任意的 $\epsilon_1 > 0$, 存在 $K_1 \in \mathbb{N}$ 使得 $\widetilde{\mathcal{L}}_\beta(\omega^*) < \widetilde{\mathcal{L}}_\beta(\omega^k) < \widetilde{\mathcal{L}}_\beta(\omega^*) + \epsilon_1$ 对所有的 $k > K_1$ 成立。又从引理 4.3 可得, $\lim_{k \rightarrow +\infty} \text{dist}(\omega^k, S^*(\bar{\omega})) = 0$ 。因此, 对任意 $\epsilon_2 > 0$, 存在 $K_2 \in \mathbb{N}$ 使得 $\text{dist}(\omega^k, S^*(\bar{\omega})) < \epsilon_2$ 对所有的 $k > K_2$ 成立。从而, 对于所有 $k > K = \max\{K_1, K_2\}$, 应用引理 2.1 中所述的一致 KL 性质得到

$$\varphi'(\widetilde{\mathcal{L}}_\beta(\omega^k) - \widetilde{\mathcal{L}}_\beta(\omega^*)) \text{dist}(0, \partial \widetilde{\mathcal{L}}_\beta(\omega^k)) \geq 1. \tag{26}$$

利用 φ 的凹性, 可得

$$\varphi(\widetilde{\mathcal{L}}_\beta(\omega^k) - \widetilde{\mathcal{L}}_\beta(\omega^*)) - \varphi(\widetilde{\mathcal{L}}_\beta(\omega^{k+1}) - \widetilde{\mathcal{L}}_\beta(\omega^*)) \geq \varphi'(\widetilde{\mathcal{L}}_\beta(\omega^k) - \widetilde{\mathcal{L}}_\beta(\omega^*)) (\widetilde{\mathcal{L}}_\beta(\omega^k) - \widetilde{\mathcal{L}}_\beta(\omega^{k+1})).$$

方便起见, 定义 $\Delta_{p,q} = \varphi(\widetilde{\mathcal{L}}_\beta(\omega^p) - \widetilde{\mathcal{L}}_\beta(\omega^*)) - \varphi(\widetilde{\mathcal{L}}_\beta(\omega^q) - \widetilde{\mathcal{L}}_\beta(\omega^*))$ 。又由引理 4.1, 引理 4.2 以及(26)式,

$$\Delta_{k,k+1} \geq \frac{\tau_1 \left(\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2 + \|y^k - \hat{y}^k\|^2 + \|y^{k+1} - y^k\|^2 + \|\hat{y}^{k-1} - \hat{y}^k\|^2 \right)}{\tau_2 \left(\|x^k - x^{k-1}\| + \|y^k - y^{k-1}\| + \|y^k - \hat{y}^{k-1}\| + \|\hat{y}^{k-1} - \hat{y}^{k-2}\| \right)},$$

整理后得,

$$\begin{aligned} & \|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2 + \|y^k - \hat{y}^k\|^2 + \|y^{k+1} - y^k\|^2 + \|\hat{y}^{k-1} - \hat{y}^k\|^2 \\ & \leq \frac{\tau_2}{\tau_1} \left(\|x^k - x^{k-1}\| + \|y^k - y^{k-1}\| + \|y^k - \hat{y}^{k-1}\| + \|\hat{y}^{k-1} - \hat{y}^{k-2}\| \right) \Delta_{k,k+1}. \end{aligned}$$

因此,

$$\begin{aligned} & \|z^{k+1} - z^k\| + \|x^{k+1} - x^k\| + \|y^k - \hat{y}^k\| + \|y^{k+1} - y^k\| + \|\hat{y}^{k-1} - \hat{y}^k\| \\ & \leq \sqrt{5 \left(\|z^{k+1} - z^k\|^2 + \|x^{k+1} - x^k\|^2 + \|y^k - \hat{y}^k\|^2 + \|y^{k+1} - y^k\|^2 + \|\hat{y}^{k-1} - \hat{y}^k\|^2 \right)}. \end{aligned}$$

再根据不等式 $a+b \geq 2\sqrt{ab}, a, b \geq 0$ 有,

$$\begin{aligned} & \|z^{k+1} - z^k\| + \|x^{k+1} - x^k\| + \|y^k - \hat{y}^k\| + \|y^{k+1} - y^k\| + \|\hat{y}^{k-1} - \hat{y}^k\| \\ & \leq 2 \sqrt{\frac{5\tau_2}{\tau_1} \Delta_{k,k+1}} \sqrt{\frac{1}{4} \left(\|x^k - x^{k-1}\| + \|y^k - y^{k-1}\| + \|y^k - \hat{y}^{k-1}\| + \|\hat{y}^{k-1} - \hat{y}^{k-2}\| \right)} \tag{27} \\ & \leq \frac{5\tau_2}{\tau_1} \Delta_{k,k+1} + \frac{1}{4} \left(\|x^k - x^{k-1}\| + \|y^k - y^{k-1}\| + \|y^k - \hat{y}^{k-1}\| + \|\hat{y}^{k-1} - \hat{y}^{k-2}\| \right). \end{aligned}$$

对(27)式从 $k = K+1$ 到 q 求和可得

$$\begin{aligned} & \sum_{k=K+1}^q \left(\|z^{k+1} - z^k\| + \frac{3}{4} \|x^{k+1} - x^k\| + \frac{1}{2} \|y^{k+1} - y^k\| + \frac{3}{4} \|\hat{y}^k - y^k\| + \frac{3}{4} \|\hat{y}^k - \hat{y}^{k-1}\| \right) \\ & \leq \sum_{k=K+1}^q \left(\frac{1}{4} \left(\|x^k - x^{k-1}\| - \|x^{k+1} - x^k\| \right) + \frac{1}{4} \left(\|y^k - y^{k-1}\| - \|y^{k+1} - y^k\| \right) \right. \\ & \quad \left. + \frac{1}{4} \left(\|\hat{y}^{k-1} - \hat{y}^{k-2}\| - \|\hat{y}^k - \hat{y}^{k-1}\| \right) + \frac{1}{4} \left(\|y^k - \hat{y}^{k-1}\| + \|y^{k+1} - y^k\| - \|y^k - \hat{y}^k\| \right) \right) \tag{28} \\ & \quad + \sum_{k=K+1}^q \frac{5\tau_2}{\tau_1} \Delta_{K+1,q+1} \\ & \leq \frac{5\tau_2}{\tau_1} \Delta_{K+1,q+1} + \frac{1}{4} \|x^{K+1} - x^K\| + \frac{1}{2} \|y^{K+1} - y^K\| + \frac{1}{4} \|\hat{y}^K - \hat{y}^{K-1}\| + \frac{1}{4} \|y^K - \hat{y}^K\|. \end{aligned}$$

由于显然有 $\varphi(\widetilde{\mathcal{F}}_\beta(\omega^{q+1}) - \widetilde{\mathcal{F}}_\beta(\omega^*)) > 0$, 在(28)中取极限 $q \rightarrow +\infty$ 得到

$$\begin{aligned} & \sum_{k=K+1}^{+\infty} \left(\|z^{k+1} - z^k\| + \frac{3}{4} \|x^{k+1} - x^k\| + \frac{1}{2} \|y^{k+1} - y^k\| + \frac{3}{4} \|\hat{y}^k - y^k\| + \frac{3}{4} \|\hat{y}^k - \hat{y}^{k-1}\| \right) \\ & \leq \frac{5\tau_2}{\tau_1} \varphi(\widetilde{\mathcal{F}}_\beta(\omega^{K+1}) - \widetilde{\mathcal{F}}_\beta(\omega^*)) + \frac{1}{4} \|x^{K+1} - x^K\| + \frac{1}{2} \|y^{K+1} - y^K\| + \frac{1}{4} \|\hat{y}^K - \hat{y}^{K-1}\| + \frac{1}{4} \|y^K - \hat{y}^K\|, \end{aligned}$$

这一不等式意味着有以下结果成立,

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

又根据(16)式以及 $\sum_{k=1}^{+\infty} \|\hat{y}^k - \hat{y}^{k-1}\|$ 这一结论可知,

$$\sum_{k=1}^{+\infty} \|\lambda^{k+1} - \lambda^k\| < +\infty.$$

因此,

$$\sum_{k=1}^{+\infty} \|\omega^{k+1} - \omega^k\| < +\infty,$$

表明 $\{\omega^k\}$ 是一个柯西列, 从而收敛。证明完毕。

5. 数值实验

本节中, 通过求解最小二乘问题验证所提出的 IMADMM 算法的有效性。所有实验均在 Intel Core i5-1035G1 处理器(主频 1.19 GHz), 内存 8 GB 的 64 位计算机上进行, 通过 MATLAB R2024b 运行实现。

考虑线性回归模型 $b_i = a_i^T x + \epsilon_i$, 其中 $i = 1, 2, \dots, m$, 令 $A = (a_1, a_2, \dots, a_m)^T \in \mathbb{R}^{m \times n}$, $b = (b_1, b_2, \dots, b_m)^T \in \mathbb{R}^m$ 以及误差向量 $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_m)^T$ 。注意特征维度 n 可能超过样本量 m , 即允许欠定模型 $n > m$ 。本实验目标是重构真实的稀疏向量 x , 采用如下的惩罚最小二乘模型[25] [26],

$$\min \sum_{i=1}^n P_\theta(|x_i|) + \frac{1}{2} \|Ax - b\|_2^2, \quad (29)$$

其中目标函数 P_θ 是带有稀疏性的非负惩罚函数。

问题(29)可以通过定义,

$$f(x) = \sum_{i=1}^n P_\theta(|x_i|), g(y) = \frac{1}{2} \|Ay - b\|_2^2,$$

转化为问题(1)的形式。在本小节中, 主要关注一种广泛使用的非凸惩罚, 即极小极大凹罚(MCP), 定义如下,

$$\text{MCP}_\theta(x) = \begin{cases} \theta x - \frac{x^2}{2a}, & x \leq \theta a, \\ \frac{a\theta^2}{2}, & x > \theta a, \end{cases}$$

其中 $a > 1$, $\theta > 0$ 。MCP $_\theta$ 的邻近算子, 即 firm-threshold 算子[27], 由下式给出,

$$\text{Prox}_{t, \text{MCP}_\theta}(w_i) = \begin{cases} w_i, & |w_i| \geq a\theta, \\ \text{sign}(w_i) \frac{\theta(|w_i| - ta)}{\theta - t}, & ta < |w_i| < a\theta, \\ 0, & |w_i| \leq ta. \end{cases}$$

现在给出最小二乘问题的实验设置。设计矩阵 $A \in \mathbb{R}^{m \times n}$ 的元素从标准正态分布中生成, 之后对 A 的所有列进行归一化处理。生成一个稀疏度为 10% 的稀疏向量 ω , 响应向量 $b = A\omega + \epsilon$, 其中噪声向量 ϵ 服从均值为 0, 协方差矩阵为 $10^{-3}I$ 的正态分布。对于惩罚函数, 根据文献[25], 为 MCP 设置参数 $\theta = 0.158$, $a = 2$ 。当满足以下条件时算法终止,

$$\text{Error} := \frac{\|x^{k+1} - x^k\|}{\|x^k\|} \leq 10^{-4}.$$

在参数的设置上, 固定 $\gamma = 0.079, \alpha = 0.142$, 并设置 IMADMM 中 $\sigma = 0.5, \delta = 0.01$ 。我们测试了两种

问题规模, 对应的 (m, n) 分别为 $(10, 100)$ 和 $(100, 1000)$ 。

本实验比较了所提出的 IMADMM, Moreau 包络交替方向乘法(MADMM)以及经典 ADMM 算法的性能。图 1 展示了在 MCP 正则下, 不同算法的目标函数对数值随迭代次数变化的趋势。结果表明, IMADMM 和 MADMM 在迭代次数和收敛速率相较于 ADMM 具有显著优势。并且, 所提出的 IMADMM 算法在高维以及低维问题上收敛均最快。这表明 Moreau 包络显著提高了算法的稳定性和效率。此外, 为了进一步观察测试算法的收敛速度, 图 2 展示了在 MCP 正则化下, 不同算法的 Error 值随迭代次数的变化趋势, 它们也表明了 IMADMM 算法优于其他算法。

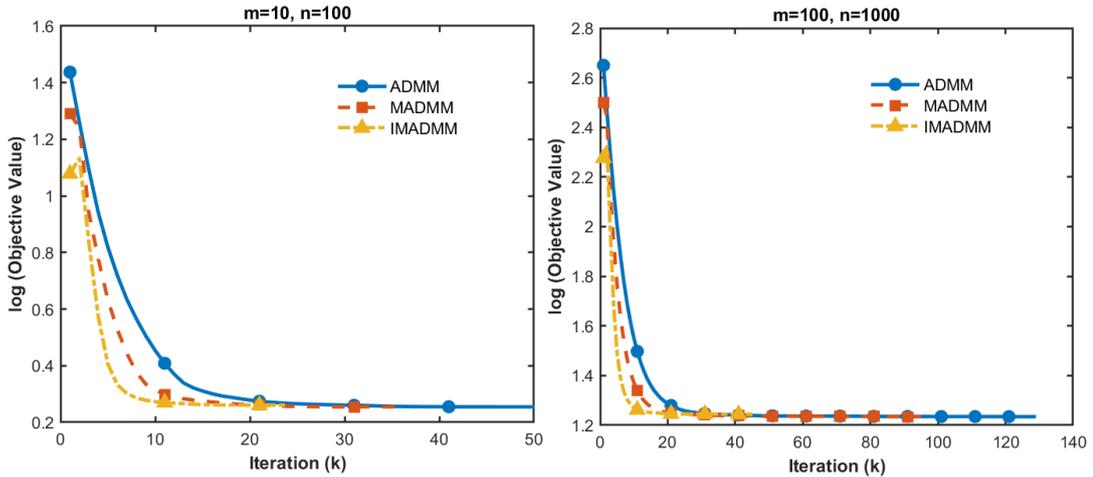


Figure 1. Graphs of the logarithmic values of the objective function under different algorithms

图 1. 不同算法下目标函数的对数值变化曲线图

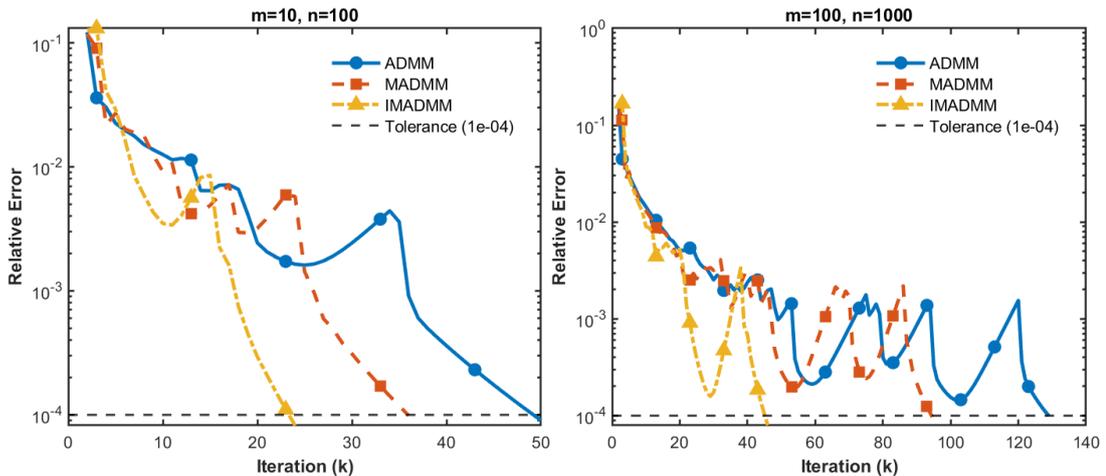


Figure 2. Error value variation curves under different algorithms

图 2. 不同算法下 Error 值变化曲线图

6. 结论

本文针对线性约束下的非凸非光滑优化问题, 提出了一种结合惯性技术与 Moreau 包络的交替方向乘法(IMADMM)。通过引入 Moreau 包络对非光滑函数进行光滑近似, 并利用惯性外推步以及 Armijo 线搜索确定步长加速迭代, 有效改善了经典 ADMM 在非凸非光滑问题中的收敛性能。理论分析表明, 在

适当的参数条件下, 算法产生的序列具有全局收敛性。数值实验以 MCP 正则化的最小二乘问题为例, 验证了 IMADMM 在低维与高维场景下均具有更快的收敛速度和更好的数值稳定性。未来工作可进一步探讨参数自适应选择策略, 并将算法拓展至更广泛的非凸优化模型中。

参考文献

- [1] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2011) 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>
- [2] Goldstein, T., Bresson, X. and Osher, S. (2010) Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction. *Journal of Scientific Computing*, **45**, 272-293. <https://doi.org/10.1007/s10915-009-9331-z>
- [3] Yang, J. and Zhang, Y. (2011) Alternating Direction Algorithms for ℓ_1 -Problems in Compressive Sensing. *SIAM Journal on Scientific Computing*, **33**, 250-278. <https://doi.org/10.1137/09077761>
- [4] Sleem, O., Ashour, M. and Aybat, N. (2024) L_p Quasi-Norm Minimization: Algorithm and Applications. *EURASIP Journal on Advances in Signal Processing*, **2024**, Article No. 22. <https://doi.org/10.1186/s13634-024-01114-6>
- [5] Huang, J., Zhang, F. and Liu, X. (2022) Stable Recovery of Sparse Signals with Non-Convex Weighted r -Norm Minus 1 -Norm. *Journal of Computational Mathematics*, **43**, 43-62. <https://doi.org/10.4208/jcm.2307-m2022-0225>
- [6] Gabay, D. and Mercier, B. (1976) A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation. *Computers & Mathematics with Applications*, **2**, 17-40. [https://doi.org/10.1016/0898-1221\(76\)90003-1](https://doi.org/10.1016/0898-1221(76)90003-1)
- [7] 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)
- [8] Alvarez, F. and Attouch, H. (2001) An Inertial Proximal Method for Maximal Monotone Operators via Discretization of a Nonlinear Oscillator with Damping. *Set-Valued Analysis*, **9**, 3-11. <https://doi.org/10.1023/a:1011253113155>
- [9] 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>
- [10] Ochs, P., Brox, T. and Pock, T. (2015) iPiasco: Inertial Proximal Algorithm for Strongly Convex Optimization. *Journal of Mathematical Imaging and Vision*, **53**, 171-181. <https://doi.org/10.1007/s10851-015-0565-0>
- [11] Boţ, R.I., Csetnek, E.R. and Hendrich, C. (2015) Inertial Douglas-Rachford Splitting for Monotone Inclusion Problems. *Applied Mathematics and Computation*, **256**, 472-487. <https://doi.org/10.1016/j.amc.2015.01.017>
- [12] Wu, Z., Li, C., Li, M. and Lim, A. (2021) Inertial Proximal Gradient Methods with Bregman Regularization for a Class of Nonconvex Optimization Problems. *Journal of Global Optimization*, **79**, 617-644. <https://doi.org/10.1007/s10898-020-00943-7>
- [13] Gao, X., Cai, X. and Han, D. (2020) A Gauss-Seidel Type Inertial Proximal Alternating Linearized Minimization for a Class of Nonconvex Optimization Problems. *Journal of Global Optimization*, **76**, 863-887. <https://doi.org/10.1007/s10898-019-00819-5>
- [14] Chen, C., Chan, R.H., Ma, S. and Yang, J. (2015) Inertial Proximal ADMM for Linearly Constrained Separable Convex Optimization. *SIAM Journal on Imaging Sciences*, **8**, 2239-2267. <https://doi.org/10.1137/15100463x>
- [15] Chao, M.T., Zhang, Y. and Jian, J.B. (2021) 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>
- [16] Xu, J.W. and Chao, M.T. (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>
- [17] Moreau, J. (1965) Proximité et dualité dans un espace hilbertien. *Bulletin de la Societe Mathematique de France*, **93**, 273-299.
- [18] Yosida, K. (1964) *Functional Analysis*. Springer.
- [19] Zeng, J.S., Yin, W.T. and Zhou, D.X. (2022) Moreau Envelope Augmented Lagrangian Method for Nonconvex Optimization with Linear Constraints. *Journal of Scientific Computing*, **91**, Article No. 61. <https://doi.org/10.1007/s10915-022-01815-w>
- [20] Qiu, T.X., Huang, W.G., Zhang, Z.C., Wang, J. and Zhu, Z.K. (2024) A New Approach for Sparse Optimization with Moreau Envelope to Extract Bearing Fault Feature. *Mechanical Systems and Signal Processing*, **216**, Article 111493. <https://doi.org/10.1016/j.ymssp.2024.111493>

-
- [21] Wu, Z. and Li, M. (2019) General Inertial Proximal Gradient Method for a Class of Nonconvex Non-Smooth Optimization Problems. *Computational Optimization and Applications*, **73**, 129-158. <https://doi.org/10.1007/s10589-019-00073-1>
- [22] Rockafellar, R. and Wets, R. (2009) *Variational Analysis*. Springer.
- [23] Attouch, H., Bolte, J., Redont, P. and Soubeyran, A. (2010) Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Łojasiewicz Inequality. *Mathematics of Operations Research*, **35**, 438-457. <https://doi.org/10.1287/moor.1100.0449>
- [24] Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal Alternating Linearized Minimization for Nonconvex and Non-Smooth Problems. *Mathematical Programming*, **146**, 459-494. <https://doi.org/10.1007/s10107-013-0701-9>
- [25] Fan, J.Q. and Li, R.Z. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348-1360. <https://doi.org/10.1198/016214501753382273>
- [26] Liu, H.C., Yao, T., Li, R.Z. and Ye, Y.Y. (2017) Folded Concave Penalized Sparse Linear Regression: Sparsity, Statistical Performance, and Algorithmic Theory for Local Solutions. *Mathematical Programming*, **166**, 207-240. <https://doi.org/10.1007/s10107-017-1114-y>
- [27] Gao, H.Y. and Bruce, A.G. (1997) Waveshrink with Firm Shrinkage. *Statistica Sinica*, **7**, 855-874.