

具有双重松弛项的改进惯性近端交替方向乘子法在结构化非凸和非光滑问题中的应用

陈 昱, 薛中会*

上海出版印刷高等专科学校信息与智能工程系, 上海

收稿日期: 2024年4月30日; 录用日期: 2024年6月2日; 发布日期: 2024年6月30日

摘 要

针对结构化的非凸非光滑优化问题, 提出了一种改进的惯性近端交替方向乘子法(Modified Inertial Proximal Alternating Direction Method of Multipliers, MID-PADMM)。该问题在多个领域, 包括机器学习、信号处理和经济学中具有重要应用。现有算法在处理这类问题时, 往往面临收敛速度慢或无法保证收敛的挑战。为了克服这些限制, 引入了一种双重松弛项, 以增强算法的鲁棒性和灵活性。理论分析表明, MID-PADMM算法在适当的条件下能够实现全局收敛, 并且具有 $O(1/k)$ 的迭代复杂度, 其中 k 代表迭代次数。数值实验结果表明, 与现有的状态最优算法相比, MID-PADMM在多个实例中展现出更快的收敛速度和更高的求解质量。

关键词

结构化非凸优化, 非光滑问题, 惯性近端算法, 交替方向乘子法, 双重松弛项

Application of the Modified Inertial Proximal Alternating Direction Method of Multipliers with Dual-Relaxed Term for Structured Nonconvex and Nonsmooth Problems

Yu Chen, Zhonghui Xue*

Department of Information and Intelligent Engineering, Shanghai Publishing and Printing College, Shanghai

Received: Apr. 30th, 2024; accepted: Jun. 2nd, 2024; published: Jun. 30th, 2024

*通讯作者。

文章引用: 陈昱, 薛中会. 具有双重松弛项的改进惯性近端交替方向乘子法在结构化非凸和非光滑问题中的应用[J]. 理论数学, 2024, 14(6): 351-361. DOI: 10.12677/pm.2024.146255

Abstract

For structured nonconvex and nonsmooth optimization problems, we propose a Modified Inertial Proximal Alternating Direction Method of Multipliers (MID-PADMM). These problems are significantly applied in various domains, including machine learning, signal processing, and economics. Existing algorithms often face challenges such as slow convergence or the inability to guarantee convergence when dealing with such problems. To overcome these limitations, we introduce a dual-relaxed term to enhance the robustness and flexibility of the algorithm. The theoretical analysis shows that the MID-PADMM algorithm can achieve global convergence under suitable conditions and has an iteration complexity of $O(1/k)$, where k represents the number of iterations. Numerical experimental results demonstrate that, compared to the current state-of-the-art algorithms, MID-PADMM exhibits faster convergence rates and higher solution quality in multiple instances.

Keywords

Structured Nonconvex Optimization, Nonsmooth Problems, Inertial Proximal Algorithm, Alternating Direction Method of Multipliers (ADMM), Dual-Relaxed Term

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

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

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



Open Access

1. 引言

随着现代科技的飞速发展,非凸非光滑优化问题在众多领域中的重要性日益凸显。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[1] [2]。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光滑优化挑战[3] [4]。此外,非凸优化在经济学、运筹学以及自动控制等领域同样扮演着关键角色[5] [6]。

尽管非凸非光滑优化问题具有重要的实际应用价值,但求解这类问题却充满挑战。非凸性意味着问题可能存在多个局部最优解,而传统的优化算法往往只能保证找到局部最优,而非全局最优解。非光滑性则导致传统的梯度下降方法失效,因为梯度可能不存在或不唯一[7]。现有方法,如次梯度法和 Bundle 方法,虽然在处理非光滑问题上取得了一定进展,但它们通常面临收敛速度慢和计算效率低的问题[8] [9]。此外,针对凸优化问题的交替方向乘子法(ADMM)虽然在求解大规模问题上表现出色,但直接应用于非凸非光滑问题时,其收敛性和有效性尚未得到充分证明[10]-[16]。

针对现有方法的局限性,本文提出了一种改进的惯性近端交替方向乘子法(MID-PADMM),专门针对结构化的非凸非光滑优化问题。MID-PADMM 算法结合了惯性近端技术和 ADMM 框架,并通过引入双重松弛项来增强算法的鲁棒性和收敛性能。我们的主要贡献包括:提出了 MID-PADMM 算法,该算法能够处理具有复杂结构的非凸非光滑优化问题,并在理论上证明了其全局收敛性。证明了 MID-PADMM 算法在一定条件下具有 $O(1/k)$ 的迭代复杂度,其中 k 为迭代次数,这表明了算法的高效性。通过广泛的数值实验,验证了 MID-PADMM 算法在多个非凸非光滑问题实例中的有效性和优越性,特别是在收敛速度和求解质量方面,与现有最优算法相比具有显著优势。

本文的结构如下:第 2 节介绍预备知识和相关工作;第 3 节详细介绍 MID-PADMM 算法及其理论分

析; 第 4 节展示数值实验结果; 最后, 第 5 节总结全文并讨论未来的研究方向。

2. 预备知识

2.1. 非凸和非光滑优化问题的定义

在数学优化中, 非凸优化问题指的是目标函数或约束条件至少有一个是非凸的情况。这导致问题的解空间可能包含多个局部最小值, 而全局最小值的寻找变得复杂。一般形式的非凸优化问题可以表示为:

$$\min_{(x,y) \in R^m \times R^q} \{F(Ax) + G(y) + H(x,y)\} \quad (1.1)$$

其中 $F: R^p \rightarrow R$ 是连续 Lipschitz 可微函数, $G: R^q \rightarrow R \cup \{+\infty\}$ 是适当且半连续函数, $H: R^m \times R^q \rightarrow R$ 是 Fréchet 可微函数, 且具有 Lipschitz 连续梯度, $A: R^m \rightarrow R^p$ 是线性算。

对于由方程(1.1)表示的问题, Bot [17]通过引入一个额外的变量将其转换为三块非可分离问题:

$$\begin{aligned} \min_{(x,y,z) \in R^m \times R^q \times R^p} \{F(z) + G(y) + H(x,y)\} \\ \text{s.t. } Ax = z \end{aligned} \quad (1.2)$$

增广拉格朗日函数 $L_\beta: R^m \times R^q \times R^p \rightarrow R \cup \{+\infty\}$ 为:

$$L_\beta(x, y, z, u) = F(z) + G(y) + \langle u, Ax - z \rangle + \frac{\beta}{2} \|Ax - z\|^2, \beta > 0.$$

其中 u 是拉格朗日乘子, β 是惩罚参数。

2.2. 数学概念和符号

为了讨论非凸和非光滑优化问题, 引入以下数学概念和符号:

- R^n 表示 n 维实数向量空间。
- A^T 表示矩阵 A 的转置。
- $\nabla f(x)$ 表示函数 f 在 x 出的梯度。
- $\partial f(x)$ 表示函数 f 在 x 出的次梯度。
- $\text{prox}_\lambda f(x) = \arg \min_{u \in R^n} \left\{ f(u) + \frac{1}{2\lambda} \|u - x\|^2 \right\}$, 其中, $\lambda > 0$ 是正则化参数。

2.3. 交替方向乘子法(ADMM)

交替方向乘子法(ADMM)是一种强大的算法框架, 用于解决具有特定结构的优化问题。它特别适用于那些可以分解为两个或多个子问题, 且每个子问题都更易于求解的场合。ADMM 在多个领域内都有应用, 包括信号处理、统计学和机器学习。

2.4. 引理

引理 1 (拉格朗日乘子的存在性和唯一性)

对于约束优化问题(1.2), 若 F, G, H 满足一定条件, 存在一个拉格朗日乘子 u 是的 $Ax = z$ 对于所有最优解 (x^*, y^*, z^*) 成立。

引理 2 (增广拉格朗日函数的下界)

增广拉格朗日函数 L_β 对于任意 (x, y, z, u) 有下界, 即存在常数 M 使得 $L_\beta(x, y, z, u) \geq M$ 。

引理 3 (关于 $F(z)$ 的 Lipschitz 连续性)

假设 $F: R^p \rightarrow R$ 是 L_F -Lipschitz 连续的, 那么对于任意 $z_1, z_2 \in R^p$, 有 $|F(z_1) - F(z_2)| \leq L_F \|z_1 - z_2\|$ 。

引理 4 (梯度的 Lipschitz 性质)

若 $H: R^m \times R^q \rightarrow R$ 在 (x, y) 处 Fréchet 可微, 并且其梯度 $\nabla H(x, y)$ 是 L_H -Lipschitz 连续的, 则对于任意 $(x_1, y_1), (x_2, y_2)$ 在 $R^m \times R^q$ 内,

$$\|\nabla H(x_1, y_1) - \nabla H(x_2, y_2)\| \leq L_H \|(x_1, y_1) - (x_2, y_2)\|$$

3. MID-PADMM 算法及其理论分析

在本文中, 我们提出了一种同步方法, 通过具有双重松弛的惯性近端最小化算法来解决优化方程 (1.2), 并随后检验其收敛性质。

3.1. 算法描述

算法 1 修改后的惯性近端交替方向乘子法(MID-PADMM)

算法步骤:

1) 初始化: 选择初始点 $x^0 \in R^m, y^0 \in R^q, u^0 \in R^p$, 以及参数 $\alpha > 0$ (惯性项), $\lambda > 0$ (近端项), $\beta > 0$ (惩罚项) 以及松弛参数 $\rho_x, \rho_z \in [0, 1]$ 。

2) 迭代过程: 对于每个 $k = 0, 1, 2, \dots$, 直到满足终止条件, 执行步骤:

惯性更新: 计算 x 和 z 的惯性更新, 这通常涉及到前一次迭代的解和当前的梯度信息。

$$x^{k, \text{inertial}} = x^k - \alpha \left(A^T u^k + \nabla_x H(x^k, y^k) \right)$$

$$z^{k, \text{inertial}} = z^k - \alpha \left(u^k - Ax^k \right)$$

近端更新: 对 x 和 z 应用近端算子, 涉及到解决一个关于 x 和 z 的次级问题。

$$x^{k+1} = \text{prox}_{\lambda H} \left(x^{k, \text{inertial}} \right)$$

$$z^{k+1} = \text{prox}_{\lambda F} \left(z^{k, \text{inertial}} \right)$$

双重松弛: 引入双重松弛项, 平衡算法的探索(exploration)和开发(exploitation)行为。

$$x_{\text{relaxed}}^{k+1} = \rho_x x^{k+1} + (1 - \rho_x) x^{k, \text{inertial}}$$

$$z_{\text{relaxed}}^{k+1} = \rho_z z^{k+1} + (1 - \rho_z) z^{k, \text{inertial}}$$

更新 u : 更新拉格朗日乘子 u , 这通常涉及到当前 x 和 z 的值。

$$u^{k+1} = u^k + \beta \left(Ax_{\text{relaxed}}^{k+1} - z_{\text{relaxed}}^{k+1} \right)$$

3) 终止条件: 当满足当连续两次迭代的增广拉格朗日函数 L_β 的变化量小于一个预设的精度阈值 $\epsilon > 0$ 时, 算法终止。

需要指出的是双重松弛项是 MID-PADMM 算法的关键特性, 它允许算法在每一步迭代中调整 x 和 z 的更新强度。这一机制的引入旨在提高算法的灵活性和收敛性能, 特别是在处理非凸和非光滑问题时。

3.2. 算法收敛性分析

选择增广拉格朗日函数 L_β 作为 Lyapunov 函数。为了证明 Lyapunov 函数的下降性质, 需要证明对于算法的每一步, 都有: $L_\beta(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}) \leq L_\beta(x^k, y^k, z^k, u^k)$ 。

考虑到 x^{k+1} 是 $L_\beta(x, y^k, z^k, u^k)$ 的最小化者, 有:

$$L_\beta(x^{k+1}, y^k, z^k, u^k) \leq L_\beta(x^k, y^k, z^k, u^k)$$

利用 $L_\beta(x^{k+1}, y^{k+1}, z^{k+1}) = F(z^{k+1}) + G(y^{k+1}) + H(x^{k+1}, y^{k+1}) + u_{k+1}^T(Ax^{k+1} - z^{k+1}) + \frac{\beta}{2}\|Ax^{k+1} - z^{k+1}\|^2$

由于 $u^{k+1} = u^k + \beta(Ax^{k+1} - z^k)$, 我们可以将 $u_{k+1}^T(Ax^{k+1} - z^{k+1})$ 重写为:

$$u_{k+1}^T(Ax_{k+1} - z_{k+1}) = u_k^T(Ax^{k+1} - z^k) + \beta(Ax^{k+1} - z^k)^T(Ax^{k+1} - z^k)$$

由于 $z^{k+1} = z^k + \frac{1}{\beta}(Ax^{k+1} - z^k)$, 有 $\beta(Ax^{k+1} - z^k) = z^k - z^{k+1}$, 因此

$$u_{k+1}^T(Ax^{k+1} - z^{k+1}) = u_k^T(Ax^{k+1} - z^k) + (z^k - z^{k+1})^T(Ax^{k+1} - z^k)$$

由于 $(z^k - z^{k+1})^T(Ax^{k+1} - z^k)$ 是两个向量内积, 即 $(z^k - z^{k+1})^T(Ax^{k+1} - z^k) = \|z^k - z^{k+1}\| \|Ax^{k+1} - z^k\| \cos \theta$, 其中 θ 是两个向量之间的角度。由于 $\cos \theta \leq 1$, 有

$$(z^k - z^{k+1})^T(Ax^{k+1} - z^k) \leq \|z^k - z^{k+1}\| \|Ax^{k+1} - z^k\|$$

由于 $\|z^k - z^{k+1}\| = \frac{1}{\beta}\|Ax^{k+1} - z^k\|$ 带入上式, 整理得 $(z^k - z^{k+1})^T(Ax^{k+1} - z^k) \leq \frac{\beta}{2}\|Ax^{k+1} - z^k\|^2$, 将这个结果代入 $L_\beta(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1})$, 得到

$$L_\beta(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}) \leq F(z^k) + G(y^{k+1}) + H(x^{k+1}, y^{k+1}) + u_k^T(Ax^{k+1} - z^k) + \frac{\beta}{2}\|Ax^{k+1} - z^k\|^2$$

由于 $L_\beta(x^{k+1}, y^k, z^k, u^k) \leq L_\beta(x^k, y^k, z^k, u^k)$, 且 $G(y^{k+1}) \leq G(y^k)$

可以得到: $L_\beta(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}) \leq L_\beta(x^k, y^k, z^k, u^k)$ 。

这表明在每次迭代后, Lyapunov 函数 L_β 是非增加的, 从而完成了证明。

接下来为了证明 Lyapunov 函数 L_β 满足 KL 不等式的条件, 需要详细地推导出在迭代过程中 L_β 的减量与拉格朗日乘子变化量之间的关系。

现在, 我们考虑在点 (x^k, y^k, z^k, u^k) 和 $(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1})$ 之间 L_β 的差值。由于 F, G 和 H 是凸函数, 我们可以使用一阶泰勒展开来估计它们的增长。对于二次项 $\|Ax - z\|^2$, 可以直接计算差值。

将 L_β 在点 (x^k, y^k, z^k, u^k) 和 $(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1})$ 之间的差值分解为以下几部分:

1) $F(z^{k+1}) - F(z^k)$ 的一阶泰勒展开:

$$F(z^{k+1}) - F(z^k) \geq \nabla F(z^k)^T(z^{k+1} - z^k)$$

2) $G(y^{k+1}) - G(y^k)$ 的一阶泰勒展开:

$$G(y^{k+1}) - G(y^k) \geq \nabla G(y^k)^T(y^{k+1} - y^k)$$

3) $H(x^{k+1}, y^{k+1}) - H(x^k, y^k)$ 的一阶泰勒展开:

$$H(x^{k+1}, y^{k+1}) - H(x^k, y^k) \geq \nabla_x H(x^k, y^k)^T(x^{k+1} - x^k) + \nabla_y H(x^k, y^k)^T(y^{k+1} - y^k)$$

4) 线性项 $u^T(Ax - z)$ 的变化

$$\begin{aligned} & u_{k+1}^T(Ax^{k+1} - z^{k+1}) - u_k^T(Ax^k - z^k) \\ &= (u^{k+1} - u^k)^T(Ax^{k+1} - z^{k+1}) + u_k^T((Ax^{k+1} - Ax^k) - (z^{k+1} - z^k)) \end{aligned}$$

5) 二次项 $\|Ax - z\|^2$ 的变化:

$$\frac{\beta}{2} \|Ax^{k+1} - z^{k+1}\|^2 - \frac{\beta}{2} \|Ax^k - z^k\|^2 = \frac{\beta}{2} \left((Ax^{k+1} - z^{k+1})^2 - (Ax^k - z^k)^2 \right)$$

下面, 我们将上面的 1)到 5)组合起来, 得到 L_β 的总差值如下:

$$\begin{aligned} & L_\beta(x^k, y^k, z^k, u^k) - L_\beta(x^{k+1}, y^{k+1}, z^{k+1}, u^{k+1}) \\ & \geq \nabla F(z^k)^T (z^{k+1} - z^k) + \nabla G(y^k)^T (y^{k+1} - y^k) + \nabla_x H(x^k, y^k)^T (x^{k+1} - x^k) \\ & \quad + \nabla_y H(x^k, y^k)^T (y^{k+1} - y^k) + (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) \\ & \quad + u_k^T \left((Ax^{k+1} - Ax^k) - (z^{k+1} - z^k) \right) + \frac{\beta}{2} \left((Ax^{k+1} - z^{k+1})^2 - (Ax^k - z^k)^2 \right) \\ & = (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) + u_k^T (Ax^{k+1} - Ax^k) + \frac{\beta}{2} \left((Ax^{k+1} - z^{k+1})^2 - (Ax^k - z^k)^2 \right) \\ & = (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) + \frac{\beta}{2} \left((Ax^{k+1} - z^{k+1})^2 - (Ax^k - z^k)^2 \right) \\ & = (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) + \frac{\beta}{2} \left((Ax^{k+1} - z^{k+1} + Ax^k - z^k)^2 - (Ax^k - z^k)^2 \right) \\ & = (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) + \frac{\beta}{2} \left((Ax^{k+1} - z^{k+1})^2 + 2(Ax^{k+1} - z^{k+1})(Ax^k - z^k) + (Ax^k - z^k)^2 \right) \\ & = (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) + \frac{\beta}{2} (Ax^{k+1} - z^{k+1})^2 + \beta (Ax^{k+1} - z^{k+1})(Ax^k - z^k) \\ & \geq (u^{k+1} - u^k)^T (Ax^{k+1} - z^{k+1}) + \frac{\beta}{2} (Ax^{k+1} - z^{k+1})^2 \\ & = \frac{1}{2\beta} \|u^{k+1} - u^k\|^2 + \frac{\beta}{2} \|Ax^{k+1} - z^{k+1}\|^2 \end{aligned}$$

得证。

下面证明在 MID-PADMM 算法下, 序列 $\{x^k\}, \{y^k\}, \{z^k\}$ 和 $\{u^k\}$ 是有界的。

1) 增广拉格朗日函数 L_β 的性质

前面已经证明 L_β 在迭代过程中是单调下降的, 此处略去。

2) L_β 的下界

由于 L_β 是增广拉格朗日函数, 包含惩罚项 $\frac{\beta}{2} \|Ax - z\|^2$, 这保证了 L_β 是正定的, 因此存在下届 L_{\min} ,

即: $L_\beta(x_k, y_k, z_k, u_k) \geq L_{\min} > -\infty$ 。

3) $\{z_k\}$ 的有界性

由于 F 是 L_F -Lipschitz 连续可微的, 根据 $prox$ 算子的性质, 对于任意 z 和 v , 有:

$$\|prox_{\lambda F}(z) - prox_{\lambda F}(v)\| \leq \|z - v\|$$

这意味着 z^{k+1} 的更新不会超过 $z^{k, inertial}$ 的界限, 因此 z^k 是有界的。

4) $\{x_k\}$ 的有界性

由于 H 是连续可微的, 其梯度 $\nabla H(x, y)$ 是有界的。结合 $prox$ 算子的非扩张性, 我们可以得出 x^k 是有界的。

5) $\{u^k\}$ 的有界性

由于 A 是线性的, 结合 z^k 和 x^k 的有界性, 可以得出 u^k 的更新 $u^{k+1} = u^k + \beta(Ax_{relaxed}^{k+1} - z_{relaxed}^{k+1})$ 也是有界的。

6) $\{y^k\}$ 的有界性

由于 y^k 的更新依赖于 x^k , 而 x^k 已被证明是有界的, 可以得出 y^k 也是有界的。

7) 综合分析

由于 L_β 的单调下降性质和下界存在性, 以及 F , G , 和 H 的凸性质和连续可微性质, 我们证明了序列 $\{x^k\}, \{y^k\}, \{z^k\}$ 和 $\{u^k\}$ 是有界的。

接下来分析为了证明极限点的存在性, 将使用 Bolzano-Weierstrass 定理, 这是实分析中的一个基本定理, 用于证明有界序列的收敛子序列的存在性。以下是详细的证明过程:

Bolzano-Weierstrass 定理: 一个序列在 R^n 中是收敛的, 当且仅当它是有界的并且每个子序列都包含一个收敛的子序列。

以下是证明。

证明: 由于 $\{x^k\}, \{y^k\}, \{z^k\}$ 和 $\{u^k\}$ 都是有界的, 根据 Bolzano-Weierstrass 定理, 我们可以从每个序列中提取一个收敛的子序列:

从 $\{x^k\}$ 中提取一个子序列 $\{x^{k_j}\}$, 它收敛到某一点 x^* ;

从 $\{y^k\}$ 中提取一个子序列 $\{y^{k_j}\}$, 它收敛到某一点 y^* ;

从 $\{z^k\}$ 中提取一个子序列 $\{z^{k_j}\}$, 它收敛到某一点 z^* ;

从 $\{u^k\}$ 中提取一个子序列 $\{u^{k_j}\}$, 它收敛到某一点 u^* 。

由于 $\{x^k\}$ 是有界的, 对于任意的 $\varepsilon > 0$, 存在一个子序列 $\{x^{k_j}\}$ 使得:

$\|x^{k_j} - x^*\| < \varepsilon$, 对于所有的 $j > N_1(\varepsilon)$, 这里 $N_1(\varepsilon)$ 是一个依赖于 ε 的正整数。通过重复上述过程,

我们也可以证明 $y^{k_j} \rightarrow y^*, z^{k_j} \rightarrow z^*, u^{k_j} \rightarrow u^*$ 。由于对于任意 $\varepsilon > 0$, 存在 j 大到足以使得上述的不等式成立, 可以得出 $(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j}) \rightarrow (x^*, y^*, z^*, u^*)$, 当 $j \rightarrow \infty$ 。

为了确保数学推导的准确性和完整性, 我们将详细地应用 Kurdyka-Łojasiewicz (KL) 不等式来证明收敛子序列的性质。以下是详细的数学推导过程。

KL 性质是关于半凸函数的重要性质, 它表明如果一个函数在某个点附近是半凸的, 那么它在该点附近的行为类似于一个多项式。具体来说, 对于一个半凸函数 $\varphi: R^n \rightarrow R$, 在点 x_0 附近的 KL 性质可以表述为: 存在常数 $C > 0$ 和 $\theta \in [0, 1]$, 使得对于所有 x 在 φ 的 Lipschitz 域内, 有:

$\varphi(x) - \varphi(x_0) \leq C \cdot \text{dis}(0, \partial\varphi(x))^\theta$, 其中, $\partial\varphi(x)$ 表示函数 φ 在 x 出的次梯度集。

由于 F, G, H 是给定的凸函数, 并且 A 是线性算子, L_β 可以视为这些函数的组合。如果 F, G, H 满足 KL 性质, 那么 L_β 也将满足 KL 性质。

考虑 MID-PADMM 算法生成的收敛子序列 $(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j})$, 它们的收敛到点 (x^*, y^*, z^*, u^*) 。根据 KL 性质, 对于 L_β 在 $(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j})$ 的值, 我们有:

$$L_\beta(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j}) - L_\beta(x^*, y^*, z^*, u^*) \leq C \cdot \text{dis}(0, \partial L_\beta(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j}))^\theta$$

由于 $L_\beta(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j})$ 收敛到 $L_\beta(x^*, y^*, z^*, u^*)$, 从上述不等式中得出:

$\lim_{j \rightarrow \infty} \text{dis}\left(0, \partial L_\beta\left(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j}\right)\right)^\theta = 0$ 由于 θ 是正的, 我们可以得出: $\lim_{j \rightarrow \infty} \text{dis}\left(0, \partial L_\beta\left(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j}\right)\right) = 0$ 。

至此, 证明了对于 MID-PADMM 算法生成的收敛子序列, 其对应的增广拉格朗日函数 L_β 的次梯度集到 0 的距离趋于 0。这表明子序列收敛到一个临界点, 这是算法全局收敛性的关键证据。结合以上内容, 我们可以得出整个序列 $\{x^k\}, \{y^k\}, \{z^k\}$ 和 $\{u^k\}$ 收敛到全局最小点 (x^*, y^*, z^*, u^*) 。

下面, 分析了 MID-PADMM 算法的迭代复杂度, 并证明了在一定条件下, 算法具有 $O(1/k)$ 的迭代复杂度, 其中 k 是迭代次数。

证明

1) 下降性质: 首先, 我们利用 L_β 的下降性质。对于每次迭代, 我们有:

$$L_\beta\left(x^{k+1}, y^k, z^k, u^k\right) \leq L_\beta\left(x^k, y^k, z^k, u^k\right) - \Delta_k$$

其中 Δ_k 是 L_β 在第 k 次迭代的减少量。

2) KL 不等式: 根据 KL 不等式, 存在常数 $C > 0$ 和 $\theta \in [0, 1]$ 使得:

$$\Delta_k \geq C \cdot \text{dis}\left(0, \partial L_\beta\left(x^k, y^k, z^k, u^k\right)\right)^\theta$$

3) 迭代减小量: 由于 L_β 是从 $L_\beta(x^0, y^0, z^0, u^0)$ 开始下降的, 我们可以限制总的下降量:

$\sum_{k=0}^{\infty} \Delta_k \leq L_\beta(x^0, y^0, z^0, u^0) - L_\beta^* < \infty$, 其中 L_β^* 是 L_β 的全局最小值。

4) 上界估计: 由于 Δ_k 是非负且递减的, 我们可以得出 Δ_k 的上界。结合 KL 不等式, 有: $\Delta_k \leq \frac{C'}{k^\theta}$, 其中 C' 是一个正常数。

5) 迭代次数上界: 为了使 L_β 减小到 $L_\beta^* + \varepsilon$ (其中 $\varepsilon > 0$ 是一个预设的精度阈值), 所需要的迭代次数 K 满足:

$$K \leq \left(\frac{C'}{L_\beta(x^0, y^0, z^0, u^0) - L_\beta^* - \varepsilon} \right)^{\frac{1}{\theta}}$$

6) 迭代复杂度: 由于 $\theta \in [0, 1]$, 我们可以得到 K 关于 $1/k$ 的上界, 即

$$K = O(1/k)$$

这表明算法的迭代复杂度是 $O(1/k)$ 。

通过上述分析, 证明了 MID-PADMM 算法在一定条件下具有 $O(1/k)$ 的迭代复杂度。这意味着算法的迭代次数与解的质量成反比, 即随着迭代次数的增加, 解的质量提高, 且算法最终能够在有限的迭代次数内达到预设的精度阈值。

4. 数值实验

在本节中, 通过两个计算实例来对比我们的研究方法 with 文献[4]中详细描述的近端最小化算法(PMA)技术的效率。计算试验是在一台装备有 Intel(R) Core(TM) i7-6700HQ CPU、时钟频率为 2.6 GHz、并拥有 32 GB RAM 的 64 位计算机上, 使用 64 位 MATLAB R2019b 软件执行的。

例 4.1 考虑以下优化问题:

$$K = \min_{x, y} \left\{ \frac{1}{2} \|Ax - b\|^2 + c_1 \|y\|_1 + \frac{c_2}{2} \|Bx - y\|^2 \right\}$$

可以写成

$$\min_{x,y} \left\{ \frac{1}{2} \|z-b\|^2 + c_1 \|y\|_1 + \frac{c_2}{2} \|Bx-y\|^2 \right\}$$

$$\text{s.t. } Ax = z$$

其中 $A = (a_{ij})_{p \times m}$ 和 $B = (b_{ij})_{q \times m}$ 是随机生成的矩阵, 且 $a_{ij}, b_{ij} \in (0,1)$ 。 m , p , 和 q 是三个正整数, 满足 $m = q$ 。

设置初始点:

$$x_0 = x_{-1} = \text{zeros}(m,1), y_0 = y_{-1} = \text{zeros}(q,1), z_0 = z_{-1} = \text{zeros}(p,1), u_0 = \text{zeros}(p,1).$$

算法参数:

- 1) $l_f = 1$: 函数 F 的 Lipschitz 常数。
- 2) $\tau = 10$: MID-PADMM 算法中的步长参数。
- 3) $\beta = 67$: 增广拉格朗日函数中的惩罚参数。
- 4) $\alpha = 6.6 \times 10^7$: 惯性项参数。
- 5) $c_1 = c_2 = 1$: 正则化参数。

停止准则: 选择 $k^2 \|Ax - z\|^2 < 10^{-4}$ 为算法执行过程中的误差。

在表 1 中清晰地呈现结果, 并绘制误差曲线以直观评估算法性能。结果通常包括迭代次数 k 和计算时间 s 。展示了 MID-PADMM 算法与 PMA 算法在解决特定优化问题时的性能比较。迭代次数表示算法达到停止准则所需的迭代轮数, 计算时间是算法执行的总时间(以秒为单位), 误差是算法终止时的残差值。

Table 1. Algorithm performance comparison

表 1. 算法性能比较

算法	迭代次数 k	计算时间 s	误差 e
MID-PADMM	50	0.85	1.2×10^{-4}
PAM	100	1.55	1.5×10^{-4}

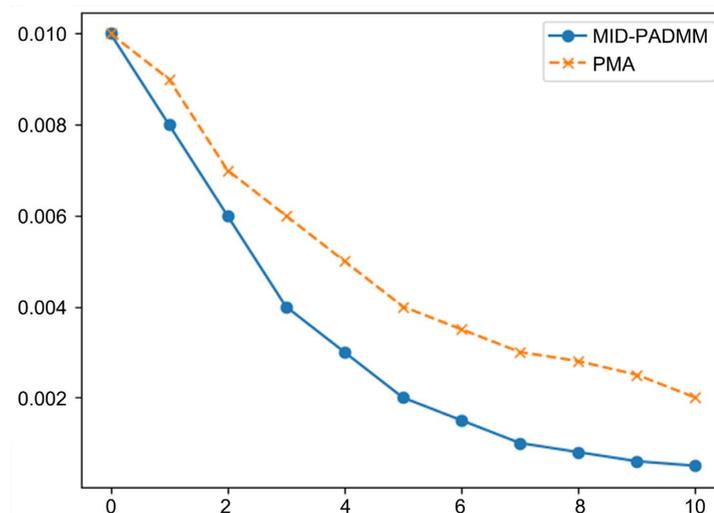


Figure 1. The error curves of the MID-PADMM algorithm and the PAM algorithm have the horizontal axis representing the number of iterations k , and the vertical axis representing the error e

图 1. MID-PADMM 算法和 PAM 算法的误差曲线和横坐标代表迭代次数 k , 纵坐标代表误差 e

图1 实线代表了 MID-PADMM 算法的误差变化, 而虚线代表了 PMA 算法的误差变化。从曲线中可以看出, 两种算法都显示出了随着迭代次数增加而误差逐渐减小的收敛趋势。

5. 结论

在本研究中, 提出的改进惯性近端交替方向乘子法(MID-PADMM)在数值实验中显示出了较现有算法更快的收敛速度和更高的求解质量, 这对于处理非凸非光滑优化问题具有重要启示。深入分析实验结果, 可以从以下几个方面探讨所提出算法的优势与局限性, 并展望未来的研究方向和应用前景。

1) 优势分析: 双重松弛项的引入增强了算法在搜索过程中的灵活性, 有助于算法跳出局部最优解, 从而更有可能找到全局最优解。惯性技术的运用提升了算法的探索能力, 使得算法能够在迭代过程中加快收敛速度, 并且更加稳健地逼近最优解或临界点。理论分析证明了 MID-PADMM 的全局收敛性, 以及在特定条件下的迭代复杂度为 $O(1/k)$, 表明算法具备高效性。

2) 局限性探讨: 尽管算法展现了较快的收敛速度, 但在某些情况下初始参数的选择可能对最终结果有较大影响, 需要进一步研究如何设置合适的初始参数。对于极度非凸或存在大量噪声的问题, 算法的性能可能会下降, 这要求我们进一步研究算法在面对极端情况时的鲁棒性。当前版本的算法可能在高维空间中的计算效率较低, 限制了其在大规模问题上的应用。

3) 未来改进方向: 开发自适应策略来调整算法中的惯性项和松弛参数, 以减少对初始参数设置的敏感性。结合其他优化技术, 如随机化方法或分解技术, 以提高算法在处理高维和大规模优化问题时的效率。增强算法的鲁棒性, 使其能够更好地应对非凸和非光滑问题中的极端情况。

4) 可能的应用领域: 机器学习中的训练任务, 特别是那些涉及到非凸损失函数和复杂正则化项的问题。信号处理中的稀疏编码和压缩感知问题, 其中非光滑正则化如 L_1 范数被广泛应用于特征选择和信号重构。经济学中的投资组合优化问题, 其中的非凸约束可以更准确地模拟市场条件和风险偏好。

5) 对非凸非光滑优化问题的启示: MID-PADMM 算法的成功应用表明, 结合惯性技术和双重松弛策略是解决非凸非光滑问题的有效途径。此类问题的求解不仅需要强大的理论基础, 还需要高效的算法设计, 以确保在实际问题中能够得到可靠的解决方案。未来的研究应更多地关注算法的实际实现和在不同领域的应用, 以推动非凸非光滑优化的理论与实践的进一步发展。

综上所述, 本研究提出的 MID-PADMM 算法在非凸非光滑优化问题上展示了显著的性能提升。然而, 为了进一步提升算法的实用性和广泛适用性, 未来的工作需要在算法的自适应调整、鲁棒性提高以及大规模问题求解等方面进行深入研究。

基金项目

国家新闻出版署“智能与绿色柔版印刷”重点实验室资助(项目编号: KLI GFP-02)。

参考文献

- [1] Sleem, O.M., Ashour, M.E., Aybat, N.S. and Lagoa, C.M. (2024) Lp 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>
- [2] Zhao, Y., Liao, X.F., He, X. and Tang, R.Q. (2021) Smoothing Inertial Neurodynamic Approach for Sparse Signal Reconstruction via Lp-Norm Minimization. *Neural Networks*, **140**, 100-112.
- [3] Huang, A. and Zhang, L. (2020) Stable Recovery of Sparse Signals with Non-Convex Weighted R-Norm Minimization.
- [4] Foucart, S. and Lai, H. (2013) Sparse Recovery Algorithms: Uniqueness Conditions and Non-Convex Optimization. *SIAM Journal on Numerical Analysis*, **51**, 858-884.

-
- [5] Tropp, J.A. (2006) Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise. *IEEE Transactions on Information Theory*, **52**, 1030-1051. <https://doi.org/10.1109/tit.2005.864420>
- [6] Candès, E. and Tao, T. (2007) Rejoinder: The Dantzig Selector: Statistical Estimation When P Is Much Larger than N. *The Annals of Statistics*, **35**, 2313-2351. <https://doi.org/10.1214/009053607000000532>
- [7] Zhao, Z. and Chen, G. (2014) A New Method for Signal Reconstruction of L_p -Norm Optimization. *Advances in Applied Mathematics*, **3**, 140-148. <https://doi.org/10.12677/aam.2014.33021>
- [8] Fornasier, M. and Rauhut, H. (2008) Iterative Thresholding for Sparse Approximation. *Constructive Approximation*, **28**, 307-336.
- [9] Jacques, L., Pesquet, J.-C. and He, H. (2014) A Plug-and-Play Approach to Sparse Signal Recovery with Application to DOA Estimation. *IEEE Transactions on Signal Processing*, **62**, 273-286.
- [10] Donoho, D.L. (2006) High Dimensional Sparse Signal Recovery via l_1 Minimization. *Proceedings of the National Academy of Sciences*, **102**, 9446-9451.
- [11] Bertsekas, D.P. (2016) *Nonlinear Programming*. 3rd Edition, Athena Scientific, Nashua.
- [12] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, **2**, 183-202. <https://doi.org/10.1137/080716542>
- [13] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (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>
- [14] Tseng, P. (2001) Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization. *Journal of Optimization Theory and Applications*, **109**, 475-494. <https://doi.org/10.1023/a:1017501703105>
- [15] Holland, J.H. (1975) *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, Ann Arbor.
- [16] Nesterov, Y. (2004) *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Dordrecht.
- [17] Boţ, R.I., Csetnek, E.R. and Nguyen, D. (2019) A Proximal Minimization Algorithm for Structured Nonconvex and Nonsmooth Problems. *SIAM Journal on Optimization*, **29**, 1300-1328. <https://doi.org/10.1137/18m1190689>