

带有绝对值形式的非线性约束下的交替方向乘子法

叶景然*, 陈跨越#, 杨 硕, 苏映梅

云南师范大学数学学院, 云南 昆明

收稿日期: 2025年2月23日; 录用日期: 2025年3月16日; 发布日期: 2025年3月25日

摘 要

交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)广泛应用于线性约束优化问题, 无论是凸目标函数还是非凸目标函数, 其约束一般是线性约束。文章研究了具有绝对值形式的非线性约束的非凸极小化问题的ADMM。在假设相关函数满足Kurdyka-Lojasiewicz (KL)不等式的情况下, 证明了由ADMM生成的迭代子序列收敛于问题的一个临界点, 并根据一些数值例子说明了ADMM是可行的。

关键词

交替方向乘子法, 非线性约束, 绝对值, 收敛性

Alternating Direction Method of Multipliers under Nonlinear Constraints with Absolute Value Form

Jingran Ye*, Kuayue Chen#, Shuo Yang, Yingmei Su

School of Mathematics, Yunnan Normal University, Kunming Yunnan

Received: Feb. 23rd, 2025; accepted: Mar. 16th, 2025; published: Mar. 25th, 2025

Abstract

The Alternating Direction Method of Multipliers (ADMM) is widely used for linear constraint optimization problems, whether the objective function is convex or non-convex, with constraints generally

*第一作者。

#通讯作者。

being linear. This article studies the ADMM applied to non-convex minimization problems with non-linear constraints in the form of absolute values. Under the assumption that the relevant functions satisfy the Kurdyka-Lojasiewicz (KL) inequality, it is proved that the iterative subsequence generated by ADMM converges to a critical point of the problem. Additionally, some numerical examples are provided to illustrate the feasibility of ADMM.

Keywords

Alternating Direction Method of Multipliers, Nonlinear Constraints, Absolute Value, Convergence

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

近年来, Gabay 等[1]提出的交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)由于能够解决套索问题[2]、矩阵完成问题[3]等凸优化问题而得到了广泛的关注。许多学者对 ADMM 的收敛性进行了分析[4] [5]。此外, 部分学者还存在对 ADMM 的收敛速度的研究。例如, He 等[6]利用变分不等式给出了 ADMM 在遍历意义上的收敛速度为 $O(1/k)$ (k 表示迭代度量), 其研究问题为:

$$\min \{f_1(x_1) + f_2(x_2) \mid A_1x_1 + A_2x_2 - b = 0\}$$

并且 Davis 等[7]指出, 在遍历意义上的收敛速度是紧的。此外, 对于非遍历意义, He 等[8]表明 ADMM 的收敛速度为 $O(1/\sqrt{k})$, Davis 等[7]也指出它是紧密的。

ADMM 最初被用于求解具有线性约束的凸优化问题。目前, 为了得到更广泛的应用, ADMM 发展了两个方面: (1) 算法的推导。如通过线性化增强项和光滑目标函数[9]的线性化版本, 以及通过将线性 ADMM 与 Nesterov 加速度[10]相结合的 Fast-ADMM (FADMM)等。(2) 应用范围的扩展。一些研究人员已经将其扩展到更一般的约束, 包括线性等式约束、线性不等式约束和非线性约束等[11]。

非凸目标函数的 ADMM 是近年来研究的热点之一。例如, Wang 等[12]利用布雷格曼距离提出了近似 ADMM。Zhang 等[13]使用指数平均布雷格曼距离到所有历史迭代近似 ADMM。最近, Guo 等[14]使用 ADMM 用来求解线性约束下的非凸目标函数, 即

$$\min \{f(x) + g(y) \mid Ax + y - b = 0\} \quad (1)$$

其迭代格式为:

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

在本文中, 受式(1)的启发, 考虑了以下形式的具有非线性约束的优化问题:

$$\min \{f(x) + g(y) \mid Ax + |y| - b = 0\} \quad (2)$$

式中, $f: R^n \rightarrow R \cup \{+\infty\}$ 是一个下半连续函数, $g: R^m \rightarrow R$ 是一个连续可微的函数, 并且 ∇g 满足 $L > 0$

的利普希兹连续, $A \in \mathbb{R}^{n \times n}$ 是一个矩阵, $b \in \mathbb{R}^n$ 是一个向量, y 中每个元素非零。另外, $\mathcal{L}_\beta(\cdot)$ 表示为问题(2)的增广拉格朗日函数, 其定义为:

$$\mathcal{L}_\beta(x, y, \lambda) = f(x) + g(y) - \langle \lambda, (Ax + |y| - b) \rangle + \frac{\beta}{2} \|Ax + |y| - b\|^2 \quad (3)$$

式中, $\|\cdot\|$ 在表示 2-范数。

目前, 还没有研究表明可以将 ADMM 用于问题(2)。因此, 本文将使用 ADMM 来解决不同凸性假设下的问题(2)。我们将通过 Kurdyka-Lojasiewicz (KL) 不等式证明, 如果增广拉格朗日函数是 KL 函数, 那么 ADMM 生成的序列收敛到问题(2)的 Karush-Kuhn-Tucker (KKT) 点。最后用一些数值实验来说明其 ADMM 的有效性。

2. 预备知识

在本节中, 将回顾一些关于非凸非光滑问题下的次梯度微分的符号和相关概念[15]。 $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ 是点对集的映射, 其图被定为:

$$\text{Graph}(F) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in F(x)\}.$$

对任意的集合 $S \subseteq \mathbb{R}^n$, 任意的点 $x \in \mathbb{R}^n$, 若 S 非空, 点 x 到 S 的距离记为:

$$d(x, S) = \inf_{y \in S} \|y - x\|. \quad (4)$$

若 $S = \emptyset$, 则令 $d(x, S) = +\infty$ 。

定义 1 [8]: 若函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 在 x_0 处满足 $f(x_0) \leq \liminf_{x \rightarrow x_0} f(x)$, 则称函数 f 在 x_0 处下半连续。若 f 在每一点处均下半连续, 则称 f 为下半连续函数。

定义 2 [16]: 设函数 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 为适当下半连续函数:

(1) f 在 $x \in \text{dom} f$ 处的 Fréchet 次微分定义为

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

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

(2) f 在 $x \in \text{dom} f$ 处的极限次微分定义为

$$\partial f(x) = \left\{ u \in \mathbb{R}^n : \exists x^k \rightarrow x, f(x^k) \rightarrow f(x) \text{ and } u^k \in \hat{\partial} f(x^k) \rightarrow u \text{ as } k \rightarrow \infty \right\}.$$

注 1

(1) 极限次微分是通过在 x 附近的点取 Fréchet 次微分的极限得到的, 即定义 2 意味着 $\hat{\partial} f(x) \subseteq \partial f(x)$ 对于每个 $u \in \mathbb{R}^n$, 其中第一个集合是闭凸的, 而第二个集合仅是闭的。并不是在所有 $x \in \text{dom} f$ 处都有 Fréchet 次微分。

(2) $x \in \text{dom} f$ 为 f 的极小值的一个必要条件是 $0 \in \partial f(x)$ 。

(3) 设 $(x^k, x^*) \in \text{Graph}(\partial f)$ 为一个序列, 收敛到 (x, x^*) 。根据 $\partial f(x)$ 的定义, 如果 $f(x^k)$ 收敛到 $f(x)$ ($k \rightarrow +\infty$), 则 $(x, x^*) \in \text{Graph}(\partial f)$ 。

(4) 如果 $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ 是一个适当的下半连续函数, $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 是连续可微的, 那么对于任意 $x \in \text{dom} f$, 有 $\partial(f + g) = \partial f(x) + \nabla g(y)$ 。

引理 1 [17]: 设函数 $G(x, y) = f(x) + h(y)$, 其中 $f: R^n \rightarrow R \cup \{+\infty\}$ 和 $h: R^m \rightarrow R$ 都是适当下半连续函数。当 $\forall (x, y) \in \text{dom}G = \text{dom}f \times \text{dom}h$, 有

$$\partial G(x, y) = \partial_x G(x, y) \times \partial_y G(x, y).$$

定义 3 [17]: (KL 不等式) 设 $f: R^n \rightarrow R \cup \{+\infty\}$ 为一个适当的下半连续函数。对于 $-\infty < \eta_1 < \eta_2 \leq +\infty$,

$$[\eta_1 < f < \eta_2] = \{x \in R^n : \eta_1 < f(x) < \eta_2\}.$$

我们说函数 f 在 $x^* \in \text{dom}f$ 处具有 KL 性质, 如果存在 $\eta \in (0, +\infty]$, x^* 的一个邻域 U , 以及一个连续凹函数 $\varphi: [0, \eta) \rightarrow R_+$, 使得

- (1) $\varphi(0) = 0$;
- (2) φ 在 $(0, +\infty)$ 上连续可微且在 0 处连续;
- (3) $\forall s \in (0, \eta), \varphi'(s) > 0$;
- (4) 对于所有 x 在 $U \cap [f(x^*) < f < f(x^*) + \eta]$ 中, KL 不等式成立

$$\varphi'(f(x) - f(x^*))d(0, \partial f(x)) \geq 1.$$

定义 4 [18]: (一致 KL 性质) 设 Ω 为一个紧集, 函数 $f: R^n \rightarrow R \cup \{+\infty\}$ 是一个适当的下半连续函数。

- (1) 如果在给定点 $\bar{x} \in \text{dom}\partial f = \{x | \partial f(x) \neq \emptyset\}$ 处, 存在 $\varepsilon > 0, \eta > 0$, 和 $\varphi \in \Phi_\eta$ 使得对于所有 $\bar{x} \in \Omega$,

$$\{x \in R^n : d(x, \Omega) < \varepsilon\} \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$$

且以下不等式成立

$$\varphi'(f(x) - f(\bar{x}))d(0, \partial f(x)) \geq 1.$$

- (2) 如果 f 在 $\text{dom}\partial f$ 上的任意点都满足(1), 则 f 是一个 KL 函数。

定义 5 [19]: (梯度利普希茨连续) 给定一个可微函数 f , 对于所有 $x, y \in \text{dom}f$, 如果存在 $L > 0$, 使得

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

引理 2 [19]: (二次上界) 设可微函数 f 的定义域为 R^n , 并且 ∇f 是利普希茨连续的, L 为其利普希茨常数, 则 f 具有二次上界:

$$|f(y) - f(x) - \nabla f(x)^T(y - x)| \leq \frac{L}{2}\|y - x\|^2, \forall x, y \in \text{dom}f.$$

定义 6: (x^*, y^*, λ^*) 是问题(2)的增广拉格朗日函数 $\mathcal{L}_\rho(\cdot)$ 的一个临界点。则其满足

$$A^T \lambda^* \in \partial f(x^*); \nabla g(y^*) = \lambda^* \cdot \text{sign}(y^*); Ax^* + |y^*| - b = 0.$$

备注 1: 定义 6 中, 我们可以看“ $\cdot$ ”表示两个向量相乘得到一个新的向量。下文中的表示意义同上。

定理 1 [20]: (有限长度性质) 设函数 Ψ 是一个 KL 函数, 并且满足下面的假设:

- (1) $f: R^n \rightarrow (-\infty, +\infty]$ 和 $g: R^m \rightarrow (-\infty, +\infty]$ 均是适当下半连续函数, 并且 $\inf_{R^n \times R^m} \Psi > -\infty, \inf_{R^n} f > -\infty$ 以

及 $\inf_{R^m} g > -\infty$;

(2) $H: R^n \times R^m \rightarrow R$ 是个连续可微函数, 并且 ∇H 在有界集上是联合利普希茨连续的, 即对任意的 $\forall B_1 \times B_2 \subset R^n \times R^m, \exists L > 0, \forall (x_i, y_i) \in B_1 \times B_2, i = 1, 2$ 有

$$\left\| (\nabla_x H(x_1, y_1) - \nabla_x H(x_2, y_2), \nabla_y H(x_1, y_1) - \nabla_y H(x_2, y_2)) \right\| \leq L \left\| (x_1 - x_2, y_1 - y_2) \right\|.$$

则有以下结论成立:

(1) 序列 $\{z^k\}$ 的长度有限, 即

$$\sum_{k=1}^{\infty} \|z^{k+1} - z^k\| < +\infty;$$

(2) 序列 $\{z^k\}$ 收敛到 Ψ 的一个临界点 $z^* = (x^*, y^*)$ 。

备注 2: 续可根据临界点的性质, $\omega(z^0)$ 是一个非空紧集, 并且 Ψ 在 $\omega(z^0)$ 上连续。在定义 4 中, 设 $\Omega = \omega(z^0)$, 对于任意的 $k > l$ (l 是一个足够大的正整数),

$$\varphi'(\Psi(z^k) - \Psi(\bar{z}))d(0, \partial\Psi(z^k)) \geq 1.$$

备注 3: 定理 1 中 $\{z^k\}$ 的迭代序列是收敛的。不难看出, 问题(2)中 $\mathcal{L}_\beta(\cdot)$ 的临界点正是其 KKT 点。

引理 3: 设迭代序列 $\{y^k\}$ 收敛到 $\{y^*\}$, $y^k \in R^n$ 。则总存在 N 使得对于 $k > N$, 使得

$$\text{sign}(y^k) = \text{sign}(y^{k+1}).$$

3. 收敛性分析

在本节中, 假设问题(2)至少存在一个 KKT 点。此外, 假设问题(2)中的两个子问题的最小化问题都具有解。因此, 问题(2)的 ADMM 是定义良好的, 并且生成一个无限迭代序列 $\{(x^k, y^k, \lambda^k)\}$ 。求解问题(2)的 ADMM 设计如下:

$$\begin{cases} x^{k+1} \in \arg \min \{ \mathcal{L}_\beta(x, y^k, \lambda^k) \} \\ y^{k+1} \in \arg \min \{ \mathcal{L}_\beta(x^{k+1}, y, \lambda^k) \} \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + |y^{k+1}| - b). \end{cases} \quad (5)$$

根据问题(2)的优化条件, 有

$$\begin{cases} 0 \in \partial f(x^{k+1}) - A^T \lambda^k + \beta A^T(Ax^{k+1} + |y^k| - b) \\ 0 = \nabla g(y^{k+1}) - \lambda^k \cdot \text{sign}(y^{k+1}) + \beta(Ax^{k+1} + |y^{k+1}| - b) \cdot \text{sign}(y^{k+1}) \end{cases} \quad (6)$$

基于(6)和(5)中的第三个等式, 得到

$$\begin{cases} A^T \lambda^k + \beta A^T(Ax^{k+1} + |y^k| - b) \in \partial f(x^{k+1}) \\ \nabla g(y^{k+1}) = \lambda^{k+1} \cdot \text{sign}(y^{k+1}) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + |y^{k+1}| - b) \end{cases} \quad (7)$$

在本文中, 我们还做了以下假设:

假设 1: 设 $f: R^n \rightarrow R \cup \{+\infty\}$ 是一个适当的下半连续函数, $g: R^n \rightarrow R$ 是一个可微函数, 并且 ∇g 满足利普希茨连续, 并且 $L > 0$ 。假设 $\beta\sigma > 2L, \sigma \in [0, 1]$ 使得 $\delta = (\beta\sigma - 2L)/2 - L^2/\beta > 0$ 。

为了后续讨论, 需要以下引理:

引理 4: 设 (x^*, y^*, λ^*) 是 $\mathcal{L}_\beta(\omega^{k+1})$ 的一个临界点, 并且 y^* 中的每个元素均不为 0。设 $\{\omega^k\}_{k \geq 0} = \{(x^k, y^k, \lambda^k)\}_{k \geq 0}$ 为算法(5)生成的序列。假设假设 1 成立并且 $\{y^k\}$ 收敛到 y^* 。存在一个 N 当 $k > N$, 则有

$$\mathcal{L}_\beta(\omega^{k+1}) \leq \mathcal{L}_\beta(\omega^k) - \delta \|y^{k+1} - y^k\|^2. \quad (8)$$

证明: 基于增广拉格朗日函数 $\mathcal{L}_\beta(\omega)$ 的定义, 有

$$\begin{aligned} \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}) &= \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) + \langle \lambda^k - \lambda^{k+1}, Ax^{k+1} + |y^{k+1}| - b \rangle \\ &= \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) + \frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|^2 \end{aligned} \quad (9)$$

和

$$\begin{aligned} &\mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) \\ &= f(x^{k+1}) + g(y^k) - \langle \lambda^k, Ax^{k+1} + |y^k| - b \rangle + \frac{\beta}{2} \|Ax^{k+1} + |y^k| - b\|^2 \\ &\quad - \left\{ f(x^{k+1}) + g(y^{k+1}) - \langle \lambda^k, Ax^{k+1} + |y^{k+1}| - b \rangle + \frac{\beta}{2} \|Ax^{k+1} + |y^{k+1}| - b\|^2 \right\} \\ &= g(y^k) - g(y^{k+1}) + \langle \lambda^k, |y^{k+1}| - |y^k| \rangle + \frac{\beta}{2} \|Ax^{k+1} + |y^k| - b\|^2 - \frac{\beta}{2} \|Ax^{k+1} + |y^{k+1}| - b\|^2. \end{aligned} \quad (10)$$

由于 ∇g 满足 $L > 0$ 的利普希茨连续, 根据引理 2 和式(7)中的第二个等式, 可得:

$$g(y^k) \geq g(y^{k+1}) + \lambda^{k+1T} \text{sign}(y^{k+1})(y^k - y^{k+1}) - \frac{L}{2} \|y^k - y^{k+1}\|^2. \quad (11)$$

将(11)带入到(10)中可得:

$$\begin{aligned} \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^k) &\geq \langle \lambda^{k+1} \cdot \text{sign}(y^{k+1}), y^k - y^{k+1} \rangle + \frac{\beta}{2} \|Ax^{k+1} + |y^k| - b\|^2 \\ &\quad - \frac{\beta}{2} \|Ax^{k+1} + |y^{k+1}| - b\|^2 - \langle \lambda^k, |y^k| - |y^{k+1}| \rangle - \frac{L}{2} \|y^k - y^{k+1}\|^2. \end{aligned} \quad (12)$$

回忆

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + |y^{k+1}| - b),$$

即

$$Ax^{k+1} + |y^k| - b = \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) + (|y^k| - |y^{k+1}|).$$

然后

$$\begin{aligned} \frac{\beta}{2} \|Ax^{k+1} + |y^k| - b\|^2 &= \frac{\beta}{2} \left\| \frac{1}{\beta}(\lambda^k - \lambda^{k+1}) + (|y^k| - |y^{k+1}|) \right\|^2 \\ &= \frac{1}{2\beta} \|\lambda^k - \lambda^{k+1}\|^2 + \frac{\beta}{2} \| |y^k| - |y^{k+1}| \|^2 + \langle \lambda^k - \lambda^{k+1}, |y^k| - |y^{k+1}| \rangle \end{aligned} \quad (13)$$

组合(12)和(13), 可得

$$\begin{aligned} \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - L_\beta(x^{k+1}, y^{k+1}, \lambda^k) &\geq \langle \lambda^{k+1} \cdot \text{sign}(y^{k+1}), y^k - y^{k+1} \rangle - \langle \lambda^k, |y^k| - |y^{k+1}| \rangle \\ &\quad + \frac{\beta}{2} \| |y^k| - |y^{k+1}| \|^2 + \langle \lambda^k - \lambda^{k+1}, |y^k| - |y^{k+1}| \rangle - \frac{L}{2} \| y^k - y^{k+1} \|^2. \end{aligned} \quad (14)$$

而

$$\begin{aligned} &\langle \lambda^{k+1} \cdot \text{sign}(y^{k+1}), y^k - y^{k+1} \rangle - \langle \lambda^k, |y^k| - |y^{k+1}| \rangle + \langle \lambda^k - \lambda^{k+1}, |y^k| - |y^{k+1}| \rangle \\ &= \langle \lambda^{k+1}, y^k \cdot \text{sign}(y^{k+1}) - |y^{k+1}| \rangle - \langle \lambda^k, |y^k| - |y^{k+1}| \rangle + \langle \lambda^k - \lambda^{k+1}, |y^k| - |y^{k+1}| \rangle \\ &= \langle \lambda^{k+1}, y^k \cdot \text{sign}(y^{k+1}) \rangle - \langle \lambda^{k+1}, |y^{k+1}| \rangle - \langle \lambda^{k+1}, |y^k| \rangle + \langle \lambda^{k+1}, |y^{k+1}| \rangle \\ &= \langle \lambda^{k+1}, y^k \cdot \text{sign}(y^{k+1}) \rangle - \langle \lambda^{k+1}, |y^k| \rangle \\ &= \langle \lambda^{k+1}, y^k \cdot \text{sign}(y^{k+1}) \rangle - \langle \lambda^{k+1}, y^k \cdot \text{sign}(y^k) \rangle \\ &= \langle \lambda^{k+1}, y^k \cdot (\text{sign}(y^{k+1}) - \text{sign}(y^k)) \rangle. \end{aligned} \quad (15)$$

从式(15)中容易看出, 存在一个 N , 当 $k > N$ 时, 根据引理 3, 可得:

$$\langle \lambda^{k+1} \cdot \text{sign}(y^{k+1}), y^k - y^{k+1} \rangle - \langle \lambda^k, |y^k| - |y^{k+1}| \rangle + \langle \lambda^k - \lambda^{k+1}, |y^k| - |y^{k+1}| \rangle = 0.$$

根据

$$\| |y^k| - |y^{k+1}| \|^2 = \sigma_k \| y^k - y^{k+1} \|^2, \sigma_k \in [0, 1]$$

可得:

$$\begin{aligned} \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - L_\beta(x^{k+1}, y^{k+1}, \lambda^k) &\geq \frac{\beta}{2} \| |y^k| - |y^{k+1}| \|^2 - \frac{L}{2} \| y^k - y^{k+1} \|^2 \\ &= \frac{\beta \sigma_k}{2} \| y^k - y^{k+1} \|^2 - \frac{L}{2} \| y^k - y^{k+1} \|^2 \\ &= \frac{\beta \sigma_k - L}{2} \| y^k - y^{k+1} \|^2. \end{aligned} \quad (16)$$

利用定义 5 和等式(7)当中的第二个等式, 可得:

$$\| \lambda^{k+1} \cdot \text{sign}(y^{k+1}) - \lambda^k \cdot \text{sign}(y^k) \| \leq L \| y^{k+1} - y^k \|. \quad (17)$$

根据式(17), 可以知道存在一个利普希茨常数 L , 使得:

$$\| \lambda^{k+1} \pm \lambda^k \| \leq L \| y^{k+1} - y^k \|. \quad (18)$$

组合(9)、(16)和(18)可得:

$$\begin{aligned} \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}) &\leq \mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) - \left(\frac{\beta \sigma_k - L}{2} - \frac{L^2}{\beta} \right) \| y^k - y^{k+1} \|^2 \\ &\leq \mathcal{L}_\beta(x^k, y^k, \lambda^k) - \left(\frac{\beta \sigma_k - L}{2} - \frac{L^2}{\beta} \right) \| y^k - y^{k+1} \|^2. \end{aligned}$$

其中, 上述第二个不等式满足 x^{k+1} 是 $\mathcal{L}_\beta(x, y^k, \lambda^k)$ 中关于 x 的最小值, 即

$$\mathcal{L}_\beta(x^{k+1}, y^k, \lambda^k) \leq \mathcal{L}_\beta(x^k, y^k, \lambda^k).$$

令 $\sigma = \inf \{ \sigma_k \}$, 则

$$\delta = \frac{\beta\sigma - 2L}{2} - \frac{L^2}{\beta}.$$

因此, 引理 4 成立。

备注 4: 由于 $\sigma = \inf \{\sigma_k\}$ 和 $\sigma_k \in (0, 1]$, 则说明 $\sigma \geq 0$ 。然而对于任意的 k , 都有 $0 < \sigma_k \leq 1$, 这就意味着 $\sigma = \inf_{k \geq 0} \{\sigma_k\} = \min_{k \geq 0} \{\sigma_k\} > 0$ 。因此, 存在 $\xi = \frac{\min_{k \geq 0} \{\sigma_k\}}{2}$ 使得 $\sigma > \xi > 0$ 。

备注 5: 由 $\delta > 0$, 根据引理 3, $\mathcal{L}_\beta(\omega^k)$ 单调递减。

接下来, 我们给出一些充分条件来保证算法(5)生成的序列 $\{\omega^k = (x^k, y^k, \lambda^k)\}$ 是有界的。

引理 5: 设 $\{\omega^k\} = \{(x^k, y^k, \lambda^k)\}$ 是算法(5)生成的序列, 并假设

$$\inf_y \{g(y) - \tau \|\nabla g(y)\|^2\} =: \bar{g} > -\infty.$$

那么以下陈述是正确的。

(1) f 是强制性的, 即 $\liminf_{\|x\|^2 \rightarrow +\infty} f(x) = +\infty$;

(2) 存在常数 $\tau = \frac{1}{2L}$ 。

证明: 根据引理 4, 可得:

$$\mathcal{L}_\beta(x^k, y^k, \lambda^k) \leq \mathcal{L}_\beta(x^0, y^0, \lambda^0).$$

然后, 结合 $\lambda^k \cdot \text{sign}(y^k) = \nabla g(y^k)$, 可得:

$$\begin{aligned} \mathcal{L}_\beta(x^0, y^0, \lambda^0) &\geq f(x^k) + g(y^k) - \langle \lambda^k, (Ax^k + |y^k| - b) \rangle + \frac{\beta}{2} \|Ax^k + |y^k| - b\|^2 \\ &= f(x^k) + g(y^k) - \frac{1}{2\beta} \|\lambda^k\|^2 + \frac{\beta}{2} \left\| Ax^k + |y^k| - b - \frac{1}{\beta} \lambda^k \right\|^2 \\ &= f(x^k) + \left(g(y^k) - \frac{1}{2L} \|\nabla g(y^k)\|^2 \right) + \left(\frac{1}{2L} - \frac{1}{2\beta} \right) \|\lambda^k\|^2 + \frac{\beta}{2} \left\| Ax^k + |y^k| - b - \frac{1}{\beta} \lambda^k \right\|^2 \\ &\geq f(x^k) + \bar{g} + \left(\tau - \frac{1}{2\beta} \right) \|\lambda^k\|^2 + \frac{\beta}{2} \left\| Ax^k + |y^k| - b - \frac{1}{\beta} \lambda^k \right\|^2. \end{aligned}$$

观察陈述(1)意味着 $\inf_x f(x) > -\infty$ 。由 $\beta > 2L$ 得到 $\{x^k\}, \{\lambda^k\}$ 和 $\left\{ \frac{\beta}{2} \left\| Ax^k + |y^k| - b - \frac{1}{\beta} \lambda^k \right\|^2 \right\}$ 是有界的。因此, $\{y^k\}$ 是有界的, 从而, $\{\omega^k\}$ 是有界的。

引理 6: 设 $\{\omega^k\} = \{(x^k, y^k, \lambda^k)\}$ 是算法(5)生成的序列, 且 $\{\omega^k\}$ 是有界的。那么,

$$\sum_{k=0}^{+\infty} \|\omega^{k+1} - \omega^k\|^2 < +\infty. \quad (19)$$

证明: 为了证明(19)成立, 考虑三种情况:

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

首先, 证明 $\sum_{k=0}^{\infty} \|y^{k+1} - y^k\|^2 < +\infty$ 。由于 $\{\omega^k\}$ 是有界的, 它至少有一个聚点。设 $\{\omega^*\}$ 是 $\{\omega^k\}$ 的一个

聚点, $\{\omega^{k^j}\}$ 是收敛到它的子序列, 即 $\omega^{k^j} \rightarrow \omega^*$ 。由于 f 是下半连续的且 g 是连续的, $\mathcal{L}_\beta(\cdot)$ 是下半连续的, 可得:

$$\mathcal{L}_\beta(\omega^*) \leq \liminf_{j \rightarrow +\infty} \mathcal{L}_\beta(\omega^{k^j}).$$

由于 $\{\mathcal{L}_\beta(\omega^{k^j})\}$ 收敛, 所以 $\{\mathcal{L}_\beta(\omega^{k^j})\}$ 有界。此外, $\{\mathcal{L}_\beta(\omega^k)\}$ 是收敛的, 并且 $\{\mathcal{L}_\beta(\omega^k)\} \geq \{\mathcal{L}_\beta(\omega^*)\}$ 。根据式(8), 可得:

$$\delta \|y^{k+1} - y^k\|^2 \leq L_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}). \quad (20)$$

根据式(20), 可得:

$$\sum_{k=0}^n \delta \|y^{k+1} - y^k\|^2 \leq L_\beta(\omega^0) - L_\beta(\omega^{n+1}) \leq L_\beta(\omega^0) - \mathcal{L}_\beta(\omega^*) < +\infty.$$

由于 $\delta > 0$, 有:

$$\sum_{k=0}^{\infty} \|y^{k+1} - y^k\|^2 < +\infty.$$

其次, 证明 $\sum_{k=0}^{\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty$ 。这显然来自于方程(18)。

最后, 证明 $\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < +\infty$ 。回忆

$$\lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + |y^{k+1}| - b), \lambda^k = \lambda^{k-1} - \beta(Ax^k + |y^k| - b).$$

那么

$$\beta(Ax^k - Ax^{k+1}) = (\lambda^{k+1} - \lambda^k) - (\lambda^k - \lambda^{k-1}) - \beta(|y^k| - |y^{k+1}|).$$

所以

$$\begin{aligned} \|\beta(Ax^k - Ax^{k+1})\|^2 &= \|(\lambda^{k+1} - \lambda^k) - (\lambda^k - \lambda^{k-1}) - \beta(|y^k| - |y^{k+1}|)\|^2 \\ &\leq 3\left(\|\lambda^{k+1} - \lambda^k\|^2 + \|\lambda^k - \lambda^{k-1}\|^2 + \beta^2 \|y^k - y^{k+1}\|^2\right) \\ &= 3\left(\|\lambda^{k+1} - \lambda^k\|^2 + \|\lambda^k - \lambda^{k-1}\|^2 + \beta^2 \sigma \|y^k - y^{k+1}\|^2\right). \end{aligned} \quad (21)$$

根据 $|A|_2 = \sqrt{\lambda_{\max}(A^T A)}$, 有

$$\|\beta(Ax^k - Ax^{k+1})\|^2 = \beta^2 \lambda_{\max} \|x^{k+1} - x^k\|^2. \quad (22)$$

然后, 根据公式(21)和公式(22), 可以看出

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < +\infty.$$

因此, 引理 6 的结果成立。

备注 4: 如果 $\mathcal{L}_\beta(\cdot)$ 于是有界的, 那么容易推 $\sum_{k=0}^{+\infty} \|\omega^{k+1} - \omega^k\|^2 < +\infty$ 而无需保证 $\{\omega^k\}$ 的有界性。

引理 7: 设 $\{\omega^k\} = \{(x^k, y^k, \lambda^k)\}$ 是算法(5)生成的序列, 且 $\{\omega^k\}$ 是有界的。定义

$$x_{k+1}^* = \beta A^T (|y^{k+1}| - |y^k|) + A^T (\lambda^k - \lambda^{k+1}); y_{k+1}^* = \lambda^k - \lambda^{k+1}; \lambda_{k+1}^* = \frac{1}{\beta} (\lambda^k - \lambda^{k+1}).$$

则有

$$(x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^*) \in \partial \mathcal{L}_\beta(\omega^{k+1}).$$

此外, 存在 $\zeta > 0$, 使得

$$d(0, \partial \mathcal{L}_\beta(\omega^{k+1})) \leq \zeta \|y^k - y^{k+1}\|.$$

证明: 根据增广拉格朗日函数 $\mathcal{L}_\beta(\cdot)$, 我们可知

$$\begin{aligned} \partial_x \mathcal{L}_\beta(\omega^{k+1}) &= \partial f(x^{k+1}) - A^T y^{k+1} \lambda^{k+1} + \beta A^T y^{k+1} (x^{k+1T} A^T y^{k+1} - b), \\ \partial_y \mathcal{L}_\beta(\omega^{k+1}) &= \nabla g(y^{k+1}) - x^{k+1T} A^T \lambda^{k+1} + \beta (x^{k+1T} A^T y^{k+1} - b), \\ \partial_\lambda \mathcal{L}_\beta(\omega^{k+1}) &= -(x^{k+1T} A^T y^{k+1} - b). \end{aligned}$$

下再结合(7), 可得

$$\begin{aligned} \beta A^T (|y^{k+1}| - |y^k|) + A^T (\lambda^k - \lambda^{k+1}) &\in \partial_x \mathcal{L}_\beta(\omega^{k+1}); \\ (\lambda^k - \lambda^{k+1}) \cdot \text{sign}(y^{k+1}) &\in \partial_y \mathcal{L}_\beta(\omega^{k+1}); \\ \frac{1}{\beta} (\lambda^k - \lambda^{k+1}) &\in \partial_\lambda \mathcal{L}_\beta(\omega^{k+1}). \end{aligned}$$

因此, 根据引理 1, 我们可知 $(x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^*) \in \partial \mathcal{L}_\beta(\omega^{k+1})$ 。除此之外, 存在 $\zeta_1, \zeta_2 > 0$ 使得

$$\|(x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^*)\| \leq \zeta_1 \sigma \|y^{k+1} - y^k\| + \zeta_2 \|\lambda^{k+1} - \lambda^k\|.$$

连通过定义 $\zeta = \zeta_1 \sigma + L \zeta_2$, 再根据(18)可知

$$d(0, \partial \mathcal{L}_\beta(\omega^{k+1})) \leq \|(x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^*)\| \leq \zeta \|y^k - y^{k+1}\|.$$

这就完成了此引理的证明。

引理 8: 设 $\{\omega^k\} = \{(x^k, y^k, \lambda^k)\}$ 是算法(5)生成的序列, 且 $\{\omega^k\}$ 是有界的。令 $S(\omega^0)$ 表示其极限点的集合。那么:

- (1) $\emptyset \neq S(\omega^0) \subset \text{crit} \mathcal{L}_\beta$;
- (2) $\mathcal{L}_\beta(\cdot)$ 在 $S(\omega^0)$ 上是有限的且为常数;
- (3) $S(\omega^0)$ 是一个非空连通的紧集, 并且

$$\lim_{k \rightarrow \infty} d(\omega^k, S(\omega^0)) = 0.$$

定理 2: 设 $\{\omega^k\} = \{(x^k, y^k, \lambda^k)\}$ 是算法(5)生成的序列, 且 $\{\omega^k\}$ 是有界的。假设 f 和 g 是半代数函数, 那么 ω^k 具有有限长度, 即

$$\sum_{k=0}^{+\infty} \|\omega^{k+1} - \omega^k\| < +\infty$$

因此, ω^k 收敛到 $\mathcal{L}_\beta(\cdot)$ 的一个临界点。

证明: 从引理 8 可知, 对于所有 $\omega^* \in S(\omega^0)$, 有 $\mathcal{L}_\beta(\omega^k) \rightarrow \mathcal{L}_\beta(\omega^*)$, 考虑两种情况:

(1) 如果存在一个整数 k_0 , 使得 $\mathcal{L}_\beta(\omega^{k_0}) = \mathcal{L}_\beta(\omega^*)$ 。根据方程(8)以及备注 3, 对于任意 $k > k_0$, 有

$$\delta \|y^{k+1} - y^k\|^2 \leq L_\beta(\omega^k) - L_\beta(\omega^{k+1}) \leq L_\beta(\omega^{k_0}) - \mathcal{L}_\beta(\omega^*) = 0。$$

对于任意 $k > k_0$, 有 $y^{k+1} = y^k$ 。结合引理 5 的证明, 对于任意 $k > k_0 + 1$, 有 $\omega^{k+1} = \omega^k$, 因此定理 2 成立。

(2) 假设对于所有 k 都有 $\mathcal{L}_\beta(\omega^k) > \mathcal{L}_\beta(\omega^*)$ 。由于 $d(\omega^k, S(\omega^0)) \rightarrow 0$, 对于任意 $\varepsilon > 0$, 存在 $k_1 > 0$, 使得对于任何 $k > k_1$, 有 $d(\omega^k, S(\omega^0)) < \varepsilon$ 。再次, 由于 $\mathcal{L}_\beta(\omega^k) \rightarrow \mathcal{L}_\beta(\omega^*)$, 对于任意 $\eta > 0$, 存在 $k_2 > 0$, 使得对于任意 $k > k_2$, 有 $\mathcal{L}_\beta(\omega^k) < \mathcal{L}_\beta(\omega^*) + \eta$ 。因此, 对于所有 $\varepsilon, \eta > 0$, 当 $k > \tilde{k} = \max\{k_1, k_2\}$ 时,

$$d(\omega^k, S(\omega^0)) < \varepsilon, \mathcal{L}_\beta(\omega^k) < \mathcal{L}_\beta(\omega^*) + \eta,$$

由于 $S(\omega^0)$ 是一个非空紧集, 并且 $\mathcal{L}_\beta(\cdot)$ 在 $S(\omega^0)$ 上是常数, 应用定义 4 中 $\Omega = S(\omega^0)$, 推导出对于任意 $k > \tilde{k}$, 有

$$\varphi'(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) d(0, \partial \mathcal{L}_\beta(\omega^k)) \geq 1。$$

由于

$$\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}) = (\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) - (\mathcal{L}_\beta(\omega^{k+1}) - \mathcal{L}_\beta(\omega^*))。$$

根据 φ 的凹性, 可得

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

关联 $d(0, \partial \mathcal{L}_\beta(\omega^k)) \leq \zeta \|y^k - y^{k-1}\|, \varphi'(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) > 0$, 得到

$$\begin{aligned} \mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}) &\leq \frac{\varphi(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) - \varphi(\mathcal{L}_\beta(\omega^{k+1}) - \mathcal{L}_\beta(\omega^*))}{\varphi(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*))} \\ &\leq \zeta \|y^k - y^{k-1}\| \left[\varphi(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) - \varphi(\mathcal{L}_\beta(\omega^{k+1}) - \mathcal{L}_\beta(\omega^*)) \right]。 \end{aligned} \quad (23)$$

设

$$\Delta_{p,q} = \varphi(\mathcal{L}_\beta(\omega^p) - \mathcal{L}_\beta(\omega^*)) - \varphi(\mathcal{L}_\beta(\omega^q) - \mathcal{L}_\beta(\omega^*))。$$

结合引理 3 和不等式(23), 得到对于所有 $k > \tilde{k}$,

$$\|y^{k+1} - y^k\| \leq \sqrt{\frac{\zeta}{\delta} \Delta_{k,k+1}} \|y^k - y^{k-1}\| \leq \frac{1}{2} \left(\|y^k - y^{k-1}\| + \frac{\zeta}{\delta} \Delta_{k,k+1} \right)。 \quad (24)$$

从方程(24)可以得出

$$\sum_{k=\tilde{k}+1}^m \|y^{k+1} - y^k\| \leq \frac{1}{2} \left(\sum_{k=\tilde{k}+1}^m \|y^k - y^{k-1}\| + \frac{\zeta}{\delta} \Delta_{\tilde{k}+1,m+1} \right)。 \quad (25)$$

注意到 $\varphi(\mathcal{L}_\beta(\omega^{m+1}) - \mathcal{L}_\beta(\omega^*)) > 0$, 重写(25), 并令 $m \rightarrow +\infty$, 有

$$\sum_{k=\bar{k}+1}^{+\infty} \|y^{k+1} - y^k\| \leq \|y^{\bar{k}+1} - y^{\bar{k}}\| + \frac{\zeta}{\delta} \varphi(\mathcal{L}_\beta(\omega^{\bar{k}+1}) - \mathcal{L}_\beta(\omega^*)). \quad (26)$$

这意味着

$$\sum_{k=0}^{+\infty} \|y^{k+1} - y^k\| < +\infty.$$

从方程(18), 知道

$$\sum_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\| < +\infty.$$

另一方面, 从方程(21)和方程(22)可以得出

$$\begin{aligned} \|x^{k+1} - x^k\| &\leq \sqrt{\frac{3}{\beta^2 \lambda_{\max}(A^T A)}} \left(\|\lambda^{k+1} - \lambda^k\|^2 + \|\lambda^k - \lambda^{k-1}\|^2 + \beta^2 \|y^k - y^{k+1}\|^2 \right)^{1/2} \\ &\leq \sqrt{\frac{3}{\beta^2 \lambda_{\max}(A^T A)}} \left(\|\lambda^{k+1} - \lambda^k\| + \|\lambda^k - \lambda^{k-1}\| + \beta \sqrt{\sigma} \|y^k - y^{k+1}\| \right). \end{aligned}$$

且 $\{\omega^k\}$ 是一个柯西序列(参见 Botel 等[18]的第 482 页的一个简单证明)。因此, 根据备注 3 可知它是收敛的。

4. 数值实验分析

例 1: 二次规划(Quadratic Programming, QP)问题: 二次规划(QP)是一个数学优化问题, 其目标是求解满足一组线性或不等式约束的二次函数的最小值或最大值。在优化问题中, QP 问题是一个非常重要的问题, 因为它广泛应用于金融、工程、计算机视觉、机器学习等领域。QP 问题的约束通常是线性约束, 将其扩展到非线性约束问题, 并通过以下例子来说明

$$\begin{aligned} \min & (x-1)^2 + (y-3)^2 \\ \text{s.t. } & 2x + |y| - 4 = 0 \\ & -7 \leq x \leq 5, -8 \leq y \leq 1. \end{aligned}$$

上述问题的增广拉格朗日函数为

$$\mathcal{L}_\beta(x, y, \lambda) = f(x) + g(y) - \langle \lambda, 2x + |y| - 4 \rangle + \frac{\beta}{2} \|2x + |y| - 4\|^2.$$

在此实验中, 每个问题的子问题根据(5)进行求解。然后, 列出了(5)的每次迭代过程中的子 x 问题和 y 子问题:

x 子问题为

$$x^{k+1} = \frac{2[\lambda^k - \beta(2|y^k| - 4)] + 2}{2 + 4\beta}.$$

y 子问题为

$$y^{k+1} = \frac{\text{sign}(y^k)[\lambda^k - \beta(2x^{k+1} - 4)] + 6}{2 + \beta}.$$

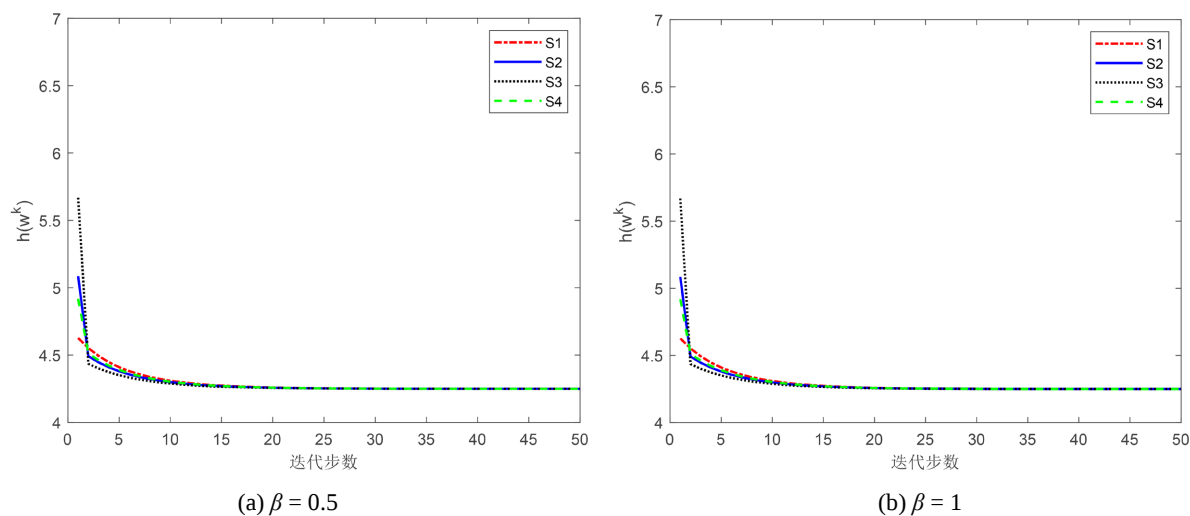


Figure 1. The relationship between function value and the number of iterations
图 1. 函数值与迭代次数的关系

Table 1. Iteration results with different initial values
表 1. 不同初始值下的迭代结果

x	1	-2	-5	7
y	1	-2	-5	-1
(x^*, y^*)	(1.5, 1.0)	(1.5, 1.0)	(1.5, 1.0)	(1.5, 1.0)
β	1	1	1	1
IT	49	48	47	48
CPU	5.691	6.238	6.274	5.346
β	0.5	0.5	0.5	0.5
IT	49	48	47	48
CPU	5.657	5.561	6.120	4.367

在计算中，令 $\mathcal{L}_\beta(x, y, \lambda)$ 中的 $\lambda = 1.75$ ，分别测试了四组 (x, y) 的初始值：S1: $(x, y) = (1, 1)$ ；S2: $(x, y) = (-2, -2)$ ；S3: $(x, y) = (-5, -5)$ ；S4: $(x, y) = (7, -1)$ ，分别取 $\beta = 0.5$ 和 $\beta = 1$ 。在这种情况下，图 1 和表 1 中显示了一些数值结果。从图 1 中可以看出，随着迭代次数的增加，函数值逐渐减少，最终达到平稳状态。通过表 1 可以看出，不同初始值达到最优解的迭代次数是不同的。在 x 和 y 的定义域下，发现 β 的变化不会影响最优解的变化，且最优解都是 $(x^*, y^*) = (1.5, 1.0)$ 。不出所料，随着参数 β 的增加，CPU 时间变长。

例 2: 矩阵 $A \in R^{n \times n}$ ，向量 $b \in R^n$ 。考虑问题(2)具体的函数形式为 $f(x) = \|x - 1\|^2$, $g(y) = \|y - 1\|^2$ ，即

$$\begin{aligned} & \min \|x - 1\|^2 + \|y - 1\|^2, \\ & \text{s.t. } Ax + |y| - b = 0, \end{aligned} \tag{27}$$

且

$$A = [a_{ij}] = \begin{cases} 10, & i = j. \\ -7, & j - i = 1, i, j = (1, 2, \dots, n), \\ 2, & i - j = 1. \end{cases} b = (0, 0, \dots, 0, 0)^\top.$$

上述问题的增广拉格朗日函数为

$$\mathcal{L}_\beta(x, y, \lambda) = \|x-1\|^2 + \|y-1\|^2 - \langle \lambda, Ax + |y| - b \rangle + \frac{\beta}{2} \|Ax + |y| - b\|^2.$$

在此实验中，每个问题的子问题根据(5)进行求解。在求解过程中，我们引入变量 $z = |y|$ ，则问题(27)可改写为

$$\min \left\{ \|x-1\|^2 + \|y-1\|^2 \mid Ax + z - b = 0, z = |y| \right\}.$$

对应的子问题为：

x 子问题为

$$x^{k+1} = (A^\top A + \beta I)^{-1} (A^\top (b - z^k) + \lambda^k).$$

y 子问题为

$$y^{k+1} = \text{sign}(y^k) \cdot \max \left(|z^k| - \frac{1}{\beta}, 0 \right).$$

z 相当于约束投影，其子问题为

$$z^{k+1} = y^{k+1} + (b + Ax^{k+1}).$$

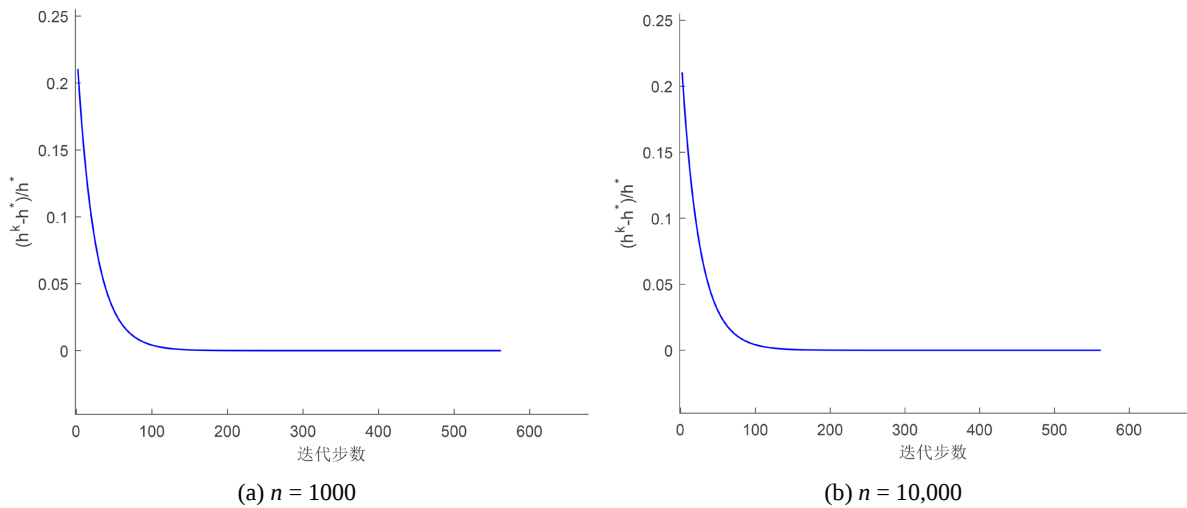


Figure 2. The relationship between $(h(w^k) - h^*)/h^*$ and the number of iterations

图 2. $(h(w^k) - h^*)/h^*$ 与迭代次数的关系

在实验中，设定 $\beta = 1$ ，初始向量为： $x_0 = (1, 1, \dots, 1, 1)^\top$, $y_0 = (1, 1, \dots, 1, 1)^\top$, $\lambda_0 = (1, 1, \dots, 1, 1)^\top$ 。在这种情况下，我们在图 2 中展示了当 $n = 1000$ 和 $n = 10,000$ 时的数值结果。很容易发现，当 $n = 1000$ 时，在迭代 562 次后得到最优解，而当 $n = 10,000$ 时，在迭代 565 次后得到最优解。虽然维数增加迭代次数也增多，但维数的变化对达到最优解的迭代次数变化的影响不是很显著。通过图 2，我们可以看到横坐标和纵坐标反映了迭代次数与 $(h(w^k) - h^*)/h^*$ 的关系 (h^* 代表函数的最优解)。随着迭代次数的增加， $(h(w^k) - h^*)/h^*$ 的值是递减的，并最终保持为平稳状态。这就表明 $h(w^k)$ 是非递增的，并最终收敛到 h^* 。

5. 结论

本文通过使用 KL 不等式, 研究了带有绝对值形式非线性约束的最小化问题的 ADMM, 并证明了 ADMM 生成的迭代子序列收敛到问题的一个关键点。更重要的是, 通过数值例子, 进一步验证了本文设计的 ADMM 在带有绝对值形式的非线性约束的最小化问题中是实用的。值得注意的是, 文章仅考虑了这种情况下 y 的每个元素都非零的情况, 而包含零元素的 y 的另一种情况则有待进一步研究。

基金项目

云南师范大学研究生科研训练基金项目(YJSJJ23-B69)。

参考文献

- [1] 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)
- [2] Yuan, M. and Lin, Y. (2005) Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **68**, 49-67. <https://doi.org/10.1111/j.1467-9868.2005.00532.x>
- [3] Han, D. (2022) A Survey on Some Recent Developments of Alternating Direction Method of Multipliers. *Journal of the Operations Research Society of China*, **10**, 1-52. <https://doi.org/10.1007/s40305-021-00368-3>
- [4] Eckstein, J. and Bertsekas, D.P. (1992) On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators. *Mathematical Programming*, **55**, 293-318. <https://doi.org/10.1007/bf01581204>
- [5] Boley, D. (2013) Local Linear Convergence of the Alternating Direction Method of Multipliers on Quadratic or Linear Programs. *SIAM Journal on Optimization*, **23**, 2183-2207. <https://doi.org/10.1137/120878951>
- [6] He, B. and Yuan, X. (2012) On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method. *SIAM Journal on Numerical Analysis*, **50**, 700-709. <https://doi.org/10.1137/110836936>
- [7] Davis, D. and Yin, W. (2016) Convergence Rate Analysis of Several Splitting Schemes. In: Glowinski, R., Osher, S. and Yin, W., Eds., *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer, 115-163. https://doi.org/10.1007/978-3-319-41589-5_4
- [8] He, B. and Yuan, X. (2014) On Non-Ergodic Convergence Rate of Douglas-Rachford Alternating Direction Method of Multipliers. *Numerische Mathematik*, **130**, 567-577. <https://doi.org/10.1007/s00211-014-0673-6>
- [9] He, B., Liao, L., Han, D. and Yang, H. (2002) A New Inexact Alternating Directions Method for Monotone Variational Inequalities. *Mathematical Programming*, **92**, 103-118. <https://doi.org/10.1007/s101070100280>
- [10] Goldstein, T., O'Donoghue, B., Setzer, S. and Baraniuk, R. (2014) Fast Alternating Direction Optimization Methods. *SIAM Journal on Imaging Sciences*, **7**, 1588-1623. <https://doi.org/10.1137/120896219>
- [11] Giesen, J. and Laue, S. (2016) Distributed Convex Optimization with Many Convex Constraints. arXiv: 1610.02967
- [12] 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. 122101. <https://doi.org/10.1007/s11432-017-9367-6>
- [13] Zhang, J. and Luo, Z. (2020) A Proximal Alternating Direction Method of Multiplier for Linearly Constrained Nonconvex Minimization. *SIAM Journal on Optimization*, **30**, 2272-2302. <https://doi.org/10.1137/19m1242276>
- [14] Guo, K., Han, D.R. and Wu, T.T. (2016) Convergence of Alternating Direction Method for Minimizing Sum of Two Nonconvex Functions with Linear Constraints. *International Journal of Computer Mathematics*, **94**, 1653-1669. <https://doi.org/10.1080/00207160.2016.1227432>
- [15] Mordukhovich, B.S. (2006) Variational Analysis and Generalized Differentiation Vol. I: Basic Theory, Vol. II: Applications. Springer.
- [16] Yang, W.H. and Han, D. (2016) Linear Convergence of the Alternating Direction Method of Multipliers for a Class of Convex Optimization Problems. *SIAM Journal on Numerical Analysis*, **54**, 625-640. <https://doi.org/10.1137/140974237>
- [17] 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>
- [18] Bolte, J., Sabach, S. and Teboulle, M. (2013) Proximal Alternating Linearized Minimization for Nonconvex and Non-smooth Problems. *Mathematical Programming*, **146**, 459-494. <https://doi.org/10.1007/s10107-013-0701-9>

-
- [19] Nesterov, Y. (2013) *Introductory Lectures on Convex Optimization: A Basic Course*. Springer Science & Business Media.
 - [20] Liu, H., Hu, J., Li, Y. and Wen, Z. (2020) *Computational Methods for Optimization*. Higher Education Press. (In Chinese)