

# 基于广义近似交替方向乘子法求解可分离凸优化问题

殷倩雯, 党亚峥\*, 向浩东

上海理工大学管理学院, 上海

Email: 1178225576@qq.com, \*jgdyz@163.com, xianghaodongss@163.com

收稿日期: 2021年3月6日; 录用日期: 2021年4月8日; 发布日期: 2021年4月15日

## 摘要

本文提出了一种广义近似交替方向乘子法(gPADMM)来求解可分离凸优化问题。和近似邻近点算法(APPA)和扩展邻近交替方向方法(ePADM)相比,新算法不仅更新自定义矩阵的结构,而且引入随机变量进行随机加速更新步长,从而克服了旧算法固定步长的不灵活性。在某些适当的假设条件下,本文证明了新算法的全局收敛性,并且初步数值实验表明该算法是有效的,收敛速度比旧算法更快。

## 关键词

广义近似交替方向乘子法, 可分离凸优化, 随机加速, 全局收敛

# Solving Separable Convex Optimization Problem Based on Generalized Proximal Alternating Direction Method of Multipliers

Qianwen Yin, Yazheng Dang\*, Haodong Xiang

Business School, University of Shanghai for Science and Technology, Shanghai

Email: 1178225576@qq.com, \*jgdyz@163.com, xianghaodongss@163.com

Received: Mar. 6<sup>th</sup>, 2021; accepted: Apr. 8<sup>th</sup>, 2021; published: Apr. 15<sup>th</sup>, 2021

## Abstract

In this paper, we propose a generalized-proximal alternating direction method of multipliers (gPADMM) for separable convex optimization problem. Compared with the approximate proximal point algorithm (APPA) and the extend proximal alternating directions method (ePADM), the new algorithm not only updates the structure of customized matrix, but also induces random variables

\*通讯作者。

for random acceleration to update the step length, which overcomes the inflexibility of the old algorithms' fixed step length. We prove the global convergence of the new algorithm under certain mild conditions. And preliminary numerical experiments show that the algorithm is effective and the gPADMM converges faster than the old algorithms.

## Keywords

Generalized-Proximal Alternating Direction Method of Multipliers, Separable Convex Optimization, Random Acceleration, Global Convergence

Copyright © 2021 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

本文考虑一个具有如下形式的可分离凸优化问题

$$\min \{f(x) + g(y) \mid Ax + By = b, x \in X, y \in Y\}, \quad (1.1)$$

其中  $X \subset \mathbb{R}^{n_1}, Y \subset \mathbb{R}^{n_2}$  是有界闭合的、凸的、非空集合,  $f: X \rightarrow \mathbb{R}$  和  $g: Y \rightarrow \mathbb{R}$  是凸函数(不一定是平滑函数)。  $A \in \mathbb{R}^{m \times n_1}, B \in \mathbb{R}^{m \times n_2}$  是给定的矩阵,  $b \in \mathbb{R}^m$  是一个向量。

Glowsinski [1]和 Gabay [2]证明了基于增广拉格朗日方法(ALM)的交替方向乘子方法对于问题(1.1)是非常有效的。针对问题(1.1), ALM 算法的迭代形式为

$$\begin{cases} x^{k+1} = \arg \min_{x \in X} \left\{ f(x) + \frac{\beta}{2} \left\| Ax + By^k - b - \frac{1}{\beta} \lambda^k \right\|^2 \right\}, \\ y^{k+1} = \arg \min_{y \in Y} \left\{ g(y) + \frac{\beta}{2} \left\| Ax^{k+1} + By - b - \frac{1}{\beta} \lambda^k \right\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \beta (Ax^{k+1} + By^{k+1} - b). \end{cases} \quad (1.2)$$

Gabay [3]表明 ADMM 算法在本质上是 Douglas-Rachford 分裂方法的一种应用[4]。Cai, Gu 和 He 在文献[5]中提供了一种新颖简单的邻近点算法(PPA)来解释 ADMM 算法, 并提出了一种广义的 ADMM 算法(gADMM), 即它首先产生了一个预测算子, 形式如下:

$$\begin{cases} \tilde{x}^k = \arg \min_{x \in X} \left\{ f(x) + \frac{\beta}{2} \left\| Ax + By^k - b - \frac{1}{\beta} \lambda^k \right\|^2 \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta (A\tilde{x}^k + By^k - b), \\ \tilde{y}^k = \arg \min_{y \in Y} \left\{ g(y) + \frac{\beta}{2} \left\| A\tilde{x}^k + By - b - \frac{1}{\beta} \tilde{\lambda}^k \right\|^2 \right\}, \\ x^{k+1} = \tilde{x}^k, \\ y^{k+1} = y^k - \gamma (y^k - \tilde{y}^k), \\ \lambda^{k+1} = \lambda^k - \gamma (\lambda^k - \tilde{\lambda}^k), \gamma \in (0, 2). \end{cases} \quad (1.3)$$

其中新的迭代步表示为  $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ 。自定义矩阵表示为:

$$Q = \begin{pmatrix} 0_{n_1 \times n_1} & 0_{n_1 \times n_2} & 0_{n_1 \times m} \\ 0_{n_2 \times n_1} & \beta B^T B & -B^T \\ 0_{m \times n_1} & -B & -\frac{1}{\beta} I_m \end{pmatrix}$$

通过以上分析,可以发现针对问题(1.1),ADMM算法有许多变体。进一步研究得到,文献[6]中提出了DRSM算法并在[7]中被进一步解释为邻近点算法的应用。Jiang, B. Q., Peng, Z., Deng, K. K.提出了两种新颖的定制邻近点算法,在[8]中将建立所提出方法的全局收敛性和  $O(1/k)$ 收敛速率。因此,在[6]中建议对PPA应用[9]中提出的加速方案来加速原始ADMM。最近,[10]中的工作表明了开发下降型方法求解的可能性,其下降方向是从ADM生成的迭代中得出的。在[11]中,李、袁等人考虑将广义交替方向乘子法与对数-二次近端正则化相结合,以解决具有可分离结构的变分不等式问题(VI),并建立由在遍历和非遍历意义上的迭代复杂度。袁晓明提出了一种改进的基于PADM的下降方法,该方法继承了[12]中ADM,PPA和下降类型方法的所有优点。在[13]中,徐使用服从高斯分布的随机数来随机扩展步长,并针对一类变分不等式提出了随机步长收缩方法。但是,在某些实际应用中,宽松的步骤可能是不可接受的,甚至是不被允许的。

为了应对这些实际情况,本文修改了预测步长,并采用随机变量来更新步长,而不是放松固定步长。然后,提出了广义的邻近交替方乘子向法。在一些适当的假设下,本文证明了算法的收敛性。同时,本文通过数值实验,还验证了与以前提出的某些算法相比,新算法在实践中具有更好的数值性能。

本文的其余部分安排如下。本文的第2节描述了一些相关的基础知识和引理,以便后续证明。在第3节中,本文针对问题(1.1)提出了广义近邻交替方向乘子方法。所提方法的收敛性将在第4节中得到证明。在第5节中,与以前提出的算法相比,一些初步的数值结果被提出来表明新方法的高效率。第六部分总结了本文的一些结论。

## 2. 预备知识

本节给出了将在后续当中使用的一些预备知识。贯穿全文,本文将  $\|u\| = \sqrt{u^T u}$  表示为Euclidean范数,而  $\langle \cdot, \cdot \rangle$  表示为内积。令  $\Omega \in R^m \times R^n \times R^r$  是  $R$  上的一个非空闭合子集。当  $P_\Omega$  的可行集  $\Omega$  是非负正数  $R_+^N$  时,向量  $x \in R$  的每个分量  $i$  的投影很简单,表示为  $(P_\Omega[x])_i = \max\{0, x_i\}$ 。

**引理 2.1** 令  $\zeta > 0$  为常数,  $C$  为非空封闭的凸集。当且仅当

$$x^* = P_C[x^* - \zeta f(x^*)], x^* \in C$$

是  $VI(f, C)$  的最优解。

记  $e(x, \zeta) = x - P_C[x - \zeta f(x)]$  为投影方程的残差,则  $VI(f, C)$  等价于找到  $e(x, \zeta)$  的零点。因此,我们可以将  $\|e(x, \zeta)\|$  记为算法的停止准则。下一个引理表明,对于任何  $\zeta > 0$ ,  $\|e(x, \zeta)\|$  是一个非递减函数。

众所周知,对于任何  $x$  和  $y$ ,正交投影算子  $P_\Omega$  具有以下特性:

**引理 2.2** [14] 令  $\Omega \in R^m \times R^n \times R^r$  是一个非空闭合凸集,那么对于任意的  $x, y \in R^n$ , 都有

- (1)  $\langle x - P_\Omega(x), P_\Omega(x) - y \rangle \geq 0, \forall y \in \Omega$
- (2)  $\langle P_\Omega(x) - P_\Omega(y), x - y \rangle \geq 0, \forall x, y \in R^m \times R^n \times R^r$
- (3)  $\|P_\Omega(x) - P_\Omega(y)\| \leq \|x - y\|, \forall x, y \in R^m \times R^n \times R^r$

$$(4) \langle x-y, P_{\Omega}(x)-P_{\Omega}(y) \rangle \geq \|P_{\Omega}(x)-P_{\Omega}(y)\|^2, \forall x, y \in R^m \times R^n \times R^r$$

在许多情况下, 可行集  $C$  具有以下形式:  $C = \{x | Ax + By = b, x \in Q\}$ , 其中  $A, B \in R^{m \times n}, b \in R^m$ , 而  $Q$  是  $R_+^n$  的简单凸子集。基本的  $VI(f, C)$  可以转换为结构化的变分不等式问题, 表示为  $VI(F, C)$ : 寻找一个  $u^* \in Q$ , 使得

$$F(u^*)^T(u-u^*) \geq 0, \forall u \in Q,$$

其中  $u := \begin{pmatrix} x \\ y \end{pmatrix}$ ,  $w := \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}$ ,  $F(w) := \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix}$ 。

关于目标函数  $f$  的以下定义是非常重要的, 并且会在后面证明会使用到。

**定义 2.3** 设  $f$  为从  $C \rightarrow R^n$  的映射。如果  $(u-v)^T(f(u)-f(v)) \geq 0, \forall u, v \in C$ , 那么  $f$  在  $C$  上是单调的。

如果  $C$  是  $R^n$  的紧凸集, 而  $F: C \rightarrow R^n$  是连续映射, 则变分不等式问题(VIP)至少具有一个解。假设函数是单调的, 那么可以保证变分不等式问题的解存在且唯一。

**引理 2.4** 如果  $C$  是  $R^n$  的非空闭凸集, 而  $F: C \rightarrow R^n$  是一个连续映射, 那么对于任给的  $x, y \in C$ , 有  $\langle x-y, F(x)-F(y) \rangle \geq 0, \forall x, y \in C$ 。

**定义 2.5** 假设序列  $\{X_n\}$  是由  $X_1, X_2, \dots, X_n, \dots$  组成的随机变量, 如果对于任意的  $\epsilon > 0$ , 有  $\lim_{n \rightarrow \infty} P\{(X_n - X) > \epsilon\} = 0$ , 那么就称序列  $\{X_n\}$  以概率收敛到  $X$ , 并且记为  $X_n \xrightarrow{P} X$ 。

**引理 2.6** 马尔可夫不等式: 对于任何  $\epsilon > 0$  和  $r > 0$ , 都有

$$P(|X| \geq \epsilon) \leq \frac{E(|X|^r)}{\epsilon^r}$$

**引理 2.7** 假设  $X_n \xrightarrow{P} X$ , 那么一定有  $X_{n_k} \rightarrow X$  a.s. (几乎必然地)。

### 3. 算法提出

本节提出了一个 gPADMM 算法来解决问题(1.1), 其中问题(1.1)的增广拉格朗日形式为

$$L(x, y, \lambda) = \theta_1(x) + \theta_2(y) - \lambda^T(Ax + By - b) + \frac{\beta}{2} \|Ax + By - b\|^2. \tag{3.1}$$

如果  $(x^*, y)^* \in X \times Y$  是问题的最优解, 那么存在一个  $\lambda^* \in R^m$ , 使得  $(x^*, y^*, \lambda^*)$  是一个鞍点, 满足

$$\begin{cases} L(x, y^*, \lambda^*) - L(x^*, y^*, \lambda^*) \geq 0, & \forall x \in X \\ L(x^*, y, \lambda^*) - L(x^*, y^*, \lambda^*) \geq 0, & \forall y \in Y \\ L(x^*, y^*, \lambda^*) - L(x^*, y^*, \lambda) \geq 0, & \forall \lambda \in R^m \end{cases} \tag{3.2}$$

因此, 由(3.2)得出

$$\begin{cases} x^* = \arg \min_{x \in X} L(x, y^*, \lambda^*), \\ y^* = \arg \min_{y \in Y} L(x^*, y, \lambda^*), \\ \lambda^* = \arg \min_{\lambda \in R^m} L(x^*, y^*, \lambda). \end{cases} \tag{3.3}$$

根据[12]中的引理 2.1, 本文可以得到问题(3.3)的一阶最优条件:

$$\begin{cases} f(x) - f(x^*) - (x - x^*)(A^T \lambda^*) \geq 0, & \forall x \in X, \\ g(y) - g(y^*) - (y - y^*)(B^T \lambda^*) \geq 0, & \forall y \in Y, \\ (\lambda - \lambda^*)^T (Ax^* + By^* - b) \geq 0, & \forall \lambda \in R^m, \end{cases} \quad (3.4)$$

因此, 问题的解集也可以表示为包括所有满足条件的  $w^* \in W^*$ , 其中  $W := X \times Y \times R^m$ ,  $\theta(u) := \theta_1(x) + \theta_2(y)$  和

$$u := \begin{pmatrix} x \\ y \end{pmatrix}, \quad w := \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) := \begin{pmatrix} -A^T \lambda \\ -B^T \lambda \\ Ax + By - b \end{pmatrix}. \quad (3.5)$$

综上所述, 针对可分离凸优化问题(1.1), 本文提出如下的 gPADMM 算法:

### 算法 3.1 gPADMM 算法

(1) 对于给定的一个  $w^k$ , 有

$$\begin{cases} \tilde{x}^k = \arg \min_{x \in X} \left\{ \theta_1(x) - \langle \lambda, Ax + By^k - b \rangle + \frac{\beta}{2} \|Ax + By^k - b\|^2 + \frac{R+S}{2} \|x - x^k\|^2 \right\}, \\ \tilde{y}^k = \arg \min_{y \in Y} \left\{ \theta_2(y) - \langle \lambda, A\tilde{x}^k + By - b \rangle + \frac{\beta}{2} \|A\tilde{x}^k + By - b\|^2 + \frac{S}{2} \|y - y^k\|^2 \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta (A\tilde{x}^k + B\tilde{y}^k - b), \end{cases} \quad (3.6)$$

(2) 计算:

$$w^{k+1} = w^k - \eta_k \alpha_k (w^k - \tilde{w}^k), \quad (3.7)$$

其中  $\xi_k = \omega(a, b)$ ,  $\eta_k = \frac{1}{k} \sum_{i=1}^k \xi_i$ ,  $\alpha_k = \frac{\|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k) B (y^k - \tilde{y}^k)}{\|w^k - \tilde{w}^k\|_Q^2}$ ,

(3) 停止准则: 计算  $stopc = \frac{\max(\|w^{k+1} - w^k\|)}{\max(\|w^0 - \tilde{w}^0\|)} \leq 10^{-6}$ , 如果  $stopc < \epsilon$ , 则停止迭代, 否则返回  $k = k + 1$ .

**注 3.1** 为了确保收敛,  $\xi_k = \omega(a, b)$  需要满足是某个独立且均匀分布的扩展序列, 并且  $\xi_k = \omega(a, b)$  的数学期望为  $E(\xi_k)$ ,  $k = 1, 2, \dots$ . 另外, 在整个迭代过程不需要计算  $\rho(A^T A)$  或  $A^{-1}$ .

## 4. 收敛性证明

在本节中, 本文充分结合变分不等式来证明算法 3.1 的全局收敛性. 因此, 公式(3.6)可变为具有如下形式的变分不等式:

$$\begin{cases} (x' - \tilde{x}^k) \left\{ \theta_1(\tilde{x}^k) - A^T [\lambda^k - \beta (A\tilde{x}^k + By^k - b)] + (R+S) \|\tilde{x}^k - x^k\|^2 \right\} \geq 0, & \forall x' \in X, \\ (y' - \tilde{y}^k) \left\{ \theta_2(\tilde{y}^k) - B^T [\lambda^k - \beta (A\tilde{x}^k + B\tilde{y}^k - b)] + S \|\tilde{y}^k - y^k\|^2 \right\} \geq 0, & \forall y' \in Y, \\ \tilde{\lambda}^k = \lambda^k - \beta (A\tilde{x}^k + B\tilde{y}^k - b). \end{cases} \quad (4.1)$$

接下来将引入一些重要的引理和定理来证明算法的收敛性, 如下所示.

**引理 4.1** 令  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  是由公式(3.6)生成得到, 那么对于任意的  $w^* \in W^*$  是问题(1.1)的最优解, 得到

$$(w^k - w^*)^T Q (w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T B (y^k - \tilde{y}^k), \quad (4.2)$$

$$\text{其中 } Q = \begin{pmatrix} (R+S)I_{n_1} & 0_{n_1 \times n_2} & 0_{n_1 \times m} \\ 0_{n_2 \times n_1} & \beta B^T B + SI_{n_2} & 0_{n_2 \times m} \\ 0 & 0 & \frac{1}{\beta} I_m \end{pmatrix}.$$

**证明** 由于  $w^* \in W^*$ ,  $\tilde{x} \in X$  并且  $\tilde{y} \in Y$ , 那么可知

$$\begin{cases} (\tilde{x}^k - x^*)^T (\theta_1(x^*) - A^T \lambda^*) \geq 0, \\ (\tilde{y}^k - y^*)^T (\theta_2(y^*) - B^T \lambda^*) \geq 0. \end{cases} \quad (4.3)$$

将(3.6)分别加入(4.3)中且算子  $\theta_1$  和  $\theta_2$  的单调性, 得出

$$\begin{cases} (A\tilde{x}^k - Ax^*)^T (\tilde{\lambda}^k - \lambda^*) + (\tilde{x}^k - x^*)^T (R+S)(x^k - \tilde{x}^k) \geq \beta (A\tilde{x}^k - Ax^*)^T (By^k - B\tilde{y}^k), \\ (B\tilde{y}^k - By^*)^T (\tilde{\lambda}^k - \lambda^*) + (\tilde{y}^k - y^*)^T S(y^k - \tilde{y}^k) \geq 0, \end{cases} \quad (4.4)$$

因此, 由(4.4) (3.6)以及  $Ax^* + By^* = b$  得知

$$\begin{aligned} & \frac{1}{\beta} (\tilde{\lambda}^k - \lambda^*)^T (\lambda^k - \tilde{\lambda}^k) + (\tilde{x}^k - x^*)^T (R+S)(x^k - \tilde{x}^k) + (\tilde{y}^k - y^*)^T S(y^k - \tilde{y}^k) \\ & \geq \beta (A\tilde{x}^k - Ax^*)^T (By^k - B\tilde{y}^k), \end{aligned} \quad (4.5)$$

在(4.5)的两边同时加上  $\beta (B\tilde{y}^k - By^*)^T (By^k - B\tilde{y}^k)$  及矩阵  $Q$  的定义和  $Ax^* + By^* = b$ , 于是得到

$$\begin{aligned} (\tilde{w}^k - w^*)^T Q(w^k - \tilde{w}^k) & \geq \beta (A\tilde{x}^k + B\tilde{y}^k - b)^T (By^k - B\tilde{y}^k) \\ & \geq (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k), \end{aligned} \quad (4.6)$$

这意味着

$$(\tilde{w}^k - w^*)^T Q(w^k - \tilde{w}^k) \geq \|w^k - \tilde{w}^k\|_Q^2 (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k).$$

因此, 引理 4.1 得证。

**引理 4.2** 令  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  是由公式(3.6)生成得到, 那么对于任意的  $w^* \in W^*$  是问题(1.1)的最优解, 有

$$\|w^k - w^*\|_Q^2 \geq \|\tilde{w}^k - w^*\|_Q^2 + 2(\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) + \|\tilde{w}^k - w^k\|_Q^2. \quad (4.7)$$

**证明** 由引理 4.1 和(3.6)式得

$$\begin{aligned} \|\tilde{w}^k - w^*\|_Q^2 & = \|w^k - w^* - (w^k - \tilde{w}^k)\|_Q^2 \\ & = \|w^k - w^*\|_Q^2 + 2(w^k - w^*)^T Q(w^k - \tilde{w}^k) + \|w^k - \tilde{w}^k\|_Q^2 \\ & \leq \|w^k - w^*\|_Q^2 - 2\left[\|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k)\right] + \|w^k - \tilde{w}^k\|_Q^2 \\ & \leq \|w^k - w^*\|_Q^2 - \|w^k - \tilde{w}^k\|_Q^2 - 2(\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) \end{aligned}$$

因此, 引理 4.2 证毕且公式(4.7)表明  $d(w^k, \tilde{w}^k) = w^k - \tilde{w}^k$  是  $\|w^k - w^*\|_Q^2$  的下降方向。

**定理 4.1** 假设  $w^*$  是问题(1.1)的解, 那么对任给的迭代点  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ , 本文定义如下的一些函数:

$$\begin{cases} w^{k+1}(\alpha) = w^k - \alpha_k (w^k - \tilde{w}^k), \\ \theta^k(\alpha) = \|w^k - w^*\|_Q^2 - \|w^{k+1}(\alpha) - w^*\|_Q^2, \\ \varphi(w^k, \tilde{w}^k) = \|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k), \\ \psi^k(\alpha) = 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2 \|w^k - \tilde{w}^k\|_Q^2. \end{cases} \quad (4.8)$$

那么对任意的  $w^* = (x^*, y^*, \lambda^*) \in W^*, \alpha \geq 0$ , 存在  $\theta^k(\alpha) \geq \psi^k(\alpha)$  (4.9)

**证明** 由(4.8)可知

$$\begin{aligned} \|w^{k+1}(\alpha) - w^*\|_Q^2 &= \|w^k - \alpha_k (w^k - \tilde{w}^k) - w^*\|_Q^2 \\ &= \|w^k - w^*\|_Q^2 - 2\alpha_k (w^k - w^*)^T Q (w^k - \tilde{w}^k) + \alpha_k^2 \|w^k - \tilde{w}^k\|_Q^2 \\ &\leq \|w^k - w^*\|_Q^2 - 2\alpha\varphi(w^k, \tilde{w}^k) + \alpha^2 \|w^k - \tilde{w}^k\|_Q^2 \\ &= \|w^k - w^*\|_Q^2 - \psi^k(\alpha). \end{aligned}$$

因此, (4.9)得证。

从定理 4.1 可知  $\psi^k(\alpha)$  是  $\theta^k(\alpha)$  的一个下界且  $\theta^k(\alpha) \geq \psi^k(\alpha)$ 。由  $\psi^k(\alpha)$  的定义可知, 它是关于  $\alpha$  的一个二次函数, 且它的最大值

$$\alpha_k = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_Q^2} = \frac{\|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k)}{\|w^k - \tilde{w}^k\|_Q^2}。$$

也是最优的步长。

**定理 4.2** 如果  $0 < a < E(\xi_k) = \rho < b$ , 序列  $\{w^k\}$  是由广义的邻近交替方向乘法生成, 那么有

$$E\left(\|w^{k+1} - w^*\|_Q^2\right) \leq E\left(\|w^k - w^*\|_Q^2\right) - \frac{\rho(2-\rho)}{4} E\left(\|w^k - \tilde{w}^k\|_Q^2\right), \quad (4.10)$$

**证明** 由引理 4.1 和(3.7), 我们得到

$$\begin{aligned} \|w^{k+1} - w^*\|_Q^2 &= \|w^k - \eta_k \alpha_k (w^k - \tilde{w}^k) - w^*\|_Q^2 \\ &= \|w^k - w^*\|_Q^2 - 2\eta_k \alpha_k (w^k - w^*)^T Q (w^k - \tilde{w}^k) + (\eta_k \alpha_k)^2 \|w^k - \tilde{w}^k\|_Q^2 \\ &\leq \|w^k - w^*\|_Q^2 - 2\eta_k \alpha_k \left( \|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) \right) + \eta_k^2 \alpha_k^2 \|w^k - \tilde{w}^k\|_Q^2 \\ &\leq \|w^k - w^*\|_Q^2 - \frac{1}{2} \eta_k (2 - \eta_k) \alpha_k \left( \|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k) \right). \end{aligned} \quad (4.11)$$

由于  $\xi_1, \xi_2, \dots, \xi_k, \dots$  是独立且均匀分布, 可以推出  $\eta_k, w^k, \tilde{w}^k$  和  $\|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k)$  也是独立的。又因为

$$E(\eta_k) = E\left(\frac{1}{k} \sum_{i=1}^k \xi_k\right) = \frac{1}{k} \sum_{i=1}^k E(\xi_k) = \rho, E(\alpha) > \frac{1}{2},$$

那么  $\|w^{k+1} - w^*\|_Q^2$  的数学期望满足

$$\begin{aligned} E\left(\|w^{k+1} - w^*\|_Q^2\right) &= E\left(\|w^k - \eta_k \alpha_k (w^k - \tilde{w}^k) - w^*\|_Q^2\right) \\ &= E\left(\|w^k - w^*\|_Q^2\right) - E\left(2\eta_k \alpha_k (w^k - w^*)^T Q (w^k - \tilde{w}^k)\right) + (\eta_k \alpha_k)^2 E\left(\|w^k - \tilde{w}^k\|_Q^2\right) \\ &\leq E\left(\|w^k - w^*\|_Q^2\right) - 2E(\eta_k) E(\alpha_k) E\left(\|w^k - \tilde{w}^k\|_Q^2 + (\lambda^k - \tilde{\lambda}^k)^T (By^k - B\tilde{y}^k)\right) \\ &\quad + E(\eta_k^2) E(\alpha_k^2) E\left(\|w^k - \tilde{w}^k\|_Q^2\right) \\ &\leq E\left(\|w^k - w^*\|_Q^2\right) - \frac{1}{4}\rho(2-\rho) E\left(\|w^k - \tilde{w}^k\|_Q^2\right). \end{aligned}$$

最后, 获得本文期望的结果。

**推论 4.1** 如果  $0 < a < E(\xi_k) = \rho < b$ , 序列  $\{w^k\}$  是由广义的邻近交替方向乘法生成, 那么有

- (1) 序列  $E(w^k)$  和  $E(\tilde{w}^k)$  都是有界的。
- (2) 序列  $E\left(\|w^{k+1} - w^*\|_Q^2\right)$  是单调且有界的。
- (3)  $\lim_{k \rightarrow \infty} E\left(\|w^k - \tilde{w}^k\|_Q^2\right) = 0$ 。

**定理 4.3** 如果  $0 < a < E(\xi_k) = \rho < b$ , 序列  $\{w^k\}$  是由广义的邻近交替方向乘法生成, 那么序列  $\{w^k\}$  依概率收敛到  $w^\infty \in W^*$ 。

**证明** 从推论 4.1 和引理 4.2, 得到

$$\lim_{k \rightarrow \infty} E\left(\|A\tilde{x}^k + B\tilde{y}^k - b\|_Q^2 + \|B(y^k - \tilde{y}^k)\|_Q^2\right) = 0. \tag{4.12}$$

因为序列  $\{E(\tilde{w}^k)\}$  是有界的, 那么它至少有一个聚集点。令  $E(w^\infty)$  是序列  $\{E(\tilde{w}^k)\}$  的一个聚集点, 并且存在一个子序列  $\{E(\tilde{w}^{k_j})\}$  收敛到  $E(w^\infty)$ , 那么存在一个子序列  $\{\tilde{w}^{k_j}\}$  依概率收敛到  $w^\infty \in W^*$ , 记为  $\tilde{w}^{k_j} \xrightarrow{P} w^\infty$ 。

从引理 2.3 得到子序列  $\{\tilde{w}^{k_j}\}$  几乎必然地收敛到  $w^\infty$ , 那么它也可标记子序列为  $\{\tilde{w}^{k_j}\}$ , 记作

$$\lim_{j \rightarrow \infty} \tilde{w}^{k_j} = w^\infty \quad a.s. \tag{4.13}$$

结合公式(3.6)、(4.12)和(4.13), 有

$$\begin{cases} \lim_{j \rightarrow \infty} (x - \tilde{x}^{k_j})^T \{\theta_1(\tilde{x}^{k_j}) - A^T \tilde{\lambda}^{k_j}\} \geq 0, & \forall x \in X \\ \lim_{j \rightarrow \infty} (y - \tilde{y}^{k_j})^T \{\theta_2(\tilde{y}^{k_j}) - B^T \tilde{\lambda}^{k_j}\} \geq 0, & \forall y \in Y \\ \lim_{j \rightarrow \infty} (A\tilde{x}^{k_j} + B\tilde{y}^{k_j} - b) = 0 \end{cases}$$

$$\text{从而有} \begin{cases} (x - x^\infty)^T \{\theta_1(x^\infty) - A^T \lambda^\infty\} \geq 0, & \forall x \in X \\ (y - y^\infty)^T \{\theta_2(y^\infty) - B^T \lambda^\infty\} \geq 0, & \forall y \in Y. \\ (Ax^\infty + By^\infty - b) = 0 \end{cases}$$



因此, 可以推出  $\omega^\infty \in W^*$ 。

又因为  $\lim_{j \rightarrow \infty} E(\tilde{w}^{k_j}) = E(w^\infty)$ , 则  $\lim_{l \rightarrow \infty} E(\|\tilde{w}^{k_l} - w^\infty\|_Q^2) = 0$ 。同时因为  $\lim_{k \rightarrow \infty} E(\|w^k - \tilde{w}^k\|_Q^2) = 0$ , 且对任给地  $\varepsilon > 0$ , 有  $l > 0$ , 于是得到

$$E(\|w^{k_l} - \tilde{w}^{k_l}\|_Q^2) = \frac{\varepsilon}{2}$$

和

$$E(\|\tilde{w}^{k_l} - w^\infty\|_Q^2) < \frac{\varepsilon}{2}. \tag{4.14}$$

因此, 对任意的  $k > k_l$ , 由(4.14), 可知

$$E(\|w^k - w^\infty\|_Q^2) \leq E(\|w^{k_l} - w^\infty\|_Q^2) \leq E(\|w^{k_l} - \tilde{w}^{k_l}\|_Q^2) + E(\|\tilde{w}^{k_l} - w^\infty\|_Q^2) < \varepsilon. \tag{4.15}$$

由此推出  $\lim_{k \rightarrow \infty} E(\|w^k - w^\infty\|_Q^2) = 0$ 。

结合引理 2.2 和公式(4.15), 可得对任给的  $\varepsilon > 0$ , 如果存在  $l > 0$ , 只要  $k > l$ , 就有

$$P(\|w^k - w^\infty\|_Q^2 \geq \varepsilon) \leq \frac{E(\|w^k - w^\infty\|_Q^2)}{\varepsilon} < \varepsilon$$

那么得到  $\lim_{k \rightarrow \infty} P(\|w^k - w^\infty\|_Q^2 \geq \varepsilon) = 0$ 。

因此, 序列  $\{w^k\}$  依概率收敛到  $w^\infty \in W^*$ , 记作  $w^k \xrightarrow{P} w^\infty$ 。

至此, 收敛性证明全部完成。

### 5. 数值实验

本节通过数值实验重点介绍算法的有效性。

**例 5.1** 首先, 考虑陶、袁和何[15]提出的财务和统计问题, 其形式如下:

$$\min \left\{ \frac{1}{2} \|X - C\|_F^2 \mid X \in S_+^n \cap S_B \right\}, \tag{5.1}$$

其中  $S_B = \{H \in R^{n \times n} \mid H_L \geq H \geq H_U\}$ 。

通过引入一个辅助变量  $Y$  使得  $X - Y = 0$ , 问题(5.1)可以重新表示为如下形式的可分离凸优化问题

$$\min \left\{ \frac{1}{2} \|X - C\|_F^2 + \frac{1}{2} \|Y - C\|_F^2 \mid X = Y, X \in S_+^n, Y \in S_B \right\}, \tag{5.2}$$

显然地, 问题(5.2)是问题(1.1)的一个特殊形式, 其中  $\rho(A^T A) = \rho(B^T B) = 1$ 。令  $\lambda \in R^n$  是线性约束  $X - Y = 0$  的拉格朗日乘子。对于给定的  $(X^k, Y^k, \lambda^k)$ , 算法能产生第  $k+1$  次迭代步, 有

$$\begin{cases} \tilde{X}^k = \arg \min_{X \in S_+^n} \left\{ \frac{1}{2} \|X - C\|^2 + \langle \lambda^k, X - Y^k \rangle + \frac{\beta}{2} \|X - Y^k\|^2 + \frac{R+S}{2} \|X - X^k\|^2 \right\}, \\ \tilde{Y}^k = \arg \min_{Y \in S_B} \left\{ \frac{1}{2} \|Y - C\|^2 + \langle \lambda^k, \tilde{X}^k - Y \rangle + \frac{\beta}{2} \|\tilde{X}^k - Y\|^2 + \frac{S}{2} \|Y - Y^k\|^2 \right\}, \\ \tilde{\lambda}^k = \lambda^k - \beta(\tilde{X}^k - \tilde{Y}^k). \end{cases} \tag{5.3}$$

迭代(5.3)的  $X$ -子问题是通过 SVD 分解来进行求解, 它承担每次迭代过程中的主要计算负荷。迭代(5.3)的  $Y$ -子问题也是一个投影, 有如下形式

$$\tilde{Y}^k = P_{S_B} \left\{ \frac{1}{S+1+\beta} (\tilde{X}^k - \lambda^k + C + SY^k) \right\},$$

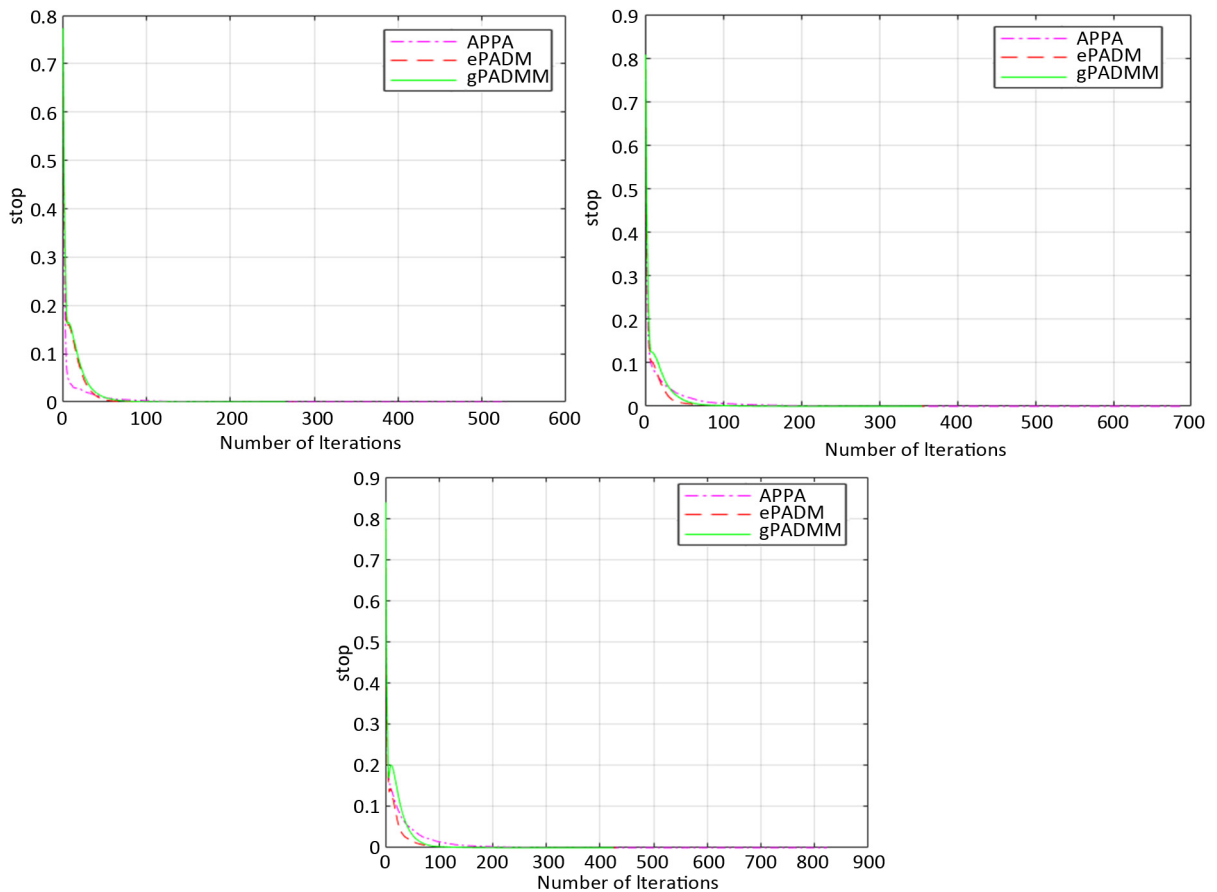
其中  $P_{S_B}$  表示根据欧几里得范数到集合  $S_B$  的投影。

在数值实验中, 对于任意的所有  $i, j$ , 设  $C = (c_{ij})_{n \times n}$  是一个满足  $c_{ij} \in (-1, 1)$  随机矩阵。对于每个给定, 将测试 20 个随机实例。公平地说, 将  $X^0$  和  $Y^0$  设置为  $n \times n$  独立性矩阵,  $\lambda^0$  是一个  $n$  维零向量, 并且将

$$\varepsilon = \frac{\max(|w^{k+1} - w^k|)}{\max(|w^0 - \tilde{w}^0|)} \leq 10^{-6}$$

记为算法的终止准则。

得到的实验结果如下图 1 所示。



**Figure 1.** Comparisons of APPA, ePADM, gPADMM:  $n = 100, 200, 500$   
**图 1.** APPA, ePADM, gPADMM 三种算法的比较:  $n$  取 100, 200, 500

**Table 1.** Numerical results of Example 5.1: Comparisons of APPA, ePADM, gPADMM  
**表 1.** 例 5.1 的实验结果: APPA, ePADM, gPADMM 三种算法的比较

		p = 1.9		
		gPADMM	APPA [16]	ePADM [16]
$n = 100$		Iter. = 266, $s = 2.139$	Iter. = 529, $s = 4.725$	Iter. = 275, $s = 2.441$
$n = 200$		Iter. = 355, $s = 12.995$	Iter. = 686, $s = 23.919$	Iter. = 358, $s = 13.013$
$n = 500$		Iter. = 425, $s = 110.273$	Iter. = 823, $s = 216.642$	Iter. = 430, $s = 107.588$

显然,从表 1 中可以看出,算法 3.1 的性能优于 APPA 算法和 ePADM 算法,因为其迭代和计算时间更少。因此,结果表明对于问题(1.1),提出的算法是有效的。

## 6. 结论

本文针对线性凸优化问题和可分离凸优化问题,提出了广义的近似交替方向乘子算法,并且通过初步实验结果验证了算法是可行的。算法 3.1 是一种简单且有希望的迭代方案,其收敛速度比原有算法的收敛速度快。它在其他加速技术和其他应用中进行分析和设计的潜力以及更全面的计算研究是我们进一步研究的方向。

## 基金项目

国家自然科学基金(72071130)。

## 参考文献

- [1] Glowinski, R. and Marrocco, A. (1974) Analyse numerique du champ magnetique d'un alternateur par elements finis et sur-relaxation ponctuelle non lineaire. *Computer Methods in Applied Mechanics and Engineering*, **3**, 55-85. [https://doi.org/10.1016/0045-7825\(74\)90042-5](https://doi.org/10.1016/0045-7825(74)90042-5)
- [2] 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)
- [3] Gabay, D. (1983) Chapter IX: Applications of the Method of Multipliers to Variational Inequalities. *Studies in Mathematics and Its Applications*, **15**, 299-331. [https://doi.org/10.1016/S0168-2024\(08\)70034-1](https://doi.org/10.1016/S0168-2024(08)70034-1)
- [4] Douglas, J. and Rachford, H.H. (1956) On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables. *Transactions of the AMS*, **82**, 421-439. <https://doi.org/10.1090/S0002-9947-1956-0084194-4>
- [5] Cai, X.J., Gu, G.Y., He, B.S. and Yuan, X.M. (2013) A Proximal Point Algorithm Revisit on the Alternating Direction Method of Multipliers. *Science China Mathematics*, **56**, 2179-2186. <https://doi.org/10.1007/s11425-013-4683-0>
- [6] Eckstein, J. and Bertsekas, D. (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>
- [7] Martinet, B. (1970) Regularisation, d'inequations variationnelles par approximations succesives. *Revue française d'informatique et de recherche opérationnelle*, **4**, 154-159. <https://doi.org/10.1051/m2an/197004R301541>
- [8] Jiang, B.Q., Peng, Z. and Deng, K.K. (2019) Two New Customized Proximal Point Algorithms without Relaxation for Linearly Constrained Convex Optimization. *Bulletin of the Iranian Mathematical Society*, **46**, 865-892. <https://doi.org/10.1007/s41980-019-00298-0>
- [9] Gol'shtein, E.G. and Tret'yakov, N.V. (1979) Modified Lagrangian in Convex Programming and Their Generalizations. In: *Point-to-Set Maps and Mathematical Programming*, Mathematical Programming Studies Vol. 10, Springer, Berlin, 86-97. <https://doi.org/10.1007/BFb0120845>
- [10] Ye, C.H. and Yuan, X.M. (2007) A Descent Method for Structured Monotone Variational Inequalities. *Optimization Methods and Software*, **22**, 329-338. <https://doi.org/10.1080/10556780600552693>
- [11] Li, M., Li, X.X. and Yuan, X.M. (2014) Convergence Analysis of the Generalized Alternating Direction Method of Multipliers with Logarithmic-Quadratic Proximal Regularization. *Journal of Optimization Theory and Applications*, **164**, 218-233. <https://doi.org/10.1007/s10957-014-0567-x>
- [12] Yuan, X.M. (2009) An Improved Proximal Alternating Direction Method for Monotone Variational Inequalities with Separable Structure. *Computational Optimization and Applications*, **49**, 17-29. <https://doi.org/10.1007/s10589-009-9293-y>
- [13] 徐海文. 一类变分不等式的随机步长收缩算法[J]. 工程数学学报, 2011, 28(4): 462-469.
- [14] Wang, Y., Yang, J., Yin, W. and Zhang, Y. (2008) A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, **1**, 248-272. <https://doi.org/10.1137/080724265>
- [15] Tao, M., Yuan, X.M. and He, B.S. (2011) Solving a Class of Matrix Minimization Problems by Linear Variational Inequality Approaches. *Linear Algebra and Its Applications*, **434**, 2343-2352. <https://doi.org/10.1016/j.laa.2010.11.041>
- [16] He, B.S., Li, O.L.Z. and Wang, X. (2012) Proximal-Like Contraction Methods for Monotone Variational Inequalities in a Unified Framework II: Effective Quadruplet and Primary Methods. *Computational Optimization and Applications*, **5**, 649-679.