

# 求解结构型凸优化问题的一种算法改进

张宇婷\*, 李 锋#

云南师范大学, 云南 昆明

收稿日期: 2021年11月23日; 录用日期: 2021年12月17日; 发布日期: 2021年12月24日

## 摘 要

本文考虑一个求解结构型凸优化问题的新算法——带随机步长的并行分裂ADMM下降算法, 它是基于并行分裂ADMM下降算法和改善步长的收缩算法产生的。新算法改变了步长和更新公式, 同时利用服从独立同分布的随机数来扩张步长, 提高了ADMM下降算法中因固定步长因子带来的收敛较慢的结果。然后在适当的条件下, 证明了该方法是依概率收敛的。最后, 通过对来自金融和统计中的一类问题进行的一系列数值实验, 验证了新算法的高效性。

## 关键词

并行分裂算法, ADMM算法, 随机分布

# An Improved Algorithm for Solving Structural Convex Optimization Problems

Yuting Zhang\*, Feng Li#

Yunnan Normal University, Kunming Yunnan

Received: Nov. 23<sup>rd</sup>, 2021; accepted: Dec. 17<sup>th</sup>, 2021; published: Dec. 24<sup>th</sup>, 2021

## Abstract

This paper considers a new algorithm for solving structural convex optimization problems: parallel split ADMM descent algorithm with random step size, which is based on parallel split ADMM descent algorithm and shrinkage algorithm with improved step size. The new algorithm changes the step size and updates formula, and expands the step size by using random numbers subject to independent and identically distributed, which improves the slow convergence result caused by fixed step size factor in ADMM descent algorithm. Then, under appropriate conditions, it is proved that the method converges according to probability. Finally, a series of numerical experiments on

\*第一作者 Email: 809667206@qq.com

#通讯作者 Email: 809776206@qq.com

a class of problems from finance and statistics verify the efficiency of the new algorithm.

## Keywords

Parallel Splitting Algorithm, ADMM Algorithm, Random Distribution

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

### 1.1. 研究背景及意义

随着科技快速的发展,如今数字化技术与网络化技术日新月异,大数据正在飞速增长,人类已经迈入一个全新的时代。最优化问题的研究和发展一直受到广泛关注。大数据时代带来许多机遇的同时,也带来许多问题,例如工程、信息、生物等领域,所建立的模型可以转化为等式约束优化问题。

实际问题中一般所求的目标函数问题规模巨大,且可能是非光滑的,传统的内点法等算法往往无法求解这类问题。而根据 ADMM 算法的核心思想:先分解协调过程,将全局问题分解为多个较容易求解的局部子问题,再求解协调子问题而得到全局问题的解;运用 ADMM 算法等一阶算法求解这类问题更为适合,所以对 ADMM 算法的研究非常具有理论意义和应用价值,并且具有巨大的发展潜力。

### 1.2. 研究现状

Glowinski & Marrocco [1]及 Gabay & Mercier [2]分别于 1975 年和 1976 年最先提出了 ADMM 算法,现在 ADMM 算法主要用于解决具有可分离变量的优化问题,它可以解决增广拉格朗日算法所不能解决的问题。50 年代中期,有大量文献分析了相似的算法的性质,但当时这一类似算法主要用于解决偏微分方程问题。

本文主要研究以下具有分离结构的约束凸优化问题:

$$\begin{aligned} \min_{x,y} f(x) + g(y) \\ \text{s.t. } Ax + By = b \\ x \in \mathcal{X}, y \in \mathcal{Y}. \end{aligned} \quad (1)$$

其中  $f$  和  $g$  是闭凸连续函数,  $A \in \mathcal{R}^{r \times n}$  和  $B \in \mathcal{R}^{r \times m}$  是一个给定的行满秩矩阵,  $b \in \mathcal{R}^r$  是一个给定的向量,  $\mathcal{X} \subset \mathcal{R}^n$  和  $\mathcal{Y} \subset \mathcal{R}^m$  是非空闭凸集。

针对求解式(1)的算法目前已经有很多种,其增广拉格朗日函数为:

$$L_{\rho}(x, y, \lambda) = f(x) + g(y) - \lambda^T (Ax + By - b) + \frac{\rho}{2} \|Ax + By - b\|_2^2$$

其中,  $\lambda \in \mathcal{R}^r$  为拉格朗日乘子,  $\rho > 0$  为罚参数。

经典的 ADMM 算法迭代步骤如下:

$$\begin{cases} x^{k+1} = \arg \min \left\{ f(x) - (\lambda^k)^T (Ax + By^k - b) + \frac{1}{2} \|Ax + By^k - b\|_H^2 \mid x \in \mathcal{X} \right\} \\ y^{k+1} = \arg \min \left\{ g(y) - (\lambda^k)^T (Ax^{k+1} + By - b) + \frac{1}{2} \|Ax^k + By - b\|_H^2 \mid y \in \mathcal{Y} \right\} \\ \lambda^{k+1} = \lambda^k - H (Ax^{k+1} + By^{k+1} - b) \end{cases}$$

其中,  $H \in \mathcal{R}^{r \times r}$  是给定的对称正定阵。

与增广拉格朗日乘子算法(ALM) [3] [4] [5]相比, ALM 算法可以同时求解两个分离的变量, 而 ADMM 算法是对变量交替进行求解, 这也是算法名字的由来。

我们可以将经典的 ADMM 算法看作是 ALM 算法的 Gauss-Seidel [6]形式, 而 Douglas-Rachford 分裂算法[7] [8]是 ADMM 的一种对偶形式, 并且邻近点算法(PPA) [9] [10]可以说明 D-R 算法。

2011 年, Boyd [11]等人重新回顾总结并证明 ADMM 算法适用于大规模优化问题后, 使得这种方法逐渐进入了学者们的视野。随后, ADMM 算法因为其良好的理论性质备受学者们的追捧, 其收敛性与计算复杂性有丰富的研究成果[12] [13] [14] [15]。此外, ADMM 算法在压缩感知[16]、图像处理[17]、信号处理[18] [19]、机器学习[20]等领域也取得了重大进展。由于其广泛的应用性, 该算法近十年来已经有了很多改进和推广。

ADMM 下降算法[21]是 ADMM 算法的一种改进, 我们定义:

$$v = \begin{pmatrix} y \\ \lambda \end{pmatrix} \in \mathcal{Y} \times \mathcal{R}^r,$$

算法的迭代步骤如下

$$\begin{cases} \tilde{x}^k = \arg \min \left\{ f(x) - (\lambda^k)^T (Ax + By^k - b) + \frac{1}{2} \|Ax + By^k - b\|_H^2 \mid x \in \mathcal{X} \right\}, \\ \tilde{y}^k = \arg \min \left\{ g(y) - (\lambda^k)^T (A\tilde{x}^k + By - b) + \frac{1}{2} \|A\tilde{x}^k + By - b\|_H^2 \mid y \in \mathcal{Y} \right\}, \\ \tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b), \\ v^{k+1} = v^k - \gamma \alpha_k^* (v^k - \tilde{v}^k), \quad \alpha_k^* = \frac{1}{2} + \frac{1}{2} \frac{\|A\tilde{x}^k + B\tilde{y}^k - b\|_H^2}{\|v^k - \tilde{v}^k\|_G^2}, \quad 0 < \gamma < 2, \end{cases}$$

其中  $\alpha_k^*$  为最优步长,  $H \in \mathcal{R}^{r \times r}$  是给定的对称正定阵,  $\gamma$  是延长因子(一般  $\gamma \in (1, 2)$  时会使收敛速度加快)。

并行分裂 ADMM 下降算法[22]的迭代步骤如下, 定义:

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix} \in \mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^r$$

$$\begin{cases} \tilde{x}^k = \arg \min \left\{ f(x) - (\lambda^k)^T (Ax + By^k - b) + \frac{1}{2} \|Ax + By^k - b\|_H^2 \mid x \in \mathcal{X} \right\}, \\ \tilde{y}^k = \arg \min \left\{ g(y) - (\lambda^k)^T (A\tilde{x}^k + By - b) + \frac{1}{2} \|A\tilde{x}^k + By - b\|_H^2 \mid y \in \mathcal{Y} \right\}, \\ \tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b), \\ w^{k+1} = w^k - \gamma \alpha_k^* G^{-1} M (w^k - \tilde{w}^k), \quad \alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|G^{-1} M (w^k - \tilde{w}^k)\|_G^2}, \quad 1 \leq \gamma < 2, \end{cases}$$

在 ADMM 下降算法中, 要得到  $\tilde{y}^k$ , 都要用到上一步的结果  $\tilde{x}^k$ , 而并行分裂 ADMM 下降算法可以弥补这一不足,  $\tilde{y}^k$  的获得是独立不依赖于  $\tilde{x}^k$  的,  $\tilde{x}^k$  和  $\tilde{y}^k$  是并行分解的, 这也是称之为并行分裂的原因。并行分裂 ADMM 下降算法在实际应用中可以大大地加快收敛的速度。

文献[23]考虑改善步长收缩算法[24] [25]的固定扩张步长没有充分利用下降量函数的不等式放缩, 将算法中不变的延长因子改为使用随机数生成的延长因子, 并称为随机步长收缩算法, 这一改进方便选定步长且其取值更加丰富。

本文将上述并行分裂 ADMM 下降算法和随机步长收缩算法的思想相结合, 提出了带随机步长的并行分裂 ADMM 下降算法, 并在一定的假设条件下证明了算法的收敛性, 且数值实验表明新算法是有效的。

## 2. 预备知识

下面简单回顾一些在随后的证明中要用到的一些基础知识。

引入拉格朗日乘子  $\lambda$ , 将式(1)转化为下列变分不等式, 也就是寻找  $w = (x, y, \lambda) \in \mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^r$ , 使得

$$\begin{cases} (x' - x)^\top (\nabla f(x) - A^\top \lambda) \geq 0 \\ (y' - y)^\top (\nabla g(y) - B^\top \lambda) \geq 0 \\ Ax + By - b = 0 \end{cases} \quad \forall w' \in \mathcal{W} \quad (2)$$

再记  $\mathcal{W}^*$  为上式的非空解集,  $w^* = (x^*, y^*, \lambda^*)$  是  $\mathcal{W}^*$  中的一个解。

为了分析的便利, 我们引入以下记号

$$G = \begin{pmatrix} A^\top HA & 0 & 0 \\ 0 & B^\top HB & 0 \\ 0 & 0 & H^{-1} \end{pmatrix}, \quad \bar{G} = \begin{pmatrix} 0 & 0 & 0 \\ -B^\top HA & B^\top HB & 0 \\ 0 & 0 & H^{-1} \end{pmatrix}, \quad F(w) = \begin{pmatrix} \nabla f(x) - A^\top \lambda \\ \nabla g(y) - B^\top \lambda \\ Ax + By - b \end{pmatrix}$$

**定义 2.1** [26] 设  $x = (x_1, x_2, \dots, x_n) \in \mathcal{R}^n$ , 对称正定阵  $G \in \mathcal{R}^{n \times n}$ , 则向量  $x$  在矩阵  $G$  意义下的范数定义为

$$\|x\|_G = (x^\top Gx)^{\frac{1}{2}}$$

**定义 2.2** [27] 设  $X_1, X_2, \dots, X_n, \dots$  是一列随机变量, 若  $\forall \varepsilon > 0$ , 恒有

$$\lim_{n \rightarrow \infty} P\{|X_n - X| > \varepsilon\} = 0$$

则称  $\{X_n\}$  依概率收敛到  $X$ , 记作  $X_n \xrightarrow{P} X$ 。

**定理 2.1** [27] 设  $X_n \xrightarrow{P} X$ , 则必有子序列  $X_{n_k} \rightarrow X$  a.s. (几乎必然地)。

**定理 2.2** [27] Markov 不等式: 对  $\forall \varepsilon > 0$ 、 $r > 0$ , 有

$$P(|x| > \varepsilon) \leq \frac{E(|x|^r)}{\varepsilon^r}$$

**定理 2.3** [28] 设  $\mathcal{X} \subset \mathcal{R}^n$  是闭凸集,  $\theta(x)$  和  $h(x)$  为可微凸函数。设  $x^*$  满足  $\min\{\theta(x) + h(x) | x \in \mathcal{X}\}$ , 即

$$x^* = \arg \min\{\theta(x) + h(x) | x \in \mathcal{X}\}$$

的充分必要条件是

$$x^* \in \mathcal{X}, (x - x^*)^\top (\nabla \theta(x^*) + \nabla h(x^*)) \geq 0, \quad \forall x \in \mathcal{X}$$

**定理 2.4** [27] Jensen 不等式: 设  $X$  是一个随机变量,  $f$  是一个凸函数, 则有

$$f(E(X)) \leq E(f(X))$$

如果  $f$  是一个凹函数, 则有

$$f(E(X)) \geq E(f(X))$$

### 3. 带随机步长的并行分裂 ADMM 下降算法及其收敛性

#### 3.1. 带随机步长的并行分裂 ADMM 下降算法

**步 0** 初始化  $w = w^0 = (x^0, y^0, z^0)$ ,  $stopc = 1$ ,  $0 \leq a < b \leq 2$ ,  $k = 0$ , 容许度  $\varepsilon$ ;

**步 1** 当  $stopc \leq \varepsilon$  时, 停机; 否则依次计算

$$\tilde{x}^k = \arg \min \left\{ f(x) - (\lambda^k)^\top (Ax + By^k - b) + \frac{1}{2} \|Ax + By^k - b\|_H^2 \mid x \in \mathcal{X} \right\} \quad (3)$$

$$\tilde{y}^k = \arg \min \left\{ g(y) - (\lambda^k)^\top (Ax^k + By - b) + \frac{1}{2} \|Ax^k + By - b\|_H^2 \mid y \in \mathcal{Y} \right\} \quad (4)$$

$$\tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b) \quad (5)$$

**步 2** 计算

$$w^{k+1} = w^k - \gamma^k \alpha_k^* (w^k - \tilde{w}^k) \quad (6)$$

其中

$$\gamma^k = \frac{1}{k} \sum_{i=1}^k \xi^i$$

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_G^2} \quad (7)$$

$$\varphi(w^k, \tilde{w}^k) = \|w^k - \tilde{w}^k\|_G^2 + (\lambda^k - \tilde{\lambda}^k)^\top [(Ax^k - A\tilde{x}^k) + (By^k - B\tilde{y}^k)] \quad (8)$$

**步 3** 计算

$$stopc = \frac{\|w^{k+1} - w^k\|}{\|w^k\|}, \quad k = k + 1$$

返回步 1。

**注 1**  $\{\xi^k\}$  为服从区间  $(a, b)$  上的某种独立同分布的随机分布序列,  $\xi^k$  的数学期望  $E(\xi^k) = \rho$ ;

**注 2** 当  $\xi^k$  取为固定值时, 该算法退化为并行分裂的 ADMM 下降算法。

#### 3.2. 收敛性

在该算法的收敛性分析中还需利用下述引理。

**引理 3.1** 当  $F(w)$  为单调连续算子时, 由新算法依次计算得到的  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  满足

$$\begin{pmatrix} x' - \tilde{x}^k \\ y' - \tilde{y}^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ F(\tilde{w}^k) + \begin{pmatrix} A^T H B (y^k - \tilde{y}^k) \\ B^T H B (y^k - \tilde{y}^k) \\ 0 \end{pmatrix} + \bar{G}(\tilde{w}^k - w^k) \right\} \geq 0 \quad (9)$$

证明 由(3)~(5)可得

$$(x' - \tilde{x}^k)^T \left\{ \nabla f(\tilde{x}^k) - A^T [\lambda^k - H(A\tilde{x}^k + B y^k - b)] \right\} \geq 0, \quad \forall x' \in \mathcal{X} \quad (10)$$

$$(y' - \tilde{y}^k)^T \left\{ \nabla g(\tilde{y}^k) - B^T [\lambda^k - H(Ax^k + B\tilde{y}^k - b)] \right\} \geq 0, \quad \forall y' \in \mathcal{Y} \quad (11)$$

$$\tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b)$$

将上述式子简单整理写为向量格式, 即可得到(9)。

**引理 3.2** 给定  $w^k = (x^k, y^k, \lambda^k)$ ,  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  是由(3)~(5)产生得到的,  $\varphi(w^k, \tilde{w}^k)$  由(8)所定义, 则有

$$\varphi(w^k, \tilde{w}^k) \geq \frac{2 - \sqrt{2}}{2} \|w^k - \tilde{w}^k\|_G^2 \quad (12)$$

证明 首先, 根据矩阵  $G$  和  $\varphi(w^k, \tilde{w}^k)$  的定义,

$$\begin{aligned} \varphi(w^k, \tilde{w}^k) &= \|A(x^k - \tilde{x}^k)\|_H^2 + \|B(y^k - \tilde{y}^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \\ &\quad + (Ax^k - A\tilde{x}^k)^T (\lambda^k - \tilde{\lambda}^k) + (By^k - B\tilde{y}^k)^T (\lambda^k - \tilde{\lambda}^k) \end{aligned} \quad (13)$$

根据 Cauchy-Schwarz 不等式, 得

$$(Ax^k - A\tilde{x}^k)^T (\lambda^k - \tilde{\lambda}^k) \geq -\frac{1}{2} \left( \sqrt{2} \|A(x^k - \tilde{x}^k)\|_H^2 + \frac{1}{\sqrt{2}} \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \right) \quad (14)$$

$$(By^k - B\tilde{y}^k)^T (\lambda^k - \tilde{\lambda}^k) \geq -\frac{1}{2} \left( \sqrt{2} \|B(y^k - \tilde{y}^k)\|_H^2 + \frac{1}{\sqrt{2}} \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \right) \quad (15)$$

把(14)(15)带入到(13)中, 得

$$\varphi(w^k, \tilde{w}^k) \geq \frac{2 - \sqrt{2}}{2} \left( \|A(x^k - \tilde{x}^k)\|_H^2 + \|B(y^k - \tilde{y}^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \right)$$

从而直接得到引理中 3.2 中的结果。

**引理 3.3** 给定  $w^k = (x^k, y^k, \lambda^k)$ ,  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  是由(3)~(5)产生得到的, 则对任意的  $w^* \in \mathcal{W}^*$ , 有

$$(\tilde{\lambda}^k - \lambda^*)^T H^{-1} (\lambda^k - \tilde{\lambda}^k) \geq (A\tilde{x}^k - Ax^*)^T H (By^k - B\tilde{y}^k) + (B\tilde{y}^k - By^*)^T H (Ax^k - A\tilde{x}^k) \quad (16)$$

证明 因为  $w^* \in \mathcal{W}^*$ ,  $\tilde{x}^k \in \mathcal{X}$ ,  $\tilde{y}^k \in \mathcal{Y}$ , 由(8)可得

$$\begin{pmatrix} \tilde{x}^k - x^* \\ \tilde{y}^k - y^* \end{pmatrix}^T \begin{pmatrix} \nabla f(x^*) - A^T \lambda^* \\ \nabla g(y^*) - B^T \lambda^* \end{pmatrix} \geq 0 \quad (17)$$

$$Ax^* + By^* - b = 0$$

另一方面, 因为  $x^* \in \mathcal{X}$ ,  $y^* \in \mathcal{Y}$ , 从(10)(11)可得到

$$\begin{pmatrix} x^* - \tilde{x}^k \\ y^* - \tilde{y}^k \end{pmatrix}^T \begin{pmatrix} \nabla f(\tilde{x}^k) - A^T \tilde{\lambda}^k + A^T H(B\tilde{y}^k - B\tilde{y}^k) \\ \nabla g(\tilde{y}^k) - B^T \tilde{\lambda}^k + B^T H(A\tilde{x}^k - A\tilde{x}^k) \end{pmatrix} \geq 0 \quad (18)$$

把(17)和(18)相加, 并利用  $\nabla f$  和  $\nabla g$  的单调性, 得

$$\begin{pmatrix} \tilde{x}^k - x^* \\ \tilde{y}^k - y^* \end{pmatrix}^T \begin{pmatrix} A^T(\tilde{\lambda}^k - \lambda^*) - A^T H(B\tilde{y}^k - B\tilde{y}^k) \\ B^T(\tilde{\lambda}^k - \lambda^*) - B^T H(A\tilde{x}^k - A\tilde{x}^k) \end{pmatrix} \geq 0$$

整理得到

$$\begin{aligned} & (\tilde{\lambda}^k - \lambda^*)^T \left[ (A\tilde{x}^k - Ax^*) + (B\tilde{y}^k - By^*) \right] \\ & \geq (A\tilde{x}^k - Ax^*)^T H(B\tilde{y}^k - B\tilde{y}^k) + (B\tilde{y}^k - By^*)^T H(A\tilde{x}^k - A\tilde{x}^k) \end{aligned} \quad (19)$$

由于  $Ax^* + By^* - b = 0$  且  $A\tilde{x}^k + B\tilde{y}^k - b = H^{-1}(\lambda^k - \tilde{\lambda}^k)$ , 从(19)我们直接得到(16)。

**引理 3.4** 给定  $w^k = (x^k, y^k, \lambda^k)$ ,  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  是由(3)~(5)产生得到的, 则对任意的  $w^* \in \mathcal{W}^*$ , 有

$$(w^k - w^*)^T G(w^k - \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k) \quad (20)$$

证明 首先, 使用矩阵  $G$  的定义, 有

$$\begin{aligned} & (\tilde{w}^k - w^*)^T G(w^k - \tilde{w}^k) \\ & = (A\tilde{x}^k - Ax^*)^T H(Ax^k - A\tilde{x}^k) + (B\tilde{y}^k - By^*)^T H(By^k - B\tilde{y}^k) \\ & \quad + (\tilde{\lambda}^k - \lambda^*)^T H^{-1}(\lambda^k - \tilde{\lambda}^k) \end{aligned} \quad (21)$$

将(16)代入(21)右边的最后一项得到

$$\begin{aligned} & (\tilde{w}^k - w^*)^T G(w^k - \tilde{w}^k) \\ & \geq (A\tilde{x}^k - Ax^*)^T H(Ax^k - A\tilde{x}^k) + (B\tilde{y}^k - By^*)^T H(By^k - B\tilde{y}^k) \\ & \quad + (B\tilde{y}^k - By^*)^T H(Ax^k - A\tilde{x}^k) + (A\tilde{x}^k - Ax^*)^T H(By^k - B\tilde{y}^k) \end{aligned}$$

利用  $Ax^* + By^* = b$ , 从上面的不等式我们得到

$$(\tilde{w}^k - w^*)^T G(w^k - \tilde{w}^k) \geq (A\tilde{x}^k + B\tilde{y}^k - b)^T H \left[ (Ax^k - A\tilde{x}^k) + (By^k - B\tilde{y}^k) \right] \quad (22)$$

因为  $H(A\tilde{x}^k + B\tilde{y}^k - b) = \lambda^k - \tilde{\lambda}^k$ , 再由  $\varphi(w^k, \tilde{w}^k)$  的定义, 我们可以直接得到(20)。

**引理 3.5** 给定  $w^k = (x^k, y^k, \lambda^k)$ ,  $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$  是由(3)~(5)产生得到的, 定义

$$w^{k+1}(\alpha) = w^k - \alpha(w^k - \tilde{w}^k) \quad (23)$$

$$\theta_k(\alpha) = \|w^k - w^*\|_G^2 - \|w^{k+1}(\alpha) - w^*\|_G^2, \forall w^* \in \mathcal{W}^* \quad (24)$$

则有

$$\theta_k(\alpha) \geq q_k(\alpha) = 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2 \|w^k - \tilde{w}^k\|_G^2, \forall \alpha > 0 \quad (25)$$

证明 由(23) (24)和(20)可知

$$\begin{aligned}
\theta_k(\alpha) &= \|w^k - w^*\|_G^2 - \|w^k - \alpha(w^k - \tilde{w}^k) - w^*\|_G^2 \\
&= 2\alpha(w^k - w^*)^T G(w^k - \tilde{w}^k) - \alpha^2 \|w^k - \tilde{w}^k\|_G^2 \\
&\geq 2\alpha\varphi(w^k, \tilde{w}^k) - \alpha^2 \|w^k - \tilde{w}^k\|_G^2
\end{aligned} \tag{26}$$

(26)的右侧是  $q_k(\alpha)$ , 得证。

**注1** 引理 3.5 中  $\theta_k(\alpha)$  是迭代点和最优点距离的下降量, 因其含有未知量  $w^*$ , 从而无法计算;  $q_k(\alpha)$  是  $\theta_k(\alpha)$  的一个下界函数, 此函数不含未知量  $w^*$ , 相对易于计算。

**注2** 注意到  $q_k(\alpha)$  是关于  $\alpha$  的二次函数, 最大值在  $\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_G^2}$  处取得, 且  $q_k(\alpha_k^*) = \alpha_k^* \varphi(w^k, \tilde{w}^k)$ 。

**定理 3.1** 当  $0 \leq a < E(\gamma^k) = \rho < b \leq 2$  时, 由(3)~(6)产生的迭代序列  $\{w^k\}$  收敛, 并且存在  $T_0$  ( $0 < T_0 < \infty$ ) 使得对任意的  $w^* \in \mathcal{W}^*$ ,  $k \rightarrow \infty$  有

$$E\left(\|w^{k+1} - w^*\|_G^2\right) \leq E\left(\|w^k - w^*\|_G^2\right) - T_0 \cdot E\left(\|w^{k+1} - w^k\|_G^2\right) \tag{27}$$

其中  $T_0 = (2 - \sqrt{2})\left(\frac{2}{\rho} - 1\right)$ 。

证明 由于  $G$  是对称正定阵, 根据(5)和  $0 \leq a < E(\gamma^k) = \rho < b \leq 2$ , 得

$$\begin{aligned}
\|w^{k+1} - w^*\|_G^2 &= \|w^k - \gamma^k \alpha_k^*(w^k - \tilde{w}^k) - w^*\|_G^2 \\
&= \|w^k - w^*\|_G^2 - 2\gamma^k \alpha_k^*(w^k - w^*)^T G(w^k - \tilde{w}^k) + (\gamma^k \alpha_k^*)^2 \|w^k - \tilde{w}^k\|_G^2 \\
&\leq \|w^k - w^*\|_G^2 - 2\gamma^k \alpha_k^* \varphi(w^k, \tilde{w}^k) + (\gamma^k \alpha_k^*)^2 \|w^k - \tilde{w}^k\|_G^2
\end{aligned} \tag{28}$$

$$\begin{aligned}
&= \|w^k - w^*\|_G^2 - \left[ 2\gamma^k \alpha_k^* \varphi(w^k, \tilde{w}^k) + (\gamma^k)^2 \alpha_k^* \alpha_k^* \|w^k - \tilde{w}^k\|_G^2 \right] \\
&= \|w^k - w^*\|_G^2 - \gamma^k (2 - \gamma^k) \alpha_k^* \varphi(w^k, \tilde{w}^k)
\end{aligned} \tag{29}$$

$$\leq \|w^k - w^*\|_G^2 - \frac{1}{2} \gamma^k (2 - \gamma^k) \varphi(w^k, \tilde{w}^k) \tag{30}$$

$$\leq \|w^k - w^*\|_G^2 - \frac{2 - \sqrt{2}}{4} \gamma^k (2 - \gamma^k) \|w^k - \tilde{w}^k\|_G^2 \tag{31}$$

其中(28)由(20)可得, 即  $2\gamma^k \alpha_k^*(w^k - w^*)^T G(w^k - \tilde{w}^k) \geq 2\gamma^k \alpha_k^* \varphi(w^k, \tilde{w}^k)$ ; (29)由  $\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|w^k - \tilde{w}^k\|_G^2}$  可得;

(30)由  $\alpha_k^* \geq 0.5$  可得; (31)由(12)可得。

根据(5)和  $\alpha_k^* \geq 0.5$ , 可得

$$w^{k+1} - w^k = -\gamma^k \alpha_k^*(w^k - \tilde{w}^k) \leq -\frac{1}{2} \gamma^k (w^k - \tilde{w}^k) = \frac{1}{2} \gamma^k (\tilde{w}^k - w^k)$$

即  $\frac{4}{(\gamma^k)^2} \|w^{k+1} - w^k\|_G^2 \leq \|w^k - \tilde{w}^k\|_G^2$ 。

故

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - (2 - \sqrt{2}) \left(\frac{2}{\gamma^k} - 1\right) \|w^{k+1} - w^k\|_G^2 \tag{32}$$



由(32)可知, 因  $\xi^1, \xi^2, \dots, \xi^k, \dots$  独立同分布, 且  $E(\gamma^k) = E\left(\frac{1}{k} \sum_{i=1}^k \xi^i\right) = \frac{1}{k} \sum_{i=1}^k E(\xi^i) = \rho$ , 所以

$$\begin{aligned} E\left(\|w^{k+1} - w^*\|_G^2\right) &\leq E\left(\|w^k - w^*\|_G^2 - (2-\sqrt{2})\left(\frac{2}{\gamma^k} - 1\right)\|w^{k+1} - w^k\|_G^2\right) \\ &= E\left(\|w^k - w^*\|_G^2\right) - E\left(\left(2-\sqrt{2}\right)\left(\frac{2}{\gamma^k} - 1\right)\|w^{k+1} - w^k\|_G^2\right) \\ &\leq E\left(\|w^k - w^*\|_G^2\right) - (2-\sqrt{2})\left(\frac{2}{E(\gamma^k)} - 1\right)E\left(\|w^{k+1} - w^k\|_G^2\right) \\ &= E\left(\|w^k - w^*\|_G^2\right) - (2-\sqrt{2})\left(\frac{2}{\rho} - 1\right)E\left(\|w^{k+1} - w^k\|_G^2\right) \end{aligned} \quad (33)$$

其中(33)是由 Jensen 不等式得到的,  $\frac{2}{\gamma^k} - 1$  是一个凸函数, 故  $E\left(\frac{2}{\gamma^k} - 1\right) \geq \frac{2}{E(\gamma^k)} - 1$ ; 再令

$T_0 = (2-\sqrt{2})\left(\frac{2}{\rho} - 1\right)$ 。即(28)得证。

**推论 3.1** 当  $0 \leq a < E(\gamma^k) = \rho < b \leq 2$  时,  $\{w^k\}$  为新算法产生的序列, 则有

- 1) 序列  $\{E(w^k)\}$  和  $\{E(\tilde{w}^k)\}$  是有界的;
- 2) 序列  $\left\{E\left(\|w^k - w^*\|_G^2\right)\right\}$  是严格单调递减且有界的;
- 3)  $\lim_{k \rightarrow \infty} E\left(\|w^k - \tilde{w}^k\|_G\right) = 0$ 。

下面给出新算法的依概率收敛的证明。

**定理 3.2** 由新算法产生的序列  $\{w^k\}$  依概率收敛于  $w^\infty \in \mathcal{W}^*$ , 即  $w^k \xrightarrow{P} w^\infty$ 。

证明 根据上面推论 3.1 中的 2), 有

$$\lim_{k \rightarrow \infty} E\left(\|A(x^k - \tilde{x}^k)\|_H\right) = 0, \quad \lim_{k \rightarrow \infty} E\left(\|B(y^k - \tilde{y}^k)\|_H\right) = 0 \quad (34)$$

$$\lim_{k \rightarrow \infty} E\left(\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}\right) = \lim_{k \rightarrow \infty} E\left(\|A\tilde{x}^k + B\tilde{y}^k - b\|_H\right) = 0 \quad (35)$$

对于  $\forall x' \in \mathcal{X}, \forall y' \in \mathcal{Y}$ , 由(10)(11)和(6)得:

$$\begin{aligned} (x' - \tilde{x}^k)^T \left\{ \nabla f(\tilde{x}^k) - A^T \left[ \tilde{\lambda}^k + H(A\tilde{x}^k + B\tilde{y}^k - b) \right] + A^T H(A\tilde{x}^k + B\tilde{y}^k - b) \right\} &\geq 0 \\ (y' - \tilde{y}^k)^T \left\{ \nabla g(\tilde{y}^k) - B^T \left[ \tilde{\lambda}^k + H(A\tilde{x}^k + B\tilde{y}^k - b) \right] + B^T H(A\tilde{x}^k + B\tilde{y}^k - b) \right\} &\geq 0 \end{aligned}$$

即:

$$(x' - \tilde{x}^k)^T \left( \nabla f(\tilde{x}^k) - A^T \tilde{\lambda}^k \right) \geq A^T H B (\tilde{y}^k - y^k) \quad (36)$$

$$(y' - \tilde{y}^k)^T \left( \nabla g(\tilde{y}^k) - B^T \tilde{\lambda}^k \right) \geq B^T H A (\tilde{x}^k - x^k) \quad (37)$$

由(34)和(36)(37)得

$$\begin{cases} \lim_{k \rightarrow \infty} (x' - \tilde{x}^k)^T \left( \nabla f(\tilde{x}^k) - A^T \tilde{\lambda}^k \right) \geq 0 \\ \lim_{k \rightarrow \infty} (y' - \tilde{y}^k)^T \left( \nabla g(\tilde{y}^k) - B^T \tilde{\lambda}^k \right) \geq 0 \end{cases} \quad (38)$$

由于  $\{E(\tilde{w}^k)\}$  有界, 故该序列至少存在一个聚点, 设  $E(w^\infty)$  是  $\{E(w^k)\}$  的一个聚点, 且存在子序列  $\{E(w^{k_j})\}$  收敛于  $E(w^\infty)$ , 那么子序列  $\{w^{k_j}\}$  依概率收敛到  $w^\infty$ , 即  $w^{k_j} \xrightarrow{P} w^\infty$ 。

再由(35)和(38)可得

$$\begin{cases} \lim_{j \rightarrow \infty} (x' - \tilde{x}^{k_j})^T (\nabla f(\tilde{x}^{k_j}) - A^T \tilde{\lambda}^{k_j}) \geq 0 \\ \lim_{j \rightarrow \infty} (y' - \tilde{y}^{k_j})^T (\nabla g(\tilde{y}^{k_j}) - B^T \tilde{\lambda}^{k_j}) \geq 0 \\ \lim_{j \rightarrow \infty} (A\tilde{x}^{k_j} + B\tilde{y}^{k_j} - b) = 0 \end{cases}$$

即

$$\begin{cases} (x' - x^\infty)^T (\nabla f(x^\infty) - A^T \lambda^\infty) \geq 0 \\ (y' - y^\infty)^T (\nabla g(y^\infty) - B^T \lambda^\infty) \geq 0 \\ (Ax^\infty + By^\infty - b) = 0 \end{cases}$$

因此,  $w^\infty \in \mathcal{W}^*$ 。

由  $\lim_{j \rightarrow \infty} E(\tilde{w}^{k_j}) = E(w^\infty)$  可知,  $\lim_{l \rightarrow \infty} E(\|\tilde{w}^{k_l} - w^\infty\|_G) = 0$ 。

又因为  $\lim_{k \rightarrow \infty} E(\|w^k - \tilde{w}^k\|_G) = 0$ , 则对任意  $\varepsilon > 0$ , 存在  $l \in \mathbb{Z}^+$  使得

$$E(\|w^{k_l} - \tilde{w}^{k_l}\|) < \frac{\varepsilon^2}{2}, \quad E(\|\tilde{w}^{k_l} - w^\infty\|) < \frac{\varepsilon^2}{2}$$

根据推论,  $\{E(\|w^k - w^*\|_G^2)\}$  是严格单调递减的。因此, 对任意  $k \geq k_l$ , 有

$$E(\|w^k - w^\infty\|_G) \leq E(\|w^{k_l} - w^\infty\|_G) \leq E(\|w^{k_l} - \tilde{w}^{k_l}\|_G) + E(\|\tilde{w}^{k_l} - w^\infty\|_G) < \varepsilon^2 \quad (39)$$

即  $\lim_{k \rightarrow \infty} E(\|w^k - w^\infty\|_G) = 0$ 。

由定理 2.2 和(39)可知, 对任意  $\varepsilon > 0$ , 存在  $l > 0$ , 只要  $k > l$ , 就有

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

即  $\lim_{k \rightarrow \infty} P(\|w^k - w^\infty\|_G \geq \varepsilon) = 0$ 。

故  $\{w^k\}$  依概率收敛到  $w^\infty \in \mathcal{W}^*$ , 即  $w^k \xrightarrow{P} w^\infty$ 。

#### 4. 数值实验

我们考虑求解下列等式约束二次规划问题

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{2} x^T P x + p^T x + \frac{1}{2} y^T Q y + q^T y \\ \text{s.t.} \quad & A x + B y = b \end{aligned}$$

其中  $P$  和  $Q$  是对称正定阵,  $A \in \mathcal{R}^{m \times n_1}$ ,  $B \in \mathcal{R}^{m \times n_2}$ ,  $x \in \mathcal{R}^{n_1}$ ,  $y \in \mathcal{R}^{n_2}$ ,  $b \in \mathcal{R}^m$ ,  $p \in \mathcal{R}^{n_1}$ ,  $q \in \mathcal{R}^{n_2}$ 。

我们令  $H = \beta I$ ，其中的最优取值是通过反复测试确定的。同时， $\gamma^k$  是在区间  $(0, 2)$  上选取的服从正态分布的随机数。在新算法运行前，我们对矩阵  $P, Q, A^T A, B^T B$  进行 Cholesky 分解，来化简对子问题的求解。我们取

$$\text{stopc} = \left\| \left( \tilde{x}^k - x^k, \tilde{y}^y - y^y, A\tilde{x}^k + B\tilde{y}^y - b \right) \right\|_{\infty} \leq 10^{-12}$$

作为停机准则。

我们在每组  $(m, n_1, n_2)$  设置下，测试了几组不同设置的问题。然后将几种算法数值实验的结果进行比较，见表 1。

**Table 1.** Numerical experimental results of four algorithms

**表 1.** 四种算法的数值实验结果

数据维度	算法	经典 ADMM 算法 (ADMM)		ADMM 下降算法 (DADMM)		带随机步长的 ADMM 下降算法 (SDADMM)		新算法 (SPDADMM)	
		迭代步数	CPU 时间	迭代步数	CPU 时间	迭代步数	CPU 时间	迭代步数	CPU 时间
(100, 200, 200)		331	0.079	270	0.068	215	0.057	202	0.045
(200, 200, 200)		420	0.255	316	0.190	304	0.181	282	0.174
(400, 500, 500)		659	0.944	508	0.789	459	0.629	430	0.547
(500, 500, 500)		706	1.437	537	1.149	463	1.104	438	1.058
(500, 1000, 1000)		793	2.689	627	2.128	604	2.074	591	1.980
(1000, 1000, 1000)		1001	6.182	816	4.825	795	4.730	764	4.271

## 5. 结语

本文基于并行分裂 ADMM 下降算法和改善步长的收缩算法，提出了一种求解结构型凸优化问题的新算法——带随机步长的并行分裂 ADMM 下降算法，文章证明了新算法的收敛性。数值实验结果显示，在求解结构型凸优化问题时，新算法的迭代步数及 CPU 运行时间均优于文章中的其他 4 类算法，并且新算法的计算效率对数据维度的增长有较好的鲁棒性，进一步说明新算法可用于处理一些高维数据的问题。

本文在 ADMM 算法基础上进行算法改进的研究，并取得了一些新的成果，但还有许多问题值得我们更深一步的研究。在随后的研究中，我们主要考虑以下几个方面：

- 1) 本文考虑的目标函数是带有两块的结构型凸优化问题，后面还可以考虑新算法能否推广到带有三块及多块的结构型凸优化问题的情况；
- 2) 能否加入 Nesterov 加速等其他加速算法，使算法的计算效率进一步提高；
- 3) 以后还可测试一些实际问题，比如图像降噪、信号处理等问题。

## 参考文献

- [1] Glowinski, R. and Marroco, A. (1975) Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis*, **9**, 41-76. <https://doi.org/10.1051/m2an/197509R200411>
- [2] Gabay, D. and Mercier, B. (1976) A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations. *Computers and 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] Li, G.Y. and Pong, T.K. (2015) Global Convergence of Splitting Methods for Nonconvex Composite Optimization. *SIAM Journal on Optimization*, **25**, 2434-2460. <https://doi.org/10.1137/140998135>
- [4] Betts, J.T. (1978) A Gradient Projection-Multiplier Method for Nonlinear Programming. *Journal of Optimization Theory and Applications*, **24**, 523-548. <https://doi.org/10.1007/BF00935298>
- [5] Root, R.R. and Ragsdell, K.M. (1980) On the Relationship between a Continuous Update Method of Multipliers and the Generalized Reduced Gradient Method. *International Journal for Numerical Methods in Engineering*, **15**, 1735-1745. <https://doi.org/10.1002/nme.1620151203>
- [6] Hestenes, M.R. (1969) Multiplier and Gradient Methods. *Journal of Optimization Theory & Applications*, **4**, 303-320. <https://doi.org/10.1007/BF00927673>
- [7] Cai, X.J., Gu, G.Y., He, B.S., et al. (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>
- [8] Giselsson, P. and Boyd, S. (2014) Linear Convergence and Metric Selection for Douglas-Rachford Splitting and ADMM. *IEEE Transactions on Automatic Control*, **62**, 532-544. <https://doi.org/10.1109/TAC.2016.2564160>
- [9] Cai, X.J., Gu, G.Y., He, B.S., et al. (2011) A Relaxed Customized Proximal Point Algorithm for Separable Convex Programming. Optimization.
- [10] He, B.S. and Yuan, X.M. (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>
- [11] Boyd, S., Parikh, N., Chu, E., et al. (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>
- [12] He, B.S. and Yuan, X.M. (2015) 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>
- [13] He, B.S., Yang, H. and Wang, S.L. (2000) Alternating Direction Method with Self-Adaptive Penalty Parameters for Monotone Variational Inequalities. *Journal of Optimization Theory & Applications*, **106**, 337-356. <https://doi.org/10.1023/A:1004603514434>
- [14] Gabay, D. and Mercier, B. (1976) A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximations. *Computers and 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)
- [15] Shi, W., Ling, Q., Yuan, K., et al. (2014) On the Linear Convergence of the ADMM in Decentralized Consensus Optimization. *IEEE Transactions on Signal Processing*, **62**, 1750-1761. <https://doi.org/10.1109/TSP.2014.2304432>
- [16] Cao, J., Liu, S., Liu, H., et al. (2021) MRI Reconstruction Based on Bayesian Group Sparse Representation. *Signal Processing*, **187**, Article ID: 108151. <https://doi.org/10.1016/j.sigpro.2021.108151>
- [17] Du, S., Shi, Y., Hu, W., et al. (2020) Robust Tensor Factorization for Color Image and Grayscale Video Recovery. *IEEE Access*, **8**, 174410-174423. <https://doi.org/10.1109/ACCESS.2020.3024635>
- [18] Chan, S.H. (2019) Performance Analysis of Plug-and-Play ADMM: A Graph Signal Processing Perspective. *IEEE Transactions on Computational Imaging*, Volume number, 1-20.
- [19] Bai, J., Liang, J., Guo, K., et al. (2021) Accelerated Symmetric ADMM and Its Applications in Large-Scale Signal Processing.
- [20] Yue, H., Chi, E.C. and Allen, G.I. (2016) ADMM Algorithmic Regularization Paths for Sparse Statistical Machine Learning. Springer International Publishing, Berlin.
- [21] 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>
- [22] He, B.S. (2009) Parallel Splitting Augmented Lagrangian Methods for Monotone Structured Variational Inequalities. *Computational Optimization and Applications*, **42**, 195-212. <https://doi.org/10.1007/s10589-007-9109-x>
- [23] Xu, H.W. (2011) A Contraction Method with Random Step Size for a Class of Variational Inequalities. *Chinese Journal of Engineering Mathematics*, **28**, 461-469.
- [24] He, B.S., Liao, L.Z., Han, D.R., et al. (2002) A New Inexact Alternating Directions Method for Monotone Variational Inequalities. *Mathematical Programming*, **92**, 103-118. <https://doi.org/10.1007/s101070100280>
- [25] He, B.S., Liao, L.Z. and Wang, X.F. (2012) Proximal-Like Contraction Methods for Monotone Variational Inequalities in a Unified Framework I: Effective Quadruplet and Primary Methods. *Computational Optimization and Applications*, **51**, 649-679. <https://doi.org/10.1007/s10589-010-9372-0>
- [26] Tao, M. and Yuan, X.M. (2012) On the  $O(1/t)$  Convergence Rate of Alternating Direction Method with Logarithm-

mic-Quadratic Proximal Regularization. *SIAM Journal on Optimization*, **22**, 1431-1448.  
<https://doi.org/10.1137/110847639>

[27] 林正炎, 等. 概率极限理论基础[M]. 北京: 高等教育出版社, 2003.

[28] 何炳生. 凸优化的一阶分裂算法——变分不等式作为工具的统一框架[EB/OL]. <http://math.nju.edu.cn/~hebma>