Appl. Set-Valued Anal. Optim. 4 (2022), No. 2, pp. 129-143
Available online at http://asvao.biemdas.com
https://doi.org/10.23952/asvao.4.2022.2.01

A BREGMAN PROXIMAL PEACEMAN-RACHFORD SPLITTING METHOD FOR CONVEX PROGRAMMING

ABDELLAH BNOUHACHEM[1,*], MICHAEL TH. RASSIAS[2]

[1]*Equipe MAISI, Ibn Zohr University, ENSA, BP 1136, Agadir, Morocco*
[2]*Institute of Mathematics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*

Abstract. In this paper, based on the Bregman distance and the proximal Peaceman-Rachford splitting method (PRSM), we propose a Bregman proximal Peaceman-Rachford splitting method for solving a separable convex minimization model. We establish the relationship between two parameters under which we prove the global convergence of the algorithm. The $O(1/t)$ convergence rate of the proposed method in the ergodic sense is also studied. Preliminary numerical experiments are included to illustrate the advantage and efficiency of the proposed method.
Keywords. Alternating direction method; Bregman distance; Peaceman-Rachford splitting method; Variational inequalities.

1. INTRODUCTION

We consider the constrained convex programming problem with the following separate structure:

$$\min\{\theta_1(x)+\theta_2(y)|Ax+By=b, x\in\mathscr{X}, y\in\mathscr{Y}\}, \tag{1.1}$$

where $\mathscr{X}\subset\mathbb{R}^{n_1}$ and $\mathscr{Y}\subset\mathbb{R}^{n_2}$ are closed convex sets, $\theta_1:\mathscr{X}\to\mathbb{R}$ and $\theta_2:\mathscr{Y}\to\mathbb{R}$ are closed proper convex functions, $A\in\mathbb{R}^{l\times n_1}$ and $B\in\mathbb{R}^{l\times n_2}$ are given matrices, and $b\in\mathbb{R}^l$.

The alternating direction multiplier method (ADMM) is one of most powerful and successful methods for solving (1.1), and it is a general framework for optimizing composite functions; see, e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9]. During the years which have been elapsed since its discovery, ADMM has been gathering much attentions because it has a broad class of applications in the fields of image and signal processing, and machine learning [5]. A number of algorithms have been established for ADMM as well as its varieties. Some of these variants included proximal terms in the subproblems of the ADMM in order to make them easier to solve. Recently, there has been an increasing interest in the study of ADMM by adding a step size parameter in the Lagrangian multiplier updating to improve the performance of the method; see, e.g., [10, 11, 12, 13, 14, 15]. Recently, various inexact of ADMM have been suggested; see, e.g., [16, 17, 18, 19] and the references therein.

*Corresponding author.
E-mail addresses: babedallah@yahoo.com (A. Bnouhachem), michail.rassias@math.uzh.ch (M.Th. Rassias).
This research is dedicated to Tiddas.
Received June 8, 2022; Accepted July 2, 2022.

The augmented Lagrangian function of (1.1) is defined by

$$L_\beta(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^\top(Ax+By-b)+\frac{\beta}{2}\|Ax+By-b\|^2,$$

where $\lambda\in\mathbb{R}^l$ is the Lagrange multiplier, and $\beta>0$ is a penalty parameter. Some operator splitting methods [8, 9, 20, 21, 22] were developed for solving the dual problem of (1.1). The Peaceman-Rachford operator splitting method (PRSM) [23], which is also a splitting method and different from the ADMM, updates the Lagrange multiplier twice at each iteration. Based on PRSM, He *et al.* [24] proposed a strictly contractive PRSM by introducing a parameter $\alpha\in(0,1)$ to the update scheme of the dual variable, yielding the following procedure:

$$\begin{cases} x^{k+1}=\arg\min\left\{L_\beta(x,y^k,\lambda^k)|x\in\mathscr{X}\right\},\\ \lambda^{k+\frac{1}{2}}=\lambda^k-\alpha\beta(Ax^{k+1}+By^k-b),\\ y^{k+1}=\arg\min\left\{L_\beta(x^{k+1},y,\lambda^{k+\frac{1}{2}})|y\in\mathscr{Y}\right\},\\ \lambda^{k+1}=\lambda^{k+\frac{1}{2}}-\alpha\beta(Ax^{k+1}+By^{k+1}-b). \end{cases}$$

Inspired by [24], Gu *et al.* [25] proposed a modification of the Peaceman-Rachford splitting method by introducing two different parameters in updating the dual variable, and by introducing semi-proximal terms to the subproblems in updating the primal variables. From a given $w^k=(x^k,y^k,\lambda^k)\in\mathscr{W}$, the new iterate $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})$ is obtained via solving the following:

$$\begin{cases} x^{k+1}=\arg\min\left\{L_\beta(x,y^k,\lambda^k)+\frac{1}{2}\|x-x^k\|_L^2|x\in\mathscr{X}\right\},\\ \lambda^{k+\frac{1}{2}}=\lambda^k-\alpha\beta(Ax^{k+1}+By^k-b),\\ y^{k+1}=\arg\min\left\{L_\beta(x^{k+1},y,\lambda^{k+\frac{1}{2}})+\frac{1}{2}\|y-y^k\|_T^2|y\in\mathscr{Y}\right\},\\ \lambda^{k+1}=\lambda^{k+\frac{1}{2}}-\gamma\beta(Ax^{k+1}+By^{k+1}-b), \end{cases}$$

where L and T are two positive semi-definite matrices. The global convergence is proved under $\alpha\in(0,1)$ and $\gamma\in\left(0,\frac{1-\alpha+\sqrt{(1-\alpha)^2+4(1-\alpha^2)}}{2}\right)$.

Recently, He *et al.* [26] proposed the strictly contractive PRSM with the step-size enlarged by Fortin and Glowinski's constant in [27, 28, 29]. Its iterative scheme is as following:

$$\begin{cases} x^{k+1}=\arg\min\left\{L_\beta(x,y^k,\lambda^k)|x\in\mathscr{X}\right\},\\ \lambda^{k+\frac{1}{2}}=\lambda^k-\alpha\beta(Ax^{k+1}+By^k-b),\\ y^{k+1}=\arg\min\left\{L_\beta(x^{k+1},y,\lambda^{k+\frac{1}{2}})|y\in\mathscr{Y}\right\},\\ \lambda^{k+1}=\lambda^{k+\frac{1}{2}}-\gamma\beta(Ax^{k+1}+By^{k+1}-b), \end{cases}$$

in which α and γ are independent constants that are restricted into the domain

$$\mathscr{D}=\left\{(\alpha,\gamma)|\alpha\in(-1,1),\gamma\in(0,\frac{1+\sqrt{5}}{2}),\alpha+\gamma>0\quad\text{and}\quad|\alpha|<1+\gamma-\gamma^2\right\}.$$

Recently, the Bregman modification of ADMM (BADMM) has been adopted by several researchers to improve the performance of the ADMM algorithm; see, e.g., [30, 31]. More specifically, BADMM takes the following iterative form:

$$\begin{cases} x^{k+1} = \arg\min\left\{L_\beta(x,y^k,\lambda^k) + B_\phi(x,x^k) | x \in \mathscr{X}\right\}, \\ y^{k+1} = \arg\min\left\{L_\beta(x^{k+1},y,\lambda^k + B_\psi(y,y^k) | y \in \mathscr{Y}\right\}, \\ \lambda^{k+1} = \lambda^k + \alpha\beta(Ax^{k+1} + By^{k+1} - b), \end{cases}$$

where $B_\phi(x,x^k)$ and $B_\psi(y,y^k)$ are the Bregman distance with respect to functions ϕ and ψ, respectively.

This present work is impelled by the research going on this field. The main contribution of this paper is to show that the recently proposed the strictly contractive PRSM [26] can be well integrated with the Bregman distance for solving problem (1.1). Since the choice of the step size selection strategies is important for the algorithms efficiency, we establish the relationship between two parameters under which we prove the global convergence of the algorithm. Our results can be viewed as significant extensions of the previously known results.

2. Preliminaries

We state some preliminaries that are useful in later analysis. Let $\mathbb{R}^n$ stand for the n-dimensional Euclidean space. For any vector $u \in \mathbb{R}^n$, $\|u\|^2 = u^\top u$. Let $D \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. We denote the D-norm of u by $\|u\|_D^2 = u^\top Du$.

2.1. **Bregman distance.**

The Bregman distance, which can be viewed as a generalization of the squared Euclidean distance, was proposed by [32]. For a convex differential function ϕ, the associated Bregman distance is defined as

$$B_\phi(x,y) := \phi(x) - \phi(y) - \nabla\phi(y)^\top(x-y).$$

In particular, if we let $\phi(x) = \|x\|^2$, then $B_\phi(x,y) = \|x-y\|^2$, i.e., the classical Euclidean distance. If $\phi(x) = x^\top Qx$ with Q a symmetric positive definite matrix, then $B_\phi(x,y) = \|x-y\|_Q^2$, i.e., Mahalanobis distance.

Let us now collect some basic properties of the Bregman distance.

Proposition 2.1. [31] *Let ϕ be a convex differentiable function and $B_\phi(x,y)$ be the associated Bregman distance.*

- *$B_\phi(x,y) \geq 0$ and $B_\phi(x,x) = 0$ for all x,y.*
- *$B_\phi(x,y)$ is convex in x, but not necessarily in y.*

Proposition 2.2. *For differentiable convex functions Φ, it holds that*

$$(a-b)^\top(\nabla\Phi(c) - \nabla\Phi(d)) = B_\Phi(a,d) - B_\Phi(a,c) + B_\Phi(b,c) - B_\Phi(b,d), \quad \textit{for all} \quad a,b,c,d.$$

Proof. Using the definition of B_Φ, we have

$$B_\Phi(a,d) = \Phi(a) - \Phi(d) - \nabla\Phi(d)^\top(a-d),$$
$$B_\Phi(a,c) = \Phi(a) - \Phi(c) - \nabla\Phi(c)^\top(a-c),$$
$$B_\Phi(b,c) = \Phi(b) - \Phi(c) - \nabla\Phi(c)^\top(b-c),$$
$$B_\Phi(b,d) = \Phi(b) - \phi(d) - \nabla\Phi(d)^\top(b-d).$$

By simple manipulations, we obtain

$$B_\Phi(a,d) - B_\Phi(a,c) + B_\Phi(b,c) - B_\Phi(b,d) = (a-b)^\top(\nabla\Phi(c) - \nabla\Phi(d)). \tag{2.1}$$

This completes the proof. □

Using the Proposition 2.2, we have the following proposition.

Proposition 2.3. *For differentiable convex functions ϕ and ψ, it holds that*

$$(x-\tilde{x}^k)^\top(\nabla\phi(x^k)-\nabla\phi(\tilde{x}^k)) = B_\phi(x,\tilde{x}^k)-B_\phi(x,x^k)+B_\phi(\tilde{x}^k,x^k)$$

for all $x, x^k, \tilde{x}^k$, and

$$(y-\tilde{y}^k)^\top(\nabla\psi(y^k)-\nabla\psi(\tilde{y}^k)) = B_\psi(y,\tilde{y}^k)-B_\psi(y,y^k)+B_\psi(\tilde{y}^k,y^k)$$

for all $y, y^k, \tilde{y}^k$.

2.2. **Variational inequality characterization.** Let the Lagrangian function of (1.1) be

$$L(x,y,\lambda) = \theta_1(x)+\theta_2(y)-\lambda^\top(Ax+By-b),$$

which is defined on $\mathscr{W} = \mathscr{X}\times\mathscr{Y}\times\mathbb{R}^l$. We call $w^* = (x^*,y^*,\lambda^*)\in\mathscr{W}$ a saddle point of the Lagrangian function if it satisfies

$$L(x^*,y^*,\lambda)\le L(x^*,y^*,\lambda^*)\le L(x,y,\lambda^*) \quad \forall (x,y,\lambda)\in\mathscr{W}.$$

Accordingly, for any$(x,y,\lambda)\in\mathscr{W}$, we deduce the following inequalities

$$\begin{cases} \theta_1(x)+\theta_1(y)-(\theta_2(x^*)+\theta_2(y^*))+(x-x^*)^\top(-A^\top\lambda^*)+(y-y^*)^\top(-B^\top\lambda^*)\ge 0, \\ (\lambda-\lambda^*)^\top(Ax^*+By^*-b)\ge 0. \end{cases}$$

Then, problem (1.1) is equivalent to the mixed variational inequalities: find $w^*\in\mathscr{W}$ such that

$$\theta(u)-\theta(u^*)+(w-w^*)^\top F(w^*)\ge 0, \quad \forall\, w\in\mathscr{W}, \tag{2.2}$$

where

$$u=\begin{pmatrix} x\\ y\end{pmatrix}, \quad \theta(u)=\theta_1(x)+\theta_2(y), \quad v=\begin{pmatrix} y\\ \lambda\end{pmatrix}, \quad w=\begin{pmatrix} x\\ y\\ \lambda\end{pmatrix}, \text{ and } F(w)=\begin{pmatrix} -A^\top\lambda\\ -B^\top\lambda\\ Ax+By-b.\end{pmatrix}.$$

Clearly, F is monotone, i.e., $(F(w_1)-F(w_2))^\top(w_1-w_2)\ge 0, \forall w_1,w_2\in\mathscr{W}$.

2.3. **Some notations.** Our analysis needs several matrices defined by

$$M=\begin{pmatrix} I_{n_2} & 0\\ -\gamma\beta B & (\alpha+\gamma)I_l\end{pmatrix}, \quad Q=\begin{pmatrix} \beta B^\top B & -\alpha B^\top\\ -B & \frac{1}{\beta}I_l\end{pmatrix}, \tag{2.3}$$

$$H=\begin{pmatrix} \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\top B & -\frac{\alpha}{\alpha+\gamma}B^T\\ -\frac{\alpha}{\alpha+\gamma}B & \frac{1}{(\alpha+\gamma)\beta}I_l\end{pmatrix} \quad G=\begin{pmatrix} (1-\gamma)\beta B^\top B & (\gamma-1)B^\top\\ (\gamma-1)B & \frac{(2-\alpha-\gamma)}{\beta}I_l\end{pmatrix} \tag{2.4}$$

Lemma 2.1. *If $\gamma>0$, $\alpha\in(-1,1)$, and $\alpha+\gamma>0$, then*

(1) *the matrices M,Q and H,G defined respectively in (2.3) and (2.4) satisfy*

$$HM = Q \text{ and } G := Q+Q^T-M^THM; \tag{2.5}$$

(2) *H is positive definite.*

Proof. (1) holds evidently. Next we prove (2) only. Since $\gamma > 0$, $\alpha \in (-1,1)$, and $\alpha+\gamma > 0$, we have $(1-\alpha)(\alpha+\gamma) > 0$ and $\alpha+\gamma-\alpha\gamma > \alpha^2$. Then, for any $v = (y,\lambda) \neq (0,0)$,

$$\begin{aligned} v^\top H v &= \frac{1}{(\alpha+\gamma)\beta}\left((\alpha+\gamma-\alpha\gamma)\beta^2\|By\|^2 - 2\alpha\beta y^\top B^\top\lambda + \|\lambda\|^2\right) \\ &> \frac{1}{(\alpha+\gamma)\beta}\left(\alpha^2\beta^2\|By\|^2 - 2\alpha\beta y^\top B^\top\lambda + \|\lambda\|^2\right) \\ &= \frac{1}{(\alpha+\gamma)\beta}\|\alpha\beta By - \lambda\|^2 \\ &\geq 0. \end{aligned}$$

Thus H is positive definite. This completes the proof. □

3. The Proposed Method

Let ϕ and ψ be convex differentiable functions. We propose the following PRSM with the Bregman distance for solving (1.1).

Algorithm 3.1. Step 0. *The initial step:*
Give $\varepsilon > 0$, $\beta > 0$, $\alpha \in (-1,1)$, $\gamma \in (0,1)$, $\alpha+\gamma > 0$, and $w^0 = (x^0, y^0, \lambda^0) \in \mathscr{X} \times \mathscr{Y} \times \mathbb{R}^l$. Set $k = 0$.

Step 1. Compute $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \mathscr{X} \times \mathscr{Y} \times \mathbb{R}^l$ by solving the following system:

$$x^{k+1} = \arg\min\Big\{\theta_1(x) - (\lambda^k)^\top(Ax + By^k - b) + \frac{\beta}{2}\|Ax + By^k - b\|^2 + B_\phi(x, x^k), \forall x \in \mathscr{X}\Big\}, \tag{3.1a}$$

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), \tag{3.1b}$$

$$y^{k+1} = \arg\min\Big\{\theta_2(y) - (\lambda^{k+\frac{1}{2}})^\top(Ax^{k+1} + By - b) + \frac{\beta}{2}\|Ax^{k+1} + By - b\|^2 + B_\psi(y, y^k), \forall y \in \mathscr{Y}\Big\}, \tag{3.1c}$$

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \gamma\beta(Ax^{k+1} + By^{k+1} - b). \tag{3.1d}$$

Step 2. If $\max\{\|By^k - By^{k+1}\|^2, \|Ax^{k+1} + By^{k+1} - b\|^2\} < \varepsilon$, stop; otherwise set $k = k+1$ and go to Step 1.

Remark 3.1. According to their first-order optimality conditions, the minimization problems in (3.1) can be characterized as: find $(x^{k+1}, y^{k+1}) \in \mathscr{X} \times \mathscr{Y}$ such that, for any $(x,y) \in \mathscr{X} \times \mathscr{Y}$,

$$\theta_1(x) - \theta_1(x^{k+1}) + (x - x^{k+1})^\top\left(-A^\top[\lambda^k - \beta(Ax^{k+1} + By^k - b)] + \nabla\phi(x^{k+1}) - \nabla\phi(x^k)\right) \geq 0, \tag{3.2}$$

$$\theta_2(y) - \theta_2(y^{k+1}) + (y - y^{k+1})^\top\left(-B^\top[\lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b)] + \nabla\psi(y^{k+1}) - \nabla\psi(y^k)\right) \geq 0, \tag{3.3}$$

Throughout this paper, we make the following standard assumptions:
Assumption A. The matrix B is full-column rank.
Assumption B. The solution set of (1.1), denoted by $\mathscr{W}^*$, is nonempty. We use $\mathscr{V}^*$ to denote the set of v^* for all subvectors of w^* in $\mathscr{W}^*$.

Our analysis needs some sequences defined as follows:

$$\tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} x^{k+1} \\ y^{k+1} \\ \lambda^k - \beta(Ax^{k+1} + By^k - b) \end{pmatrix} \quad \text{and} \quad \tilde{v}^k = \begin{pmatrix} \tilde{y}^k \\ \tilde{\lambda}^k \end{pmatrix}. \tag{3.4}$$

$$\begin{aligned} \lambda^{k+1} &= \lambda^{k+\frac{1}{2}} - \gamma\beta(A\tilde{x}^k + B\tilde{y}^k - b) \\ &= \lambda^k - (\alpha+\gamma)\beta(A\tilde{x}^k + By^k - b) + \gamma\beta(By^k - B\tilde{y}^k) \\ &= \lambda^k - [(\alpha+\gamma)(\lambda^k - \tilde{\lambda}^k) - \gamma\beta B(y^k - \tilde{y}^k)]. \end{aligned}$$

Thus $v^{k+1} = v^k - M(v^k - \tilde{v}^k)$, where M is defined in (2.3).

Lemma 3.1. *Let w^k be generated by Algorithm 3.1 and $\tilde{w}^k$ be defined by (3.4). Then, for any $v = (y, \lambda) \in \mathscr{V}$,*

$$(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) + \frac{1}{2}\|v^k - \tilde{v}^k\|_G^2.$$

Proof. From (2.5), it follows that $(v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k) = (v - \tilde{v}^k)^T H(v^k - v^{k+1})$. Using the following identity $(a-b)^\top H(c-d) = \frac{1}{2}(\|a-d\|_H^2 - \|a-c\|_H^2) + \frac{1}{2}(\|c-b\|_H^2 - \|d-b\|_H^2)$, for $a = v, b = \tilde{v}^k, c = v^k, d = v^{k+1}$, we obtain

$$\begin{aligned} (v - \tilde{v}^k)^T H(v^k - v^{k+1}) &= \frac{1}{2}(\|v - v^{k+1}\|_H^2 - \|v - v^k\|_H^2) \\ &\quad + \frac{1}{2}(\|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2). \end{aligned} \tag{3.5}$$

For the last term in (3.5), we have

$$\begin{aligned} \|v^k - \tilde{v}^k\|_H^2 - \|v^{k+1} - \tilde{v}^k\|_H^2 &= \|v^k - \tilde{v}^k\|_H^2 - \|(v^k - \tilde{v}^k) - M(v^k - \tilde{v}^k)\|_H^2 \\ &= 2(v^k - \tilde{v}^k)^T HM(v^k - \tilde{v}^k) - (v^k - \tilde{v}^k)^T M^T HM(v^k - \tilde{v}^k) \\ &= (v^k - \tilde{v}^k)^T (Q^T + Q - M^T HM)(v^k - \tilde{v}^k) \\ &= \|v^k - \tilde{v}^k\|_G^2. \end{aligned} \tag{3.6}$$

Combining (3.5), and (3.6), we can obtain the assertion of this lemma. □

In the following, we prove some properties which are useful for establishing the main result. The first lemma presents a lower bound of $\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) + (v - \tilde{v}^k)^\top Q(\tilde{v}^k - v^k)$.

Lemma 3.2. *Let w^k be generated by Algorithm 3.1. Then, for any $w \in \mathscr{W}$,*

$$\begin{aligned} &\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) + (v - \tilde{v}^k)^\top Q(\tilde{v}^k - v^k) \\ &\geq (u - \tilde{u}^k)^\top (\nabla\varphi(u^k) - \nabla\varphi(\tilde{u}^k)), \end{aligned} \tag{3.7}$$

where $\varphi(u) = \phi(x) + \psi(y)$.

Proof. It follows from (3.2) that

$$\begin{aligned}&\theta_1(x^{k+1})-\theta_1(x)+(x^{k+1}-x)^\top\Big\{-A^\top\lambda^k+\beta A^\top(Ax^{k+1}+By^k-b)\Big\}\\&\le(x-x^{k+1})^\top(\nabla\phi(x^{k+1})-\nabla\phi(x^k)),\end{aligned}$$

and it can be written as

$$\theta_1(\tilde{x}^k)-\theta_1(x)+(\tilde{x}^k-x)^\top(-A^\top\tilde{\lambda}^k)\le(x-\tilde{x}^k)^\top(\nabla\phi(\tilde{x}^k)-\nabla\phi(x^k)). \tag{3.8}$$

From (3.3), we have

$$\begin{aligned}&\theta_2(y^{k+1})-\theta_2(y)+(y^{k+1}-y)^\top\Big\{-B^\top[\lambda^{k+\frac{1}{2}}-\beta(Ax^{k+1}+By^{k+1}-b)]\Big\}\\&\le(y-y^{k+1})^\top(\nabla\psi(y^{k+1})-\nabla\psi(y^k)).\end{aligned}$$

Using (3.1b), we obtain

$$\begin{aligned}&\theta_2(\tilde{y}^k)-\theta_2(y)+(\tilde{y}^k-y)^\top\Big\{-B^\top[\tilde{\lambda}^k+(\alpha-1)(\tilde{\lambda}^k-\lambda^k)-\beta(A\tilde{x}^k+B\tilde{y}^k-b)]\Big\}\\&\le(y-\tilde{y}^k)^\top(\nabla\psi(\tilde{y}^k)-\nabla\psi(y^k)).\end{aligned}$$

Thus

$$\begin{aligned}&\theta_2(\tilde{y}^k)-\theta_2(y)+(\tilde{y}^k-y)^\top\Big\{-B^\top[\tilde{\lambda}^k+(\alpha-1)(\tilde{\lambda}^k-\lambda^k)+(\tilde{\lambda}^k-\lambda^k)+\beta(By^k-B\tilde{y}^k)]\Big\}\\&\le(y-\tilde{y}^k)^\top(\nabla\psi(\tilde{y}^k)-\nabla\psi(y^k)),\end{aligned}$$

and

$$\begin{aligned}&\theta_2(\tilde{y}^k)-\theta_2(y)+(\tilde{y}^k-y)^\top[-B^\top\tilde{\lambda}^k-\alpha B^\top(\tilde{\lambda}^k-\lambda^k)-\beta B^\top B(y^k-\tilde{y}^k)]\\&\le(y-\tilde{y}^k)^\top(\nabla\psi(\tilde{y}^k)-\nabla\psi(y^k)).\end{aligned} \tag{3.9}$$

It follows from (3.4) that

$$A\tilde{x}^k+B\tilde{y}^k-b+(By^k-B\tilde{y}^k)-\frac{1}{\beta}(\lambda^k-\tilde{\lambda}^k)=0. \tag{3.10}$$

Combining (3.8), (3.9), and (3.10) and using the definition of the matrix Q, we can obtain the assertion of this lemma immediately. □

With Lemma 3.1 and Lemma 3.2 at hand, we can find the result below.

Theorem 3.1. *Let w^k be generated by Algorithm 3.1. Then, for any $w^*=(x^*,y^*,\lambda^*)\in\mathcal{W}^*$, $\Gamma_H(w^*,w^k)-\Gamma_H(w^*,w^{k+1})\ge\Gamma_G(\tilde{w}^k,w^k)\ge\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2$. where $\Gamma_J(w,w')=B_\varphi(u,u'^k)+\frac{1}{2}\|v-v'\|_J^2$, $J\in\{H,G\}$.*

Proof. Setting $w=w^*$ in (3.7), it follows from Lemma 3.1 and Proposition 2.3 that

$$\begin{aligned}&\theta(u^*)-\theta(\tilde{u}^k)+(w^*-\tilde{w}^k)^\top F(\tilde{w}^k)\\\ge\;&\frac{1}{2}(\|v^*-v^{k+1}\|_H^2-\|v^*-v^k\|_H^2)+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2+(u^*-\tilde{u}^k)^\top(\nabla\varphi(u^k)-\nabla\varphi(\tilde{u}^k))\\\ge\;&\frac{1}{2}(\|v^*-v^{k+1}\|_H^2-\|v^*-v^k\|_H^2)+\frac{1}{2}\|v^k-\tilde{v}^k\|_G^2+B_\varphi(u^*,\tilde{u}^k)-B_\varphi(u^*,u^k)+B_\varphi(\tilde{u}^k,u^k)\\=\;&\Gamma_H(w^*,w^{k+1})-\Gamma_H(w^*,w^k)+\Gamma_G(\tilde{w}^k,w^k).\end{aligned} \tag{3.11}$$

Setting $w=\tilde{w}^k$ in (2.2), we obtain

$$\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^\top F(w^*)\ge 0. \tag{3.12}$$

Combining (3.11) and (3.12), we have

$$(w^*-\tilde{w}^k)^\top(F(\tilde{w}^k)-F(w^*)) \geq \Gamma_H(w^*,w^{k+1})-\Gamma_H(w^*,w^k)+\Gamma_G(\tilde{w}^k,w^k).$$

In view of the monotonicity of F, we arrive at

$$\Gamma_H(w^*,w^k)-\Gamma_H(w^*,w^{k+1}) \geq \Gamma_G(\tilde{w}^k,w^k) \geq \frac{1}{2}\|v^k-\tilde{v}^k\|_G^2.$$

The assertion of this lemma is proved. □

Remark 3.2. It follows from (3.1b) and (3.1d) that

$$\lambda^k = \lambda^{k+1}+(\alpha+\gamma)\beta(Ax^{k+1}+By^{k+1}-b)+\alpha\beta(By^k-By^{k+1}).$$

In view of (3.10), we have $Ax^{k+1}+By^{k+1}-b+(By^k-By^{k+1})-\frac{1}{\beta}(\lambda^k-\tilde{\lambda}^k)=0$. If

$$\begin{cases} x^k-x^{k+1}=0, \\ By^k-By^{k+1}=0, \\ Ax^{k+1}+By^{k+1}-b=0, \end{cases}$$

then, it follows from (3.7) and Assumption A that

$$\theta(u)-\theta(u^{k+1})+(w-w^{k+1})^\top F(w^{k+1}) \geq 0, \forall w \in \mathscr{W},$$

which indicates that $(x^{k+1},y^{k+1},\lambda^{k+1})$ is a solution point of (2.2). Hence, the stopping criterion adopted here is reasonable: if it is satisfied with a small ε, we can regard the current iterate as an approximate solution.

4. Convergence of The Proposed Method

In this section, we prove the global convergence for the proposed method. Before proceeding, we need the following lemma.

Lemma 4.1. *Let w^k be generated by Algorithm 3.1 and $\tilde{w}^k$ be defined by (3.4). Then there exist constants $C_{11}>0$ and $C_{12}>0$ such that*

$$\|v^k-\tilde{v}^k\|_G^2 \geq C_{11}\|By^k-By^{k+1}\|^2+C_{12}\|Ax^{k+1}+By^{k+1}-b\|^2.$$

Proof. By using the definition of the matrix G, we obtain

$$\begin{aligned}\|v^k-\tilde{v}^k\|_G^2 &= (1-\gamma)\beta\|By^k-By^{k+1}\|^2-2(1-\gamma)(\lambda^k-\tilde{\lambda}^k)^\top(By^k-By^{k+1}) \\ &\quad +\tfrac{(2-\alpha-\gamma)}{\beta}\|\lambda^k-\tilde{\lambda}^k\|^2.\end{aligned}$$

Since $\lambda^k-\tilde{\lambda}^k=\beta(Ax^{k+1}+By^{k+1}-b)+\beta(By^k-By^{k+1})$, we have

$$\begin{aligned}\|v^k-\tilde{v}^k\|_G^2 &= (1-\alpha)\beta\|By^k-By^{k+1}\|^2+(2-\alpha-\gamma)\beta\|Ax^{k+1}+By^{k+1}-b\|^2 \\ &\quad +2(1-\alpha)\beta(Ax^{k+1}+By^{k+1}-b)^\top(By^k-By^{k+1})\end{aligned}$$

and it can be written as

$$\begin{aligned}
&\|v^k-\tilde{v}^k\|_G^2\\
= & \begin{pmatrix} By^k-By^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{pmatrix}^\top \begin{pmatrix} (1-\alpha)\beta I_l & (1-\alpha)\beta I_l \\ (1-\alpha)\beta I_l & (2-\alpha-\gamma)\beta I_l \end{pmatrix} \begin{pmatrix} By^k-By^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{pmatrix}\\
= & \left\| \begin{matrix} By^k-By^{k+1} \\ Ax^{k+1}+By^{k+1}-b \end{matrix} \right\|_S^2,
\end{aligned}$$

where

$$S=\begin{pmatrix} (1-\alpha)\beta I_l & (1-\alpha)\beta I_l \\ (1-\alpha)\beta I_l & (2-\alpha-\gamma)\beta I_l \end{pmatrix}.$$

Since $\alpha\in(-1,1),\gamma\in(0,1)$, and $\alpha+\gamma>0$, then S is positive definite. Therefore, we obtain the assertion of this Lemma. □

Combining Theorem 3.1 and Lemma 4.1, we can obtain the result below.

Theorem 4.1. *Let w^k be generated by Algorithm 3.1, and let $\tilde{w}^k$ be defined by (3.4). Then, there exist constants $C_1>0$ and $C_2>0$ such that*

$$\Gamma_H(w^*,w^{k+1}) \;\le\; \Gamma_H(w^*,w^k)-\left(C_1\|By^k-By^{k+1}\|^2+C_2\|Ax^{k+1}+By^{k+1}-b\|^2\right). \tag{4.1}$$

Now we are ready to prove the global convergence of the proposed method.

Theorem 4.2. *Let w^k be generated by Algorithm 3.1. Then,*

$$\lim_{k\to\infty}(\|By^k-By^{k+1}\|^2+\|Ax^{k+1}+By^{k+1}-b\|)=0. \tag{4.2}$$

Moreover, $\{v^k\}$ converges to a point $v^\infty\in\mathscr{V}^*$.

Proof. It follows from (4.1) that

$$\sum_{k=0}^{\infty}\left(C_1\|By^k-By^{k+1}\|^2+C_2\|Ax^{k+1}+By^{k+1}-b\|^2\right)\le\Gamma_H(w^*,w^0),$$

and then

$$\lim_{k\to\infty}\left(\|By^k-By^{k+1}\|^2+\|Ax^{k+1}+By^{k+1}-b\|^2\right)=0.$$

Moreover, it is easy to know from (4.1) that $\{\Gamma_H(w^*,w^k)\}$ is monotonically non-increasing, and hence is bounded. Thus it is convergent. Since $\{\|\Gamma_H(w^*,w^k)\|\}$ is bounded, then $\{\|v^*-v^k\|_H\}$ is bounded. From Lemma 2.1 and the fact that H is positive definite, the sequence $\{v^k\}$ is also bounded, and it has a subsequence $\{v^{k_j}\}$ converging to a cluster point, say v^∞. Due to (4.2) and the definition of $\tilde{w}^k$ in (3.4), we have $\lim_{k\to\infty}\|v^k-\tilde{v}^k\|=0$. Let $\tilde{x}^\infty$ be induced by (3.1a) with given $(y^\infty,\lambda^\infty)$. From (4.2), we have $B(y^\infty-\tilde{y}^\infty)=0$ and $\lambda^\infty-\tilde{\lambda}^\infty=0$. Since B is full column rank, we can obtain $y^\infty=\tilde{y}^\infty$. Hence, taking limits in (3.7) along the subsequence $\{w^{k_j}\}$, it follows that

$$\theta(u)-\theta(\tilde{u}^\infty)+(w-\tilde{w}^\infty)^\top F(\tilde{w}^\infty)\ge 0,\quad \forall\, w\in\mathscr{W}. \tag{4.3}$$

Therefore $\tilde{w}^\infty=w^\infty$ is a solution of (2.2). Since $\{\Gamma_H(w^\infty,w^k)\}$ is convergent and $\frac{1}{2}\|v^k-v^\infty\|_H^2\le\Gamma_H(w^\infty,w^k)$, then the sequence $\{\|v^k-v^\infty\|_H\}$ is convergent. Since $v^{k_j}\to v^\infty$, it follows that

$v^k \to v^\infty$. In view of $w^\infty \in \mathscr{W}^*$ and (4.1), we have $\Gamma_H(w^\infty, w^{k+1}) \le \Gamma_H(w^\infty, w^k)$. Then, for any other stationary point w_1^∞,

$$\begin{aligned}\Gamma_H(w^\infty, w_1^\infty) &\le \liminf_k \Gamma_H(w^\infty, w^k) \le \lim_j \Gamma_H(w^\infty, w^{k_j}) \\ &= \Gamma_H(w^\infty, w^\infty) = 0,\end{aligned}$$

which indicates that $w^\infty = w_1^\infty$ from $\frac{1}{2}\|w^\infty - w_1^\infty\|^2 \le \Gamma_H(w^\infty, w_1^\infty)$. That is also to say that the sequence $\{v^k\}$ cannot have another cluster point and thus it converges to a solution point $v^\infty \in \mathscr{V}^*$. The proof is completed. □

Combining (3.11) and Lemma 4.1, and using the monotonicity of F, we obtain the following results.

Theorem 4.3. *Let w^k be generated by Algorithm 3.1, and let $\tilde{w}^k$ be defined by (3.4). Thus*

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(w) \ge \Gamma_H(w, w^{k+1}) - \Gamma_H(w, w^k), \forall w \in \mathscr{W}. \tag{4.4}$$

5. $O(1/t)$ Convergence Rate

In this section, we show that the proposed method has the $O(1/t)$ convergence rate. Recall that $\mathscr{W}^*$ can be characterized as (see (2.3.2) on pp. 159 of [33])

$$\mathscr{W}^* = \bigcap_{w \in \mathscr{W}} \{\tilde{w} \in \mathscr{W} : (\Theta(u) - \Theta(\tilde{u})) + (w - \tilde{w})^\top F(w) \ge 0\}.$$

This implies that $\tilde{w}$ is an approximate solution of (2.2) with the accuracy $\varepsilon > 0$ if it satisfies

$$\tilde{w} \in \mathscr{W} \quad \text{and} \quad \sup_{w \in \mathscr{W}} \{(\Theta(\tilde{u}) - \Theta(u)) + (\tilde{w} - w)^\top F(w)\} \le \varepsilon. \tag{5.1}$$

In the rest, our purpose is to show that after t iterations of the proposed method, and we can find a $\tilde{w} \in \mathscr{W}$ such that (5.1) is satisfied with $\varepsilon = O(1/t)$.

Theorem 5.1. *Let w^k be generated by Algorithm 3.1, and let $\tilde{w}^k$ be defined by (3.4). Then*

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^\top F(w) \le \frac{1}{t+1} \Gamma_H(w, w^0), \forall w \in \mathscr{W},$$

where

$$\tilde{w}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{w}^k \quad \text{and} \quad \tilde{u}_t = \frac{1}{t+1} \sum_{k=0}^{t} \tilde{u}^k.$$

Proof. From (4.4), we have

$$\theta(\tilde{u}^k) - \theta(u) + (\tilde{w}^k - w)^\top F(w) \le \Gamma_H(w, w^k) - \Gamma_H(w, w^{k+1}), \forall w \in \mathscr{W}. \tag{5.2}$$

Summing the inequality (5.2) over $k = 0, \cdots, t$, we obtain

$$\sum_{k=0}^{t} \theta(\tilde{u}^k) - (t+1)\theta(u) + \left(\sum_{k=0}^{t} \tilde{w}^k - (t+1)w\right)^\top F(w) \le \Gamma_H(w, w^0), \forall w \in \mathscr{W}. \tag{5.3}$$

The convexity of θ indicates

$$\theta(\tilde{u}_t) = \theta\left(\frac{1}{t+1} \sum_{k=0}^{t} \tilde{u}^k\right) \le \frac{1}{t+1} \sum_{k=0}^{t} \theta(\tilde{u}^k). \tag{5.4}$$

Substituting (5.4) into (5.3) and using the definition of $\tilde{w}_t$, we have

$$\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^\top F(w) \le \frac{1}{t+1}\Gamma_H(w, w^0), \forall w \in \mathscr{W},$$

which proves the desired result. Indeed, $\tilde{w}_t \in \mathscr{W}$ because it is a convex combination of $\tilde{w}^0, \tilde{w}^1, \cdots, \tilde{w}^t$. The proof is completed. □

For any compact set $\mathscr{D} \subset \mathscr{W}$, let $d = \sup\{\Gamma_H(w, w^0) | w \in \mathscr{D}\}$. For any given $\varepsilon > 0$, after most

$$t = (\frac{d}{\varepsilon} - 1)$$

iterations, we have $\theta(\tilde{u}_t) - \theta(u) + (\tilde{w}_t - w)^\top F(w) \le \varepsilon, \forall w \in \mathscr{D}$. That is, the $O(1/t)$ convergence rate is established in an ergodic sense.

6. Preliminary Computational Results

In order to verify the theoretical assertions, we consider the LASSO model in [5, 34]

$$\min\left\{\frac{1}{2}\|x-b\|^2 + \sigma\|y\|_1 | x - Ay = 0, x \in \mathbb{R}^{n_1}, y \in \mathbb{R}^{n_2}\right\}, \tag{6.1}$$

where $\|x\|_1 = \Sigma_i |x_i|$. Applying the proposed method to the problem (6.1) with $B_\phi(x,y) = \frac{1}{2}\|x - y\|_C^2$ and $B_\psi(x,y) = \frac{1}{2}\|x-y\|_D^2$, where $C = r_1 I_{n_1} - \beta A^\top A$, $D = \tau r_2 I_{n_2} - \beta B^\top B$ with $r_1 \ge \beta\|A^\top A\|$, $r_2 \ge \beta\|B^\top B\|, \tau \in (\frac{1+\alpha}{2}, 1)$, and B is full column rank. $r_1 \ge \beta\|A^\top A\|$ implies the positive semi-definiteness of matrix C, and $\tau < 1$ indicates the indefiniteness of matrix D. Note that similar matrices was tested in [7]. The iterative scheme is given as

$$x^{k+1} = \frac{1}{r_1 + \beta}(b + \lambda^k + \beta A y^k + (r_1 - 1)x^k)$$

$$\lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - Ay^k),$$

$$y^{k+1} = \text{shrink}\left(y^k - \frac{A^\top(\lambda^{k+\frac{1}{2}} - \beta(x^{k+1} - Ay^k))}{\tau r_2}, \frac{\sigma}{\tau r_2}\right)$$

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \gamma\beta(x^{k+1} - Ay^{k+1}),$$

where the shrink is the soft thresholding operator (see [35]) defined as

$$(S_\delta(t))_i = (1 - \delta/|t_i|)_+ \cdot t_i, i = 1, 2, \cdots, n_2.$$

We use the same notations as in [7]. We compare the performance of the proposed method (BPRSM) with CADMM in [5], IDSADMM in [36], IPGADMM in [37], and GLADMM) in [38]. In all tests, we take $r_1 = 1.001, r_2 = \beta\|A^\top A\| + 0.001$ and $(y^0, \lambda^0) = (0, 0)$ as the initial point. We compare the results of each algorithm when the parameters are optimal. The parameters in each algorithm are listed as below.
CADMM: $\beta = 1, \sigma = 0.1\|A^\top b\|_\infty, \tau = 1.001, \alpha = 0.3, \gamma = 1.618.$
IDSADMM: $\beta = 1, \sigma = 0.1\|A^\top b\|_\infty, \alpha = 0.3, \tau = \frac{\alpha^2 - \alpha + 4}{\alpha^2 - 2\alpha + 5} + 0.001, \gamma = 1.$
IPGADMM: $\beta = 1, \sigma = 0.1\|A^\top b\|_\infty, \alpha = 0.3, \tau = \frac{3+\alpha}{4} + 0.001, \gamma = 1.$
GLADMM: $\beta = 1, \sigma = 0.1\|A^\top b\|_\infty, \alpha = 0.3, \tau = \frac{4(\alpha+\gamma)^2 - 5(\alpha+\gamma) + 10}{4(\alpha+\gamma)^2 - 8(\alpha+\gamma) + 16} + 0.001, \gamma = 0.4.$

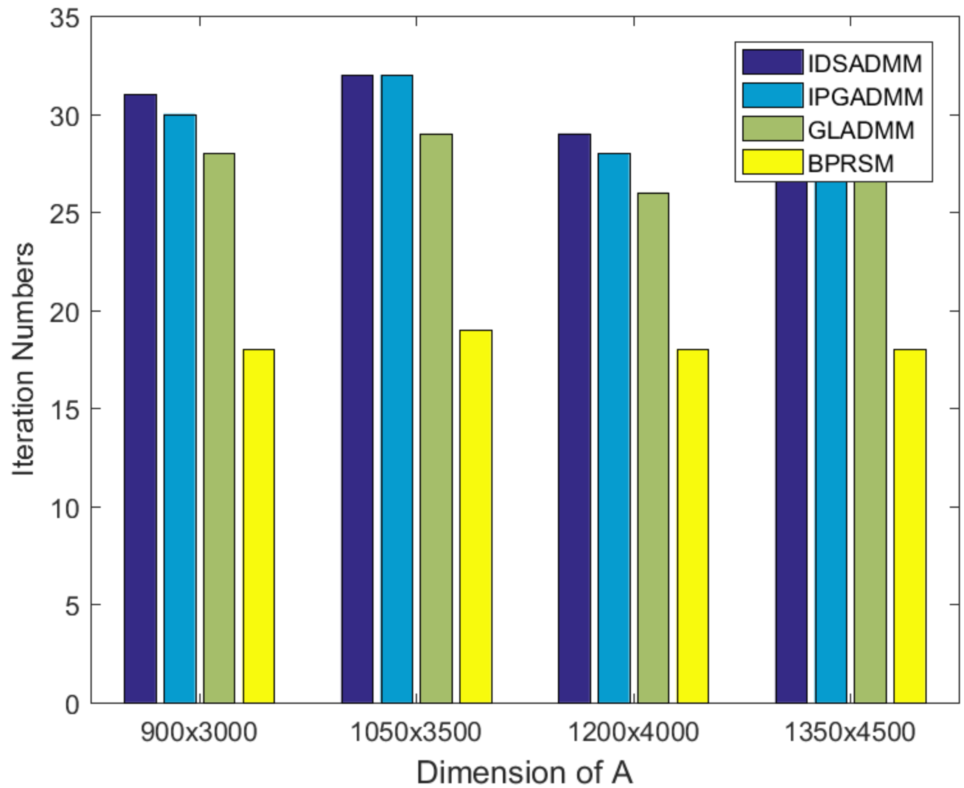

FIGURE 6.1. Comparison results for problem (6.1)

BPRSM: $\beta = 1, \sigma = 0.1\|A^\top b\|_\infty, \alpha = -0.4, \tau = \frac{1+\alpha}{2} + 0.001, \gamma = 0.9$.
The iteration is stopped as soon as

$$\|x^{k+1} - Ay^{k+1}\| \le \varepsilon^{\text{pri}} \quad \text{and} \quad \|A(y^{k+1} - y^k)\| \le \varepsilon^{\text{dual}}$$

where $\varepsilon^{\text{pri}} = \sqrt{n}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}} \max\{\|x^{k+1}\|, \|Ay^{k+1}\|\}$ and $\varepsilon^{\text{dual}} = \sqrt{n}\varepsilon^{\text{abs}} + \varepsilon^{\text{rel}}\|y^{k+1}\|$, with ε^{abs} and ε^{rel} set to be 10^{-4} and 10^{-2}. For a given dimension $m \times n$, we generate the data randomly as follows. $p = 1/n, x^0 = \text{sprandn}(n,1,p), A = \text{randn}(m,n), b = A * x^0 + sqrt(0.0001) * \text{randn}(m,1)$. All codes were written in Matlab; we compare the proposed method with those in [5, 36, 37, 38]. The iteration numbers, denoted by k, and the computational time for problem (6.1) with different dimensions are given in Table 6.1.

Table 6.1: Numerical results for problem (6.1)

		CADMM		IDSADMM		IPGADMM		GLADMM		BPRSM	
m	n	k	CPU(s)	k	CPU(s)	k	CPU(s)	k	CPU(s)	k	CPU(s)
900	300	34	1.67	31	1.58	30	1.53	28	1.67	18	0.92
1050	3500	36	2.49	32	2.40	32	2.44	29	2.38	19	1.38
1200	4000	32	2.80	29	2.57	28	2.48	26	2.88	18	1.76
1350	4500	35	3.92	31	3.88	31	3.61	28	3.92	18	2.16
1500	5000	29	3.97	25	3.57	25	3.58	24	3.96	17	2.40

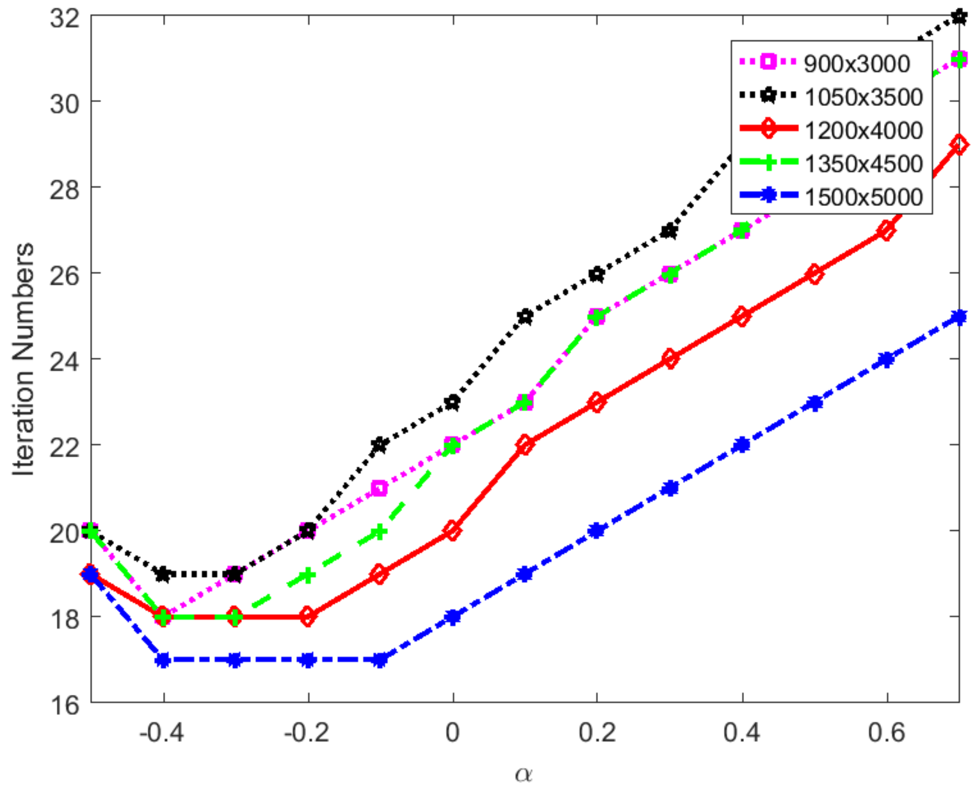

FIGURE 6.2. Sensitivity tests on different α of BPRSM for problem (6.1)

Figure 6.1 reports the comparison between the methods in [5, 36, 37, 38] and the proposed method for solving problem (6.1). Tables 6.1 and Figure 6.1 show the high efficiency and robustness of the proposed method. We observe that the proposed method is faster and need fewer iteration.

Figure 6.2 shows the sensitivity tests of the proposed method for solving different dimensional with different α values. When α takes the value between intervals [-0.5, 0.1], the number of iterations under different dimensions is not more than 24, of which the minimum value is 17.

7. CONCLUSIONS

In this paper, by using the Bregman distance and the the proximal Peaceman-Rachford splitting method, we proposed a Bregman proximal Peaceman-Rachford splitting method for solving a separable convex minimization model. The new algorithm only needs to solve two unconstrained subproblems at each iteration. We proved the global convergence of the algorithm under standard assumptions. Finally, we reported some numerical results, indicating the efficiency of the proposed algorithm.

REFERENCES

[1] A. Bnouhachem, A descent SQP alternating direction method for minimizing the sum of three convex functions, J. Nonlin. and Var. Anal. 4 (2020), 469-482.

[2] A. Bnouhachem, M.T. Rassias, On descent alternating direction method with LQP regularization for the structured variational inequalities, Optim. Lett. 14 (2020), 1353-1369.

[3] Q. Bnouhachem, A self-adaptive descent LQP alternating direction method for the structured variational inequalities, Numer. Algo. 86 (2021), 303-324.

[4] A. Bnouhachem, X. Qin, An inertial proximal Peaceman-Rachford splitting method with SQP Regularization for Convex Program, J. Nonlin. Func. Anal. 2020, 1-17 (2020).

[5] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (2010), 1-122.

[6] W. Deng, M. Lai, P. Peng, W. Yin, Parallel multi-block ADMM with $o(1/k)$ convergence. J. Sci. Comput. 71 (2017), 712-736.

[7] Z. Deng, S. Liu, Inertial proximal strictly contractive Peaceman-Rachford splitting method with an indefinite term for convex optimization, J. Comp. Appl. Math. 374 (2020), Article ID 112772.

[8] D. Gabay, B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, Comput. Math. Appl. 2 (1976), 17-40.

[9] R. Glowinski, A. Marrocco, Sur l'approximation par éléments finis d'ordre un et la résolution par pénalisation-dualité d'une classe de problémes de Dirichlet non linéaires. Revue Fr. Autom. Inform. Rech. Opér. Anal. Numér. 2 (1975), 41-76.

[10] Y. Cui, X. Li, D. Sun, K.C. Toh, On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions, J. Optim. Theory Appl. 169 (2016), 1013-1041.

[11] W. Deng, W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, J. Sci. Comput. 66 (2016), 889-916.

[12] B. He, L. Liao, D. Han, H. Yang, A new inexact alternating directions method for monotone variational inequalities, Math. Program. 92 (2002), 103-118.

[13] B. He, X. Yuan, On the $o(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, SIAM J. Numer. Anal. 50 (2012), 700-709.

[14] B. He, X. Yuan, On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers, Numer. Math. 130 (2015), 567-577.

[15] M. Xu, Proximal alternating directions method for structured variational inequalities. J. Optim. Theory Appl. 134 (2007), 107-117.

[16] J. Eckstein, P.J.S. Silva, A practical relative error criterion for augmented Lagrangians, Math. Program. 141 (2013), 319-348.

[17] J. Eckstein, W. Yao, Approximate ADMM algorithms derived from Lagrangian splitting, Comput. Optim. Appl. 68 (2017, 363-405).

[18] J. Eckstein W. Yao, Relative-error approximate versions of Douglas-Rachford splitting and special cases of the ADMM, Math. Program. 170 (2018), 417-444.

[19] J. Xie, A. Liao, X. Yang, An inexact alternating direction method of multipliers with relative error criteria, Optim. Lett. 11 (2017), 583-596.

[20] J. Douglas, H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, Trans. Am. Math. Soc. 82 (1956), 421-439.

[21] J. Eckstein, D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, Math. Program. 55 (1992), 293-318.

[22] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979), 964-979.

[23] D.W. Peaceman, H.H. Rachford, The numerical solution of parabolic and elliptic differential equations, J. Soc. Ind. Appl. Math. 3 (1955), 28-41.

[24] B. He, H. Liu, Z. Wang, X. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, SIAM J. Optim. 24 (2014), 1011-1040.

[25] Y, Gu, B. Jiang, D.R. Han, A semi-proximal-based strictly contractive Peaceman-Rachford splitting method, ArXiv: 1506.02221, 2015.

[26] B. He, F. Ma, X. Yuan, Convergence study on the symmetric version of ADMM with larger step sizes, SIAM J. Imaging Sci. 9 (2016), 1467-1501.

[27] M. Fortin, R. Glowinski, eds., Méthodes de Lagrangien Augmenté: Application à la résolution numérique de problèmes aux limites, Dunod, Paris, 1982.
[28] M. Fortin, R. Glowinski, eds., Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, North-Holland, Amsterdam, The Netherlands, 1983.
[29] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, Springer-Verlag, New York, 1984.
[30] K. Tu, H. Zhang, H. Gao, J. Feng, A hybrid Bregman alternating direction method of multipliers for the linearly constrained difference-of-convex problems, J. Global Optim. 76 (2020), 665-693.
[31] F. Wang, Z. Xu, H.K. Xu, Convergence of Bregman alternating direction method with multipliers for non-convex composite problems (2014). arXiv preprint arXiv:1410.8625.
[32] I. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, USSR Comput. Math. Math. Phys. 7 (1967), 200-217.
[33] F. Facchinei, J.S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I and II. Springer Series in Operations Research. Springer, New York, 2003.
[34] R. Tibshirani, Regression shrinkage and selection via the LASSO: a retrospective, J. R. Stat. Soc. Ser. B-Stat. Methodol. 73 (2011), 273-282.
[35] D.L. Donoho, Y. Tsaig, Fast solution of l_1-norm minimization problems when the solution may be sparse, IEEE Trans. Info. Theory 54 (2008), 4789-4812.
[36] B. Gao, F. Ma, Symmetric alternating direction method with indefinite proximal regularization for linearly constrained convex optimization, J. Optim. Theory Appl. 176 (2018), 178-204.
[37] F. Jiang, Z. Wu, X. Cai, Generalized admm with optimal indefinite proximal term for linearly constrained convex optimization, J. Indust. Manag. Optim. 13 (2017), 1-22.
[38] X. Chang, S. Liu P. Zhao, D. Song, A generalization of linearized alternating direction method of multipliers for solving two-block separable convex programming, J. Comput. Appl. Math. 357 (2019), 251-272.