Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(6): 10513–10533.
DOI: 10.3934/math.2022586
Received: 18 December 2021
Revised: 11 March 2022
Accepted: 15 March 2022
Published: 28 March 2022

Research article

A linearly convergent proximal ADMM with new iterative format for BPDN in compressed sensing problem

Bing Xue[1], **Jiakang Du**[2], **Hongchun Sun**[1,*] **and Yiju Wang**[2,*]

[1] School of Mathematics and Statistics, Linyi University, Linyi 276005, China
[2] School of Management Science, Qufu Normal University, Rizhao 276800, China

* **Correspondence:** Email: hcsun68@126.com, wang-yiju@163.com.

Abstract: In recent years, compressive sensing (CS) problem is being popularly applied in the fields of signal processing and statistical inference. The alternating direction method of multipliers (ADMM) is applicable to the equivalent forms of basis pursuit denoising (BPDN) in CS problem. However, the solving speed and accuracy are adversely affected when the dimension increases greatly. In this paper, a new iterative format of proximal ADMM, which has fast solving speed and pinpoint accuracy when the dimension increases , is proposed to solve BPDN problem. Global convergence of the new type proximal ADMM is established in detail, and we exhibit a $R-$ linear convergence rate under suitable condition. Moreover, we apply this new algorithm to solve different types of BPDN problems. Compared with the state-of-the-art of algorithms in BPDN problem, the proposed algorithm is more accurate and efficient.

Keywords: new type proximal ADMM; global convergence; $R-$ linear convergence rate; BPDN problem
Mathematics Subject Classification: 90C30, 90C33

1. Introduction

Compressive sensing (CS) problem is to recover a sparse signal $\bar{x}$ from an undetermined linear system $A\bar{x} = b$, where $\bar{x} \in \mathcal{R}^n$, A is the sensing matrix and $A \in \mathcal{R}^{m\times n}$ $(m \ll n)$, b is the observed signal and $b \in \mathcal{R}^m$. A fundamental decoding model of CS problem is the following basis pursuit denoising (termed as BPDN) problem:

$$\min_{x\in R^n} \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1 \tag{1.1}$$

where μ (> 0) is a parameter, and the norm $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the Euclidean 1-norm and 2-norm, respectively.

Recently, a lot of numerical algorithms about BPDN problem have been extensively developed. In fact, the BPDN problem can be equivalently converted into a separable convex programming by introducing auxiliary variable. Thus the numerical methods, which can be used to solve the separable convex programming, are applicable to BPDN problem, such as the alternating direction method of multipliers and its linearized version [1–4], the Peaceman-Rachford splitting method (PRSM) or Douglas-Rachford splitting method (DRSM) of multipliers [5–8], the symmetric alternating direction method of multipliers [9], etc. Yang and Zhang [1] investigate the use of alternating direction algorithms for several ℓ_1-norm minimization problems arising from sparse solution recovery in CS, including the basis pursuit problem, the basis-pursuit denoising problems, and so on. Yuan [2] presents a descent-like method, which can obtain a descent direction and an appropriate step size and improve the proximal alternating direction method. Yu et al. [5] apply the primal DRSM to solve equivalent transformation form of BPDN problem. He and Dan [6] furtherly study the multi-block separable convex minimization problem with linear constraints along the way by the primal application of DRSM, and present the exact and inexact versions of the new method in a unified framework. Compared to the DRSM, the PRSM requires more restrictive assumptions to ensure its convergence, while it is always faster whenever it is convergent. He and Liu et al. [7] illustrate the reason for this difference, and develop a modified PRSM for separable convex programming, which includes BPDN problem as a special case. Sun and Liu [8] develop a generalized PRSM for BPDN problem, of which both subproblems are approximated by the linearization technique. He et al. [9] obtain the convergence of the symmetric version of ADMM with step sizes, where step sizes can be enlarged by Fortin and Glowinski's constant. On the other hand, BPDN problem can be equivalently transformed into an equation or variational inequality problem by splitting technique [10–15], which can be solved by some standard methods such as conjugate gradient methods, proximal point algorithms and projection-type algorithms. Xiao and Zhu [10] transform BPDN problem into a convex constrained monotone equation, and present a conjugate gradient method for the equivalent form of the problem. Sun and Tian [11] propose a class of derivative-free conjugate gradient (CG) projection methods for nonsmooth equations with convex constraints, including the BPDN problem. Sun et al. [12] reformulate BPDN problem as variational inequality problem by splitting the decision variable into two nonnegative auxiliary variables, and propose a novel inverse matrix-free proximal point algorithm for BPDN problem. Base on the same transformation of BPDN problem, Feng and Wang [13] also propose a projection-type algorithm without any backtracking line search. Although there are so many ways to solve the problem, it is still needed to improve the solving speed and accuracy. In particular, as the dimension increases greatly, the solving speed and accuracy are adversely affected. In the paper, we will establish a new iterative format of proximal ADMM, which has closed-form solutions. The motivation behind this is for the better numerical performance when the dimension increases. Furthermore, the linear rate convergence result for new algorithm is established, which is also one of the most important motivations.

The rest of this paper is organized as follows. In Section 2, some equivalent reformulations of (1.1) and related theories of (1.1), which are the basis of our analysis, are given. In Section 3, basing on a special iterative format, we present a new type proximal ADMM, in which the corresponding subproblems have closed-form solutions. The global convergence of new method is discussed in detail. We establish a $R-$linear rate convergence theorem under suitable condition. In Section 4, some numerical experiments on sparse signal recovery are given, and we compare the CPU time and the relative error among the Peaceman-Rachford splitting method of multipliers [8], the conjugate gradient methods [11], the proximal point algorithms [12], the projection-type algorithms [13] and our algorithm, and show that our algorithm is more accurate and efficient than other algorithms. Finally, some conclusive remarks are drawn in the last section.

We end this section with some notations used in this paper. Vectors considered in this paper are all taken in Euclidean space equipped with the standard inner product. The notation $\mathcal{R}^{m\times n}$ stands for the set of all $m \times n$ real matrices. For $x, y \in \mathcal{R}^n$, we use $(x; y)$ to denote the column vector $(x^\top, y^\top)^\top$. If $G \in \mathcal{R}^{n\times n}$ is a symmetric positive definite matrix, we denote by $\|x\|_G = \sqrt{x^\top G x}$ the G-norm of the vector x.

2. Equivalent reformulations of BPDN problem and preliminaries

2.1. Equivalent reformulations of BPDN problem

In this section, we first establish some equivalent reformulations to the BPDN problem via some the related optimality theories.

It is obvious that the BPDN problem can be equivalently reformulated as the following optimization problem [8]

$$\begin{array}{ll} \min & \frac{1}{2}\|Ax_1 - b\|_2^2 + \mu\|x_2\|_1 \\ \text{s.t.} & x_1 - x_2 = 0 \\ & x_1 \in \mathcal{R}^n, \quad x_2 \in \mathcal{R}^n. \end{array} \tag{2.1}$$

Then the augmented Lagrangian function of the convex programming (2.1) can be written as

$$\mathcal{L}_\beta(x_1, x_2, \lambda) := \theta_1(x_1) + \theta_2(x_2) - \left\langle \lambda, x_1 - x_2 \right\rangle + \frac{\beta}{2}\left\|x_1 - x_2\right\|^2, \tag{2.2}$$

where λ is the Lagrangian multiplier for the linear constraints of (2.1) and $\lambda \in \mathcal{R}^n$,

$$\theta_1(x_1) = \frac{1}{2}\|Ax_1 - b\|_2^2, \quad \theta_2(x_2) = \mu\|x_2\|_1.$$

By invoking the first-order optimality condition for convex programming, we can equivalently reformulate problem (2.1) as the following variational inequality problem: finding vector $x^* = (x_1^*; x_2^*) \in \mathcal{R}^n \times \mathcal{R}^n$ and $\lambda^* \in \mathcal{R}^n$ such that

$$\begin{cases} \theta_1(x_1) - \theta_1(x_1^*) - (x_1 - x_1^*)^\top \lambda^* \ge 0, & \forall x_1 \in \mathcal{R}^n, \\ \theta_2(x_2) - \theta_2(x_2^*) + (x_2 - x_2^*)^\top \lambda^* \ge 0, & \forall x_2 \in \mathcal{R}^n, \\ x_1^* - x_2^* = 0. \end{cases} \tag{2.3}$$

Obviously, the system (2.3) is equivalent to the following problem: Find a vector $w^* \in \mathcal{W}$ such that

$$\theta(x) - \theta(x^*) + (w - w^*)^\top F(w^*) \ge 0, \quad \forall w \in \mathcal{W}, \tag{2.4}$$

where $w = (x_1; x_2; \lambda) \in \mathcal{W} = \mathcal{R}^n \times \mathcal{R}^n \times \mathcal{R}^n$, $\theta(x) = \theta_1(x_1) + \theta_2(x_2)$, and

$$F(w) := \begin{pmatrix} -\lambda \\ \lambda \\ x_1 - x_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -I_n \\ 0 & 0 & I_n \\ I_n & -I_n & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix}. \tag{2.5}$$

We denote the solution set of (2.4) by $\mathcal{W}^*$.

It is easy to verify that the mapping $F(\cdot)$ is not only monotone but also satisfies the following nice property

$$(w' - w)^\top (F(w') - F(w)) = 0, \ \ \forall w', w \in \mathcal{W}.$$

2.2. Preliminaries

To proceed, we present the following definition, which will be needed in the sequel.

Definition 2.1. *For sequence vector $\mu^k = (\mu_1^k, \mu_2^k, \cdots, \mu_n^k)^\top \in \mathcal{R}^n$ $(k = 1, 2, \cdots)$, we define two new functions $\psi(\mu^k)$ and $\delta(\mu^k)$.*

(i) The function $\psi(\mu^k) = (\psi_1(\mu_1^k), \psi_1(\mu_2^k), \cdots, \psi_1(\mu_n^k))^\top$, and

$$\psi_1(\mu_i^k) = \begin{cases} 0, & \text{if } \ |\mu_i^k| \le \frac{C}{n2^k}, \\ \mu_i^k, & \text{if } \ |\mu_i^k| > \frac{C}{n2^k}. \end{cases} \qquad (i = 1, 2, \cdots, n) \tag{2.6}$$

where k is positive integer, and $C > 0$ is a constant.

(ii) The function $\delta(\mu^k) = \mu^k - \psi(\mu^k)$.

3. Algorithm and convergence

In this section, we present a new type proximal ADMM for solving (2.1) by a special iterative format, and the global convergence of new method is also established in detail.

Algorithm 3.1.

Step 0. Select constants $\beta, \gamma, \varepsilon > 0$, two positive semi-definite matrices $R_i \in \mathcal{R}^{n\times n} (i = 1, 2)$. Choose an arbitrarily initial point $w^0 = (x_1^0; x_2^0; \lambda^0) \in \mathcal{R}^n \times \mathcal{R}^n \times \mathcal{R}^n$. Take

$$\eta = \begin{cases} \gamma, & \text{if } \ 0 < \gamma \le 1, \\ \frac{1}{\gamma}, & \text{if } \ \gamma > 1. \end{cases} \tag{3.1}$$

Set $k = 0$.

Step 1. By current iterate w^k, compute the new iterate $\hat{w}^k = (\hat{x}_1^k; \hat{x}_2^k; \hat{\lambda}^k)$ via

$$\begin{cases} \hat{x}_1^k \in \text{argmin}_{x_1 \in \mathcal{R}^n} \mathcal{L}_\beta(x_1, x_2^k, \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{R_1}^2, \\ \hat{x}_2^k \in \text{argmin}_{x_2 \in \mathcal{R}^n} \mathcal{L}_\beta(\hat{x}_1^k, x_2, \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_{R_2}^2, \\ \hat{\lambda}^k = \lambda^k - \gamma\beta\left(\hat{x}_1^k - \hat{x}_2^k\right), \end{cases} \tag{3.2}$$

Step 2. If $\|w^k - \hat{w}^k\| \le \varepsilon$, then stop; otherwise, go to Step 3.

Step 3. Set $w^{k+1} = \rho\hat{w}^k + (1 - \rho)\psi(w^k)$, where $\rho \in (0, \eta)$ and η is a constant defined in (3.1). Go to Step 1.

In the following, we show that it is reasonable to use $\|w^k - \hat{w}^k\| \le \varepsilon$ to terminate Algorithm 3.1.

Lemma 3.1. *If $w^k = \hat{w}^k$, then the vector w^k is a solution of (2.4).*

Proof. For x_1-subproblem in (3.2), using the first-order optimality condition, for any $x_1 \in \mathcal{R}^n$, we obtain

$$\begin{aligned}
0 &\leq \theta_1(x_1) - \theta_1(\hat{x}_1^k) + (x_1 - \hat{x}_1^k)^\top \left\{-\lambda^k + \beta\left(\hat{x}_1^k - x_2^k\right) + R_1(\hat{x}_1^k - x_1^k)\right\} \\
&= \theta_1(x_1) - \theta_1(\hat{x}_1^k) + (x_1 - \hat{x}_1^k)^\top \left\{-\hat{\lambda}^k + \beta(1-\gamma)\left(\hat{x}_1^k - \hat{x}_2^k\right) - \beta(x_2^k - \hat{x}_2^k)\right. \\
&\quad \left. + R_1(\hat{x}_1^k - x_1^k)\right\} \\
&= \theta_1(x_1) - \theta_1(\hat{x}_1^k) + (x_1 - \hat{x}_1^k)^\top(-\hat{\lambda}^k) \\
&\quad + (x_1 - \hat{x}_1^k)^\top \left\{\tfrac{1-\gamma}{\gamma}(\lambda^k - \hat{\lambda}^k) - \beta(x_2^k - \hat{x}_2^k)\right\} + (x_1 - \hat{x}_1^k)^\top R_1(\hat{x}_1^k - x_1^k),
\end{aligned}$$

where the first and the second equalities are by $\lambda^k = \hat{\lambda}^k + \gamma\beta(\hat{x}_1^k - \hat{x}_2^k)$, i.e.,

$$\begin{aligned}
&\theta_1(x_1) - \theta_1(\hat{x}_1^k) + (x_1 - \hat{x}_1^k)^\top(-\hat{\lambda}^k) \\
&\geq (x_1 - \hat{x}_1^k)^\top R_1(x_1^k - \hat{x}_1^k) + \beta(x_1 - \hat{x}_1^k)^\top(x_2^k - \hat{x}_2^k) - \tfrac{1-\gamma}{\gamma}(x_1 - \hat{x}_1^k)^\top(\lambda^k - \hat{\lambda}^k).
\end{aligned} \tag{3.3}$$

For x_2-subproblem in (3.2), similar to discussion above, one has

$$\begin{aligned}
0 &\leq \theta_2(x_2) - \theta_2(\hat{x}_2^k) + (x_2 - \hat{x}_2^k)^\top \left\{\lambda^k - \beta\left(\hat{x}_1^k - \hat{x}_2^k\right) + R_2(\hat{x}_2^k - x_2^k)\right\} \\
&= \theta_2(x_2) - \theta_2(\hat{x}_2^k) + (x_2 - \hat{x}_2^k)^\top \left\{\hat{\lambda}^k - \beta(1-\gamma)\left(\hat{x}_1^k - \hat{x}_2^k\right) + R_2(\hat{x}_2^k - x_2^k)\right\} \\
&= \theta_2(x_2) - \theta_2(\hat{x}_2^k) + (x_2 - \hat{x}_2^k)^\top \hat{\lambda}^k \\
&\quad - (x_2 - \hat{x}_2^k)^\top \left\{\tfrac{1-\gamma}{\gamma}(\lambda^k - \hat{\lambda}^k)\right\} + (x_2 - \hat{x}_2^k)^\top R_2(\hat{x}_2^k - x_2^k),
\end{aligned}$$

where the first and second equalities are by $\lambda^k = \hat{\lambda}^k + \gamma\beta(\hat{x}_1^k - \hat{x}_2^k)$, i.e.,

$$\theta_2(x_2) - \theta_2(\hat{x}_2^k) + (x_2 - \hat{x}_2^k)^\top \hat{\lambda}^k \geq (x_2 - \hat{x}_2^k)^\top R_2(x_2^k - \hat{x}_2^k) + \frac{1-\gamma}{\gamma}(x_2 - \hat{x}_2^k)^\top(\lambda^k - \hat{\lambda}^k). \tag{3.4}$$

For λ-subproblem in (3.2), for any $\lambda \in \mathcal{R}^n$, one has

$$(\lambda - \hat{\lambda}^k)^\top \left(\hat{x}_1^k - \hat{x}_2^k\right) = \frac{1}{\beta\gamma}(\lambda - \hat{\lambda}^k)^\top(\lambda^k - \hat{\lambda}^k). \tag{3.5}$$

By (3.3)–(3.5), we get

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(\hat{w}^k) \geq (w - \hat{w}^k)^\top G(w^k - \hat{w}^k), \quad \forall w \in \mathcal{W}, \tag{3.6}$$

where

$$G = \begin{pmatrix} R_1 & \beta I_n & -\frac{1-\gamma}{\gamma} I_n \\ 0 & R_2 & \frac{1-\gamma}{\gamma} I_n \\ 0 & 0 & \frac{1}{\beta\gamma} I_n \end{pmatrix}. \tag{3.7}$$

Combining $w^k = \hat{w}^k$ with (3.6), one has

$$\theta(x) - \theta(\hat{x}^k) + (w - \hat{w}^k)^\top F(\hat{w}^k) \geq 0, \quad \forall w \in \mathcal{W}. \tag{3.8}$$

Substituting $\hat{w}^k$ and $\hat{x}^k$ in (3.8) with w^k and x^k, respectively, we obtain

$$\theta(x) - \theta(x^k) + (w - w^k)^\top F(w^k) \geq 0, \quad \forall w \in \mathcal{W},$$

which indicates that w^k is a solution of (2.4). □

Remark 3.1. From Lemma 3.1, if Algorithm 3.1 stops at Step 2, then w^k is a proximal solution of (2.4).

In the following, we assume that Algorithm 3.1 generates infinite sequences $\{w^k\}$ and $\{\hat{w}^k\}$. For convenience, we define two matrices to simplify our notation in the later analysis.

$$M = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \beta I_n + R_2 & 0 \\ 0 & 0 & \frac{1}{\beta\gamma} I_n \end{pmatrix}, \quad Q = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & \beta I_n + R_2 & \frac{-1}{2\gamma} I_n \\ 0 & \frac{-1}{2\gamma} I_n & \frac{1}{\beta\gamma^2} I_n \end{pmatrix}. \tag{3.9}$$

The following lemma gives some interesting properties of the two matrices M, Q just defined. These properties are crucial in the convergence analysis of Algorithm 3.1.

Lemma 3.2. *If R_1 and R_2 are two positive semi-definite matrices, then we have*
(i) *Both matrices M and Q are positive semi-definite;*
(ii) *The matrix $H_1 := 2Q - \gamma M$ is positive semi-definite if $0 < \gamma \leq 1$;*
(iii) *The matrix $H_2 := 2\gamma Q - M$ is positive semi-definite if $\gamma > 1$.*

Proof. (i) For any $w = (x_1; x_2; \lambda)$, one has

$$w^\top M w = \|x_1\|_{R_1}^2 + \beta\|x_2\|^2 + \|x_2\|_{R_2}^2 + \frac{1}{\beta\gamma}\|\lambda\|^2 \geq 0.$$

So the matrix M is positive semi-definite.
The matrix Q can be written as

$$Q = \begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta I_n & -\frac{1}{2\gamma} I_n \\ 0 & -\frac{1}{2\gamma} I_n & \frac{1}{\beta\gamma^2} I_n \end{pmatrix} := Q_1 + Q_2.$$

Obviously, the matrix Q_1 is positive semi-definite, and we have

$$\begin{aligned} w^\top Q_2 w &= \beta x_2^\top x_2 + \tfrac{1}{\beta\gamma^2}\lambda^\top\lambda - \tfrac{1}{\gamma} x_2^\top \lambda \\ &\geq \beta x_2^\top x_2 + \tfrac{1}{\beta\gamma^2}\lambda^\top\lambda - \tfrac{1}{\gamma}\|x_2\|\|\lambda\| \\ &\geq \beta x_2^\top x_2 + \tfrac{1}{\beta\gamma^2}\lambda^\top\lambda - (\tfrac{\beta}{4}\|x_2\|^2 + \tfrac{1}{\beta\gamma^2}\|\lambda\|^2) \\ &\geq \tfrac{3\beta}{4}\|x_2\|^2 \\ &\geq 0, \end{aligned}$$

where the first inequality is obtained by the Cauchy-Schwartz inequality, the second inequality follows from the fact that $a^2 + b^2 \geq 2ab, \forall a, b \in \mathcal{R}_+$, and the desired result follows.

(ii) For the matrix H_1. By a direct computation, it yields that

$$\begin{aligned} H_1 = 2Q - \gamma M &= \begin{pmatrix} (2-\gamma)R_1 & 0 & 0 \\ 0 & (2-\gamma)(\beta I_n + R_2) & -\frac{1}{\gamma} I_n \\ 0 & -\frac{1}{\gamma} I_n & \frac{2-\gamma^2}{\beta\gamma^2} I_n \end{pmatrix} \\ &= (2-\gamma)\begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2-\gamma)\beta I_n & -\frac{1}{\gamma} I_n \\ 0 & -\frac{1}{\gamma} I_n & \frac{2-\gamma^2}{\beta\gamma^2} I_n \end{pmatrix} \end{aligned}$$

$$:= (2-\gamma)Q_1 + Q_3.$$

Obviously, by $0 < \gamma \le 1$, the first part is positive semi-definite, and

$$\begin{aligned} w^\top Q_3 w &= (2-\gamma)\beta x_2^\top x_2 + \tfrac{2-\gamma^2}{\beta\gamma^2}\lambda^\top\lambda - \tfrac{2}{\gamma}x_2^\top\lambda \\ &\ge (2-\gamma)\beta x_2^\top x_2 + \tfrac{2-\gamma^2}{\beta\gamma^2}\lambda^\top\lambda - \tfrac{2}{\gamma}\|x_2\|\|\lambda\| \\ &\ge (2-\gamma)\beta x_2^\top x_2 + \tfrac{2-\gamma^2}{\beta\gamma^2}\lambda^\top\lambda - (\tfrac{\beta}{2-\gamma^2}\|x_2\|^2 + \tfrac{2-\gamma^2}{\beta\gamma^2}\|\lambda\|^2) \\ &\ge (2-\gamma)\tfrac{1}{2-\gamma^2}\beta\|x_2\|^2 \\ &\ge 0, \end{aligned}$$

and thus the desired result follows.

(iii) For the matrix H_2. By a direct computation, it yields that

$$\begin{aligned} H_2 = 2\gamma Q - M &= \begin{pmatrix} (2\gamma-1)R_1 & 0 & 0 \\ 0 & (2\gamma-1)(\beta I_n + R_2) & -I_n \\ 0 & -I_n & \frac{1}{\beta\gamma}I_n \end{pmatrix} \\ &= (2\gamma-1)\begin{pmatrix} R_1 & 0 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & (2\gamma-1)\beta I_n & -I_n \\ 0 & -I_n & \frac{1}{\beta\gamma}I_n \end{pmatrix} \\ &:= (2\gamma-1)Q_1 + Q_4. \end{aligned}$$

Similar to discussion above, using $\gamma > 1$, we obtain

$$\begin{aligned} w^\top Q_4 w &= (2\gamma-1)\beta x_2^\top x_2 + \tfrac{1}{\beta\gamma}\lambda^\top\lambda - 2x_2^\top\lambda \\ &\ge (2\gamma-1)\beta x_2^\top x_2 + \tfrac{1}{\beta\gamma}\lambda^\top\lambda - 2|x_2\|\|\lambda\| \\ &\ge (2\gamma-1)\beta x_2^\top x_2 + \tfrac{1}{\beta\gamma}\lambda^\top\lambda - (\beta\gamma\|x_2\|^2 + \tfrac{1}{\beta\gamma}\|\lambda\|^2) \\ &\ge (\gamma-1)\beta\|x_2\|^2 \\ &\ge 0. \end{aligned}$$

Combining this with the positive semi-definite of Q_1, and the desired result follows. □

Lemma 3.3. *Let $\{w^k\}$ and $\{\hat{w}^k\}$ be two sequences generated by Algorithm 3.1. Then we have*

$$(w^k - w^*)^\top M(w^k - \hat{w}^k) \ge \|w^k - \hat{w}^k\|_Q^2, \quad \forall w^* \in \mathcal{W}^*. \tag{3.10}$$

Proof. From the definitions of M in (3.9) and G in (3.7), one has

$$G = M + \tilde{M},$$

where $\tilde{M} = \begin{pmatrix} 0 & \beta I_n & -\frac{1-\gamma}{\gamma} I_n \\ 0 & -\beta I_n & \frac{1-\gamma}{\gamma} I_n \\ 0 & 0 & 0 \end{pmatrix}$. By a direct computation, it yields that

$$\begin{aligned} 0 \ & \leq \theta(\hat{x}^k) - \theta(x^*) + (\hat{w}^k - w^*)^\top F(\hat{w}^k) \\ & \leq (\hat{w}^k - w^*)^\top G(w^k - \hat{w}^k) \\ & = (\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) + (\hat{w}^k - w^*)^\top \tilde{M}(w^k - \hat{w}^k) \\ & = (\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) + \beta[(\hat{x}_1^k - x_1^*) - (\hat{x}_2^k - x_2^*)]^\top (x_2^k - \hat{x}_2^k) \\ & \quad - \tfrac{1-\gamma}{\gamma}[(\hat{x}_1^k - x_1^*) - (\hat{x}_2^k - x_2^*)]^\top (\lambda^k - \hat{\lambda}^k) \\ & = (\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) + \beta(\hat{x}_1^k - \hat{x}_2^k)^\top (x_2^k - \hat{x}_2^k) \\ & \quad - \tfrac{1-\gamma}{\gamma}(\hat{x}_1^k - \hat{x}_2^k)^\top (\lambda^k - \hat{\lambda}^k) \\ & = (\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) + \tfrac{1}{\gamma}(\lambda^k - \hat{\lambda}^k)^\top (x_2^k - \hat{x}_2^k) \\ & \quad - \tfrac{1-\gamma}{\gamma}[\tfrac{1}{\beta\gamma}(\lambda^k - \hat{\lambda}^k)^\top](\lambda^k - \hat{\lambda}^k) \\ & = (\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) + \tfrac{1}{\gamma}(\lambda^k - \hat{\lambda}^k)^\top (x_2^k - \hat{x}_2^k) - \tfrac{1-\gamma}{\beta\gamma^2}\|\lambda^k - \hat{\lambda}^k\|^2, \end{aligned} \tag{3.11}$$

where the first inequality is by (2.4) with $w^* \in \mathcal{W}^*$, $w^k \in \mathcal{W}$, since $w^* \in \mathcal{W}^* \subseteq \mathcal{W}$, using (3.6) with $x = x^*$ and $w = w^*$, we have that the second inequality holds, the third equality follows from $x_1^* = x_2^*$, and the four equality comes from the fact $(\hat{x}_1^k - \hat{x}_2^k) = \frac{1}{\gamma\beta}(\lambda^k - \hat{\lambda}^k)$. Applying (3.11), we get

$$(\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) \geq -\frac{1}{\gamma}(\lambda^k - \hat{\lambda}^k)^\top (x_2^k - \hat{x}_2^k) + \frac{1-\gamma}{\beta\gamma^2}\|\lambda^k - \hat{\lambda}^k\|^2,$$

and one has

$$\begin{aligned} & (w^k - w^*)^\top M(w^k - \hat{w}^k) \\ & = (w^k - \hat{w}^k)^\top M(w^k - \hat{w}^k) + (\hat{w}^k - w^*)^\top M(w^k - \hat{w}^k) \\ & \geq (w^k - \hat{w}^k)^\top M(w^k - \hat{w}^k) - \tfrac{1}{\gamma}(\lambda^k - \hat{\lambda}^k)^\top (x_2^k - \hat{x}_2^k) + \tfrac{1-\gamma}{\beta\gamma^2}\|\lambda^k - \hat{\lambda}^k\|^2 \\ & = \|w^k - \hat{w}^k\|_Q^2, \end{aligned}$$

where the second equality follows from the definition of matrix Q in (3.9), and the desired result follows. □

Lemma 3.4. *For any solution $w^* = (x_1^*; x_2^*; \lambda^*)$ of (2.4), the sequence $\{w^k\}$ generated by Algorithm 3.1 satisfies*

$$\|\rho\hat{w}^k + (1-\rho)w^k - w^*\|_M^2 \leq \|w^k - w^*\|_M^2 - \rho\|\hat{w}^k - w^k\|_H^2, \tag{3.12}$$

where the matrix H is defined by

$$H = (\eta - \rho)M, \tag{3.13}$$

and η is defined in (3.1).

Proof. From (3.1), if $0 < \gamma \leq 1$, then $\eta = \gamma$. A direct computation yields that

$$\begin{array}{rl} & \|\rho(\hat{w}^k - w^k) + (w^k - w^*)\|_M^2 \\ = & \|w^k - w^*\|_M^2 + 2\rho(w^k - w^*)^\top M(\hat{w}^k - w^k) + \rho^2\|\hat{w}^k - w^k\|_M^2 \\ \leq & \|w^k - w^*\|_M^2 - 2\rho\|\hat{w}^k - w^k\|_Q^2 + \rho^2\|\hat{w}^k - w^k\|_M^2 \\ = & \|w^k - w^*\|_M^2 - 2\rho\|\hat{w}^k - w^k\|^2_{(\frac{1}{2}H_1 + \frac{\gamma}{2}M)} + \rho^2\|\hat{w}^k - w^k\|_M^2 \\ = & \|w^k - w^*\|_M^2 - (\hat{w}^k - w^k)^\top(\rho H_1 + (\rho\gamma - \rho^2)M)(\hat{w}^k - w^k) \\ = & \|w^k - w^*\|_M^2 - (\hat{w}^k - w^k)^\top\rho(\gamma - \rho)M(\hat{w}^k - w^k)) \\ & -(\hat{w}^k - w^k)^\top\rho H_1(\hat{w}^k - w^k) \\ \leq & \|w^k - w^*\|_M^2 - \rho\|\hat{w}^k - w^k\|^2_{(\gamma-\rho)M}, \end{array} \tag{3.14}$$

where the first inequality follows from (3.10), the second equality follows from $H_1 = 2Q - \gamma M$ in Lemma 3.2 (*ii*), and the second inequality follows from the fact that the matrix H_1 is positive semi-definite in Lemma 3.2.
If $\gamma \geq 1$, then $\eta = \frac{1}{\gamma}$. Similar to discussion (3.14), we can also obtain

$$\begin{array}{rl} & \|\rho(\hat{w}^k - w^k) + (w^k - w^*)\|_M^2 \\ \leq & \|w^k - w^*\|_M^2 - 2\rho\|\hat{w}^k - w^k\|_Q^2 + \rho^2\|\hat{w}^k - w^k\|_M^2 \\ = & \|w^k - w^*\|_M^2 - 2\rho\|\hat{w}^k - w^k\|^2_{(\frac{1}{2\gamma}H_2 + \frac{1}{2\gamma}M)} + \rho^2\|\hat{w}^k - w^k\|_M^2 \\ = & \|w^k - w^*\|_M^2 - (\hat{w}^k - w^k)^\top(\frac{\rho}{\gamma}H_2 + \rho(\frac{1}{\gamma} - \rho)M(\hat{w}^k - w^k) \\ = & \|w^k - w^*\|_M^2 - \rho\|\hat{w}^k - w^k\|^2_{(\frac{1}{\gamma}-\rho)M} \\ & -\frac{\rho}{\gamma}(\hat{w}^k - w^k)^\top H_2(\hat{w}^k - w^k) \\ \leq & \|w^k - w^*\|_M^2 - \rho\|\hat{w}^k - w^k\|^2_{(\frac{1}{\gamma}-\rho)M}, \end{array} \tag{3.15}$$

where the first equality comes from $H_2 = 2\gamma Q - M$ in Lemma 3.2 (*iii*), and the second inequality follows from the fact that the matrix H_2 is positive semi-definite in Lemma 3.2.
The desired result follows by combining above. □

Remark 3.2. By the definition of η in (3.1), the matrix H in (3.13) is positive semi-definite.

Theorem 3.1. For any solution $w^* = (x_1^*; x_2^*; \lambda^*)$ of (2.4), the sequence $\{w^k\}$ generated by Algorithm 3.1 satisfies

$$\|w^{k+1} - w^*\|_M \leq \|w^k - w^*\|_M + (1 - \rho)\|\delta(w^k)\|_M. \tag{3.16}$$

Proof. Since the matrix H in (3.13) is positive semi-definite, by (3.12), one has

$$\begin{aligned} \|w^{k+1} - w^*\|_M &= \|\rho\hat{w}^k + (1 - \rho)\psi(w^k) - w^*\|_M \\ &= \|\rho(\hat{w}^k - w^k) + (w^k - w^*) + (\rho - 1)(w^k - \psi(w^k))\|_M \\ &\leq \|\rho(\hat{w}^k - w^k) + (w^k - w^*)\|_M + (1 - \rho)\|\delta(w^k)\|_M \\ &\leq \|w^k - w^*\|_M + (1 - \rho)\|\delta(w^k)\|_M. \end{aligned} \tag{3.17}$$

Thus, the desired result follows. □

Theorem 3.2. *Assume that the matrix R_1 is positive definite. Then the sequence $\{w^k\}$ generated by Algorithm 3.1 converges to some $\bar{w} \in \mathcal{W}^*$.*

Proof. Using Definition 2.1, there exists a constant $c_1 > 0$ such that $\|\delta(w^k)\|_M \le \frac{c_1}{2^k}$. Combining this with (3.16), we obtain

$$\begin{aligned} \|w^{k+1} - w^*\|_M &\le \|w^k - w^*\|_M + (1-\rho)c_1 \tfrac{1}{2^k} \\ &\le \|w^{k-1} - w^*\|_M + (1-\rho)c_1 \left(\tfrac{1}{2^k} + \tfrac{1}{2^{k-1}}\right) \\ &\le \cdots\cdots\cdots \\ &\le \|w^1 - w^*\|_M + (1-\rho)c_1 \textstyle\sum_{m=1}^{k} \tfrac{1}{2^m} \\ &\le \|w^1 - w^*\|_M + (1-\rho)c_1 \textstyle\sum_{m=1}^{\infty} \tfrac{1}{2^m}, \end{aligned} \tag{3.18}$$

where $w^* \in \mathcal{W}^*$. Since that the matrix R_1 is positive definite, it is easy to obtain that the matrix M is positive definite. Combining this with (3.18), we have that $\{w^k\}$ is bounded.

Now, we break up the discussion into two cases.

Case 1. If there exists a subsequence $\{w^{k_j}\}$ such that $\|w^{k_j} - w^*\|_M \le (1-\rho)\|\delta(w^{k_j})\|_M$, i.e., $\|w^{k_j} - w^*\| \le (1-\rho)c_1 \frac{1}{2^{k_j}}$, Then $\{w^{k_j}\}$ converges to w^*. Since the series of positive terms $\sum_{k=1}^{\infty} \frac{1}{2^k}$ is convergent, by Cauchy convergence criterion, for any $\epsilon > 0$, there exists a positive integer m such that

$$\frac{1}{2^k} + \frac{1}{2^{k-1}} + \cdots + \frac{1}{2^{k_j}} \le \frac{\epsilon}{2c_1(1-\rho)} \tag{3.19}$$

as positive integer $k, k_j \ge m$ ($k \ge k_j$). From $\lim_{j\to\infty} w^{k_j} = w^*$, for $\epsilon > 0$ above, there exists an integer j, such that

$$\|w^{k_j} - w^*\|_M < \frac{\epsilon}{2}. \tag{3.20}$$

Combining (3.19) with (3.20), for sufficiently large positive integer $k, k_j (k \ge k_j)$, similar to discussion (3.18), we obtain

$$\begin{aligned} \|w^{k+1} - w^*\|_M &\le \|w^k - w^*\|_M + (1-\rho)c_1 \tfrac{1}{2^k} \\ &\le \|w^{k-1} - w^*\|_M + (1-\rho)c_1 \left(\tfrac{1}{2^k} + \tfrac{1}{2^{k-1}}\right) \\ &\le \cdots\cdots\cdots \\ &\le \|w^{k_j} - w^*\|_M + (1-\rho)c_1 \left(\tfrac{1}{2^k} + \tfrac{1}{2^{k-1}} + \cdots + \tfrac{1}{2^{k_j}}\right) \\ &< \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon, \end{aligned}$$

which indicates that the sequence $\{w^k\}$ converges globally to the point w^*.

Case 2. For any subsequence $\{w^{k_j}\}$ such that $\|w^{k_j} - w^*\|_M > (1-\rho)\|\delta(w^{k_j})\|_M$, i.e., $\|w^{k+1} - w^*\|_M > (1-\rho)\|\delta(w^k)\|_M$.

Since $\{w^k\}$ is bounded, there exists constant $c_2 > 0$ such that $\|w^{k+1} - w^*\| \le c_2$. By definition of w^{k+1} in Step 3 of Algorithm 3.1, we have

$$\begin{aligned} & \|\rho(\hat{w}^k - w^k) + (w^k - w^*)\|_M^2 \\ = \;& \|w^{k+1} - w^* + (1-\rho)\delta(w^k)\|_M^2 \\ \ge \;& (\|w^{k+1} - w^*\| - (1-\rho)\|\delta(w^k)\|_M)^2 \\ = \;& \|w^{k+1} - w^*\|_M^2 + (1-\rho)^2\|\delta(w^k)\|_M^2 - 2(1-\rho)\|w^{k+1} - w^*\|\|\delta(w^k)\| \\ \ge \;& \|w^{k+1} - w^*\|_M^2 + (1-\rho)^2\|\delta(w^k)\|_M^2 - 2(1-\rho)c_2\|\delta(w^k)\|. \end{aligned} \tag{3.21}$$

Combining this with (3.12), one has

$$\begin{aligned}
&\textstyle\sum_{k=0}^{\infty}\|w^k-\hat{w}^k\|_H^2\\
\le\ &\rho^{-1}\textstyle\sum_{k=0}^{\infty}(\|w^k-w^*\|_M^2-\|\rho\hat{w}^k+(1-\rho)w^k-w^*)\|_M^2)\\
\le\ &2\rho^{-1}(1-\rho)c_2\textstyle\sum_{k=0}^{\infty}\|\delta(w^k)\|-\rho^{-1}(1-\rho)^2\sum_{k=0}^{\infty}\|\delta(w^k)\|_M^2\\
&+\rho^{-1}\textstyle\sum_{k=0}^{\infty}(\|w^k-w^*\|_M^2-\|w^{k+1}-w^*\|_M)\\
\le\ &2\rho^{-1}(1-\rho)c_1c_2\textstyle\sum_{k=0}^{\infty}\frac{1}{2^k}+\rho^{-1}\|w^0-w^*\|_M^2,
\end{aligned} \tag{3.22}$$

which together with the positive definiteness of H yields

$$\lim_{k\to\infty}\|w^k-\hat{w}^k\|=0, \tag{3.23}$$

and one has

$$\lim_{k\to\infty}\|x_1^k-\hat{x}_1^k\|=0, \tag{3.24}$$

$$\lim_{k\to\infty}\|x_2^k-\hat{x}_2^k\|=0, \tag{3.25}$$

$$\lim_{k\to\infty}\|\lambda^k-\hat{\lambda}^k\|=0. \tag{3.26}$$

By (3.23), we know that the sequence $\{\hat{w}^k\}$ is also bounded since $\{w^k\}$ is bounded. Thus, it has at least a cluster point, saying $w^\infty := (x_1^\infty; x_2^\infty; \lambda^\infty)$, and suppose that the subsequence $\{\hat{w}^{k_i}\}$ converges to w^∞. By (3.26), one has $\lim_{k_i\to\infty}\|\lambda^{k_i}-\hat{\lambda}^{k_i}\|=0$. Taking limits on both sides of

$$\hat{x}_1^{k_i}-\hat{x}_2^{k_i}=\frac{1}{\gamma\beta}(\lambda^{k_i}-\hat{\lambda}^{k_i}),$$

we have

$$x_1^\infty-x_2^\infty=0.$$

Furthermore, taking limits on both sides of (3.3) and (3.4), and using (3.24)–(3.26), we obtain

$$\theta_1(x_1)-\theta_1(x_1^\infty)+(x_1-x_1^\infty)^\top(-\lambda^\infty)\ge 0,\quad \forall x_1\in R^n,$$

and

$$\theta_2(x_2)-\theta_2(x_2^\infty)+(x_2-x_2^\infty)^\top\lambda^\infty\ge 0,\quad \forall x_2\in R^n.$$

Therefore, $(x_1^\infty, x_2^\infty, \lambda^\infty)\in\mathcal{W}^*$.
Since the series of positive terms $\sum_{k=1}^{\infty}\frac{1}{2^k}$ is convergent, by Cauchy convergence criterion, for any $\epsilon>0$, there exists a positive integer m such that

$$\frac{1}{2^k}+\frac{1}{2^{k-1}}+\cdots+\frac{1}{2^{k_l}}\le\frac{\epsilon}{3c_1(1-\rho)} \tag{3.27}$$

as positive integer $k, k_l\ge m(k\ge k_l)$. By (3.23) and $\lim_{j\to\infty}\hat{w}^{k_j}=w^\infty$, for $\epsilon>0$ above, there exists an integer l, such that

$$\|w^{k_l}-\hat{w}^{k_l}\|_M<\frac{\epsilon}{3},\quad \|\hat{w}^{k_l}-w^\infty\|_M<\frac{\epsilon}{3}. \tag{3.28}$$

Combining (3.27) with (3.28), for sufficiently large positive integer $k, k_l (k \geq k_l)$, similar to discussion (3.18), we obtain

$$\begin{aligned}
\|w^{k+1} - w^{\infty}\|_M &\leq \|w^k - w^{\infty}\|_M + (1-\rho)c_1 \tfrac{1}{2^k} \\
&\leq \|w^{k-1} - w^{\infty}\|_M + (1-\rho)c_1 \left(\tfrac{1}{2^k} + \tfrac{1}{2^{k-1}}\right) \\
&\leq \cdots\cdots\cdots \\
&\leq \|w^{k_l} - w^{\infty}\|_M + (1-\rho)c_1 \left(\tfrac{1}{2^k} + \tfrac{1}{2^{k-1}} + \cdots + \tfrac{1}{2^{k_l}}\right) \\
&\leq \|w^{k_l} - \hat{w}^{k_l}\|_M + \|\hat{w}^{k_l} - w^{\infty}\|_M + (1-\rho)c_1 \left(\tfrac{1}{2^k} + \tfrac{1}{2^{k-1}} + \cdots + \tfrac{1}{2^{k_l}}\right) \\
&< \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} = \epsilon,
\end{aligned}$$

which indicates that the sequence $\{w^k\}$ converges globally to the point w^{∞}. The proof is completed. □

In the end of this section, we discuss the convergence rate of Algorithm 3.1. To this end, the following assumptions is needed.

Assumption 3.1. *For two sequences $\{w^k\}$ and $\{\hat{w}^k\}$, assume that there exist a positive constant σ such that, for w^k, there exists $w^* \in \mathcal{W}^*$ such that*

$$\|w^k - w^*\| \leq \sigma \|\nabla L_k(w^k)\|,$$

where

$$\begin{aligned}
L_k(w) := \;& \theta_1(x_1) + \mu\partial(\|\hat{x}_2^k\|_1)^\top x_2 - \langle \lambda^k, x_1 - x_2\rangle + \tfrac{\beta}{2}\|x_1 - x_2^k\|^2 + \tfrac{\beta}{2}\|\hat{x}_1^k - x_2\|^2 \\
& + \tfrac{1}{2}\|x_1 - x_1^k\|_{R_1}^2 + \tfrac{1}{2}\|x_2 - x_2^k\|_{R_2}^2 + \tfrac{1}{2}\lambda^\top\lambda - \langle \lambda, \lambda^k - \gamma\beta(\hat{x}_1^k - \hat{x}_2^k)\rangle,
\end{aligned} \tag{3.29}$$

and $\partial(\|\hat{x}_2^k\|_1)$ denotes the subdifferential of the function $\|x_2\|_1$ at the point $\hat{x}_2^k$.

Lemma 3.5. *Suppose that Assumption 3.1 holds. Then there exists a positive constant $\hat{\mu}$ such that*

$$\|w^k - w^*\| \leq \hat{\mu}\|w^k - \hat{w}^k\|.$$

Proof. From (3.2), we have

$$\begin{cases}
\hat{x}_1^k = (A^\top A + \beta I + R_1)^{-1}(A^\top b + \lambda^k + \beta x_2^k + R_1 x_1^k) \\
\hat{x}_2^k = (\beta I + R_2)^{-1}(-\mu\partial(\|\hat{x}_2^k\|_1) - \lambda^k + \beta\hat{x}_1^k + R_2 x_2^k) \\
\hat{\lambda}^k = \lambda^k - \gamma\beta(\hat{x}_1^k - \hat{x}_2^k),
\end{cases} \tag{3.30}$$

By (3.29) and (3.30), a direct computation yields that

$$\begin{aligned}
\nabla L_k(w) &= \left(\tfrac{\partial L}{\partial x_1}; \tfrac{\partial L}{\partial x_2}; \tfrac{\partial L}{\partial \lambda}\right) \\
&= \begin{pmatrix} A^\top(Ax_1 - b) - \lambda^k + \beta(x_1 - x_2^k) + R_1(x_1 - x_1^k) \\ \mu\partial(\|\hat{x}_2^k\|_1) + \lambda^k - \beta(\hat{x}_1^k - x_2) + R_2(x_2 - x_2^k) \\ \lambda - [\lambda^k - \gamma\beta(\hat{x}_1^k - \hat{x}_2^k)] \end{pmatrix} \\
&= \begin{pmatrix} (A^\top A + \beta I + R_1)[x_1 - (A^\top A + \beta I + R_1)^{-1}(A^\top b + \lambda^k + \beta x_2^k + R_1 x_1^k)] \\ (\beta I + R_2)[x_2 - (\beta I + R_2)^{-1}(-\mu\partial(\|\hat{x}_2^k\|_1) - \lambda^k + \beta\hat{x}_1^k + R_2 x_2^k)] \\ \lambda - [\lambda^k - \gamma\beta(\hat{x}_1^k - \hat{x}_2^k)] \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{pmatrix} A^\top A + \beta I + R_1 & 0 & 0 \\ 0 & \beta I + R_2 & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} x_1 - \hat{x}_1^k \\ x_2 - \hat{x}_2^k \\ \lambda - \hat{\lambda}^k \end{pmatrix} \\
&= \bar{Q}(w - \hat{w}^k),
\end{aligned}
$$

where $\bar{Q} = \begin{pmatrix} A^\top A + \beta I + R_1 & 0 & 0 \\ 0 & \beta I + R_2 & 0 \\ 0 & 0 & I \end{pmatrix}$. Combining this with Assumption 3.1, we obtain

$$
\|w^k - w^*\| \le \sigma \|\nabla L_k(w^k)\| = \sigma \|\bar{Q}(w^k - \hat{w}^k)\| \le \sigma \|\bar{Q}\| \|w^k - \hat{w}^k\|.
$$

Let $\hat{\mu} = \sigma \|\bar{Q}\|$. Then the desired result follows. □

Theorem 3.3 *Suppose that the hypothesis of Theorem 3.2 and Assumption 3.1 hold, and $\frac{3}{4}\hat{\mu}^2 < (\eta - \rho)\rho < \frac{5}{4}\hat{\mu}^2$. Then the sequence $\{w^k\}$ generated by Algorithm 3.1 converges to a solution of (2.4) R−linearly.*

Proof. From Theorem 3.2. Without loss of generality, we assume that the sequence $\{w^k\}$ converges to $w^* \in \mathcal{W}^*$. A direct computation yields that

$$
\begin{aligned}
\|w^{k+1} - w^*\|_M^2 &= \|\rho \hat{w}^k + (1-\rho)\psi(w^k) - w^*\|_M^2 \\
&= \|\rho(\hat{w}^k - w^k) + (w^k - w^*) + (\rho - 1)(w^k - \psi(w^k))\|_M^2 \\
&\le (\|\rho(\hat{w}^k - w^k) + (w^k - w^*)\|_M + (1-\rho)\|\delta(w^k)\|_M)^2 \\
&\le 2\|\rho(\hat{w}^k - w^k) + (w^k - w^*)\|_M^2 + 2(1-\rho)^2\|\delta(w^k)\|_M^2 \\
&\le 2\|w^k - w^*\|_M^2 - 2\rho\|\hat{w}^k - w^k\|_H^2 + 2(1-\rho)^2\|\delta(w^k)\|_M^2 \\
&\le 2\|w^k - w^*\|_M^2 - \tfrac{2\rho}{\hat{\mu}^2}\|w^k - w^*\|_H^2 + 2(1-\rho)^2\|\delta(w^k)\|_M^2 \\
&\le 2(1 - \tfrac{(\eta-\rho)\rho}{\hat{\mu}^2})\|w^k - w^*\|_M^2 + 2(1-\rho)^2\|\delta(w^k)\|_M^2.
\end{aligned}
\tag{3.31}
$$

where the second inequality follows from the fact that $a^2 + b^2 \ge 2ab, \forall a, b \in R$, the third inequality is obtained by (3.12), the fourth inequality follows by Lemma 3.5, and the fifth inequality is by (3.13). Since the sequence $\{w^k\}$ converges to w^*, it follows that there exists a positive integer k_0, for all $k \ge k_0$, we obtain the following conclusions.
If $\|w^k - w^*\|_M \le 2(1-\rho)\|\delta(w^k)\|_M$. By Definition 2.1, the following holds

$$
\|w^k - w^*\| \le 2(1-\rho)c_1 \frac{1}{2^k}. \tag{3.32}
$$

If $\|w^k - w^*\|_M > 2(1-\rho)\|\delta(w^k)\|_M$. Combining this with (3.31), we obtain

$$
\frac{\|w^{k+1} - w^*\|_M^2}{\|w^k - w^*\|_M^2} \le \left[\frac{2(1 - \frac{(\eta-\rho)\rho}{\hat{\mu}^2})\|w^k - w^*\|_M^2}{\|w^k - w^*\|_M^2} + \frac{2(1-\rho)^2\|\delta(w^k)\|_M^2}{\|w^k - w^*\|_M^2} \right] \le 2(1 - \frac{(\eta-\rho)\rho}{\hat{\mu}^2}) + \frac{1}{2}.
$$

Let $\hat{\tau} = 2(1 - \frac{(\eta-\rho)\rho}{\hat{\mu}^2}) + \frac{1}{2}$. Then $0 < \hat{\tau} < 1$ by $\frac{3}{4}\hat{\mu}^2 < (\eta - \rho)\rho < \frac{5}{4}\hat{\mu}^2$. Therefore, we have

$$\begin{aligned} \|w^{k+1} - w^*\|_M &\le \sqrt{\hat{\tau}}\|w^k - w^*\|_M \\ &\le \sqrt{\hat{\tau}}^2\|w^{k-1} - w^*\|_M \\ &\cdots\cdots\cdots \\ &\le \sqrt{\hat{\tau}}^{k-k_0+1}\|w^{k_0} - w^*\|_M \\ &\le c_2\sqrt{\hat{\tau}}^k. \end{aligned}$$

where c_2 is positive constant, i.e., $\|w^k - w^*\|_M \le c_2\sqrt{\hat{\tau}}^{-1}\sqrt{\hat{\tau}}^k$. Combining this with (3.32), one has

$$\|w^k - w^*\|_M \le c_3 \max\{\frac{1}{2^k}, \sqrt{\hat{\tau}}^k\}.$$

where c_3 is positive constant, and thus the desired result follows. □

4. Numerical results

In this section, we provide some numerical tests about BPDN problem to show the efficiency of method proposed in this paper. All codes are written by MATLAB 9.2.0.538062 and performed on a Windows 10 PC with an AMD FX-7500 Radeon R7,10 Computer Cores 4C+6G CPU, 2.10GHz CPU and 8GB of memory. In experiment, we set $\mu = 0.001$, and the measurement matrix A is generated by MATLAB scripts:

$$[Q, R] = qr(A', 0); A = Q'.$$

The original signal $\bar{x}$ is generated by $p = randperm(n); x(p(1:k)) = randn(k, 1)$.

To simplify calculations of (3.2), we set $R_1 = \tau I_n - A^\top A$, where $\tau > 0$ is a constant. Combining this with the first equality in (3.30), one has

$$\hat{x}_1^k = \frac{1}{\beta + \tau}(\lambda^k + \tau x_1^k + \beta x_2^k - g^k), \tag{4.1}$$

where $g^k = A^\top(Ax_1^k - b)$.

For the second equality in (3.30), we set $R_2 = 0$, then

$$\hat{x}_2^k = \hat{x}_1^k - \frac{1}{\beta}\lambda^k - \frac{\mu}{\beta}\partial(\|\hat{x}_2^k\|_1). \tag{4.2}$$

The subdifferential of the absolute value function $|t|$ is given as follows

$$\partial(|t|) = \begin{cases} -1 & \text{if } t < 0, \\ [-1, 1] & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases}$$

Combining this with (4.2), for $i = 1, 2, \cdots, n$, we obtain

$$\begin{aligned} (\hat{x}_2^k)_i &= \begin{cases} (\hat{x}_1^k - \frac{1}{\beta}\lambda^k)_i + \frac{\mu}{\beta} & \text{if } (\hat{x}_1^k - \frac{1}{\beta}\lambda^k)_i < -\frac{\mu}{\beta} \\ 0 & \text{if } |\hat{x}_1^k - \frac{1}{\beta}\lambda^k)_i| < \frac{\mu}{\beta} \\ (\hat{x}_1^k - \frac{1}{\beta}\lambda^k)_i - \frac{\mu}{\beta} & \text{if } (\hat{x}_1^k - \frac{1}{\beta}\lambda^k)_i > \frac{\mu}{\beta} \end{cases} \\ &= \left(\text{shrink}_{\frac{\mu}{\beta}}(\hat{x}_1^k - \tfrac{1}{\beta}\lambda^k)\right)_i, \end{aligned}$$

where $\text{shrink}_c(*)$ is the soft-thresholding operator defined as

$$\text{shrink}_c(k) := k - \min\{c, |k|\}\frac{k}{|k|}, \quad \forall k \in R,$$

and $c > 0$ is a constant. In addition, when $k = 0$, $k/|k|$ should be taken 0. Therefore, we have the following formula to calculate $\hat{x}_2^k$, i.e.,

$$\hat{x}_2^k = \text{shrink}_{\frac{\mu}{\beta}}(\hat{x}_1^k - \frac{1}{\beta}\lambda^k). \tag{4.3}$$

Applying (4.1) and (4.3), then (3.2) in Algorithm 3.1 can be written as follow

$$\begin{cases} \hat{x}_1^k = \frac{1}{\beta+\tau}(\lambda^k + \tau x_1^k + \beta x_2^k - g^k), \\ \hat{x}_2^k = \text{shrink}_{\frac{\mu}{\beta}}(\hat{x}_1^k - \frac{1}{\beta}\lambda^k), \\ \hat{\lambda}^k = \lambda^k - \gamma\beta\left(\hat{x}_1^k - \hat{x}_2^k\right). \end{cases} \tag{4.4}$$

For any methods, the stop criterion is

$$\frac{\|f_k - f_{k-1}\|}{\|f_{k-1}\|} < 10^{-6},$$

where f_k denotes the objective value of (1.1) at iteration x_k.

In each test, we calculate the relative error

$$\text{RelErr} = \frac{\|x^{k+1} - \bar{x}\|}{\|\bar{x}\|},$$

where $\bar{x}$ denotes the recovery signal.

4.1. Test on additive Gaussian white noise

In this subsection, we apply Algorithm 2.3 in [11] (DFCGPM), Algorithm 2 in [12] (IMFPPA) and Algorithms 3.1 in this paper to recover a simulated sparse signal of which observation data is corrupted by additive Gaussian white noise, respectively. We set $n = 2^{12}, m = 2^{10}, k = 2^7$, and some parameters about tested algorithms are listed as follows:
Algorithms 3.1: $\beta = 0.2, \mu = 0.01, \gamma = 1.9, \tau = 0.5$;
IMFPPA: $\rho = 0.01, \gamma = 0.01, \tau = 0.2$;
DFCGPM: $C = 1, r = 0, \eta = 1, \rho = 0.4, \sigma = 0.01, \gamma = 1.9$.

The original signal, the measurement and the reconstructed signal(marked by red point) by Algorithm 3.1, DFCGPM and IMPPA are given in Figure 1. Obviously, from the first, third, fourth and the last subplots in Figure 1, all elements in the original signal are circled by the red points, which indicates that the three methods can recover the original signal quite well. Furthermore, we record the number of iterations (abbreviated as Iter), the CPU time in seconds (abbreviated as CPU Time), the relative error (abbreviated as RelErr) of the three methods. Figure 1 indicates that Algorithm 3.1 have higher accuracy than IMPPA method, and Algorithm 3.1 is also always faster than DFCGPM and IMPPA methods. Thus, Algorithm 3.1 is an efficient method for sparse signal recovery.

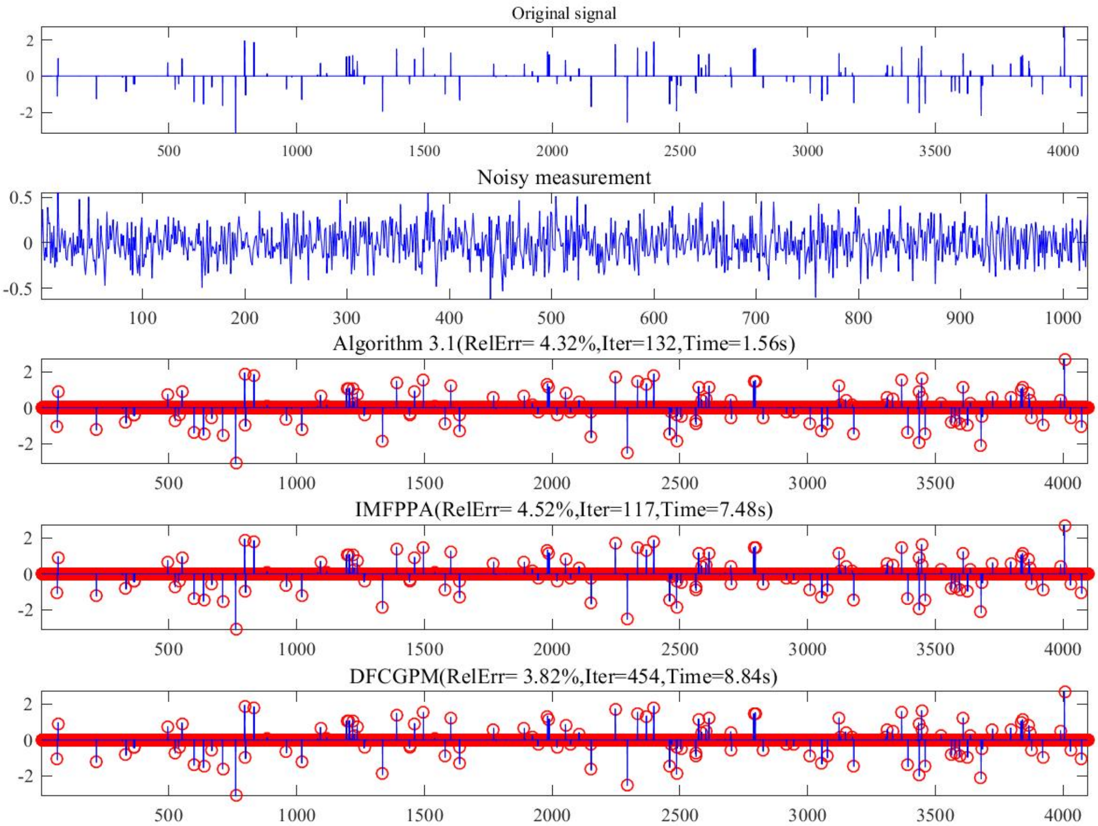

Figure 1. Signal recovery result.

4.2. Compared from different k-Sparse signal ($n = 2^{12}, m = 2^{10}$)

In this subsection, Algorithm 3.1 proposed in this paper is compared with IMFPPA [12], DFCGPM [11], Algorithm 3.1 in [8] (PPRSM) and Algorithm 3.1 in [13] (LAPM) from the CPU Time and the RelErr, where some parameters about PPRSM and LAPM are listed as follows:
PPRSM: $\gamma = 0.2, \sigma = 0.1$;
LAPM: $\beta = 0.25, \tau = 0.6$.

All algorithms have run 5 times, respectively, and the average of the the CPU Time and the RelErr are obtained. The numerical results are listed in Table 1. From the Table1, It is obvious that the CPU time of Algorithm 3.1 is less than other algorithms in different k-Sparse signal whether it is Free noise or Gaussian noise, which shows that Algorithm 3.1 is faster. In addition, we find that the accuracy of algorithm 3.1 are also better than other algorithms. So, Algorithm 3.1 is a more efficient method for different k-Sparse signal recovery.

Table 1. Compared Free noise with Gaussian noise from different k-Sparse signal (Subsection 4.2).

		No noise		Gaussian noise	
k-Sparse signal	Methods	CPU Time	RelErr	CPU Time	RelErr
80	Algorithm 3.1	1.0937	3.3782	1.1250	3.4187
	IMFPPA	4.6253	4.1837	4.7031	4.7528
	DFCGPM	4.9585	3.8375	5.2744	3.8501
	PPRSM	7.2164	4.6751	7.5363	4.6751
	LAPM	12.8906	4.5521	13.4894	4.5095
120	Algorithm 3.1	1.4531	3.0063	1.4688	3.1883
	IMFPPA	6.8872	4.9351	7.2431	5.2315
	DFCGPM	7.1964	3.5026	7.2173	3.8449
	PPRSM	8.3361	4.3693	8.5913	4.3559
	LAPM	13.2656	4.1577	14.7813	4.4311
140	Algorithm 3.1	1.6719	3.8609	1.7813	3.7207
	IMFPPA	7.1722	5.1312	8.8313	4.8873
	DFCGPM	7.6563	3.9326	8.3519	4.4939
	PPRSM	8.9961	4.5807	9.1320	4.5431
	LAPM	14.1875	4.8401	15.6406	4.9417
160	Algorithm 3.1	1.9844	4.4808	2.1875	4.3251
	IMFPPA	8.2192	4.9317	9.2268	4.9718
	DFCGPM	8.3548	3.7487	8.8897	4.4939
	PPRSM	9.4201	4.6442	9.8942	4.4337
	LAPM	16.3594	4.9220	17.6375	4.9417

4.3. Compared with DFCGPM and IMFPPA in Different Dimensions

In this subsection, Algorithm 3.1 proposed in this paper is compared with DFCGPM [11] and IMFPPA [12] from aspects of the CPU Time and the RelErr in different dimension, where some parameters about DFCGPM and IMFPPA are given in Subsection 4.1. We set $m = \frac{n}{4}, k = \frac{n}{32}$ and no additive Gaussian white noise. All algorithms have run 5 times, respectively. The average of the CPU Time and the RelErr are obtained. Some numerical results are listed in Table 2, where the IMFPPA and DFCGPM are difficult to solve the problem in our computer when $n \geq 10,000$ since our computer configuration constraints, and it is also drawn in Figure 2. The numerical results in Table 2 and Figure 2 indicates that: The CPU Time and RelErr of Algorithm 3.1 are less than that of the other two tested methods, which shows that Algorithm 3.1 is more effective for large scale problem. Thus, Algorithm 3.1 is very suitable for solving large-scale problems.

Table 2. Compared with DFCGPM method and IMFPPA method from results (Subsection 4.3).

Dimension (n)	Algorithm 3.1		IMFPPA		DFCGPM	
	CPU Time	RelErr(%)	CPU Time	RelErr(%)	CPU Time	RelErr(%)
1024	0.0625	2.6592	0.6250	4.2036	1.1205	2.9219
2048	0.2813	3.6120	2.3281	4.0544	2.1701	3.6048
3072	0.7188	3.2769	3.7188	4.2890	5.2639	3.1776
4096	1.5313	3.2059	7.6563	4.8450	9.7066	3.6737
5120	2.4688	3.3400	12.6094	4.7120	13.0390	4.2521
6144	3.6094	3.5239	16.8594	4.8617	16.3323	4.7696
7168	4.3594	3.4662	22.7031	5.0325	24.6479	3.3939
8192	6.1153	3.2848	29.1563	4.8082	32.3971	3.5910
9216	7.7188	3.4343	39.4844	4.9214	41.6752	3.8263
10238	9.5469	3.6712	-	-	-	-
11262	10.4688	3.4449	-	-	-	-
12286	12.6875	3.2772	-	-	-	-
13310	14.2969	3.3361	-	-	-	-
14334	17.5313	3.2459	-	-	-	-

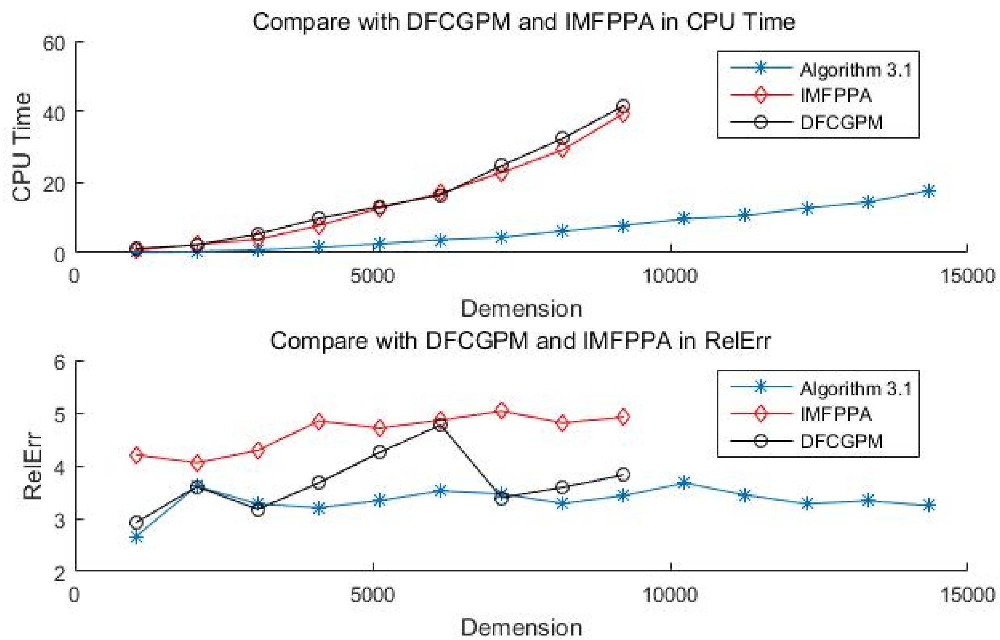

Figure 2. Compared with DFCGPM and IMFPPA in CPU Time and RelErr (Subsection 4.3).

5. Discussion

The method proposed in this work has several possible extensions. Firstly, it could be numerically beneficial to tune the parameter, and thus it is meaningful to investigate the global convergence of the proposed method with adaptively adjusted parameter. Secondly, we may establish global error bound for (1.1) just as was done for generalized linear complementarity problem in [16–18], and may use the error bound estimation to establish quick convergence rate of the new Algorithm 3.1 for solving (1.1). This is a topic for future research.

Since the RNNM model is a convex program, we explore the possibility of the proposed algorithm developed for BPDN model to solve the RNNM model from theoretical results and numerical experiments. This will be our further research direction.

The Regularized Nuclear Norm Minimization (RNNM) model is defined as follows [19, 20]:

$$\min_X \quad \frac{1}{2}\|\mathcal{A}(X) - b\|^2 + \mu\|X\|_*,$$

where $\mu > 0$ is a parameter, $b \in R^m$ is an observed vector, $\mathcal{A} : R^{n_1 \times n_2} \longrightarrow R^m$ is a known linear measurement map defined as

$$\mathcal{A}(X) = [tr(X^\top A^{(1)}), tr(X^\top A^{(2)}), \cdots, tr(X^\top A^{(m)})]^\top.$$

Here, $A^{(i)}$ for $i = 1, 2, \cdots, m$ is denoted as a matrix with size $n_1 \times n_2$, and $tr(\cdot)$ is the trace function, the norm $\|\cdot\|_*$ denote the Euclidean nuclear norm.

6. Conclusions

In this paper, by choosing a special iterative format, we have developed a new iterative format of proximal ADMM, which has closed-form solutions. Thus, it has fast solving speed and pinpoint accuracy when the dimension increases. It makes new algorithm very attractive for solving large-scale problems. The global convergence of new method is discussed in detail. Furthermore, the linear rate convergence result for new algorithm is established. Some numerical experiments on sparse signal recovery are given, and compared with the state-of-the-art of algorithms in [8, 11–13], the method proposed in this paper is more accurate and efficient.

Acknowledgments

The authors wish to give their sincere thanks to the Editor-in-chief, the Associated Editor and the anonymous referees for their valuable suggestions and helpful comments which improved the presentation of the paper.

This study was funded by the Natural Science Foundation of China (Nos. 12071250, 11801309), and Shandong Provincial Natural Science Foundation (ZR2021MA088).

Conflict of interest

No potential conflict of interest was reported by the authors.

References

1. J. F. Yang, Y. Zhang, Alternating direction algorithms for ℓ_1-problems in compressive sensing, *SIAM J. Sci. Comput.*, **33** (2011), 250–278. https://doi.org/10.1137/090777761
2. X. M. Yuan, An improved proximal alternating direction method for monotone variational inequalities with separable structure, *Comput. Optim. Appl.*, **49** (2011), 17–29. https://doi.org/10.1007/s10589-009-9293-y
3. M. H. Xu, T. Wu, A class of linearized proximal alternating direction methods, *J. Optim. Theory Appl.*, **151** (2011), 321–337. https://doi.org/10.1007/s10957-011-9876-5
4. Y. H. Xiao, H. N. Song, An inexact alternating directions algorithm for constrained total variation regularized compressive sensing problems, *J. Math. Imaging Vis.*, **44** (2012), 114–127. https://doi.org/10.1007/s10851-011-0314-y
5. Y. C. Yu, J. G. Peng, X. L. Han, A. G. Cui, A primal Douglas-Rachford splitting method for the constrained minimization problem in compressive sensing, *Circuits Syst. Signal Process*, **36** (2017), 4022–4049. https://doi.org/10.1007/s00034-017-0498-5
6. H. J. He, D. R. Han, A distributed Douglas-Rachford splitting method for multi-block convex minimization problems, *Adv. Comput. Math.*, **42** (2016), 27–53. https://doi.org/10.1007/s10444-015-9408-1
7. B. S. He, H. Liu, Z. R. Wang, X. M. Yuan, A strictly contractive Peaceman-Rachford splitting method for convex programming, *SIAM J. Optim.*, **24** (2014), 1011–1040. https://doi.org/10.1137/13090849X
8. M. Sun, J. Liu, A proximal Peaceman-Rachford splitting method for compressive sensing, *J. Appl. Math. Comput.*, **50** (2016), 349–363. https://doi.org/10.1007/s12190-015-0874-x
9. B. S. He, F. Ma, X. M. Yuan, Convergence study on the symmetric version of ADMM with larger step sizes, *SIAM J. Imaging Sci.*, **9** (2016), 1467–1501. https://doi.org/10.1137/15M1044448
10. Y. H. Xiao, H. Zhu, A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing, *J. Math. Anal. Appl.*, **405** (2013), 310–319. https://doi.org/10.1016/j.jmaa.2013.04.017
11. M. Sun, M. Y. Tian, A class of derivative-free CG projection methods for nonsmooth equations with an application to the LASSO problem, *Bull. Iran. Math. Soc.*, **46** (2020), 183–205. https://doi.org/10.1007/s41980-019-00250-2
12. H. C. Sun, M. Sun, B. H. Zhang, An inverse matrix-Free proximal point algorithm for compressive sensing, *ScienceAsia*, **44** (2018), 311–318. https://doi.org/10.2306/scienceasia1513-1874.2018.44.311
13. D. X. Feng, X. Y. Wang, A linearly convergent algorithm for sparse signal reconstruction, *J. Fixed Point Theory Appl.*, **20** (2018), 154. https://doi.org/10.1007/s11784-018-0635-1
14. W. T. Yin, S. Osher, D. Goldfarb, J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, *SIAM J. Imaging Sci.*, **1** (2008), 143–168. https://doi.org/10.1137/070703983

15. Y. H. Xiao, Q. Y. Wang, Q. J. Hu, Non-smooth equations based method for ℓ_1-norm problems with applications to compressed sensing, *Nonlinear Anal.-Theor.*, **74** (2011), 3570–3577. https://doi.org/10.1016/j.na.2011.02.040

16. H. C. Sun, Y. J. Wang, L. Q. Qi, Global error bound for the generalized linear complementarity problem over a polyhedral cone, *J. Optim. Theory Appl.*, **142** (2009), 417–429. https://doi.org/10.1007/s10957-009-9509-4

17. H. C. Sun, Y. J. Wang, Further discussion on the error bound for generalized linear complementarity problem over a polyhedral cone, *J. Optim. Theory Appl.*, **159** (2013), 93–107. https://doi.org/10.1007/s10957-013-0290-z

18. H. C. Sun, Y. J. Wang, S. J. Li, M. Sun, A sharper global error bound for the generalized linear complementarity problem over a polyhedral cone under weaker conditions, *J. Fixed Point Theory Appl.*, **20** (2018), 75. https://doi.org/10.1007/s11784-018-0556-z

19. E. J. Candès, Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements, *IEEE T. Inform. Theory*, **57** (2011), 2342–2359. https://doi.org/10.1109/TIT.2011.2111771

20. W. D. Wang, F. Zhang, J. J. Wang, Low-rank matrix recovery via regularized nuclear norm minimization, *Appl. Comput. Harmon Anal.*, **54** (2021), 1–19. https://doi.org/10.1016/j.acha.2021.03.001

AIMS Press