A semi-proximal-based strictly contractive Peaceman-Rachford splitting method*

Yan Gu†, Bo Jiang†‡ and Deren Han†§

June 7, 2015

Abstract

The Peaceman-Rachford splitting method is very efficient for minimizing sum of two functions each depends on its variable, and the constraint is a linear equality. However, its convergence was not guaranteed without extra requirements. Very recently, He *et al.* (SIAM J. Optim. 24: 1011 - 1040, 2014) proved the convergence of a strictly contractive Peaceman-Rachford splitting method by employing a suitable underdetermined relaxation factor. In this paper, we further extend the so-called strictly contractive Peaceman-Rachford splitting method by using two different relaxation factors, and to make the method more flexible, we introduce semi-proximal terms to the subproblems. We characterize the relation of these two factors, and show that one factor is always underdetermined while the other one is allowed to be larger than 1. Such a flexible conditions makes it possible to cover the Glowinski's ADMM whith larger stepsize. We show that the proposed modified strictly contractive Peaceman-Rachford splitting method is convergent and also prove $O(1/t)$ convergence rate in ergodic and nonergodic sense, respectively. The numerical tests on an extensive collection of problems demonstrate the efficiency of the proposed method.

Key words: Semi-proximal, strictly contractive, Peaceman-Rachford splitting method, convex minimization, convergence rate.

1 Introduction

We consider the convex minimization problem with linear constraints and a separable objective function

$$\min \theta_1(x) + \theta_2(y), \quad \text{s.t.} \quad Ax + By = b,\ x \in \mathcal{X},\ y \in \mathcal{Y}, \tag{1.1}$$

where $\theta_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ and $\theta_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ are continuous closed convex (could be nonsmooth) functions; $A \in \mathbb{R}^{m\times n_1}$ and $B \in \mathbb{R}^{m\times n_2}$ are given matrices; $b \in \mathbb{R}^m$ is a given vector; $\mathcal{X}$ and $\mathcal{Y}$ are nonempty closed convex subsets of $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively. Throughout, the solution set of (1.1) is assumed to be nonempty; and $\mathcal{X}$ and $\mathcal{Y}$ are assumed to be simple in the sense that it is easy to compute the projections under the Euclidean norm onto them (e.g., positive orthant, spheroidal or box areas).

*This research is supported by a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

†School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing 210023, P.R. China.

‡Email: jiangbo@njnu.edu.cn.

§Email: handeren@njnu.edu.cn.

Let $\mathcal{L}_\beta(x, y, \lambda)$ be the augmented Lagrangian function for (1.1) that defined by

$$\mathcal{L}_\beta(x, y, \lambda) := \theta_1(x) + \theta_2(y) - \langle \lambda, Ax + By - b \rangle + \frac{\beta}{2}\|Ax + By - b\|^2, \tag{1.2}$$

in which $\lambda \in \mathbb{R}^m$ is the multiplier associated to the linear constraint and $\beta > 0$ is a penalty parameter. Based on the classic Douglas-Rachford operator splitting method [4], the alternating direction method of multipliers was proposed by Gabay and Mercier [8], Glowinski and Marrocco [10] in the mid-1970s, which generates the iterative sequence via the following recursion:

$$\begin{cases} x^{k+1} = \arg\min_{x \in \mathcal{X}} \mathcal{L}_\beta(x, y^k, \lambda^k), & (1.3a) \\ y^{k+1} = \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, \lambda^k), & (1.3b) \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b). & (1.3c) \end{cases}$$

Based on another classic operator splitting method, i.e., Peaceman-Rachford operator splitting method [18], one can derive the following method for (1.1):

$$\begin{cases} x^{k+1} = \arg\min_{x \in \mathcal{X}} \mathcal{L}_\beta(x, y^k, \lambda^k), & (1.4a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta(Ax^{k+1} + By^{k+1} - b), & (1.4b) \\ y^{k+1} = \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}), & (1.4c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} + By^{k+1} - b). & (1.4d) \end{cases}$$

While the global convergence of the alternating direction method of multipliers (1.3a)-(1.3c) can be established under very mild conditions [1], the convergence of the Peaceman-Rachford-based method (1.4a)-(1.4d) can not be guaranteed without further conditions [2]. Most recently, He et al. [12] propose a modification of (1.4a)-(1.4d) by introducing a parameter α to the update scheme of the dual variable λ in (1.4b) and (1.4d), yielding the following procedure:

$$\begin{cases} x^{k+1} = \arg\min_{x \in \mathcal{X}} \mathcal{L}_\beta(x, y^k, \lambda^k), & (1.5a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^{k+1} - b), & (1.5b) \\ y^{k+1} = \arg\min_{y \in \mathcal{Y}} \mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}), & (1.5c) \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \alpha\beta(Ax^{k+1} + By^{k+1} - b). & (1.5d) \end{cases}$$

Note that when $\alpha = 1$, (1.5a)-(1.5d) is exactly the same as (1.4a)-(1.4d). They explained the nonconvergence behavior of (1.4a)-(1.4d) from the contract perspective, i.e., the distance from the iterative point to the solution set is merely nonexpansive, but not contractive. The parameter α in (1.5a)-(1.5d) plays the essential role in forcing the strict contractiveness of the generated sequence. Under the condition that $\alpha \in (0, 1)$, they proved the same sublinear convergence rate as that for ADMM [14]. Particularly, they showed that (1.5a)-(1.5d) achieves an approximate solution of (1.1) with the accuracy of $O(1/t)$ after t iterations[1], both in the ergodic sense and the nonergodic sense.

Note that the parameter α plays different roles in (1.4b) and (1.4d): the former only affects the update of the variable y in (1.3b) while the latter is for the update of the dual variable λ. Hence, it is natural to choose different parameters in these two equalities. In this paper, we give such a scheme by introducing a new parameter γ in (1.4d), i.e., the dual variable is updated by the following manner:

$$\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \gamma\beta(Ax^{k+1} + By^{k+1} - b). \tag{1.6}$$

[1]As the work [16, 17] and many others, a worst-case $O(1/t)$ convergence rate means the accuracy to a solution under certain criteria is of the order $O(1/t)$ after t iterations of an iterative scheme; or equivalently, it requires at most $O(1/\epsilon)$ iterations to achieve an approximate solution with an accuracy of ϵ.

For convenience, we first introduce the whole update scheme of the *modified strictly contractive semi-proximal Peaceman-Rachford splitting method* (sP-PRSM) as

$$
\begin{cases}
x^{k+1} = \arg\min_{x\in\mathcal{X}} \ \mathcal{L}_\beta(x, y^k, \lambda^k) + \frac{1}{2}\|x - x^k\|_S^2, & \text{(1.7a)}\\
\lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(Ax^{k+1} + By^k - b), & \text{(1.7b)}\\
y^{k+1} = \arg\min_{y\in\mathcal{Y}} \ \mathcal{L}_\beta(x^{k+1}, y, \lambda^{k+\frac{1}{2}}) + \frac{1}{2}\|y - y^k\|_T^2, & \text{(1.7c)}\\
\lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \gamma\beta(Ax^{k+1} + By^{k+1} - b) & \text{(1.7d)}
\end{cases}
$$

where S and T are two positive semi-definite matrices. In applications, by choosing different matrices S and T customizing the problems' structures, we can obtain different efficient methods.

Our main contributions are twofold.

1. Motivated by the nice analysis techniques in [12] and [19], we proved that the sequence generated by sP-PRSM is strictly contractive and thus convergent, under the requirement that

$$
\alpha \in [0,\, 1), \quad \gamma \in \left(0,\, \frac{1-\alpha+\sqrt{(1+\alpha)^2+4(1-\alpha^2)}}{2}\right). \tag{1.8}
$$

   Moreover, we proved that sP-PRSM is $O(1/t)$ sublinearly convergent both in the ergodic and nonergodic sense. Note that the nonergodic convergence rate requires that $\gamma \in (0, 1]$. We remark that the convergence of sP-PRSM (1.7) can unify that of several existing splitting methods.

   - Choosing $\alpha = 0, \gamma \in (0, (1+\sqrt{5})/2)$ and $T = 0$, $S = 0$, sP-PRSM (1.7) reduces to the classical ADMM while the convergence coincides with that of ADMM [9];
   - Setting $\alpha = 0, \gamma \in (0, (1+\sqrt{5})/2)$ and $T \succeq 0, S \succeq 0$, sP-PRSM (1.7) covers the semi-proximal ADMM considered in [3, 7, 14, 19] and the corresponding convergence results;
   - Setting $\alpha = \gamma \in (0, 1)$, and $T = 0$, $S = 0$, sP-PRSM (1.7) reduces to the strictly contractive PRSM proposed in [12] and the convergence of the two methods is identical.

2. We added a proximal term to each of the two main subproblems in updating the x and the y variables. In fact, for ADMM, Eckstein [5] and He et al. [11] have already considered to add proximal terms to the subproblems for different purpose. Recently, Fazel et al. [7] proposed to allow S and T to be positive semi-definite, in contrast to the positive definite requirements in the classical algorithms, which makes the algorithm more flexible.

The rest of this paper is organized as follows. In Sect. 2, we give the optimality condition of (1.1) by using the variational inequality and also list some assertions which will be used in later analysis. In Sect. 3, we first give the contraction analysis of sP-PRSM (1.7), and then establish the global convergence. We discuss the sublinear and linear convergence rate in Sects. 4 and 5, respectively. In Sect. 6, we test a variety of problems to show the efficiency of the proposed sP-PRSM (1.7). Finally, we make some conclusions in Sect. 7.

On May 25, after attending the first author's thesis defense, Prof. He told us that they had also considered to use different stepsizes α and γ in (1.7) with $T = 0$ and $S = 0$, and we got their manuscript on May 26 [13]. We found that they established the convergence when

$$
\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right), \ \alpha \in (-1, 1), \ \alpha + \gamma > 0, \ |\alpha| < 1 + \gamma - \gamma^2. \tag{1.9}
$$

Notice that (1.9) covers the case when $\alpha < 0$. If we restrict $\alpha \geq 0$ in (1.9), then the corresponding relation will be

$$\alpha \in [0,1), \quad \gamma \in \left(0, \frac{1+\sqrt{5-4\alpha}}{2}\right). \tag{1.10}$$

By some simple calculations, it is easy to see that the domain defined by (1.10) is a bit smaller than that defined by (1.8). By some private communications with Prof. He, we learned that they also discovered the formula (1.8). However, to provide the intuitive understanding and the unified convergence analysis, they presented (1.9) in [13] which connects α and $1-\gamma+\gamma^2$ directly.

2 Preliminaries

In this section, we give the optimality condition of (1.1) and some notations or relations which will be frequently used in our analysis. Let $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$. Throughout this paper, we make the following assumption.

Assumption 2.1. *Let $\Omega^* \subset \Omega$ be the set whose elements are the optimal solutions of (1.1) and the associating dual solutions of (1.1). Throughout the paper, we assume that Ω^* is non-empty.*

2.1 Optimality condition of (1.1)

Owing to the convexity of $\theta_1(\cdot)$ and $\theta(\cdot)$, there exist two positive semidefinite matrices Σ_1 and Σ_2 such that for any $x, x' \in \mathbb{R}^{n_1}$

$$\langle x - x', \xi_x - \xi_x' \rangle \geq \|x - x'\|_{\Sigma_1}^2, \tag{2.1}$$

where $\xi_x \in \partial\theta_1(x)$, $\xi_x' \in \partial\theta_1(x')$, and for any $y, y' \in \mathbb{R}^{n_2}$

$$\langle y - y', \xi_y - \xi_y' \rangle \geq \|y - y'\|_{\Sigma_2}^2, \tag{2.2}$$

where $\xi_y \in \partial\theta_2(y)$, $\xi_y' \in \partial\theta_1(y')$.

Denote

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \; v = \begin{pmatrix} y \\ \lambda \end{pmatrix}, \; w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \text{ and } F(w) = \begin{pmatrix} \xi_x - A^\mathsf{T}\lambda \\ \xi_y - B^\mathsf{T}\lambda \\ Ax + By - b \end{pmatrix}.$$

Due to the convexity of $\theta_1(\cdot)$ and $\theta_2(\cdot)$, it is easy to show that the operator $F(\cdot)$ is monotonic. Specifically, for any $w, w' \in \Omega$, we have

$$\langle w - w', F(w) - F(w') \rangle = \left\langle \begin{pmatrix} x - x' \\ y - y' \end{pmatrix}, \begin{pmatrix} \xi_x - \xi_x' \\ \xi_y - \xi_y' \end{pmatrix} \right\rangle \geq \|u - u'\|_\Sigma^2, \tag{2.3}$$

where $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$ and the inequality is due to (2.1) and (2.2). As shown in [12], we say that $w^* \in \Omega^*$ if there holds that

$$\langle w - w^*, F(w^*) \rangle \geq 0, \quad \forall w \in \Omega. \tag{2.4}$$

From Theorem 2.3.5 in [6] or Theorem 2.1 in [14], we can see that Ω^* is closed and convex, and it can be reformulated as

$$\Omega^* = \bigcap_{w \in \Omega} \{\tilde{w} \in \Omega \colon \langle w - \tilde{w}, F(w) \rangle \geq 0\}. \tag{2.5}$$

Let $\mathcal{S}$ denote the feasible set of (1.1), namely,

$$\mathcal{S} = \{(x, y) : Ax + By = b, x \in \mathcal{X}, y \in \mathcal{Y}\}$$

and denote $\mathcal{D} = \mathcal{S} \times \mathbb{R}^m$. Let us restrict $w \in \Omega$ in (2.4) and (2.5) to be $w \in \mathcal{D} \subset \Omega$. By some suitable modification of the proof of Theorem 2.1 in [14], it is easy to show that the optimal solution set Ω^* can be characterized as

$$\Omega^* = \bigcap_{w \in \mathcal{D}} \{\tilde{w} \in \Omega \colon \ \langle w - \tilde{w}, F(w)\rangle \geq 0\}. \tag{2.6}$$

Similarly, we say that $w^* \in \Omega^*$ if there holds that

$$\langle w - w^*, F(w^*)\rangle \geq 0, \quad \forall w \in \mathcal{D}. \tag{2.7}$$

Note that the optimality condition (2.7) will be frequently used in the convergence analysis. Owing to (2.7), we call $\tilde{w}$ to be an ϵ solution of (1.1) if

$$\sup_{w \in \mathcal{D}} \langle \tilde{w} - w, F(w)\rangle \leq \epsilon. \tag{2.8}$$

2.2 Some notations

Given two real matrices C and D of the same dimension, we define $\langle C, D\rangle = \mathrm{tr}(C^\mathsf{T} D)$, where $\mathrm{tr}(\cdot)$ is the trace operator. We use $\|\cdot\|$ to denote the 2-norm of a vector. We denote $\|z\|_G^2 = z^\mathsf{T} G z$ for $z \in \mathbb{R}^n$ and $G \in \mathbb{R}^{n\times n}$. For a real symmetric matrix S, we mark $S \succeq 0$ ($S \succ 0$) if S is positive semidefinite (positive definite). To make the analysis more elegant, we use $r^k = Ax^k + By^k - b$ for short. Similarly, for any $w \in \mathcal{D}$, we denote $r(w) = Ax + By - b$. Obviously, there holds that $r(w) = 0$ for any $w \in \mathcal{D}$. For ease of the analysis, we define the following matrices as

$$H = \frac{1}{\alpha+\gamma}\begin{pmatrix} (\alpha+\gamma-\alpha\gamma)\beta B^\mathsf{T} B & -\alpha B^\mathsf{T} \\ -\alpha B & \frac{1}{\beta} I_m \end{pmatrix} \tag{2.9}$$

and

$$M = \begin{pmatrix} I_{n_2} & 0 \\ \alpha\beta B & (\alpha+\gamma)\beta I_m \end{pmatrix}. \tag{2.10}$$

We can easily verify that H is positive semidefinite if $0 \leq \alpha \leq 1$ and $\gamma > 0$, and

$$M^\mathsf{T} H M = \begin{pmatrix} (1-\alpha)\beta B^\mathsf{T} B & 0 \\ 0 & (\alpha+\gamma)\beta I_m \end{pmatrix}. \tag{2.11}$$

Denote $P = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix}$ and define

$$G := \begin{pmatrix} P & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & H \end{pmatrix} = \begin{pmatrix} S & 0 & 0 \\ 0 & T + \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\mathsf{T} B & -\frac{\alpha}{\alpha+\gamma} B^\mathsf{T} \\ 0 & -\frac{\alpha}{\alpha+\gamma} B & \frac{1}{(\alpha+\gamma)\beta} I_m \end{pmatrix} \tag{2.12}$$

and

$$\widehat{G} := \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + G = \begin{pmatrix} S + \Sigma_1 & 0 & 0 \\ 0 & T + \Sigma_2 + \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\mathsf{T} B & -\frac{\alpha}{\alpha+\gamma} B^\mathsf{T} \\ 0 & -\frac{\alpha}{\alpha+\gamma} B & \frac{1}{(\alpha+\gamma)\beta} I_m \end{pmatrix}, \tag{2.13}$$

where the dimension of zero matrix can be easily identified from the context. For any $w, w' \in \Omega$, there hold that

$$\|w - w'\|_G^2 = \|u - u'\|_P^2 + \|v - v'\|_H^2 \tag{2.14}$$

and

$$\|w - w'\|_{\widehat{G}}^2 = \|u - u'\|_\Sigma^2 + \|w - w'\|_G^2. \tag{2.15}$$

Finally, we present some relations of the iterates. They will play a crucial role in the convergence analysis. With the update scheme (1.7b) and (1.7d), it is easy to have

$$\lambda^k = \lambda^{k+1} + (\alpha + \gamma)\beta r^{k+1} + \alpha\beta B(y^k - y^{k+1}). \tag{2.16}$$

Recalling the definition of v, with (2.16) and (2.10), we conclude that

$$v^k - v^{k+1} = \begin{pmatrix} y^k - y^{k+1} \\ \lambda^k - \lambda^{k+1} \end{pmatrix} = M \begin{pmatrix} y^k - y^{k+1} \\ r^{k+1} \end{pmatrix}. \tag{2.17}$$

With (2.11) and (2.17), we thus have

$$\|v^k - v^{k+1}\|_H^2 = (1 - \alpha)\beta\|B(y^k - y^{k+1})\|^2 + (\alpha + \gamma)\beta\|r^{k+1}\|^2. \tag{2.18}$$

3 Convergence of sP-PRSM

In this section, we first show that the sequence $\{w_k\}$ generated by sP-PRSM (1.7) is strictly contractive and thus can establish the convergence of the method. With the help of the contraction property, we further discuss the convergence rate in the ergodic and nonergodic sense.

3.1 Contraction analysis

We first give the contraction result as follows, and then complete its proof at the end of this section. Note that the following result will play a key role in proving the convergence of (1.7).

Theorem 3.1. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). If we choose α and γ according to (1.8), then for any $w^* \in W^*$, there holds that*

$$\begin{aligned} &\left(\|w^k - w^*\|_{\widehat{G}}^2 + \rho_{\alpha,\gamma}\|r^k\|^2 + \eta_{\alpha,\gamma}\|y^k - y^{k-1}\|_T^2\right) \\ -&\left(\|w^{k+1} - w^*\|_{\widehat{G}}^2 + \rho_{\alpha,\gamma}\|r^{k+1}\|^2 + \eta_{\alpha,\gamma}\|y^{k+1} - y^k\|_T^2\right) \geq \widehat{\tau}_{\alpha,\gamma}\|w^k - w^{k+1}\|_{\widehat{G}}^2, \end{aligned} \tag{3.1}$$

where $\widehat{\tau}_{\alpha,\gamma} = \min(\frac{1}{2}, \tau_{\alpha,\gamma})$, and $\rho_{\alpha,\gamma}, \eta_{\alpha,\gamma} \geq 0$ and $0 < \tau_{\alpha,\gamma} < 1$ are some constants which only depend on α and γ.

To prove Theorem 3.1, several lemmas should be established first. The following Lemma 3.1 is mainly based on the optimality conditions of (1.7a) and (1.7b).

Lemma 3.1. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). If $\alpha \geq 0$ and $\gamma > 0$, then for any $w \in \mathcal{D}$, there holds that*

$$\begin{aligned} (w - w^{k+1})^\mathsf{T} G(w^{k+1} - w^k) \geq\; &(1 - \alpha - \gamma)\beta \left\|r^{k+1}\right\|^2 + (1 - \alpha)\beta \left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle \\ &+ \left\langle w^{k+1} - w, F(w^{k+1})\right\rangle. \end{aligned} \tag{3.2}$$

Proof. Consider any $\xi_x^{k+1} \in \partial\theta_1(x^{k+1})$ and $\xi_y^{k+1} \in \partial\theta_2(y^{k+1})$. We then obtain respectively from the optimality conditions for (1.7a) and (1.7c) that

$$\left\langle x - x^{k+1}, S(x^{k+1} - x^k) + \xi_x^{k+1} - A^\top \lambda^k + \beta A^\top r^{k+1} + \beta A^\top B(y^k - y^{k+1}) \right\rangle \geq 0, \quad \forall x \in \mathcal{X}$$

and

$$\left\langle y - y^{k+1}, T(y^{k+1} - y^k) + \xi_y^{k+1} - B^\top \lambda^{k+\frac{1}{2}} + \beta B^\top r^{k+1} \right\rangle \geq 0, \quad \forall y \in \mathcal{Y}.$$

Substituting (2.16) and $\lambda^{k+\frac{1}{2}} = \lambda^{k+1} + \gamma\beta r^{k+1}$ which follows from (1.7d) into the above two inequalities, respectively, we obtain the new formulations of the optimality conditions for (1.7a) and (1.7c) as

$$\left\langle x - x^{k+1}, S(x^{k+1} - x^k) + \xi_x^{k+1} - A^\top \lambda^{k+1} + (1 - \alpha - \gamma)\beta A^\top r^{k+1} + (1 - \alpha)\beta A^\top B(y^k - y^{k+1}) \right\rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{3.3}$$

and

$$\left\langle y - y^{k+1}, T(y^{k+1} - y^k) + \xi_y^{k+1} - B^\top \lambda^{k+1} + (1 - \gamma)\beta B^\top r^{k+1} \right\rangle \geq 0, \quad \forall y \in \mathcal{Y}. \tag{3.4}$$

Rewriting (2.16) to be

$$r^{k+1} - \frac{\alpha}{\alpha + \gamma} B(y^{k+1} - y^k) + \frac{1}{(\alpha + \gamma)\beta}(\lambda^{k+1} - \lambda^k) = 0. \tag{3.5}$$

Combing (3.3), (3.4) and (3.5) in a suitable way, and recalling the definition of $F(w)$, we see that for any $w \in \Omega$ there holds that

$$\begin{aligned} &\left\langle w - w^{k+1}, \begin{pmatrix} S(x^{k+1} - x^k) \\ T(y^{k+1} - y^k) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \alpha\beta B^\top r^{k+1} + (1 - \alpha)\beta B^\top B(y^{k+1} - y^k) \\ -\frac{\alpha}{\alpha+\gamma} B(y^{k+1} - y^k) + \frac{1}{(\alpha+\gamma)\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} \right\rangle \\ \geq &\left\langle w^{k+1} - w, \begin{pmatrix} A^\top \\ B^\top \\ 0 \end{pmatrix} \left[(1 - \alpha - \gamma)\beta r^{k+1} + (1 - \alpha)\beta B(y^k - y^{k+1})\right] \right\rangle + \left\langle w^{k+1} - w, F(w^{k+1}) \right\rangle. \end{aligned} \tag{3.6}$$

With the relation (3.5) and the definition (2.9) of H, by some easy calculations, we obtain that

$$\begin{pmatrix} \alpha\beta B^\top r^{k+1} + (1 - \alpha)\beta B^\top B(y^{k+1} - y^k) \\ -\frac{\alpha}{\alpha+\gamma} B(y^{k+1} - y^k) + \frac{1}{(\alpha+\gamma)\beta}(\lambda^{k+1} - \lambda^k) \end{pmatrix} = H(v^{k+1} - v^k). \tag{3.7}$$

Considering the definition of r^{k+1} and $r(w)$, we have $(A,\, B,\, 0)(w^{k+1} - w) = r^{k+1} - r(w)$. This, together with (3.6) and (3.7) indicates that

$$\begin{aligned} (w - w^{k+1})^\top G(w^{k+1} - w^k) \geq &\left\langle r^{k+1} - r(w), (1 - \alpha - \gamma)\beta r^{k+1} + (1 - \alpha)\beta B(y^k - y^{k+1}) \right\rangle \\ &+ \left\langle w^{k+1} - w, F(w^{k+1}) \right\rangle. \end{aligned} \tag{3.8}$$

Notice that $r(w) = 0$ for any $w \in \mathcal{D}$, we can immediately obtain (3.2) from (3.8) since The proof is completed. □

Lemma 3.2. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). If $\alpha \geq 0$ and $\gamma > 0$, then for any $w \in \mathcal{D}$, there holds that*

$$\begin{aligned} \|w^k - w\|_G^2 - \|w^{k+1} - w\|_G^2 \geq &\|u^k - u^{k+1}\|_P^2 + (1 - \alpha)\beta\|B(y^k - y^{k+1})\|^2 + (2 - \alpha - \gamma)\beta \left\|r^{k+1}\right\|^2 \\ &+ 2(1 - \alpha)\beta \left\langle r^{k+1}, B(y^k - y^{k+1}) \right\rangle + 2\left\langle w^{k+1} - w, F(w^{k+1}) \right\rangle. \end{aligned} \tag{3.9}$$

Proof. Using the identity

$$\|a\|_G^2 - \|b\|_G^2 = \|a-b\|_G^2 + 2b^\mathsf{T} G(a-b),$$

with $a = w - w^k$ and $b = w - w^{k+1}$, we have

$$\|w^k - w\|_G^2 - \|w^{k+1} - w\|_G^2 = \|w^k - w^{k+1}\|_G^2 + 2(w - w^{k+1})^\mathsf{T} G(w^{k+1} - w^k). \tag{3.10}$$

Plugging (3.2) into (3.10) with (2.14) and (2.18), we have (3.9). The proof is completed. □

Denote $K = \begin{pmatrix} (1-\alpha)\beta B^\mathsf{T} B & (1-\alpha)\beta B^\mathsf{T} \\ (1-\alpha)\beta B & (2-\alpha-\gamma)\beta I_m \end{pmatrix}$. With (2.10) and (2.17) , the inequality (3.9) can be expressed as

$$\begin{aligned} &\|w^k - w\|_G^2 - \|w^{k+1} - w\|_G^2 \\ \geq\ &\|u^k - u^{k+1}\|_P^2 + (v^k - v^{k+1})^\mathsf{T} M^{-\mathsf{T}} K M^{-1}(v^k - v^{k+1}) + 2\left\langle w^{k+1} - w, F(w^{k+1})\right\rangle. \end{aligned} \tag{3.11}$$

Our task now is to estimate the first term in the righthand side of the inequality in (3.11).

Consider the case where $0 \leq \alpha < 1$ and $0 < \gamma < 1$. It is easy to see that $K \succ 0$. Moreover, by some tedious calculations, we conclude that

$$K \geq c \cdot M^\mathsf{T} H M,$$

where $c = \frac{1-\sqrt{1-(\alpha+\gamma)(1-\gamma)}}{\alpha+\gamma} \in (0,1)$. Thus we obtain from (3.11) and (2.14) that

$$\|w^k - w\|_G^2 - \|w^{k+1} - w\|_G^2 \geq c\|w^k - w^{k+1}\|_G^2 + 2\left\langle w^{k+1} - w, F(w^{k+1})\right\rangle, \tag{3.12}$$

which is sufficient to establish that the sequence $\{\|w^k - w^*\|_G^2\}$ is strictly contractive by choosing $w = w^*$ in (3.12) and further to show the convergence of sP-PRSM (1.7).

Consider the case when $0 < \alpha < 1$ but $\gamma \geq 1$. The matrix K is not positive semidefinite any more. Instead of establishing the strictly contractive property of the sequence $\{\|w^k - w^*\|_G^2\}$, we consider to prove the property of another sequence related to the sequence $\{\|w^k - w^*\|_G^2\}$. To do so, we need to give a more careful estimation of the intersection term $\left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle$.

Lemma 3.3. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). If $\alpha \geq 0$ and $\gamma > 0$, then for any $w \in \Omega$, there holds that*

$$\begin{aligned} \left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle \geq\ &\frac{1-\gamma}{1+\alpha}\left\langle r^k, B(y^k - y^{k+1})\right\rangle - \frac{\alpha}{1+\alpha}\|B(y^k - y^{k+1})\|^2 \\ &+ \frac{1}{2(1+\alpha)} \cdot \frac{1}{\beta}\left(\|y^{k+1} - y^k\|_T^2 - \|y^k - y^{k-1}\|_T^2\right). \end{aligned} \tag{3.13}$$

Proof. Note that the optimality condition (3.4) also holds with $k := k-1$, then we have

$$\left\langle y - y^k, T(y^k - y^{k-1}) + \xi_y^k - B^\mathsf{T}\lambda^k + (1-\gamma)\beta B^\mathsf{T} r^k\right\rangle \geq 0, \quad \forall y \in \mathcal{Y}, \tag{3.14}$$

where $\xi_y^k \in \partial\theta_2(y^k)$. Choosing y to be y^k and y^{k+1} in (3.4) and (3.14) and then rearranging the obtained inequalities, respectively, we have that

$$\left\langle B(y^k - y^{k+1}), -\lambda^{k+1} + (1-\gamma)\beta r^{k+1}\right\rangle \geq \|y^{k+1} - y^k\|_T^2 + \left\langle y^{k+1} - y^k, \xi_y^{k+1}\right\rangle \tag{3.15}$$

and

$$\left\langle B(y^k - y^{k+1}), \lambda^k - (1-\gamma)\beta r^k\right\rangle \geq -\left\langle y^{k+1} - y^k, T(y^k - y^{k-1})\right\rangle - \left\langle y^{k+1} - y^k, \xi_y^k\right\rangle. \tag{3.16}$$

Summing (3.15) and (3.16) over the both sides and then using the relation (2.16), we obtain that

$$
\begin{aligned}
&(1+\alpha)\beta\left\langle B(y^k-y^{k+1}),r^{k+1}\right\rangle+\alpha\beta\|B(y^k-y^{k+1})\|^2-(1-\gamma)\beta\left\langle B(y^k-y^{k+1}),r^k\right\rangle\\
\geq\ &\|y^{k+1}-y^k\|_T^2+\left\langle y^{k+1}-y^k,\xi_y^{k+1}-\xi_y^k\right\rangle-\left\langle y^{k+1}-y^k,T(y^k-y^{k-1})\right\rangle\\
\geq\ &\frac{1}{2}(\|y^{k+1}-y^k\|_T^2-\|y^k-y^{k-1}\|_T^2),
\end{aligned}
\tag{3.17}
$$

where the second inequality is due to (2.2) and the Cauchy-Schwarz inequality. It is trivial to obtain (3.13) from (3.17). The proof is completed. □

With the help of Lemma 3.3, we are ready to have the following theorem.

Theorem 3.2. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). If we choose α and γ according to (1.8) , then for any $w\in\mathcal{D}$, there holds that*

$$
\begin{aligned}
&\left(\|w^k-w\|_G^2+\rho_{\alpha,\gamma}\|r^k\|^2+\eta_{\alpha,\gamma}\|y^k-y^{k-1}\|_T^2\right)-\left(\|w^{k+1}-w\|_G^2+\rho_{\alpha,\gamma}\|r^{k+1}\|^2+\eta_{\alpha,\gamma}\|y^{k+1}-y^k\|_T^2\right)\\
\geq\ &\tau_{\alpha,\gamma}\|w^k-w^{k+1}\|_G^2+2(w^{k+1}-w)^{\mathsf{T}}F(w^{k+1}),
\end{aligned}
\tag{3.18}
$$

where $\rho_{\alpha,\gamma},\eta_{\alpha,\gamma}\geq 0$ and $0<\tau_{\alpha,\gamma}<1$ are some constants which only depend on α and γ.

Proof. We consider three cases.

I). $0<\gamma<1$. With $\rho_{\alpha,\gamma}=\eta_{\alpha,\gamma}=0$ and $\tau_{\alpha,\gamma}=\frac{1-\sqrt{1-(\alpha+\gamma)(1-\gamma)}}{\alpha+\gamma}$, the inequality (3.18) coincides with (3.12) which has been proved.

Before we proceed, by combining (3.13) and (3.9), we derive the useful inequality as

$$
\begin{aligned}
&\left(\|w^k-w\|_G^2+\frac{1-\alpha}{1+\alpha}\|y^{k+1}-y^k\|_T^2\right)-\left(\|w^{k+1}-w\|_G^2+\frac{1-\alpha}{1+\alpha}\|y^k-y^{k-1}\|_T^2\right)\\
\geq\ &\|u^k-u^{k+1}\|_P^2+\frac{(1-\alpha)^2}{1+\alpha}\beta\|B(y^k-y^{k+1})\|^2+(2-\alpha-\gamma)\beta\left\|r^{k+1}\right\|^2\\
&+2(1-\gamma)\frac{1-\alpha}{1+\alpha}\beta\left\langle r^k,B(y^k-y^{k+1})\right\rangle+2\left\langle w^{k+1}-w,F(w^{k+1})\right\rangle.
\end{aligned}
\tag{3.19}
$$

II). $\gamma=1$. It is not difficult to see from (3.19), (2.14) and(2.18) that (3.18) holds with $\rho_{\alpha,\gamma}=0$ and $\eta_{\alpha,\gamma}=\tau_{\alpha,\gamma}=\frac{1-\alpha}{1+\alpha}$.

III). $1<\gamma<\frac{1-\alpha+\sqrt{(1+\alpha)^2+4(1-\alpha^2)}}{2}$. Let us choose any $\delta\in\left(\frac{\gamma-1}{1-\alpha},\frac{1+\alpha}{\gamma-1}-\frac{1+\alpha}{1-\alpha}\right)$. Note that this interval is well-defined due to the range of γ. With the Cauchy-Schwarz inequality, we know that

$$
-2\left\langle r^k,B(y^k-y^{k+1})\right\rangle\geq-\delta\cdot\|r^k\|^2-\frac{1}{\delta}\cdot\|B(y^k-y^{k+1})\|^2.
$$

Plugging the above inequality into (3.19), we obtain that

$$
\begin{aligned}
&\left(\|w^k-w\|_G^2+\frac{1-\alpha}{1+\alpha}\|y^{k+1}-y^k\|_T^2\right)-\left(\|w^{k+1}-w\|_G^2+\frac{1-\alpha}{1+\alpha}\|y^k-y^{k-1}\|_T^2\right)\\
\geq\ &\|u^k-u^{k+1}\|_P^2+\frac{(1-\alpha)^2}{1+\alpha}\left(1-\frac{\gamma-1}{1-\alpha}\cdot\frac{1}{\delta}\right)\beta\|B(y^k-y^{k+1})\|2+(\gamma-1)\frac{1-\alpha}{1+\alpha}\left(\frac{1+\alpha}{\gamma-1}-\frac{1+\alpha}{1-\alpha}-\delta\right)\left\|r^{k+1}\right\|^2\\
&-\delta(\gamma-1)\frac{1-\alpha}{1+\alpha}\left(\|r^k\|^2-\|r^{k+1}\|^2\right)+2\left\langle w^{k+1}-w,F(w^{k+1})\right\rangle,
\end{aligned}
$$

which with (2.14) and (2.18) implies that (3.18) holds with $\rho_{\alpha,\gamma}=\delta(\gamma-1)\frac{1-\alpha}{1+\alpha}$, $\eta_{\alpha,\gamma}=\frac{1-\alpha}{1+\alpha}$ and

$$
\tau_{\alpha,\gamma}=\frac{1-\alpha}{1+\alpha}\cdot\min\left\{\left(1-\frac{\gamma-1}{1-\alpha}\cdot\frac{1}{\delta}\right),\frac{\gamma-1}{\alpha+\gamma}\left(\frac{1+\alpha}{\gamma-1}-\frac{1+\alpha}{1-\alpha}-\delta\right)\right\}.
\tag{3.20}
$$

The proof is completed. □

Finally, we end this subsection by finishing the proof of Theorem 3.1.

Proof of Theorem 3.1. Consider any $w^* \in \Omega^*$, recalling the monotonicity (2.3) of $F(w)$, we thus obtain from the optimality (2.4) of w^* that

$$\langle w^{k+1} - w^*, F(w^{k+1})\rangle \geq \langle w^{k+1} - w^*, F(w^*)\rangle + \|u^{k+1} - u^*\|_\Sigma^2 \geq \|u^{k+1} - u^*\|_\Sigma^2.$$

With the above statement, choosing $w = w^*$ in (3.18), we can easily obtain (3.1).

$$\begin{aligned}&\left(\|w^k - w^*\|_G^2 + \rho_{\alpha,\gamma}\|r^k\|^2 + \eta_{\alpha,\gamma}\|y^k - y^{k-1}\|_T^2\right) - \left(\|w^{k+1} - w^*\|_G^2 + \rho_{\alpha,\gamma}\|r^{k+1}\|^2 + \eta_{\alpha,\gamma}\|y^{k+1} - y^k\|_T^2\right)\\ &\geq \tau_{\alpha,\gamma}\|w^k - w^{k+1}\|_G^2 + 2\|u^{k+1} - u^*\|_\Sigma^2,\end{aligned} \tag{3.21}$$

Adding the term $\|u^k - u^*\|_\Sigma^2 - \|u^{k+1} - u^*\|_\Sigma^2$ on both sides of (3.21), we have

$$\begin{aligned}&\left(\|u^k - u^*\|_\Sigma^2 + \|w^k - w^*\|_G^2 + \rho_{\alpha,\gamma}\|r^k\|^2 + \eta_{\alpha,\gamma}\|y^k - y^{k-1}\|_T^2\right)\\ &-\left(\|u^{k+1} - u^*\|_\Sigma^2 + \|w^{k+1} - w^*\|_G^2 + \rho_{\alpha,\gamma}\|r^{k+1}\|^2 + \eta_{\alpha,\gamma}\|y^{k+1} - y^k\|_T^2\right)\\ &\geq \tau_{\alpha,\gamma}\|w^k - w^{k+1}\|_G^2 + \|u^{k+1} - u^*\|_\Sigma^2 + \|u^k - u^*\|_\Sigma^2\\ &\geq \tau_{\alpha,\gamma}\|w^k - w^{k+1}\|_G^2 + \frac{1}{2}\|u^k - u^{k+1}\|_\Sigma^2,\end{aligned} \tag{3.22}$$

which with (2.13) and (2.15) leads to (3.1). The proof is completed.

3.2 Global convergence

Theorem 3.3. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). If we choose α and γ according to (1.8) , then $\{w^k\}$ converges to an optimal solution of (1.1).*

Proof. We divide the proof into three steps.

(i) To show that the sequences $\{w^k\}$ is bounded. It is trivial to know from (3.1) that $\lim_{k\to\infty}\|w^k - w^{k+1}\|_{\widehat{G}}^2 = 0$, which with (2.14), (2.15) and (2.18) indicates that

$$\lim_{k\to\infty}\|x^k - x^{k+1}\|_{S+\Sigma_1} = \lim_{k\to\infty}\|y^k - y^{k+1}\|_{T+\Sigma_2} = \lim_{k\to\infty}\|r^k\| = \lim_{k\to\infty}\|B(y^k - y^{k+1})\| = 0. \tag{3.23}$$

By some simple calculations, there holds that there exists a constant $c_0 = \frac{2(\alpha+\gamma)-\alpha\gamma-\alpha\sqrt{\gamma^2+4(\alpha+\gamma)}}{2(\alpha+\gamma)} \in (0, 1]$ such that

$$\widehat{G} \succeq c_0 \begin{pmatrix} S+\Sigma_1 & 0 & 0\\ 0 & T+\Sigma_2+\beta B^\mathsf{T}B & 0\\ 0 & 0 & \frac{1}{\beta}I_m\end{pmatrix}. \tag{3.24}$$

Besides, it is straightforward to see from (3.1) that $\|w^k - w^*\|_{\widehat{G}}^2$ is bounded. This together with (3.24) shows that

$$\|x^k - x^*\|_{S+\Sigma_1}, \quad \|y^k - y^*\|_{T+\Sigma_2+\beta B^\mathsf{T}B}, \quad \|\lambda^k - \lambda^*\|$$

are all bounded. Obviously, we claim that $\{\lambda^k\}$ is bounded. Moreover, we see that $\{y^k\}$ is bounded as $T + \Sigma_2 + \beta B^\mathsf{T}B \succ 0$. Notice that $\|Ax^k - Ax^*\| = \|r^k + B(y^k - y^*)\| \leq \|r^k\| + \|B(y^k - y^*)\|$, with (3.23), this implies that $\|x^k - x^*\|_{\beta A^\mathsf{T}A}$ is bounded. Recalling that $S + \Sigma_1 + \beta A^\mathsf{T}A \succ 0$ and $\|x^k - x^*\|_{S+\Sigma_1}$ is bounded, it is safe to say that $\{x^k\}$ is also bounded.

(ii) To show that any cluster point of the sequence $\{w^k\}$ is an optimal solution of (1.1). Let $\{w^{k_i}\}$ be a subsequence of the sequence $\{w^k\}$ and $\lim\limits_{k_i\to\infty} w^{k_i} = w^\infty$. Since the graphs of $\partial\theta_1(\cdot)$ and $\partial\theta_2(\cdot)$ are both closed, taking the limit with respect $k_i \to \infty$ on both sides of (3.2) and using (3.23), we have that

$$(w - w^\infty)^\mathsf{T}F(w^\infty) \geq 0, \quad \forall w \in \mathcal{D},$$

which means that w^∞ is an optimal solution of (1.1).

(iii) To show that the sequence $\{w^k\}$ has only one cluster point. We first replace x^* with x^∞ in the analysis of Steps (i) and (ii). It follows from $w^{k_i} \to w^\infty$ that $\|w^{k_i} - w^\infty\|^2_{\widehat{G}} + \rho_{\alpha,\gamma}\|r^{k_i}\|^2 + \eta_{\alpha,\gamma}\|y^{k_i} - y^{k_i-1}\|^2_T \to 0$. Owing to the monotonicity of the sequence $\|w^k - w^\infty\|^2_{\widehat{G}} + \rho_{\alpha,\gamma}\|r^k\|^2 + \eta_{\alpha,\gamma}\|y^k - y^{k-1}\|^2_T$, we can see that

$$\lim_{k\to\infty} \|w^k - w^\infty\|^2_{\widehat{G}} + \rho_{\alpha,\gamma}\|r^k\|^2 + \eta_{\alpha,\gamma}\|y^k - y^{k-1}\|^2_T = 0.$$

This together with (3.24) shows that

$$\lim_{k\to\infty} \|x^k - x^\infty\|_{S+\Sigma_1} = \lim_{k\to\infty} \|y^k - y^\infty\|_{T+\Sigma_2+\beta B^\top B} = \lim_{k\to\infty} \quad \|\lambda^k - \lambda^\infty\| = 0. \tag{3.25}$$

Using again the inequality $\|Ax^k - Ax^\infty\| = \|r^k + B(y^k - y^\infty)\| \le \|r^k\| + \|B(y^k - y^\infty)\|$ and $\lim_{k\to\infty}\|r_k\| = 0$, we have

$$\lim_{k\to\infty} \|A(x^k - x^\infty)\| = 0. \tag{3.26}$$

Combing (3.25) and (3.26), and using that $S + \Sigma_1 + \beta A^\top A \succ 0$ and $T + \Sigma_2 + \beta B^\top B \succ 0$, we immediately have

$$\lim_{k\to\infty} w^k = w^\infty.$$

The proof is completed. □

4 Sublinear convergence of sP-PRSM

The rate of convergence of an algorithm can help us have a deeper understanding of the algorithm. Thus, in this section, we establish the sublinear rate of convergence of sP-PRSM, in ergodic sense and nonergodic sense, respectively.

4.1 Convergence rate in the ergodic sense

We now give the sublinear rate of convergence of sP-PRSM in the ergodic sense, which is very easy due to the key inequality (3.18).

Theorem 4.1. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). For any integer $t > 0$, define*

$$\bar{w}_t = \frac{1}{t}\sum_{k=1}^{t} w^{k+1}.$$

Then for any $w \in \mathcal{D}$, we have that

$$\sup_{w\in\Omega} \langle \bar{w}_t - w, F(w)\rangle \le \frac{1}{2t}\left(\|w^1 - w^0\|^2_G + \rho_{\alpha,\gamma}\|r^1\|^2 + \eta_{\alpha,\gamma}\|y^1 - y^0\|^2_T\right). \tag{4.1}$$

Proof. It follows from (3.18) and $(w^{k+1} - w)^\top F(w^{k+1}) \ge (w^{k+1} - w)^\top F(w)$ following from (2.3) that

$$\begin{aligned}\left\langle w^{k+1} - w, F(w)\right\rangle \le\ & \frac{1}{2}\left(\|w^k - w\|^2_G + \rho_{\alpha,\gamma}\|r^k\|^2 + \eta_{\alpha,\gamma}\|y^k - y^{k-1}\|^2_T\right) \\ & - \frac{1}{2}\left(\|w^{k+1} - w\|^2_G + \rho_{\alpha,\gamma}\|r^{k+1}\|^2 + \eta_{\alpha,\gamma}\|y^{k+1} - y^k\|^2_T\right)\end{aligned}$$

For all $k = 1, \ldots, t$, summing the above inequality from both sides and noting the notation of $\bar{w}_t$, we derive that

$$t\,\langle \bar{w}_t - w, F(w)\rangle \le \frac{1}{2}\left(\|w^1 - w^0\|^2_G + \rho_{\alpha,\gamma}\|r^1\|^2 + \eta_{\alpha,\gamma}\|y^1 - y^0\|^2_T\right). \tag{4.2}$$

Since (4.2) holds for any $w \in \mathcal{D}$, we can easily have (4.1). The proof is completed. □

Remark: Consider the case when the set $\mathcal{D}$ is compact, (4.1) implies the $O(1/t)$ ergodic convergence of sP-PRSM (1.7).

4.2 Convergence rate in the nonergodic sense

Lemma 4.1. *Let the sequence* $\{w^k\}$ *be generated by sP-PRSM* (1.7). *If* $0 \le \alpha \le 1$ *and* $\gamma > 0$, *then* $w^{k+1} \in \Omega^*$, *namely,* w^{k+1} *is one optimal solution of* (1.1), *if*

$$\|w^k - w^{k+1}\|_G = 0.$$

Proof. Following from (2.14) and (2.18), the statement $\|w^k - w^{k+1}\|_G = 0$ indicates that $r^{k+1} = 0$. Noting that G is always positive semidefinite, it is trivial to have $G(w^k - w^{k+1}) = 0$. Plugging the above two assertions into (3.2), we see that $\langle w - w^{k+1}, F(w^{k+1})\rangle \ge 0$ holds for any $w \in \mathcal{D}$. Following the optimality condition (2.7), we immediately see that $w^{k+1} \in \Omega^*$. The proof is completed. □

Based on this lemma, we can use $\|w^k - w^{k+1}\|_G$ to measure the accuracy of w^{k+1}. Similarly to the definition (2.8) of ϵ-solution of (1.1), we also call w^{k+1} as an ϵ-solution of (1.1) when $\|w^k - w^{k+1}\|_G \le \epsilon$. Together with (2.14) and (2.18), this also provide a practical stopping condition for sP-PRSM (1.7) as

$$\max\{\|x^{k+1} - x^k\|_S, \|y^{k+1} - y^k\|_T, \|B(y^{k+1} - y^k)\|, \|r^{k+1}\|\} \le \text{tol}, \tag{4.3}$$

where tol is some tolerance. Note that in (4.3), we may pay some more price to compute $\|x^{k+1} - x^k\|_S, \|y^{k+1} - y^k\|_T$. Thus, when the spectral norm of S or T is bounded, we can also simply stop the method when

$$\max\{\|B(y^{k+1} - y^k)\|, \|r^{k+1}\|\} \le \text{tol}. \tag{4.4}$$

Lemma 4.2. *Let the sequence* $\{w^k\}$ *be generated by sP-PRSM* (1.7). *If we choose* α *and* γ *according to* (1.8), *then there holds that*

$$\|w^{k+1} - w^k\|_G^2 - \|w^{k+2} - w^{k+1}\|_G^2 \ge \frac{1-\gamma}{\alpha+\gamma}\|(w^{k+1} - w^k) - (w^{k+2} - w^{k+1})\|_G^2. \tag{4.5}$$

Proof. Note that (3.8) also holds with $k := k+1$, then we have

$$\begin{aligned}(w - w^{k+2})^\mathsf{T} G(w^{k+2} - w^{k+1}) \ge{} & \left\langle r^{k+2} - r(w), (1-\alpha-\gamma)\beta r^{k+2} + (1-\alpha)\beta B(y^{k+1} - y^{k+2})\right\rangle \\ & + \left\langle w^{k+2} - w, F(w^{k+2})\right\rangle. \end{aligned}\tag{4.6}$$

Choosing w to be w^{k+2} and w^{k+1}, respectively, in (3.8) and (4.6) leads to

$$\begin{aligned}(w^{k+2} - w^{k+1})^\mathsf{T} G(w^{k+1} - w^k) \ge{} & \left\langle r^{k+1} - r^{k+2}, (1-\alpha-\gamma)\beta r^{k+1} + (1-\alpha)\beta B(y^k - y^{k+1})\right\rangle \\ & + \left\langle w^{k+1} - w^{k+2}, F(w^{k+1})\right\rangle. \end{aligned}\tag{4.7}$$

and

$$\begin{aligned}(w^{k+1} - w^{k+2})^\mathsf{T} G(w^{k+2} - w^{k+1}) \ge{} & \left\langle r^{k+2} - r^{k+1}, (1-\alpha-\gamma)\beta r^{k+2} + (1-\alpha)\beta B(y^{k+1} - y^{k+2})\right\rangle \\ & + \left\langle w^{k+2} - w^{k+1}, F(w^{k+2})\right\rangle. \end{aligned}\tag{4.8}$$

Adding (4.7) and (4.8) and noting $\langle w^{k+2} - w^{k+1}, F(w^{k+2}) - F(w^{k+1})\rangle \ge 0$, following from (2.3), we obtain that

$$\begin{aligned}& (w^{k+2} - w^{k+1})^\mathsf{T} G\left[(w^{k+1} - w^k) - (w^{k+2} - w^{k+1})\right] \\ \ge{} & (1-\alpha-\gamma)\beta\|r^{k+1} - r^{k+2}\|^2 + (1-\alpha)\beta\left\langle B\left[(y^k - y^{k+1}) - (y^{k+2} - y^{k+1})\right], r^{k+1} - r^{k+2}\right\rangle. \end{aligned}\tag{4.9}$$

Following the deriving process of (2.18), we have that

$$\begin{aligned}&\|(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\|_G^2\\ =\ &(1-\alpha)\beta\left\|B[(y^{k+1}-y^k)-(y^{k+2}-y^{k+1})]\right\|^2+(\alpha+\gamma)\beta\|r^{k+1}-r^{k+2}\|^2.\end{aligned}\tag{4.10}$$

Thus we conclude that

$$\begin{aligned}&\|w^{k+1}-w^k\|_H^2-\|w^{k+2}-w^{k+1}\|_H^2\\ =\ &2(w^{k+2}-w^{k+1})^\mathsf{T}G\left[(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\right]+\|(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\|_H^2\\ \geq\ &(2-\alpha-\gamma)\beta\|r^{k+1}-r^{k+2}\|^2+2(1-\alpha)\beta\left\langle B\left[(y^k-y^{k+1})-(y^{k+2}-y^{k+1})\right],r^{k+1}-r^{k+2}\right\rangle\\ &+(1-\alpha)\beta\left\|B[(y^{k+1}-y^k)-(y^{k+2}-y^{k+1})]\right\|^2\\ \geq\ &(1-\gamma)\beta\|r^{k+1}-r^{k+2}\|^2\geq\frac{1-\gamma}{\alpha+\gamma}\|(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\|_G^2,\end{aligned}$$

where the first inequality is due to (4.9) and (4.10), the second inequality is trivial and the last inequality owes to (4.10). The proof is completed. □

Theorem 4.2. *Let the sequence $\{v^k\}$ be generated by sP-PRSM (1.7). If $\gamma\in(0,1]$, then there holds that*

$$\|w^{t+1}-w^t\|_G^2\leq\frac{1}{\tau_{\alpha,\gamma}}\cdot\frac{1}{t}\cdot\left(\|w^1-w^0\|_G^2+\rho_{\alpha,\gamma}\|r^1\|^2+\eta_{\alpha,\gamma}\|y^1-y^0\|_T^2\right).\tag{4.11}$$

Proof. Summing (3.21)over $k=1,\dots,t$ leads to

$$\tau_{\alpha,\gamma}\cdot\sum_{k=1}^{t}\|w^{k+1}-w^k\|_G^2\leq\|w^1-w^*\|_G^2+\rho_{\alpha,\gamma}\|r^1\|^2+\eta_{\alpha,\gamma}\|y^1-y^0\|_T^2.\tag{4.12}$$

Observe that $\gamma\in(0,1]$, it follows from (4.5) that $\|w^{k+1}-w^k\|_G^2\geq\|w^{k+2}-w^{k+1}\|_G^2$, which together with (4.12) can easily suggest (4.11). The proof is completed. □

5 Linear convergence of sP-PRSM (1.7)

For simplicity, we only consider the case when $T=0$ and $S=0$. In this case, we always have that

$$G=\begin{pmatrix}0&0\\0&H\end{pmatrix}\quad\text{and}\quad\|w-w'\|_G^2=\|v-v'\|_H^2.\tag{5.1}$$

Assumption 5.1. *The set $\mathcal{Y}=\mathbb{R}^{n_2}$.*

Assumption 5.2. *The gradient $\nabla\theta_2(\cdot)$ is Lipschtiz continuous on $\mathbb{R}^{n_2}$ with $L>0$. Thus for any $y_1,y_2\in\mathbb{R}^{n_2}$ there holds that*

$$\|\nabla\theta_2(y_1)-\nabla\theta_2(y_2)\|\leq L\|y_1-y_2\|\tag{5.2}$$

and

$$\langle y_1-y_2,\nabla\theta_2(y_1)-\nabla\theta_2(y_2)\rangle\geq\frac{1}{L}\|\nabla\theta_2(y_1)-\nabla\theta_2(y_2)\|^2.\tag{5.3}$$

Assumption 5.3. *The function $\theta_2(\cdot)$ is strongly convex on $\mathbb{R}^{n_2}$ with constant $\mu>0$. Thus for any $y_1,y_2\in\mathbb{R}^{n_2}$ there holds that*

$$\langle y_1-y_2,\nabla\theta_2(y_1)-\nabla\theta_2(y_2)\rangle\geq\mu\|y_1-y_2\|^2.\tag{5.4}$$

Since $\mathcal{Y} = \mathbb{R}^{n_2}$, it follows from (3.4) that

$$\nabla\theta_2(y^{k+1}) = B^\mathsf{T}\lambda^{k+1} - (1-\gamma)\beta B^\mathsf{T} r^{k+1}. \tag{5.5}$$

Similarly, we obtain from (2.7) that

$$\nabla\theta_2(y^*) = B^\mathsf{T}\lambda^*. \tag{5.6}$$

Theorem 5.1. *Let the sequence $\{w^k\}$ be generated by sP-PRSM (1.7). Under Assumptions 5.1, 5.3, 5.2, there must exists some constant $c > 0$ such that*

$$\|v^k - v^*\|_H^2 + \rho_{\alpha,\gamma}\|r^k\|^2 \geq (1+c)\cdot\left(\|v^{k+1} - v^*\|_H^2 + \rho_{\alpha,\gamma}\|r^{k+1}\|^2\right).$$

Proof. Letting $v = v^*$ in (3.18), with (5.1), we can obtain

$$\begin{aligned}&\left(\|v^k - v^*\|_H^2 + \rho_{\alpha,\gamma}\|r^k\|^2\right) - \left(\|v^{k+1} - v^*\|_H^2 + \rho_{\alpha,\gamma}\|r^{k+1}\|^2\right)\\ &\geq \tau_{\alpha,\gamma}\|v^k - v^{k+1}\|_H^2 + 2\left\langle w^{k+1} - w^*, F(w^{k+1})\right\rangle.\end{aligned} \tag{5.7}$$

Next we will estimate the two terms of the righthand side of (5.7), respectively. Firstly, with (2.18), we see that

$$\|v^k - v^{k+1}\|_H^2 \geq (\alpha+\gamma)\beta\|r^{k+1}\|^2. \tag{5.8}$$

It follows from (2.3), (2.4) and the convexity of $\theta_1(\cdot)$ that

$$\left\langle w^{k+1} - w^*, F(w^{k+1})\right\rangle \geq \left\langle y^{k+1} - y^*, \nabla\theta_2(y^{k+1}) - \nabla\theta_2(y^*)\right\rangle \tag{5.9}$$

For any $0 \leq t \leq 1$, consider the convex combination of (5.4) and (5.3), we know from (5.9), (5.5) and (5.6) that there holds that

$$\left\langle w^{k+1} - w^*, F(w^{k+1})\right\rangle \geq \mu(1-t)\|y^{k+1} - y^*\|^2 + \frac{\lambda_{\min}(BB^\mathsf{T})}{L}t\cdot\|(\lambda^{k+1} - \lambda^*) - (1-\gamma)\beta r^{k+1}\|^2. \tag{5.10}$$

By applying the inequality

$$\|a+b\|^2 \geq \left(1 - \frac{1}{\rho}\right)\|a\|^2 + (1-\rho)\|b\|^2, \quad \rho > 1$$

with $a = \lambda^{k+1} - \lambda^*$, $b = -(1-\gamma)r^{k+1}$ and $\tau > 1$, we have

$$\|(\lambda^{k+1} - \lambda^*) - (1-\gamma)r^{k+1}\|^2 \geq \left(1 - \frac{1}{\rho}\right)\|\lambda^{k+1} - \lambda^*\|^2 + (1-\rho)|\gamma - 1|\beta\cdot\|r^{k+1}\|^2. \tag{5.11}$$

Plugging (5.11) into (5.10) , we further have

$$\left\langle w^{k+1} - w^*, F(w^{k+1})\right\rangle \geq c_1\left(\beta\|B\|^2\cdot\|y^{k+1} - y^*\|^2 + \frac{1}{\beta}\|\lambda^{k+1} - \lambda^*\|^2\right) + c_2\beta\|r^{k+1}\|^2, \tag{5.12}$$

where $c_1 = \min\left(\frac{\mu}{\beta\|B\|^2}(1-t), \left(1-\rho^{-1}\right)\frac{\lambda_{\min}(BB^\mathsf{T})}{L}\beta t\right)$, $c_2 = (1-\rho)|\gamma - 1|\frac{\lambda_{\min}(BB^\mathsf{T})}{L}t$. With the definition (2.9) of H, we can easily see that there exits a positive constant $c_{\alpha,\gamma} = \frac{2(\alpha+\gamma) - \alpha\gamma - \alpha\sqrt{\gamma^2 + 4(\alpha+\gamma)}}{2(\alpha+\gamma)(1-\alpha)}$ such that

$$\beta\|B\|^2\cdot\|y^{k+1} - y^*\|^2 + \frac{1}{\beta}\|\lambda^{k+1} - \lambda^*\|^2 \geq c_{\alpha,\gamma}\cdot\|v^{k+1} - v^*\|_H^2. \tag{5.13}$$

Combing (5.8), (5.12) and (5.13), we see that

$$\begin{aligned}\text{RHS of (5.7)} &\geq 2(c_1\cdot c_{\alpha,\gamma})\|v^{k+1} - v^*\|_H^2 + (2c_2 + \alpha + \gamma)\beta\|r^{k+1}\|^2\\ &\geq c\left(\|v^{k+1} - v^*\|_H^2 + \rho_{\alpha,\gamma}\|r^{k+1}\|^2\right),\end{aligned} \tag{5.14}$$

where $c = \min\left(2c_1 c_{\alpha,\gamma}, \frac{2c_2 + \alpha + \gamma}{\rho_{\alpha,\gamma}}\beta\right)$. The proof is completed. ☐

Remark. Let us consider as a special case, namely, $0 \le \alpha < 1, \gamma = 1$. Here, $\rho_{\alpha,\gamma} = 0$ and (5.11) reduces to $\|(\lambda^{k+1} - \lambda^*) - (1-\gamma)r^{k+1}\|^2 = \|\lambda^{k+1} - \lambda^*\|^2$. Hence $c = 2c_1c_3$. Letting $t = \frac{\mu L}{\mu L + \beta^2\|B\|^2\lambda_{\min}(BB^\mathsf{T})}$, we have that

$$c_1 = \frac{\mu\beta\lambda_{\min}(BB^\mathsf{T})}{\mu L + \beta^2\|B\|^2\lambda_{\min}(BB^\mathsf{T})} \le \frac{1}{2\kappa_B\sqrt{\kappa_{\theta_2}}}$$

where $\kappa_B = \sqrt{\lambda(BB^\mathsf{T})/\lambda_{\min}(BB^\mathsf{T})}$, $\kappa_{\theta_2} = L/\mu$ and the equality attains when $\beta = \sqrt{\frac{\mu L}{\|B\|^2\lambda_{\min}(BB^\mathsf{T})}}$. The largest c is

$$c_{\max} = \frac{c_{\alpha,\gamma}}{\kappa_B\sqrt{\kappa_{\theta_2}}}. \tag{5.15}$$

which reduces to (3.16) in [3] when $\alpha = 0$.

6 Numerical results

In this section, we demonstrate the potential efficiency of our method sP-PRSM (1.7) by solving the l_1-regularized least square problem:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1, \quad \mu > 0, \tag{6.1}$$

where $A \in \mathbb{R}^{m\times n}$ is the data matrix, m is the number of the data points, n is the number of features, $x \in \mathbb{R}^n$ is the vector of feature coefficients to be estimated and $\|x\|_1 := \sum_{i=1}^m |x_i|$, $b \in \mathbb{R}^m$ is the observation vector and $\mu \in \mathbb{R}$ is the regularization parameter. Given n, $m = n/2$ and a p-sparse vector $\bar{x} \in \mathbb{R}^m$ (p is the number of nonzero elements in x over m), the Matlab codes for generating the data of (6.1) are given as

```
xbar = sprandn(n,1,p);  D = randn(m,n);
D = D*spdiags(1./sqrt(sum(D.^2))',0,n,n); % normalize columns
b = D*xbar + sqrt(0.001)*randn(m,1);
mu = 0.1* norm(D'*b, 'inf').
```

By introducing an auxiliary variable $y \in \mathbb{R}^n$, we reformulate the problem (6.1) as

$$\min_{x\in\mathbb{R}^n,\ y\in\mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \mu\|y\|_1, \quad \text{s.t.} \quad x - y = 0. \tag{6.2}$$

We consider to apply the sP-PRSM (1.7) to solve (6.2). In our implementation, we always choose $T = 0$ and $\beta = 1$. Starting from $x^0 = y^0 = 0$ and $\lambda^0 = 0$, with some suitably chosen proximal matrix S, the iterative scheme is given as

$$\begin{cases} x^{k+1} = (A^TA + \beta I + S)^{-1}(A^Tb + \beta y^k + \lambda^k + Sx^k), \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} - y^k), \\ y^{k+1} = \mathcal{S}_{\mu/\beta}(x^{k+1} - \lambda^{k+\frac{1}{2}}/\beta), \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \gamma\beta(x^{k+1} - y^{k+1}), \end{cases} \tag{6.3}$$

where the shrinkage operator is defined as $\mathcal{S}_\nu(y)_i = \text{sgn}(y_i)\cdot\max\{|y_i| - \nu, 0\}$ and $\text{sgn}(y_i) = 1$ if $y_i \ge 0$ and $\text{sgn}(y_i) = -1$ if $y_i < 0$. According to (4.4), we stop the iterative scheme when

$$\max\{\|x^{k+1} - y^{k+1}\|, \|y^{k+1} - y^k\|\} \le 10^{-6} \tag{6.4}$$

or the iterative counter $k \ge 1000$.

Considering that the role of α and γ in the sP-PRSM (1.7) are equal, we can easily extend the convergence domain to

$$\alpha \in \left[0, \frac{1+\sqrt{5}}{2}\right), \quad \gamma \in \left(0, \frac{1-\alpha+\sqrt{(1+\alpha)^2+4(1-\alpha^2)}}{2}\right). \tag{6.5}$$

Note that [13] also establish the convergence of the sP-PRSM (1.7) with $S = 0$ and $T = 0$ when

$$\alpha \in (-1, 0), \gamma \in (0, \alpha^2 - \alpha - 1), \text{ or } \alpha \in (0, 1), \gamma \in (\alpha^2 - \alpha - 1, 0). \tag{6.6}$$

The union domain defined by (6.5) and (6.6) will be considered below. We generate the mesh grid with respect to α and γ by equally dividing the intervals $[-1, 0]$ and $[0, 1.618]$ into 10 parts, respectively. We set $n = 2000$, $p = 0.2$. The corresponding scatter diagram is depicted in Figure 1. For consideration of space, only the detailed results for (6.5) are shown in Table 1. From this table, we see that when $\alpha + \gamma = 1.618$ and $\min(\alpha, \gamma) \geq 0.485$, the sP-PRSM (1.7) always perform best; when α and γ satisfy (6.6), the results are not so good.

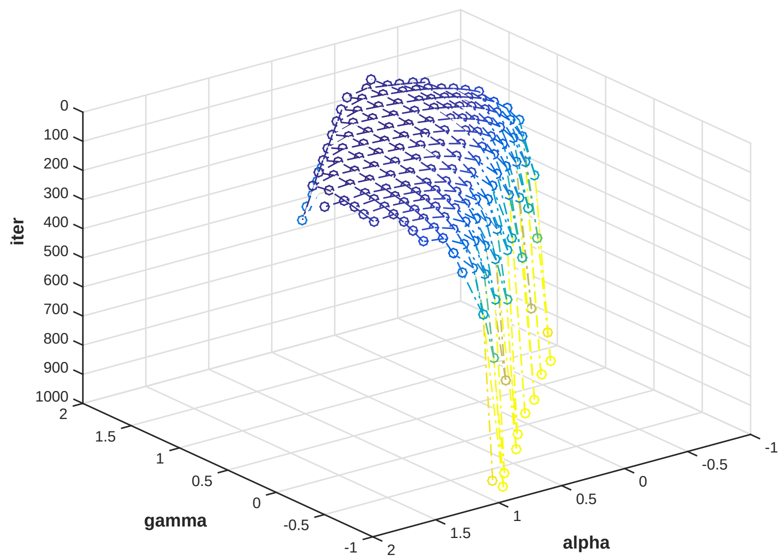

Figure 1: $n = 2000$, $p = 0.2$, the intervals $[-1, 0]$ and $[0, 1.618]$ are both equally divided into 10 parts. Note that the 'iter'-axis and 'alpha'-axis are both in reverse order.

Based on the above observation, we below focus on the choice of $\alpha = 0.618$ and $\gamma = 1$. For simplicity, we call the corresponding sP-PRSM (1.7) as sP-PRSM*. Recall that the classical ADMM always take the dual stepsize as 1 or 1.618 (named by ADMM-1, ADMM-2), covered by our sP-PRSM by taking $\alpha = 0, \gamma = 1$ or $\alpha = 0, \gamma = 1.618$, we next compare these three methods for solving (6.1).

We first consider the sP-PRSM (1.7) without proximal terms, namely, $S = 0$. The comparison results are shown in Table 2. Comparing with ADMM-1, sP-PRSM* can always reduce the number of iterations in 40%. While as compared to ADMM-2, sP-PRSM* can reduce the number of iterations at about 40% for the hard case when the sparsity level p is 0.2 or 0.3, and lead to about 2-7% reduction for the easy case when p is 0.1.

We now consider the sP-PRSM (1.7) with semi-positive proximal matrix S, which takes the form as

$$S = \beta\left((\xi - 1)I - \frac{1}{2\beta}A^\mathsf{T}A\right) \text{ with } \xi = 1.01\lambda_{\max}(I + \frac{1}{2\beta}A^\mathsf{T}A). \tag{6.7}$$

Table 1: The number of iterations taken by sP-PRSM (1.7) with different α and γ for solving (6.1) with $n = 2000$, $p = 0.2$, '-' means that the corresponding α and γ do not satisfy (1.8).

γ \ α	0.000	0.162	0.324	0.485	0.647	0.809	0.971	1.133	1.294	1.456	1.618
0.000	–	384	189	124	92	76	66	61	56	62	65
0.162	384	189	124	92	73	63	56	52	48	51	–
0.324	189	124	92	73	61	55	49	45	43	60	–
0.485	124	92	73	61	53	48	44	40	60	–	–
0.647	92	73	61	53	48	44	40	61	246	–	–
0.809	75	63	54	48	44	40	61	246	–	–	–
0.971	64	55	48	43	40	60	246	–	–	–	–
1.133	60	52	45	40	60	243	–	–	–	–	–
1.294	56	48	43	61	242	–	–	–	–	–	–
1.456	60	50	62	–	–	–	–	–	–	–	–
1.618	63	–	–	–	–	–	–	–	–	–	–

Table 2: Comparison among the number of iterations taken by ADMM-1, ADMM-2 and sP-PRSM* for solving (6.1) with $S = 0$

n	p	ADMM-1	ADMM-2	sP-PRSM*	ratio1(%)	ratio2(%)
2000	0.1	65	43	40	61.5	93.0
2000	0.2	63	63	39	61.9	61.9
2000	0.3	85	93	53	62.4	57.0
4000	0.1	68	43	42	61.8	97.7
4000	0.2	67	62	42	62.7	67.7
4000	0.3	80	88	49	61.3	55.7
8000	0.1	70	44	43	61.4	97.7
8000	0.2	66	62	41	62.1	66.1
8000	0.3	69	79	42	60.9	53.2

Now we name ADMM-1 and ADMM-2 as sPADMM-1 and sPADMM-2, respectively. The results are shown in Table 3. Note that in this case, sPADMM-1 performs better than sPADMM-2, while sP-PRSM* is slightly better than sPADMM-1.

Table 3: Comparison among the number of iterations taken by sPADMM-1, sPADMM-2 and sP-PRSM for solving (6.1) with semi-positive S

n	p	sPADMM-1	sPADMM-2	sP-PRSM	ratio1(%)	ratio2(%)
2000	0.1	236	245	231	97.9	94.3
2000	0.2	401	408	385	96.0	94.4
2000	0.3	574	579	549	95.6	94.8
4000	0.1	195	201	189	96.9	94.0
4000	0.2	395	402	380	96.2	94.5
4000	0.3	542	548	518	95.6	94.5
8000	0.1	210	220	206	98.1	93.6
8000	0.2	392	399	377	96.2	94.5
8000	0.3	491	498	471	95.9	94.6

Finally, we follow the way in [15] to generate an indefinite proximal matrix S as

$$S = \beta(\xi - 1)I - A^\top A, \text{ with } \xi = \lambda_{\max}(I + \frac{1}{\beta}A^\top A). \tag{6.8}$$

Now we name ADMM-1 and ADMM-2 as iPADMM-1 and iPADMM-2, respectively. We list the results in Table 4. From this table, we can see that sP-PRSM* performs best among the three methods. On average, it can lead to 5% reduction in the number of iterations as compared to iPADMM-1 and 10% reduction as compared to iPADMM-2. We also remark that for our test instances, no matter choosing the semi-definite or indefinite S, ADMM-1 is always better than ADMM-2. Besides, for each method, the version with indefinite S in (6.8) can always bring out 40-50% improvement in the number of iterations as compared to the version with semi-definite S in (6.7). This coincides with the observation in [15].

Table 4: Comparison among the number of iterations taken by iPADMM-1, iPADMM-2 and sP-PRSM for solving (6.1) with indefinite S

n	p	iPADMM-1	iPADMM-2	sP-PRSM	ratio1(%)	ratio2(%)
2000	0.1	124	138	122	98.4	88.4
2000	0.2	231	240	216	93.5	90.0
2000	0.3	342	348	315	92.1	90.5
4000	0.1	105	116	101	96.2	87.1
4000	0.2	228	237	213	93.4	89.9
4000	0.3	322	328	297	92.2	90.5
8000	0.1	107	122	106	99.1	86.9
8000	0.2	226	235	211	93.4	89.8
8000	0.3	288	296	267	92.7	90.2

From Table 4 we can observe that we can use indefinite matrices in the subproblems (1.7a) and (1.7c), which can further improve the efficiency of the algorithm. In fact, under certain conditions on the problems' data and suitable requirements on the matrices, we can theoretically establish its convergence. Since the analysis is very similar to those in this paper, we do not describe them here.

7 Conclusions

In this paper, we 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. We established the relationship between the two parameters under which we proved the global convergence of the algorithm. We also analyzed the sublinear rate convergence under ergodic and non-ergodic senses, respectively. Under further conditions of one of the objective functions, we proved the linear convergence of the algorithm. Finally, we reported extensive numerical results, indicating the efficiency of the proposed algorithm.

Note that the parameters α and γ are essential to the efficiency of the algorithm, which should be variable along with the iteration. Allowing the parameter α and γ varying with the process of the iterate may give us the freedom of choosing them in a self-adaptive manner. Suitable updating rules are among our future research tasks.

On the other hand, considering that the main task in the algorithm is to solve the x-optimization problem (e.g., (1.5a)) or the y-optimization problem (e.g., (1.5c)), solving them in an inexact manner may improve the efficiency of the algorithm. An approximate version of the proposed sP-PRSM with practical accuracy criteria is also our future research topic.

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine Learning, 3 (2011), pp. 1–122.

[2] E. Corman and X. Yuan, *A generalized proximal point algorithm and its convergence rate*, SIAM Journal on Optimization, 24 (2014), pp. 1614–1638.

[3] W. Deng and W. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, tech. report, DTIC Document, 2012.

[4] J. Douglas and H. Rachford, *On the numerical solution of heat conduction problems in two and three space variables*, Transactions of the American mathematical Society, (1956), pp. 421–439.

[5] J. Eckstein and M. Fukushima, *Some reformulation and applications of the alternating directions method of multipliers.* In: Hager, W.W. et al., eds., Large Scale Optimization: State of the Art, pp. 115-134. Kluwer Academic Publishers, 1994.

[6] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, vol. 1, Springer, 2003.

[7] M. Fazel, T. K. Pong, D. Sun, and P. Tseng, *Hankel matrix rank minimization with applications to system identification and realization*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 946–977.

[8] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, 2 (1976), pp. 17–40.

[9] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer, 1984.

[10] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires*, ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 9 (1975), pp. 41–76.

[11] B.S. He, L.Z. Liao, D.R. Han, H. Yang, *A new inexact alternating directions method for monotone variational inequalities*, Mathematical Programming, 92 (2002), pp. 103–118.

[12] B. He, H. Liu, Z. Wang, and X. Yuan, *A strictly contractive Peaceman–Rachford splitting method for convex programming*, SIAM Journal on Optimization, 24 (2014), pp. 1011–1040.

[13] B. He, F. Ma, and X. Yuan, *On the step size of symmetric alternating directions method of multipliers*, tech. report, Optimization online, 2015.

[14] B. He and X. Yuan, *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 700–709.

[15] M. Li, D. Sun, and K.-C. Toh, *A majorized admm with indefinite proximal terms for linearly constrained convex composite optimization*, arXiv preprint arXiv:1412.1911, (2014).

[16] Y. Nesterov, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$*, in Soviet Mathematics Doklady, vol. 27, 1983, pp. 372–376.

[17] ———, *Gradient methods for minimizing composite objective function*, 2007.

[18] D. W. Peaceman and H. H. Rachford, Jr, *The numerical solution of parabolic and elliptic differential equations*, Journal of the Society for Industrial & Applied Mathematics, 3 (1955), pp. 28–41.

[19] M. Xu and T. Wu, *A class of linearized proximal alternating direction methods*, Journal of Optimization Theory and Applications, 151 (2011), pp. 321–337.