Journal of Computational Mathematics
Vol.41, No.6, 2023, 1017–1040.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2112-m2020-0023

AN INDEFINITE-PROXIMAL-BASED STRICTLY CONTRACTIVE PEACEMAN-RACHFORD SPLITTING METHOD[*1)]

Yan Gu
Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China
Key Laboratory of Mathematical Modelling and High Performance Computing of Air Vehicles (NUAA), MIIT, Nanjing, 210016, China
Email: guyanmath@nuaa.edu.cn
Bo Jiang
School of Mathematical Sciences, Key Laboratory for NSLSCS of Jiangsu Province, Nanjing Normal University, Nanjing, 210023, China
Email: jiangbo@njnu.edu.cn
Deren Han[2)]
LMIB, School of Mathematics and Systems Science, Beihang University, Beijing, 100191, China
Key Laboratory for NSLSCS of Jiangsu Province, China
Email: handr@buaa.edu.cn

Abstract

The Peaceman-Rachford splitting method is efficient for minimizing a convex optimization problem with a separable objective function and linear constraints. However, its convergence was not guaranteed without extra requirements. 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. Besides, motivated by the recent advances on the ADMM type method with indefinite proximal terms, we employ the indefinite proximal term in the strictly contractive Peaceman-Rachford splitting method. We show that the proposed indefinite-proximal strictly contractive Peaceman-Rachford splitting method is convergent and also prove the $o(1/t)$ convergence rate in the nonergodic sense. The numerical tests on the l_1 regularized least square problem demonstrate the efficiency of the proposed method.

Mathematics subject classification: 90C25, 90C30, 65K05.
Key words: Indefinite 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:

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

[*] Received January 22, 2020 / Revised version received June 9, 2021 / Accepted December 31, 2021 / Published online April 1, 2022 /

[1)] This paper is an improved version of our earlier paper: "Y. Gu, B. Jiang, D. Han, A semi-proximal-based strictly contractive Peaceman-Rachford splitting method, arXiv:1506.02221, 2015".

[2)] Corresponding author

where $A \in \mathbb{R}^{m\times n_1}$ and $B \in \mathbb{R}^{m\times n_2}$, $b \in \mathbb{R}^m$, $\mathcal{X} := \mathbb{R}^{n_1}$, $\mathcal{Y} := \mathbb{R}^{n_2}$. The functions $\theta_1(x) := p(x) + f(x)$ and $\theta_2(y) := h(y) + g(y)$, where $p : \mathcal{X} \to (-\infty, +\infty]$ and $h : \mathcal{Y} \to (-\infty, +\infty]$ are proper closed convex (could be nonsmooth) functions; $f : \mathcal{X} \to (-\infty, +\infty)$ and $g : \mathcal{Y} \to (-\infty, +\infty)$ are two convex functions with Lipschitz continuous gradients on $\mathcal{X}$ and $\mathcal{Y}$. Throughout, the solution set of (1.1) is assumed to be nonempty. Note that one can also consider $\mathcal{X}$ and $\mathcal{Y}$ as general real finite dimensional Euclidean or Hilbert spaces, see [3,15,35] for instances. For ease of presentation, we adopt the $\mathcal{X}$ and $\mathcal{Y}$ as the ordinary $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ in this paper.

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.

A well-known method called alternating direction method of multipliers (ADMM) is efficient to minimize such problems. It was observed in Gabay and Mercier [11], Glowinski and Marrocco [13] that ADMM can be derived from applying the Douglas-Rachford operator splitting method [8] to the dual of the problem (1.1). The iterative sequence is given as the following recursion:

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

Based on another classical operator splitting method, i.e., the Peaceman-Rachford operator splitting method [30], one can derive the following similar method for (1.1):

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

While the global convergence of the alternating direction method of multipliers (1.3a)-(1.3c) can be established under very mild conditions [2], the convergence of the Peaceman-Rachford-based method (1.4a)-(1.4d) can not be guaranteed without further conditions [5].

He *et al.* [17] proposed 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). Note that when $\alpha = 1$, it is the same as (1.4a)-(1.4d). They explained the non-convergence behavior of (1.4a)-(1.4d) from the contractive perspective, i.e., the distance from the iterative point to the solution set is merely nonexpansive, but not contractive. Under the condition that $\alpha \in (0, 1)$, they proved the same sublinear convergence rate as that for ADMM [20]. Particularly, they showed that it achieves an approximate solution of (1.1) with the accuracy of $\mathcal{O}(1/t)$ after t iterations[1)] , both in the ergodic and nonergodic sense. Besides, Gu [14] and He *et al.* [18] took two different constants α and γ to different step sizes in (1.4b) and (1.4d). The convergence results, including global

[1)] A worst-case $\mathcal{O}(1/t)$ convergence rate means the accuracy to a solution under certain criteria is of the order $\mathcal{O}(1/t)$ after t iterations of an iterative scheme; or equivalently, it requires at most $\mathcal{O}(1/\epsilon)$ iterations to achieve an approximate solution with an accuracy of ϵ. See, e.g., [27,28].

convergence, the worst-case $\mathcal{O}(1/t)$ convergence rate in the ergodic sense, have been established in it but without the worst-case $\mathcal{O}(1/t)$ convergence rate in the nonergodic sense. Chen *et al.* [4] proposed a variant Peaceman-Rachford splitting method in a prediction-correction framework. For some recent advances of the Peaceman-Rachford splitting method, one can refer to [1, 12, 21, 23, 24, 33], to name a few.

Considering that in many cases the subproblem in (1.3a)-(1.3c) and (1.4a)-(1.4d) might be difficult to solve and that in some applications θ_1 or θ_2 is a convex quadratic function, Eckstein [9] and He *et al.* [16] considered to add proximal terms to the subproblems for different purpose. Fazel *et al.* [10] proposed the following semi-proximal ADMM scheme:

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

where $\gamma \in (0, (1+\sqrt{5})/2)$. They allowed S and T to be positive semidefinite which makes the algorithm more flexible. We refer the reader to [7,10,20,34] for a brief history of the development of the semi-proximal ADMM and the corresponding convergence results. To further relax the requirements of the proximal terms, Li *et al.* [25] considered majorized ADMM with indefinite proximal S and T. They established the convergence and the sublinear convergence rate under some mild assumptions. The numerical results in [25] showed that the (majorized) ADMM with indefinite proximal term always performs better than that with semidefinite proximal terms. Very recently, He *et al.* [19] obtained a linearized ADMM with an optimal indefinite proximal term. In their method, $S = 0$ and T in (1.5b) is chosen by

$$T = \tau r I_{n_2} - \beta B^\top B \ \text{ with } \ r > \beta\|B^\top B\|, \quad \tau \in (0.75, 1), \tag{1.6}$$

where $\|B^\top B\|$ means the spectral norm of $B^\top B$. Note that they require that the dual stepsize $\gamma = 1$ in (1.5). The small value $\tau \in (0.75, 1)$ can ensure the proximal term has less weight for the y-subproblem (1.5b), and thus allows for a larger step. Solving a general problem, i.e., finding zeros of a maximal operator, using a proximal point algorithm with indefinite proximal term, was recently developed in [22].

It is natural to extend the proximal ADMM to the proximal Peaceman-Rachford splitting method. For convenience, we first introduce the whole update scheme of the *indefinite-proximal-based strictly contractive Peaceman-Rachford splitting method (iPSPR)*

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

where S and T are symmetric and possibly indefinite. Gao *et al.* [12] considered the generalized ADMM with indefinite proximal term, which corresponds to (1.7) with $S = 0$ and $\gamma = 1$. The proximal term T takes a similar formulation as (1.6) but with $\tau \in [\frac{\alpha^2-\alpha+4}{\alpha^2-2\alpha+5}, 1)$. Jiang *et al.* [23] considered the same generalized ADMM as in [12], but they give an optimal bound of τ as $\tau \in (\frac{3+\alpha}{4}, 1)$. For other related works one can refer to [26, 32].

In this paper, we focus on (1.7) with indefinite S and T. Our main contributions are two-fold. Firstly, motivated by the nice analysis techniques in [17] and [34], we prove the global convergence of iPSPR under some assumptions on S and T, see (3.32) and (3.33), in which the stepsizes α and γ are in the range

$$(\alpha,\gamma)\in\mathbb{D}:=\left\{(\alpha,\gamma):0\le\alpha<1,0\le\gamma<\frac{1-\alpha+\sqrt{(1+\alpha)^2+4(1-\alpha^2)}}{2},\alpha+\gamma>0\right\}. \tag{1.8}$$

With some additional mild requirements, see (4.6), we prove that the iPSPR is $o(1/t)$ sublinearly convergent in the nonergodic sense. Secondly, our proposed requirements on the proximal T can cover some existing results, such as the special linearized choice (1.6) in [19, 23]. More importantly, our proposed requirements on the proximal T employs both the Hessian information of the objective function and the information of $\beta B^{\mathsf T}B$ for the first time. Note that He *et al.* [19] only uses the information of $\beta B^{\mathsf T}B$, while Li *et al.* [25] only considers the Hessian information of the objective function.

The rest of this paper is organized as follows. In Section 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 Section 3, we first give the contraction analysis of iPSPR (1.7), and then establish the global convergence. We will discuss how to choose T in the end of Section 3. The detailed formulae will be given for the different ranges of the parameters α and γ. We discuss the nonergodic sublinear convergence rate in Section 4. In Section 5, we test the l_1 regularized least square problem to show the efficiency of the proposed iPSPR (1.7). Finally, we make some conclusions in Section 6.

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. Denote $\Omega=\mathcal{X}\times\mathcal{Y}\times\mathbb{R}^m$. Let $\mathcal{U}$ be the feasible set of (1.1), namely, $\mathcal{U}=\{(x,y):Ax+By=b,x\in\mathcal{X},y\in\mathcal{Y}\}$ and denote $\mathcal{D}=\mathcal{U}\times\mathbb{R}^m$. Throughout this paper, we make the following assumption.

Assumption 2.1. *Let $\Omega^*\subset\mathcal{D}$ 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 $\theta_2(\cdot)$, there exist two positive semidefinite matrices Σ_1 and Σ_2 such that for all $x,\ x'\in\mathbb{R}^{n_1}$ and $\xi_x\in\partial\theta_1(x)$, $\xi_x'\in\partial\theta_1(x')$,

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

and for all $y,\ y'\in\mathbb{R}^{n_2}$, $\xi_y\in\partial\theta_2(y)$, $\xi_y'\in\partial\theta_2(y')$,

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

Denote $u=\begin{pmatrix}x\\y\end{pmatrix}$, $v=\begin{pmatrix}y\\\lambda\end{pmatrix}$ and $w=\begin{pmatrix}x\\y\\\lambda\end{pmatrix}$. For given w, and some specific subgradients $\xi_x\in\partial\theta_1(x)$ and $\xi_y\in\partial\theta_2(x)$, we define $F(w,\xi_x,\xi_y)=\begin{pmatrix}\xi_x-A^{\mathsf T}\lambda\\\xi_y-B^{\mathsf T}\lambda\\Ax+By-b\end{pmatrix}$. Due to the convexity of $\theta_1(\cdot)$ and $\theta_2(\cdot)$, it is easy to show that the operator $F(\cdot)$ is monotone. Specifically, for any $w, w' \in \mathcal{D}$, we have

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

where $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$ and the inequality is due to (2.1) and (2.2).

Following Theorem 3.1.24 in [29], we say that $w^* \in \Omega^*$ if and only if there exists $\xi_x^* \in \partial\theta_1(x^*)$ and $\xi_y^* \in \partial\theta_2(y^*)$ such that $\langle x - x^*, \xi_x^*\rangle + \langle y - y^*, \xi_y^*\rangle \geq 0$, which is further equivalent to

$$\big\langle w - w^*, F(w^*,\xi_x^*,\xi_y^*)\big\rangle \geq 0, \quad \forall w \in \mathcal{D}, \tag{2.4}$$

because

$$\big\langle w - w^*, F(w^*,\xi_x^*,\xi_y^*)\big\rangle = \langle x - x^*, \xi_x^*\rangle + \langle y - y^*, \xi_y^*\rangle \geq 0, \ \forall w \in \mathcal{D}.$$

2.2. Some notations

We use the symbol 0 to denote a zero matrix, whose size can be always easily identified form the context. 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 Z, we mark $Z \succeq 0$ (resp. $Z \succ 0$) if Z is positive semidefinite (resp. positive definite). For any given symmetric matrix T, we decompose it as

$$T = T_+ - T_- \quad \text{with} \quad T_+ \succeq 0 \quad \text{and} \quad T_- \succeq 0.$$

To make the analysis more elegantly, we use $r^k = Ax^k + By^k - b$ for short. Similarly, for any $w \in \Omega$, 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

$$H = \frac{1}{\alpha+\gamma}\begin{pmatrix} (\alpha+\gamma-\alpha\gamma)\beta B^\mathsf{T} B & -\alpha B^\mathsf{T} \\ -\alpha B & \frac{1}{\beta} I_m \end{pmatrix}, \tag{2.5}$$

$$\hat{H} := \begin{pmatrix} T + \Sigma_2 & 0 \\ 0 & 0 \end{pmatrix} + H = \begin{pmatrix} T + \Sigma_2 + \dfrac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\mathsf{T} B & -\dfrac{\alpha}{\alpha+\gamma} B^\mathsf{T} \\ -\dfrac{\alpha}{\alpha+\gamma} B & \dfrac{1}{(\alpha+\gamma)\beta} I_m \end{pmatrix}. \tag{2.6}$$

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

$$G := \begin{pmatrix} P & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & H \end{pmatrix} = \begin{pmatrix} S & 0 & 0 \\ 0 & T + \dfrac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\mathsf{T} B & -\dfrac{\alpha}{\alpha+\gamma} B^\mathsf{T} \\ 0 & -\dfrac{\alpha}{\alpha+\gamma} B & \dfrac{1}{(\alpha+\gamma)\beta} I_m \end{pmatrix}, \tag{2.7}$$

and

$$\begin{aligned} \widehat{G} &:= \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} + G = \begin{pmatrix} S + \Sigma_1 & 0 & 0 \\ 0 & T + \Sigma_2 + \dfrac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\mathsf{T} B & -\dfrac{\alpha}{\alpha+\gamma} B^\mathsf{T} \\ 0 & -\dfrac{\alpha}{\alpha+\gamma} B & \frac{1}{(\alpha+\gamma)\beta} I_m \end{pmatrix} \\ &= \begin{pmatrix} S + \Sigma_1 & 0 \\ 0 & \hat{H} \end{pmatrix}. \end{aligned} \tag{2.8}$$

It follows from (2.7) and (2.8) that for any $w, w' \in \Omega$,

$$\|w - w'\|_G^2 = \|u - u'\|_P^2 + \|v - v'\|_H^2, \tag{2.9}$$

$$\|w - w'\|_{\widehat{G}}^2 = \|u - u'\|_\Sigma^2 + \|w - w'\|_G^2 = \|x - x'\|_{S+\Sigma_1}^2 + \|v - v'\|_{\hat{H}}^2. \tag{2.10}$$

With the update scheme (1.7b) and (1.7d), it is easy to have

$$\lambda^k = \lambda^{k+1} + (\alpha + \gamma)\beta r^{k+1} + \alpha\beta B(y^k - y^{k+1}). \tag{2.11}$$

With (2.5) and (2.11), we thus have

$$\|v^k - v^{k+1}\|_H^2 = (1-\alpha)\beta\|B(y^k - y^{k+1})\|^2 + (\alpha+\gamma)\beta\|r^{k+1}\|^2. \tag{2.12}$$

Finally, it is easy to have the following proposition.

Proposition 2.1. *If $0 \le \alpha \le 1$ and $\gamma > 0$, then $H \succeq 0$. If $T + \Sigma_2 + (1-\alpha)\beta B^\mathsf{T} B \succ 0$, then $\hat{H} \succ 0$. If $T + \Sigma_2 + (1-\alpha)\beta B^\mathsf{T} B \succ 0$ and $S + \Sigma_1 \succeq 0$, then $\widehat{G} \succeq 0$.*

3. Convergence of iPSPR

In this section, we first show that a sequence related to $\{w_k\}$ generated by iPSPR (1.7) is strictly contractive in section 3.1 and then establish the global convergence of the method in Section 3.2, and discuss the choices of the proximal terms in section 3.3. Note that the contraction property is also helpful to establish the convergence rate in the nonergodic sense.

3.1. Contraction analysis

To establish the strictly contractive property of the sequence $\{\Phi_{\alpha,\gamma}^k(w^*)\}$ (see (3.21) for the definition), we first give a rough estimation of $\|w^k - w^*\|_{\widehat{G}}^2 - \|w^{k+1} - w^*\|_{\widehat{G}}^2$ based on the optimality conditions of (1.7a) and (1.7c).

Lemma 3.1. *Let the sequence $\{w^k\}$ be generated by iPSPR (1.7). If we choose $(\alpha, \gamma) \in \mathbb{D}$, then there holds that*

$$\begin{aligned}
&\|w^k - w^*\|_{\widehat{G}}^2 - \|w^{k+1} - w^*\|_{\widehat{G}}^2 \\
\ge\ &\|x^k - x^{k+1}\|_{S+\frac{1}{2}\Sigma_1}^2 + \|y^k - y^{k+1}\|_{T+\frac{1}{2}\Sigma_2}^2 + (1-\alpha)\beta\|B(y^k - y^{k+1})\|^2 \\
&+ (2-\alpha-\gamma)\beta\|r^{k+1}\|^2 + 2(1-\alpha)\beta\left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle,
\end{aligned} \tag{3.1}$$

and

$$\begin{aligned}
&\|w^k - w^*\|_{\widehat{G}}^2 - \|w^{k+1} - w^*\|_{\widehat{G}}^2 \\
\ge\ &\|x^k - x^{k+1}\|_{S+\frac{1}{2}\Sigma_1}^2 + \|y^k - y^{k+1}\|_{T+\frac{1}{2}\Sigma_2}^2 \\
&+ \frac{\alpha^2(1-\gamma) + \gamma^2(1-\alpha)}{(\alpha+\gamma)^2}\beta\|B(y^k - y^{k+1})\|^2 + \frac{2-\alpha-\gamma}{(\alpha+\gamma)^2\beta}\left\|\lambda^k - \lambda^{k+1}\right\|^2 \\
&+ \frac{2(\gamma-\alpha)}{(\alpha+\gamma)^2}\left\langle B(y^k - y^{k+1}), \lambda^k - \lambda^{k+1}\right\rangle.
\end{aligned} \tag{3.2}$$

Proof. The proof of (3.1) consists of three steps.

I). We give a rough lower bound estimation of the term $\|w^k-w\|^2_{\widehat{G}}-\|w^{k+1}-w\|^2_{\widehat{G}}$. Following from the first equality of (2.10), we have

$$\begin{aligned}&\|w^k-w\|^2_{\widehat{G}}-\|w^{k+1}-w\|^2_{\widehat{G}}\\=&\|w^k-w\|^2_G-\|w^{k+1}-w\|^2_G+\|u^k-u\|^2_\Sigma-\|u^{k+1}-u\|^2_\Sigma.\end{aligned}\tag{3.3}$$

The Cauchy-Schwartz inequality ensures $\|u^k-u\|^2_\Sigma+\|u^{k+1}-u\|^2_\Sigma\geq\frac{1}{2}\|u^k-u^{k+1}\|^2_\Sigma$. Thus, we have

$$\|u^k-u\|^2_\Sigma-\|u^{k+1}-u\|^2_\Sigma\geq\frac{1}{2}\|u^k-u^{k+1}\|^2_\Sigma-2\|u^{k+1}-u\|^2_\Sigma.\tag{3.4}$$

Using the identity $\|a\|^2_G-\|b\|^2_G=\|a-b\|^2_G+2b^\top G(a-b)$ with $a=w-w^k$ and $b=w-w^{k+1}$, we have

$$\|w^k-w\|^2_G-\|w^{k+1}-w\|^2_G=\|w^k-w^{k+1}\|^2_G+2(w-w^{k+1})^\top G(w^{k+1}-w^k).\tag{3.5}$$

Substituting (3.4) and (3.5) into (3.3), and using (2.9) and (2.12), we have that for any $w\in\Omega$

$$\begin{aligned}&\|w^k-w\|^2_{\widehat{G}}-\|w^{k+1}-w\|^2_{\widehat{G}}\\\geq&\|x^k-x^{k+1}\|^2_{S+\frac{1}{2}\Sigma_1}+\|y^k-y^{k+1}\|^2_{T+\frac{1}{2}\Sigma_2}+(1-\alpha)\beta\|B(y^k-y^{k+1})\|^2\\&+(\alpha+\gamma)\beta\|r^{k+1}\|^2+2(w-w^{k+1})^\top G(w^{k+1}-w^k)-2\|u^{k+1}-u\|^2_\Sigma.\end{aligned}\tag{3.6}$$

II). We focus on the estimation of $(w-w^{k+1})^\top G(w^{k+1}-w^k)$. From the optimality conditions of (1.7a) and (1.7c), we know that there exist $\xi_x^{k+1}\in\partial\theta_1(x^{k+1})$ and $\xi_y^{k+1}\in\partial\theta_2(y^{k+1})$ such that

$$\begin{aligned}&\left\langle x-x^{k+1},S(x^{k+1}-x^k)+\xi_x^{k+1}-A^\top\lambda^k+\beta A^\top(Ax^{k+1}+By^k-b)\right\rangle\geq0,\quad\forall x\in\mathcal{X},\\&\left\langle y-y^{k+1},T(y^{k+1}-y^k)+\xi_y^{k+1}-B^\top\lambda^{k+\frac{1}{2}}+\beta B^\top r^{k+1}\right\rangle\geq0,\quad\forall y\in\mathcal{Y}.\end{aligned}\tag{3.7}$$

Substituting (2.11) into (3.7) and noting $r^{k+1}=Ax^{k+1}+By^{k+1}-b$, we have

$$\begin{aligned}&\langle x-x^{k+1},S(x^{k+1}-x^k)+\xi_x^{k+1}-A^\top\lambda^{k+1}+(1-\alpha-\gamma)\beta A^\top r^{k+1}\\&\quad+(1-\alpha)\beta A^\top B(y^k-y^{k+1})\rangle\geq0,\qquad\forall x\in\mathcal{X}.\end{aligned}\tag{3.8}$$

Substituting $\lambda^{k+\frac{1}{2}}=\lambda^{k+1}+\gamma\beta r^{k+1}$ into (3.7), we have

$$\left\langle y-y^{k+1},T(y^{k+1}-y^k)+\xi_y^{k+1}-B^\top\lambda^{k+1}+(1-\gamma)\beta B^\top r^{k+1}\right\rangle\geq0,\quad\forall y\in\mathcal{Y}.\tag{3.9}$$

Rewrite (2.11) to

$$r^{k+1}-\frac{\alpha}{\alpha+\gamma}B(y^{k+1}-y^k)+\frac{1}{(\alpha+\gamma)\beta}(\lambda^{k+1}-\lambda^k)=0.\tag{3.10}$$

Combing (3.8)–(3.10) in a suitable way, and recalling the definitions of w and $F(\cdot)$, for any $w\in\Omega$ there holds that

$$\left\langle w-w^{k+1},\begin{pmatrix}S(x^{k+1}-x^k)\\T(y^{k+1}-y^k)\\0\end{pmatrix}+\begin{pmatrix}0\\\alpha\beta B^\top r^{k+1}+(1-\alpha)\beta B^\top B(y^{k+1}-y^k)\\-\dfrac{\alpha}{\alpha+\gamma}B(y^{k+1}-y^k)+\dfrac{1}{(\alpha+\gamma)\beta}(\lambda^{k+1}-\lambda^k)\end{pmatrix}\right\rangle$$

$$\geq \left\langle w^{k+1} - w, \begin{pmatrix} A^\mathsf{T} \\ B^\mathsf{T} \\ 0 \end{pmatrix} \left[(1-\alpha-\gamma)\beta r^{k+1} + (1-\alpha)\beta B(y^k - y^{k+1})\right] \right\rangle$$
$$+ \left\langle w^{k+1} - w, F(w^{k+1}, \xi_x^{k+1}, \xi_y^{k+1}) \right\rangle. \tag{3.11}$$

With the assertion (3.10), we have

$$\begin{aligned} &\alpha\beta B^\mathsf{T} r^{k+1} + (1-\alpha)\beta B^\mathsf{T} B(y^{k+1} - y^k) \\ =& \frac{\alpha+\gamma-\alpha\gamma}{\alpha+\gamma}\beta B^\mathsf{T} B(y^{k+1} - y^k) - \frac{\alpha}{\alpha+\gamma} B^\mathsf{T}(\lambda^{k+1} - \lambda^k). \end{aligned}$$

Using the definition (2.5) of H, the definition (2.7) of G and the definition of r^{k+1} and $r(w)$, we can rewrite (3.11) as

$$\begin{aligned} &(w - w^{k+1})^\mathsf{T} G(w^{k+1} - w^k) \\ \geq& \left\langle r^{k+1} - r(w), (1-\alpha-\gamma)\beta r^{k+1} + (1-\alpha)\beta B(y^k - y^{k+1}) \right\rangle \\ &+ \left\langle w^{k+1} - w, F(w^{k+1}, \xi_x^{k+1}, \xi_y^{k+1}) \right\rangle. \end{aligned} \tag{3.12}$$

Noting that $r(w) = 0$ for any $w \in \mathcal{D}$, we have from (3.12) that for any $w \in \mathcal{D}$

$$\begin{aligned} &(w - w^{k+1})^\mathsf{T} G(w^{k+1} - w^k) \\ \geq& (1-\alpha-\gamma)\beta\|r^{k+1}\|^2 + (1-\alpha)\beta \left\langle r^{k+1}, B(y^k - y^{k+1}) \right\rangle \\ &+ \left\langle w^{k+1} - w, F(w^{k+1}, \xi_x^{k+1}, \xi_y^{k+1}) \right\rangle. \end{aligned} \tag{3.13}$$

III). Plugging (3.13) into (3.6), we have for any $w \in \mathcal{D}$

$$\begin{aligned} &\|w^k - w\|_{\widehat{G}}^2 - \|w^{k+1} - w\|_{\widehat{G}}^2 \\ \geq& \|x^k - x^{k+1}\|_{S+\frac{1}{2}\Sigma_1}^2 + \|y^k - y^{k+1}\|_{T+\frac{1}{2}\Sigma_2}^2 + (1-\alpha)\beta\|B(y^k - y^{k+1})\|^2 \\ &+ (2-\alpha-\gamma)\beta\|r^{k+1}\|^2 + 2(1-\alpha)\beta \left\langle r^{k+1}, B(y^k - y^{k+1}) \right\rangle + \Delta(w^{k+1}, w), \end{aligned} \tag{3.14}$$

where

$$\Delta(w^{k+1}, w) := 2 \left\langle w^{k+1} - w, F(w^{k+1}, \xi_x^{k+1}, \xi_y^{k+1}) \right\rangle - 2\|u^{k+1} - u\|_\Sigma^2.$$

Taking $w = w^{k+1}$ and $w' = w^*$ in (2.3), we have from (2.3) that

$$\begin{aligned} &\left\langle w^{k+1} - w^*, F(w^{k+1}, \xi_x^{k+1}, \xi_y^{k+1}) \right\rangle \\ \geq& \left\langle w^{k+1} - w^*, F(w^*, \xi_x^*, \xi_y^*) \right\rangle + \|u^{k+1} - u^*\|_\Sigma^2 \geq \|u^{k+1} - u^*\|_\Sigma^2, \end{aligned}$$

where $\xi_x^* \in \partial\theta_1(x^*)$, $\xi_y^* \in \partial\theta_2(y^*)$ and the second inequality is due to the optimality condition (2.4) of w^*. This further means that $\Delta(w^{k+1}, w^*) \geq 0$. Setting $w = w^*$ in (3.14), we have (3.1).

The proof of (3.2) follows directly from (3.1) and

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

which comes from (3.10). The proof is complete. $\square$

We now need to give a careful estimation of the crossing term $\left\langle r^{k+1}, B(y^k - y^{k+1}) \right\rangle$, which is useful to establish the strictly contractive property of $\{\Phi_{\alpha,\gamma}^k(w^*)\}$ when $(\alpha, \gamma) \in \mathbb{D}_1 \cup \mathbb{D}_2$ (see (3.20) for the definition).

Lemma 3.2. *Let the sequence $\{w^k\}$ be generated by iPSPR (1.7). If $\alpha \geq 0$ and $\gamma > 0$, then there holds that*

$$\begin{aligned}
&\left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle \\
\geq\ & \frac{1-\gamma}{1+\alpha}\left\langle r^k, B(y^k - y^{k+1})\right\rangle - \frac{\alpha}{1+\alpha}\|B(y^k - y^{k+1})\|^2 + \frac{1}{1+\alpha}\cdot\frac{1}{\beta}\|y^k - y^{k+1}\|^2_{-2T_-+\Sigma_2} \\
&+ \frac{1}{2(1+\alpha)}\cdot\frac{1}{\beta}\left(\|y^k - y^{k+1}\|^2_{T_++T_-} - \|y^{k-1} - y^k\|^2_{T_++T_-}\right).
\end{aligned} \tag{3.15}$$

Proof. From the optimality conditions of (1.7c) with $k := k-1$, we know that there exist $\xi_y^k \in \partial\theta_2(y^k)$ such that

$$\left\langle y - y^k, T(y^k - y^{k-1}) + \xi_y^k - B^\mathsf{T}\lambda^k + (1-\gamma)\beta B^\mathsf{T} r^k\right\rangle \geq 0, \quad \forall y \in \mathcal{Y}. \tag{3.16}$$

Setting y to be y^k and y^{k+1} in (3.9) respectively, and (3.16) and then rearranging the obtained inequalities suitably, we have that

$$\left\langle B(y^k - y^{k+1}), -\lambda^{k+1} + (1-\gamma)\beta r^{k+1}\right\rangle \geq \|y^k - y^{k+1}\|_T^2 - \left\langle y^k - y^{k+1}, \xi_y^{k+1}\right\rangle, \tag{3.17}$$

$$\left\langle B(y^k - y^{k+1}), \lambda^k - (1-\gamma)\beta r^k\right\rangle \geq -\left\langle y^k - y^{k+1}, T(y^{k-1} - y^k)\right\rangle + \left\langle y^k - y^{k+1}, \xi_y^k\right\rangle. \tag{3.18}$$

Summing (3.17) and (3.18) over the both sides yields

$$\begin{aligned}
&\left\langle B(y^k - y^{k+1}), \lambda^k - \lambda^{k+1}\right\rangle + (1-\gamma)\beta\left\langle B(y^k - y^{k+1}), r^{k+1}\right\rangle \\
&\quad - (1-\gamma)\beta\left\langle B(y^k - y^{k+1}), r^k\right\rangle \\
\geq\ & \|y^k - y^{k+1}\|_T^2 - \left\langle y^k - y^{k+1}, T(y^{k-1} - y^k)\right\rangle + \left\langle y^k - y^{k+1}, \xi_y^k - \xi_y^{k+1}\right\rangle.
\end{aligned} \tag{3.19}$$

Recalling that $T = T_+ - T_-$, we know from the Cauchy-Schwarz inequality that

$$\begin{aligned}
& -\left\langle y^k - y^{k+1}, T(y^{k-1} - y^k)\right\rangle \\
=\ & -\left\langle y^k - y^{k+1}, T_+(y^{k-1} - y^k)\right\rangle + \left\langle y^k - y^{k+1}, T_-(y^{k-1} - y^k)\right\rangle \\
\geq\ & -\frac{1}{2}\|y^k - y^{k+1}\|^2_{T_++T_-} - \frac{1}{2}\|y^{k-1} - y^k\|^2_{T_++T_-},
\end{aligned}$$

which with (2.2) implies that

$$\text{RHS of (3.19)} \geq \frac{1}{2}\left(\|y^k - y^{k+1}\|^2_{T_++T_-} - \|y^{k-1} - y^k\|^2_{T_++T_-}\right) + \|y^k - y^{k+1}\|^2_{-2T_-+\Sigma_2}.$$

This with relations (2.11) and (3.19) implies that (3.15). The proof is complete. $\square$

We now decompose the domain $\mathbb{D}$ (see (1.8) for its definition) as $\mathbb{D} = \mathbb{D}_1 \cup \mathbb{D}_2 \cup \mathbb{D}_3 \cup \mathbb{D}_4$ with

$$\begin{aligned}
&\mathbb{D}_1 = \left\{(\alpha,\gamma) : 0 \leq \alpha < 1 < \gamma < \frac{1-\alpha+\sqrt{(1+\alpha)^2 + 4(1-\alpha^2)}}{2}\right\}, \\
&\mathbb{D}_2 = \{(\alpha,\gamma) : 0 \leq \alpha < 1, \gamma = 1\}, \\
&\mathbb{D}_3 = \{(\alpha,\gamma) : 0 \leq \alpha < 1, 0 \leq \gamma < 1, \alpha + \gamma > 0, \alpha \neq \gamma\}, \\
&\mathbb{D}_4 = \{(\alpha,\gamma) : 0 < \alpha = \gamma < 1\}.
\end{aligned} \tag{3.20}$$

For a given $w \in \mathcal{D}$, we define $\Phi^k_{\alpha,\gamma}(w)$ as

$$\Phi^k_{\alpha,\gamma}(w) := \|w^k - w\|^2_{\widehat{G}} + \rho_1^{\alpha,\gamma}\|y^{k-1} - y^k\|^2_{T_++T_-} + \rho_2^{\alpha,\gamma}\beta\|r^k\|^2, \tag{3.21}$$

where the constants

$$\rho_1^{\alpha,\gamma}=\begin{cases}\dfrac{1-\alpha}{1+\alpha}, & (\alpha,\gamma)\in\mathbb{D}_1,\\ \dfrac{1-\alpha}{2(1+\alpha)}, & (\alpha,\gamma)\in\mathbb{D}_2,\\ 0, & (\alpha,\gamma)\in\mathbb{D}_3\cup\mathbb{D}_4,\end{cases}\qquad \rho_2^{\alpha,\gamma}=\begin{cases}\dfrac{(\gamma-1)^2}{(1-c^{\alpha,\gamma})(1+\alpha)}, & (\alpha,\gamma)\in\mathbb{D}_1,\\ 0, & (\alpha,\gamma)\in\mathbb{D}_2\cup\mathbb{D}_3\cup\mathbb{D}_4,\end{cases} \tag{3.22}$$

in which the constant $c^{\alpha,\gamma}$ is defined as

$$c^{\alpha,\gamma}\in\begin{cases}\left(0,\dfrac{1-\alpha^2+\alpha-(\alpha-1)\gamma-\gamma^2}{(2-\alpha-\gamma)(1+\alpha)}\right), & (\alpha,\gamma)\in\mathbb{D}_1,\\ (0,1), & (\alpha,\gamma)\in\mathbb{D}_2\cup\mathbb{D}_3.\end{cases}$$

We are now ready to have the following theorem.

Theorem 3.1. *Given $w^*\in\Omega^*$, let the sequence $\{w^k\}$ be generated by iPSPR* (1.7). *If we choose $(\alpha,\gamma)\in\mathbb{D}$, then there holds that*

$$\begin{aligned}&\Phi^k_{\alpha,\gamma}(w^*)-\Phi^{k+1}_{\alpha,\gamma}(w^*)\\ &\geq\|x^k-x^{k+1}\|^2_{S+\frac{1}{2}\Sigma_1}+\|y^k-y^{k+1}\|^2_{T+\frac{1}{2}\Sigma_2+\kappa_1^{\alpha,\gamma}(-2T_-+\Sigma_2)+\kappa_2^{\alpha,\gamma}\beta B^\intercal B}\\ &\quad+\frac{\kappa_3^{\alpha,\gamma}}{\beta}\|\lambda^k-\lambda^{k+1}\|^2+\kappa_4^{\alpha,\gamma}\|r^{k+1}\|^2,\end{aligned} \tag{3.23}$$

where the constants

$$\kappa_1^{\alpha,\gamma}=\begin{cases}\dfrac{2(1-\alpha)}{1+\alpha}, & (\alpha,\gamma)\in\mathbb{D}_1,\\ \dfrac{1-\alpha}{1+\alpha}, & (\alpha,\gamma)\in\mathbb{D}_2,\\ 0, & (\alpha,\gamma)\in\mathbb{D}_3\cup\mathbb{D}_4,\end{cases}\qquad \kappa_2^{\alpha,\gamma}=\begin{cases}\dfrac{c^{\alpha,\gamma}(1-\alpha)^2}{1+\alpha}, & (\alpha,\gamma)\in\mathbb{D}_1,\\ \dfrac{c^{\alpha,\gamma}(1-\alpha)(3-\alpha)}{4(1+\alpha)}, & (\alpha,\gamma)\in\mathbb{D}_2,\\ \dfrac{c^{\alpha,\gamma}(1-\alpha)(1-\gamma)}{(2-\gamma-\alpha)}, & (\alpha,\gamma)\in\mathbb{D}_3,\\ \dfrac{1-\alpha}{2}, & (\alpha,\gamma)\in\mathbb{D}_4,\end{cases} \tag{3.24}$$

$$\kappa_3^{\alpha,\gamma}=\begin{cases}0, & (\alpha,\gamma)\in\mathbb{D}_1\cup\mathbb{D}_2,\\ \dfrac{(1-c^{\alpha,\gamma})(1-\alpha)(1-\gamma)(2-\alpha-\gamma)}{(\gamma-\alpha)^2+(1-c^{\alpha,\gamma})(1-\alpha)(1-\gamma)(\alpha+\gamma)^2}, & (\alpha,\gamma)\in\mathbb{D}_3,\\ \dfrac{1-\alpha}{2\alpha^2}, & (\alpha,\gamma)\in\mathbb{D}_4,\end{cases} \tag{3.25}$$

$$\kappa_4^{\alpha,\gamma}=\begin{cases}2-\alpha-\gamma-\dfrac{(\gamma-1)^2}{(1-c^{\alpha,\gamma})(1+\alpha)}, & (\alpha,\gamma)\in\mathbb{D}_1,\\ \dfrac{(1-c^{\alpha,\gamma})(1-\alpha)(3-\alpha)}{(1+\alpha)+(1-c^{\alpha,\gamma})(3-\alpha)}, & (\alpha,\gamma)\in\mathbb{D}_2,\\ 0, & (\alpha,\gamma)\in\mathbb{D}_3\cup\mathbb{D}_4.\end{cases} \tag{3.26}$$

Proof. We consider four cases.

I). $(\alpha,\gamma)\in\mathbb{D}_1$. By combining (3.15) and (3.1), we derive

$$\left(\|w^k-w^*\|^2_{\widehat{G}}+\frac{1-\alpha}{1+\alpha}\|y^{k-1}-y^k\|^2_{T_++T_-}\right)-\left(\|w^{k+1}-w^*\|^2_{\widehat{G}}+\frac{1-\alpha}{1+\alpha}\|y^k-y^{k+1}\|^2_{T_++T_-}\right)$$

$$
\begin{aligned}
&\geq \|x^k - x^{k+1}\|^2_{S+\frac{1}{2}\Sigma_1} + \|y^k - y^{k+1}\|^2_{T+\frac{1}{2}\Sigma_2} + \frac{2(1-\alpha)}{1+\alpha}\|y^k - y^{k+1}\|^2_{-2T_-+\Sigma_2} \\
&\quad + \frac{(1-\alpha)^2}{1+\alpha}\beta\|B(y^k - y^{k+1})\|^2 + (2-\alpha-\gamma)\beta\|r^{k+1}\|^2 \\
&\quad -2(\gamma-1)\frac{1-\alpha}{1+\alpha}\beta\left\langle r^k, B(y^k - y^{k+1})\right\rangle.
\end{aligned} \tag{3.27}
$$

Note that in this case $0 < c^{\alpha,\gamma} < \frac{1-\alpha^2+\alpha-(\alpha-1)\gamma-\gamma^2}{(2-\alpha-\gamma)(1+\alpha)} < 1$, with the Cauchy-Schwarz inequality, we have

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

Plugging the above inequality into (3.27), we obtain (3.23) in this case.

II). $(\alpha,\gamma)\in\mathbb{D}_2$. For this case, (3.15) reduces to

$$
\begin{aligned}
\left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle \geq\ & -\frac{\alpha}{1+\alpha}\|B(y^k - y^{k+1})\|^2 + \frac{1}{(1+\alpha)\beta}\|y^k - y^{k+1}\|^2_{-2T_-+\Sigma_2} \\
& + \frac{1}{2(1+\alpha)\beta}\left(\|y^k - y^{k+1}\|^2_{T_++T_-} - \|y^{k-1} - y^k\|^2_{T_++T_-}\right).
\end{aligned} \tag{3.28}
$$

On the other hand, by the Cauchy-Schwartz inequality, we have

$$
\left\langle r^{k+1}, B(y^k - y^{k+1})\right\rangle \geq -\delta\|B(y^k - y^{k+1})\|^2 - \frac{1}{4\delta}\|r^{k+1}\|^2, \tag{3.29}
$$

where

$$
\delta = \frac{(1+\alpha) + (1-c^{\alpha,\gamma})(3-\alpha)}{4(1+\alpha)}.
$$

Combing (3.28) and (3.29), and using (3.1), we have

$$
\begin{aligned}
&\left(\|w^k - w^*\|^2_{\widehat{G}} + \frac{1-\alpha}{2(1+\alpha)}\|y^{k-1} - y^k\|^2_{T_++T_-}\right) \\
&\quad - \left(\|w^{k+1} - w^*\|^2_{\widehat{G}} + \frac{1-\alpha}{2(1+\alpha)}\|y^k - y^{k+1}\|^2_{T_++T_-}\right) \\
\geq\ & \|x^k - x^{k+1}\|^2_{S+\frac{1}{2}\Sigma_1} + \|y^k - y^{k+1}\|^2_{T+\frac{1}{2}\Sigma_2} + \frac{1-\alpha}{1+\alpha}\|y^k - y^{k+1}\|^2_{-2T_-+\Sigma_2} \\
&\quad + \frac{c^{\alpha,\gamma}(1-\alpha)(3-\alpha)}{4(1+\alpha)}\beta\|B(y^k - y^{k+1})\|^2 + \frac{(1-c^{\alpha,\gamma})(1-\alpha)(3-\alpha)}{(1+\alpha)+(1-c^{\alpha,\gamma})(3-\alpha)}\|r^{k+1}\|^2.
\end{aligned} \tag{3.30}
$$

This means (3.23) holds in this case.

III). $(\alpha,\gamma)\in\mathbb{D}_3$. Noting that $c^{\alpha,\gamma}\in(0,1)$ and letting

$$
\tilde{c} = \frac{(\gamma-\alpha)^2}{(\gamma-\alpha)^2 + (1-c^{\alpha,\gamma})(1-\alpha)(1-\gamma)(\alpha+\gamma)^2},
$$

we have from the Cauchy-Schwarz inequality that

$$
\begin{aligned}
&2(\gamma-\alpha)\left\langle B(y^k - y^{k+1}), \lambda^k - \lambda^{k+1}\right\rangle \\
&\geq -\frac{(\alpha-\gamma)^2\beta}{\tilde{c}(2-\gamma-\alpha)}\|B(y^k - y^{k+1})\|^2 - \frac{\tilde{c}(2-\gamma-\alpha)}{\beta}\|\lambda^k - \lambda^{k+1}\|^2,
\end{aligned}
$$

which with (3.2) and the equality

$$\left[\alpha^2(1-\gamma)+\gamma^2(1-\alpha)\right](2-\alpha-\gamma)=(\gamma-\alpha)^2+(1-\alpha)(1-\gamma)(\alpha+\gamma)^2$$

implies that (3.23) holds in this case.

IV). $(\alpha,\gamma)\in\mathbb{D}_4$. Note that $\alpha=\gamma$ in this case. It is easy to see from (3.2) that

$$\begin{aligned}&\|w^k-w^*\|_{\widehat{G}}^2-\|w^{k+1}-w^*\|_{\widehat{G}}^2\\ \geq\;&\|x^k-x^{k+1}\|^2_{S+\frac{1}{2}\Sigma_1}+\|y^k-y^{k+1}\|^2_{T+\frac{1}{2}\Sigma_2}\\ &+\frac{1-\alpha}{2}\beta\|B(y^k-y^{k+1})\|^2+\frac{1-\alpha}{2\alpha^2\beta}\left\|\lambda^k-\lambda^{k+1}\right\|^2,\end{aligned}\tag{3.31}$$

which means that (3.23) holds in this case. The proof is complete. □

3.2. Global convergence

We are now ready to state the global convergence results formally.

Theorem 3.2. *Let the sequence $\{w^k\}$ be generated by iPSPR* (1.7). *If the stepsizes $(\alpha,\gamma)\in\mathbb{D}$ and the proximal terms S, T are chosen such that*

$$S+\frac{1}{2}\Sigma_1\succeq 0,\quad S+\frac{1}{2}\Sigma_1+\beta A^\mathsf{T}A\succ 0,\tag{3.32}$$

$$T+\Sigma_2+(1-\alpha)\beta B^\mathsf{T}B\succ 0,\quad T+\frac{1}{2}\Sigma_2+\kappa_1^{\alpha,\gamma}(-2T_-+\Sigma_2)+\kappa_2^{\alpha,\gamma}\beta B^\mathsf{T}B\succ 0,\tag{3.33}$$

then $\{w^k\}$ converges to an optimal solution of (1.1).

Proof. The first conditions in (3.32) and (3.33) guarantee $\hat{G}\succeq 0$ and $\hat{H}\succ 0$, see Proposition 2.1. We divide the proof into three steps.

I) We show that the sequences $\{w^k\}$ is bounded. It is straightforward to see from (3.23), (3.32) and (3.33) that $\Phi^k_{\alpha,\gamma}(w^*)$ is monotone decreasing. This with $T_+,T_-\succeq 0$ and the definition (3.21) means that $\|w^k-w^*\|^2_{\widehat{G}}$ is bounded. With the second equality of (2.10), we have $\|w^k-w^*\|^2_{\widehat{G}}=\|x^k-x^*\|^2_{S+\Sigma_1}+\|v^k-v^*\|^2_{\hat{H}}$, which means that $\|x^k-x^*\|_{S+\Sigma_1}$ and $\|v^k-v^*\|_{\hat{H}}$ are all bounded. Besides, with the positiveness of $\hat{H}$, we know that the sequences $\{\lambda^k\}$ and $\{y^k\}$ are bounded. Following from (3.23), (3.32) and (3.33), we also have

$$\lim_{k\to\infty}\frac{\kappa_3^{\alpha,\gamma}}{\beta}\|\lambda^k-\lambda^{k+1}\|^2+\kappa_4^{\alpha,\gamma}\|r^{k+1}\|^2=0.\tag{3.34}$$

Noting that $\kappa_3^{\alpha,\gamma}+\kappa_4^{\alpha,\gamma}>0$, with (2.11) and (3.34) and the boundness of y^k, we can see that $\{r^k\}$ is bounded. With the definition of r^k, we know that $\|Ax^k-Ax^*\|=\|r^k-B(y^k-y^*)\|\leq\|r^k\|+\|B(y^k-y^*)\|$, which with the boundness of r^k and y^k implies that $\|x^k-x^*\|_{\beta A^\mathsf{T}A}$ is bounded. Recalling that $S+\frac{1}{2}\Sigma_1+\beta A^\mathsf{T}A\succ 0$ and $\|x^k-x^*\|_{S+\Sigma_1}$ is bounded, it is safe to say that $\{x^k\}$ is also bounded.

II) We argue 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 $\lim_{k_i\to\infty}w^{k_i}=w^\infty$. Following from (3.23), (3.32) and (3.33), we have

$$\lim_{k\to\infty}\|x^k-x^{k+1}\|_{S+\frac{1}{2}\Sigma_1}=\lim_{k\to\infty}\|y^k-y^{k+1}\|_{T+\frac{1}{2}\Sigma_2+\kappa_1^{\alpha,\gamma}(-2T_-+\Sigma_2)+\kappa_2^{\alpha,\gamma}\beta B^\mathsf{T}B}=0.\tag{3.35}$$

With the second condition on T in (3.33), we know from the second equality in (3.35) that

$$\lim_{k\to\infty} \|y^k - y^{k+1}\| = 0. \tag{3.36}$$

Again using $\kappa_3^{\alpha,\gamma} + \kappa_4^{\alpha,\gamma} > 0$, with (2.11) and (3.34), it is easy to see that

$$\lim_{k\to\infty} \|r^k\| = \lim_{k\to\infty} \|\lambda^k - \lambda^{k+1}\| = 0. \tag{3.37}$$

On the other hand, with the definition of r^k, we have

$$A(x^k - x^{k+1}) = r^k - r^{k+1} - B(y^k - y^{k+1}).$$

Therefore, we know from (3.36) and (3.37) that $\lim_{k\to\infty} \|A(x^k - x^{k+1})\| = 0$, which with the first equality in (3.35) implies

$$\lim_{k\to\infty} \|x^k - x^{k+1}\|_{S+\frac{1}{2}\Sigma_1+\beta A^T A} = 0.$$

This with the second condition on S in (3.32) implies

$$\lim_{k\to\infty} \|x^k - x^{k+1}\| = 0. \tag{3.38}$$

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 (3.11) and by using (3.36)–(3.38), we know that there exists ξ_x^∞ and ξ_y^∞ such that

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

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

III) We finally prove that the sequence $\{w^k\}$ has only one cluster point. We first replace w^* with w^∞ in the analysis of Steps I) and II). It follows from $\lim_{k_i\to\infty} w^{k_i} = w^\infty$ and (3.36), (3.37) that $\lim_{k_i\to\infty} \Phi_{\alpha,\gamma}^{k_i}(w^\infty) = 0$. Owing to the decreasing monotonicity of the sequence $\Phi_{\alpha,\gamma}^{k}(w^\infty)$, we can see that

$$\lim_{k\to\infty} \Phi_{\alpha,\gamma}^{k}(w^\infty) = 0.$$

This together with $T_+, T_- \succeq 0$ and

$$\|w^k - w^\infty\|_{\widehat{G}}^2 = \|x^k - x^\infty\|_{S+\Sigma_1}^2 + \|v^k - v^\infty\|_{\hat{H}}^2$$

and $\hat{H} \succ 0$ shows that

$$\lim_{k\to\infty} \|x^k - x^\infty\|_{S+\Sigma_1} = \lim_{k\to\infty} \|y^k - y^\infty\| = \lim_{k\to\infty} \|\lambda^k - \lambda^\infty\|. \tag{3.39}$$

With (2.12), we further have $\lim_{k\to\infty} \|r^k\| = 0$. Using again the inequality

$$\|Ax^k - Ax^\infty\| = \|r^k - B(y^k - y^\infty)\| \leq \|r^k\| + \|B(y^k - y^\infty)\|,$$

which with (3.39) and (3.37) implies

$$\lim_{k\to\infty} \|A(x^k - x^\infty)\| = 0. \tag{3.40}$$

Combing (3.39) and (3.40), and using that $S + \frac{1}{2}\Sigma_1 + \beta A^\mathsf{T} A \succ 0$, we immediately have

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

The proof is complete. $\square$

Remark 3.1. If the condition (3.32) is replaced by $S \succeq 0$, we can have from $\lim_{k\to\infty} \|x^k - x^{k+1}\|_{S+\frac{1}{2}\Sigma_1} = 0$ that $\lim_{k\to\infty} S(x^k - x^{k+1}) = 0$. Using similar analysis to the above proof, we can show that $\{v^k\}$ converges to some $v^* = (y^*, \lambda^*)^\tau$, where $w^* = (x^*, v^*)^\tau$ is an optimal solution of problem (1.1).

3.3. Choices of proximal terms

When the proximal terms S and T satisfy conditions (3.32) and (3.33), it is easy to see that the objective functions of subproblems (1.7a) and (1.7c) are strongly convex, which make the corresponding problems more easier to solve. Note that by allowing S or T indefinite, we can always take a larger step on updating the variable x or y. Besides, we next show that for some special cases, with particularly chosen proximal term T, the subproblem (1.7c) is easy to solve or even takes a closed form solution. Note that the discussion for the proximal term S is omitted since it is similar.

We consider to choose T as

$$T = rI - \left(\Sigma_2 + \beta B^\mathsf{T} B\right) \quad \text{with} \quad r = \lambda_{\max}\left(\frac{1}{2}\Sigma_2 + \tau\beta B^\mathsf{T} B\right), \tag{3.41}$$

where $\tau \in (0, 1]$. We decompose $T = T_+ - T_-$ with

$$\begin{aligned} T_+ &= rI - \left(\frac{1}{2}\Sigma_2 + \tau\beta B^\mathsf{T} B\right), \\ T_- &= \frac{1}{2}\Sigma_2 + (1-\tau)\beta B^\mathsf{T} B. \end{aligned}$$

Note that $T_+, T_- \succeq 0$. By some direct calculations, we have

$$T + \Sigma_2 + (1-\alpha)\beta B^\mathsf{T} B = rI - \alpha\beta B^\mathsf{T} B \tag{3.42}$$

and

$$\begin{aligned} &T + \frac{1}{2}\Sigma_2 + \kappa_1^{\alpha,\gamma}(-2T_- + \Sigma_2) + \kappa_2^{\alpha,\gamma}\beta B^\mathsf{T} B \\ =& rI - \left(\frac{1}{2}\Sigma_2 + (1 + 2\kappa_1^{\alpha,\gamma}(1-\tau) - \kappa_2^{\alpha,\gamma})\beta B^\mathsf{T} B\right). \end{aligned} \tag{3.43}$$

For given $(\alpha, \gamma) \in \mathbb{D}$ and a fixed $c^{\alpha,\gamma}$, by (3.42) and (3.43), we know that if we choose $\tau > \alpha$ and $\tau > 1 - \frac{\kappa_2^{\alpha,\gamma}}{1+2\kappa_1^{\alpha,\gamma}}$, then (3.33) must hold. Note that the number $1 - \frac{\kappa_2^{\alpha,\gamma}}{1+2\kappa_1^{\alpha,\gamma}}$ is decreasing with respect to $c^{\alpha,\gamma}$ which is defined over an open interval. Hence, we can argue that if

$$1 \ge \tau > \max\left\{\alpha, \inf_{c^{\alpha,\gamma}}\left\{1 - \frac{\kappa_2^{\alpha,\gamma}}{1 + 2\kappa_1^{\alpha,\gamma}}\right\}\right\},$$

namely,

$$1 \ge \tau > \underline{\tau}^{\alpha,\gamma} := \begin{cases} 1 - (1-\alpha)^2 \dfrac{1 - \alpha^2 - (\gamma - 1)(\alpha + \gamma)}{(2 - \alpha - \gamma)(1 + \alpha)(5 - 3\alpha)}, & (\alpha,\gamma) \in \mathbb{D}_1, \\ \dfrac{3+\alpha}{4}, & (\alpha,\gamma) \in \mathbb{D}_2, \\ \dfrac{1-\alpha\gamma}{2-\alpha-\gamma}, & (\alpha,\gamma) \in \mathbb{D}_3, \\ \dfrac{1+\alpha}{2}, & (\alpha,\gamma) \in \mathbb{D}_4, \end{cases} \tag{3.44}$$

then (3.33) must hold.

Consider the case when $\theta_2(y) = \frac{1}{2}y^\top My + h(y)$, where M is symmetric positive semidefinite and the nonsmooth convex function $h(y)$ is simple in the sense that $\min_{y\in\mathcal{Y}} h(y) + \frac{1}{2}\|y-d\|^2$ is easy to compute. Here $d \in \mathbb{R}^{n_2}$ is a given vector. In this case, we have $\Sigma_2 = M$ and the subproblem (1.7b) with T chosen according to (3.41) and (3.44) takes the following form:

$$y^{k+1} = \arg\min_{y\in\mathcal{Y}}\ h(y) + \frac{1}{2}\|y - d^k\|^2$$

where

$$d^k = Ty^k + B^\top\left(\lambda^{k+\frac{1}{2}} - \beta(Ax^{k+1} - b)\right).$$

To end this subsection, some comments are listed in order. Firstly, if $\alpha = 0, \gamma = 1$ and $\Sigma_2 = 0$, (3.44) becomes $0.75 < \tau \le 1$, which recovers the optimal bound of τ for the linearized ADMM in [19]; if $\alpha \in (0,1), \gamma = 1$ and $\Sigma_2 = 0$, (3.44) becomes $(3+\alpha)/4 < \tau \le 1$, which partially recovers the optimal bound of τ for the linearized version of the generalized ADMM in [23]. Note that in [23], they allowed $\alpha \in (-1,1)$. Secondly, if $(\alpha,\gamma) \in \mathbb{D}_2 \cup \mathbb{D}_3$, it is easy to see that $\frac{1-\alpha\gamma}{2-\alpha-\gamma} \ge \frac{2+\alpha+\gamma}{4}$ and the equality holds if and only if $\alpha = \gamma$, namely, $(\alpha,\gamma) \in \mathbb{D}_4$. Thirdly, when the subproblem (1.7c) does not take a closed form solution, as done in [10,25,35], we can consider the majorized version of iPSPR. The techniques for constructing the indefinite proximal T in [3,25] can be explored to construct T. We leave this for future investigation.

4. Sublinear Convergence of iPSPR

The rate of convergence of an algorithm can help us have a deeper understanding of the algorithm. In this section, motivated by [3,6,15,25,35], we establish the $o(1/t)$ sublinear rate of convergence of iPSPR in the nonergodic sense.

We first give a new optimality condition of (1.1) as follows.

Lemma 4.1. *Let the sequence $\{w^k\}$ be generated by iPSPR* (1.7). *We choose $(\alpha,\gamma) \in \mathbb{D}$ and the proximal terms S, T are chosen such that* (3.32) *and* (3.33) *hold. Then $w^{k+1} \in \Omega^*$, namely, w^{k+1} is one optimal solution of* (1.1), *if*

$$\|w^k - w^{k+1}\|_{\widehat{G}} = 0.$$

Proof. The proof is similar to the second part of the proof of Theorem 3.2, we omit the details here. ☐

Following from (2.6), (2.10) and (2.12), we have

$$\|w^k - w^{k+1}\|^2_{\widehat{G}} = \|x^k - x^{k+1}\|^2_{S+\Sigma_1} + \|y^k - y^{k+1}\|^2_{T+\Sigma_2+(1-\alpha)\beta B^TB} + (\alpha+\gamma)\beta\|r^{k+1}\|^2.$$

Hence, Lemma 4.1 provides a practical stopping condition for iPSPR (1.7), which is shown as

$$\max\left\{\|x^k - x^{k+1}\|_{S+\Sigma_1}, \|y^k - y^{k+1}\|^2_{T+\Sigma_2+(1-\alpha)\beta B^TB}, \beta\|r^{k+1}\|^2\right\} \le \text{tol}, \tag{4.1}$$

where tol is some tolerance.

Theorem 4.1. *Let the sequence $\{w^k\}$ be generated by iPSPR* (1.7) *with $(\alpha,\gamma) \in \mathbb{D}$. Suppose that the proximal terms S, T are chosen such that* (3.32), (3.33) *and*

$$S + \frac{1}{2}\Sigma_1 \succeq \frac{1}{2}c\Sigma_1 \tag{4.2}$$

hold, where c is a positive constant. We have that

$$\min_{1\le i\le t}\|w^i - w^{i+1}\|^2_{\widehat{G}} = o(1/t). \tag{4.3}$$

Proof. With conditions (4.2) on S, we know that $S+\Sigma_1 \preceq (1+c^{-1})(S+\frac{1}{2}\Sigma_1)$. With the condition (3.33) on T, we know that there exists a positive constant c_1 such that

$$\hat{H} \preceq c_1 \begin{pmatrix} T+\frac{1}{2}\Sigma_2+\kappa_1^{\alpha,\gamma}(-2T_- +\Sigma_2)+\kappa_2^{\alpha,\gamma}\beta B^\top B & 0 \\ 0 & \frac{\bar{\kappa}_3^{\alpha,\gamma}}{\beta} I \end{pmatrix},$$

which with (2.8) implies that

$$\widehat{G} \preceq \max\{1+c^{-1}, c_1\} \begin{pmatrix} S+\frac{1}{2}\Sigma_1 & 0 & 0 \\ 0 & T+\frac{1}{2}\Sigma_2+\kappa_1^{\alpha,\gamma}(-2T_- +\Sigma_2)+\kappa_2^{\alpha,\gamma}\beta B^\top B & 0 \\ 0 & 0 & \frac{\bar{\kappa}_3^{\alpha,\gamma}}{\beta} I \end{pmatrix}.$$

This with (3.23) implies that

$$\Phi^k_{\alpha,\gamma}(w^*) - \Phi^{k+1}_{\alpha,\gamma}(w^*) \ge \frac{1}{\max\{c, c_1\}}\|w^k - w^{k+1}\|^2_{\widehat{G}}. \tag{4.4}$$

Summing (4.4) over $k = 1, \ldots, +\infty$ leads to

$$\frac{1}{c}\cdot\sum_{k=1}^{+\infty}\|w^{k+1} - w^k\|^2_{\widehat{G}} \le \Phi^1_{\alpha,\gamma}(w^*). \tag{4.5}$$

Using Lemma 3 in [25], we have (4.3). □

Now we show that if $(\alpha,\gamma) \in \mathbb{D}_2\cup\mathbb{D}_3\cup\mathbb{D}_4$ and some additional requirement is made on T, we can have a stronger result than (4.3). We first show that the sequence $\{\|w^k - w^{k+1}\|^2_{\widehat{G}}\}$ is non-increasing.

Lemma 4.2. *Let the sequence $\{w^k\}$ be generated by iPSPR* (1.7)*. If $(\alpha,\gamma) \in \mathbb{D}_2\cup\mathbb{D}_3\cup\mathbb{D}_4$ and the proximal terms S, T are chosen such that* (3.32)*,* (3.33) *and*

$$T + \frac{1}{2}\Sigma_2 + \frac{(1-\alpha)(1-\gamma)}{2-\alpha-\gamma}\beta B^\top B \succeq 0 \tag{4.6}$$

hold, then there holds that

$$\|w^k - w^{k+1}\|^2_{\widehat{G}} \ge \|w^{k+1} - w^{k+2}\|^2_{\widehat{G}}. \tag{4.7}$$

Proof. Note that (3.12) also holds with $k := k+1$, then we have

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

where $\xi_x^{k+2} \in \partial\theta_1(x^{k+2})$ and $\xi_y^{k+2} \in \partial\theta_2(y^{k+2})$. Choosing w to be w^{k+2} and w^{k+1}, respectively, in (3.12) and (4.8) leads to

$$\begin{aligned} &(w^{k+2} - w^{k+1})^\top G(w^{k+1} - w^k) \\ &\ge \left\langle r^{k+1} - r^{k+2}, (1-\alpha-\gamma)\beta r^{k+1} + (1-\alpha)\beta B(y^k - y^{k+1})\right\rangle \\ &\quad + \left\langle w^{k+1} - w^{k+2}, F(w^{k+1}, \xi_x^{k+1}, \xi_y^{k+1})\right\rangle. \end{aligned} \tag{4.9}$$

and

$$\begin{aligned}&(w^{k+1}-w^{k+2})^{\mathsf T}G(w^{k+2}-w^{k+1})\\ \geq&\left\langle r^{k+2}-r^{k+1},(1-\alpha-\gamma)\beta r^{k+2}+(1-\alpha)\beta B(y^{k+1}-y^{k+2})\right\rangle\\ &+\left\langle w^{k+2}-w^{k+1},F(w^{k+2},\xi_x^{k+2},\xi_y^{k+2})\right\rangle.\end{aligned}\tag{4.10}$$

Adding (4.9) and (4.10) and noting

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

which follows from (2.3), we obtain that

$$\begin{aligned}&(w^{k+2}-w^{k+1})^{\mathsf T}G\left[(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\right]\\ \geq&(1-\alpha-\gamma)\beta\|r^{k+1}-r^{k+2}\|^2+(1-\alpha)\beta\left\langle B\left[(y^k-y^{k+1})-(y^{k+1}-y^{k+2})\right],r^{k+1}-r^{k+2}\right\rangle.\\ &+\|u^{k+2}-u^{k+1}\|_\Sigma^2.\end{aligned}\tag{4.11}$$

Following the deriving process of (2.9) and (2.12), we have

$$\begin{aligned}&\|(w^k-w^{k+1})-(w^{k+1}-w^{k+2})\|_G^2\\ =&\|(u^k-u^{k+1})-(u^{k+1}-u^{k+2})\|_P^2+(1-\alpha)\beta\left\|B[(y^k-y^{k+1})-(y^{k+1}-y^{k+2})]\right\|^2\\ &+(\alpha+\gamma)\beta\|r^{k+1}-r^{k+2}\|^2.\end{aligned}\tag{4.12}$$

Thus we conclude that

$$\begin{aligned}&\|w^k-w^{k+1}\|_{\widehat G}^2-\|w^{k+1}-w^{k+2}\|_{\widehat G}^2\\ =&\left(\|w^k-w^{k+1}\|_G^2+\|u^k-u^{k+1}\|_\Sigma^2\right)-\left(\|w^{k+1}-w^{k+2}\|_G^2+\|u^{k+1}-u^{k+2}\|_\Sigma^2\right)\\ =&2(w^{k+2}-w^{k+1})^{\mathsf T}G\left[(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\right]+\|(w^{k+1}-w^k)-(w^{k+2}-w^{k+1})\|_G^2\\ &+\|u^k-u^{k+1}\|_\Sigma^2-\|u^{k+1}-u^{k+2}\|_\Sigma^2\\ \geq&(2-\alpha-\gamma)\beta\|r^{k+1}-r^{k+2}\|^2+2(1-\alpha)\beta\left\langle B\left[(y^k-y^{k+1})-(y^{k+1}-y^{k+2})\right],r^{k+1}-r^{k+2}\right\rangle\\ &+(1-\alpha)\beta\left\|B[(y^k-y^{k+1})-(y^{k+1}-y^{k+2})]\right\|^2+\|(u^k-u^{k+1})-(u^{k+1}-u^{k+2})\|_P^2\\ &+\|u^k-u^{k+1}\|_\Sigma^2+\|u^{k+1}-u^{k+2}\|_\Sigma^2\\ \geq&\frac{(1-\alpha)(1-\gamma)}{2-\alpha-\gamma}\beta\left\|B[(y^k-y^{k+1})-(y^{k+1}-y^{k+2})]\right\|^2+\|(u^k-u^{k+1})-(u^{k+1}-u^{k+2})\|_P^2\\ &+\frac12\|(u^k-u^{k+1})-(u^{k+1}-u^{k+2})\|_\Sigma^2\\ =&\|(x^k-x^{k+1})-(x^{k+1}-x^{k+2})\|_{S+\frac12\Sigma_1}^2+\|(y^k-y^{k+1})-(y^{k+1}-y^{k+2})\|_{T+\frac12\Sigma_2+\frac{(1-\alpha)(1-\gamma)}{2-\alpha-\gamma}\beta B^{\mathsf T}B}^2\\ \geq&0,\end{aligned}\tag{4.13}$$

where the first inequality is due to (4.11) and (4.12), the second inequality follows from the Cauchy-Schwarz inequality and the last inequality is due to $P=\begin{pmatrix}S&0\\0&T\end{pmatrix}$, $S+\frac12\Sigma_1\succeq0$ and (4.6). The proof is complete. □

Theorem 4.2. *Let the sequence* $\{w^k\}$ *be generated by iPSPR* (1.7) *with* $(\alpha,\gamma)\in\mathbb D_2\cup\mathbb D_3\cup\mathbb D_4$. *Suppose that the proximal term* S *is chosen according to* (3.32) *and* (4.2) *and the proximal term* T *is chosen according to* (3.33) *and* (4.6). *We have*

$$\|w^t-w^{t+1}\|_{\widehat G}^2=o(1/t).\tag{4.14}$$

Proof. It follows from Theorem 4.1 that $\min_{1\le i\le t}\|w^i - w^{i+1}\|^2_{\widehat{G}} = o(1/t)$, which with (4.7) implies (4.14). The proof is complete. $\square$

Define the KKT mapping $R\colon \Omega \to \Omega$ as

$$R(w) = \begin{pmatrix} x - \mathrm{Pr}_p[x - (\nabla f(x) - A^\top\lambda)] \\ y - \mathrm{Pr}_h[y - (\nabla g(y) - B^\top\lambda)] \\ Ax + By - b \end{pmatrix}, \qquad \forall w \in \Omega, \tag{4.15}$$

where $\mathrm{Pr}_p(\cdot)$ denotes the Moreau-Yosida proximal mapping [31] defined as

$$\mathrm{Pr}_p(z) := \arg\min_{x\in\mathcal{X}} \left\{ p(x) + \frac{1}{2}\|x - z\|^2 \right\},$$

and $\mathrm{Pr}_h(\cdot)$ is defined accordingly. It is well known that (see [3, 15, 35] for instance)

$$\forall w \in \Omega, \quad R(w) = 0 \Longleftrightarrow w \in \Omega^*. \tag{4.16}$$

Inspired by [15, Lemma 1] and [35, Lemma 3.1], we now characterize the relations between $\|R(w^{k+1})\|^2$ and $\|w^{k+1} - w^k\|^2_{\widehat{G}}$.

Lemma 4.3. *Let the sequence $\{w^k\}$ be generated by iPSPR (1.7). Then there exists a positive constant ϱ such that for any $k \ge 0$,*

$$\|R(w^{k+1})\|^2 \le \varrho\|w^{k+1} - w^k\|^2_{\widehat{G}}. \tag{4.17}$$

Proof. From the optimality condition for (1.7) and $\lambda^{k+\frac{1}{2}} = \lambda^{k+1} + \gamma\beta r^{k+1}$, we have

$$x^{k+1} = \mathrm{Pr}_p\left[x^{k+1} - \left(\nabla f(x^{k+1}) - A^\top\lambda^k + \beta A^\top r^{k+1} + \beta A^\top B(y^k - y^{k+1}) + S(x^{k+1} - x^k)\right)\right], \tag{4.18}$$

$$y^{k+1} = \mathrm{Pr}_h\left[y^{k+1} - \left(\nabla g(y^{k+1}) - B^\top\lambda^{k+1} + (1-\gamma)\beta B^\top r^{k+1} + T(y^{k+1} - y^k)\right)\right]. \tag{4.19}$$

Substituting the Eqs. (4.18) and (4.19) into $R(w^{k+1})$ and noting that the Moreau-Yosida proximal mapping is globally Lipschitz continuous with modulus one, we get

$$\begin{aligned}
\|R(w^{k+1})\| \le\ & \left\|A^\top(\lambda^{k+1} - \lambda^k) + \beta A^\top r^{k+1} + \beta A^\top B(y^k - y^{k+1}) + S(x^{k+1} - x^k)\right\| \\
& + \left\|(1-\gamma)\beta B^\top r^{k+1} + T(y^{k+1} - y^k)\right\| + \|r^{k+1}\| \\
\le\ & \|A\|\|\lambda^{k+1} - \lambda^k\| + \left((\|A\| + |1-\gamma|\|B\|)\beta + 1\right)\|r^{k+1}\| + \beta\|A^\top B\|\|y^k - y^{k+1}\| \\
& + \|S(x^{k+1} - x^k)\| + \|T(y^k - y^{k+1})\|.
\end{aligned}$$

Notice that

$$\|S(x^{k+1} - x^k)\|^2 \le \|S\|\|x^{k+1} - x^k\|^2_{S+\Sigma_1},$$

by using (3.10), we thus have

$$\|R(w^{k+1})\| \le \iota_1\|x^{k+1} - x^k\|_{S+\Sigma_1} + \iota_2\|y^{k+1} - y^k\| + \iota_3\|\lambda^{k+1} - \lambda^k\|, \tag{4.20}$$

where

$$\begin{aligned}
\iota_1 &= \sqrt{\|S\|}, \\
\iota_2 &= \left((\|A\| + |1-\gamma|\|B\|)\beta + 1\right)\frac{\alpha\|B\|}{\alpha+\gamma} + \beta\|A^\top B\| + \|T\|, \\
\iota_3 &= \|A\| + \frac{(\|A\| + |1-\gamma|\|B\|)\beta + 1}{(\alpha+\gamma)\beta}.
\end{aligned}$$

On the other hand, we know from (2.10) that

$$\|w^{k+1} - w^k\|_{\widehat{G}} = \|x^{k+1} - x^k\|^2_{S+\Sigma_1} + \|v^{k+1} - v^k\|^2_{\widehat{H}}. \tag{4.21}$$

Since $\widehat{H} \succ 0$ (see Proposition 2.1), combining (4.20) and (4.21) together, we know that there must exist a positive constant such that (4.17) holds. The proof is complete. □

With Theorems 4.1-4.2 and Lemma 4.3, we can immediately arrive at the sublinear convergence rate results based on the KKT residual $\|R(w^k)\|$.

Theorem 4.3. *Let the sequence $\{w^k\}$ be generated by iPSPR* (1.7) *with $(\alpha, \gamma) \in \mathbb{D}$. Suppose that the proximal term S is chosen according to* (3.32) *and* (4.2) *and the proximal term T is chosen according to* (3.33) *hold, we have*

$$\min_{1 \le i \le t} \|R(w^{i+1})\|^2 = o(1/t).$$

If we restrict $(\alpha, \gamma) \in \mathbb{D}_2 \cup \mathbb{D}_3 \cup \mathbb{D}_4$ and suppose that the proximal term S is chosen according to (3.32) *and* (4.2) *and the proximal term T is chosen according to* (3.33) *and* (4.6)*, we have*

$$\|R(w^{t+1})\|^2 = o(1/t).$$

5. Numerical Results

In this section, we demonstrate the potential efficiency of our method iPSPR (1.7) by solving the following ℓ_1 regularized least square problem

$$\min \ \frac{1}{2}\|Qy - c\|^2 + \rho\|y\|_1, \quad \text{s.t.} \quad By \le b, \tag{5.1}$$

where $y \in \mathbb{R}^n, c \in \mathbb{R}^p$, $Q \in \mathbb{R}^{p\times n}$ and $B \in \mathbb{R}^{m\times n}$. Problem (5.1) is a constrained extension of the ordinary unconstrained ℓ_1 regularized least square problem and it was considered in [25]. By introducing an auxiliary variable $x \in \mathbb{R}^m$, we rewrite (5.1) as

$$\min \ \frac{1}{2}\|Qy - c\|^2 + \rho\|y\|_1, \quad \text{s.t.} \quad x + By = b, \ x \ge 0, \tag{5.2}$$

which is a special instance of (1.1).

For our method iPSPR (1.7), we set $S = 0$ and choose T according to (3.41), namely,

$$T = rI - (Q^\mathsf{T} Q + \beta B^\mathsf{T} B) \quad \text{with} \quad r = \lambda_{\max}\left(\frac{1}{2}Q^\mathsf{T} Q + \tau\beta B^\mathsf{T} B\right) \tag{5.3}$$

with $\tau = 1.001\underline{\tau}^{\alpha,\gamma}$, where $\underline{\tau}^{\alpha,\gamma}$ is defined in (3.44). Our method iPSPR (1.7) for solving (5.2) is then given as

$$\begin{cases} x^{k+1} = \mathcal{P}_+\left[b - By^k + \lambda^k/\beta\right], \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \alpha\beta(x^{k+1} + By^k - b), \\ y^{k+1} = \mathcal{S}_{\rho/r}\left[y^k + \frac{1}{r}\left(B^\mathsf{T}\left(\lambda^{k+\frac{1}{2}} - \beta(x^{k+1} + By^k - b)\right) + Q^\mathsf{T}(c - Qy^k)\right)\right], \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - \gamma\beta(x^{k+1} + By^{k+1} - b), \end{cases} \tag{5.4}$$

where the projection operator $\mathcal{P}_+(z) = \max(z, 0)$ and the shrinkage operator $\mathcal{S}_\nu(z) := \mathrm{sgn}(z) \odot \max\{|z| - \nu, 0\}$. Note that for problem (5.2), the majorized indefinite proximal ADMM in [25] coincides with our iPSPR (5.4) with $\alpha = 0$ and $(\alpha, \gamma) \in \mathbb{D}_1$ since the smooth part of the objective function is quadratic with respect to y. If the proximal parameter $r = 1.001\lambda_{\max}(Q^\mathsf{T} Q+\beta B^\mathsf{T} B)$, iPSPR becomes the semidefinite proximal-based strictly contractive Peaceman-Rachford splitting method (sPSPR). To make a fair comparison, as done in (85) of [25], we stop iPSPR or sPSPR when the KKT residual is less than 10^{-6}.

All the experiments were preformed in Ubuntu 16.04 LTS a Dell workstation with a 3.5-GHz Intel Xeon E3-1240 v5 processor with access to 32 GB of RAM. All the methods were implemented in MATLAB (R2016b). Given m and n, as done in [25], we set $p = 0.1n$, $\rho = 5\sqrt{n}$ and generate the data as

```
B = sprandn(m, n, 0.2);  yy = randn(n, 1);  b = B * yy + max(randn(m,1), 0);
Q = sprandn(p, n, 0.1);  c = Q * yy;
```

In our tests, we set $m = 2000$ and $n = 1000, 2000, 4000, 8000$. For each m and n, we use the above scheme to generate 50 groups of instances and will always report the average performance for methods iPSPR and sPSPR. For each instance, we fix the sum $\alpha + \gamma$ to be $\{1.9, 1.8, 1.618, 1\}$ and always choose the special cases with $\alpha = \gamma$, $\alpha = 0$ or $\gamma = 1$. In total, we have nine groups of choices of α and γ. In our tests, the penalty parameter β is fixed during the iterations. Generally, choosing the best penalty parameter β is not easy and it might be problem dependent [17]. We spent some efforts to choose the penalty parameter β from a large number of candidates. For each given m, n and α, γ, we report the performance of iPSPR or sPSPR with four choices of β. Note that in our tests, the second choice is the best choice in the candidates for $\alpha+\gamma = 1.9$ and almost the best choice in the candidates for $\alpha+\gamma \in \{1.618, 1.8\}$; the third choice of β is the best choice in the candidates for $\alpha + \gamma = 1$.

The numerical results are presented in Tables 5.1-5.4. In the tables, "iter" means the averaged iteration numbers, "r" denotes the proximal parameter in (5.4), and "t" means the CPU time in seconds. From the tables, we can make the following observations. First, iPSPR always perform better than the sPSPR. In particular, for $n = 4000, \beta = 0.15$ and $n = 8000, \beta = 0.07$, iPSPR can bring about 40% - 50% reduction in the number of iterations and the CPU time over the sPSPR. For $n = 1000$ and 2000, iPSPR with large sum $\alpha + \gamma$ performs only slightly better than sPSPR with the same α and γ. This might be due to that $\beta B^\mathsf{T} B$ takes a major part in computing r and the parameter τ of iPSPR is near to 1 in this case. Second, iPSPR (resp. sPSPR) with $\alpha = \gamma$ performs slightly better among the choices of α and γ with fixed sum. Third, a large $\alpha + \gamma$ sum (near to 2) always performs better than a small sum for a relatively small β, while a small sum works better than a large sum for a relatively large β. However, if we choose the best β (the corresponding results are marked in bold in each table) for each α and γ, we can see that iPSPR (resp. sPSPR) a large $\alpha + \gamma$ sum always performs better than iPSPR (resp. sPSPR) with a small sum.

6. Conclusions

In this paper, we proposed a modification of the Peaceman-Rachford splitting method by introducing two different parameters α and γ in updating the dual variable, and by introducing indefinite proximal terms to the subproblems in updating the primal variables. We established the relationship between the two parameters α and γ and proved the global convergence of

Table 5.1: The results for $m = 2000, n = 1000$ over 50 runs. The CPU time is in seconds.

		$\beta = 0.50$			$\beta = 1.50$			$\beta = 3.00$			$\beta = 5.00$		
(α, γ)	method	iter	r	t	iter	r	t	iter	r	t	iter	r	t
(0.950, 0.950)	iPSPR	8769.0	5.23e2	4.9	**4750.1**	1.56e3	2.7	5876.0	3.11e3	3.3	8576.2	5.18e3	4.8
(0.950, 0.950)	sPSPR	8877.3	5.47e2	5.0	**4815.4**	1.60e3	2.7	6000.3	3.19e3	3.4	8810.7	5.31e3	5.0
(0.900, 1.000)	iPSPR	8769.8	5.23e2	5.0	**4752.9**	1.56e3	2.7	5878.7	3.11e3	3.4	8579.8	5.18e3	4.9
(0.900, 1.000)	sPSPR	8878.4	5.47e2	5.0	**4817.9**	1.60e3	2.8	6001.5	3.19e3	3.4	8811.9	5.31e3	5.0
(0.900, 0.900)	iPSPR	9090.5	5.10e2	5.1	**4884.2**	1.52e3	2.8	5815.1	3.03e3	3.3	8375.5	5.04e3	4.7
(0.900, 0.900)	sPSPR	9252.9	5.47e2	5.2	**4982.8**	1.60e3	2.8	6075.2	3.19e3	3.4	8880.5	5.31e3	5.0
(0.800, 1.000)	iPSPR	9094.7	5.10e2	5.2	**4887.6**	1.52e3	2.8	5823.3	3.03e3	3.3	8405.3	5.04e3	4.8
(0.800, 1.000)	sPSPR	9256.5	5.47e2	5.2	**4985.6**	1.60e3	2.8	6079.5	3.19e3	3.5	8894.2	5.31e3	5.0
(0.809, 0.809)	iPSPR	9706.1	4.86e2	5.5	**5250.0**	1.44e3	3.0	5667.2	2.88e3	3.2	8094.9	4.80e3	4.6
(0.809, 0.809)	sPSPR	10058.7	5.47e2	5.7	**5388.7**	1.60e3	3.1	6240.7	3.19e3	3.6	8960.9	5.31e3	5.1
(0.618, 1.000)	iPSPR	9722.5	4.86e2	5.5	**5245.5**	1.44e3	3.0	5694.4	2.88e3	3.2	8124.6	4.80e3	4.6
(0.618, 1.000)	sPSPR	10074.1	5.47e2	5.7	**5392.1**	1.60e3	3.1	6261.2	3.19e3	3.6	9007.0	5.31e3	5.1
(0.000, 1.618)	iPSPR	10216.2	5.37e2	5.8	**5511.2**	1.60e3	3.1	6602.8	3.19e3	3.8	9583.0	5.31e3	5.4
(0.000, 1.618)	sPSPR	10264.2	5.47e2	5.8	**5517.4**	1.60e3	3.1	6605.5	3.19e3	3.8	9583.1	5.31e3	5.4
(0.000, 1.000)	iPSPR	13521.7	4.05e2	7.6	7761.5	1.20e3	4.4	**5967.7**	2.39e3	3.4	7667.3	3.98e3	4.3
(0.000, 1.000)	sPSPR	14513.1	5.47e2	8.2	8041.4	1.60e3	4.6	**7311.9**	3.19e3	4.2	9930.7	5.31e3	5.6
(0.500, 0.500)	iPSPR	13308.9	4.05e2	7.5	7724.5	1.20e3	4.4	**5728.7**	2.39e3	3.3	7265.2	3.98e3	4.1
(0.500, 0.500)	sPSPR	14336.6	5.47e2	8.1	7964.6	1.60e3	4.5	**7097.4**	3.19e3	4.0	9540.8	5.31e3	5.4

Table 5.2: The results for $m = 2000, n = 2000$ over 50 runs. The CPU time is in seconds.

		$\beta = 0.10$			$\beta = 0.30$			$\beta = 0.50$			$\beta = 1.00$		
(α, γ)	method	iter	r	t	iter	r	t	iter	r	t	iter	r	t
(0.950, 0.950)	iPSPR	2192.4	2.18e2	3.2	**1012.0**	4.43e2	1.5	1240.6	7.22e2	1.8	2328.8	1.43e3	3.4
(0.950, 0.950)	sPSPR	2357.0	3.83e2	3.4	**1121.2**	5.22e2	1.7	1309.0	7.66e2	1.9	2417.4	1.48e3	3.5
(0.900, 1.000)	iPSPR	2192.8	2.18e2	3.2	**1012.2**	4.43e2	1.5	1240.9	7.22e2	1.9	2329.3	1.43e3	3.4
(0.900, 1.000)	sPSPR	2357.3	3.83e2	3.4	**1121.4**	5.22e2	1.7	1309.2	7.66e2	2.0	2417.9	1.48e3	3.5
(0.900, 0.900)	iPSPR	2294.6	2.17e2	3.3	**1028.0**	4.32e2	1.5	1226.5	7.04e2	1.8	2274.1	1.39e3	3.3
(0.900, 0.900)	sPSPR	2474.9	3.83e2	3.6	**1146.1**	5.22e2	1.7	1324.5	7.66e2	2.0	2423.7	1.48e3	3.5
(0.800, 1.000)	iPSPR	2295.3	2.17e2	3.4	**1028.4**	4.32e2	1.6	1227.1	7.04e2	1.8	2276.1	1.39e3	3.3
(0.800, 1.000)	sPSPR	2475.5	3.83e2	3.6	**1146.8**	5.22e2	1.7	1325.2	7.66e2	2.0	2425.9	1.48e3	3.5
(0.809, 0.809)	iPSPR	2509.4	2.14e2	3.6	**1074.5**	4.13e2	1.6	1201.1	6.71e2	1.8	2170.0	1.33e3	3.2
(0.809, 0.809)	sPSPR	2722.4	3.83e2	3.9	**1201.1**	5.22e2	1.8	1351.8	7.66e2	2.0	2437.7	1.48e3	3.5
(0.618, 1.000)	iPSPR	2511.7	2.14e2	3.6	**1075.4**	4.13e2	1.6	1204.1	6.71e2	1.8	2177.7	1.33e3	3.2
(0.618, 1.000)	sPSPR	2724.1	3.83e2	3.9	**1202.5**	5.22e2	1.8	1354.7	7.66e2	2.0	2445.8	1.48e3	3.6
(0.000, 1.618)	iPSPR	2541.2	2.20e2	3.7	**1138.7**	4.53e2	1.7	1366.0	7.40e2	2.0	2547.4	1.47e3	3.7
(0.000, 1.618)	sPSPR	2733.0	3.83e2	3.9	**1228.7**	5.22e2	1.8	1406.8	7.66e2	2.1	2571.5	1.48e3	3.7
(0.000, 1.000)	iPSPR	3737.8	2.05e2	5.4	1551.7	3.50e2	2.3	**1242.2**	5.60e2	1.8	1910.1	1.10e3	2.8
(0.000, 1.000)	sPSPR	4152.4	3.83e2	5.9	1649.0	5.22e2	2.4	**1542.5**	7.66e2	2.3	2559.0	1.48e3	3.7
(0.500, 0.500)	iPSPR	3716.2	2.05e2	5.3	1546.0	3.50e2	2.3	**1213.4**	5.60e2	1.8	1831.6	1.10e3	2.7
(0.500, 0.500)	sPSPR	4137.3	3.83e2	5.9	1637.5	5.22e2	2.4	**1509.1**	7.66e2	2.2	2481.4	1.48e3	3.6

Table 5.3: The results for $m = 2000, n = 4000$ over 50 runs. The CPU time is in seconds.

		$\beta = 0.08$			$\beta = 0.15$			$\beta = 0.25$			$\beta = 0.50$		
(α, γ)	method	iter	r	t	iter	r	t	iter	r	t	iter	r	t
(0.950, 0.950)	iPSPR	905.6	3.71e2	3.0	**672.0**	4.24e2	2.3	831.3	5.65e2	2.8	1657.8	1.06e3	5.4
(0.950, 0.950)	sPSPR	1207.2	7.03e2	4.0	**1103.8**	7.39e2	3.7	1233.3	8.09e2	4.1	1813.6	1.15e3	5.9
(0.900, 1.000)	iPSPR	905.7	3.71e2	3.0	**672.0**	4.24e2	2.3	831.4	5.65e2	2.8	1657.9	1.06e3	5.4
(0.900, 1.000)	sPSPR	1207.2	7.03e2	4.0	**1103.9**	7.39e2	3.7	1233.3	8.09e2	4.1	1813.7	1.15e3	5.9
(0.900, 0.900)	iPSPR	943.5	3.70e2	3.2	**681.6**	4.21e2	2.3	812.2	5.54e2	2.8	1615.6	1.03e3	5.3
(0.900, 0.900)	sPSPR	1234.3	7.03e2	4.1	**1103.9**	7.39e2	3.7	1231.0	8.09e2	4.1	1815.3	1.15e3	6.0
(0.800, 1.000)	iPSPR	943.7	3.70e2	3.2	**681.7**	4.21e2	2.3	812.3	5.54e2	2.8	1616.2	1.03e3	5.3
(0.800, 1.000)	sPSPR	1234.5	7.03e2	4.1	**1104.4**	7.39e2	3.7	1231.3	8.09e2	4.1	1816.0	1.15e3	6.0
(0.809, 0.809)	iPSPR	1029.0	3.68e2	3.4	**706.1**	4.14e2	2.4	779.1	5.34e2	2.7	1537.7	9.86e2	5.1
(0.809, 0.809)	sPSPR	1294.0	7.03e2	4.3	**1108.9**	7.39e2	3.7	1226.0	8.09e2	4.1	1817.5	1.15e3	6.0
(0.618, 1.000)	iPSPR	1029.6	3.68e2	3.4	**706.8**	4.14e2	2.4	780.9	5.34e2	2.7	1540.7	9.86e2	5.1
(0.618, 1.000)	sPSPR	1294.5	7.03e2	4.3	**1109.9**	7.39e2	3.7	1227.7	8.09e2	4.1	1820.4	1.15e3	6.0
(0.000, 1.618)	iPSPR	1035.6	3.73e2	3.5	**731.1**	4.28e2	2.5	876.7	5.77e2	3.0	1764.3	1.09e3	5.8
(0.000, 1.618)	sPSPR	1300.3	7.03e2	4.3	**1125.7**	7.39e2	3.7	1256.0	8.09e2	4.2	1876.5	1.15e3	6.2
(0.000, 1.000)	iPSPR	1594.1	3.61e2	5.2	927.6	3.95e2	3.1	**759.7**	4.74e2	2.6	1300.8	8.25e2	4.3
(0.000, 1.000)	sPSPR	1727.8	7.03e2	5.7	1232.1	7.39e2	4.1	**1223.3**	8.09e2	4.1	1847.1	1.15e3	6.1
(0.500, 0.500)	iPSPR	1589.2	3.61e2	5.2	922.7	3.95e2	3.1	**747.5**	4.74e2	2.6	1264.4	8.25e2	4.2
(0.500, 0.500)	sPSPR	1724.1	7.03e2	5.7	1224.8	7.39e2	4.1	**1205.5**	8.09e2	4.0	1810.6	1.15e3	6.0

Table 5.4: The results for $m = 2000, n = 8000$ over 50 runs. The CPU time is in seconds.

		$\beta = 0.04$			$\beta = 0.07$			$\beta = 0.15$			$\beta = 0.30$		
(α, γ)	method	iter	r	t	iter	r	t	iter	r	t	iter	r	t
(0.950, 0.950)	iPSPR	889.6	6.80e2	6.6	**759.9**	6.95e2	5.7	861.4	7.55e2	6.4	1236.9	1.05e3	9.1
(0.950, 0.950)	sPSPR	**1487.6**	1.34e3	10.7	1556.1	1.36e3	11.2	1658.1	1.40e3	11.9	1819.2	1.52e3	13.1
(0.900, 1.000)	iPSPR	889.7	6.80e2	6.6	**760.0**	6.95e2	5.7	861.5	7.55e2	6.4	1236.9	1.05e3	9.1
(0.900, 1.000)	sPSPR	**1487.6**	1.34e3	10.7	1556.2	1.36e3	11.2	1658.1	1.40e3	11.9	1819.2	1.52e3	13.1
(0.900, 0.900)	iPSPR	913.9	6.80e2	6.8	**761.9**	6.94e2	5.7	854.8	7.51e2	6.4	1210.7	1.03e3	8.9
(0.900, 0.900)	sPSPR	1488.1	1.34e3	10.7	**1550.8**	1.36e3	11.2	1655.9	1.40e3	11.9	1819.2	1.52e3	13.1
(0.800, 1.000)	iPSPR	913.9	6.80e2	6.8	**762.0**	6.94e2	5.8	854.9	7.51e2	6.4	1211.0	1.03e3	9.0
(0.800, 1.000)	sPSPR	1488.3	1.34e3	10.7	**1550.9**	1.36e3	11.3	1656.0	1.40e3	12.0	1819.6	1.52e3	13.2
(0.809, 0.809)	iPSPR	968.8	6.79e2	7.1	**770.7**	6.93e2	5.8	842.1	7.44e2	6.3	1164.0	9.89e2	8.7
(0.809, 0.809)	sPSPR	1497.5	1.34e3	10.8	**1539.4**	1.36e3	11.2	1650.8	1.40e3	12.0	1819.0	1.52e3	13.2
(0.618, 1.000)	iPSPR	969.2	6.79e2	7.1	**771.2**	6.93e2	5.8	842.9	7.44e2	6.3	1165.4	9.89e2	8.6
(0.618, 1.000)	sPSPR	1497.7	1.34e3	10.8	**1539.8**	1.36e3	11.2	1651.7	1.40e3	12.0	1820.6	1.52e3	13.2
(0.000, 1.618)	iPSPR	972.6	6.81e2	7.1	**779.9**	6.96e2	5.9	873.9	7.59e2	6.6	1289.3	1.07e3	9.5
(0.000, 1.618)	sPSPR	1501.2	1.34e3	10.8	**1546.2**	1.36e3	11.2	1664.1	1.40e3	12.0	1845.1	1.52e3	13.4
(0.000, 1.000)	iPSPR	1406.3	6.76e2	10.2	922.8	6.87e2	6.9	**799.3**	7.25e2	6.1	1030.0	8.76e2	7.7
(0.000, 1.000)	sPSPR	1732.8	1.34e3	12.4	**1500.3**	1.36e3	11.0	1625.9	1.40e3	11.8	1825.7	1.52e3	13.3
(0.500, 0.500)	iPSPR	1403.0	6.76e2	10.1	919.9	6.87e2	6.9	**790.6**	7.25e2	6.0	1014.0	8.76e2	7.6
(0.500, 0.500)	sPSPR	1730.7	1.34e3	12.4	**1496.1**	1.36e3	10.9	1617.5	1.40e3	11.8	1809.7	1.52e3	13.1

the algorithm under some mild requirements on the proximal matrices S and T. Moreover, we provided a specific construction of the proximal matrix T and discussed the detailed performance for the variants parameters α and γ which can unify several existing results. We also analyzed the $o(1/t)$ sublinear rate convergence in the nonergodic sense. Finally, we reported some preliminary 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. Such suitable updating rules are among our future research tasks. Besides, an approximate version of the proposed iPSPR with practical accuracy criteria is also our future research topic.

Acknowledgments. The authors thank the referee for their valuable comments which improves the early version of the article. Part of this work was done during Y. Gu's Master study in Nanjing Normal University and PhD study in Kyoto University. Y. Gu was partially supported by the Natural Science Foundation of Jiangsu Province (Grant No. BK20210267). B. Jiang was supported by the National Natural Science Foundation of China (Grant No. 11971239) and the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (Grant No. 21KJA110002). D. Han was supported by the National Natural Science Foundation of China (Grant Nos. 12131004, 11625105).

References

[1] J. Bai, J. Li, F. Xu, and H. Zhang, Generalized symmetric ADMM for separable convex optimization, *Computational Optimization and Applications*, **70** (2018), 129–170.

[2] 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), 1–122.

[3] L. Chen, X. Li, D. Sun, and K.C. Toh, On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming, *Mathematical Programming*, **185** (2021), 111–161.

[4] Y. Chen, Y.H. Dai, and D. Han, Fiber orientation distribution estimation using a Peaceman-Rachford splitting method, *SIAM Journal on Imaging Sciences*, **9** (2016), 573–604.

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

[6] W. Deng, M.J. Lai, Z. Peng, and W. Yin, Parallel multi-block ADMM with $o(1/k)$ convergence, *Journal of Scientific Computing*, **71** (2017), 712–736.

[7] W. Deng and W. Yin, On the global and linear convergence of the generalized alternating direction method of multipliers, *Journal of Scientific Computing*, **66** (2016), 889–916.

[8] J. Douglas and H. H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, *Transactions of the American Mathematical Society*, **82** (1956), 421–439.

[9] J. Eckstein and M. Fukushima, Some reformulations and applications of the alternating direction method of multipliers, in Large scale optimization, Springer, 1994, 115–134.

[10] 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), 946–977.

[11] D. Gabay and B. Mercier, A dual algorithm for the solution of non linear variational problems via finite element approximation, Institut de recherche d'informatique et d'automatique, 1975.

[12] B. Gao and F. Ma, Symmetric alternating direction method with indefinite proximal regularization for linearly constrained convex optimization, *Journal of Optimization Theory and Applications*, **176** (2018), 178–204.

[13] 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, *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, **9** (1975), 41–76.

[14] Y. Gu, Modified strictly contractive Peaceman–Rachford splitting method (in Chinese), 2015.

[15] D. Han, D. Sun, and L. Zhang, Linear rate convergence of the alternating direction method of multipliers for convex composite programming, *Mathematics of Operations Research*, **43** (2017), 622–637.

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

[17] 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), 1011–1040.

[18] B. He, F. Ma, and X. Yuan, Convergence study on the symmetric version of ADMM with larger step sizes, *SIAM Journal on Imaging Sciences*, **9** (2016), 1467–1501.

[19] B. He, F. Ma, and X. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, *Computational Optimization and Applications*, (2019), 1–28.

[20] 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), 700–709.

[21] Y. He, H. Li, and X. Liu, Relaxed inertial proximal Peaceman–Rachford splitting method for separable convex programming, *Frontiers of Mathematics in China*, **13** (2018), 555–578.

[22] F. Jiang, X. Cai, and D. Han, The indefinite proximal point algorithms for maximal monotone operators, *Optimization*, **70** (2021), 1759–1790.

[23] F. Jiang, Z. Wu, and X. Cai, Generalized ADMM with optimal indefinite proximal term for linearly constrained convex optimization, *Journal of Industrial & Management Optimization*, (2018), 183–202.

[24] G. Li, T. Liu, and T. K. Pong, Peaceman–Rachford splitting for a class of nonconvex optimization problems, *Computational Optimization and Applications*, **68** (2017), 407–436.

[25] M. Li, D. Sun, and K.C. Toh, A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization, *SIAM Journal on Optimization*, **26** (2016), 922–950.

[26] X. Li and X. Yuan, A proximal strictly contractive Peaceman–Rachford splitting method for convex programming with applications to imaging, *SIAM Journal on Imaging Sciences*, **8** (2015), 1332–1365.

[27] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, *in Dokl. Akad. Nauk SSSR*, **269** (1983), 543–547.

[28] Y. Nesterov, Gradient methods for minimizing composite functions, *Mathematical Programming*, **140** (2013), 125–161.

[29] Y. Nesterov, Lectures on Convex Optimization, Springer, **137** (2018).

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

[31] R. Rockafellar and R. Wets, Variational Analysis, Springer, **317** (2004).

[32] H. Sun, M. Tian, and M. Sun, The symmetric ADMM with indefinite proximal regularization and its application, *Journal of Inequalities and Applications*, **172** (2017).

[33] Z. Wu, M. Li, D.Z. Wang, and D. Han, A symmetric alternating direction method of multipliers for separable nonconvex minimization problems, *Asia-Pacific Journal of Operational Research*, **34** (2017), 1750030.

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

[35] N. Zhang, J. Wu, and L. Zhang, A linearly convergent majorized ADMM with indefinite proximal terms for convex composite programming and its applications, *Mathematics of Computation*, **89** (2020), 1867–1894.