

# Peaceman-Rachford Splitting Method Converges Ergodically for Solving Convex Optimization Problems

Kaihuang Chen · Defeng Sun · Yancheng Yuan · Guojun Zhang · Xinyuan Zhao

Received: date / Accepted: date

**Abstract** In this paper, we prove that the ergodic sequence generated by the Peaceman-Rachford (PR) splitting method with semi-proximal terms converges for convex optimization problems (COPs). Numerical experiments on the linear programming benchmark dataset further demonstrate that, with a restart strategy, the ergodic sequence of the PR splitting method with semi-proximal terms consistently outperforms both the point-wise and ergodic sequences of the Douglas-Rachford (DR) splitting method. These findings indicate that the restarted ergodic PR splitting method is a more effective choice for tackling large-scale COPs compared to its DR counterparts.

**Keywords** Peaceman-Rachford splitting method · Ergodic convergence · Degenerate proximal point algorithm · Preconditioned ADMM

**Mathematics Subject Classification (2020)** 90C05 · 90C06 · 90C25

Kaihuang Chen  
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong  
E-mail: kaihuang.chen@connect.polyu.hk

Defeng Sun  
Corresponding author, Department of Applied Mathematics and Research Center for Intelligent Operations Research, The Hong Kong Polytechnic University, Hung Hom, Hong Kong  
E-mail: defeng.sun@polyu.edu.hk

Yancheng Yuan  
Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hung Hom, Hong Kong  
E-mail: yancheng.yuan@polyu.edu.hk

Guojun Zhang  
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong  
E-mail: guojun.zhang@connect.polyu.hk

Xinyuan Zhao  
Department of Mathematics, Beijing University of Technology, Beijing, P.R. China  
E-mail: xyzhao@bjut.edu.cn

## 1 Introduction

Let  $\mathbb{U}$ ,  $\mathbb{W}$ ,  $\mathbb{X}$ ,  $\mathbb{Y}$ , and  $\mathbb{Z}$  be finite-dimensional real Hilbert spaces, each equipped with an inner product  $\langle \cdot, \cdot \rangle$  and its corresponding norm  $\| \cdot \|$ . In this paper, we focus on the ergodic convergence of the Peaceman-Rachford (PR) splitting method for solving the following convex optimization problem (COP):

$$\begin{aligned} \min_{y \in \mathbb{Y}, z \in \mathbb{Z}} \quad & f_1(y) + f_2(z) \\ \text{s.t.} \quad & B_1 y + B_2 z = c, \end{aligned} \quad (1)$$

where  $f_1 : \mathbb{Y} \rightarrow (-\infty, +\infty]$  and  $f_2 : \mathbb{Z} \rightarrow (-\infty, +\infty]$  are proper closed convex functions,  $B_1 : \mathbb{Y} \rightarrow \mathbb{X}$  and  $B_2 : \mathbb{Z} \rightarrow \mathbb{X}$  are given linear operators, and  $c \in \mathbb{X}$  is a given vector. Let  $\sigma > 0$  be a given positive parameter. The augmented Lagrangian function for problem (1) is defined, for any  $(y, z, x) \in \mathbb{Y} \times \mathbb{Z} \times \mathbb{X}$ , as

$$L_\sigma(y, z; x) := f_1(y) + f_2(z) + \langle x, B_1 y + B_2 z - c \rangle + \frac{\sigma}{2} \|B_1 y + B_2 z - c\|^2.$$

The dual of problem (1) is

$$\max_{x \in \mathbb{X}} \{-f_1^*(-B_1^* x) - f_2^*(-B_2^* x) - \langle c, x \rangle\}. \quad (2)$$

Under mild constraint qualifications [19, Theorem 23.8], solving the dual problem (2) is equivalent to solving the following monotone inclusion problem (MIP):

$$0 \in (\mathcal{G}_1 + \mathcal{G}_2)x, \quad x \in \mathbb{X}, \quad (3)$$

where  $\mathcal{G}_1(x) := \partial(f_1^* \circ (-B_1^*))(x) + c$  and  $\mathcal{G}_2(x) := \partial(f_2^* \circ (-B_2^*))(x)$ . For any given maximal monotone operator  $\mathcal{G} : \mathbb{X} \rightarrow 2^{\mathbb{X}}$ , we denote its resolvent by  $\mathcal{J}_{\mathcal{G}} := (\mathcal{I} + \mathcal{G})^{-1}$ , where  $\mathcal{I}$  is the identity operator. The relaxed Douglas-Rachford (DR) splitting method with the relaxation factor  $\rho \in (0, 2]$  [15] for solving problem (3) is given by

$$u^{k+1} = \widehat{\mathcal{F}}_\rho^{\text{DR}}(u^k) := \left( \mathcal{I} + \rho(\mathcal{J}_{\sigma \mathcal{G}_1} \circ (2\mathcal{J}_{\sigma \mathcal{G}_2} - \mathcal{I}) - \mathcal{J}_{\sigma \mathcal{G}_2}) \right) u^k, \quad k = 0, 1, \dots, \quad (4)$$

where  $u^0 \in \mathbb{X}$  is an initial point. This is equivalent to the following iterative scheme: for any  $k \geq 0$ ,

$$\begin{cases} \bar{x}_1^k = \mathcal{J}_{\sigma \mathcal{G}_2}(u^k), \\ \bar{x}_2^k = \mathcal{J}_{\sigma \mathcal{G}_1}(2\bar{x}_1^k - u^k), \\ u^{k+1} = u^k + \rho(\bar{x}_2^k - \bar{x}_1^k). \end{cases}$$

The case  $\rho = 2$  is referred to as the PR splitting method, while  $\rho = 1$  corresponds to the widely studied DR splitting method.

Building on the relaxed DR splitting method, Eckstein and Bertsekas [11] introduced the generalized alternating direction method of multipliers (GADMM) for solving the COP (1). Later, in extending the semi-proximal ADMM (sPADMM) framework with larger dual step lengths studied in [12] to

**Algorithm 1** A pADMM method for solving COP (1)

- 
- 1: Input: Let  $\mathcal{T}_1$  and  $\mathcal{T}_2$  be two self-adjoint, positive semidefinite linear operators on  $\mathbb{Y}$  and  $\mathbb{Z}$ , respectively. Denote  $w = (y, z, x)$  and  $\bar{w} = (\bar{y}, \bar{z}, \bar{x})$ . Select an initial point  $w^0 = (y^0, z^0, x^0) \in \text{dom}(f_1) \times \text{dom}(f_2) \times \mathbb{X}$ . Set the parameters  $\sigma > 0$  and  $\rho \in (0, 2]$ .
  - 2: **for**  $k = 0, 1, \dots$ , **do**
  - 3:   Step 1.  $\bar{z}^k = \arg \min_{z \in \mathbb{Z}} \left\{ L_\sigma(y^k, z; x^k) + \frac{1}{2} \|z - z^k\|_{\mathcal{T}_2}^2 \right\}$ ;
  - 4:   Step 2.  $\bar{x}^k = x^k + \sigma(B_1 y^k + B_2 \bar{z}^k - c)$ ;
  - 5:   Step 3.  $\bar{y}^k = \arg \min_{y \in \mathbb{Y}} \left\{ L_\sigma(y, \bar{z}^k; \bar{x}^k) + \frac{1}{2} \|y - y^k\|_{\mathcal{T}_1}^2 \right\}$ ;
  - 6:   Step 4.  $w^{k+1} = (1 - \rho)w^k + \rho\bar{w}^k$ ;
  - 7: **end for**
- 

encompass the GADMM, Xiao et al. [24] proposed a preconditioned ADMM (pADMM) with semi-proximal terms (or degenerate proximal ADMM) for solving the COP (1), as described in Algorithm 1.

When  $\mathcal{T}_1 = 0$  and  $\mathcal{T}_2 = 0$ , the pADMM with  $\rho = 2$  reduces to the PR splitting method [11, 24], while the pADMM with  $\rho = 1$  corresponds to the DR splitting method [13]. For general positive semidefinite linear operators  $\mathcal{T}_1$  and  $\mathcal{T}_2$ , we refer to the pADMM as the PR splitting method with semi-proximal terms when  $\rho = 2$ , or as the DR splitting method with semi-proximal terms when  $\rho = 1$ . These semi-proximal terms are crucial for simplifying subproblem solutions and improving the scalability of the PR and DR splitting methods, especially for large-scale COPs (1).

It has long been observed that the PR splitting method is typically faster than the DR splitting method whenever it converges point-wisely [13, 15]. Unfortunately, since  $\hat{\mathcal{F}}_2^{\text{DR}}$  defined in (4) is merely a non-expansive operator, the PR splitting method does not necessarily converge point-wisely for general MIPs (3) (see e.g., [16]). Consequently, previous studies have primarily focused on studying the point-wise convergence of the PR splitting method under strong monotonicity assumptions on  $\mathcal{G}_1$  or  $\mathcal{G}_2$  [10, 14, 16, 22]. However, these strong monotonicity conditions are rarely met in practice, particularly for important problem classes like COPs (1), which significantly limit the practical applicability of the PR splitting method. Moreover, motivated by the impressive performance of the restarted ergodic primal-dual hybrid gradient method for linear programming (LP) [2], we evaluate the restarted ergodic PR splitting method with semi-proximal terms on LP instances (see Section 4). Numerical results show that it significantly outperforms both the point-wise and ergodic sequences of the DR splitting method with semi-proximal terms.

Let the sequences  $\{w^t = (y^t, z^t, x^t)\}$  and  $\{\bar{w}^t = (\bar{y}^t, \bar{z}^t, \bar{x}^t)\}$  be generated by the pADMM in Algorithm 1. Without imposing any strong monotonicity assumption, here we focus on analyzing the convergence of the two ergodic sequences  $\{w_a^k = (y_a^k, z_a^k, x_a^k)\}$  and  $\{\bar{w}_a^k = (\bar{y}_a^k, \bar{z}_a^k, \bar{x}_a^k)\}$ , where for each  $k \geq 0$ ,

$$(y_a^k, z_a^k, x_a^k) := \frac{1}{k+1} \sum_{t=0}^k (y^t, z^t, x^t) \quad \text{and} \quad (\bar{y}_a^k, \bar{z}_a^k, \bar{x}_a^k) := \frac{1}{k+1} \sum_{t=0}^k (\bar{y}^t, \bar{z}^t, \bar{x}^t).$$

By applying Baillon's non-linear ergodic theorem [3] to the non-expansive operator  $\widehat{\mathcal{F}}_2^{\text{DR}}$  in (4), the convergence of the ergodic sequence  $\{u_a^k := \frac{1}{k+1} \sum_{t=0}^k u^t\}$  can be directly obtained in the PR splitting method to solve MIP (3). Furthermore, the continuity of the resolvent operator  $\mathcal{J}_{\sigma\mathcal{G}_2}(\cdot)$  ensures the convergence of the sequence  $\{\mathcal{J}_{\sigma\mathcal{G}_2}(u_a^k)\}$ . However, these results do not imply the convergence of  $\{\bar{w}_a^k\}$  or  $\{w_a^k\}$  of the PR splitting method to solve COP (1). One key reason is that averaging and applying the resolvent operator cannot generally be interchanged. An analytical example in Appendix A further highlights the challenges of analyzing the ergodic sequence of the PR splitting method by showing that the sequence  $\{w^k\}$  generated by the PR splitting method for solving COP (1) may be unbounded. To our knowledge, whether the ergodic sequence of the PR splitting method converges to solve COP (1) remains an unsolved question.

In this paper, we address this unsolved question by proving the convergence of the ergodic sequence  $\{\bar{w}_a^k\}$  of the pADMM with  $\rho \in (0, 2]$ , which includes the PR splitting method with semi-proximal terms as a special case. Specifically, by reformulating the pADMM as a degenerate proximal point algorithm (dPPA) with a positive semidefinite preconditioner, as proposed in [4], we establish the ergodic convergence of the pADMM by analyzing the ergodic convergence of the dPPA with a relaxation factor  $\rho \in (0, 2]$ . Furthermore, numerical results on the LP benchmark dataset show that, with a restart strategy, the ergodic sequence of the PR splitting method with semi-proximal terms outperforms both the point-wise and ergodic sequences of the DR splitting method with semi-proximal terms. This empirical evidence highlights the practical advantages of the ergodic sequence of the PR splitting method with semi-proximal terms for solving large-scale COPs (1). To provide theoretical justification for these findings, in Appendix B, we establish an ergodic iteration complexity of  $O(1/k)$  for the PR splitting method with semi-proximal terms, evaluated in terms of the objective error, the feasibility violation, and the Karush-Kuhn-Tucker residual using the  $\varepsilon$ -subdifferential.

The remainder of this paper is organized as follows: Section 2 establishes the ergodic convergence of the dPPA for  $\rho \in (0, 2]$ . Section 3 builds on this result to prove the ergodic convergence of the PR splitting method for COPs (1). Section 4 presents numerical results on the LP benchmark dataset, highlighting the superior performance of the PR splitting method's ergodic sequence. Finally, Section 5 concludes the paper.

*Notation.* For any convex function  $f : \mathbb{X} \rightarrow (-\infty, +\infty]$ , we define its effective domain as  $\text{dom}(f) := \{x \in \mathbb{X} : f(x) < \infty\}$ , its conjugate as  $f^*(x) := \sup_{z \in \mathbb{X}} \{\langle x, z \rangle - f(z)\}$ ,  $x \in \mathbb{X}$ , and its subdifferential  $\partial f(x) = \{v \mid f(z) \geq f(x) + \langle v, x - z \rangle, \forall z \in \mathbb{X}\}$ ,  $x \in \mathbb{X}$ . Furthermore, consider a closed convex set  $C \subseteq \mathbb{X}$ . We express the Euclidean projection of  $x$  onto  $C$  as  $\Pi_C(x) := \arg \min\{\|x - z\| \mid z \in C\}$ . Moreover, for a linear operator  $A : \mathbb{X} \rightarrow \mathbb{Y}$ , its adjoint is denoted by  $A^*$  and  $\lambda_1(AA^*)$  represents the largest eigenvalue of  $AA^*$ . Additionally, for any self-adjoint, positive semidefinite linear operator  $\mathcal{M} : \mathbb{X} \rightarrow \mathbb{X}$ , we define the semi-norm as  $\|x\|_{\mathcal{M}} := \sqrt{\langle x, x \rangle_{\mathcal{M}}} = \sqrt{\langle x, \mathcal{M}x \rangle}$  for

any  $x \in \mathbb{X}$ . Finally, for a non-expansive operator  $\widehat{\mathcal{F}} : \mathbb{X} \rightarrow \mathbb{X}$ , the set of fixed points of  $\widehat{\mathcal{F}}$  is denoted by  $\text{Fix } \widehat{\mathcal{F}}$ .

## 2 Ergodic convergence of the dPPA

Let  $\mathcal{T}$  be a maximal monotone operator from  $\mathbb{W}$  into itself. Consider the following inclusion problem:

$$\text{find } w \in \mathbb{W} \text{ such that } 0 \in \mathcal{T}w. \quad (5)$$

Assume that  $\mathcal{M} : \mathbb{W} \rightarrow \mathbb{W}$  is an admissible preconditioner for the maximal monotone operator  $\mathcal{T}$ , that is,  $\mathcal{M}$  is a linear, bounded, self-adjoint and positive semidefinite operator such that  $\widehat{\mathcal{T}} := (\mathcal{M} + \mathcal{T})^{-1}\mathcal{M}$  is single-valued and has a full domain. For a given  $w^0 \in \mathbb{W}$ , the dPPA [4] for solving the inclusion problem (5) is given by:

$$w^{k+1} = (1 - \rho)w^k + \rho\bar{w}^k, \quad \bar{w}^k = \widehat{\mathcal{T}}w^k = (\mathcal{M} + \mathcal{T})^{-1}\mathcal{M}w^k, \quad k = 0, 1, \dots, \quad (6)$$

where  $\rho \in (0, 2]$ . When  $\mathcal{M}$  is the identity mapping, the dPPA reduces to the classical PPA introduced by Rockafellar in [20]. By carefully selecting  $\mathcal{M}$ , the computation of  $\widehat{\mathcal{T}}$  can be greatly simplified, as discussed in [4, 23].

The global point-wise convergence of the dPPA for  $\rho \in (0, 2)$  has been established by Bredies et al. [4] under the assumption that  $(\mathcal{M} + \mathcal{T})^{-1}$  is  $L$ -Lipschitz continuous, i.e., there exists a constant  $L \geq 0$  such that for all  $v_1, v_2$  in the domain of  $(\mathcal{M} + \mathcal{T})^{-1}$ ,

$$\|(\mathcal{M} + \mathcal{T})^{-1}v_1 - (\mathcal{M} + \mathcal{T})^{-1}v_2\| \leq L\|v_1 - v_2\|.$$

This Lipschitz continuity assumption is met by many splitting algorithms, including the pADMM, under mild conditions (see [4, 23]). For  $\rho = 2$ , the dPPA may fail to converge point-wisely for solving the inclusion problem (5) without imposing additional assumptions, such as the strong monotonicity of  $\mathcal{T}$ . Instead, we focus on analyzing the global convergence of the following ergodic sequences generated by the dPPA for  $\rho = 2$ , without requiring the strong monotonicity of  $\mathcal{T}$ :

$$w_a^k = \frac{1}{k+1} \sum_{t=0}^k w^t, \quad \bar{w}_a^k = \frac{1}{k+1} \sum_{t=0}^k \bar{w}^t, \quad \forall k \geq 0, \quad (7)$$

where the sequences  $\{w^t\}$  and  $\{\bar{w}^t\}$  are generated by the dPPA (6). Recall that, for a maximal monotone operator  $\mathcal{T} : \mathbb{W} \rightarrow 2^{\mathbb{W}}$  and  $\varepsilon \geq 0$ , the  $\varepsilon$ -enlargement of  $\mathcal{T}$  at  $w$  [5] is defined as

$$\mathcal{T}^\varepsilon(w) = \{v \in \mathbb{W} : \langle w - w', v - v' \rangle \geq -\varepsilon, \forall (w', v') \in \text{gph}(\mathcal{T})\},$$

where  $\text{gph}(\mathcal{T}) := \{(w, v) \in \mathbb{W} \times \mathbb{W} \mid v \in \mathcal{T}w\}$ . Using the  $\varepsilon$ -enlargement of  $\mathcal{T}$ , we can establish the following proposition on the ergodic convergence properties of the dPPA.

**Proposition 1** Let  $\mathcal{T} : \mathbb{W} \rightarrow 2^{\mathbb{W}}$  be a maximal monotone operator with  $\mathcal{T}^{-1}(0) \neq \emptyset$ , and let  $\mathcal{M}$  be an admissible preconditioner. Then the ergodic sequences  $\{\bar{w}_a^k\}$  and  $\{w_a^k\}$ , generated by the dPPA (6) with  $\rho \in (0, 2]$ , satisfy for all  $k \geq 0$  and  $w^* \in \mathcal{T}^{-1}(0)$  that

- (a)  $\|\bar{w}_a^k - w_a^k\|_{\mathcal{M}} \leq \frac{2}{\rho(k+1)}\|w^0 - w^*\|_{\mathcal{M}}$ ;  
(b)  $\mathcal{M}(w_a^k - \bar{w}_a^k) \in \mathcal{T}^{\varepsilon_a^k}(\bar{w}_a^k)$ , where  $\varepsilon_a^k := \frac{1}{k+1} \sum_{t=0}^k \langle \bar{w}^t - \bar{w}_a^k, w^t - \bar{w}^t \rangle_{\mathcal{M}}$  and  $0 \leq \varepsilon_a^k \leq \frac{1}{2\rho(k+1)}\|w^0 - w^*\|_{\mathcal{M}}^2$ .

*Proof* Note that  $\widehat{\mathcal{T}}$  is  $\mathcal{M}$ -firmly non-expansive, as stated in [23, Proposition 2.3], meaning that

$$\|\widehat{\mathcal{T}}w - \widehat{\mathcal{T}}w'\|_{\mathcal{M}}^2 + \|(\mathcal{I} - \widehat{\mathcal{T}})w - (\mathcal{I} - \widehat{\mathcal{T}})w'\|_{\mathcal{M}}^2 \leq \|w - w'\|_{\mathcal{M}}^2, \forall w, w' \in \mathbb{W}.$$

From this property, we derive the following key inequality for the dPPA (6):

$$\|w^{k+1} - w^*\|_{\mathcal{M}}^2 \leq \|w^k - w^*\|_{\mathcal{M}}^2 - \rho(2 - \rho)\|w^k - \bar{w}^k\|_{\mathcal{M}}^2, \forall k \geq 0, w^* \in \mathcal{T}^{-1}(0). \quad (8)$$

Using this inequality and the scheme (6), we can complete the proof of statement (a) by noticing that for any  $k \geq 0$  and  $w^* \in \mathcal{T}^{-1}(0)$ ,

$$\begin{aligned} \|\bar{w}_a^k - w_a^k\|_{\mathcal{M}} &= \left\| \frac{1}{k+1} \sum_{t=0}^k (\bar{w}^t - w^t) \right\|_{\mathcal{M}} = \left\| \frac{1}{k+1} \sum_{t=0}^k \frac{(w^{t+1} - w^t)}{\rho} \right\|_{\mathcal{M}} \\ &= \frac{1}{\rho(k+1)} \|w^{k+1} - w^0\|_{\mathcal{M}} \leq \frac{2}{\rho(k+1)} \|w^0 - w^*\|_{\mathcal{M}}. \end{aligned}$$

Moreover, for any  $(w', v') \in \text{gph}(\mathcal{T})$ , the definitions of  $w_a^k$  and  $\bar{w}_a^k$  in (7), along with the monotonicity of  $\mathcal{T}$ , yield the following for any  $k \geq 0$ :

$$\begin{aligned} \langle \bar{w}_a^k - w', \mathcal{M}(w_a^k - \bar{w}_a^k) - v' \rangle &= \frac{1}{k+1} \sum_{t=0}^k \langle \bar{w}^t - w', \mathcal{M}(w_a^k - \bar{w}_a^k) - v' \rangle \\ &= \frac{1}{k+1} \sum_{t=0}^k \left( \langle \bar{w}^t - w', (w_a^k - \bar{w}_a^k) - (w^t - \bar{w}^t) \rangle_{\mathcal{M}} \right. \\ &\quad \left. + \langle \bar{w}^t - w', \mathcal{M}(w^t - \bar{w}^t) - v' \rangle \right) \\ &\geq \frac{1}{k+1} \sum_{t=0}^k \langle \bar{w}^t - w', (w_a^k - \bar{w}_a^k) - (w^t - \bar{w}^t) \rangle_{\mathcal{M}} \quad (9) \\ &= \frac{1}{k+1} \sum_{t=0}^k \langle \bar{w}^t - \bar{w}_a^k + \bar{w}_a^k - w', (w_a^k - \bar{w}_a^k) - (w^t - \bar{w}^t) \rangle_{\mathcal{M}} \\ &= \frac{1}{k+1} \sum_{t=0}^k \langle \bar{w}^t - \bar{w}_a^k, (w_a^k - \bar{w}_a^k) - (w^t - \bar{w}^t) \rangle_{\mathcal{M}} \\ &= -\frac{1}{k+1} \sum_{t=0}^k \langle \bar{w}^t - \bar{w}_a^k, w^t - \bar{w}^t \rangle_{\mathcal{M}} = -\varepsilon_a^k. \end{aligned}$$

It follows from the definition of the  $\varepsilon$ -enlargement of  $\mathcal{T}$  that  $\mathcal{M}(w_a^k - \bar{w}_a^k) \in \mathcal{T}^{\varepsilon_a^k}(\bar{w}_a^k)$ . Next, we prove that  $\varepsilon_a^k \geq 0$  for all  $k \geq 0$  by contradiction. Suppose that  $\varepsilon_a^k < 0$  for some  $k \geq 0$ . Then, for any  $(w', v') \in \text{gph}(\mathcal{T})$ , from (9), we have

$$\langle \bar{w}_a^k - w', \mathcal{M}(w_a^k - \bar{w}_a^k) - v' \rangle \geq -\varepsilon_a^k > 0,$$

which, combined with the maximality of  $\mathcal{T}$ , implies that  $(\bar{w}_a^k, \mathcal{M}(w_a^k - \bar{w}_a^k)) \in \text{gph}(\mathcal{T})$ . Taking  $(w', v') = (\bar{w}_a^k, \mathcal{M}(w_a^k - \bar{w}_a^k))$ , we obtain  $0 \geq -\varepsilon_a^k$ , which contradicts the assumption  $\varepsilon_a^k < 0$ . Thus,  $\varepsilon_a^k \geq 0$  for all  $k \geq 0$ .

Now, we establish an upper bound for  $\varepsilon_a^k$  for any  $k \geq 0$ . By the scheme of dPPA (6) and  $\rho \in (0, 2]$ , we have

$$\begin{aligned} \varepsilon_a^k &= \frac{1}{k+1} \sum_{t=0}^k \langle w^t - \bar{w}^t, \bar{w}^t - \bar{w}_a^k \rangle_{\mathcal{M}} \\ &= \frac{1}{k+1} \sum_{t=0}^k \left( \langle \frac{1}{\rho}(w^t - w^{t+1}), w^t - \frac{1}{\rho}(w^t - w^{t+1}) - \bar{w}_a^k \rangle_{\mathcal{M}} \right) \\ &= \frac{1}{k+1} \sum_{t=0}^k \left( -\frac{1}{\rho^2} \|w^t - w^{t+1}\|_{\mathcal{M}}^2 + \frac{1}{2\rho} \|w^t - w^{t+1}\|_{\mathcal{M}}^2 \right. \\ &\quad \left. + \frac{1}{2\rho} (\|w^t - \bar{w}_a^k\|_{\mathcal{M}}^2 - \|w^{t+1} - \bar{w}_a^k\|_{\mathcal{M}}^2) \right) \tag{10} \\ &\leq \frac{1}{k+1} \sum_{t=0}^k \left( \frac{1}{2\rho} (\|w^t - \bar{w}_a^k\|_{\mathcal{M}}^2 - \|w^{t+1} - \bar{w}_a^k\|_{\mathcal{M}}^2) \right) \\ &= \frac{1}{2\rho(k+1)} \left( \|w^0 - \bar{w}_a^k\|_{\mathcal{M}}^2 - \|w^{k+1} - \bar{w}_a^k\|_{\mathcal{M}}^2 \right) \\ &= \frac{1}{2\rho(k+1)} \left( -\|w^0 - w^{k+1}\|_{\mathcal{M}}^2 - 2\langle w^{k+1} - w^0, w^0 - \bar{w}_a^k \rangle_{\mathcal{M}} \right) \\ &\leq \frac{1}{2\rho(k+1)} \left( -\|w^0 - w^{k+1}\|_{\mathcal{M}}^2 + 2\|w^{k+1} - w^0\|_{\mathcal{M}} \|w^0 - \bar{w}_a^k\|_{\mathcal{M}} \right). \end{aligned}$$

Using the convexity of  $\|\cdot\|_{\mathcal{M}}$  and (8), we obtain that

$$\|w^0 - \bar{w}_a^k\|_{\mathcal{M}} = \left\| \frac{1}{k+1} \sum_{t=0}^k (w^0 - \bar{w}^t) \right\|_{\mathcal{M}} \leq \frac{1}{k+1} \sum_{t=0}^k \|w^0 - \bar{w}^t\|_{\mathcal{M}} \leq 2\|w^0 - w^*\|_{\mathcal{M}}.$$

Combining this with (10), we derive for any  $k \geq 0$  that

$$\begin{aligned} \varepsilon_a^k &\leq \frac{1}{2\rho(k+1)} \left( -\|w^0 - w^{k+1}\|_{\mathcal{M}}^2 + 2\|w^{k+1} - w^0\|_{\mathcal{M}} \|w^0 - w^*\|_{\mathcal{M}} \right) \\ &\leq \frac{1}{2\rho(k+1)} \|w^0 - w^*\|_{\mathcal{M}}^2. \end{aligned}$$

This completes the proof.

Based on Proposition 1, we can establish the convergence of ergodic sequence  $\{\bar{w}_a^k\}$  generated by the dPPA (6) in Theorem 1.

**Theorem 1** Let  $\mathcal{T} : \mathbb{W} \rightarrow 2^{\mathbb{W}}$  be a maximal monotone operator with  $\mathcal{T}^{-1}(0) \neq \emptyset$ , and let  $\mathcal{M}$  be an admissible preconditioner such that  $(\mathcal{M} + \mathcal{T})^{-1}$  is  $L$ -Lipschitz continuous. Then, the ergodic sequence  $\{\bar{w}_a^k\}$  generated by the dPPA (6) with  $\rho \in (0, 2]$  converges to a point in  $\mathcal{T}^{-1}(0)$ .

*Proof* Suppose that  $\mathcal{M} = \mathcal{C}\mathcal{C}^*$  is a decomposition of  $\mathcal{M}$  according to [4, Proposition 2.3], where  $\mathcal{C} : \mathbb{U} \rightarrow \mathbb{W}$  is an injective operator. Since  $(\mathcal{M} + \mathcal{T})^{-1}$  is  $L$ -Lipschitz continuous and  $\|\mathcal{C}^*w\| = \|w\|_{\mathcal{M}}$  for every  $w \in \mathbb{W}$ , we derive for all  $w' \in \mathbb{W}$  and  $w^* \in \mathcal{T}^{-1}(0)$  that

$$\|\widehat{\mathcal{T}}w' - \widehat{\mathcal{T}}w^*\| = \|(\mathcal{M} + \mathcal{T})^{-1}\mathcal{C}\mathcal{C}^*w' - (\mathcal{M} + \mathcal{T})^{-1}\mathcal{C}\mathcal{C}^*w^*\| \leq L\|\mathcal{C}\| \|w' - w^*\|_{\mathcal{M}}.$$

Combining this with the  $\mathcal{M}$ -firm non-expansiveness of  $\widehat{\mathcal{T}}$  [23, Proposition 2.3], we conclude that

$$\|\bar{w}^k - w^*\| = \|\widehat{\mathcal{T}}w^k - w^*\| \leq L\|\mathcal{C}\| \|w^k - w^*\|_{\mathcal{M}} \leq L\|\mathcal{C}\| \|w^0 - w^*\|_{\mathcal{M}}.$$

Thus, both sequences  $\{\bar{w}^k\}$  and  $\{\bar{w}_a^k\}$  are bounded. Furthermore, according to Proposition 1 and the maximality of  $\mathcal{T}$ , any cluster point of  $\{\bar{w}_a^k\}$  belongs to  $\mathcal{T}^{-1}(0)$ .

To establish the uniqueness of cluster points, we define two shadow sequences as follows:

$$u^k = \mathcal{C}^*w^k \quad \text{and} \quad u_a^k = \frac{1}{k+1} \sum_{t=0}^k u^t, \quad \forall k \geq 0. \quad (11)$$

A straightforward calculation shows that  $u^{k+1} = \widetilde{\mathcal{F}}_{\rho}u^k$ ,  $\forall k \geq 0$ , where  $\widetilde{\mathcal{F}}_{\rho} := (1 - \rho)\mathcal{I} + \rho(\mathcal{C}^*(\mathcal{M} + \mathcal{T})^{-1}\mathcal{C})$  with  $\rho \in (0, 2]$  is a non-expansive operator, as shown in [23, Proposition 2.5]. By Baillon's non-linear ergodic theorem [3], the sequence  $\{u_a^k\}$  converges to a point in  $\text{Fix}(\widetilde{\mathcal{F}}_{\rho})$ . Using the equivalence between  $\text{Fix}(\widetilde{\mathcal{F}}_{\rho})$  and  $\mathcal{C}^*\mathcal{T}^{-1}(0)$  as stated in [23, Proposition 2.5], we conclude that there exists  $w_a^* \in \mathcal{T}^{-1}(0)$  such that

$$\|u_a^k - \mathcal{C}^*w_a^*\| \rightarrow 0.$$

Therefore, by the definition of  $\{u_a^k\}$  in (11), we have

$$\|w_a^k - w_a^*\|_{\mathcal{M}} = \left\| \frac{1}{k+1} \sum_{t=0}^k \mathcal{C}^*w^t - \mathcal{C}^*w_a^* \right\| = \|u_a^k - \mathcal{C}^*w_a^*\| \rightarrow 0,$$

which, together with part (a) of Proposition 1, implies that

$$\|\bar{w}_a^k - w_a^*\|_{\mathcal{M}}^2 = \|\bar{w}_a^k - w_a^k\|_{\mathcal{M}}^2 + \|w_a^k - w_a^*\|_{\mathcal{M}}^2 + 2\langle \bar{w}_a^k - w_a^k, w_a^k - w_a^* \rangle_{\mathcal{M}} \rightarrow 0. \quad (12)$$

Since the sequence  $\{\bar{w}_a^k\}$  is bounded, it must have at least one cluster point. Assume that there is a subsequence  $\{\bar{w}_a^{k_i}\}$  converging to  $w^*$ . Suppose that  $\|w^* - w_a^*\|_{\mathcal{M}} > 0$ . By an Opial-type argument [18, Lemma 1], we have

$$\liminf_{i \rightarrow \infty} \|\bar{w}_a^{k_i} - w^*\|_{\mathcal{M}} < \liminf_{i \rightarrow \infty} \|\bar{w}_a^{k_i} - w_a^*\|_{\mathcal{M}},$$

which, combined with (12), implies  $\liminf_{i \rightarrow \infty} \|\bar{w}_a^{k_i} - w^*\|_{\mathcal{M}} < 0$ , a contradiction to the positive semidefiniteness of  $\mathcal{M}$ . Hence,  $\|w^* - w_a^*\|_{\mathcal{M}} = 0$ . It follows that

$$w^* = (\mathcal{M} + \mathcal{T})^{-1} \mathcal{M} w^* = (\mathcal{M} + \mathcal{T})^{-1} \mathcal{M} w_a^* = w_a^*.$$

Taking any other cluster point  $w^{**}$ , we can similarly show that  $w^{**} = w_a^*$ . Hence, the cluster point is unique, and the sequence  $\{\bar{w}_a^k\}$  converges to  $w_a^*$ .

### 3 Ergodic convergence of the PR splitting method

The Karush–Kuhn–Tucker (KKT) system of COP (1) is given by:

$$-B_1^* x^* \in \partial f_1(y^*), \quad -B_2^* x^* \in \partial f_2(z^*), \quad B_1 y^* + B_2 z^* - c = 0. \quad (13)$$

As shown in [19, Corollary 28.3.1],  $(y^*, z^*) \in \mathbb{Y} \times \mathbb{Z}$  is an optimal solution to problem (1) if and only if there exists  $x^* \in \mathbb{X}$  such that  $(y^*, z^*, x^*)$  satisfies the KKT system. To analyze the ergodic convergence of the pADMM including the PR splitting method, we make the following assumption:

**Assumption 1** *The KKT system (13) has a nonempty solution set.*

Under Assumption 1, solving the COP (1) is equivalent to finding  $w \in \mathbb{W} = \mathbb{Y} \times \mathbb{Z} \times \mathbb{X}$  such that  $0 \in \mathcal{T}w$ , where the maximal monotone operator  $\mathcal{T}$  is

$$\mathcal{T}w = \begin{pmatrix} \partial f_1(y) + B_1^* x \\ \partial f_2(z) + B_2^* x \\ c - B_1 y - B_2 z \end{pmatrix}, \quad \forall w = (y, z, x) \in \mathbb{W}. \quad (14)$$

Additionally, since  $f_1$  and  $f_2$  are proper closed convex functions, there exist two self-adjoint and positive semidefinite operators  $\Sigma_{f_1}$  and  $\Sigma_{f_2}$  such that:

$$\begin{aligned} f_1(y) &\geq f_1(\hat{y}) + \langle \hat{\phi}, y - \hat{y} \rangle + \frac{1}{2} \|y - \hat{y}\|_{\Sigma_{f_1}}^2, \quad \forall y, \hat{y} \in \text{dom}(f_1), \hat{\phi} \in \partial f_1(\hat{y}), \\ f_2(z) &\geq f_2(\hat{z}) + \langle \hat{\varphi}, z - \hat{z} \rangle + \frac{1}{2} \|z - \hat{z}\|_{\Sigma_{f_2}}^2, \quad \forall z, \hat{z} \in \text{dom}(f_2), \hat{\varphi} \in \partial f_2(\hat{z}). \end{aligned}$$

We make the following assumption to ensure that each step of the pADMM is well defined.

**Assumption 2**  *$\Sigma_{f_1} + B_1^* B_1 + \mathcal{T}_1$  and  $\Sigma_{f_2} + B_2^* B_2 + \mathcal{T}_2$  are positive definite.*

By defining the self-adjoint linear operator  $\mathcal{M} : \mathbb{W} \rightarrow \mathbb{W}$  as

$$\mathcal{M} = \begin{bmatrix} \sigma B_1^* B_1 + \mathcal{T}_1 & 0 & B_1^* \\ 0 & \mathcal{T}_2 & 0 \\ B_1 & 0 & \sigma^{-1} \mathcal{I} \end{bmatrix}, \quad (15)$$

Sun et al. [23] demonstrated the following equivalence between the pADMM in Algorithm 1 and the dPPA in (6).

**Proposition 2 ([23])** *Suppose that Assumption 2 holds. Consider the operators  $\mathcal{T}$  defined in (14) and  $\mathcal{M}$  defined in (15), respectively. Then the sequence  $\{w^k\}$  generated by the pADMM in Algorithm 1 coincides with the sequence  $\{w^k\}$  generated by the dPPA in (6) with the same initial point  $w^0 \in \mathbb{W}$ . Additionally,  $\mathcal{M}$  is an admissible preconditioner such that  $(\mathcal{M} + \mathcal{T})^{-1}$  is Lipschitz continuous.*

The equivalence in Proposition 2 allows us to use the ergodic convergence results of the dPPA in Theorem 1 to establish the ergodic convergence of the pADMM, including the PR splitting method ( $\rho = 2$ ) with semi-proximal terms, for solving COP (1), as shown in the following corollary.

**Corollary 1** *Suppose that Assumptions 1 and 2 hold. Then the ergodic sequence  $\{\bar{w}_a^k\} = \{(\bar{y}_a^k, \bar{z}_a^k, \bar{x}_a^k)\}$ , generated by the pADMM with  $\rho \in (0, 2]$  in Algorithm 1, converges to the point  $w^* = (y^*, z^*, x^*)$ , where  $(y^*, z^*)$  is a solution to problem (1), and  $x^*$  is a solution to problem (2).*

*Remark 1* The example in Appendix A demonstrates that the ergodic sequence  $\{w_a^k\}$  of the PR splitting method may fail to converge for general COPs (1). In this context, the ergodic convergence result established in Corollary 1 represents the best achievable.

#### 4 Numerical experiment

In this section, we use the following LP as an example to evaluate the performance of the ergodic sequence of the PR splitting method with semi-proximal terms:

$$\min\{\langle c, x \rangle \mid A_1 x = b_1, A_2 x \geq b_2, x \in C\}, \quad (16)$$

where  $A_1 \in \mathbb{R}^{m_1 \times n}$ ,  $A_2 \in \mathbb{R}^{m_2 \times n}$ ,  $b_1 \in \mathbb{R}^{m_1}$ ,  $b_2 \in \mathbb{R}^{m_2}$ , and  $c \in \mathbb{R}^n$ . The set  $C := \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ , with  $l \in (\mathbb{R} \cup \{-\infty\})^n$  and  $u \in (\mathbb{R} \cup \{+\infty\})^n$ . Let  $A = [A_1; A_2] \in \mathbb{R}^{m \times n}$  with  $m = m_1 + m_2$  and  $b = [b_1; b_2] \in \mathbb{R}^m$ . We assume that  $A$  is a non-zero matrix. The dual of problem (16) can be expressed as:

$$\min\{-\langle b, y \rangle + \delta_D(y) + \delta_C^*(-z) \mid A^* y + z = c, y \in \mathbb{R}^m, z \in \mathbb{R}^n\}, \quad (17)$$

where  $\delta_D(\cdot)$  is the indicator function over  $D := \{y = (y_1, y_2) \in \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}\}$ .

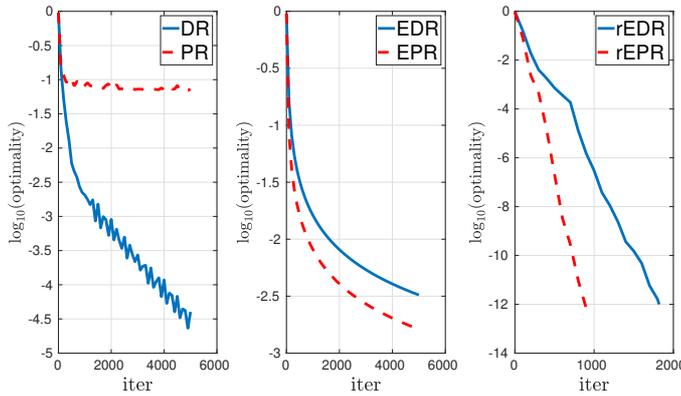
We apply the pADMM from Algorithm 1 with  $\mathcal{T}_1 = \sigma(\lambda_1(AA^*)I_m - AA^*)$  and  $\mathcal{T}_2 = 0$  to solve problem (17), where  $I_m$  is the identity matrix in  $\mathbb{R}^m$ . In this experiment, the pADMM with  $\rho = 2$  and  $\rho = 1$  are denoted as ‘‘PR’’ and ‘‘DR,’’ respectively, with their ergodic sequences referred to as ‘‘EPR’’ and ‘‘EDR.’’ To enhance the performance of the ergodic sequences, we apply the restart strategy from [2, 7], using the following merit function based on primal and dual infeasibility:

$$\tilde{R}_k = \sqrt{\sigma^{-1} \|II_D(b - A\bar{x}_a^k)\|^2 + \sigma \|c - A^*\bar{y}_a^k - \bar{z}_a^k\|^2}.$$

The restarted variants are referred to as “rEPR” and “rEDR”. We implement all tested algorithms in Julia and run them on an NVIDIA A100-SXM4-80GB GPU with CUDA 12.3. The tested algorithms terminate when the following “optimality” measure falls below a specified tolerance  $\epsilon > 0$ :

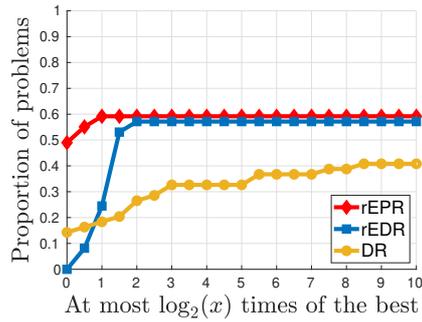
$$\max \left\{ \frac{|\langle b, y \rangle - \delta_C^*(-z) - \langle c, x \rangle|}{1 + |\langle b, y \rangle - \delta_C^*(-z)| + |\langle c, x \rangle|}, \frac{\|H_D(b - Ax)\|}{1 + \|b\|}, \frac{\|c - A^*y - z\|}{1 + \|c\|} \right\} \leq \epsilon.$$

Figure 1 illustrates the performance comparison of tested algorithms with  $\sigma = 1$  on the “ex10” instance from Mittelmann’s LP benchmark. In particular, the left subfigure indicates that PR does not necessarily converge, while the middle subfigure shows that EPR does converge. Furthermore, to reach a solution with a tolerance of  $10^{-2.5}$ , EPR requires roughly half of the iterations of EDR, which is consistent with the ratio of the upper bounds for the ergodic iteration complexity results of the PR and DR splitting methods, as shown in (21) and (22) in Appendix B. More importantly, with the restart strategy, rEPR significantly outperforms both DR and EDR.



**Fig. 1** Performance comparison of algorithms ( $\sigma = 1$ ) on the “ex10” instance from Mittelmann’s LP benchmark

To further evaluate rEPR, rEDR, and DR, we test them on 49 instances from Mittelmann’s LP benchmark set. Each algorithm runs with a tolerance of  $\epsilon = 10^{-8}$  and a 3,600-second time limit. Figure 2 shows performance profiles for solving times. We observe that rEPR is the fastest solver for 50% of the problems, solves 60% of them using half of the time required by rEDR, and solves 20% more problems than DR. In summary, with the restart strategy, the ergodic sequence of the PR splitting method with semi-proximal terms outperforms both the point-wise and ergodic sequences of the DR splitting method with semi-proximal terms.



**Fig. 2** Performance profiles of solving times for tested algorithms ( $\sigma = 1$ ) on 49 instances from Mittelmann's LP benchmark set

## 5 Conclusion

To establish the ergodic convergence of the PR splitting method for solving the COP (1), we first proved the ergodic convergence of the dPPA with  $\rho \in (0, 2]$ . Leveraging the equivalence between the pADMM and the dPPA, we then demonstrated the ergodic convergence of the pADMM with  $\rho \in (0, 2]$ , including the PR splitting method with semi-proximal terms. Numerical results on the LP benchmark dataset revealed that, with a restart strategy, the ergodic sequence of the PR splitting method with semi-proximal terms consistently outperforms both the point-wise and ergodic sequences of the DR splitting method with semi-proximal terms. These results suggested that the restarted ergodic PR splitting method might be a more effective approach for solving large-scale COPs compared to its DR counterparts. For future research, it would be interesting to investigate why the restart strategy significantly enhances the performance of the PR splitting method's ergodic sequence.

## Statements and Declarations

**Funding** The work of Defeng Sun was supported by the Research Center for Intelligent Operations Research, RGC Senior Research Fellow Scheme No. SRFS2223-5S02, and GRF Project No. 15304721. The work of Yancheng Yuan was supported by Early Career Scheme Project No. 25305424 and the Research Center for Intelligent Operations Research. The work of Xinyuan Zhao was supported in part by the National Natural Science Foundation of China under Project No. 12271015.

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Adona, V.A., Gonçalves, M.L., Melo, J.G.: Iteration-complexity analysis of a generalized alternating direction method of multipliers. *J. Global Optim.* **73**, 331–348 (2019)

2. Applegate, D., Díaz, M., Hinder, O., Lu, H., Lubin, M., O'Donoghue, B., Schudy, W.: Practical large-scale linear programming using primal-dual hybrid gradient. In: *Advances in Neural Information Processing System*, vol. 34, pp. 20243–20257 (2021)
3. Baillon, J.B.: Un théorème de type ergodique pour les contractions non linéaires dans un espace de hilbert. *CR Acad. Sci. Paris Sér. AB* **280**, 1511–1514 (1975)
4. Bredies, K., Chenchene, E., Lorenz, D.A., Naldi, E.: Degenerate preconditioned proximal point algorithms. *SIAM J. Optim.* **32**(3), 2376–2401 (2022)
5. Burachik, R.S., Iusem, A.N., Svaiter, B.F.: Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.* **5**, 159–180 (1997)
6. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal–dual algorithm. *Math. Program.* **159**(1), 253–287 (2016)
7. Chen, K., Sun, D., Yuan, Y., Zhang, G., Zhao, X.: HPR-LP: An implementation of an HPR method for solving linear programming. arXiv preprint arXiv:2408.12179 (2024)
8. Cui, Y., Li, X., Sun, D., Toh, K.C.: On the convergence properties of a majorized alternating direction method of multipliers for linearly constrained convex optimization problems with coupled objective functions. *J. Optim. Theory Appl.* **169**(3), 1013–1041 (2016)
9. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. In: *Splitting Methods in Communication, Imaging, Science, and Engineering*, pp. 115–163. Springer (2016)
10. Dong, Y., Fischer, A.: A family of operator splitting methods revisited. *Nonlinear Anal.* **72**(11), 4307–4315 (2010)
11. Eckstein, J., Bertsekas, D.P.: On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* **55**(1), 293–318 (1992)
12. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications to system identification and realization. *SIAM J. Matrix Anal. Appl.* **34**(3), 946–977 (2013)
13. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: M. Fortin, R. Glowinski (eds.) *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, *Studies in Mathematics and Its Applications*, vol. 15, pp. 299–331. North-Holland Publishing Company (1983)
14. Giselsson, P., Boyd, S.: Linear convergence and metric selection for Douglas-Rachford splitting and ADMM. *IEEE Trans. Automat. Control* **62**(2), 532–544 (2016)
15. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**(6), 964–979 (1979)
16. Monteiro, R.D., Sim, C.K.: Complexity of the relaxed Peaceman–Rachford splitting method for the sum of two maximal strongly monotone operators. *Comput. Optim. Appl.* **70**, 763–790 (2018)
17. Monteiro, R.D., Svaiter, B.F.: Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.* **23**(1), 475–507 (2013)
18. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings (1967)
19. Rockafellar, R.T.: *Convex Analysis*, vol. 18. Princeton University Press (1970)
20. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**(5), 877–898 (1976)
21. Shen, L., Pan, S.: Weighted iteration complexity of the sPADMM on the KKT residuals for convex composite optimization. arXiv preprint arXiv:1611.03167 (2016)
22. Sim, C.K.: Convergence rates for the relaxed Peaceman-Rachford splitting method on a monotone inclusion problem. *J. Optim. Theory Appl.* **196**(1), 298–323 (2023)
23. Sun, D., Yuan, Y., Zhang, G., Zhao, X.: Accelerating preconditioned ADMM via degenerate proximal point mappings. arXiv preprint arXiv:2403.18618 (2024)
24. Xiao, Y., Chen, L., Li, D.: A generalized alternating direction method of multipliers with semi-proximal terms for convex composite conic programming. *Math. Program. Comput.* **10**, 533–555 (2018)

## A An analytical example of the ergodic convergence of the PR splitting method

The following example illustrates the challenge of analyzing the ergodic sequence of the PR splitting method.

*Example 1*

$$\begin{aligned} \min_{y \in \mathbb{R}, z \in \mathbb{R}} \quad & \delta_{\{0\}}(z) \\ \text{s.t.} \quad & y + z = 0. \end{aligned} \quad (18)$$

The optimal solution to problem (18) and its dual problem is  $(y^*, z^*, x^*) = (0, 0, 0)$ . Choosing  $\mathcal{T}_1 = 0$ ,  $\mathcal{T}_2 = 0$ , and initializing with  $(y^0, z^0, x^0) = (1, 0, 1)$  and  $\sigma = 1$ , we apply the pADMM in Algorithm 1 to solve problem (18). Through direct calculations, we obtain the following for any  $k \geq 0$ :

$$\left\{ \begin{array}{l} \bar{y}^k = -2(1-\rho)^k, \\ \bar{z}^k = 0, \\ \bar{x}^k = 2(1-\rho)^k, \\ y^k = -2k\rho(1-\rho)^{k-1} + (1-\rho)^k, \\ z^k = 0, \\ x^k = 2k\rho(1-\rho)^{k-1} + (1-\rho)^k, \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} \bar{y}_a^k = -\frac{2(1-(1-\rho)^{k+1})}{\rho(k+1)}, \\ \bar{z}_a^k = 0, \\ \bar{x}_a^k = \frac{2(1-(1-\rho)^{k+1})}{\rho(k+1)}, \\ y_a^k = 2(1-\rho)^k - \frac{1-(1-\rho)^{k+1}}{\rho(k+1)}, \\ z_a^k = 0, \\ x_a^k = -2(1-\rho)^k + \frac{3(1-(1-\rho)^{k+1})}{\rho(k+1)}, \end{array} \right.$$

where  $0^0$  is defined as 1. In particular, if  $\rho = 2$ , then for  $k \geq 0$ , we have

$$\left\{ \begin{array}{l} \bar{y}^k = -2(-1)^k, \\ \bar{z}^k = 0, \\ \bar{x}^k = 2(-1)^k, \\ y^k = -4k(-1)^{k-1} + (-1)^k, \\ z^k = 0, \\ x^k = 4k(-1)^{k-1} + (-1)^k, \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{l} \bar{y}_a^k = -\frac{1-(-1)^{k+1}}{k+1}, \\ \bar{z}_a^k = 0, \\ \bar{x}_a^k = \frac{1-(-1)^{k+1}}{k+1}, \\ y_a^k = 2(-1)^k - \frac{1-(-1)^{k+1}}{2(k+1)}, \\ z_a^k = 0, \\ x_a^k = -2(-1)^k + \frac{3(1-(-1)^{k+1})}{2(k+1)}. \end{array} \right.$$

It is evident that the point-wise sequence  $\{(\bar{x}^k, \bar{y}^k)\}$  of the PR splitting method ( $\rho = 2$ ) oscillates, while  $\{(x^k, y^k)\}$  diverges to infinity. In contrast, the ergodic sequence  $\{(\bar{x}_a^k, \bar{y}_a^k)\}$  of the PR splitting method converges to the optimal solution, whereas  $\{(x_a^k, y_a^k)\}$  still diverges. Moreover, the ergodic sequence  $\{(\bar{x}_a^k, \bar{y}_a^k)\}$  of the PR splitting method performs better than that of the DR splitting method ( $\rho = 1$ ), in the sense that the odd ergodic iterations of the PR splitting method reach the solution directly:

$$(\bar{x}_a^k, \bar{y}_a^k) = \begin{cases} \left( -\frac{1-(-1)^{k+1}}{k+1}, \frac{1-(-1)^{k+1}}{k+1} \right), & \rho = 2, \\ \left( -\frac{2}{k+1}, \frac{2}{k+1} \right), & \rho = 1. \end{cases}$$

In summary, this example illustrates that the PR splitting method may fail to achieve point-wise convergence when solving COP (1), while clearly highlighting the superiority of its ergodic sequences compared to those of the DR splitting method. Moreover, it reveals that the ergodic sequence  $\{w_a^k\}$  of the PR splitting method does not necessarily converge for general COPs (1). In this sense, the ergodic convergence result of the PR splitting method established in Corollary 1 represents the best achievable outcome.

## B Ergodic iteration complexity of the PR splitting method for COPs

To theoretically justify the superior performance of the ergodic PR splitting method compared to the ergodic DR splitting method, this section focuses on analyzing the ergodic

iteration complexity of the pADMM framework, which includes the PR splitting method. We begin by introducing the concept of the  $\varepsilon$ -subgradient of a convex function  $f$  [19]:

**Definition 1** Let  $f : \mathbb{X} \rightarrow (-\infty, +\infty]$  be a proper convex function, and let  $\bar{x} \in \text{dom}(f)$ . Given  $\varepsilon \geq 0$ , the  $\varepsilon$ -subgradient of  $f$  at  $\bar{x}$  is defined as

$$\partial_\varepsilon f(\bar{x}) := \{x^* \in \mathbb{X}^* \mid \langle x^*, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) + \varepsilon, \forall x \in \mathbb{X}\}.$$

Using the optimality conditions of each subproblem in the pADMM and the concept of  $\varepsilon$ -subgradient, we derive the following lemma for the ergodic sequence  $\{\bar{w}_a^k\}$ .

**Lemma 1** Suppose that Assumptions 1 and 2 hold. Let  $w^* = (y^*, z^*, x^*)$  be a solution to the KKT system (13). Then the sequence  $\{(\bar{y}_a^k, \bar{z}_a^k, \bar{x}_a^k)\}$  generated by the pADMM in Algorithm 1 with  $\rho \in (0, 2]$  satisfies for any  $k \geq 0$ ,

$$\begin{cases} -B_2^* \bar{x}_a^k - \mathcal{T}_2(\bar{z}_a^k - z_a^k) \in \partial_{\bar{\varepsilon}_z^k} f_2(\bar{z}_a^k), \\ -B_1^*(\bar{x}_a^k + \sigma(B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c)) - \mathcal{T}_1(\bar{y}_a^k - y_a^k) \in \partial_{\bar{\varepsilon}_y^k} f_1(\bar{y}_a^k), \end{cases}$$

where

$$\begin{cases} \bar{\varepsilon}_z^k = \frac{1}{k+1} \sum_{t=0}^k \langle -B_2^* \bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), \bar{z}^t - \bar{z}_a^k \rangle \geq 0, \\ \bar{\varepsilon}_y^k = \frac{1}{k+1} \sum_{t=0}^k \langle -B_1^*(\bar{x}^t + \sigma(B_1 \bar{y}^t + B_2 \bar{z}^t - c)) - \mathcal{T}_1(\bar{y}^t - y^t), \bar{y}^t - \bar{y}_a^k \rangle \geq 0, \end{cases} \quad (19)$$

and

$$\bar{\varepsilon}_z^k + \bar{\varepsilon}_y^k \leq \frac{1}{2\rho(k+1)} \|w^0 - w^*\|_{\mathcal{M}}^2. \quad (20)$$

*Proof* From the optimality conditions of the subproblems in Algorithm 1, we have, for any  $t \geq 0$ ,

$$\begin{cases} f_2(z) \geq f_2(\bar{z}^t) + \langle -B_2^* \bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), z - \bar{z}^t \rangle, \quad \forall z \in \mathbb{Z}, \\ f_1(y) \geq f_1(\bar{y}^t) + \langle -B_1^*(\bar{x}^t + \sigma(B_1 \bar{y}^t + B_2 \bar{z}^t - c)) - \mathcal{T}_1(\bar{y}^t - y^t), y - \bar{y}^t \rangle, \quad \forall y \in \mathbb{Y}. \end{cases}$$

Summing these from  $t = 0$  to  $k$ , and dividing by  $k + 1$ , we obtain for any  $z \in \mathbb{Z}$  and  $y \in \mathbb{Y}$ ,

$$\begin{cases} f_2(z) \geq \frac{1}{k+1} \sum_{t=0}^k \left( f_2(\bar{z}^t) + \langle -B_2^* \bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), z - \bar{z}^t \rangle \right), \\ f_1(y) \geq \frac{1}{k+1} \sum_{t=0}^k \left( f_1(\bar{y}^t) + \langle -B_1^*(\bar{x}^t + \sigma(B_1 \bar{y}^t + B_2 \bar{z}^t - c)) - \mathcal{T}_1(\bar{y}^t - y^t), y - \bar{y}^t \rangle \right). \end{cases}$$

By the convexity of  $f_2(\cdot)$ , we have for any  $z \in \mathbb{Z}$ ,

$$\begin{aligned} f_2(z) &\geq f_2(\bar{z}_a^k) + \frac{1}{k+1} \sum_{t=0}^k \langle -B_2^* \bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), z - \bar{z}^t \rangle \\ &= f_2(\bar{z}_a^k) + \frac{1}{k+1} \sum_{t=0}^k \langle -B_2^* \bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), z - \bar{z}_a^k \rangle - \bar{\varepsilon}_z^k, \end{aligned}$$

where

$$\bar{\varepsilon}_z^k = \frac{1}{k+1} \sum_{t=0}^k \langle -B_2^* \bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), \bar{z}^t - \bar{z}_a^k \rangle$$

is non-negative by substituting  $z = \bar{z}_a^k$  in the first inequality. Hence, we have for any  $k \geq 0$ ,

$$-(B_2^* \bar{x}_a^k + \mathcal{T}_2(\bar{z}_a^k - z_a^k)) = \frac{1}{k+1} \sum_{t=0}^k -(B_2^* \bar{x}^t + \mathcal{T}_2(\bar{z}^t - z^t)) \in \partial_{\bar{\varepsilon}_z^k} f_2(\bar{z}_a^k).$$

Similarly, for  $f_1(\cdot)$ , we obtain that for all  $k \geq 0$ ,

$$-B_1^*(\bar{x}_a^k + \sigma(B_1\bar{y}_a^k + B_2\bar{z}_a^k - c)) - \mathcal{T}_1(\bar{y}_a^k - y_a^k) \in \partial_{\bar{\varepsilon}_y^k} f_1(\bar{y}_a^k),$$

where

$$\bar{\varepsilon}_y^k = \frac{1}{k+1} \sum_{t=0}^k \langle -B_1^*(\bar{x}^t + \sigma(B_1\bar{y}^t + B_2\bar{z}^t - c)) - \mathcal{T}_1(\bar{y}^t - y^t), \bar{y}^t - \bar{y}_a^k \rangle \geq 0.$$

Now, we show the upper bound of  $\bar{\varepsilon}_z^k + \bar{\varepsilon}_y^k$  for all  $k \geq 0$ . According to the definitions of  $\bar{\varepsilon}_z^k$  and  $\bar{\varepsilon}_y^k$  in (19), Step 2 of Algorithm 1, and the definition of  $\mathcal{M}$  in (15), we obtain for any  $k \geq 0$ ,

$$\begin{aligned} & \bar{\varepsilon}_z^k + \bar{\varepsilon}_y^k \\ &= \frac{1}{k+1} \sum_{t=0}^k \left( \langle -B_2^*\bar{x}^t - \mathcal{T}_2(\bar{z}^t - z^t), \bar{z}^t - \bar{z}_a^k \rangle \right. \\ & \quad \left. + \langle -B_1^*(\bar{x}^t + \sigma(B_1\bar{y}^t + B_2\bar{z}^t - c)) - \mathcal{T}_1(\bar{y}^t - y^t), \bar{y}^t - \bar{y}_a^k \rangle \right) \\ &= \frac{1}{k+1} \sum_{t=0}^k \left( \langle \mathcal{M}(w^t - \bar{w}^t), \bar{w}^t - \bar{w}_a^k \rangle \right. \\ & \quad \left. - \langle B_1^*\bar{x}^t, \bar{y}^t - \bar{y}_a^k \rangle - \langle B_2^*\bar{x}^t, \bar{z}^t - \bar{z}_a^k \rangle - \langle c - B_1\bar{y}^t - B_2\bar{z}^t, \bar{x}^t - \bar{x}_a^k \rangle \right) \\ &= \frac{1}{k+1} \sum_{t=0}^k \langle \mathcal{M}(w^t - \bar{w}^t), \bar{w}^t - \bar{w}_a^k \rangle. \end{aligned}$$

Thus, by the definition of  $\bar{\varepsilon}_a^k$  in Proposition 1, and the equivalence between the pADMM and the dPPA in Proposition 2, we can derive

$$\bar{\varepsilon}_z^k + \bar{\varepsilon}_y^k = \bar{\varepsilon}_a^k \leq \frac{1}{2\rho(k+1)} \|w^0 - w^*\|_{\mathcal{M}}^2.$$

This completes the proof.

To estimate the objective error, we define

$$h(\bar{y}_a^k, \bar{z}_a^k) := f_1(\bar{y}_a^k) + f_2(\bar{z}_a^k) - f_1(y^*) - f_2(z^*), \quad \forall k \geq 0,$$

where  $(y^*, z^*)$  is a solution to the COP (1). Based on the ergodic properties established in Lemma 1, we derive the following iteration complexity with respect to the objective error, the feasibility violation, and the KKT residual based on  $\varepsilon$ -subdifferential in Theorem 2.

**Theorem 2** *Suppose that Assumptions 1 and 2 hold. Let  $w^* = (y^*, z^*, x^*)$  be a solution to the KKT system (13), and define  $R_0 = \|w^0 - w^*\|_{\mathcal{M}}$ . Then, the sequence  $\{(\bar{y}_a^k, \bar{z}_a^k, \bar{x}_a^k)\}$  generated by the pADMM in Algorithm 1 with  $\rho \in (0, 2]$  satisfies the following iteration complexity bounds for all  $k \geq 0$ :*

$$\begin{aligned} & \text{dist}(0, \partial_{\bar{\varepsilon}_y^k} f_1(\bar{y}_a^k) + B_1^*\bar{x}_a^k) + \text{dist}(0, \partial_{\bar{\varepsilon}_z^k} f_2(\bar{z}_a^k) + B_2^*\bar{x}_a^k) + \|B_1\bar{y}_a^k + B_2\bar{z}_a^k - c\| \\ & \leq \left( \frac{\sigma\|B_1^*\| + 1}{\sqrt{\sigma}} + \|\sqrt{\mathcal{T}_2}\| + \|\sqrt{\mathcal{T}_1}\| \right) \frac{2R_0}{\rho(k+1)}, \end{aligned} \quad (21)$$

where  $\bar{\varepsilon}_z^k + \bar{\varepsilon}_y^k \leq \frac{1}{2\rho(k+1)} \|w^0 - w^*\|_{\mathcal{M}}^2$ . Moreover, the following bound holds for the objective function:

$$\begin{aligned} & \left( -\frac{\|x^*\|}{\sqrt{\sigma}} \right) \frac{2R_0}{\rho(k+1)} \leq h(\bar{y}_a^k, \bar{z}_a^k) \\ & \leq (R_0 + 4\sqrt{\sigma}\|B_1y^*\|) \frac{R_0}{2\rho(k+1)} + \frac{\|x^0 + \sigma B_1y^0\|^2}{2\rho(k+1)}. \end{aligned} \quad (22)$$

*Proof* According to Propositions 1 and 2, we have

$$\|\bar{w}_a^k - w_a^k\|_{\mathcal{M}}^2 \leq \frac{4R_0^2}{\rho^2(k+1)^2}, \quad \forall k \geq 0.$$

By the definition of  $\mathcal{M}$  in (15), this can be rewritten as

$$\|\bar{y}_a^k - y_a^k\|_{\mathcal{T}_1}^2 + \frac{1}{\sigma} \|\sigma B_1(\bar{y}_a^k - y_a^k) + (\bar{x}_a^k - x_a^k)\|^2 + \|\bar{z}_a^k - z_a^k\|_{\mathcal{T}_2}^2 \leq \frac{4R_0^2}{\rho^2(k+1)^2}, \quad \forall k \geq 0. \quad (23)$$

From Step 2 of Algorithm 1, we can deduce that for any  $k \geq 0$ ,

$$\begin{aligned} \|\sigma B_1(\bar{y}_a^k - y_a^k) + (\bar{x}_a^k - x_a^k)\| &= \|\sigma B_1(\bar{y}_a^k - y_a^k) + \sigma(B_1 y_a^k + B_2 \bar{z}_a^k - c)\| \\ &= \sigma \|B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c\|, \end{aligned}$$

which, together with (23), yields that

$$\|B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c\| \leq \frac{2R_0}{\sqrt{\sigma}\rho(k+1)}, \quad \forall k \geq 0. \quad (24)$$

Furthermore, according to the Lemma 1, we have for  $k \geq 0$ ,

$$\begin{cases} -B_2^* \bar{x}_a^k - \mathcal{T}_2(\bar{z}_a^k - z_a^k) \in \partial_{\bar{\varepsilon}^k} f_2(\bar{z}_a^k), \\ -B_1^*(\bar{x}_a^k + \sigma(B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c)) - \mathcal{T}_1(\bar{y}_a^k - y_a^k) \in \partial_{\bar{\varepsilon}^k} f_1(\bar{y}_a^k), \end{cases}$$

which, together with (23) and (24), implies

$$\text{dist}(0, \partial_{\bar{\varepsilon}^k} f_2(\bar{z}_a^k) + B_2^* \bar{x}_a^k) \leq \|\mathcal{T}_2(\bar{z}_a^k - z_a^k)\| \leq \|\sqrt{\mathcal{T}_2}\| \|\bar{z}_a^k - z_a^k\|_{\mathcal{T}_2} \leq \|\sqrt{\mathcal{T}_2}\| \frac{2R_0}{\rho(k+1)} \quad (25)$$

and

$$\begin{aligned} \text{dist}(0, \partial_{\bar{\varepsilon}^k} f_1(\bar{y}_a^k) + B_1^* \bar{x}_a^k) &\leq \sigma \|B_1^*\| \|B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c\| + \|\mathcal{T}_1(\bar{y}_a^k - y_a^k)\| \\ &\leq \left( \frac{\sigma \|B_1^*\|}{\sqrt{\sigma}} + \|\sqrt{\mathcal{T}_1}\| \right) \frac{2R_0}{\rho(k+1)}. \end{aligned} \quad (26)$$

Thus, combining (20), (24), (25), and (26), we derive the iteration complexity bound in (21).

We now estimate the ergodic iteration complexity results for the objective error. From the KKT conditions in (13), we have, for any  $k \geq 0$ ,

$$f_1(\bar{y}_a^k) - f_1(y^*) \geq \langle -B_1^* x^*, \bar{y}_a^k - y^* \rangle \text{ and } f_2(\bar{z}_a^k) - f_2(z^*) \geq \langle -B_2^* x^*, \bar{z}_a^k - z^* \rangle.$$

Thus, it follows from (24) that for all  $k \geq 0$ ,

$$h(\bar{y}_a^k, \bar{z}_a^k) \geq \langle B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c, -x^* \rangle \geq -\|x^*\| \|B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c\| \geq -\frac{2R_0 \|x^*\|}{\sqrt{\sigma}\rho(k+1)}.$$

For the upper bound of the objective error, from [23, Lemma 3.6], we first have the following upper bounds:

$$\begin{aligned} h(\bar{y}^k, \bar{z}^k) &\leq \langle \sigma B_1(y^* - \bar{y}^k) - \bar{x}^k, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle \\ &\quad + \langle y^* - \bar{y}^k, \mathcal{T}_1(\bar{y}^k - y^k) \rangle + \langle z^* - \bar{z}^k, \mathcal{T}_2(\bar{z}^k - z^k) \rangle. \end{aligned} \quad (27)$$

Note that from Step 4 of Algorithm 1 and  $\rho \in (0, 2]$ , we have for any  $k \geq 0$ ,

$$\begin{aligned} \langle y^* - \bar{y}^k, \mathcal{T}_1(\bar{y}^k - y^k) \rangle &= \langle y^* - (y^k + \frac{y^{k+1} - y^k}{\rho}), \mathcal{T}_1(\frac{y^{k+1} - y^k}{\rho}) \rangle \\ &= \frac{1}{2\rho} (\|y^k - y^*\|_{\mathcal{T}_1}^2 - \|y^{k+1} - y^*\|_{\mathcal{T}_1}^2) + \frac{\rho - 2}{2\rho^2} \|y^{k+1} - y^k\|_{\mathcal{T}_1}^2 \\ &\leq \frac{1}{2\rho} (\|y^k - y^*\|_{\mathcal{T}_1}^2 - \|y^{k+1} - y^*\|_{\mathcal{T}_1}^2). \end{aligned} \quad (28)$$

Similarly, we also have

$$\langle z^* - \bar{z}^k, \mathcal{T}_2(\bar{z}^k - z^k) \rangle \leq \frac{1}{2\rho} (\|z^k - z^*\|_{\mathcal{T}_1}^2 - \|z^{k+1} - z^*\|_{\mathcal{T}_1}^2), \quad \forall k \geq 0. \quad (29)$$

Additionally, for ease of notation, we define

$$\Delta_k := x^k + \sigma B_1 y^k, \quad \forall k \geq 0.$$

From Step 4 of Algorithm 1 and  $\rho \in (0, 2]$ , we can derive that for any  $k \geq 0$ ,

$$\begin{aligned} & \langle \sigma B_1(y^* - \bar{y}^k) - \bar{x}^k, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle \\ &= \langle \sigma B_1 y^*, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle - \langle \bar{x}^k + \sigma B_1 \bar{y}^k, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle \\ &= \langle \sigma B_1 y^*, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle - \langle \Delta_k + \frac{\Delta_{k+1} - \Delta_k}{\rho}, \frac{\Delta_{k+1} - \Delta_k}{\rho} \rangle \\ &= \langle \sigma B_1 y^*, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle - \frac{1}{2\rho} (\|\Delta_{k+1}\|^2 - \|\Delta_k\|^2) + \frac{\rho - 2}{2\rho^2} \|\Delta_{k+1}\|^2 \\ &\leq \langle \sigma B_1 y^*, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle - \frac{1}{2\rho} (\|\Delta_{k+1}\|^2 - \|\Delta_k\|^2). \end{aligned} \quad (30)$$

Thus, combing with (27), (28), (29), and (30), we conclude that for all  $k \geq 0$ ,

$$\begin{aligned} h(\bar{y}^k, \bar{z}^k) &\leq \frac{1}{2\rho} (\|y^k - y^*\|_{\mathcal{T}_1}^2 - \|y^{k+1} - y^*\|_{\mathcal{T}_1}^2) + \frac{1}{2\rho} (\|z^k - z^*\|_{\mathcal{T}_1}^2 - \|z^{k+1} - z^*\|_{\mathcal{T}_1}^2) \\ &\quad + \langle \sigma B_1 y^*, B_1 \bar{y}^k + B_2 \bar{z}^k - c \rangle - \frac{1}{2\rho} (\|\Delta_{k+1}\|^2 - \|\Delta_k\|^2). \end{aligned}$$

It follows from the convexity of  $h(\cdot)$  and (24) that for any  $k \geq 0$ ,

$$\begin{aligned} h(\bar{y}_a^k, \bar{z}_a^k) &\leq \frac{1}{k+1} \sum_{t=0}^k h(\bar{y}^t, \bar{z}^t) \\ &\leq \frac{1}{2\rho(k+1)} (\|y^0 - y^*\|_{\mathcal{T}_1}^2 + \|z^0 - z^*\|_{\mathcal{T}_1}^2) + \langle \sigma B_1 y^*, B_1 \bar{y}_a^k + B_2 \bar{z}_a^k - c \rangle + \frac{1}{2\rho(k+1)} \|\Delta_0\|^2 \\ &\leq \frac{R_0^2}{2\rho(k+1)} + \sqrt{\sigma} \|B_1 y^*\| \frac{2R_0}{\rho(k+1)} + \frac{1}{2\rho(k+1)} \|x^0 + \sigma B_1 y^0\|^2. \end{aligned}$$

This completes the proof.

*Remark 2* From the ergodic iteration complexity results in (21) and (22), the ergodic sequence of the PR splitting method with semi-proximal terms has a worst-case upper bound that is half that of the DR splitting method with semi-proximal terms. This provides theoretical support for the superior performance of the ergodic PR splitting method with semi-proximal terms compared to the ergodic DR splitting method with semi-proximal terms, as demonstrated in the numerical experiments in Section 4.

*Remark 3* We summarize some existing complexity results for ergodic sequences of ADMM-type algorithms closely related to our work in Table 1. For more results, one can refer to [1, 6] and the references in.

**Table 1** Ergodic iteration complexity results of ADMM-type algorithms

Paper	Algorithm	Proximal Operators	$\rho$	Dual Step-size	Feasibility Violation	Objective Error	KKT Residual
[17]	ADMM	$\mathcal{T}_1 = 0, \mathcal{T}_2 = 0$	1	1	$O(1/k)$	-	$O_\varepsilon(1/k)^1$
[9]	GADMM	$\mathcal{T}_1 = 0, \mathcal{T}_2 = 0$	(0, 2]	1	$O(1/k)$	$O(1/k)$	-
[8]	(Majorized) sPADMM	$\mathcal{T}_1 \geq 0, \mathcal{T}_2 \geq 0$	1	$(0, \frac{1+\sqrt{5}}{2})$	$O(1/k)$	$O(1/k)$	-
[21]	sPADMM	$\mathcal{T}_1 \geq 0, \mathcal{T}_2 \geq 0$	1	$(0, \frac{1+\sqrt{5}}{2})$	$O(1/k)$	-	$O_\varepsilon(1/k)$
Ours	pADMM	$\mathcal{T}_1 \geq 0, \mathcal{T}_2 \geq 0$	(0, 2]	1	$O(1/k)$	$O(1/k)$	$O_\varepsilon(1/k)$

When  $\mathcal{T}_1 = 0$  and  $\mathcal{T}_2 = 0$ , pADMM with  $\rho = 2$  reduces to the PR splitting method, which corresponds to the GADMM [11] with  $\rho = 2$ . According to the results of Davis and Yin [9], the ergodic sequence of the PR splitting method achieves an  $O(1/k)$  complexity in terms of the objective error and the feasibility violation. In contrast, we establish the ergodic iteration complexity result for the PR splitting method with semi-proximal terms with respect to the objective error, the feasibility violation, and the KKT residual based on  $\varepsilon$ -subdifferential. This generalization is particularly significant, as carefully chosen  $\mathcal{T}_1$  and  $\mathcal{T}_2$  can simplify the solution of subproblems in some important convex optimization problems, such as general LPs.

---

<sup>1</sup>  $O_\varepsilon(1/k)$  of the KKT residual: an  $O(1/k)$  iteration complexity of the KKT residual based on  $\varepsilon$ -subdifferential in (21).