CrossMark

Complexity of the relaxed Peaceman–Rachford splitting method for the sum of two maximal strongly monotone operators

Renato D. C. Monteiro[1] · **Chee-Khian Sim**[2]

Abstract This paper considers the relaxed Peaceman–Rachford (PR) splitting method for finding an approximate solution of a monotone inclusion whose underlying operator consists of the sum of two maximal strongly monotone operators. Using general results obtained in the setting of a non-Euclidean hybrid proximal extragradient framework, we extend a previous convergence result on the iterates generated by the relaxed PR splitting method, as well as establish new pointwise and ergodic convergence rate results for the method whenever an associated relaxation parameter is within a certain interval. An example is also discussed to demonstrate that the iterates may not converge when the relaxation parameter is outside this interval.

Keywords Relaxed Peaceman–Rachford splitting method · Strongly monotone operators · Non-Euclidean hybrid proximal extragradient framework

✉ Chee-Khian Sim
chee-khian.sim@port.ac.uk

Renato D. C. Monteiro
monteiro@isye.gatech.edu

[1] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

[2] Department of Mathematics, University of Portsmouth, Lion Gate Building, Lion Terrace, Portsmouth PO1 3HF, UK

🖄 Springer

1 Introduction

In this paper, we consider the relaxed Peaceman–Rachford (PR) splitting method for solving the monotone inclusion

$$0 \in (A + B)(u) \tag{1}$$

where $A : \mathcal{X} \rightrightarrows \mathcal{X}$ and $B : \mathcal{X} \rightrightarrows \mathcal{X}$ are maximal β-strongly monotone (point-to-set) operators for some $\beta \geq 0$ (with the convention that 0-strongly monotone means simply monotone, and β-strongly monotone with $\beta > 0$ means strongly monotone in the usual sense). Recall that the relaxed PR splitting method is given by

$$x_k = x_{k-1} + \theta(J_B(2J_A(x_{k-1}) - x_{k-1}) - J_A(x_{k-1})), \tag{2}$$

where $\theta > 0$ is a fixed relaxation parameter and $J_T := (I + T)^{-1}$. The special case of the relaxed PR splitting method in which $\theta = 2$ is known as the Peaceman–Rachford (PR) splitting method and the one with $\theta = 1$ is the widely-studied Douglas–Rachford (DR) splitting method. Convergence results for them are studied for example in [1–4,8,13,14,22].

The analysis of the relaxed PR splitting method for the case in which $\beta = 0$ has been undertaken in a number of papers which are discussed in this paragraph. Convergence of the sequence of iterates generated by the relaxed PR splitting method is well-known when $\theta < 2$ (see for example [1,7,14]) and, according to [16], its limiting behavior for the case in which $\theta \geq 2$ is not known. We actually show in Sect. 5.2 that the sequence (2) does not necessarily converge when $\theta \geq 2$. An $\mathcal{O}(1/\sqrt{k})$ (strong) pointwise convergence rate result is established in [18] for the relaxed PR splitting method when $\theta \in (0, 2)$. Moreover, when $A = \partial f$ and $B = \partial g$ where f and g are proper lower semi-continuous convex functions, papers [9–11] derive strong pointwise (resp., ergodic) convergence rate bounds for the relaxed PR method when $\theta \in (0, 2)$ (resp., $\theta \in (0, 2]$) under different assumptions on the functions. Assuming only β-strong monotonicity of $A = \partial f$, where $\beta > 0$, some smoothness property on f, and maximal monotonicity of B, [16] shows that the relaxed PR splitting method has linear convergence rate for $\theta \in (0, 2 + \tau)$ for some $\tau > 0$. Linear rate of convergence of the relaxed PR splitting method and its two special cases, namely, the DR splitting and PR splitting methods, are established in [2–4,11,15,16,22] under relatively strong assumptions on A and/or B (see also Table 2).

This paper assumes that $\beta \geq 0$, and hence its analysis applies to the case in which both A and B are monotone ($\beta = 0$) and the case in which both A and B are strongly monotone ($\beta > 0$). This paragraph discusses papers dealing with the latter case. Paper [12] establishes convergence of the sequence generated by the relaxed PR splitting method for any $\theta \in (0, 2 + \beta)$ and, under some strong assumptions on A and B, establishes its linear convergence rate. We complement the convergence results in [12] by showing that for $\theta = 2 + \beta$, the sequence of iterates generated by the relaxed PR splitting method also converge, and describe an instance showing its nonconvergence when $\theta \geq \min\{2 + 2\beta, 2 + \beta + 1/\beta\}$. Moreover, we establish strong pointwise and

ergodic convergence rate results (Theorems 4.6 and 4.8) for the relaxed PR splitting method when $\theta \in (0, 2 + \beta)$ and $\theta \in (0, 2 + \beta]$, respectively.

Finally, by imposing strong assumptions requiring one of the operators to be strong monotone and one of them to be Lipschitz (and hence point-to-point), [11,15,16] establish linear convergence rate of the relaxed PR splitting method. As opposed to these papers, the assumptions in [12] and this paper do not imply the operators A or B to be point-to-point.

Our analysis of the relaxed PR splitting method for solving (1) is based on viewing it as an inexact proximal point method, more specifically, as an instance of a non-Euclidean hybrid proximal extragradient (HPE) framework for solving the monotone inclusion problem. The proximal point method, proposed by Rockafellar [29], is a classical iterative scheme for solving the latter problem. Paper [30] introduces an Euclidean version of the HPE framework which is an inexact version of the proximal point method based on a certain relative error criterion. Iteration-complexities of the latter framework are established in [25] (see also [26]). Generalizations of the HPE framework to the non-Euclidean setting are studied in [17,21,31]. Applications of the HPE framework can be found for example in [19,20,25,26].

This paper is organized as follows. Section 2 describes basic concepts and notation used in the paper. Section 3 discusses the non-Euclidean HPE framework which is used to the study the convergence properties of the relaxed PR splitting method in Sects. 4 and 5. Section 4 derives convergence rate bounds for the relaxed Peaceman–Rachford (PR) splitting method. Section 5, which consists of two subsections, discusses a convergence result of the relaxed PR splitting method in the first subsection and provides an example showing that its iterates may not converge when $\theta \geq \min\{2 + 2\beta, 2 + \beta + 1/\beta\}$ in the second subsection. Finally, Sect. 6 discusses the numerical performance of the relaxed PR splitting method for solving the weighted Lasso minimization problem. Section 7 gives some concluding remarks.

2 Basic concepts and notation

This section presents some definitions, notation and terminology which will be used in the paper.

We denote the set of real numbers by $\mathbb{R}$ and the set of non-negative real numbers by $\mathbb{R}_+$. Let f and g be functions with the same domain and whose values are in $\mathbb{R}_+$. We write that $f(\cdot) = \Omega(g(\cdot))$ if there exists constant $K > 0$ such that $f(\cdot) \geq Kg(\cdot)$. Also, we write $f(\cdot) = \Theta(g(\cdot))$ if $f(\cdot) = \Omega(g(\cdot))$ and $g(\cdot) = \Omega(f(\cdot))$.

Let $\mathcal{Z}$ be a finite-dimensional real vector space with inner product denoted by $\langle \cdot, \cdot \rangle$ (an example of $\mathcal{Z}$ is $\mathbb{R}^n$ endowed with the standard inner product) and let $\|\cdot\|$ denote an arbitrary seminorm in $\mathcal{Z}$. Its dual (extended) seminorm, denoted by $\|\cdot\|_*$, is defined as $\|\cdot\|_* := \sup\{\langle \cdot, z \rangle : \|z\| \leq 1\}$. It is easy to see that

$$\langle z, v \rangle \leq \|z\|\|v\|_* \quad \forall z, v \in \mathcal{Z}. \tag{3}$$

The following straightforward result states some basic properties of the dual seminorm associated with a matrix seminorm. Its proof can be found for example in Lemma A.1(b) of [23].

Proposition 2.1 *Let $A : \mathcal{Z} \to \mathcal{Z}$ be a self-adjoint positive semidefinite linear operator and consider the seminorm $\| \cdot \|$ in $\mathcal{Z}$ given by $\|z\| = \langle Az, z \rangle^{1/2}$ for every $z \in \mathcal{Z}$. Then,* dom $\| \cdot \|_* = Im\,(A)$ *and* $\|Az\|_* = \|z\|$ *for every $z \in \mathcal{Z}$.*

Given a set-valued operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$, its domain is denoted by $\mathrm{Dom}(T) := \{z \in \mathcal{Z} : T(z) \neq \emptyset\}$ and its inverse operator $T^{-1} : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is given by $T^{-1}(v) := \{z : v \in T(z)\}$. The graph of T is defined by $\mathrm{Gr}(T) := \{(z, t) : t \in T(z)\}$. The operator T is said to be monotone if

$$\langle z - z', t - t' \rangle \geq 0 \quad \forall (z, t), (z', t') \in \mathrm{Gr}(T).$$

Moreover, T is maximal monotone if it is monotone and, additionally, if T' is a monotone operator such that $T(z) \subset T'(z)$ for every $z \in \mathcal{Z}$, then $T = T'$. The sum $T + T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$ of two set-valued operators $T, T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is defined by $(T + T')(z) := \{t + t' \in \mathcal{Z} : t \in T(z), \ t' \in T'(z)\}$ for every $z \in \mathcal{Z}$. Given a scalar $\varepsilon \geq 0$, the ε-enlargement $T^{[\varepsilon]} : \mathcal{Z} \rightrightarrows \mathcal{Z}$ of a monotone operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is defined as

$$T^{[\varepsilon]}(z) := \{t \in \mathcal{Z} : \langle t - t', z - z' \rangle \geq -\varepsilon, \ \forall z' \in \mathcal{Z}, \ \forall t' \in T(z')\} \quad \forall z \in \mathcal{Z}. \quad (4)$$

3 A non-Euclidean hybrid proximal extragradient framework

This section discusses the non-Euclidean hybrid proximal extragradient (NE-HPE) framework and describes its associated convergence and iteration complexity results. The results of the section will be used in Sects. 4 and 5 to study the convergence and iteration complexity properties of the relaxed PR splitting method (2). It contains two subsections. The first one describes a class of distance generating functions introduced in [17]. The second one describes the NE-HPE framework and its corresponding convergence and iteration complexity results.

For the sake of shortness, all the results in this section are stated without proofs which, in turn, can be found in Section 3 of the technical report version of this paper (see [24]).

3.1 A class of distance generating functions

We start by introducing a class of distance generating functions (and its corresponding Bregman distances) which is needed for the presentation of the NE-HPE framework in Sect. 3.2.

Definition 3.1 For a given convex set $Z \subset \mathcal{Z}$, a seminorm $\| \cdot \|$ in $\mathcal{Z}$ and scalars $0 < m \leq M$, we let $\mathcal{D}_Z(m, M)$ denote the class of real-valued functions w which are differentiable on Z and satisfy

$$w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \geq \frac{m}{2} \|z - z'\|^2 \quad \forall z, z' \in Z, \quad (5)$$

$$\|\nabla w(z) - \nabla w(z')\|_* \leq M \|z - z'\| \quad \forall z, z' \in Z. \quad (6)$$

A function $w \in \mathcal{D}_Z(m, M)$ is referred to as a distance generating function with respect to the seminorm $\|\cdot\|$ and its associated Bregman distance $dw : Z \times Z \to \mathbb{R}$ is defined as

$$(dw)(z'; z) = (dw)_z(z') := w(z') - w(z) - \langle \nabla w(z), z' - z \rangle \quad \forall z, z' \in Z. \quad (7)$$

Throughout our presentation, we use the second notation $(dw)_z(z')$ instead of the first one $(dw)(z'; z)$ although the latter one makes it clear that (dw) is a function of two arguments, namely, z and z'. Clearly, it follows from (5) that w is a convex function on Z which is in fact m-strongly convex on Z whenever $\|\cdot\|$ is a norm.

Note that if the seminorm in Definition 3.1 is a norm, then (5) implies that w is strongly convex on Z, in which case the corresponding dw is said to be nondegenerate on Z. However, since Definition 3.1 does not necessarily assume that $\|\cdot\|$ is a norm, it admits the possibility of w being not strongly convex on Z, or equivalently, dw being degenerate on Z.

Finally, some useful relations about the above class of Bregman distances can be found in Sect. 3.1 of the technical report version of this paper (see Lemmas 3.2 and 3.3 of [24]).

3.2 The NE-HPE framework

This subsection describes the NE-HPE framework and its corresponding convergence and iteration complexity results.

Throughout this subsection, we assume that scalars $0 < m \leq M$, convex set $Z \subset \mathcal{Z}$, seminorm $\|\cdot\|$ and distance generating function $w \in \mathcal{D}_Z(m, M)$ with respect to $\|\cdot\|$ are given. Our problem of interest in this section is the MIP

$$0 \in T(z) \quad (8)$$

where $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is a maximal monotone operator satisfying the following conditions:

(A0) Dom $(T) \subset Z$;
(A1) the solution set $T^{-1}(0)$ of (8) is nonempty.

We now state a non-Euclidean HPE (NE-HPE) framework for solving the MIP (8) which generalizes its Euclidean counterparts studied in the literature (see for example in [25,27,30]).

Framework 1 (An NE-HPE framework for solving (8)).

(0) Let $z_0 \in Z$ and $\sigma \in [0, 1]$ be given, and set $k = 1$;

(1) choose $\lambda_k > 0$ and find $(\tilde{z}_k, z_k, \varepsilon_k) \in Z \times Z \times \mathbb{R}_+$ such that

$$r_k := \frac{1}{\lambda_k} \nabla (dw)_{z_k} (z_{k-1}) \in T^{[\varepsilon_k]}(\tilde{z}_k), \tag{9}$$

$$(dw)_{z_k}(\tilde{z}_k) + \lambda_k \varepsilon_k \leq \sigma (dw)_{z_{k-1}}(\tilde{z}_k); \tag{10}$$

(2) set $k \leftarrow k + 1$ and go to step 1.

end

We now make some remarks about Framework 1. First, it does not specify how to find λ_k and $(\tilde{z}_k, z_k, \varepsilon_k)$ satisfying (9) and (10). The particular scheme for computing λ_k and $(\tilde{z}_k, z_k, \varepsilon_k)$ will depend on the instance of the framework under consideration and the properties of the operator T. Second, if w is strongly convex on Z and $\sigma = 0$, then (10) implies that $\varepsilon_k = 0$ and $z_k = \tilde{z}_k$ for every k, and hence that $r_k \in T(z_k)$ in view of (9). Therefore, the HPE error conditions (9)–(10) can be viewed as a relaxation of an iteration of the exact non-Euclidean proximal point method, namely,

$$0 \in \frac{1}{\lambda_k} \nabla (dw)_{z_{k-1}}(z_k) + T(z_k).$$

We observe that NE-HPE frameworks have already been studied in [17,21] and [31]. The approach presented in this section differs from these three papers as follows. Assuming that Z is an open convex set, w is continuously differentiable on Z and continuous on its closure, [31] studies a special case of the NE-HPE framework in which $\varepsilon_k = 0$ for every k, and presents results on convergence of sequences rather than iteration complexity. Paper [21] deals with distance generating functions w which do not necessarily satisfy conditions (5) and (6), and as consequence, obtains results which are more limited in scope, i.e., only an ergodic convergence rate result is obtained for operators with bounded feasible domains (or, more generally, for the case in which the sequence generated by the HPE framwework is bounded). Paper [17] introduces the class of distance generating functions $\mathcal{D}_Z(m, M)$ but only analyzes the behavior of a HPE framework for solving inclusions whose operators are strongly monotone with respect to a fixed $w \in \mathcal{D}_Z(m, M)$ (see condition A1 in Section 2 of [17]). This section on the other hand assumes that $w \in \mathcal{D}_Z(m, M)$ but it does not assume any strong monotonicity of T with respect to w.

Before presenting the main results about the the NE-HPE framework, namely, Theorems 3.3 and 3.4 establishing its pointwise and ergodic iteration complexities, respectively, and Propositions 3.5 and 3.6 showing that $\{z_k\}$ and/or $\{\tilde{z}_k\}$ approach $T^{-1}(0)$ in terms of the Bregman distance (dw), we have the following result.

Proposition 3.2 *For every $k \geq 1$ and $z^* \in T^{-1}(0)$, we have*

$$(dw)_{z_{k-1}}(z^*) - (dw)_{z_k}(z^*) - (1-\sigma)(dw)_{z_{k-1}}(\tilde{z}_k) \geq \lambda_k \left[\langle r_k, \tilde{z}_k - z^* \rangle + \varepsilon_k \right] \geq 0. \tag{11}$$

As a consequence, the following statements hold:

(a) $\{(dw)_{z_k}(z^*)\}$ *is non-increasing;*
(b) $\lim_{k \to \infty} \lambda_k \left[\langle r_k, \tilde{z}_k - z^* \rangle + \varepsilon_k \right] = 0$;
(c) $(1 - \sigma) \sum_{i=1}^{k} (dw)_{z_{i-1}}(\tilde{z}_i) \leq (dw)_{z_0}(z^*)$.

Proof See the proof of Proposition 3.5 of [24]. □

For the purpose of stating the convergence rate results below, define

$$(dw)_0 := \inf\{(dw)_{z_0}(z^*) : z^* \in T^{-1}(0)\}. \tag{12}$$

The following pointwise convergence rate result describes the convergence rate of the sequence $\{(r_k, \varepsilon_k)\}$ of residual pairs associated to the sequence $\{\tilde{z}_k\}$. Note that its convergence rate bounds are derived on the best residual pair among (r_i, ε_i) for $i = 1, \ldots, k$ rather than on the last residual pair (r_k, ε_k).

Theorem 3.3 (Pointwise convergence) *Let $(dw)_0$ be as in* (12) *and assume that $\sigma < 1$. Then, the following statements hold:*

(a) *if $\underline{\lambda} := \inf \lambda_k > 0$, then for every $k \in \mathbb{N}$ there exists $i \leq k$ such that*

$$\|r_i\|_* \leq M(1 + \sqrt{\sigma}) \sqrt{\frac{2(dw)_0}{(1 - \sigma)m} \left(\frac{\underline{\lambda}^{-1}}{\sum_{j=1}^{k} \lambda_j} \right)} \leq \frac{M(1 + \sqrt{\sigma})}{\underline{\lambda} \sqrt{k}} \sqrt{\frac{2(dw)_0}{(1 - \sigma)m}},$$

$$\varepsilon_i \leq \frac{\sigma (dw)_0}{1 - \sigma} \frac{1}{\sum_{i=1}^{k} \lambda_i} \leq \frac{\sigma (dw)_0}{(1 - \sigma)\underline{\lambda} k};$$

(b) *for every $k \in \mathbb{N}$, there exists an index $i \leq k$ such that*

$$\|r_i\|_* \leq M(1 + \sqrt{\sigma}) \sqrt{\frac{2(dw)_0}{(1 - \sigma)m} \left(\frac{1}{\sum_{j=1}^{k} \lambda_j^2} \right)}, \qquad \varepsilon_i \leq \frac{\sigma (dw)_0 \lambda_i}{(1 - \sigma) \sum_{j=1}^{k} \lambda_j^2}. \tag{13}$$

Proof See the proof of Theorem 3.8 of [24]. □

From now on, we focus on the ergodic convergence rate of the NE-HPE framework. For $k \geq 1$, define $\Lambda_k := \sum_{i=1}^{k} \lambda_i$ and the ergodic sequences

$$\tilde{z}_k^a = \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i \tilde{z}_i, \quad r_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i r_i, \quad \varepsilon_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i \left(\varepsilon_i + \langle r_i, \tilde{z}_i - \tilde{z}_k^a \rangle \right). \tag{14}$$

The following ergodic convergence result describes the association between the ergodic iterate $\tilde{z}_k^a$ and the residual pair (r_k^a, ε_k^a), and gives a convergence rate bound on the latter residual pair.

Theorem 3.4 (Ergodic convergence) *Let $(dw)_0$ be as in* ([12](#)). *Then, for every $k \geq 1$, we have*

$$\varepsilon_k^a \geq 0, \quad r_k^a \in T^{[\varepsilon_k^a]}(\tilde{z}_k^a)$$

and

$$\|r_k^a\|_* \leq \frac{2M}{\Lambda_k} \sqrt{\frac{2(dw)_0}{m}}, \quad \varepsilon_k^a \leq \left(\frac{3M}{m}\right) \frac{2(dw)_0 + \rho_k}{\Lambda_k}$$

where

$$\rho_k := \max_{i=1,\ldots,k} (dw)_{z_i}(\tilde{z}_i). \tag{15}$$

Moreover, the sequence $\{\rho_k\}$ is bounded under either one of the following situations:

(a) $\sigma < 1$, in which case

$$\rho_k \leq \frac{\sigma(dw)_0}{1-\sigma}; \tag{16}$$

(b) Dom T is bounded, in which case

$$\rho_k \leq \frac{2M}{m}[(dw)_0 + D] \tag{17}$$

where $D := \sup\{\min\{(dw)_y(y'), (dw)_{y'}(y)\} : y, y' \in \text{Dom } T\}$ is the diameter of $\text{Dom } T$ with respect to dw.

Proof See the proof of Theorem 3.9 of [24]. □

In the remaining part of this subsection, we state some results about the sequence generated by an instance of the NE-HPE framework. We assume from now on that such instance generates an infinite sequence of iterates, i.e., the instance does not terminate in a finite number of steps and no termination criterion is checked. Since we are not assuming that the distance generating function w is nondegenerate on Z, it is not possible to establish convergence of the sequence $\{z_k\}$ generated by the NE-HPE framework to a solution of ([8](#)). However, under some mild assumptions, it is possible to establish that $\{z_k\}$ approaches a point $\tilde{z} \in T^{-1}(0)$ if the proximity measure used is the actual Bregman distance.

Proposition 3.5 *Assume that for some infinite index set $\mathcal{K}$ and some $\tilde{z} \in \mathcal{Z}$, we have*

$$\lim_{k \to \mathcal{K}} (r_k, \varepsilon_k) = (0, 0), \quad \lim_{k \to \mathcal{K}} \tilde{z}_k = \tilde{z}. \tag{18}$$

Then, $\tilde{z} \in T^{-1}(0) \subset Z$. If, in addition, $\lim_{k \in \mathcal{K}} (dw)_{z_k}(\tilde{z}_k) = 0$, then $\lim_{k \to \infty} (dw)_{z_k}(\tilde{z}) = 0$.

Proof See the proof of Proposition 3.10 of [24]. □

Proposition 3.6 *Assume that $\sigma < 1$, $\sum_{i=1}^{\infty} \lambda_k^2 = \infty$ and $\{\tilde{z}_k\}$ is bounded. Then, there exists $\tilde{z} \in T^{-1}(0) \subset Z$ such that*

$$\lim_{k \to \infty} (dw)_{z_k}(\tilde{z}) = \lim_{k \to \infty} (dw)_{\tilde{z}_k}(\tilde{z}) = 0. \tag{19}$$

Proof See the proof of Proposition 3.11 of [24] □

Clearly, if w is a nondegenerate distance generating function, then the results above give sufficient conditions for the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ to converge to some $\tilde{z} \in T^{-1}(0)$.

4 The relaxed Peaceman–Rachford splitting method

This section derives convergence rate bounds for the relaxed Peaceman–Rachford (PR) splitting method for solving the monotone inclusion (1) under the assumption that A and B are maximal β-strongly monotone operators for any $\beta \geq 0$. More specifically, its pointwise iteration-complexity is obtained in Theorem 4.6 and its ergodic iteration-complexity is derived in Theorem 4.8. These results are obtained as by-products of the corresponding ones (i.e, Theorems 3.3 and 3.4) in Sect. 3.2 and the fact that the relaxed Peaceman–Rachford (PR) splitting method can be viewed as a special instance of the NE-HPE framework.

Throughout this section, we assume that $\mathcal{X}$ a finite-dimensional real vector space with inner product and associated inner product norm denoted by $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ and $\| \cdot \|_{\mathcal{X}}$, respectively. For a given $\beta \geq 0$, an operator $T : \mathcal{X} \rightrightarrows \mathcal{X}$ is said to be β-strongly monotone if

$$\langle w - w', x - x' \rangle_{\mathcal{X}} \geq \beta \|x - x'\|_{\mathcal{X}}^2 \quad \forall (x, w), (x', w') \in \mathrm{Gr}(T).$$

In what follows, we refer to monotone operators as 0-strongly monotone operators. This terminology has the benefit of allowing us to treat both the monotone and strongly monotone case simultaneously.

Throughout this section, we consider the monotone inclusion (1) where $A, B : \mathcal{X} \rightrightarrows \mathcal{X}$ satisfy the following assumptions:

(B0) for some $\beta \geq 0$, A and B are maximal β-strongly monotone operators;
(B1) the solution set $(A + B)^{-1}(0)$ is non-empty.

We start by observing that (1) is equivalent to solving the following augmented system of inclusions/equation

$$0 \in \gamma A(u) + u - x,$$
$$0 \in \gamma B(v) + x - v,$$
$$0 = u - v$$

where $\gamma > 0$ is an arbitrary scalar. Another way of writing the above system is as

$$0 \in \gamma A(u) + u - x,$$

$$0 \in \gamma B(v) + v + x - 2u,$$
$$0 = u - v.$$

Note that the first and second inclusions are equivalent to

$$u = u(x) := J_{\gamma A}(x), \quad v = v(x) := J_{\gamma B}(2u - x) = J_{\gamma B}(2J_{\gamma A}(x) - x) \quad (20)$$

so that the third equation reduces to

$$0 = u(x) - v(x) = J_{\gamma A}(x) - J_{\gamma B}(2J_{\gamma A}(x) - x).$$

The Douglas–Rachford (DR) splitting method is the iterative procedure $x_k = x_{k-1} + v(x_{k-1}) - u(x_{k-1}), k \geq 1$, started from some $x_0 \in \mathcal{X}$. It is known that the DR splitting method is an exact proximal point method for some maximal monotone operator [13,14]. Hence, convergence of its sequence of iterates is guaranteed.

This section is concerned with a natural generalization of the DR splitting method, namely, the relaxed Peaceman–Rachford (PR) splitting method with relaxation parameter $\theta > 0$, which iterates as

$$(u_k, v_k) := (u(x_{k-1}), v(x_{k-1})) \quad x_k = x_k^{\theta} := x_{k-1} + \theta(v_k - u_k) \quad \forall k \geq 1. \quad (21)$$

We now make a few remarks about the above method. First, it reduces to the DR splitting method when $\theta = 1$, and to the PR splitting method when $\theta = 2$. Second, it reduces to (2) when $\gamma = 1$ but it is not more general than (2) since (21) is equivalent to (2) with $(A, B) = (\gamma A, \gamma B)$. Third, as presented in (21), it can be viewed as an iterative process in the (u, v, x)-space rather than only in the x-space as suggested by (2).

Our analysis of the relaxed DR splitting method is based on further exploring the last remark above, i.e., viewing it as an iterative method in the (u, v, x)-space. We start by introducing an inclusion which plays an important role in our analysis. For a fixed $\tilde{\theta} > 0$ and $\gamma > 0$, consider the inclusion

$$0 \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z) \quad (22)$$

where $\mathcal{L}_{\tilde{\theta}} : \mathcal{X} \times \mathcal{X} \times \mathcal{X} \to \mathcal{X} \times \mathcal{X} \times \mathcal{X}$ is the linear map defined as

$$\mathcal{L}_{\tilde{\theta}}(z) = \mathcal{L}_{\tilde{\theta}}(u, v, x) := \begin{bmatrix} (1 - \tilde{\theta})I & \tilde{\theta}I & -I \\ (\tilde{\theta} - 2)I & (1 - \tilde{\theta})I & I \\ I & -I & 0 \end{bmatrix} \begin{pmatrix} u \\ v \\ x \end{pmatrix} \quad (23)$$

and $C : \mathcal{X} \times \mathcal{X} \times \mathcal{X} \rightrightarrows \mathcal{X} \times \mathcal{X} \times \mathcal{X}$ is the maximal monotone operator defined as

$$C(z) = C(u, v, x) := A(u) \times B(v) \times \{0\}. \quad (24)$$

It is easy to verify that the inclusion (22) is equivalent to the two systems of inclusions/equation following conditions B0 and B1. Hence, it suffices to solve (22) in order

to solve (1). The following simple but useful result explicitly show the relationship between the solution sets of (22) and (1).

Lemma 4.1 *For any $\tilde{\theta} > 0$, the solution set $(\mathcal{L}_{\tilde{\theta}} + \gamma C)^{-1}(0)$ is given by*

$$
\begin{aligned}
(\mathcal{L}_{\tilde{\theta}} + \gamma C)^{-1}(0) &= \{(u^*, u^*, x^*) : \gamma^{-1}(x^* - u^*) \in A(u^*) \cap (-B(u^*))\} \\
&= \{(u^*, u^*, u^* + \gamma a^*) : a^* \in A(u^*), -a^* \in B(u^*)\}.
\end{aligned}
$$

As a consequence, if $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)^{-1}(0)$, then $u^* \in (A + B)^{-1}(0)$ and $u^* = J_{\gamma A}(x^*)$.

Proof The conclusion of the lemma follows immediately from the definitions of $\mathcal{L}_{\tilde{\theta}}$ and C in (23) and (24), respectively, and some simple algebraic manipulations. □

The key idea of our analysis is to show that the relaxed PR splitting method is actually a special instance of the NE-HPE framework for solving inclusion (22) and then use the results discussed in Sect. 3.2 to derive convergence and iteration-complexity results for it. With this goal in mind, the next result gives a sufficient condition for (22) to be a maximal monotone inclusion.

Proposition 4.2 *Assume that $A, B : \mathcal{X} \rightrightarrows \mathcal{X}$ satisfy B0 and let $\tilde{\theta} > 0$ be given. Then,*

(a) for every $z = (u, v, x) \in \mathcal{X} \times \mathcal{X} \times \mathcal{X}$, $z' = (u', v', x') \in \mathcal{X} \times \mathcal{X} \times \mathcal{X}$, $r \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z)$ and $r' \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z')$, we have

$$
\langle \mathcal{L}_{\tilde{\theta}}(z - z'), z - z' \rangle = (1 - \tilde{\theta}) \| (u - u') - (v - v') \|_{\mathcal{X}}^2 \tag{25}
$$

$$
\langle r - r', z - z' \rangle \geq (1 - \tilde{\theta}) \| (u - u') - (v - v') \|_{\mathcal{X}}^2
$$
$$
+ \gamma \beta (\| u - u' \|_{\mathcal{X}}^2 + \| v - v' \|_{\mathcal{X}}^2); \tag{26}
$$

(b) $\mathcal{L}_{\tilde{\theta}} + \gamma C$ is maximal monotone whenever $\tilde{\theta} \in (0, \theta_0]$ where

$$
\theta_0 := 1 + \frac{\gamma \beta}{2}. \tag{27}
$$

Proof (a) Identity (25) follows from the definition of $\mathcal{L}_{\tilde{\theta}}$ in (23). To show inequality (26), assume that $r \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z)$ and $r' \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z')$. Then, $r = \mathcal{L}_{\tilde{\theta}}(z) + \gamma c$ and $r' = \mathcal{L}_{\tilde{\theta}}(z) + \gamma c'$ for some $c \in C(z)$ and $c' \in C(z')$. Using the definition of C and assumption B0, we easily see that

$$
\langle c' - c, z' - z \rangle \geq \beta (\| u - u' \|_{\mathcal{X}}^2 + \| v - v' \|_{\mathcal{X}}^2),
$$

which together with (25), and the fact that $r = \mathcal{L}_{\tilde{\theta}}(z) + \gamma c$ and $r' = \mathcal{L}_{\tilde{\theta}}(z) + \gamma c'$, imply (26).
(b) Monotonicity of $\mathcal{L}_{\tilde{\theta}} + \gamma C$ is due to the fact that the right hand side of (26) is nonnegative for every $(u, v), (u', v') \in \mathcal{X} \times \mathcal{X}$ whenever $\tilde{\theta} \in (0, \theta_0]$. To show $\mathcal{L}_{\tilde{\theta}} + \gamma C$ is maximal monotone, write $\mathcal{L}_{\tilde{\theta}} + \gamma C = (\mathcal{L}_{\tilde{\theta}} + \gamma \bar{C}) + \gamma (C - \bar{C})$

where $\bar{C} := \beta(I, I, 0)$. As a consequence of (a) with $(A, B) = \beta(I, I)$ and the definition of C, we conclude that $\mathcal{L}_{\tilde{\theta}} + \gamma \bar{C}$ is a monotone linear operator for every $\tilde{\theta} \in (0, \theta_0]$. Moreover, Assumption B0 easily implies that $\gamma(C - \bar{C})$ is maximal monotone. The statement now follows by noting that the sum of a monotone linear map and a maximal monotone operator is a maximal monotone operator [1,28]. $\qquad \square$

Note that θ_0 in (27) depends on γ and β and that $\theta_0 = 1$ when $\beta = 0$.

The following technical result states some useful identities and inclusions needed to analyze the the sequence generated by the relaxed PR splitting method.

Lemma 4.3 *For a given $x_{k-1} \in \mathcal{X}$ and $\tilde{\theta} > 0$, define*

$$\tilde{x}_k = x_k^{\tilde{\theta}} := x_{k-1} + \tilde{\theta}(v_k - u_k), \quad \tilde{z}_k = z_k^{\tilde{\theta}} := (u_k, v_k, \tilde{x}_k) \tag{28}$$

where u_k, v_k are as in (21), and set

$$a_k := \frac{1}{\gamma}(x_{k-1} - u_k), \quad b_k := \frac{1}{\gamma}(2u_k - v_k - x_{k-1}). \tag{29}$$

Then, we have:

$$-(1 - \tilde{\theta})u_k - \tilde{\theta}v_k + \tilde{x}_k = \gamma a_k \in \gamma A(u_k), \tag{30}$$

$$(2 - \tilde{\theta})u_k - (1 - \tilde{\theta})v_k - \tilde{x}_k = \gamma b_k \in \gamma B(v_k). \tag{31}$$

As a consequence, we have

$$u_k - v_k = \gamma(a_k + b_k) \in \gamma A(u_k) + \gamma B(v_k) \tag{32}$$

$$(0, 0, u_k - v_k) = \mathcal{L}_{\tilde{\theta}}(\tilde{z}_k) + \gamma c_k \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(\tilde{z}_k) \tag{33}$$

where

$$c_k := (a_k, b_k, 0). \tag{34}$$

Proof Using the definition of $(u(\cdot), v(\cdot))$ in (20), the definition of $(u_k, v_k, x_k^{\tilde{\theta}})$ in (21), and the definitions of a_k and b_k in (29), we easily see that (30) and (31) hold. The equality and the inclusion in (32) follow by adding (30) and (31). Clearly, (33) follows as an immediate consequence of (30) and (31), definitions (23) and (24), and the definition of c_k. $\qquad \square$

The following result shows that the relaxed PR splitting method with $\theta \in (0, 2\theta_0]$ can be viewed as an inexact instance of the NE-HPE framework for solving (22) where from now on we assume that

$$\tilde{\theta} := \min\{\theta, \theta_0\}. \tag{35}$$

Proposition 4.4 *Consider the (degenerate) distance generating function given by*

$$w(z) = w(u, v, x) = \frac{\|x\|_{\mathcal{X}}^2}{2\theta} \quad \forall z = (u, v, x) \in \mathcal{X} \times \mathcal{X} \times \mathcal{X} \tag{36}$$

and the sequence $\{z_k = (u_k, v_k, x_k)\}$ generated according to the relaxed PR splitting method (21) *with any $\theta > 0$. Also, define the sequences $\{\varepsilon_k\}$, $\{\lambda_k\}$ and $\{r_k\}$ as*

$$\varepsilon_k := 0, \quad \lambda_k := 1, \quad r_k := \nabla(dw)_{z_k}(z_{k-1}) \quad \forall k \geq 1, \tag{37}$$

and the sequence $\{\tilde{z}_k = (u_k, v_k, \tilde{x}_k)\}$ as in (28) *with $\tilde{\theta}$ given by* (35). *Then, for every $k \geq 1$, we have:*

(a) $r_k = (0, 0, x_{k-1} - x_k)/\theta = (0, 0, u_k - v_k) = \gamma(0, 0, a_k + b_k)$;
(b) (λ_k, z_{k-1}) and $(z_k, \tilde{z}_k, \varepsilon_k)$ satisfy (9) *with $T = \mathcal{L}_{\tilde{\theta}} + \gamma C$, i.e., $r_k \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(\tilde{z}_k)$;*
(c) (λ_k, z_{k-1}) and $(z_k, \tilde{z}_k, \varepsilon_k)$ satisfy (10) *with $\sigma = (\theta/\tilde{\theta} - 1)^2$ and w as in* (36).

As a consequence, the relaxed PR splitting method with $\theta \in (0, 2\theta_0)$ (resp., $\theta = 2\theta_0$) is an NE-HPE instance with respect to the monotone inclusion $0 \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z)$ in which $\sigma < 1$ (resp., $\sigma = 1$), $\varepsilon_k = 0$ and $\lambda_k = 1$ for every k.

Proof (a) The first identity in (a) follows from (36) and the definition of r_k in (37). The second and third equalities in (a) are due to the second identity in (21) and relation (32), respectively.
(b) This statement follows from (a) and (33).
(c) Using the second identity in (21), relation (36) and the definition of $\tilde{x}_k$ in (28), we conclude that for any $\theta \in (0, 2\theta_0]$,

$$(dw)_{z_k}(\tilde{z}_k) = \frac{\|\tilde{x}_k - x_k\|_{\mathcal{X}}^2}{2\theta} = \left(\frac{\theta}{\tilde{\theta}} - 1\right)^2 \frac{\|\tilde{x}_k - x_{k-1}\|_{\mathcal{X}}^2}{2\theta}$$
$$= \left(\frac{\theta}{\tilde{\theta}} - 1\right)^2 (dw)_{z_{k-1}}(\tilde{z}_k)$$

and hence that (10) is satisfied with $\sigma = (1 - \theta/\tilde{\theta})^2$.
The last conclusion follows from statements (b) and (c), and Proposition 4.2(b). $\qquad\square$

We now make a remark about the special case of Proposition 4.4 in which $\theta \in (0, \theta_0]$. Indeed, in this case, $\tilde{\theta} = \theta$, and hence $\sigma = 0$ and $\tilde{z}_k = z_k$ for every $k \geq 1$. Thus, the relaxed PR splitting method with $\theta \in (0, \theta_0]$ can be viewed as an exact non-Euclidean proximal point method with distance generating function w as in (36) with respect to the monotone inclusion $0 \in T(z) := (\mathcal{L}_\theta + \gamma C)(z)$. Note also that the latter inclusion depends on θ.

As a consequence of Proposition 4.4, we are now ready to describe the pointwise and ergodic convergence rate for the relaxed PR splitting method. We first endow the space $\mathcal{Z} := \mathcal{X} \times \mathcal{X} \times \mathcal{X}$ with the semi-norm $\|(u, v, x)\| := \|x\|_{\mathcal{X}}$ and hence Proposition 2.1 implies that

$$\|(0, 0, x)\|_* = \|x\|_{\mathcal{X}}. \tag{38}$$

It is also easy to see that the distance generating function w defined in (36) is in $\mathcal{D}_{\mathcal{Z}}(m, M)$ with respect to $\|\cdot\|$ where $M = m = 1/\theta$ (see Definition 3.1).

Our next goal is to state a pointwise convergence rate bound for the relaxed PR splitting method. We start by stating a technical result which is well-known for the case where $\beta = 0$ (see for example Lemma 2.4 of [18]). The proof for the general case, i.e., $\beta \geq 0$, is similar and is given in the "Appendix" for the sake of completeness.

Lemma 4.5 *Assume that $\theta \in (0, 2\theta_0]$. Then, for every $k \geq 2$, we have $\|\Delta x_k\|_{\mathcal{X}} \leq \|\Delta x_{k-1}\|_{\mathcal{X}}$ where $\Delta x_k := x_k - x_{k-1}$.*

We now state the pointwise convergence rate result for the relaxed PR splitting method.

Theorem 4.6 *Consider the sequence $\{z_k = (u_k, v_k, x_k)\}$ generated by the relaxed PR splitting method with $\theta \in (0, 2\theta_0)$. Then, for every $k \geq 1$ and $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)^{-1}(0)$,*

$$a_k + b_k \in A(u_k) + B(v_k), \quad \gamma \|a_k + b_k\|_{\mathcal{X}} = \|u_k - v_k\|_{\mathcal{X}} \leq \frac{\sqrt{2}\|x_0 - x^*\|_{\mathcal{X}}}{\sqrt{k}\sqrt{2\tilde{\theta} - \theta}}.$$

Proof The inclusion and the equality in the theorem follows from (32). Since by Proposition 4.4, the relaxed PR splitting method with $\theta \in (0, 2\theta_0)$ is an NE-HPE instance for solving the monotone inclusion $0 \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z)$ in which $\sigma = (\theta/\tilde{\theta} - 1)^2 < 1$, $\varepsilon_k = 0$ and $\lambda_k = 1$ for all $k \geq 1$, it follows from Lemma 4.5, Theorem 3.3, the fact that $M = m = 1/\theta$, and relation (12) that

$$\|r_k\|_* \leq \frac{\sqrt{2}M}{\sqrt{m}}(1 + \sqrt{\sigma})\sqrt{\frac{(dw)_0}{1 - \sigma}\left(\frac{1}{\sum_{j=1}^{k} \lambda_j^2}\right)} \leq \frac{\sqrt{2}\|x_0 - x^*\|_{\mathcal{X}}}{\sqrt{k}\sqrt{2\tilde{\theta} - \theta}}.$$

The inequality of the theorem then follows by Proposition 4.4(a) and relation (38). □

Our main goal in the remaining part of this section is to derive ergodic convergence rate bounds for the relaxed PR splitting method for any $\theta \in (0, 2\theta_0]$. We start by stating the following variation of the transportation lemma for maximal β-strongly monotone operators.

Proposition 4.7 *Assume that T is a maximal β-strongly monotone operator for some $\beta \geq 0$. Assume also that $t_i \in T(u_i)$ for $i = 1, \ldots, k$, and define*

$$\bar{t}_k = \frac{1}{k}\sum_{i=1}^{k} t_i, \quad \bar{u}_k = \frac{1}{k}\sum_{i=1}^{k} u_i, \quad \varepsilon_k = \frac{1}{k}\sum_{i=1}^{k}\langle t_i - \beta u_i, u_i - \bar{u}_k\rangle_{\mathcal{X}} \quad (39)$$

Then, $\varepsilon_k \geq 0$ and $\bar{t}_k \in T^{[\varepsilon_k]}(\bar{u}_k)$.

Proof The assumption that T is a maximal β-strongly monotone operator implies that $T - \beta I$ is maximal monotone. Hence, it follows from the weak transportation formula (see Theorem 2.3 of [5]) applied to $T - \beta I$ that $\varepsilon_k \geq 0$ and $\bar{t}_k - \beta\bar{u}_k \in (T - \beta I)^{[\varepsilon_k]}(\bar{u}_k)$. The result then follows by observing that $(T - \beta I)^{[\varepsilon_k]}(\bar{u}_k) + \beta\bar{u}_k \subseteq T^{[\varepsilon_k]}(\bar{u}_k)$. □

In order to state the ergodic iteration complexity bound for the relaxed PR splitting method, we introduce the ergodic sequences

$$\bar{u}_k = \frac{1}{k}\sum_{i=1}^{k} u_i, \quad \bar{v}_k = \frac{1}{k}\sum_{i=1}^{k} v_i, \quad \bar{a}_k = \frac{1}{k}\sum_{i=1}^{k} a_i, \quad \bar{b}_k = \frac{1}{k}\sum_{i=1}^{k} b_i \tag{40}$$

and the scalar sequences

$$\varepsilon_k' := \frac{1}{k}\sum_{i=1}^{k} \langle a_i - \beta u_i, u_i - \bar{u}_k\rangle_{\mathcal{X}}, \quad \varepsilon_k'' := \frac{1}{k}\sum_{i=1}^{k} \langle b_i - \beta v_i, v_i - \bar{v}_k\rangle_{\mathcal{X}}. \tag{41}$$

Theorem 4.8 *Assume that $\theta \in (0, 2\theta_0]$ and consider the ergodic sequences above. Then, for every $k \geq 1$ and $z^* = (u^*, u^*, x^*) \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)^{-1}(0)$,*

$$\bar{a}_k \in A^{[\varepsilon_k']}(\bar{u}_k), \quad \bar{b}_k \in B^{[\varepsilon_k'']}(\bar{v}_k),$$

$$\gamma \left\|\bar{a}_k + \bar{b}_k\right\|_{\mathcal{X}} = \|\bar{u}_k - \bar{v}_k\|_{\mathcal{X}} \leq \frac{2\|x_0 - x^*\|_{\mathcal{X}}}{k\theta},$$

$$\varepsilon_k' + \varepsilon_k'' \leq \frac{3(1 + 2(1 - \tilde{\theta}/\theta)^2)\|x_0 - x^*\|_{\mathcal{X}}^2}{k\gamma\theta}.$$

Proof The first two inclusions follow from the two inclusions in (30) and (31), relation (40), Assumption B0 and Proposition 4.7. We will now derive the equality and the two inequalities of the theorem using the fact that the relaxed PR splitting method with $\theta \in (0, 2\theta_0]$ is an instance of the NE-HPE method. Letting $\lambda_k = 1$, $\varepsilon_k = 0$ for every k and $\tilde{z}_k^a$, r_k^a and ε_k^a be as in (14), we easily see from Proposition 4.4(a) and (40) that

$$r_k^a = (0, 0, \bar{u}_k - \bar{v}_k) = \gamma(0, 0, \bar{a}_k + \bar{b}_k), \tag{42}$$

$$\varepsilon_k^a = \frac{1}{k}\sum_{i=1}^{k} \langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle. \tag{43}$$

We claim that

$$\varepsilon_k^a \geq \gamma(\varepsilon_k' + \varepsilon_k''). \tag{44}$$

Before proving this claim, we will use it to complete the proof of the theorem. Indeed, using the definition of w in (36), relations (12), (38), (42) and (44), the conclusion of Proposition 4.4, and Theorem 3.4 with $T = \mathcal{L}_{\tilde{\theta}} + \gamma C$, $M = m = 1/\theta$ and $\lambda_k = 1$ for all k, we conclude that

$$\gamma \left\|\bar{a}_k + \bar{b}_k\right\|_{\mathcal{X}} = \|\bar{u}_k - \bar{v}_k\|_{\mathcal{X}} = \|r_k^a\|_* \leq \frac{2\sqrt{2}(dw)_0^{1/2}}{\sqrt{\theta}\Lambda_k} \leq \frac{2\|x_0 - x^*\|_{\mathcal{X}}}{k\theta}$$

and

$$\gamma(\varepsilon_k' + \varepsilon_k'') \leq \varepsilon_k^a \leq \frac{3(2(dw)_0 + \rho_k)}{\Lambda_k} \leq \frac{3\left(\|x_0 - x^*\|_{\mathcal{X}}^2 + \rho_k\right)}{k\theta} \tag{45}$$

where ρ_k is defined in (15). Moreover, using (15), the definition of w in (36), the definition of x_i and $\tilde{x}_i$ in (21) and (28), respectively, the triangle inequality, and Proposition 3.2(a), we conclude that

$$
\begin{aligned}
\rho_k &:= \max_{i=1,\ldots,k} (dw)_{z_i}(\tilde{z}_i) = \max_{i=1,\ldots,k} \frac{\|x_i - \tilde{x}_i\|_{\mathcal{X}}^2}{2\theta} \\
&= (1 - \tilde{\theta}/\theta)^2 \max_{i=1,\ldots,k} \frac{\|x_i - x_{i-1}\|_{\mathcal{X}}^2}{2\theta} \leq \frac{2(1 - \tilde{\theta}/\theta)^2 \|x_0 - x^*\|_{\mathcal{X}}^2}{\theta}.
\end{aligned}
$$

The inequalities of the theorem now follows from the above three relations.

In the remaining part of the proof, we establish our previous claim (44). By Proposition 4.4(a) and relations (33) and (43), we have

$$
k\varepsilon_k^a = \sum_{i=1}^{k} \langle r_i, \tilde{z}_i - \tilde{z}_k^a \rangle = \sum_{i=1}^{k} \langle \mathcal{L}_{\tilde{\theta}}(\tilde{z}_i) + \gamma c_i, \tilde{z}_i - \tilde{z}_k^a \rangle \tag{46}
$$

where c_i is defined in (34). Moreover, we have

$$
\begin{aligned}
\sum_{i=1}^{k} \langle \mathcal{L}_{\tilde{\theta}}(\tilde{z}_i) + \gamma c_i, \tilde{z}_i - \tilde{z}_k^a \rangle &= \sum_{i=1}^{k} \langle \mathcal{L}_{\tilde{\theta}}(\tilde{z}_i - \tilde{z}_k^a), \tilde{z}_i - \tilde{z}_k^a \rangle + \gamma \sum_{i=1}^{k} \langle c_i, \tilde{z}_i - \tilde{z}_k^a \rangle \\
&= (1 - \tilde{\theta}) \sum_{i=1}^{k} \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \\
&\quad + \gamma \sum_{i=1}^{k} \langle c_i, \tilde{z}_i - \tilde{z}_k^a \rangle \\
&\geq -\frac{\gamma \beta}{2} \sum_{i=1}^{k} \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \\
&\quad + \gamma \sum_{i=1}^{k} \langle c_i, \tilde{z}_i - \tilde{z}_k^a \rangle, \tag{47}
\end{aligned}
$$

where the second equality follows from (25) and the definitions of $\tilde{z}_k^a$ in (14), $\tilde{z}_k$ in (28), and $\bar{u}_k$ and $\bar{v}_k$ in (40), and the inequality follows from (27) and the fact that $\tilde{\theta} \leq \theta_0$ in view of (35). Finally, using the definitions of $\tilde{z}_k^a$ in (14), and $\tilde{z}_i$ and c_i in Lemma 4.3, and the straightforward relation

$$
-\frac{1}{2} \sum_{i=1}^{k} \|(u_i - \bar{u}_k) - (v_i - \bar{v}_k)\|_{\mathcal{X}}^2 \geq -\sum_{i=1}^{k} \left(\langle u_i, u_i - \bar{u}_k \rangle_{\mathcal{X}} + \langle v_i, v_i - \bar{v}_k \rangle_{\mathcal{X}} \right),
$$

we conclude from (46) and (47) that

$$\varepsilon_k^a \geq \frac{\gamma}{k} \sum_{i=1}^{k} (\langle a_i - \beta u_i, u_i - \bar{u}_k \rangle_{\mathcal{X}} + \langle b_i - \beta v_i, v_i - \bar{v}_k \rangle_{\mathcal{X}}),$$

and hence that the claim holds in view of (41). $\qquad \square$

We now make some remarks about the convergence rate bounds obtained in Theorem 4.8. In view of Lemma 4.1, x^* depends on γ according to

$$x^* = \gamma a^* + u^*, \quad a^* \in A(u^*) \cap -B(u^*).$$

Hence, letting

$$d_0 := \inf\{\|x_0 - u^*\|_{\mathcal{X}} : u^* \in (A + B)^{-1}(0)\},$$
$$S := \sup\{\|a^*\| : a^* \in A(u^*) \cap -B(u^*), \, u^* \in (A + B)^{-1}(0)\},$$

and assuming that $S < \infty$, it is easy to see that Theorem 4.8 and (27) imply that the relaxed PR splitting method with $\theta = 2\theta_0$ satisfies

$$\|\bar{a}_k + \bar{b}_k\|_{\mathcal{X}} \leq \frac{C_1(\gamma)}{\gamma k}, \quad \|\bar{u}_k - \bar{v}_k\|_{\mathcal{X}} \leq \frac{C_1(\gamma)}{k}, \quad \varepsilon_k' + \varepsilon_k'' \leq \frac{C_2(\gamma)}{k}$$

where

$$C_1(\gamma) = C_1(\gamma; \beta, d_0) = \Theta\left(\frac{d_0 + \gamma S}{1 + \beta \gamma}\right),$$
$$C_2(\gamma) = C_2(\gamma; \beta, d_0) = \Theta\left(\frac{(d_0 + \gamma S)^2}{\gamma(1 + \beta \gamma)}\right).$$

When $S/\beta \geq d_0$, then $\gamma = d_0/S$ minimizes both $C_1(\cdot)$ and $C_2(\cdot)$ up to a multiplicative constant, in which case $C_1^* = \Theta(d_0)$, $C_1^*/\gamma = \Theta(S)$ and $C_2^* = \Theta(Sd_0)$ where

$$C_1^* = C_1^*(\beta, d_0) := \inf\{C_1(\gamma) : \gamma > 0\}, \quad C_2^* = C_2^*(\beta, d_0) := \inf\{C_2(\gamma) : \gamma > 0\}.$$

Note that this case includes the case in which $\beta = 0$. On the other hand, when $S/\beta < d_0$, then both C_1 and C_2 are minimized up to a multiplicative constant by any $\gamma \geq d_0/S$, in which case $C_1^* = \Theta(S/\beta)$ and $C_2^* = \Theta(S^2/\beta)$. Clearly, in this case, C_1^*/γ converges to zero as γ tends to infinity.

Indeed, assume first that $S/\beta \geq d_0$. Then, up to some multiplicative constants, we have

$$C_1(\gamma) \geq \frac{d_0 + \gamma S}{1 + \beta \gamma} \geq \frac{d_0 + \gamma S}{1 + S\gamma/d_0} = d_0,$$

$$C_2(\gamma) \geq \frac{(d_0 + \gamma S)^2}{\gamma(1 + \beta\gamma)} \geq \frac{(d_0 + \gamma S)^2}{\gamma(1 + S\gamma/d_0)} = \frac{d_0(d_0 + \gamma S)}{\gamma} = \frac{d_0^2}{\gamma} + Sd_0,$$

and hence that $C_1^* = \Omega(d_0)$ and $C_2^* = \Omega(Sd_0)$. Moreover, if $\gamma = d_0/S$, then the assumption $S/\beta \geq d_0$ implies that $\beta\gamma \leq 1$, and hence that $C_1^* = \Theta(d_0)$ and $C_2^* = \Theta(Sd_0)$.

Assume now that $S/\beta < d_0$. Then, up to multiplicative constants, it is easy to see that

$$C_1(\gamma) \geq \frac{d_0 + \gamma S}{1 + \beta\gamma} \geq \frac{S}{\beta}$$

$$C_2(\gamma) \geq \frac{(d_0 + \gamma S)^2}{\gamma(1 + \beta\gamma)} \geq \frac{(S/\beta + \gamma S)^2}{\gamma(1 + \beta\gamma)} = \frac{S^2}{\gamma\beta^2}(1 + \gamma\beta),$$

and hence that $C_1^* = \Omega(S/\beta)$ and $C_2^* = \Omega(S^2/\beta)$. Moreover, if $\gamma \geq d_0/S$, then it is easy to see that $C_1^* = \Theta(S/\beta)$ and $C_2^* = \Theta(S^2/\beta)$.

Based on the above discussion, the choice $\gamma = d_0/S$ is optimal but has the disadvantage that d_0 is generally difficult to compute. One possibility around this difficulty is to use $\gamma = D_0/S$ where D_0 is an upper bound on d_0.

5 On the convergence of the relaxed PR splitting method

This section discusses some new convergence results about the sequence generated by the relaxed PR splitting method for the case in which $\beta > 0$. It contains two subsections. As observed in the Introduction, [12] already establishes the convergence of the relaxed PR sequence for the case in which $\beta \geq 0$ and $\theta < 2\theta_0$. The first subsection establishes convergence of the relaxed PR sequence for the case in which $\beta > 0$ and $\theta = 2\theta_0$. The second subsection describes an instance showing that the relaxed PR spliting method may diverge when $\beta \geq 0$ and $\theta \geq \min\{2(1 + \gamma\beta), 2 + \gamma\beta + 1/(\gamma\beta)\}$. (Here, we assume that $1/0 = \infty$.) Note that this instance, specialized to the case $\beta = 0$, shows that the sequence $\{z_k = (u_k, v_k, x_k)\}$ generated by the relaxed PR splitting method with $\beta = 0$ may diverge for any $\theta \geq 2$, and hence that the convergence result obtained for any $\theta \in (0, 2)$ in [12] cannot be improved.

5.1 Convergence result about the relaxed PR sequence

It is known that the sequence $\{z_k = (u_k, v_k, x_k)\}$ generated by the relaxed PR splitting method with $\theta \in (0, 2\theta_0)$ and $\beta \geq 0$ converges [12]. The main result of this subsection, namely Theorem 5.2, establishes convergence of this sequence for $\theta = 2\theta_0$ when $\beta > 0$.

We start by giving a lemma which is used in the proof of Theorem 5.2.

Lemma 5.1 *Consider the sequence $\{z_k = (u_k, v_k, x_k)\}$ generated by the relaxed PR splitting method with $\theta \in (0, 2\theta_0]$ and the sequence $\{\tilde{z}_k = (u_k, v_k, \tilde{x}_k)\}$ defined in (28). Then, the sequences $\{z_k\}$ and $\{\tilde{z}_k\}$ are bounded.*

Proof The assumption that $\theta \in (0, 2\theta_0]$ together with the last conclusion of Proposition 4.4 imply that the relaxed PR splitting method is an NE-HPE instance with $\sigma \leq 1$. Hence, for any $z^* \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)^{-1}(0)$, it follows from Proposition 3.2(a) that the sequence $\{(dw)_{z_k}(z^*)\}$ is non-increasing where w is the distance generating function given by (36). Clearly, this observation implies that $\{x_k\}$ is bounded. This conclusion together with (20) and the nonexpansiveness of $J_{\gamma A}$, $J_{\gamma B}$ imply that $\{u_k\}$ and $\{v_k\}$ are also bounded. Finally, $\{\tilde{x}_k\}$ is bounded due to the definition of $\tilde{x}_k$ in (28), and the boundedness of $\{x_k\}$, $\{u_k\}$ and $\{v_k\}$. □

As mentioned at the beginning of this subsection, the convergence of $\{(u_k, v_k)\}$ to some pair (u^*, u^*) where $u^* \in (A + B)^{-1}(0)$ has been established in [12] for the case in which $\beta > 0$ and $\theta < 2\theta_0$. The following result shows that the latter conclusion can also be extended to $\theta = 2\theta_0$.

Theorem 5.2 *In addition to Assumption B1, assume that Assumption B0 holds with $\beta > 0$. Then, the sequence $\{z_k = (u_k, v_k, x_k)\}$ generated by the relaxed PR splitting method with $\theta = 2\theta_0$ converges to some point lying in $(\mathcal{L}_{\theta_0} + \gamma C)^{-1}(0)$.*

Proof We assume that $\theta = 2\theta_0$ and without any loss of generality that $\gamma = 1$. In view of (35), we have $\tilde{\theta} = \theta_0$. Let $z^* \in (\mathcal{L}_{\theta_0} + C)^{-1}(0)$. Then, by Lemma 4.1, we have $z^* = (u^*, u^*, x^*)$ where

$$u^* = (A + B)^{-1}(0), \quad x^* - u^* \in A(u^*), \quad -x^* + u^* \in B(u^*). \tag{48}$$

Since $\tilde{\theta} = \theta_0$, it follows from Proposition 4.4(b) that $r_k \in (\mathcal{L}_{\theta_0} + C)(\tilde{z}_k)$. This together with the fact that $0 \in (\mathcal{L}_{\theta_0} + C)(z^*)$, inequality (26) with $(z, r) = (\tilde{z}_k, r_k)$ and $(z', r') = (z^*, 0)$, and the fact that $\theta_0 = 1 + \beta/2$, then imply that

$$\langle r_k, \tilde{z}_k - z^* \rangle \geq (1 - \theta_0)\|(u_k - u^*) - (v_k - u^*)\|_{\mathcal{X}}^2 + \beta(\|u_k - u^*\|_{\mathcal{X}}^2 + \|v_k - u^*\|_{\mathcal{X}}^2)$$
$$= \frac{\beta}{2}\|u_k + v_k - 2u^*\|_{\mathcal{X}}^2 \geq 0. \tag{49}$$

Since the last conclusion of Proposition 4.4 states that the relaxed PR splitting method with $\theta = 2\theta_0$ is an NE-HPE instance with respect to the monotone inclusion $0 \in (\mathcal{L}_{\tilde{\theta}} + \gamma C)(z)$ in which $\sigma = 1$, $\lambda_k = 1$ and $\varepsilon_k = 0$ for every k, it follows from Proposition 3.2(b), (49) and the assumption that $\beta > 0$ that

$$\lim_{k \to \infty} \langle r_k, \tilde{z}_k - z^* \rangle = \lim_{k \to \infty} \|u_k + v_k - 2u^*\|_{\mathcal{X}} = 0. \tag{50}$$

By Lemma 5.1, $\{z_k\}$, and hence $\{x_k\}$, is bounded. Therefore, there exist an infinite index set $\mathcal{K}$ and $\bar{x} \in \mathcal{X}$ such that $\lim_{k \in \mathcal{K}} x_{k-1} = \bar{x}$, from which we conclude that

$$\lim_{k \in \mathcal{K}} u_k = \bar{u} := J_A(\bar{x}), \quad \lim_{k \in \mathcal{K}} v_k = \bar{v} := J_B(2\bar{u} - \bar{x}) \tag{51}$$

in view of (20), (21) and the continuity of the point-to-point maps J_A and J_B. Clearly, relations (21), (50) and (51), Proposition 4.4(a), the definitions of J_B following (2)

and $\tilde{z}_k$ in (28), and the fact that $\tilde{\theta} = \theta_0$, imply that

$$\bar{u} + \bar{v} = 2u^*, \quad 2\bar{u} - \bar{v} - \bar{x} \in B(\bar{v}), \quad \lim_{k \in \mathcal{K}} z_k = (\bar{u}, \bar{v}, \bar{x} + \theta(\bar{v} - \bar{u})), \tag{52}$$

$$\lim_{k \in \mathcal{K}} r_k = (0, 0, \bar{u} - \bar{v}), \quad \lim_{k \in \mathcal{K}} \tilde{z}_k = \tilde{z} := (\bar{u}, \bar{v}, \bar{x} + \theta_0(\bar{v} - \bar{u})). \tag{53}$$

Clearly, (50) and (53) imply that

$$0 = \lim_{k \in \mathcal{K}} \langle r_k, \tilde{z}_k - z^* \rangle = \langle \bar{u} - \bar{v}, \bar{x} + \theta_0(\bar{v} - \bar{u}) - x^* \rangle_{\mathcal{X}} = -\theta_0 \|\bar{u} - \bar{v}\|_{\mathcal{X}}^2 + \langle \bar{u} - \bar{v}, \bar{x} - x^* \rangle_{\mathcal{X}}. \tag{54}$$

Using the second inclusion in (48), the identity and the inclusion in (52), the β-strong monotonicity of B, and relation (54), we then conclude that

$$\begin{aligned}
\frac{\beta}{4} \|\bar{v} - \bar{u}\|_{\mathcal{X}}^2 = \beta \|\bar{v} - u^*\|_{\mathcal{X}}^2 &\leq \langle (2\bar{u} - \bar{v} - \bar{x}) - (u^* - x^*), \bar{v} - u^* \rangle_{\mathcal{X}} \\
&= \frac{1}{2} \left\langle \frac{3}{2}(\bar{u} - \bar{v}) - \bar{x} + x^*, \bar{v} - \bar{u} \right\rangle_{\mathcal{X}} \\
&= \frac{1}{2} \langle \bar{x} - x^*, \bar{u} - \bar{v} \rangle_{\mathcal{X}} - \frac{3}{4} \|\bar{v} - \bar{u}\|_{\mathcal{X}}^2 = \left(\frac{\theta_0}{2} - \frac{3}{4} \right) \|\bar{v} - \bar{u}\|_{\mathcal{X}}^2.
\end{aligned}$$

The latter inequality together with the fact that $\theta_0 = 1 + (\beta/2)$ then imply that $\bar{u} = \bar{v} = u^*$ where the last equality is due to the identity in (52). We have thus shown that $\{u_k\}_{k \in \mathcal{K}}$ and $\{v_k\}_{k \in \mathcal{K}}$ both converge to $u^* = (A + B)^{-1}(0)$. Since $\bar{u} = \bar{v} = u^*$, it follows from (52) and (53) that

$$\lim_{k \in \mathcal{K}} r_k = 0, \quad \lim_{k \in \mathcal{K}} z_k = \lim_{k \in \mathcal{K}} \tilde{z}_k = \tilde{z} = (u^*, u^*, \bar{x}),$$

$$\lim_{k \in \mathcal{K}} (dw)_{z_k}(\tilde{z}_k) = \lim_{k \in \mathcal{K}} \|x_k - \tilde{x}_k\|_{\mathcal{X}}^2 / (2\theta) = \|\bar{x} - \bar{x}\|_{\mathcal{X}}^2 / (2\theta) = 0.$$

Hence, Proposition 3.5 with $T = (\mathcal{L}_{\theta_0} + \gamma C)$ implies that $\tilde{z} \in (\mathcal{L}_{\theta_0} + \gamma C)^{-1}(0)$ and $0 = \lim_{k \to \infty} (dw)_{z_k}(\tilde{z}) = \lim_{k \to \infty} \|\bar{x} - x_k\|_{\mathcal{X}}^2 / (2\theta)$. We thus conclude that $\{z_k\}$ converges to $(u^*, u^*, \bar{x}) = \tilde{z} \in (\mathcal{L}_{\theta_0} + \gamma C)^{-1}(0)$. $\qquad \square$

Before ending this subsection, we make two remarks. First, for a fixed $\tau > 0$, consider the set

$$R(\tau) := \{(\beta, \theta) \in \mathbb{R}^2 : \beta > 0, \, 0 < \theta \leq 2 + \tau\beta\}.$$

Then, it follows from Theorem 5.2 and the observation in the paragraph preceding it that the sequence generated by the relaxed PR splitting method with relaxation parameter θ to solve (1) with A, B maximal β-strongly monotone converges for any $(\beta, \theta) \in R(1)$. Second, it follows from the example presented in the next subsection that the above conclusion fails if $R(1)$ is enlarged to the region $R(\tau)$ for any $\tau > 1$.

5.2 Non-convergent instances for $\theta \geq \min\{2 + 2\gamma\beta, 2 + \gamma\beta + 1/(\gamma\beta)\}$

By [12] and Theorem 5.2, the sequence $\{x_k\}$ generated by the relaxed PR splitting method converges whenever either $\theta \in (0, 2 + \gamma\beta)$ or $\theta = 2 + \gamma\beta$ and $\beta > 0$. This subsection gives an instance of (1), where A, B are maximal β-strongly monotone, for which the sequence $\{x_k\}$ generated by the relaxed PR splitting method with relaxation parameter θ does not converge when $\beta \geq 0$ and $\theta \geq \min\{2(1+\gamma\beta), 2+\gamma\beta+1/(\gamma\beta)\}$.

Recall from (20) and (21) that the relaxed PR splitting method iterates as

$$x_{k+1} = x_k + \theta(J_{\gamma B}(2J_{\gamma A}(x_k) - x_k) - J_{\gamma A}(x_k)) \tag{55}$$

where $\theta > 0$. Without any loss of generality, we assume that $\gamma = 1$ in (55).

We now describe our instance. First, let $\mathcal{X} := \tilde{\mathcal{X}} \times \tilde{\mathcal{X}}$ where $\tilde{\mathcal{X}}$ is a finite-dimensional real vector space, and let $A_0, B_0 : \mathcal{X} \rightrightarrows \mathcal{X}$ be defined as

$$A_0(\tilde{x}_1, \tilde{x}_2) = (0, 0), \quad B_0(\tilde{x}_1, \tilde{x}_2) = N_{\{0\}}(\tilde{x}_1) \times \{0\}, \quad \forall x = (\tilde{x}_1, \tilde{x}_2) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{X}}$$

where $N_{\{0\}}(\cdot)$ denotes the normal cone operator of the set $\{0\}$. Clearly, A_0 and B_0 are both maximal monotone operators and

$$J_{A_0}(\tilde{x}_1, \tilde{x}_2) = (\tilde{x}_1, \tilde{x}_2), \quad J_{B_0}(\tilde{x}_1, \tilde{x}_2) = (0, \tilde{x}_2), \quad \forall x = (\tilde{x}_1, \tilde{x}_2) \in \tilde{\mathcal{X}} \times \tilde{\mathcal{X}}.$$

Now define $A := A_0 + \beta I$ and $B := B_0 + \bar{\beta} I$ where $\bar{\beta} \geq \beta \geq 0$. It follows that A is a β-strongly maximal monotone operator and B is a $\bar{\beta}$-strongly maximal monotone operator. Hence, the instance we are describing is slightly more general in that A and B have different strong monotonicity parameters.

Moreover, for any $x = (\tilde{x}_1, \tilde{x}_2) \in \mathcal{X}$, it is easy to see that

$$J_A(x) = J_{\frac{1}{1+\beta} A_0}\left(\frac{1}{1+\beta}(\tilde{x}_1, \tilde{x}_2)\right) = \frac{1}{1+\beta}(\tilde{x}_1, \tilde{x}_2),$$

$$J_B(x) = J_{\frac{1}{1+\bar{\beta}} B_0}\left(\frac{1}{1+\bar{\beta}}(\tilde{x}_1, \tilde{x}_2)\right) = \frac{1}{1+\bar{\beta}}(0, \tilde{x}_2).$$

and hence that

$$x + \theta[J_B(2J_A(x) - x) - J_A(x)] = \left(\frac{(1+\beta-\theta)}{1+\beta}\tilde{x}_1, \left[1 - \frac{(\beta+\bar{\beta})\theta}{(1+\beta)(1+\bar{\beta})}\right]\tilde{x}_2\right). \tag{56}$$

From (55) and (56), we easily see that the sequence $\{x_k\}$ generated by the relaxed PR splitting method diverges whenever

$$\frac{(1+\beta-\theta)}{1+\beta} \leq -1, \quad \text{or} \quad 1 - \frac{(\beta+\bar{\beta})\theta}{(1+\beta)(1+\bar{\beta})} \leq -1,$$

or equivalently, whenever

$$\theta \geq \min \left\{ 2(1 + \beta), \, 2 + \frac{2(1 + \beta\bar{\beta})}{\beta + \bar{\beta}} \right\}.$$

Note that when $\beta = \bar{\beta}$, the above inequality reduces to $\theta \geq \min\{2(1+\beta), 2+\beta+1/\beta\}$.

Before ending this subsection, we make two remarks. First, when $\beta = 0$ and hence A is not strongly monotone, the sequence $\{x_k\}$ for the above example diverges for any $\theta \geq 2$ even if B is strongly monotone, i.e., $\bar{\beta} > 0$. Second, the above example specialized to the case in which $\beta = \bar{\beta}$ easily shows that the sequence generated by the relaxed PR splitting method does not necessarily converge for any $(\beta, \theta) \in R(\tau)$ if $\tau > 1$ where $R(\tau)$ is defined at the end of Sect. 5.1.

6 Numerical study

This section illustrates the behavior of the relaxed PR splitting method for solving the weighted Lasso minimization problem [16] (see also [6])

$$\min_{u \in \mathbb{R}^n} f(u) + g(u), \tag{57}$$

where $f(u) := \frac{1}{2}\|Cu - b\|_{\mathcal{X}}^2$ and $g(u) := \|Wu\|_1$ for every $u \in \mathbb{R}^n$. Our numerical experiments consider instances where $n = 200$, $b \in \mathbb{R}^{300}$ and $C \in \mathbb{R}^{300 \times 200}$ is a sparse data matrix with an average of 10 nonzero entries per row. Each component of b and each nonzero element of C is drawn from a Gaussian distribution with zero mean and unit variance, while $W \in \mathbb{R}^{200 \times 200}$ is a diagonal matrix with positive diagonal elements drawn from a uniform distribution on the interval $[0, 1]$. This setup follows that of [16]. Note that $\mathcal{X} = \mathbb{R}^{300}$ and $\|\cdot\|_1$ is the 1-norm on $\mathbb{R}^{200}$. Observe that f is α-strongly convex[1] on $\mathbb{R}^{200}$ where $\alpha = \lambda_{\min}(C^T C)$ is the minimum eigenvalue of $C^T C$. Also, f is differentiable and its gradient is κ-Lipschitz continuous on $\mathbb{R}^{200}$ where $\kappa = \lambda_{\max}(C^T C)$ is the maximum eigenvalue of $C^T C$. The function g is clearly convex on $\mathbb{R}^{200}$.

We consider solving (57) by apply the relaxed PR splitting method (55) to solve the inclusion (1) with

$$A = \partial f - \alpha' I, \quad B = \partial g + \alpha' I, \tag{58}$$

where $0 \leq \alpha' \leq \alpha = \lambda_{\min}(C^T C)$. Since A (resp., B) is $(\alpha - \alpha')$-strongly (resp., α'-strongly) maximal monotone, the results developed in Sects. 4 and 5 for the relaxed PR splitting method with (A, B) as above applies with $\beta = \min\{\alpha - \alpha', \alpha'\}$. Our goal in this section is to gain some intuition of how the relaxed PR splitting method performs as α' (and hence β), γ and θ change. In our numerical experiments, we start the relaxed PR splitting algorithm with $x_0 = 0$ and terminate it when $\|x_{k+1} - x_k\|_{\mathcal{X}} \leq 10^{-5}$. The paragraphs below report the results of three experiments.

---

[1] A function f is α-strongly convex on $\mathcal{X}$ if $f - \frac{\alpha}{2}\|\cdot\|_{\mathcal{X}}^2$ is convex on $\mathcal{X}$.

Table 1 Average number of iterations performed by the relaxed PR splitting method (55) to solve (57) based on the partition (A, B) given by (58) for 4 pairs (γ, α') and 6 different values of θ

θ	Average number of iterations			
	$\gamma = 1$		$\gamma = 1/\sqrt{\alpha\kappa}$	
	$\alpha' = 0$	$\alpha' = \alpha/2$	$\alpha' = 0$	$\alpha' = \alpha/2$
1	141.79	140.64	60.10	60.11
1.25	115.96	115.06	48.47	48.48
1.50	98.31	97.48	40.51	40.49
1.75	85.33	84.64	34.67	34.70
2	264.80	75.08	58.54	42.11
$2 + \gamma\alpha/2$	> 500	73.25	74.73	49.60

In the first experiment, we generate 100 random instances of (C, W, b) and we observed that the condition $\lambda_{\min}(C^T C) > 0$ holds for all instances. The relaxed PR splitting method is used to solve these instances of (57) for various values of θ and with the pair (γ, α') taking on the values $(1, 0)$, $(1, \alpha/2)$, $(1/\sqrt{\alpha\kappa}, 0)$ and $(1/\sqrt{\alpha\kappa}, \alpha/2)$. Note that it follows from Proposition 3 of [16] that when $\alpha' = 0$ and $\theta = 2$, the choice of $\gamma = 1/\sqrt{\alpha\kappa}$ has been shown to be optimal for the relaxed PR splitting method. Our results are shown in Table 1. We see from the table that, except when $\theta = 2$ and $\theta = 2 + \gamma\alpha/2$, the average number of iterations for $\alpha' = 0$ and $\alpha' = \alpha/2$ are similar. However, when $\theta = 2$ and $\theta = 2 + \gamma\alpha/2$, the choice $\alpha' = \alpha/2$ outperforms the one with $\alpha' = 0$. One possible explanation for this behavior is due to the fact that when $\theta = 2$ and $\theta = 2 + \gamma\alpha/2$, the relaxed PR sequence converges when both operators are strongly monotone, while it does not necessary converge when either one of the operators is only monotone. Note also that the results in the last row of the table confirm the convergence result of the relaxed PR splitting method (see Theorem 5.2) for the case in which A and B are β-strongly maximal monotone operators with $\beta > 0$ and $\theta = 2 + \gamma\beta$. Finally, our results (the last two rows of table) suggest that, if A is maximal α-strongly monotone and B is only maximal monotone, it might be advantageous to use the relaxed PR splitting method with $0 < \alpha' < \alpha$ (and hence $\beta > 0$) instead of $\alpha' = 0$ (and hence $\beta = 0$).

In our second experiment, we use relaxed PR splitting method with (θ, γ) equal to $(2, 1)$ and $(2, 1/\sqrt{\alpha\kappa})$, and with α' varying from 0 to α, to solve (57) for a randomly generated (C, W, b). In this instance, $\alpha = \lambda_{\min}(C^T C) = 0.3792$ and $\kappa = \lambda_{\max}(C^T C) = 57.6624$. Our results are shown in Fig. 1. We see from Fig. 1 that the number of iterations decreases as α' increases in both cases. These graphs again suggest that it might be advantageous to have A and B maximal β-strongly monotone with $\beta > 0$. We also observe that as α' approaches α, the number of iterations does not increase even though the operator A is losing its strong monotonicity.

In our third experiment, we performed the same numerical experiments as the ones mentioned above but with $(A, B) = (\partial g + \alpha' I, \partial f - \alpha' I)$ instead of $(A, B) = (\partial f - \alpha' I, \partial g + \alpha' I)$ and note that the results obtained were very similar to the ones
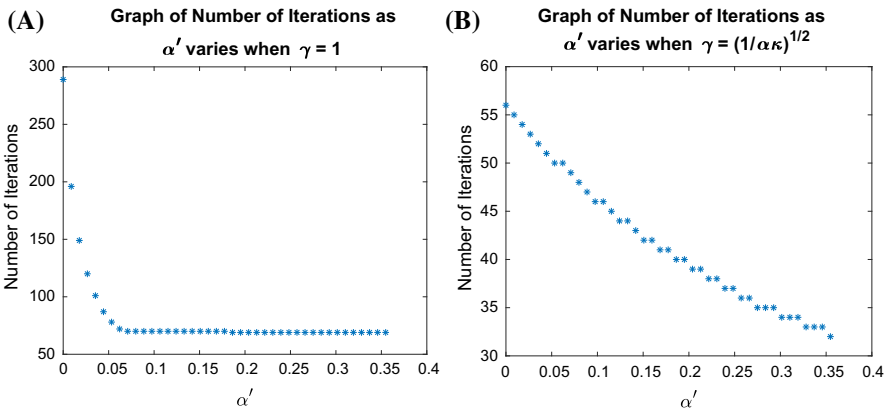
(A)

Graph of Number of Iterations as α' varies when $\gamma = 1$

(B)

Graph of Number of Iterations as α' varies when $\gamma = (1/\alpha\kappa)^{1/2}$

Fig. 1 Two graphs showing how the number of iterations performed by (55) changes with varying α' using the partition (A, B) given (58)

Table 2 Summary of convergence rate results for the relaxed PR splitting method on the sum of two convex functions f, g

Convergence				Additional conditions on f, g besides convexity
θ	Rate	Type	References	
$(0, 2)$	$\mathcal{O}(1/\sqrt{k})$	Pointwise	[18]	–
$(0, 2)$	$o(1/\sqrt{k})$	Pointwise	[10]	–
$(0, 2]$	$\mathcal{O}(1/k)$	Ergodic	[10]	–
$(0, 2)$	$o(1/k)$	Best iterate	[11]	∇f or ∇g Lipschitz
$1, 2$	–	R-linear	[22]	f strongly convex and ∇f Lipschitz
$(0, 2)$	–	R-linear	[15]	g strongly convex and ∇f Lipschitz
$(0, 2]$	–	R-linear	[11]	(f or g strongly convex) and (∇f or ∇g Lipschitz)
$(0, 2]$	–	R-linear	[12]	(f or g strongly convex) and ∇f Lipschitz
$(0, \frac{4}{1+\delta})$ $\delta \in (0, 1)$	–	R-linear	[16] (resp., [15])	f (resp., g) strongly convex and ∇f (resp., ∇g) Lipschitz

reported above. Hence interchanging A and B in the implementation of the relaxed PR splitting method have little impact on its performance.

7 Concluding remarks

This paper establishes convergence of the sequence of iterates and an $\mathcal{O}(1/k)$ ergodic convergence rate bound for the relaxed PR splitting method for any $\theta \in (0, 2 + \gamma\beta]$

Table 3 Summary of convergence rate results for the relaxed PR splitting method on the sum of two β-strongly convex functions f, g, with $\beta > 0$

Convergence				Additional conditions on f, g besides strong convexity
θ	Rate	Type	References	
$(0, 2 + \gamma\beta)$	$\mathcal{O}(1/\sqrt{k})$	Pointwise	This paper	–
$(0, 2 + \gamma\beta]$	$\mathcal{O}(1/k)$	Ergodic	This paper	–
$(0, 2]$	–	R-linear	[12]	∇f Lipschitz

by viewing it as an instance of a non-Euclidean HPE framework. It also establishes an $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate bound for it for any $\theta \in (0, 2 + \gamma\beta)$. Furthermore, an example showing that PR iterates do not necessarily converge for $\beta \geq 0$ and $\theta \geq \min\{2(1 + \gamma\beta), 2 + \gamma\beta + 1/(\gamma\beta)\}$ is given.

Table 2 (resp., Table 3) for the case in which $\beta = 0$ (resp., $\beta > 0$) provides a summary of the convergence rate results known so far for the relaxed PR splitting method when $(A, B) = (\partial f, \partial g)$ for some convex functions f and g. However, we observe that some of these results also hold for pairs (A, B) of maximal monotone operators which are not subdifferentials. The term "R-linear" in the tables below stands for linear convergence of the sequence $\{x_k\}$ generated by the relaxed PR splitting method.

We observe that our analysis in Sects. 4 and 5, in contrast to the ones in [11,15, 16,22], does not impose any regularity condition on A and B such as assuming one of them to be a Lipschitz, and hence point-to-point, operator. Also, if only one of the operators, say A, is assumed to be maximal β-strongly monotone, (1) is equivalent to $0 \in (A' + B')(u)$ where $A' := A - (\beta/2)I$ and $B' := B + (\beta/2)I$ are now both $(\beta/2)$-strongly monotone. Thus, to solve (1), the relaxed PR method with (A, B) replaced by (A', B') can be applied, thereby ensuring convergence of the iterates, as well as pointwise and ergodic convergence rate bounds, for values of $\theta \geq 2$. This idea was tested in our computational results of Sect. 6 where a weighted Lasso minimization problem [6] is solved using the partitions (A, B) and (A', B'). Our conclusion is that the partition (A', B') substantially outperforms the other one for values of $\theta \geq 2$.

Appendix

Proof of Lemma 4.5 To simplify notation, let

$$\Delta x = \Delta x_k, \quad \Delta x^- := \Delta x_{k-1}, \quad \Delta u = u_k - u_{k-1},$$
$$\Delta v = v_k - v_{k-1}, \quad \Delta a = a_k - a_{k-1}, \quad \Delta b = b_k - b_{k-1}.$$

Then, it follows from the second identity in (21) and relation (29) that

$$\Delta x = \Delta x^- + \theta(\Delta v - \Delta u), \quad \gamma \Delta a = \Delta x^- - \Delta u, \quad \gamma \Delta b = 2\Delta u - \Delta v - \Delta x^-. \tag{59}$$

Also, the two inclusions in (30) and (31) together with the β-strong monotonicity of A and B imply that

$$\langle \Delta a, \Delta u \rangle_{\mathcal{X}} \geq \beta \|\Delta u\|_{\mathcal{X}}^2, \quad \langle \Delta b, \Delta v \rangle_{\mathcal{X}} \geq \beta \|\Delta v\|_{\mathcal{X}}^2.$$

Combining the last two identities in (59) with the above inequalities, we obtain

$$\langle \Delta x^- - \Delta u, \Delta u \rangle_{\mathcal{X}} \geq \gamma \beta \|\Delta u\|_{\mathcal{X}}^2, \quad \langle 2\Delta u - \Delta v - \Delta x^-, \Delta v \rangle_{\mathcal{X}} \geq \gamma \beta \|\Delta v\|_{\mathcal{X}}^2.$$

Adding these two last inequalities and simplifying the resulting expression, we obtain

$$\langle \Delta x^-, \Delta u - \Delta v \rangle_{\mathcal{X}} + 2\langle \Delta u, \Delta v \rangle_{\mathcal{X}} \geq (1 + \gamma \beta)[\|\Delta u\|_{\mathcal{X}}^2 + \|\Delta v\|_{\mathcal{X}}^2] \tag{60}$$

From the first equality in (59), we have

$$2\theta \langle \Delta x^-, \Delta u - \Delta v \rangle_{\mathcal{X}} = \|\Delta x^-\|_{\mathcal{X}}^2 - \|\Delta x\|_{\mathcal{X}}^2 + \theta^2 \|\Delta v - \Delta u\|_{\mathcal{X}}^2,$$

which upon substituting into (60), the following is true:

$$\|\Delta x^-\|_{\mathcal{X}}^2 - \|\Delta x\|_{\mathcal{X}}^2 \geq 2\theta \left(\left[1 + \gamma \beta - \frac{\theta}{2}\right] (\|\Delta u\|_{\mathcal{X}}^2 + \|\Delta v\|_{\mathcal{X}}^2) \right.$$
$$\left. + 2\left[\frac{\theta}{2} - 1\right] \langle \Delta u, \Delta v \rangle_{\mathcal{X}} \right).$$

Note that the right-hand side in the above inequality is greater than or equal to zero if $\theta \in (0, 2\theta_0]$. Hence, we have if $\theta \in (0, 2\theta_0]$,

$$\|\Delta x\|_{\mathcal{X}} \leq \|\Delta x^-\|_{\mathcal{X}}.$$

$\square$

References

1. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
2. Bauschke, H.H., Bello Cruz, J.Y., Nghia, T.T.A., Phan, H.M., Wang, X.: The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle. J. Approx. Theory **185**, 63–79 (2014)
3. Bauschke, H.H., Bello Cruz, J.Y., Nghia, T.T.A., Phan, H.M., Wang, X.: Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas–Rachford methods for two subspaces. Numer. Algorithms **73**, 33–76 (2016)
4. Bauschke, H.H., Noll, D., Phan, H.M.: Linear and strong convergence of algorithms involving averaged nonexpansive operators. J. Math. Anal. Appl. **421**, 1–20 (2015)
5. Burachik, R.S., Sagastizábal, C.A., Svaiter, B.F.: ε-enlargements of maximal monotone operators: theory and applications. Reformulation: Nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne 1997), vol. **22** of Applied Optimization, pp. 25–43. Kluwer Academic Publishers, Dordrecht, The Netherlands (1999)
6. Candes, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted l_1 minimization. J. Fourier Anal. Appl. **14**(5–6), 877–905 (2008)
7. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**(5–6), 475–504 (2004)
8. Combettes, P.L.: Iterative construction of the resolvent of a sum of maximal monotone operators. J. Convex Anal. **16**(3), 727–748 (2009)
9. Davis, D.: Convergence rate analysis of the forward-Douglas–Rachford splitting scheme. SIAM J. Optim. **25**, 1760–1786 (2015)
10. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. arXiv preprint arXiv:1406.4834v3 (2015)
11. Davis, D., Yin, W.: Faster convergence rates of relaxed Peaceman–Rachford and ADMM under regularity assumptions. Math. Oper. Res. **42**(3), 783–805 (2017)
12. Dong, Y., Fischer, A.: A family of operator splitting methods revisited. Nonlinear Anal. **72**, 4307–4315 (2010)
13. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
14. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problem, vol. II. Springer, New York (2003)
15. Giselsson, P.: Tight global linear convergence rate bounds for Douglas–Rachford splitting. J. Fixed Point Theory Appl. **19**(4), 2241–2270 (2017)
16. Giselsson, P., Boyd, S.: Linear convergence and metric selection for Douglas–Rachford splitting and ADMM. IEEE Trans. Autom. Control **62**, 532–544 (2017)
17. Goncalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: Improved pointwise iteration-complexity of a regularized ADMM and of a regularized non-Euclidean HPE framework. arXiv preprint arXiv:1601.01140v1 (2016)
18. He, B., Yuan, X.: On the convergence rate of Douglas–Rachford operator splitting method. Math. Program. Ser. A **153**, 715–722 (2015)
19. He, Y., Monteiro, R.D.C.: Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player Nash equilibrium problems. SIAM J. Optim. **25**, 2182–2211 (2015)
20. He, Y., Monteiro, R.D.C.: An accelerated HPE-type algorithm for a class of composite convex–concave saddle-point problems. SIAM J. Optim. **26**, 29–56 (2016)
21. Kolossoski, O., Monteiro, R.D.C.: An accelerated non-Euclidean hybrid proximal extragradient-type algorithm for convex-concave saddle-point problems. Preprint (2015)
22. Lions, P.-L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
23. Monteiro, R.D.C., Sicre, M.R., Svaiter, B.F.: A hybrid proximal extragradient self-concordant primal barrier method for monotone variational inequalities. SIAM J. Optim. **25**(4), 1965–1996 (2015)
24. Monteiro, R.D.C, Sim, C.-K.: Complexity of the relaxed Peaceman–Rachford splitting method for the sum of two maximal strongly monotone operators. http://arxiv.org/abs/1611.03567 (2017). arXiv preprint arXiv:1611.03567v2

25. Monteiro, R.D.C., Svaiter, B.F.: On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. SIAM J. Optim. **20**, 2755–2787 (2010)

26. Monteiro, R.D.C., Svaiter, B.F.: Complexity of variants of Tseng's modified F–B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. SIAM J. Optim. **21**(4), 1688–1720 (2011)

27. Monteiro, R.D.C., Svaiter, B.F.: Complexity of variants of Tseng's modified F–B splitting and Korpelevich's methods of hemi-variational inequalities with applications to saddle point and convex optimization problems. SIAM J. Optim. **21**, 1688–1720 (2011)

28. Rockafellar, R.T.: On the maximality of sums of nonlinear monotone operators. Trans. Am. Math. Soc. **149**, 75–88 (1970)

29. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)

30. Solodov, M.V., Svaiter, B.F.: A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Var. Anal. **7**, 323–345 (1999)

31. Solodov, M.V., Svaiter, B.F.: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. Math. Oper. Res. **25**, 214–230 (2000)