Comput Optim Appl (2018) 70:827–839
https://doi.org/10.1007/s10589-018-9992-3

The generalized proximal point algorithm with step size 2 is not necessarily convergent

Min Tao[1] · **Xiaoming Yuan**[2]

Received: 18 April 2017 / Published online: 3 March 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract The proximal point algorithm (PPA) is a fundamental method in optimization and it has been well studied in the literature. Recently a generalized version of the PPA with a step size in (0, 2) has been proposed. Inheriting all important theoretical properties of the original PPA, the generalized PPA has some numerical advantages that have been well verified in the literature by various applications. A common sense is that larger step sizes are preferred whenever the convergence can be theoretically ensured; thus it is interesting to know whether or not the step size of the generalized PPA can be as large as 2. We give a negative answer to this question. Some counterexamples are constructed to illustrate the divergence of the generalized PPA with step size 2 in both generic and specific settings, including the generalized versions of the very popular augmented Lagrangian method and the alternating direction method of multipliers. A by-product of our analysis is the failure of convergence of the Peaceman–Rachford splitting method and a generalized version of the forward–backward splitting method with step size 1.5.

Min Tao was supported by the Natural Science Foundation of China: NSFC-11301280 and the sponsorship of Jiangsu overseas research and training program for university prominent young and middle-aged teachers and presidents. Xiaoming Yuan was supported by the General Research Fund from Hong Kong Research Grants Council: HKBU 12313516.

✉ Min Tao
taom@nju.edu.cn

Xiaoming Yuan
xmyuan@hku.hk

1 Department of Mathematics, National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

2 The University of Hong Kong, Hong Kong, China

Keywords Proximal point algorithm · Step size · Convergence · Augmented Lagrangian method · Alternating direction method of multipliers · Peaceman–Rachford splitting method · Forward–backward splitting method

Mathematics Subject Classification 90C25 · 90C30

1 Introduction

A basic mathematical problem is finding a zero point, denoted by z^*, of a maximal monotone set-valued mapping T:

$$0 \in T(z), \tag{1}$$

where $T : \mathcal{H} \to 2^{\mathcal{H}}$ and $\mathcal{H}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$. Throughout, the set of T's zero points, denoted by $zer(T)$, is assumed to be nonempty. Given an initial point z^0 in $\mathcal{H}$, the proximal point algorithm (PPA) proposed originally in [22,23] for (1) iteratively generates a sequence $\{z^k\}$ by the scheme

$$0 \in cT(z^{k+1}) + z^{k+1} - z^k, \tag{2}$$

where the proximal parameter c is positive. We refer to, e.g. [30], for the study of convergence of the PPA (2) and the restriction on the proximal parameter c. Let $J_{cT} := (I + cT)^{-1}$ be the resolvent operator of the set-valued mapping T. Then, the PPA (2) can be rewritten as

$$z^{k+1} = J_{cT}(z^k). \tag{3}$$

The PPA (3) has been playing a fundamental role both theoretically and algorithmically in optimization. In particular, when the abstract operator T is specified, the PPA (3) can be specified as a number of fundamental algorithms in different settings such as the augmented Lagrangian method (ALM) in [19,28], alternating direction method of multipliers (ADMM) in [16], Douglas–Rachford splitting method (DRSM) in [9,21], projected gradient method [27], extragradient method in [20], and so on. We refer to, e.g., [18,24,30,33], for some important literatures in which the importance of PPA was well elaborated on.

In [17], the generalized version of PPA

$$z^{k+1} = z^k - \gamma(z^k - J_{cT}(z^k)) \tag{4}$$

was proposed with $\gamma \in (0, 2)$. Both the generic setting (4) and its special settings have been further analyzed in some recent literatures. Note that we can discuss the more general case for (4) with a sequence of dynamical step sizes $\{\gamma_k\}$ under the restriction $0 < \inf_k\{\gamma_k\} \le \gamma_k \le \sup_k\{\gamma_k\} < 2$. But to expose our main idea with lighter notation, we focus on the constant step size case: $\gamma_k \equiv \gamma$. Obviously, the original PPA (3) is a special case of (4) with $\gamma = 1$. Moreover, the generalized PPA scheme (4) includes other interesting methods as special cases such as the generalized ALM in [34], the generalized ADMM in [12] and the generalized DRSM in [7,8]. For the convergence of (4) with $\gamma \in (0, 2)$, it was shown in [34] that under the assumption that T^{-1} is Lipschitz continuous at the origin; this is an extension of the analysis in [30] for the original PPA (3). Numerically, the step size $\gamma \in (0, 2)$, particularly when γ is close to 2, could accelerate the original PPA (3) as widely verified in the literature, see, e.g., [3,11,13,35] for some numerical verification in the particular ALM and ADMM contexts[1].

A common sense is that larger step sizes are preferred whenever the convergence can be theoretically ensured. We are thus interested in the question of whether or not the step size γ can be as large as 2 in the generalized PPA scheme (4). Theoretically, it seems that very little is known about this question—a clear case is when the PPA (3) corresponds to the DRSM. In this case, the scheme (4) with $\gamma = 2$ reduces to the Peaceman–Rachford splitting method (PRSM) in [21,26] and as analyzed in [7,10], the convergence of PRSM is not guaranteed. We will further discuss this case in Sect. 6. Empirically, this case works occasionally; and indeed sometimes it is more efficiently—as shown in [15] for the comparison between DRSM and PRSM.

In this note, we answer this perplexing question negatively: the convergence of the generalized PPA (4), if no additional assumption on T is assumed, is not necessarily convergent when $\gamma = 2$. In addition to the generic setting where T represents an abstract maximal monotone operator, we discuss several specific settings of (1) where T represents some particular operators and the scheme (4) can be specified as generalized versions of some very popular algorithms in the literature such as the ALM and the ADMM. Some counterexamples are constructed to show the divergence of (4) with $\gamma = 2$ in different settings. Our analysis thus ascertain that there is no convergence for the generalized versions of a series of popular algorithms with step size 2. A by-product of our analysis is that we further show the failure of convergence for the PRSM (even though it is known already in [7,10]) and a generalized version of the forward–backward splitting method (FBSM) in [25] with a step size 1.5 (its upper bound).

2 Preliminaries

This section recalls some definitions and known results for further discussions.

Definition 1 [2] Let $T : \mathcal{H} \to 2^{\mathcal{H}}$ be set-valued and maximal monotone. Then, T is called β-cocoercive if there exists a positive scalar β such that

$$\langle T(z) - T(z'), z - z' \rangle \geq \beta \|T(z) - T(z')\|^2, \ \forall\, z, z' \in \mathcal{H}.$$

[1] Note that it is even possible to consider $\gamma > 2$ in (4); but this case usually requires much stronger assumptions on T to ensure the convergence, see, e.g., [7] for details. We thus do not consider the case of $\gamma > 2$ in this note.

Definition 2 [2] Let $T : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be uniformly monotone with modulus $\Phi : [0, +\infty) \rightarrow [0, +\infty)$ if Φ is increasing, vanishes only at 0, and

$$\langle T(x) - T(y), x - y \rangle \geq \Phi(\|x - y\|).$$

It is clear that the strong monotonicity implies the uniform monotonicity, which itself implies the monotonicity, see e.g., [2].

In the following, we recall some known results regarding the convergence of (4) with $\gamma \in (0, 2)$. Their proofs can be found in, e.g., [34].

Theorem 1 [34] *Let $\{z^k\}$ be the sequence generated by the generalized PPA (4) with $\gamma \in (0, 2)$. For a solution point of (1) z^*, then we have*

$$\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 - \gamma(2 - \gamma) \left\| z^k - \tilde{z}^k \right\|^2, \tag{5}$$

where $\tilde{z}^k = J_{cT}(z^k)$.

Theorem 2 [34] *Let $\{z^k\}$ be the sequence generated by the generalized PPA (4) with $\gamma \in (0, 2)$. If $zer(T)$ is not empty, then $\{z^k\}$ globally converges to a solution point of (1).*

The following lemma is crucial for constructing the counterexamples to illustrate the divergence of (4) with $\gamma = 2$.

Lemma 1 [1,4] *Assume that $A \in \Re^{n \times n}$ $(n \geq 2)$, $A^\top = -A$ and the scalar $c > 0$. Then the matrix $(I + cA)^{-1}(I - cA)$ is orthogonal.*

3 Divergence of the generalized PPA (4) with $\gamma = 2$

We first show the divergence of the generalized PPA (4) with $\gamma = 2$ in the generic setting of finding a zero point of a maximal monotone operator, i.e., the problem (1).

We consider the case of (1) where T is a skew symmetric matrix in the space $\Re^{n \times n}$, denoted by A. That is, we solve the system of linear equation $Ax = 0$. In this case, the generalized PPA (4) with $\gamma = 2$ is specified as

$$z^{k+1} = 2\tilde{z}^k - z^k = (2(I + cA)^{-1} - I)(z^k) = (I + cA)^{-1}(I - cA)(z^k). \tag{6}$$

If $n = 2$, then the solution of $Ax = 0$ is unique, i.e., $x^* = \mathbf{0}$. We take $z^0 \neq \mathbf{0}$ as the initial iterate and implement the iteration (6). It follows from Lemma 1 that $\|z^k\|_2 = \|z^0\|_2 \neq 0$ for any k. Thus, the sequence $\{z^k\}$ is divergent.

Indeed, more examples for the case where $n > 2$ can also be constructed. For example, we choose A as

$$A(i, j) = \begin{cases} 0, & \text{if } i = j; \\ 1, & \text{if } i > j; \\ -1, & \text{if } i < j; \end{cases} \tag{7}$$

and fix $c = 1$ and $z^0 = e_n$. Hereafter, $e_i \in \Re^n$ denotes the vector whose i-th entry is 1 and the others are 0. Then, if $n = 2m - 1$ ($m > 1$), it is easy to verify that the sequence $\{z^k\}$ generated by (6), starting with z^0, is n-periodic satisfying

$$z^k = \begin{cases} e_1, & k = nt + 1; \\ (-1)^{i-1} e_i, & k = nt + i, \quad i = 2, \ldots, n; \end{cases} \qquad \forall\, t \in \mathbb{Z}.$$

Thus, the sequence $\{z^k\}$ is not convergent. Meanwhile, if $n = 2m$ ($m > 1$), then the sequence $\{z^k\}$ is $2n$-periodic satisfying

$$z^k = \begin{cases} e_1, & k = nt + 1; \\ (-1)^{i-1} e_i, & k = nt + i, \quad i = 2, \ldots, n; \\ -e_1, & k = nt + n + 1; \\ (-1)^i e_i, & k = nt + n + i, \quad i = 2, \ldots, n; \end{cases} \qquad \forall\, t \in \mathbb{Z},$$

which is clearly divergent.

4 Divergence of the generalized ALM with step size $\gamma = 2$

Now we consider some special cases of the abstract model (1) in which T represents some specific operators and accordingly the PPA (3) can be specified as some well-known algorithms in the literature. We show that the generalized PPA (4) with $\gamma = 2$ may be divergent for these special cases.

We first consider the canonical linearly constrained convex programming problem

$$\begin{array}{ll} \min_x & f(x) \\ s.t. & Ax = b, \end{array} \tag{8}$$

where $f : \Re^n \to \Re$ is a convex function, $A \in \Re^{m\times n}$, $b \in \Re^m$. Assume that $ri(dom(f)) \cap \mathcal{F} \neq \phi$, where $\mathcal{F} := \{u \in \Re^n |\ Ax = b\}$. As analyzed in [30], the dual problem of (8) is

$$\max_{\lambda} \{-f^*(A^\top \lambda) + \langle b, \lambda \rangle\}, \tag{9}$$

where "*" denotes the conjugate of a convex function, see [29]. Thus, the problem (8) can be regarded as a special case of (1) in sense of the optimality condition of (9). Note that the solution set of (9) is nonempty if the solution set of (8) is assumed to be nonempty.

When the ALM in [19,28] is applied to solve (8), the iterative scheme reads as

$$\begin{cases} x^{k+1} = \arg\min_x \{f(x) - \langle \lambda^k, Ax - b \rangle + \frac{\varrho}{2} \|Ax - b\|^2\}, \\ \lambda^{k+1} = \lambda^k - \varrho(Ax^{k+1} - b), \end{cases} \tag{10}$$

where $\varrho > 0$ is a penalty parameter. It was shown in [31] that the scheme (10) is a special case of the PPA (3) with $T := A \cdot \partial f^* \cdot (A^\top \cdot) - b$ and $c := \varrho$. Therefore, applying the generalized PPA (4), we obtain a generalized version of the ALM:

$$\begin{cases} x^{k+1} = \arg\min_x\{f(x) - \langle\lambda^k, Ax - b\rangle + \frac{\varrho}{2}\|Ax - b\|^2\}, \\ \lambda^{k+1} = \lambda^k - \gamma\varrho(Ax^{k+1} - b), \end{cases} \tag{11}$$

where the step size $\gamma \in (0, 2)$, see more analysis in, e.g., [17,34]. We refer to [3] for the verification of numerical acceleration of the generalized ALM (11).

To analyze the convergence of the generalized ALM (11), we recall a theorem in [34].

Theorem 3 *Let $\{(x^k, \lambda^k)\}$ be the sequence generated by the generalized ALM with $\gamma \in (0, 2)$ and $\varrho > 0$. Then, we have*

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2 - \gamma(2 - \gamma)\varrho^2\|Ax^{k+1} - b\|^2. \tag{12}$$

This theorem implies the strict contraction of the sequence $\{\lambda^k\}$ and thus essentially ensures the convergence of the generalized ALM (11) with $\gamma \in (0, 2)$ in sense of the dual variable λ. If $\gamma = 2$ in (11), the conclusion (12) reduces to

$$\|\lambda^{k+1} - \lambda^*\|^2 \leq \|\lambda^k - \lambda^*\|^2,$$

which does not guarantee the strict contraction of the sequence $\{\lambda^k\}$ and thus the convergence of (11) may not be ensured. It is also our starting point to construct an example to show that the generalized ALM (11) with $\gamma = 2$ is not necessarily convergent. Indeed, in (8), we specifically take

$$f(x) = w^\top x \;\text{ with } w = (1, 1)^\top, \; A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and } \; b = (0, 0)^\top.$$

Obviously, for this example the optimal primal and dual solutions are $x^* = (0, 0)^\top$ and $\lambda^* = (1, -1)^\top$, respectively. Furthermore, we set the penalty parameter $\varrho = 1$ in (11) and the initial iterate $\lambda^0 = (0, 0)^\top$. Then we have

$$x^k = \begin{cases} (-1, -1)^\top, & k \text{ is odd;} \\ (1, 1)^\top, & k \text{ is even.} \end{cases} \qquad \lambda^k = \begin{cases} 2(1, -1)^\top, & k \text{ is odd;} \\ (0, 0)^\top, & k \text{ is even.} \end{cases}$$

This means the sequence $\{\lambda^k\}$ generated by the generalized ALM (11) with $\gamma = 2$ is not convergent.

5 Divergence of the generalized ADMM with step size $\gamma = 2$

In this section, we show that when the PPA (3) reduces to the ADMM in [16], its generalized version (4) with $\gamma = 2$ may not be convergent. Thus, the generalized ADMM with step size 2 is not necessarily convergent.

We consider the following linearly constrained two-block separable convex minimization problem

$$\begin{array}{ll} \min_{x,y} & f(x) + g(y) \\ s.t. & Ax + By = b, \end{array} \tag{13}$$

where $f : \Re^n \to \Re$ and $g : \Re^l \to \Re$ are convex functions, $A \in \Re^{m\times n}$, $B \in \Re^{m\times l}$ and $b \in \Re^m$. Its dual problem can be written as

$$\min_{\lambda}\{f^*(A^\top\lambda) + g^*(B^\top\lambda) - b^\top\lambda\}. \tag{14}$$

Thus, finding a solution point of (13) can be regarded as a special case of (1). Recall that the iterative scheme of ADMM in [16] for (13) reads as

$$\begin{cases} x^{k+1} = \arg\min_x\{f(x) - \langle\lambda^k, Ax\rangle + \frac{\alpha}{2}\|Ax + By^k - b\|^2\}, \\ y^{k+1} = \arg\min_y\{g(y) - \langle\lambda^k, By\rangle + \frac{\alpha}{2}\|Ax^{k+1} + By - b\|^2\}, \\ \lambda^{k+1} = \lambda^k - \alpha(Ax^{k+1} + By^{k+1} - b), \end{cases} \tag{15}$$

where λ^k is the Lagrange multiplier and $\alpha > 0$ is a penalty parameter. In [12,14], it was shown that the ADMM (15) is a special case of the PPA (3) with $c = 1$ and $T = S_{\alpha,\mathcal{A},\mathcal{B}}$ where $S_{\alpha,\mathcal{A},\mathcal{B}}$ is defined as

$$S_{\alpha,\mathcal{A},\mathcal{B}} := G^{-1}_{\alpha,\mathcal{A},\mathcal{B}} - I \tag{16}$$

in which I is the identity mapping,

$$G_{\alpha,\mathcal{A},\mathcal{B}} = J_{\alpha\mathcal{A}} \circ (2J_{\alpha\mathcal{B}} - I) + (I - J_{\alpha\mathcal{B}}); \tag{17}$$

$$\mathcal{A} := A\partial f^* A^\top(\cdot) - b, \quad \mathcal{B} := B\partial g^* B^\top(\cdot); \tag{18}$$

and $J_{\alpha\mathcal{A}}$ and $J_{\alpha\mathcal{B}}$ are the resolvent operators of $\mathcal{A}$ and $\mathcal{B}$, respectively.

Hence, the generalized PPA (4) was specified in [12] (see Theorem 8 therein) to generate the generalized ADMM for model (13):

$$\begin{cases} x^{k+1} = \arg\min_x\{f(x) - \langle\lambda^k, Ax\rangle + \frac{\alpha}{2}\|Ax + By^k - b\|^2\}, \\ y^{k+1} = \arg\min_y\{g(y) - \langle\lambda^k, By\rangle + \frac{\alpha}{2}\|\gamma(Ax^{k+1} - b) - (1-\gamma)By^k + By\|^2\}, \\ \lambda^{k+1} = \lambda^k - \alpha(\gamma(Ax^{k+1} - b) - (1-\gamma)By^k + By^{k+1}), \end{cases} \tag{19}$$

where $\gamma \in (0, 2)$. We refer to [11,13,35] for some theoretical and numerical studies for the generalized ADMM (19) with $\gamma \in (0, 2)$.

Next, we construct an example to show the divergence of the generalized ADMM (19) with $\gamma = 2$. Indeed, we consider the special quadratic programming case of (13):

$$\begin{aligned} \min\ & \tfrac{1}{2}x^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x + \tfrac{1}{2}y^\top \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} y \\ s.t.\ & \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned} \tag{20}$$

It is easy to see that the optimal solution of (20) is $x^* = (0, 0)^\top$ and $y^* = (0, 0, 0)^\top$; and the optimal solution of its dual problem is $\lambda^* = (0, \lambda_2, \lambda_3)^\top$ with arbitrary λ_2 and λ_3 in $\Re$.

Applying (19) with $\gamma = 2$ to (20), if we fix $\alpha = 1$ and take the initial iterate $y^0 = \lambda^0 = (1, 1, 0)^\top$, then it is easy to see that the sequence $\{y^k\}$ is given by

$$y^k = \begin{cases} (0, -1, 0)^\top, & k \text{ is odd;} \\ (0, 1, 0)^\top, & k \text{ is even.} \end{cases} \tag{21}$$

Thus, it is clear that the sequence generated by the generalized ADMM (19) with $\gamma = 2$ is divergent for the example (20).

6 Divergence of the PRSM

Now, we consider the special case of (1) where T is the sum of two maximal monotone operators:

$$0 \in \mathcal{A}(z) + \mathcal{B}(z), \tag{22}$$

where both $\mathcal{A} : \mathcal{H} \to 2^{\mathcal{H}}$ and $\mathcal{B} : \mathcal{H} \to 2^{\mathcal{H}}$ are maximal monotone. To solve (22), it was shown in [12] that the well-known DRSM in [9,21] can be written as a special case of the PPA (3) with $T = S_{\alpha,\mathcal{A},\mathcal{B}}$, i.e.,

$$z^{k+1} = J_{S_{\alpha,\mathcal{A},\mathcal{B}}}(z^k), \tag{23}$$

where $S_{\alpha,\mathcal{A},\mathcal{B}}$ is defined in (16) (except that $\mathcal{A}$ and $\mathcal{B}$ are not necessarily given by (18)). Thus, applying the generalized PPA (4) results in the generalized DRSM for the special setting (22):

$$z^{k+1} = z^k - \gamma(z^k - J_{S_{\alpha,\mathcal{A},\mathcal{B}}}(z^k)) \;\; \gamma \in (0, 2), \tag{24}$$

with $\gamma \in (0, 2)$. The convergence analysis of this generalized DRSM (24) can be found in, e.g., [7,8,12,34]. Moreover, the PRSM in [21,26] is known to be the case where $\gamma = 2$ in (24), see, e.g., [7]. Indeed, for the PRSM, its divergence is known already, see [7,10,21].

A by-product of our previous analysis is that we can show that the PRSM is divergent for a class of problems in form of (22):

$$(A + B)z = 0, \tag{25}$$

whenever $A \in \Re^{n\times n}$ and $B \in \Re^{n\times n}$ are matrices. In the following, we present a theorem.

Theorem 4 *Assume that $A, B \in \Re^{n\times n}$ ($n \geq 2$), $A^\top = -A$, $B^\top = -B$ and a scalar $\alpha > 0$. Then the matrix $(I + \alpha A)^{-1}(I - \alpha A)(I + \alpha B)^{-1}(I - \alpha B)$ is orthogonal.*

Proof Note that

$$
\begin{aligned}
&(I+\alpha A)^{-1}(I-\alpha A)(I+\alpha B)^{-1}(I-\alpha B)\left[(I+\alpha A)^{-1}(I-\alpha A)(I+\alpha B)^{-1}(I-\alpha B)\right]^{\top}\\
&=(I+\alpha A)^{-1}(I-\alpha A)(I+\alpha B)^{-1}(I-\alpha B)(I-\alpha B^{\top})(I+\alpha B^{\top})^{-1}(I-\alpha A^{\top})(I+\alpha A^{\top})^{-1}\\
&=(I+\alpha A)^{-1}(I-\alpha A)(I+\alpha B)^{-1}(I-\alpha B)(I+\alpha B)(I-\alpha B)^{-1}(I+\alpha A)(I-\alpha A)^{-1}\\
&=(I+\alpha A)^{-1}(I-\alpha A)(I+\alpha B)^{-1}(I+\alpha B)(I-\alpha B)(I-\alpha B)^{-1}(I+\alpha A)(I-\alpha A)^{-1}\\
&=(I+\alpha A)^{-1}(I-\alpha A)(I+\alpha A)(I-\alpha A)^{-1}\\
&=I.
\end{aligned}
$$

The assertion follows directly. □

Based on this theorem, we will see that the PRSM is divergent for a class of problems defined in (25) for $n = 2$. Indeed, according to (24) with $\gamma = 2$, we have

$$
\begin{aligned}
z^{k+1} &= 2\left[(I+\alpha A)^{-1}\left(2(I+\alpha B)^{-1}-I\right)+I-(I+\alpha B)^{-1}\right]z^k - z^k\\
&= 2\left[(I+\alpha A)^{-1}\left(2(I+\alpha B)^{-1}-I\right)-\frac{1}{2}(2(I+\alpha B)^{-1}-I)\right]z^k\\
&= \left[2(I+\alpha A)^{-1}-I\right]\left[2(I+\alpha B)^{-1}-I\right]z^k\\
&= (I+\alpha A)^{-1}(I-\alpha A)(I+\alpha B)^{-1}(I-\alpha B)z^k. \qquad (26)
\end{aligned}
$$

For the case $n = 2$, the solution of (25) is unique, i.e., $z^* = \mathbf{0}$. If we take $z^0 \neq \mathbf{0}$ as the initial iterate and implement PRSM for (25), then we have $\|z^k\|_2 = \|z^0\|_2 \neq 0$ for any k because of Theorem 4. Thus, the sequence $\{z^k\}$ generated by the PRSM is divergent for the model (25).

For the case where $n > 2$, the solution set of (25) may not be unique, meaning the null space of $A + B$ may include nonzero vectors. For this case, we can construct the following example to show the divergence of PRSM. Let us choose

$$
A(i,j) = \begin{cases} 0, & \text{if } i = j;\\ 1, & \text{if } i > j;\\ -1, & \text{if } i < j; \end{cases} \quad \text{and} \quad B(i,j) = \begin{cases} -1, & \text{if } (i,j) = (1,n);\\ 1, & \text{if } (i,j) = (n,1);\\ 0, & \text{otherwise.} \end{cases}
$$

Then we fix $\alpha = 1$ in (24) and set the initial iterate as $z^0 = e_n$. For the case where $n = 2m + 1$ and $m > 1$, it is easy to verify that the sequence $\{z^k\}$ generated by (24) is $(n-1)$-periodic satisfying

$$
z^k = (-1)^i e_{i+1}, \ \text{ for } k = (n-1)t + i, \ i = 1,\ldots,n-1, \ \forall\, t \in \mathbb{Z}.
$$

For the case where $n = 2m$ and $m \geq 1$, the sequence $\{z^k\}$ generated by (24) is $2(n-1)$-periodic satisfying

$$
z^k = \begin{cases} (-1)^i e_{i+1}, & \text{if } k = 2t(n-1)+i;\\ (-1)^{i+1} e_{i+1}, & \text{if } k = (2t+1)(n-1)+i; \end{cases} \qquad i = 1,\ldots,n-1, \ \forall\, t \in \mathbb{Z}.
$$

Therefore, the PRSM is not convergent for a class of problems in form of (25).

7 Divergence of the generalized FBSM

To discuss the convergence/divergence of a generalized version of the FBSM for the problem (22), we follow the literature (see, e.g., [2,5,6]) and assume that $\mathcal{B}$ in (22) is single-valued and β-cocoercive. The FBSM alternates an explicit step (*forward step*) using the operator $\mathcal{B}$ with an implicit resolvent step (*backward step*) involving the operator $\mathcal{A}$. Its iterative scheme for (22) reads as

$$z^{k+1} = J_{c\mathcal{A}}(z^k - c\mathcal{B}(z^k)), \tag{27}$$

where c is a judiciously chosen step size to ensure its convergence. For example, the restriction $c \in (0, 2\beta)$ has been well studied in the literature, see [2,5,6,32] to just mention a few. Technically, the FBSM (27) is not a specific case of the PPA (2). But similarly as our previous analysis, we can consider a generalized version of (27) as the following

$$z^{k+1} = z^k - \gamma(z^k - J_{c\mathcal{A}}(z^k - c\mathcal{B}(z^k))), \tag{28}$$

with $\gamma > 0$ and $c \in (0, 2\beta)$. Note that we can also consider the case with dynamically changed step sizes γ_k for (28); but for simplicity we still fix it as a constant.

Another by-product of our previous analysis is that we can extend the presented analysis for the generalized PPA (4) to the generalized FBSM (28) and show that the latter is not necessarily convergent with a step size 1.5 (which is indeed the upper bound). In the next theorem, we first show some conditions in [2] that are sufficient to ensure the convergence of the generalized FBSM (28). Note that the condition on γ in [2] is "$\gamma_k \subseteq [0, \delta]$ and $\sum_{k=1}^{\infty} \gamma_k(\delta - \gamma_k) = +\infty$" because the case with dynamically chosen step sizes γ_k was considered therein.

Theorem 5 [2] *Let $\mathcal{A} : \mathcal{H} \to 2^{\mathcal{H}}$ and $\mathcal{B} : \mathcal{H} \to 2^{\mathcal{H}}$ be β-cocoercive with $\beta > 0$. Choose $c \in (0, 2\beta)$ and set $\delta := \min(1, \beta/c)) + 1/2$. Let $0 < \gamma < \delta$. Suppose the solution set of (22) is nonempty. Let $\{z^k\}$ be the sequence generated by the generalized FBSM (28) starting with $z^0 \in \mathcal{H}$. Then we have the following assertions.*

i. *There exists a solution point z^* of (22) such that*

$$\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\| - \gamma(\delta - \gamma)\|J_{c\mathcal{A}}(z^k - c\mathcal{B}(z^k))\|^2. \tag{29}$$

ii. *If one of the following holds:*
   a. *$\mathcal{A}$ is uniformly monotone on every nonempty bounded subset of dom$\mathcal{A}$,*
   b. *$\mathcal{B}$ is uniformly monotone on every nonempty bounded subset of dom$\mathcal{B}$,*

   Then, the sequence $\{z^k\}$ converges strongly to the unique solution point.

Proof (i) Combining Proposition 5.15 and the proof of Theorem 25.8 in [2], the first assertion follows directly. (ii) It follows directly from Theorem 25.8 in [2]. □

It follows from Theorem 5 that δ attains the maximum value 1.5 if $c = \beta$. This inspires us to construct the following example to show the divergence of the generalized

FBSM (28) when $\gamma = 1.5$. Again, we consider a special case of (22) in form of

$$(A + B)z = 0, \tag{30}$$

where both $\mathcal{A}$ and $\mathcal{B}$ are matrices. Let us choose

$$A = \begin{pmatrix} -0.3074 & 0 & 0 & 1.0208 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 12.0540 & 0 \\ -0.4253 & 0 & 0 & 1.1372 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0.1945 & 0 & 0 & -0.4719 \\ 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 0.6663 & 0 \\ 0.4719 & 0 & 0 & 0.1945 \end{pmatrix}. \tag{31}$$

Then, it is easy to see B is positive definite; and thus uniformly monotone. Thus, Case (b) of Assertion (ii) in Theorem 5 holds. Obviously, the solution point of the problem (30)–(31) is $z^* = \mathbf{0}$ since $(A + B)$ is nonsingular. According to Definition 1, by some elementary calculations, we know that the operator B is cocoercive with $\beta = 0.7466$. Thus, the iterative scheme (28) with $c = \beta$ and $\gamma = \delta = 1.5$ for (30) can be specified as

$$z^{k+1} = M \cdot z^k, \text{ with } M = 1.5(I + \beta A)^{-1}(I - \beta B) - 0.5 \cdot I. \tag{32}$$

Indeed, it follows from (31) that

$$M = \begin{pmatrix} 1.1642 & 0 & 0 & 0 \\ 0 & -0.1200 & 0 & 0 \\ 0 & 0 & -0.4246 & 0 \\ 0 & 0 & 0 & 0.1934 \end{pmatrix} \text{ and } \rho(M) = 1.1642 > 1,$$

where $\rho(M)$ is the spectral radius of M. Obviously, if we implement the generalized FBSM (28) with the initial iterate $z^0 = (1, 0, 0, 0)^\top$ for the problem (30)–(31), then the sequence $\{z^k\}$ is divergent.

8 Conclusion

The proximal point algorithm (PPA) is fundamental and it has been generalized with a step size less than 2 in the literature. But it was unknown whether or not the step size of the generalized PPA can be further enlarged to 2; a case with potential advantages in numerics. We construct some examples to show that in general it is not possible to enlarge the step size as large as 2 for the generalized PPA, for both the generic setting of finding a root of a maximal monotone operator and some specific convex programming settings. A problem complementary to our negative assertion regarding the step size 2 is that under what conditions the generalized PPA in some specific settings is convergent even if its step size is 2.

References

1. Bai, Z.Z., Hadjidimos, A.: Optimization of extrapolated Cayley transform with non-Hermitian positive definite matrix. Linear Algebra Appl. **463**, 322–339 (2014)
2. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
3. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)
4. Cayley, A.: On the theory of linear transformations. Camb. J. Math. **4**, 1–16 (1845)
5. Chen, G.H.-G., Rockafellar, R.T.: Convergence rate in forward–backward splitting. SIAM J. Optim. **7**, 1–25 (1997)
6. Combetters, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. Multiscale Model. Simul. **4**, 1168–1200 (2006)
7. Corman, E., Yuan, X.M.: A generalized proximal point algorithm and its convergence rate. SIAM J. Optim. **24**, 1614–1638 (2014)
8. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. http://arxiv.org/pdf/1406.4834.pdf
9. Douglas, J., Rachford, H.H.: On the numerical solution of the heat conduction problem in 2 and 3 space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
10. Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Ph.D. Dissertation, MIT, Cambridge (1989)
11. Eckstein, J.: Parallel alternating direction multiplier decomposition of convex programs. J. Optim. Theory Appl. **80**, 39–62 (1994)
12. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal points algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
13. Fang, E.X., Liu, H., He, B.S., Yuan, X.M.: The generalized alternating direction method of multipliers: new theoretical insights and applications. Math. Prog. Comput. **7**, 149–187 (2015)
14. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, pp. 299–331. North Holland, Amsterdam (1983)
15. Glowinski, R., Kärkkäinen, T., Majava, K.: On the convergence of operator-splitting methods. In: Kuznetsov, Y., Neittanmaki, P., Pironneau, O. (eds.) Numerical Methods for Scienfic Computing, Variational Problems and Applications, pp. 67–79. CIMNE, Barcelona (2003)
16. Glowinski, R., Marrocco, A.: Approximation par *élé*ments finis d'ordre un et r*é*solution par p*é*nalisation-dualit*é* d'une classe de probl*è*mes non lin*é*aires. RAIRO **R2**, 41–76 (1975)
17. Gol'shtein, E.G., Tret'yakov, N.V.: Modified Lagrangian in convex programming and their generalizations. Math. Progr. Study **10**, 86–97 (1979)
18. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Optim. **1**, 403–419 (1991)
19. Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)
20. Korpelevich, G.M.: The extragradient method for finding saddle points and other problems. Ekon. i Mat. Metody **12**, 747–756 (1976)
21. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
22. Martinet, B.: Regularization d'inequations variationelles par approximations successives. Rev. Fr. d'Inf. et de Rech. Opér. **4**, 154–159 (1970)
23. Moreau, J.J.: Proximité et dualit 'e dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
24. Nemirovski, A.: Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. **15**, 229–251 (2005)
25. Passty, G.B.: Ergodic convergence to a zero of the sume of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**, 383–390 (1979)
26. Peaceman, D.H., Rachford, H.H.: The numerical solution of parabolic elliptic differential equations. J. Soc. Ind. Appl. Math. **3**, 28–41 (1955)
27. Polyak, B.T.: Introduction to Optimization. Translations Series in Mathematics and Engineering, Optimization Software. Publications Division, New York (1987)

28. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)
29. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
30. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 97–116 (1976)
31. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 877–898 (1976)
32. Shen, L., Pan, S.H.: Linear convergence of the generalized PPA and several splitting methods for the composite inclusion problem. http://arxiv.org/abs/1508.05156v2
33. Teboulle, M.: Convergence of proximal-like algorithms. SIAM J. Optim. **7**, 1069–1083 (1997)
34. Tao, M., Yuan, X.M.: On the optimal linear convergence rate of a generalized proximal point algorithm. J. Sci. Comput. **74**, 826–850 (2018)
35. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented Lagrangian methods for semidefinite programming. Math. Progr. Comput. **2**, 203–230 (2010)