A note on the inertial proximal point method

Zhenguo Mu [1, *] and Yang Peng [2]

[1] *School of Mathematical Sciences, Nanjing Normal University, Nanjing, China.*
[2]*Department of Mathematics, Nanjing University, China.*

Abstract   The *proximal point method* (PPM) for solving maximal monotone operator inclusion problem is a highly powerful tool for algorithm design, analysis and interpretation. To accelerate convergence of the PPM, *inertial PPM* (iPPM) was proposed in the literature. In this note, we show that some of the attractive properties of the PPM, e.g., the generated sequence is contractive with the set of solutions, do not hold in general for iPPM. To partially inherit the advantages of the PPM and meanwhile incorporate inertial extrapolation steps, we propose an iPPM with alternating inertial steps. Our analyses show that the even subsequence generated by the proposed iPPM is contractive with the set of solutions. Moreover, we establish global convergence result under much relaxed conditions on the inertial extrapolation stepsizes, e.g., monotonicity is no longer needed and the stepsizes are significantly enlarged compared to existing methods. Furthermore, we establish certain nonasymptotic $O(1/k)$ and asymptotic $o(1/k)$ convergence rate results, where k denotes the iteration counter. These features are new to inertial type PPMs.

1. Introduction

Let $T : \Re^n \rightrightarrows \Re^n$ be a set-valued maximal monotone operator. In this paper, we consider the following operator inclusion problem

$$\text{find } x^* \in \Re^n \text{ such that } 0 \in T(x^*). \tag{1}$$

It is well known that (1) serves as a unified model for many problems of fundamental importance, including the fixed point problem, the variational inequality problem, the minimization of closed proper convex functions, and their variants and extensions. Therefore, its efficient solution is of practical interests in many situations.

The *proximal point method* (PPM, [25, 24, 31]) converts (1) to a fixed point problem of a firmly nonexpansive resolvent operator. Let $\lambda > 0$ be a constant. The resolvent operator of T is defined by $J_{\lambda T} := (I + \lambda T)^{-1}$, i.e., for any $x \in \Re^n$, $J_{\lambda T}(x)$ is the unique solution of $0 \in x + \lambda T(x)$. Initialized at any $x^0 \in \Re^n$, the PPM iterates for $k \geq 0$ as

$$x^{k+1} = J_{\lambda T}(x^k). \tag{2}$$

---

*Correspondence to: Zhenguo Mu (Email: zgmu.nju@gmail.com). School of Mathematical Sciences, Nanjing Normal University, Nanjing, China.

It turns out that the PPM is a very powerful algorithmic tool and contains many well known algorithms as special cases, including the classical augmented Lagrangian method [20, 29], the Douglas-Rachford splitting method [15] and the alternating direction method of multipliers [18, 17]. Interested readers are referred to the classical references [31, 30, 16, 19] for analysis and generalizations of the PPM. An equivalent representation of the PPM (2) is given by

$$0 \in \frac{x^{k+1} - x^k}{\lambda} + T(x^{k+1}),$$

which can be viewed as an implicit discretization of the evolution differential inclusion problem

$$0 \in \frac{\mathrm{d}x}{\mathrm{d}t} + T(x(t)). \tag{3}$$

It is known that the solution trajectory of (3) converges to a solution of (1) provided that T satisfies certain conditions, see e.g., [10]. To speed up convergence, the following second order evolution differential inclusion problem was studied in the literature:

$$0 \in \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \gamma \frac{\mathrm{d}x}{\mathrm{d}t} + T(x(t)), \tag{4}$$

where $\gamma > 0$ is a friction parameter. For the special case $n = 2$ and $T = \nabla f$, where $f : \Re^2 \to \Re$ is a differentiable convex function with attainable minimum, the system (4) characterizes roughly the motion of a heavy ball which rolls under its own inertia over the graph of f until friction stops it at a stationary point of f. In this case, the three terms in (4) denote, respectively, inertial force, friction force and gravity force. Therefore, the system (4) is usually referred as the *heavy-ball with friction* (HBF) system. In theory the convergence of the solution trajectories of the HBF system to a solution of (1) can be faster than those of the first-order system (3), while in practice the second order inertial term $\mathrm{d}^2 x / \mathrm{d}t^2$ can be exploited to design faster algorithms [1, 5]. Motivated by the properties of (4), an implicit discretization method was proposed in [2, 4]. Specifically, given x^{k-1} and x^k, the next point x^{k+1} is determined via

$$0 \in \frac{x^{k+1} - 2x^k + x^{k-1}}{h^2} + \gamma \frac{x^{k+1} - x^k}{h} + T(x^{k+1}),$$

which results to an iterative algorithm of the form

$$x^{k+1} = J_{\lambda T}(x^k + \alpha(x^k - x^{k-1})), \tag{5}$$

where $\lambda = h^2/(1 + \gamma h)$ and $\alpha = 1/(1 + \gamma h)$. Note that (5) is no more than a proximal point step applied to the extrapolated point $x^k + \alpha(x^k - x^{k-1})$, rather than x^k itself as in the classical PPM. Thus the resulting iterative scheme (5) is a two-step method and is usually referred as *inertial PPM* (iPPM). Convergence properties of (5) were studied in [2, 4] under some assumptions on the parameters α and λ. Subsequently, inexact and hybrid type iPPMs were studied in [26, 3, 23, 22]. Recently, there are increasing interests in studying inertial type algorithms, see, e.g., inertial forward-backward splitting methods [28, 27, 6], inertial Douglas-Rachford splitting method [9], inertial ADMM [7], and inertial forward-backward-forward method [8]. See also the latest references [11, 21, 14, 12, 13], which analyzed the convergence properties of inertial type algorithms for maximal monotone inclusion problem, variational inequality and structured convex optimization, and demonstrated their performance numerically on some imaging and data analysis problems.

In this note, we first give examples to illustrate that some of the attractive properties of the PPM do not hold anymore for iPPM. We then propose an iPPM with alternating inertial steps, which inherits the contractive property of the PPM to some extent. Our analyses show that, under much relaxed conditions, global convergence of the proposed iPPM can be guaranteed. In particular, the inertial extrapolation stepsizes do not need to be monotonically nondecreasing and can be significantly enlarged compared to existing methods. Furthermore, we establish certain nonasymptotic $O(1/k)$ and asymptotic $o(1/k)$ convergence rate results on a subsequence generated by the proposed iPPM. To the best of our knowledge, these features are new to the variants of PPM with inertial steps.

2. Features of PPM and iPPM

In practice, the proximal parameter λ in PPM usually varies step by step, i.e., given a sequence of positive parameters $\{\lambda_k\}$, the PPM appears as

$$x^{k+1} = J_{\lambda_k T}(x^k). \tag{6}$$

Similarly, the iPPM takes the form

$$x^{k+1} = J_{\lambda_k T}(x^k + \alpha_k(x^k - x^{k-1})), \tag{7}$$

where $\{\alpha_k\}$ is a sequence of nonnegative inertial extrapolation stepsizes. Classical requirements on the parameters to ensure the global convergence of iPPM are (i) $\lambda_k \geq \lambda$ for some $\lambda > 0$, and (ii) $0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < 1/3$ for all $k \geq 0$, see [4]. In the rest of this paper, we denote the set of solutions of (1) by $T^{-1}(0)$. The following lemma is very useful in our analysis, and its proof is elementary and is thus omitted.

Lemma 1 (Contractive property of PPM)
Let $T : \Re^n \rightrightarrows \Re^n$ be any set-valued maximal monotone operator and $\lambda > 0$. Suppose that x and x^+ satisfy $x^+ = J_{\lambda T}(x)$. Then, for any $x^* \in T^{-1}(0)$, it holds that $(x^+ - x^*)^T(x^+ - x) \leq 0$. As a consequence, it holds that $\|x^+ - x^*\|^2 \leq \|x - x^*\|^2 - \|x^+ - x\|^2$.

Let $\{x^k\}$ be the sequence generated by the PPM (6) and $x^* \in T^{-1}(0)$. Lemma 1 implies that the PPM is contractive with $T^{-1}(0)$. In particular, there holds

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^{k+1} - x^k\|^2. \tag{8}$$

A direct consequence of (8) is that if $x^k \in T^{-1}(0)$ for some k then all subsequent points will be freezed at x^k. In fact, by setting $x^* = x^k \in T^{-1}(0)$ in (8) we obtain the implication "$x^k \in T^{-1}(0) \implies x^{k+1} = x^k$". The converse is also true, i.e., if two consecutive points generated by the PPM are identical, then a solution is already reached. In short, the following property holds for PPM:

$$x^{k+1} = x^k \text{ if and only if } x^k \in T^{-1}(0). \tag{9}$$

Due to (9), it is natural to terminate PPM in practice by $\|x^{k+1} - x^k\| \leq \epsilon$ for some tolerance $\epsilon > 0$. We next give very simple examples to show that the sequence generated by iPPM violates the attractive properties (8) and (9).

Example 1 (violates (8) and the "if" direction in (9))
Suppose that we are minimizing $f(x) = x^2/2$ by iPPM (7), where λ_k is fixed at a positive constant λ. Given $x^{-1} = x^0 \neq 0$ and a nonnegative sequence $\{\alpha_k\}$, for $k \geq 0$ iterates as follows

$$x^{k+1} = \arg\min_x f(x) + \frac{1}{2\lambda}\|x - \bar{x}^k\|^2,$$

where $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$. The update appears as $x^{k+1} = \bar{x}^k/(1 + \lambda)$. We let $\alpha_k \equiv \alpha = 1/\lambda$. Then simple calculations show that

$$\bar{x}^0 = x^0, \ x^1 = \frac{x^0}{1 + \lambda}, \ \bar{x}^1 = \frac{(1 - \alpha\lambda)x^0}{1 + \lambda}, \ x^2 = \frac{(1 - \alpha\lambda)x^0}{(1 + \lambda)^2} = 0.$$

Clearly, x^2 is already the unique solution. However, further calculations show that

$$\bar{x}^2 = \frac{-\alpha x^0}{1 + \lambda}, \ x^3 = \frac{-\alpha x^0}{(1 + \lambda)^2} \neq 0.$$

This violates (9) since $x^k \in T^{-1}(0)$ does not imply $x^{k+1} = x^k$. This also implies that for iPPM x^{k+1} can be farther away to the set of solutions than x^k since in this example $|x^3 - x^*| > |x^2 - x^*| = 0$.

The following example shows that $x^{k+1} = x^k$ does not imply $x^k \in T^{-1}(0)$ for iPPM.

Example 2 (violates the "only if" direction in (9))
Again suppose that we are minimizing $f(x) = x^2/2$ by iPPM (7) with $\lambda_k \equiv \lambda = 4$. Let $x^{-1} = x^0 = -1$. The update appears as $x^{k+1} = \bar{x}^k/(1 + \lambda)$ with $\bar{x}^k = x^k + \alpha_k(x^k - x^{k-1})$. Set $\alpha_1 = 3/10$ and $\alpha_2 = 2/13$. Then simple calculation shows $x^0 = -1$, $x^1 = -1/5$ and $x^2 = x^3 = 1/125$. However, $1/125$ is not the minimizer of f. Thus, in general $x^{k+1} = x^k$ does not imply $x^k \in T^{-1}(0)$ for iPPM.

We give the following additional remarks on iPPM.

  (i) It is worth pointing out that in Examples 1 and 2 the extrapolation parameters $\{\alpha_k\}$ are less than $1/3$ (simply set $\lambda > 3$ in Example 1), as required by [4]. Though the monotonicity condition $\alpha_k \leq \alpha_{k+1}$ is violated in Example 2, we should note that this monotonically nondecreasing requirement is rather unreasonable since the iPPM reduces to the original PPM if $\alpha_k \equiv 0$.
 (ii) For iPPM, if $x^{k+1} = x^k \notin T^{-1}(0)$ then the next point x^{k+2} will be closer than x^{k+1} to the set of solutions $T^{-1}(0)$, because in this case conforms to a normal PPM step since $x^{k+2} = J_{\lambda_{k+1}T}(\bar{x}^{k+1}) = J_{\lambda_{k+1}T}(x^{k+1})$.
(iii) For iPPM, if $x^{k+1} = x^k \in T^{-1}(0)$, then all subsequent points will be equal to $x^{k+1} \in T^{-1}(0)$ because $\|x^{k+2} - x^*\| \leq \|x^{k+1} - x^*\|$ holds for any $x^* \in T^{-1}(0)$. This follows from setting $x^* = x^{k+1}$.

3. An iPPM with alternating inertial steps

Given the unsatisfactory properties of iPPM presented in Section 2, in this section we propose an iPPM with alternating inertial steps. This new iPPM has the advantage that the produced even subsequence is contractive with $T^{-1}(0)$. Furthermore, the inertial extrapolation stepsizes do not need to be monotonically nondecreasing and can vary freely in $[0, 1]$. These requirements are much less restrictive than those in [4], i.e., $0 \leq \alpha_k \leq \alpha_{k+1} \leq \alpha < 1/3$ for all $k \geq 0$.

For simplicity, in the following we assume that $\lambda_k \equiv \lambda$ for some $\lambda > 0$, and our algorithm and analyses can be simply generalized to the case with varying $\{\lambda_k\}$ as long as it is bounded below by some positive constant. Given $x^0 \in \Re^n$, $\lambda > 0$ and a sequence of nonnegative parameters $\{\alpha_k\}$. For $k \geq 0$, the proposed alternating inertial PPM iterates as

$$x^{k+1} = J_{\lambda T}(\bar{x}^k), \tag{10}$$

where $\bar{x}^k$ is defined as

$$\bar{x}^k := \begin{cases} x^k, & \text{if } k \text{ is even}; \\ x^k + \alpha_k(x^k - x^{k-1}), & \text{if } k \text{ is odd}. \end{cases} \tag{11}$$

We first give a lemma before the main convergence results.

Lemma 2 (Monotonicity property of two consecutive PPM iterations)
Let $\{x^k\}$ be the sequence generated by algorithm (10)-(11) from any initial point $x^0 \in \Re^n$. Then, it holds for all $k \geq 0$ that $\|x^{2k+3} - x^{2k+2}\| \leq \|x^{2k+2} - \bar{x}^{2k+1}\|$.

Proof
Let $k \geq 0$. It follows from $x^{2k+2} - x^{2k+3} \in \lambda T(x^{2k+3})$, $\bar{x}^{2k+1} - x^{2k+2} \in \lambda T(x^{2k+2})$ and the monotonicity of T that

$$\langle x^{2k+3} - x^{2k+2}, (x^{2k+2} - x^{2k+3}) - (\bar{x}^{2k+1} - x^{2k+2}) \rangle \geq 0. \tag{12}$$

Note that for any vectors $a, b \in \Re^n$ it holds that $\|a\|^2 - \|b\|^2 \geq 2\langle b, a - b \rangle$. Thus,

$$\begin{aligned} &\|x^{2k+2} - \bar{x}^{2k+1}\|^2 - \|x^{2k+3} - x^{2k+2}\|^2 \\ \geq\ & 2\langle x^{2k+3} - x^{2k+2}, (x^{2k+2} - x^{2k+3}) - (\bar{x}^{2k+1} - x^{2k+2}) \rangle. \end{aligned} \tag{13}$$

Then, the conclusion follows directly from (12) and (13). $\qquad\square$

Now, we are ready to present our main convergence results.

Theorem 1 (Global convergence and convergence rate)
Suppose that $0 \leq \alpha_k \leq 1$ for all $k \geq 0$ and $\lambda > 0$. Let $\{x^k\}$ be the sequence generated by algorithm (10)-(11) from any initial point $x^0 \in \Re^n$. Then, the following results hold.

(i) For any $x^* \in T^{-1}(0)$ and $k \geq 0$, there hold

$$\|x^{2k+2} - x^*\|^2 \leq \|x^{2k} - x^*\|^2 - \|x^{2k+2} - \bar{x}^{2k+1}\|^2, \tag{14}$$

$$\|x^{2k+3} - x^*\|^2 \leq \|x^{2k} - x^*\|^2 - \|x^{2k+2} - \bar{x}^{2k+1}\|^2. \tag{15}$$

(ii) $\lim_{k \to \infty} x^k$ exists and belongs to $T^{-1}(0)$.
(iii) For any $k \geq 1$ and $x^* \in T^{-1}(0)$, it holds that

$$\min_{0 \leq i \leq k-1} \|x^{2i+2} - \bar{x}^{2i+1}\|^2 \leq \frac{\|x^0 - x^*\|^2}{k}. \tag{16}$$

Moreover, as $k \to \infty$ we have

$$\min_{0 \leq i \leq k-1} \|x^{2i+2} - \bar{x}^{2i+1}\|^2 = o\left(\frac{1}{k}\right). \tag{17}$$

Proof
(i) Let $x^* \in T^{-1}(0)$ and $k \geq 0$. It follows from (10)-(11) and Lemma 1 that

$$\|x^{2k+1} - x^*\|^2 \leq \|x^{2k} - x^*\|^2 - \|x^{2k+1} - x^{2k}\|^2, \tag{18}$$

$$\|x^{2k+2} - x^*\|^2 \leq \|\bar{x}^{2k+1} - x^*\|^2 - \|x^{2k+2} - \bar{x}^{2k+1}\|^2. \tag{19}$$

Plug the identity $\bar{x}^{2k+1} = x^{2k+1} + \alpha_{2k+1}(x^{2k+1} - x^{2k})$ into the right-hand-side of (19), expand, and note that $\alpha_{2k+1} \geq 0$ and $(x^{2k+1} - x^*)^T(x^{2k+1} - x^{2k}) \leq 0$, we obtain

$$\begin{aligned}
\|x^{2k+2} - x^*\|^2 \leq{}& \|x^{2k+1} - x^*\|^2 - \|x^{2k+2} - x^{2k+1}\|^2 \\
&+ 2\alpha_{2k+1}(x^{2k+2} - x^{2k+1})^T(x^{2k+1} - x^{2k}).
\end{aligned} \tag{20}$$

The addition of (18) and (20) yields

$$\begin{aligned}
& \|x^{2k+2} - x^*\|^2 - \|x^{2k} - x^*\|^2 + \|x^{2k+2} - x^{2k+1}\|^2 \\
\leq{}& -\|x^{2k+1} - x^{2k}\|^2 + 2\alpha_{2k+1}(x^{2k+2} - x^{2k+1})^T(x^{2k+1} - x^{2k}) \\
\leq{}& -\alpha_{2k+1}^2\|x^{2k+1} - x^{2k}\|^2 + 2\alpha_{2k+1}(x^{2k+2} - x^{2k+1})^T(x^{2k+1} - x^{2k}) \\
={}& -\|x^{2k+2} - \bar{x}^{2k+1}\|^2 + \|x^{2k+2} - x^{2k+1}\|^2,
\end{aligned} \tag{21}$$

where the second "$\leq$" is due to $0 \leq \alpha_{2k+1} \leq 1$. This gives the relation (14). On the other hand, by using again (18) with $k := k+1$ and (21), we obtain (15):

$$\|x^{2k+3} - x^*\|^2 \leq \|x^{2k+2} - x^*\|^2 \leq \|x^{2k} - x^*\|^2 - \|x^{2k+2} - \bar{x}^{2k+1}\|^2.$$

(ii) Now we prove that $\lim_{k \to \infty} x^k$ exists and belongs to $T^{-1}(0)$. The following results are easy to derive based on (21).

(a) The sequence $\{x^{2k}\}$ is bounded and thus must have a limit point;
(b) The sequence $\{\|x^{2k} - x^*\|^2\}$ is monotonically decreasing, unless $x^{2k+2} = \bar{x}^{2k+1}$, in which case $x^{2k+2} \in T^{-1}(0)$ since $x^{2k+2} = \bar{x}^{2k+1} = J_{\lambda T}(\bar{x}^{2k+1})$;

(c) Let $n \geq 1$. The sum of (21) over $k = 1, \ldots, n-1$ gives

$$\sum_{k=0}^{n-1} \|x^{2k+2} - \bar{x}^{2k+1}\|^2 \leq \|x^0 - x^*\|^2. \tag{22}$$

By taking the limit $n \to \infty$ in (22), we obtain

$$\sum_{k=0}^{\infty} \|x^{2k+2} - \bar{x}^{2k+1}\|^2 \leq \|x^0 - x^*\|^2, \tag{23}$$

which implies that $\lim_{k \to \infty} (x^{2k+2} - \bar{x}^{2k+1}) = 0$.

Assume that the subsequence $\{x^{2k_j}\}_{j=0}^{\infty}$ converges to $x^\star$ as $j \to \infty$. By taking the limit $k_j \to \infty$ in $0 \in \lambda T(x^{2k_j}) + x^{2k_j} - \bar{x}^{2k_j-1}$ and noting that $x^{2k_j} - \bar{x}^{2k_j-1} \to 0$ and T is closed, we obtain $0 \in T(x^\star)$, i.e., $x^\star \in T^{-1}(0)$. Since x^* is arbitrarily chosen in $T^{-1}(0)$, we can replace x^* by $x^\star$ in (21), which yields

$$\|x^{2k+2} - x^\star\| \leq \|x^{2k} - x^\star\|.$$

Therefore, we have $\lim_{k \to \infty} x^{2k} = x^\star$. Furthermore, by setting $k := k-1$ and $x^* = x^\star$ in (15) we obtain $\|x^{2k+1} - x^\star\| \leq \|x^{2k} - x^\star\|$. As a result, $\lim_{k \to \infty} x^{2k+1} = x^\star$. In summary, we have $\lim_{k \to \infty} x^k = x^\star \in T^{-1}(0)$.

(iii) Now we prove the convergence rate results (16) and (17). Clearly, (16) follows immediately from (22). Moreover, by considering (23) and using the Cauchy principle, we obtain as $k \to \infty$ that

$$\frac{k-1}{2} \min_{0 \leq i \leq k-1} \|x^{2i+2} - \bar{x}^{2i+1}\|^2 \leq \sum_{i=\lfloor \frac{k-1}{2} \rfloor}^{k-1} \|x^{2i+2} - \bar{x}^{2i+1}\|^2 \to 0, \tag{24}$$

where $\lfloor \frac{k-1}{2} \rfloor$ denotes the greatest integer no greater than $\frac{k-1}{2}$. Thus, the relation (17) holds.   □

It follows from (14) that the sequence $\{x^{2k}\}$ is contractive with the set of solutions. Furthermore, we have the following corollaries.

Corollary 1 (Convergence rate)
Suppose that $0 \leq \alpha_k \leq 1$ for all $k \geq 0$ and $\lambda > 0$. Let $\{x^k\}$ be the sequence generated by algorithm (10)-(11) from any initial point $x^0 \in \Re^n$. Then, for any $k \geq 1$ and $x^* \in T^{-1}(0)$, it holds that

$$\min_{1 \leq i \leq 2k} \|x^{i+1} - \bar{x}^i\|^2 \leq \frac{\|x^0 - x^*\|^2}{k}. \tag{25}$$

Moreover, as $k \to \infty$ we have

$$\min_{1 \leq i \leq 2k} \|x^{i+1} - \bar{x}^i\|^2 = o\left(\frac{1}{k}\right). \tag{26}$$

Proof
It is obvious from the definition of $\bar{x}^i$ in (11) and Lemma 2 that

$$\min_{1 \leq i \leq 2k} \|x^{i+1} - \bar{x}^i\|^2 \leq \min_{0 \leq i \leq k-1} \|x^{2i+3} - x^{2i+2}\|^2 \leq \min_{0 \leq i \leq k-1} \|x^{2i+2} - \bar{x}^{2i+1}\|^2.$$

Then the conclusion of this corollary follows directly from (iii) of Theorem 1.   □

Since, for any $k \geq 0$, $x^{k+1} = \bar{x}^k$ would imply that $x^{k+1} \in T^{-1}(0)$, the results given in Corollary 1 can be viewed as convergence rate results on the optimality residue $\min_{1 \leq i \leq 2k} \|x^{i+1} - \bar{x}^i\|^2$. As such, this can be used in practice to terminate the algorithm. We also note that, different from the complexity results given in [14, 12, 13], the results given in Theorem 1 and Corollary 1 do not depend on the upper bound of $\{\alpha_k\}$. Also, the results presented here are measured by the Euclidean ℓ_2-norm, rather than certain weighted norm as given in [14, 12, 13]. More importantly, our convergence results do not depend on monotonicity of $\{\alpha_k\}$ and are established under the much relaxed condition $0 \leq \alpha_k \leq 1$ for all k. These features were previously not known for inertial type PPMs.

4. Concluding remarks

In this short note we have shown via providing concrete examples that some of the attractive properties of the classical PPM do not maintain in iPPM. To partially inherit the advantages of the PPM and meanwhile incorporate inertial extrapolation steps in the algorithm, we proposed an iPPM with alternating inertial steps. Our analyses have shown that the proposed algorithm generates a sequence whose even subsequence is contractive with the set of solutions. Our global convergence results and convergence rate results are established under much relaxed conditions compared to those required in the literature. In particular, we have removed the monotonicity requirement on the inertial extrapolation stepsizes and enlarged them from $[0, 1/3)$ to $[0, 1]$. Till far our discussion is only focused on the theoretical aspects, and a practically more important question is how to select the extrapolation stepsizes adaptively in computation so that the overall performance of inertial type PPMs can be significantly faster than the corresponding conventional PPMs. This topic is interesting for future investigation.

REFERENCES

1. F. Aluffi-Pentini, V. Parisi, and F. Zirilli. Algorithm 617. DAFNE: a differential-equations algorithm for nonlinear equations. *ACM Trans. Math. Software*, 10(3):317–324, 1984.
2. F. Alvarez. On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM J. Control Optim.*, 38(4):1102–1119 (electronic), 2000.
3. F. Alvarez. Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space. *SIAM J. Optim.*, 14(3):773–782 (electronic), 2004.
4. F. Alvarez and H. Attouch. An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.*, 9(1-2):3–11, 2001. Wellposedness in optimization and related topics (Gargnano, 1999).
5. A. S. Antipin. Minimization of convex functions on convex sets by means of differential equations. *Differentsial Equations*, 30(9):1475–1486, 1652, 1994.
6. H. Attouch, J. Peypouquet, and P. Redont. A dynamical approach to an inertial forward-backward algorithm for convex minimization. *SIAM J. Optim.*, 24(1):232–256, 2014.
7. R. I. Bot and E. R. Csetnek. An inertial alternating direction method of multipliers. *arXiv preprint arXiv:1404.4582*, 2014.
8. R. I. Bot and E. R. Csetnek. An inertial tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems. *arXiv preprint arXiv:1406.0724*, 2014.
9. R. I. Bot, E. R. Csetnek, and C. Hendrich. Inertial douglas-rachford splitting for monotone inclusion problems. *arXiv preprint arXiv:1403.3330*, 2014.
10. R. E. Bruck, Jr. Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Funct. Anal.*, 18:15–26, 1975.
11. A. Chambolle and T. Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *preprint*, September, 2014.
12. R. H. Chan, S. Q. Ma, and J. F. Yang. Inertial primal-dual algorithms for structured convex optimization. *arXiv preprint arXiv:1409.2992*, 2014.
13. C. H. Chen, R. H. Chan, S. Q. Ma, and J. F. Yang. Inertial Proximal ADMM for Linearly Constrained Separable Convex Optimization. SIAM Journal on Imaging Sciences, To appear.
14. C. H. Chen, S. Q. Ma, and J. F. Yang. A general inertial proximal point method for mixed variational inequality problem. *arXiv preprint arXiv:1407.8238*, 2014.
15. J. Douglas, Jr. and H. H. Rachford, Jr. On the numerical solution of heat conduction problems in two and three space variables. *Trans. Amer. Math. Soc.*, 82:421–439, 1956.
16. J. Eckstein and D. P. Bertsekas. On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Programming*, 55(3, Ser. A):293–318, 1992.
17. D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17–40, 1976.
18. R. Glowinski and A. Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires. *R.A.I.R.O., R2*, 9(R-2):41–76, 1975.
19. O. Güler. New proximal point algorithms for convex minimization. *SIAM J. Optim.*, 2(4):649–664, 1992.
20. M. R. Hestenes. Multiplier and gradient methods. *J. Optimization Theory Appl.*, 4:303–320, 1969.
21. D. A. Lorenz and T. Pock. An inertial forward-backward algorithm for monotone inclusions. *J. Math. Imaging Vis.*, 2014.

22. P. E. Maingé and N. Merabet. A new inertial-type hybrid projection-proximal algorithm for monotone inclusions. *Applied Mathematics of Computation*, 215:3149–3162, 2010.

23. P. E. Maingé and A. Moudafi. A proximal method for maximal monotone operators via discretization of a first order dissipative dynamical system. *J. Convex Anal.*, 14(4):869–878, 2007.

24. B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Ser. R-3):154–158, 1970.

25. J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

26. A. Moudafi and E. Elissabeth. Approximate inertial proximal methods using the enlargement of maximal monotone operators. *International Journal of Pure and Applied Mathemtics*, 5(3):283–299, 2003.

27. P. Ochs, T. Brox, and T. Pock. ipiasco: Inertial proximal algorithm for strongly convex optimization. *manuscript*, 2014.

28. P. Ochs, Y. Chen, T. Brox, and T. Pock. ipiano: Inertial proximal algorithm for non-convex optimization. *manuscript*, 2014.

29. M. J. D. Powell. A method for nonlinear constraints in minimization problems. In *Optimization (Sympos., Univ. Keele, Keele, 1968)*, pages 283–298. Academic Press, London, 1969.

30. R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.

31. R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.