On inexact relative-error hybrid proximal extragradient, forward-backward and Tseng's modified forward-backward methods with inertial effects

M. Marques Alves *   Raul T. Marcavillaca †

June 26, 2018

Abstract

In this paper, we propose and study the asymptotic convergence and nonasymptotic global convergence rates (iteration-complexity) of an inertial under-relaxed version of the relative-error hybrid proximal extragradient (HPE) method for solving monotone inclusion problems. We analyze the proposed method under more flexible assumptions than existing ones on the extrapolation and relative-error parameters. As applications, we propose and/or study inertial under-relaxed forward-backward and Tseng's modified forward-backward type methods for solving structured monotone inclusions.

2000 Mathematics Subject Classification: 90C25, 90C30, 47H05.

Key words: inertial, relaxed, proximal point method, HPE method, pointwise, ergodic, iteration-complexity, forward-backward algorithm, Tseng's modified forward-backward algorithm.

Introduction

Inertial proximal point-type algorithms for monotone inclusions gained a lot of attention in research recently (see, e.g., [3, 4] and the references therein). The first method of this type – the inertial proximal point (PP) method – for solving generalized equations with monotone operators was proposed and studied by Alvarez and Attouch in [2]. The intense research activity in the subject in the last years is in part due to its connections with fast first-order algorithms for convex programming (see, e.g., [3, 4, 5, 6, 7, 23]).

Since the inertial PP method of Alvarez and Attouch has been used as the hidden engine for the design and analysis of various first-order proximal algorithms with inertial effects, including inertial versions of ADMM, forward-backward and Douglas-Rachford algorithms (see, e.g., [3, 4, 10, 15, 16]), it is natural to attempt to design inexact versions of it. In [9], Bot and Csetnek proposed and studied the asymptotic convergence of an inertial version of the hybrid proximal extragradient (HPE) method

*Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (maicon.alves@ufsc.br). The work of this author was partially supported by CNPq grants no. 405214/2016-2 and 304692/2017-4.

†Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (rtm1111@hotmail.com). The work of this author was partially supported by CAPES.

of Solodov and Svaiter [30, 35]. The HPE method is an inexact PP algorithm for which, at each iteration, the corresponding proximal subproblems are supposed to be (inexactly) solved within a relative error criterion (this contrasts to the summable error criterion proposed by Rockafellar [34]).

In this paper, we propose and study the asymptotic convergence and nonasymptotic global convergence rates (iteration-complexity) of an inertial under-relaxed HPE method for solving monotone inclusions. The proposed method (Algorithm 1) differs from the existing inertial HPE-type method of Bot and Csetnek in the sense it is based on a different mechanism of iteration. Moreover, we prove its convergence and iteration-complexity under more flexible assumptions than those proposed in [9] on the extrapolation and relative-error parameters. As applications, we study inertial (under-relaxed) versions of the Tseng's modified forward-backward and forward-backward algorithms (see Algorithms 3 and 4) for solving structured monotone inclusions problems.

The main contributions of this paper will be further discussed in Section 1.

This paper is organized as follows. In Section 1, we present some preliminaries and basic results, review some existing algorithms and discuss in detail the main contributions of this paper. The inertial under-relaxed HPE method (Algorithm 1) is presented in Section 2; the main results are Theorems 2.5 (asymptotic convergence), and 2.7 and 2.8 (iteration-complexity). Sections 3 is devoted to present and study the inertial versions of the Tseng's modified forward-backward and forward-backward algorithms; the main results are Theorems 3.2 and 3.4. We finish the paper in Section 4 with some concluding remarks.

1 Preliminaries, basic results and general notation

1.1 Problem statement

Let $\mathcal{H}$ be a real Hilbert space and consider the general monotone inclusion problem (MIP) of finding $z \in \mathcal{H}$ such that

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

as well as the *structured* MIP

$$0 \in F(z) + B(z) \tag{2}$$

where T and B are (set-valued) maximal monotone operators on $\mathcal{H}$ and $F : D(F) \subset \mathcal{H} \to \mathcal{H}$ is a (point-to-point) monotone operator which is either *Lipschitz continuous* or *cocoercive* (see Subsections 3.1 and 3.2 for the precise statement). Problems (1) and (2) appear in different fields of applied mathematics and optimization including convex optimization, signal processing, PDEs, inverse problems, among others (see, e.g.,[8, 19]). We mention that under mild conditions on the operators F and B, problem (2) becomes a special instance of (1) with $T := F + B$.

In this paper, we propose and study the asymptotic convergence and the iteration-complexity of inertial under-relaxed versions of the *hybrid proximal extragradient* (HPE) method (Algorithm 1), and *Tseng's modified forward-backward* (Algorithm 3) and *forward-backward* (Algorithm 4) methods for solving (1), and (2), respectively.

The main contributions of (as well as the most related works with) this paper will be discussed along the next subsections, the main contributions being further summarized in Subsection 1.5.

1.2 The Alvarez–Attouch's inertial proximal point method

The *proximal point* (PP) *method* is an iterative scheme for seeking approximate solutions of (1). It was first proposed by Martinet [25] for solving monotone variational inequalities (with point-to-point operators) and further studied and developed by Rockafellar in his pioneering work [34]. In its exact formulation, an iteration of the PP method can be described by

$$z_k := (\lambda_k T + I)^{-1} z_{k-1} \qquad \forall k \geq 1, \tag{3}$$

where $\lambda_k > 0$ is a stepsize parameter and z_{k-1} is the current iterate.

The *inertial* PP *method* is a modification of (3) proposed and studied by Alvarez and Attouch in [2] as follows: for all $k \geq 1$,

$$\begin{cases} w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}), \\ z_k := (\lambda_k T + I)^{-1} w_{k-1}, \end{cases} \tag{4}$$

where $\{\alpha_k\}$ is a sequence of extrapolation parameters; note that if $\alpha_k \equiv 0$, then it follows that (4) reduces to the Rockafellar's PP method (3). Inertial PP-type methods deserve a lot of attention in nowadays research due the possibility of extending this methodology to different practical algorithms and, in part, as we mentioned earlier, due to its connections with fast first-order methods in convex programming. Asymptotic (weak) convergence of $\{z_k\}$ generated in (4) to a solution of (1) was first obtained in [2] under the assumptions that $\lambda_k \geq \underline{\lambda} > 0$ and

$$0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < 1/3 \qquad \forall k \geq 1. \tag{5}$$

The above upper bound $1/3$ on $\{\alpha_k\}$ has become standard in the analysis of inertial-like proximal algorithms (see, e.g., [15, 16, 23, 32]). It seems that (5) was first improved by Alvarez in [1, Proposition 2.5] in the setting of projective-proximal point-type methods and, more recently, by Attouch and Cabot in [4] with relaxation playing a central role. One of the main goals of this contribution is the analysis of an inertial under-relaxed HPE-type method under the assumption (actually more general than) (5) on $\{\alpha_k\}$; see Assumption (**A**).

1.3 The hybrid proximal extragradient method of Solodov and Svaiter

It is of course important to design and study inexact versions of known (exact) numerical algorithms, and this also applies to (3). In [34], Rockafellar proved that if, at each iteration $k \geq 1$, z_k is computed satisfying

$$\|z_k - (\lambda_k T + I)^{-1} z_{k-1}\| \leq e_k, \quad \sum_{k=1}^{\infty} e_k < \infty, \tag{6}$$

and $\{\lambda_k\}$ is bounded away from zero, then $\{z_k\}$ converges (weakly) to a solution of (1). Many modern inexact versions of the PP method (3), as opposed to the summable error criterion (6), use *relative error tolerances* for solving the associated subproblems. The first methods of this type were proposed by Solodov and Svaiter in [35, 36] and subsequently studied in [29, 30, 31, 37, 38]. The key idea consists of observing that (3) can be decoupled as

$$v_k \in T(z_k), \quad \lambda_k v_k + z_k - z_{k-1} = 0, \tag{7}$$

and then relaxing (7) within relative error tolerance criteria. Among these new methods, the HPE method [35] has been shown to be very effective as a framework for the design and analysis of many concrete algorithms (see, e.g., [9, 14, 17, 20, 21, 24, 26, 27, 28, 31, 35, 37, 38]). It can be described as follows: for all $k \geq 1$,

$$\begin{cases} v_k \in T^{\varepsilon_k}(\tilde{z}_k), \quad \|\lambda_k v_k + \tilde{z}_k - z_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2, \\ z_k := z_{k-1} - \lambda_k v_k, \end{cases} \tag{8}$$

where $\sigma \in [0,1[$. (see Subsection 1.6 for the general notation on ε-enlargements $T^{\varepsilon}(\cdot)$.) Note that if $\sigma = 0$, then it follows that (8) reduces to the exact PP method (3). As we mentioned before, recently Bot and Csetnek [9] proposed and studied an inertial proximal-like algorithm which combines ideas from (4) and (8). They have proved asymptotic convergence of their method under the assumption $\alpha(5+4\sigma^2)+\sigma^2 < 1$ on α and σ, where $\sigma \in [0,1[$ is as in (8) and $0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < 1$ for all $k \geq 1$ (cf. (5)). This condition enforces $\alpha \approx 0$ whenever $\sigma \approx 1$. This would, in particular, degenerate the desired inertial effect in many important applications of HPE-type methods for which $\sigma = 0.99$ is known (experimentally) to be the best choice among all possible $\sigma \in [0,1[$ (see, e.g., [17, 18, 26, 27]).

In this paper, we propose an inertial under-relaxed HPE-type method (Algorithm 1) with guarantee of asymptotic convergence and iteration-complexity (both pointwise and ergodic) under the assumption (actually more general than) (5) on $\{\alpha_k\}$; see Assumption (**A**). The price to pay is to perform, in addition to inertial, under-relaxed steps. On the other hand, the under-relaxed parameter $\tau \in]0,1]$ is explicitly computed and, in the case of (5), $\tau \geq 0.5$, the latter lower bound being uniform on $\sigma \in [0,1[$ (see the third remark following Assumption (**A**)). We also emphasize that our algorithm is different of the corresponding one in [9], in the sense it is based on a different mechanism of iteration.

The main convergence results on Algorithm 1 are Theorems 2.5, 2.7 and 2.8. It seems it is the first time in the literature that global (ergodic) $\mathcal{O}(1/k)$ convergence rates are obtained for inertial-like proximal algorithms (see Theorem 2.8).

1.4 Forward-backward and Tseng's modified forward-backward methods

With its roots in the projected gradient algorithm for convex optimization, the *forward-backward method* (see, e.g., [22, 33]) is one of the most popular numerical algorithms for solving the structured monotone inclusion problem (2), having numerous applications in modern applied mathematics (see, e.g., [8]). It can be described as follows: for all $k \geq 1$,

$$z_k := (\lambda_k B + I)^{-1}(z_{k-1} - \lambda_k F(z_{k-1})), \tag{9}$$

where $\lambda_k > 0$ is a stepsize parameter and z_{k-1} is the current iterate. Under the assumption that $F : \mathcal{H} \to \mathcal{H}$ is *cocoercive* and $\{\lambda_k\}$ is within a certain range, it follows that the sequence $\{z_k\}$ generated in (9) is weakly convergent to a solution of (2) (see, e.g., [8]). In the seminal paper [41], Tseng proposed and studied the following modification of (9) – known as the *Tseng's modified forward-backward method*: for all $k \geq 1$,

$$\begin{cases} \tilde{z}_k := (\lambda_k B + I)^{-1}(z_{k-1} - \lambda_k F(z_{k-1})), \\ z_k := \tilde{z}_k - \lambda_k (F(\tilde{z}_k) - F(z_{k-1})). \end{cases} \tag{10}$$

We clearly see that (10) generalizes (9) by performing an additional forward step to define the next iterate z_k. This is crucial to obtain convergence under the (weaker than cocoercivity) assumption of *Lipschitz continuity* on F (see, e.g., [8, 41]). Since both forward-backward and Tseng's modified forward-backward methods are known to be special instances of the HPE method (8) for solving (1) with $T := F + B$ (see, e.g., [30, 35, 39]), we have managed to propose and/or study inertial under-relaxed versions of (9) and (10) – namely, Algorithms 3 and 4, respectively – as special instances of the proposed inertial under-relaxed HPE method (Algorithm 1). As a by-product of the results obtained for Algorithm 1, we prove their asymptotic convergence as well as their global $\mathcal{O}(1/\sqrt{k})$ pointwise and $\mathcal{O}(1/k)$ ergodic convergence rates/iteration-complexity (see Theorems 3.2 and 3.4). We discuss some existing inertial/relaxed variants of (9) and (10) as well as how they are related to Algorithms 3 and 4 in the remarks following them. We also emphasize that, since Algorithms 3 and 4 will be analyzed within the framework of Algorithm 1, they will automatically inherit all the possible benefits from the proposed policy of choosing the upper bound on the sequence of inertial parameters and the relaxation parameter (see Assumption **(A)**, the remarks following it, and the remarks following Algorithms 3 and 4).

1.5 The main contributions of this work

We summarize the main contributions of this work are as follows:

(i) Asymptotic convergence and nonasymptotic global $\mathcal{O}(1/\sqrt{k})$ *pointwise* and $\mathcal{O}(1/k)$ *ergodic* convergence rates (iteration-complexity) of an inertial under-relaxed HPE method (Algorithm 1) for solving (1) under more flexible than existing assumptions on the choice of inertial $\{\alpha_k\}$ and relative-error $\sigma \in [0, 1[$ parameters (see Assumption **(A)** and the remarks following it). We show, in particular, that it is possible to assume the upper bound $1/3$ on the sequence of inertial parameters $\{\alpha_k\}$, which became standard in the analysis of inertial-type proximal algorithms, at the price of performing under-relaxed iterations with explicitly computed parameter $\tau \geq 0.5$, where the latter lower bound is uniform on the relative-error parameter $\sigma \in [0, 1[$. We also emphasize that, up to the authors knowledge, it is the first time in the literature that an iteration-complexity analysis is performed for inertial HPE-type methods (see Theorems 2.7 and 2.8) and it seems it is also the first time that *ergodic* iteration-complexity results are established for inertial proximal-type algorithms.

(ii) Asymptotic convergence and pointwise and ergodic iteration-complexity of inertial under-relaxed versions of the Tseng's modified forward-backward method (Algorithm 3) and forward-backward method (Algorithm 4) for solving (2) under the assumption that F is monotone and either Lipschitz continuous or cocoercive. Analogously to (i), in this case, the proposed methods also benefit from the more flexible than standard assumptions on the choice of inertial parameters (see Subsections 3.1 and 3.2 for a discussion).

1.6 General notation and basics on monotone operators and ε–enlargements

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. A set-valued map $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is said to be a *monotone operator* if $\langle z - z', v - v' \rangle \geq 0$ for all $v \in T(z)$ and $v' \in T(z')$. On the other hand, $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is *maximal monotone* if T is monotone and its graph $G(T) := \{(z, v) \in \mathcal{H} \times \mathcal{H} \,|\, v \in T(z)\}$ is not properly contained in the graph of any other monotone operator on $\mathcal{H}$. The inverse of $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is $T^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$, defined at any $z \in \mathcal{H}$ by $v \in T^{-1}(z)$ if and only if $z \in T(v)$. The resolvent of a maximal monotone operator $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is $(T + I)^{-1}$ and $z = (T + I)^{-1}x$ if and only if $x - z \in T(z)$. The operator $\gamma T : \mathcal{H} \rightrightarrows \mathcal{H}$, where $\gamma > 0$, is defined by $(\gamma T)z := \gamma T(z) := \{\gamma v \mid v \in T(z)\}$.

For $T : \mathcal{H} \rightrightarrows \mathcal{H}$ maximal monotone and $\varepsilon \geq 0$, the ε-enlargement [11] of T is the operator $T^{\varepsilon} : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$T^{\varepsilon}(z) := \{v \in \mathcal{H} \mid \langle z - z', v - v'\rangle \geq -\varepsilon \;\; \forall (z', v') \in G(T)\} \quad \forall z \in \mathcal{H}. \tag{11}$$

Note that $T(z) \subset T^{\varepsilon}(z)$ for all $z \in \mathcal{H}$.

The following summarizes some useful properties of T^{ε} (see, e.g., [13, Lemma 3.1 and Proposition 3.4(b)]).

Proposition 1.1. *Let $T, S : \mathcal{H} \rightrightarrows \mathcal{H}$ be set-valued maps. Then,*

(a) *If $\varepsilon \leq \varepsilon'$, then $T^{\varepsilon}(z) \subseteq T^{\varepsilon'}(z)$ for every $z \in \mathcal{H}$.*

(b) *$T^{\varepsilon}(z) + S^{\varepsilon'}(z) \subseteq (T + S)^{\varepsilon+\varepsilon'}(z)$ for every $z \in \mathcal{H}$ and $\varepsilon, \varepsilon' \geq 0$.*

(c) *T is monotone, if and only if $T \subseteq T^0$.*

(d) *T is maximal monotone, if and only if $T = T^0$.*

(e) *If T is maximal monotone, $\{(\tilde{z}_k, v_k, \varepsilon_k)\}$ is such that $v_k \in T^{\varepsilon_k}(\tilde{z}_k)$, for all $k \geq 1$, $w-\lim_{k\to\infty} \tilde{z}_k = z$, $\lim_{k\to\infty} v_k = v$ and $\lim_{k\to\infty} \varepsilon_k = \varepsilon$, then $v \in T^{\varepsilon}(z)$.*

Next we present the transportation formula for ε-enlargements.

Theorem 1.2. ([12, Theorem 2.3]) *Suppose $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is maximal monotone and let $\tilde{z}_\ell, v_\ell \in \mathcal{H}$, $\varepsilon_\ell, \alpha_\ell \in \mathbb{R}_+$, for $\ell = 1, \ldots, k$, be such that*

$$v_\ell \in T^{\varepsilon_\ell}(\tilde{z}_\ell), \quad \ell = 1, \ldots, k, \quad \sum_{\ell=1}^{k} \alpha_\ell = 1,$$

and define

$$\tilde{z}_k^a := \sum_{\ell=1}^{k} \alpha_\ell \, \tilde{z}_\ell \,, \quad v_k^a := \sum_{\ell=1}^{k} \alpha_\ell \, v_\ell \,, \quad \varepsilon_k^a := \sum_{\ell=1}^{k} \alpha_\ell \left(\varepsilon_\ell + \langle z_\ell - \tilde{z}_k^a, v_\ell - v_k^a \rangle \right).$$

Then, the following hold:

(a) *$\varepsilon_k^a \geq 0$ and $v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$.*

(b) *If, in addition, $T = \partial f$ for some proper, convex and closed function f and $v_\ell \in \partial_{\varepsilon_\ell} f(\tilde{z}_\ell)$ for $\ell = 1, \ldots, k$, then $v_k^a \in \partial_{\varepsilon_k^a} f(\tilde{z}_k^a)$.*

The following well-known property will be also useful in this paper. For any $w, z \in \mathcal{H}$ and $p \in [0, 1]$, we have

$$\|pw + (1 - p)z\|^2 = p\|w\|^2 + (1 - p)\|z\|^2 - p(1 - p)\|w - z\|^2. \tag{12}$$

2 An inertial under-relaxed hybrid proximal extragradient method

Consider the monotone inclusion problem (1), i.e., the problem of finding $z \in \mathcal{H}$ such that

$$0 \in T(z) \tag{13}$$

where T is a maximal monotone operator on $\mathcal{H}$ for which $T^{-1}(0) \neq \emptyset$.

In this section, we propose and study the asymptotic convergence and nonasymptotic global convergence rates (iteration-complexity) of an inertial under-relaxed hybrid proximal extragradient (HPE) method (Algorithm 1) for solving (13). Regarding the iteration-complexity analysis, we consider the following notion of approximate solution for (13): given tolerances $\rho, \epsilon > 0$, find $z, v \in \mathcal{H}$ and $\varepsilon \geq 0$ such that

$$v \in T^{\varepsilon}(z), \quad \|v\| \leq \rho, \quad \varepsilon \leq \epsilon. \tag{14}$$

Note that $\rho = \epsilon = 0$ in (14) gives $0 \in T(z)$, i.e., in this case $z \in \mathcal{H}$ is a solution of (13) (for a more detailed discussion on (14), see, e.g., [30]).

The main results in this section are Theorems 2.5, 2.7 and 2.8. We refer the reader to the remarks and comments following each of the above mentioned theorems for a discussion regarding the contribution of each of them in the light of related results available in the current literature.

Algorithm 1. An inertial under-relaxed HPE method for solving (13)

Input: $z_0 = z_{-1} \in \mathcal{H}$ and $0 \leq \alpha, \sigma < 1$ and $0 < \tau \leq 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}). \tag{15}$$

3: Find $(\tilde{z}_k, v_k, \varepsilon_k) \in \mathcal{H} \times \mathcal{H} \times \mathbb{R}_+$ and $\lambda_k > 0$ such that

$$v_k \in T^{\varepsilon_k}(\tilde{z}_k), \quad \|\lambda_k v_k + \tilde{z}_k - w_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - w_{k-1}\|^2. \tag{16}$$

4: Define

$$z_k := w_{k-1} - \tau \lambda_k v_k. \tag{17}$$

Remarks.

(i) Algorithm 1 clearly combines the inertial proximal point (PP) and the HPE methods (4) and (8), respectively. It reduces to (4) when $\sigma = 0$ and $\tau = 1$. Indeed, in this case, using (16), (17) and Proposition 1.1(d), we find $0 \in \lambda_k T(z_k) + z_k - [z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2})]$ for all $k \geq 1$ (cf. iteration $(\mathcal{A}_0)$–$(\mathcal{A}_2)$ in [2]).

(ii) A similar inertial relaxed relative-error PP algorithm was proposed and analyzed by Alvarez in [1]. We emphasize that in contrast to Algorithm 1, the algorithm proposed by Alvarez is a projective-type algorithm (see, e.g., [36]) and it is based on a different mechanism of iteration.

(iii) Algorithm 1 generalizes the HPE method of Solodov and Svaiter [30] and (a special instance of) the under-relaxed HPE method of Svaiter [40]. Indeed, the HPE method is obtained by letting $\alpha = 0$ and $\tau = 1$, in which case $w_{k-1} = z_{k-1}$, while the under-relaxed HPE method (with $t_k \equiv \tau$, in the notation of the latter reference) appears whenever $\alpha = 0$ in Algorithm 1.

(iv) As we mentioned in Subsection 1.3, an inertial HPE-type method was recently proposed and studied by Bot and Csetnek in [9]. We refer the reader to Subsection 1.3 for a discussion of the contributions of this paper in the light of the latter reference, regarding the HPE-type methods.

(v) We emphasize that, in contrast to the analysis presented in this work – see Theorems 2.7 and 2.8 –, in all cases of inertial-type algorithms which were mentioned in remarks (i)–(iv) no iteration-complexity analysis has been obtained.

The next three results, especially Proposition 2.3, will be important for proving the main results on convergence and iteration-complexity of Algorithm 1.

Proposition 2.1. *Let $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{w_k\}$ be generated by* Algorithm 1 *and define, for all $k \geq 1$,*

$$s_k := \max\left\{\eta\|z_k - w_{k-1}\|^2, (1-\sigma^2)\tau\|\tilde{z}_k - w_{k-1}\|^2\right\} \tag{18}$$

where

$$\eta := \eta(\sigma,\tau) := \frac{2}{(1+\sigma)\tau} - 1 > 0. \tag{19}$$

Then, for any $z^* \in T^{-1}(0)$,

$$\|z_k - z^*\|^2 + s_k \leq \|w_{k-1} - z^*\|^2 \qquad \forall k \geq 1. \tag{20}$$

Proof. Using (16), (17) and Lemma A.1(b) we obtain

$$\|z_k - z^*\|^2 + (1-\sigma^2)\tau\|\tilde{z}_k - w_{k-1}\|^2 + \tau(1-\tau)\|\lambda_k v_k\|^2 \leq \|w_{k-1} - z^*\|^2. \tag{21}$$

Note now that from (17) and (16) we have

$$\begin{aligned}\tau^{-1}\|z_k - w_{k-1}\| = \|\lambda_k v_k\| &\leq \|\lambda_k v_k + \tilde{z}_k - w_{k-1}\| + \|\tilde{z}_k - w_{k-1}\| \\ &\leq (1+\sigma)\|\tilde{z}_k - w_{k-1}\|,\end{aligned}$$

which, in turn, gives

$$(1-\sigma^2)\tau\|\tilde{z}_k - w_{k-1}\|^2 \geq \frac{(1-\sigma)}{\tau(1+\sigma)}\|z_k - w_{k-1}\|^2. \tag{22}$$

On the other hand, (17) yields

$$\tau(1-\tau)\|\lambda_k v_k\|^2 = \tau^{-1}(1-\tau)\|\tau\lambda_k v_k\|^2 = \tau^{-1}(1-\tau)\|z_k - w_{k-1}\|^2. \tag{23}$$

To finish the proof, note that (20) is a direct consequence of (18), (21)–(23) and (19). $\square$

Lemma 2.2. *Let* $\{z_k\}$*,* $\{w_k\}$ *and* $\{\alpha_k\}$ *be generated by* Algorithm 1 *and let* $z \in \mathcal{H}$*. Then, for all* $k \geq 1$*,*

$$\|w_{k-1} - z\|^2 = (1+\alpha_{k-1})\|z_{k-1} - z\|^2 - \alpha_{k-1}\|z_{k-2} - z\|^2 + \alpha_{k-1}(1+\alpha_{k-1})\|z_{k-1} - z_{k-2}\|^2.$$

Proof. From (15) we have $z_{k-1} - z = (1+\alpha_{k-1})^{-1}(w_{k-1} - z) + \alpha_{k-1}(1+\alpha_{k-1})^{-1}(z_{k-2} - z)$ and $w_{k-1} - z_{k-2} = (1+\alpha_{k-1})(z_{k-1} - z_{k-2})$, which combined with the property (12) yield the desired identity. $\square$

Proposition 2.3. *Let* $\{z_k\}$*,* $\{w_k\}$ *and* $\{\alpha_k\}$ *be generated by* Algorithm 1 *and let* $\{s_k\}$ *be as in* (18)*. Let also* $z^* \in T^{-1}(0)$ *and define*

$$(\forall k \geq -1) \quad \varphi_k := \|z_k - z^*\|^2 \;\; \textit{and} \;\; (\forall k \geq 1) \quad \delta_k := \alpha_{k-1}(1+\alpha_{k-1})\|z_{k-1} - z_{k-2}\|^2. \tag{24}$$

Then, $\varphi_0 = \varphi_{-1}$ *and*

$$\varphi_k - \varphi_{k-1} + s_k \leq \alpha_{k-1}(\varphi_{k-1} - \varphi_{k-2}) + \delta_k \qquad \forall k \geq 1, \tag{25}$$

i.e., the sequences $\{\varphi_k\}$*,* $\{s_k\}$*,* $\{\alpha_k\}$ *and* $\{\delta_k\}$ *satisfy the assumptions of* Lemma A.4.

Proof. Using Lemma 2.2 with $z = z^*$ and (24) we obtain, for all $k \geq 1$,

$$\|w_{k-1} - z^*\|^2 = (1+\alpha_{k-1})\varphi_{k-1} - \alpha_{k-1}\varphi_{k-2} + \delta_k,$$

which combined with Proposition 2.1 and the definition of φ_k in (24) yields (25). The identity $\varphi_0 = \varphi_{-1}$ follows from the fact that $z_0 = z_{-1}$ and the first definition in (24). $\square$

Next we present the first result on the asymptotic convergence of Algorithm 1.

Theorem 2.4 (first result on the weak convergence of Algorithm 1)**.** *Let* $\{z_k\}$*,* $\{\lambda_k\}$ *and* $\{\alpha_k\}$ *be generated by* Algorithm 1*. If the following holds*

$$\sum_{k=0}^{\infty} \alpha_k \|z_k - z_{k-1}\|^2 < +\infty \tag{26}$$

and, additionally, $\lambda_k \geq \underline{\lambda} > 0$*, for all* $k \geq 1$*, then the sequence* $\{z_k\}$ *converges weakly to a solution of the monotone inclusion problem* (13)*.*

Proof. Using Proposition 2.3, (26), the fact that $\alpha_k \leq \alpha$ for all $k \geq 1$ and Lemma A.4, one concludes that (i) $\lim_{k\to\infty} \|z_k - z^*\|$ exist for every $z^* \in \Omega := T^{-1}(0)$, and $\sum_{k=1}^{\infty} s_k < +\infty$, which gives (ii) $\lim_{k\to\infty} s_k = 0$, where $\{s_k\}$ is as in (18). In particular, $\{z_k\}$ is bounded. Using (ii), (16)–(18) and the assumption $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$, we find

$$\lim_{k\to\infty} \|z_k - w_{k-1}\| = \lim_{k\to\infty} \|\tilde{z}_k - w_{k-1}\| = \lim_{k\to\infty} \|v_k\| = \lim_{k\to\infty} \varepsilon_k = 0. \tag{27}$$

Now let $z^\infty \in \mathcal{H}$ be a weak cluster point of $\{z_k\}$ (recall that it is bounded) and let $\{z_{k_j}\}$ be such that $z_{k_j} \rightharpoonup z^\infty$. Using (27) and the inclusion in (16) we obtain

$$(\forall j \geq 1) \;\; v_{k_j} \in T^{\varepsilon_{k_j}}(\tilde{z}_{k_j}), \;\; \lim_{j\to\infty} v_{k_j} = 0, \;\; \lim_{j\to\infty} \varepsilon_{k_j} = 0 \;\text{ and }\; w - \lim_{j\to\infty} z_{k_j} = z^\infty, \tag{28}$$

which, in turn, combined with Proposition 1.1(e) yields $z^\infty \in \Omega = T^{-1}(0)$, and so the desired result follows form (i) and Lemma A.3. $\square$

Remark. Condition (26) appeared for the first time in [2], and since then it has become a standard assumption in the asymptotic convergence analysis of different inertial PP-type algorithms. Next, we present a sufficient condition on the input parameters (α, σ, τ) in Algorithm 1 to ensure (26) holds (see Theorems 2.5, 2.7 and 2.8).

Assumption (A): $(\alpha, \sigma, \tau) \in [0, 1[\times[0, 1[\times]0, 1]$ and $\{\alpha_k\}$ satisfy the following (for some $\beta > 0$):

$$0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < \beta < 1 \qquad \forall k \geq 1 \tag{29}$$

and

$$\tau = \tau(\sigma, \beta) := \frac{2(\beta' - 1)^2}{(1+\sigma)\left[2(\beta'-1)^2 + 3\beta' - 1\right]}, \tag{30}$$

where

$$\beta' := \max\left\{\beta, \frac{2(1-\sigma)}{3 - \sigma + \sqrt{9 + 2\sigma - 7\sigma^2}}\right\} \in \left[\frac{2(1-\sigma)}{3 - \sigma + \sqrt{9 + 2\sigma - 7\sigma^2}}, 1\right[. \tag{31}$$

Remarks.

(i) Conditions (29)–(31) will be crucial to prove convergence and iteration-complexity of the algorithms presented and studied in this paper; see, e.g., Theorems 2.5, 2.7 and 2.8, and Section 3.

(ii) Note that by letting $\sigma = 0$, which by the first remark following Algorithm 1 means that it reduces to an under-relaxed version of the (exact) Alvarez–Attouch's inertial PP method, we obtain that (29)–(31) are now simply given by: $0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < \beta < 1$, for all $k \geq 1$, and

$$\tau = \tau(\beta) := \frac{2(\beta' - 1)^2}{2(\beta'-1)^2 + 3\beta' - 1}, \qquad \beta' := \max\{\beta, 1/3\} \in [1/3, 1[\,. \tag{32}$$

In particular, in this case, we have $\tau = \tau(0, 1/3) = 1$ whenever $\beta = 1/3$ in (29), which corresponds to the standard upper bound on $\{\alpha_k\}$ which has been used in different works in the current literature (see Subsection 1.2 for a discussion). Hence, even in the setting of *exact* inertial PP methods, conditions (29)–(31) generalize the usual assumption (5). See Figure 1.

(iii) As we mentioned earlier, an inertial HPE-type method was proposed and studied by Bot and Csetnek in [9], where asymptotic convergence is proved under the assumption $\alpha(5+4\sigma^2)+\sigma^2 < 1$ on $\alpha, \sigma \in [0, 1[$. Note that, in this case, $\alpha \approx 0$ whenever $\sigma \approx 1$. This contrasts to the conditions (29)–(31), which, in particular yield $\tau = \tau(\sigma, 1/3) = 1/(1+\sigma) > 0.5$ (uniformly on σ) when $\beta = 1/3$ in (29). This may become especially useful in numerical implementations of Algorithm 1, since $\sigma = 0.99$ has been usually employed in the recent literature on HPE-type methods (see, e.g., [17, 18, 26, 27]). Further, (29)–(31) allow the upper bound α on $\{\alpha_k\}$ to be chosen arbitrarily close to 1, at the price of performing under-relaxed steps with the explicitly computed $\tau = \tau(\sigma, \beta)$ as in (30). See Figure 1.

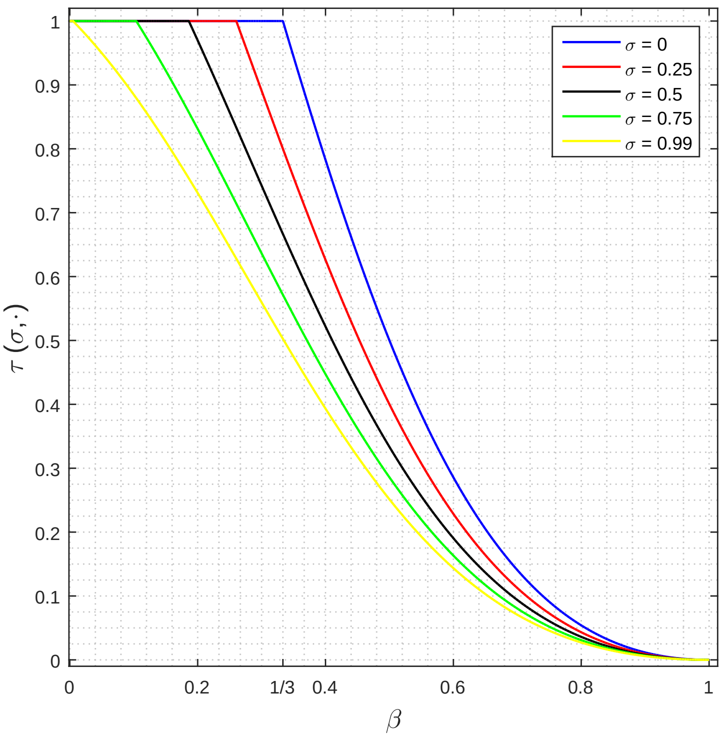

Figure 1: Function $]0,1[\ni \beta \mapsto \tau(\sigma,\beta) \in]0,1[$ as in (30) for $\sigma \in \{0, 0.25, 0.5, 0.75.0.99\}$. Note that $\tau(\sigma, 1/3) \geq 0.5$ for all $\sigma \in [0,1[$. See the second and third remarks following Assumption **(A)**.

Theorem 2.5 (second result on the weak convergence of Algorithm 1)**.** *Under the* Assumption $(\mathbf{A})$ *on* Algorithm 1*, let $\eta > 0$ be as in* (19) *and define the quadratic real function:*

$$q(\alpha') := (\eta - 1)\alpha'^{\,2} - (1 + 2\eta)\alpha' + \eta \quad \forall \alpha' \in \mathbb{R}. \tag{33}$$

Then, $q(\alpha) > 0$ and, for every $z^* \in T^{-1}(0)$,

$$\sum_{j=1}^{k} \|z_j - z_{j-1}\|^2 \leq \frac{2\,\|z_0 - z^*\|^2}{(1-\alpha)q(\alpha)} \qquad \forall k \geq 1. \tag{34}$$

As a consequence, it follows that under the assumption $(\mathbf{A})$ *the sequence $\{z_k\}$ generated by* Algorithm 1 *converges weakly to a solution of the monotone inclusion problem* (13) *whenever $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$.*

Proof. Using (15), the Cauchy-Schwarz inequality and the Young inequality $2ab \leq a^2 + b^2$ with $a := \|z_k - z_{k-1}\|$ and $b := \|z_{k-1} - z_{k-2}\|$ we find

$$\begin{aligned}
\|z_k - w_{k-1}\|^2 &= \|z_k - z_{k-1}\|^2 + \alpha_{k-1}^2\|z_{k-1} - z_{k-2}\|^2 - 2\alpha_{k-1}\langle z_k - z_{k-1}, z_{k-1} - z_{k-2}\rangle \\
&\geq \|z_k - z_{k-1}\|^2 + \alpha_{k-1}^2\|z_{k-1} - z_{k-2}\|^2 - \alpha_{k-1}\left(2\|z_k - z_{k-1}\|\|z_{k-1} - z_{k-2}\|\right) \\
&\geq (1 - \alpha_{k-1})\|z_k - z_{k-1}\|^2 - \alpha_{k-1}(1 - \alpha_{k-1})\|z_{k-1} - z_{k-2}\|^2,
\end{aligned}$$

which combined with (25), and after some algebraic manipulations, yields

$$\varphi_k - \varphi_{k-1} - \alpha_{k-1}(\varphi_{k-1} - \varphi_{k-2}) - \gamma_{k-1}\|z_{k-1} - z_{k-2}\|^2 \leq -\eta(1 - \alpha_{k-1})\|z_k - z_{k-1}\|^2 \quad \forall k \geq 1, \tag{35}$$

where

$$\gamma_k := (1-\eta)\alpha_k^2 + (1+\eta)\alpha_k \qquad \forall k \geq 0. \tag{36}$$

Define,

$$\mu_0 := (1-\alpha_0)\varphi_0 \geq 0, \qquad \mu_k := \varphi_k - \alpha_{k-1}\varphi_{k-1} + \gamma_k \|z_k - z_{k-1}\|^2 \quad \forall k \geq 1, \tag{37}$$

where φ_k is as in (24). Using (33), the assumption that $\{\alpha_k\}$ is nondecreasing (see (29)) and (35)–(37) we obtain, for all $k \geq 1$,

$$\begin{aligned}
\mu_k - \mu_{k-1} &\leq \left[\varphi_k - \varphi_{k-1} - \alpha_{k-1}(\varphi_{k-1} - \varphi_{k-2}) - \gamma_{k-1}\|z_{k-1} - z_{k-2}\|^2\right] + \gamma_k \|z_k - z_{k-1}\|^2 \\
&\leq \left[\gamma_k - \eta(1-\alpha_k)\right] \|z_k - z_{k-1}\|^2 \\
&= -\left[(\eta - 1)\alpha_k^2 - (1+2\eta)\alpha_k + \eta\right] \|z_k - z_{k-1}\|^2 \\
&= -q(\alpha_k)\|z_k - z_{k-1}\|^2.
\end{aligned} \tag{38}$$

Note now that from (30) and Lemma A.2 we have

$$\beta' = \frac{4 - 2(1+\sigma)\tau}{4 - (1+\sigma)\tau + \sqrt{(1+\sigma)\tau\left[16 - 7(1+\sigma)\tau\right]}},$$

which, in turn, combined with the definition of $\eta > 0$ in (19), and after some algebraic calculations, gives

$$\beta' = \frac{2\eta}{2\eta + 1 + \sqrt{8\eta + 1}}.$$

The latter identity implies, in particular, that β' is either the smallest or the largest root of the quadratic function $q(\cdot)$. Hence, from (29) and the fact that $\beta' \geq \beta$ (see (31)) we obtain

$$q(\alpha_k) \geq q(\alpha) > q(\beta') = 0.$$

The above inequalities combined with (38) yield

$$\|z_k - z_{k-1}\|^2 \leq \frac{1}{q(\alpha)}(\mu_{k-1} - \mu_k), \quad \forall k \geq 1, \tag{39}$$

which, in turn, combined with (29) and the definition of μ_k in (37), gives

$$\begin{aligned}
\sum_{j=1}^{k} \|z_j - z_{j-1}\|^2 &\leq \frac{1}{q(\alpha)}(\mu_0 - \mu_k), \\
&\leq \frac{1}{q(\alpha)}(\mu_0 + \alpha\varphi_{k-1}) \quad \forall k \geq 1.
\end{aligned} \tag{40}$$

Note now that (39), (29) and (37) also yield

$$\begin{aligned}
\mu_0 \geq \ldots \geq \mu_k =& \varphi_k - \alpha_{k-1}\varphi_{k-1} + \gamma_k \|z_k - z_{k-1}\|^2 \\
\geq& \varphi_k - \alpha\varphi_{k-1}, \quad \forall k \geq 1,
\end{aligned}$$

and so,

$$\varphi_k \leq \alpha^k \varphi_0 + \frac{\mu_0}{1-\alpha} \leq \varphi_0 + \frac{\mu_0}{1-\alpha} \qquad \forall k \geq 0. \tag{41}$$

Hence, (34) follows directly from (40), (41), the definition of μ_0 in (37) and the definition of φ_0 in (24). On the other hand, the second statement of the theorem follows from (34) and Theorem 2.4 (recall that $\alpha_k \leq \alpha < 1$ for all $k \geq 0$). □

Remark. A quadratic function similar to $q(\cdot)$, as defined in (33), was also considered by Alvarez in [1]. As we mentioned in the second remark following Algorithm 1, the algorithm studied in the later reference is different of the corresponding algorithm presented in this work, namely Algorithm 1. Moreover, note that if $\eta = 1$, then $q(\alpha') = 1 - 3\alpha'$ (cf. [2]).

Corollary 2.6. *Under the* Assumption (**A**) *on* Algorithm 1, *let* $\eta > 0$ *and* $q(\cdot)$ *be as in* (19) *and* (33)*, respectively, and let* $z^* \in T^{-1}(0)$*. Then, for all* $k \geq 1$,

$$\|z_k - z^*\|^2 + \sum_{j=1}^{k} \tau\Big(\max\left\{\eta\tau\|\lambda_j v_j\|^2, (1-\sigma^2)\|\tilde{z}_j - w_{j-1}\|^2\right\} \Big) \leq \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right) \|z_0 - z^*\|^2.$$

Proof. Using Proposition 2.3 and Lemma A.4(a) we conclude that (75) holds with $\{s_k\}$, $\{\varphi_k\}$ and $\{\delta_k\}$ as in (18) and (24), which gives that the desired result follows from (75) and (34). □

Next we present the first result on nonasymptotic global convergence rates/iteration-complexity of Algorithm 1.

Theorem 2.7 (global $\mathcal{O}(1/\sqrt{k})$ pointwise convergence rate of Algorithm 1)**.** *Under the* Assumption (**A**) *on* Algorithm 1*, let* $\eta > 0$ *and* $q(\cdot)$ *be as in* (19) *and* (33)*, respectively, and let* d_0 *denote the distance of* z_0 *to* $T^{-1}(0)$*. Assume that* $\lambda_k \geq \underline{\lambda} > 0$ *for all* $k \geq 1$*. Then, for every* $k \geq 1$*, there exists* $i \in \{1, \ldots, k\}$ *such that*

$$v_i \in T^{\varepsilon_i}(\tilde{z}_i), \tag{42}$$

$$\|v_i\| \leq \frac{d_0}{\underline{\lambda}\tau\sqrt{k}} \sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}, \tag{43}$$

$$\varepsilon_i \leq \frac{\sigma d_0^2}{2(1-\sigma^2)\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right). \tag{44}$$

Proof. Let $z^* \in T^{-1}(0)$ be such that $d_0 = \|z_0 - z^*\|$. It follows from Corollary 2.6 that, for every $k \geq 1$, there exists $i \in \{1, \ldots, k\}$ such that

$$\tau\, k\Big(\max\left\{\eta\tau\|\lambda_i v_i\|^2, (1-\sigma^2)\|\tilde{z}_i - w_{i-1}\|^2\right\} \Big) \leq \left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right) d_0^2,$$

which combined with the assumption $\lambda_i \geq \underline{\lambda} > 0$ and (16), and after some simple algebraic manipulations, yields the desired result. □

Remarks.

(i) Theorem 2.7 provides a global $\mathcal{O}(1/\sqrt{k})$ *pontwise* convergence rate and ensures, in particular, that for given tolerances $\rho, \epsilon > 0$, Algorithm 1 finds a triple (z, v, ε) satisfying (14) after performing at most

$$\mathcal{O}\left(\max\left\{\left\lceil \frac{d_0^2}{\underline{\lambda}^2\rho^2}\right\rceil, \left\lceil \frac{d_0^2}{\underline{\lambda}\epsilon}\right\rceil\right\}\right)$$

iterations.

(ii) If $\alpha = 0$ and $\tau = 1$, in which case Algorithm 1 reduces to the HPE method of Solodov and Svaiter, then it follows that Theorem 2.7 reduces to [30, Theorem 4.4(a)].

(iii) Analogous global $\mathcal{O}(1/\sqrt{k})$ *pontwise* convergence rates were also obtained in [15, 16] for inertial-type algorithms for variational inequality and convex optimization problems.

In order to study the *ergodic* iteration-complexity of Algorithm 1, we need to define the following.

The *aggregate stepsize sequence* $\{\Lambda_k\}$ and the *ergodic sequences* $\{\tilde{z}_k^a\}$, $\{\tilde{v}_k^a\}$, $\{\varepsilon_k^a\}$ associated to $\{\lambda_k\}$ and $\{\tilde{z}_k\}$, $\{v_k\}$, and $\{\varepsilon_k\}$ are, respectively, for $k \geq 1$,

$$\begin{aligned}
&\Lambda_k := \sum_{j=1}^{k} \lambda_j\,,\\
&\tilde{z}_k^{\,a} := \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\,\tilde{z}_j, \quad v_k^{\,a} := \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\,v_j,\\
&\varepsilon_k^{\,a} := \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^{\,a}, v_j - v_k^{\,a}\rangle) = \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^{\,a}, v_j\rangle).
\end{aligned} \tag{45}$$

Next we study the *ergodic* iteration-complexity of Algorithm 1 under the assumption that $\alpha_k \equiv \alpha$ in (15).

Theorem 2.8 (global $\mathcal{O}(1/k)$ ergodic convergence rate of Algorithm 1)**.** *Under the* Assumption (**A**) *on* Algorithm 1 *and, additionally, the assumption that $\alpha_k \equiv \alpha$, let $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ be as in* (45) *and let d_0 denote the distance of z_0 to $T^{-1}(0)$. Let also $\eta > 0$ and $q(\cdot)$ be as in* (19) *and* (33), *respectively, and assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$.*

Then, for all $k \geq 1$,

$$v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \tag{46}$$

$$\|v_k^a\| \leq \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k}\sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \tag{47}$$

$$\varepsilon_k^a \leq \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1 + \frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta\tau^2}}\right). \tag{48}$$

Proof. Let $z^* \in T^{-1}(0)$ be such that $d_0 = \|z_0 - z^*\|$. Using Algorithm 1's definition and Lemma A.1(a) with $z = \tilde{z}_k^a$ we find, for all $j \geq 1$,

$$\|w_{j-1} - \tilde{z}_k^a\|^2 - \|z_j - \tilde{z}_k^a\|^2 \geq 2\tau\lambda_j\left(\varepsilon_j + \langle \tilde{z}_j - \tilde{z}_k^a, v_j\rangle\right). \tag{49}$$

On the other hand, Lemma 2.2 yields

$$\|w_{j-1}-\tilde{z}_k^a\|^2=(1+\alpha_{j-1})\|z_{j-1}-\tilde{z}_k^a\|^2-\alpha_{j-1}\|z_{j-2}-\tilde{z}_k^a\|^2+\alpha_{j-1}(1+\alpha_{j-1})\|z_{j-1}-z_{j-2}\|^2,$$

which, in turn, combined with (49) gives, for all $j\geq 1$,

$$\|z_j-\tilde{z}_k^a\|^2-\|z_{j-1}-\tilde{z}_k^a\|^2+2\tau\lambda_j\left(\varepsilon_j+\langle\tilde{z}_j-\tilde{z}_k^a,v_j\rangle\right)\leq\alpha_{j-1}\left(\|z_{j-1}-\tilde{z}_k^a\|^2-\|z_{j-2}-\tilde{z}_k^a\|^2\right)+\delta_j,$$

where the sequence $\{\delta_j\}$ is as in (24). Summing the latter inequality over all $j=1,\dots,k$ and using (45) as well as the assumption $\alpha_k\equiv\alpha$, we obtain

$$\|z_k-\tilde{z}_k^a\|^2-\|z_0-\tilde{z}_k^a\|^2+2\tau\Lambda_k\varepsilon_k^a\leq\alpha\left(\|z_{k-1}-\tilde{z}_k^a\|^2-\|z_{-1}-\tilde{z}_k^a\|^2\right)+\sum_{j=1}^k\delta_j,$$

which combined with the definition of $\{\delta_j\}$ and (34) yields (recall that $z_0=z_{-1}$)

$$\begin{aligned}2\tau\Lambda_k\varepsilon_k^a-\frac{2\alpha(1+\alpha)d_0^2}{(1-\alpha)q(\alpha)}&\leq(1-\alpha)\left(\|z_0-\tilde{z}_k^a\|^2-\|z_k-\tilde{z}_k^a\|^2\right)\\&\quad+\alpha\left(\|z_{k-1}-\tilde{z}_k^a\|^2-\|z_k-\tilde{z}_k^a\|^2\right)\\&\leq 2\max\left\{\|z_0-\tilde{z}_k^a\|\|z_0-z_k\|,\|z_{k-1}-\tilde{z}_k^a\|\|z_{k-1}-z_k\|\right\},\end{aligned}\tag{50}$$

where we have also used the inequality $\|a\|^2-\|b\|^2\leq 2\|a\|\|a-b\|$ for all $a,b\in\mathcal{H}$. Now, define

$$(\forall j\geq 1)\quad \hat{z}_j:=w_{j-1}-\lambda_j v_j\quad\text{and}\quad\hat{z}_k^a:=\frac{1}{\Lambda_k}\sum_{j=1}^k\lambda_j\hat{z}_j.\tag{51}$$

From Corollary 2.6, the first definition in (51), (45), (17) and the convexity of $\|\cdot\|^2$ we find

$$\|z_\ell-z_j\|\leq\|z_\ell-z^*\|+\|z_j-z^*\|\leq 2d_0\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}}\qquad\forall\ell,j\geq 0,\tag{52}$$

$$(1-\tau)^{-2}\sum_{j=1}^k\|z_j-\hat{z}_j\|^2=\sum_{j=1}^k\|\lambda_j v_j\|^2\leq\frac{d_0^2}{\eta\tau^2}\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right)\tag{53}$$

and

$$\begin{aligned}\|\tilde{z}_k^a-\hat{z}_k^a\|^2\leq\frac{1}{\Lambda_k}\sum_{j=1}^k\lambda_j\|\tilde{z}_j-\hat{z}_j\|^2&\leq\sum_{j=1}^k\|\lambda_j v_j+\tilde{z}_j-w_{j-1}\|^2\\&\leq\sigma^2\sum_{j=1}^k\|\tilde{z}_j-w_{j-1}\|^2\\&\leq\frac{\sigma^2d_0^2}{(1-\sigma^2)\tau}\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2q(\alpha)}\right).\end{aligned}\tag{54}$$

From (52), (53), the convexity of $\|\cdot\|^2$ and the inequality $\|a-b\|^2 \leq 2\left(\|a\|^2+\|b\|^2\right)$ (for all $a, b \in \mathcal{H}$), we find

$$
\begin{aligned}
\|z_\ell - \hat{z}_k^a\|^2 &\leq \frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\|z_\ell - \hat{z}_j\|^2 \\
&\leq 2\left(\frac{1}{\Lambda_k}\sum_{j=1}^{k}\lambda_j\|z_\ell - z_j\|^2 + \sum_{j=1}^{k}\|z_j - \hat{z}_j\|^2\right) \\
&\leq 2d_0^2\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(4+\frac{(1-\tau)^2}{\eta\tau^2}\right) \qquad \forall \ell \geq 0.
\end{aligned}
$$

Using the above inequality and (54) we obtain, for all $\ell \geq 0$,

$$
\begin{aligned}
\|z_\ell - \tilde{z}_k^a\| &\leq \|z_\ell - \hat{z}_k^a\| + \|\tilde{z}_k^a - \hat{z}_k^a\| \\
&\leq \sqrt{2}d_0\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}\left(\frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4+\frac{(1-\tau)^2}{\eta\tau^2}}\right).
\end{aligned} \tag{55}
$$

Hence, (50), (52) with $\ell = 0, k-1$ and $j = k$, and (55) with $\ell = 0, k-1$ yield

$$
\begin{aligned}
2\tau\Lambda_k\varepsilon_k^a &\leq 4\sqrt{2}d_0^2\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(\frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4+\frac{(1-\tau)^2}{\eta\tau^2}}\right) \\
&\quad + \frac{2\alpha(1+\alpha)d_0^2}{(1-\alpha)q(\alpha)} \\
&\leq 4\sqrt{2}d_0^2\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1+\frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4+\frac{(1-\tau)^2}{\eta\tau^2}}\right),
\end{aligned}
$$

which, combined with the assumption $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$, clearly finishes the proof of (48).

Now note that using (17), (15) and the assumption $\alpha_k \equiv \alpha$ we find

$$
\tau\lambda_j v_j = z_{j-1} - z_j + \alpha(z_{j-1} - z_{j-2}) \qquad \forall j \geq 1.
$$

Summing the above identity over $j = 1, \ldots, k$ and using (45) and (52) with $\ell = 0$ and $j = k-1, k$ we find (recall that $z_0 = z_{-1}$)

$$
\begin{aligned}
\tau\Lambda_k\|v_k^a\| &\leq \|z_0 - z_k\| + \alpha\|z_0 - z_{k-1}\| \\
&\leq 2(1+\alpha)d_0\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}},
\end{aligned}
$$

which, combined with the assumption $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$, yields (47). To finish the proof of the theorem, note that (46) is a direct consequence of the inclusion in (16) and Theorem 1.2(a). $\square$

Remark. We mention that, up to the authors knowledge, this is the first time in the literature that $\mathcal{O}(1/k)$ global convergence rates are established for inertial PP-type algorithms.

2.1 On the under-relaxed inertial proximal point method

In this subsection, we analyze the convergence and iteration-complexity of the under-relaxed inertial proximal point (PP) method (see, e.g., [1, 4]) with constant under-relaxation (Algorithm 2) for solving (13). The analysis is performed by viewing Algorithm 2 within the framework of Algorithm 1, for which asymptotic convergence and iteration-complexity were obtained in Theorems 2.5, 2.7 and 2.8.

Algorithm 2. Under-relaxed inertial proximal point method for solving (13)

Input: $z_0 = z_{-1} \in \mathcal{H}$ and $0 \le \alpha < 1$ and $0 < \tau \le 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}). \tag{56}$$

3: Compute

$$\tilde{z}_k = (\lambda_k T + I)^{-1} w_{k-1}. \tag{57}$$

4: Define

$$z_k := \tau \tilde{z}_k + (1-\tau) w_{k-1}. \tag{58}$$

Proposition 2.9. Algorithm 2 *is a special instance of* Algorithm 1 *with* $\sigma = 0$ *in the* Input*, in which case* $\varepsilon_k = 0$ *and* $v_k = (w_{k-1} - \tilde{z}_k)/\lambda_k \in T(\tilde{z}_k)$ *for all* $k \ge 1$.

Proof. The proof follows from the well-known fact that $\tilde{z} = (\lambda T + I)^{-1} w$ if and only if $v := (w - \tilde{z})/\lambda \in T(\tilde{z})$ and Algorithms 2 and 1's definitions. □

Theorem 2.10 (convergence and iteration-complexity of Algorithm 2)**.** *Under the* Assumption **(A)** *with* $\sigma = 0$ *on* Algorithm 2*, let* $\{z_k\}$, $\{v_k\}$, $\{\tilde{z}_k\}$ *and* $\{\lambda_k\}$ *be generated by* Algorithm 2 *and let the ergodic sequences* $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ *and* $\{\varepsilon_k^a\}$ *be as in* (45)*. Let also* $q(\cdot)$ *be as in* (33) *and let* d_0 *denote the distance of* z_0 *to* $T^{-1}(0)$*. Assume that* $\lambda_k \ge \underline{\lambda} > 0$ *for all* $k \ge 1$*. Then, the following statements hold:*

(a) *The sequence* $\{z_k\}$ *converges weakly to a solution of the monotone inclusion problem* (13)*.*

(b) *For all* $k \ge 1$*, there exists* $i \in \{1, \ldots, k\}$ *such that*

$$v_i \in T(\tilde{z}_i), \qquad \|v_i\| \le \frac{d_0}{\underline{\lambda}\tau\sqrt{k}} \sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}. \tag{59}$$

(c) *If, additionally, $\alpha_k \equiv \alpha$, then, for all $k \geq 1$,*

$$v_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a), \tag{60}$$

$$\|v_k^a\| \leq \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k}\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \tag{61}$$

$$\varepsilon_k^a \leq \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1+\sqrt{4+\frac{(1-\tau)^2}{\tau(2-\tau)}}\right). \tag{62}$$

Proof. The results in (a), (b) and (c) follow directly from Proposition 2.9 and Theorems 2.5, 2.7 and 2.8. □

3 Inertial under-relaxed forward-backward and Tseng's modified forward-backward methods

Consider the structured monotone inclusion problem (2), i.e., the problem of finding $z \in \mathcal{H}$ such that

$$0 \in F(z) + B(z) =: T(z) \tag{63}$$

where $F : D(F) \subset \mathcal{H} \to \mathcal{H}$ is point-to-point monotone and $B : \mathcal{H} \rightrightarrows \mathcal{H}$ is a (set-valued) maximal monotone operator for which $T^{-1}(0) \neq \emptyset$ (precise assumption on F and B will be stated later).

In this section, we study the convergence and iteration-complexity of inertial (under-relaxed) versions of the forward-backward and Tseng's modified forward-backward methods (9) and (10), respectively, for solving (63), by viewing them within the framework of Algorithm 1, for which asymptotic convergence and iteration-complexity were studied in Section 2.

3.1 An inertial under-relaxed Tseng's modified forward-backward method

In this subsection, we consider the monotone inclusion problem (63) where the following assumptions are assumed to hold:

(C1) $F : D(F) \subset \mathcal{H} \to \mathcal{H}$ is monotone and L-Lipschitz continuous on a (nonempty) closed convex set Ω such that $D(B) \subset \Omega \subset D(F)$, i.e., F is monotone on Ω and there exists $L \geq 0$ such that

$$\|F(z) - F(z')\| \leq L\|z - z'\| \qquad \forall z, z' \in \Omega.$$

(C2) B is a (set-valued) maximal monotone operators on $\mathcal{H}$.

(C3) The solution set of (63) is nonempty.

We mention that it was proved in [29, Proposition A.1] that under assumptions $(C1)$–$(C3)$ the operator $T(\cdot)$ defined in (63) is maximal monotone, which guarantee that (63) is a special instance of (13). In particular, it follows that Algorithm 1 can be used to solving the structured monotone inclusion (63).

As we mentioned earlier, in this subsection, we shall study the convergence and iteration-complexity of the following inertial under-relaxed version of the Tseng's modified forward-backward method for solving (63).

Algorithm 3. An inertial under-relaxed Tseng's modified forward-backward method for solving (63)

Input: $z_0 = z_{-1} \in \mathcal{H}$, $0 \leq \alpha < 1$, $0 < \sigma < 1$ and $0 < \tau \leq 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}).$$

3: Choose $\lambda_k \in]0, \sigma/L]$, let $w'_{k-1} = P_\Omega(w_{k-1})$ and compute

$$\tilde{z}_k = (\lambda_k B + I)^{-1}(w_{k-1} - \lambda_k F(w'_{k-1})),$$
$$\hat{z}_k = \tilde{z}_k - \lambda_k \left(F(\tilde{z}_k) - F(w'_{k-1})\right).$$

4: Define

$$z_k := (1 - \tau)w_{k-1} + \tau \hat{z}_k.$$

Remarks.

(i) Algorithm 3 reduces to the Tseng's modified forward-backward method [41] for solving (63) if $\alpha = 0$ and $\tau = 1$, in which case $w_{k-1} = z_{k-1}$ and $z_k = \hat{z}_k$.

(ii) An inertial Tseng's modified forward-backward-type method (based on a different mechanism of iteration) was proposed and studied in [9]. The proposed Tseng's modified forward-backward type method in the latter reference tends to suffer from similar limitations as the inertial HPE-type method proposed in [9], as we discussed in the third remark following Assumption (**A**). Moreover, in contrast to this paper which performs the iteration-complexity analysis of Algorithm 3 (see Theorem 3.2), [9] has focused on asymptotic convergence.

Since the proof of the next proposition follows the same outline of [30, Proposition 6.1], we omit it here.

Proposition 3.1. *Let* $\{w_k\}$, $\{w'_k\}$, $\{z_k\}$, $\{\alpha_k\}$, $\{\tilde{z}_k\}$ *and* $\{\lambda_k\}$ *be generated by* Algorithm 3 *and define*

$$\varepsilon_k := 0 \;\; \text{and} \;\; v_k := F(\tilde{z}_k) - F(w'_{k-1}) + \frac{1}{\lambda_k}(w_{k-1} - \tilde{z}_k) \qquad \forall k \geq 1. \tag{64}$$

Then, the sequences $\{w_k\}$, $\{z_k\}$, $\{\alpha_k\}$, $\{\tilde{z}_k\}$, $\{v_k\}$, $\{\varepsilon_k\}$ *and* $\{\lambda_k\}$ *satisfy the conditions* (15)–(17) *in* Algorithm 1. *As a consequence, it follows that* Algorithm 3 *is a special instance of* Algorithm 1 *for solving* (63).

Next we present the convergence and iteration-complexity of Algorithm 3 under the Assumption (**A**) on the Input $(\alpha, \sigma, \tau) \in [0, 1[\times]0, 1[\times]0, 1]$ and on the sequence $\{\alpha_k\}$. We also mention that the observations regarding the parameter τ in the third remark following Assumption (**A**) obviously apply to Algorithm 3.

Theorem 3.2 (convergence and iteration-complexity of Algorithm 3)**.** *Under the* Assumption (**A**) *on* $(\alpha, \sigma, \tau) \in [0, 1[\times]0, 1[\times]0, 1]$ *and* $\{\alpha_k\}$*, let* $\{z_k\}$*,* $\{\tilde{z}_k\}$ *and* $\{\lambda_k\}$ *be generated by* Algorithm 3*, let* $\{v_k\}$ *and* $\{\varepsilon_k\}$ *be as in* (64) *and let the ergodic sequences* $\{\tilde{z}_k^a\}$*,* $\{v_k^a\}$ *and* $\{\varepsilon_k^a\}$ *be as in* (45)*. Let also* $\eta > 0$ *and* $q(\cdot)$ *be as in* (19) *and* (33)*, respectively, let* d_0 *denote the distance of* z_0 *to* $(F + B)^{-1}(0)$ *and assume that* $\lambda_k \geq \underline{\lambda} > 0$ *for all* $k \geq 1$*. Then, the following statements hold:*

(a) *The sequence* $\{z_k\}$ *converges weakly to a solution of the monotone inclusion problem* (63)*.*

(b) *For all* $k \geq 1$*, there exists* $i \in \{1, \dots, k\}$ *such that*

$$v_i \in (F + B)(\tilde{z}_i), \qquad \|v_i\| \leq \frac{d_0}{\underline{\lambda}\tau\sqrt{k}} \sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}. \tag{65}$$

(c) *If, additionally,* $\alpha_k \equiv \alpha$*, then, for all* $k \geq 1$*,*

$$\begin{aligned} & v_k^a \in (F + B)^{\varepsilon_k^a}(\tilde{z}_k^a), \\ & \|v_k^a\| \leq \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k} \sqrt{1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \\ & \varepsilon_k^a \leq \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1 + \frac{\sigma}{\sqrt{(1-\sigma^2)\tau}} + \sqrt{4 + \frac{(1-\tau)^2}{\eta\tau^2}}\right). \end{aligned} \tag{66}$$

Proof. The proof follows directly from Proposition 3.1 and Theorems 2.5, 2.7 and 2.8. $\square$

Remarks.

(i) Itens (b) and (c) ensure, respectively, global pointwise $\mathcal{O}(1/\sqrt{k})$ and ergodic $\mathcal{O}(1/k)$ convergence rates for Algorithm 3. On the other hand, note that the inclusion in (66) is potentially weaker than the corresponding one in (65).

(ii) If $\lambda_k \equiv \sigma/L$ in Step 3 of Algorithm 3, in which case $\underline{\lambda} = \sigma/L$, then $d_0/\underline{\lambda}$ in (65) and (66) can be replaced by $d_0 L/\sigma$. In this case, Item (b) gives that for a given tolerance $\rho > 0$, Algorithm 3 finds a pair (z, v) such that (cf. (14))

$$v \in (F + B)(z), \quad \|v\| \leq \rho$$

in at most

$$\mathcal{O}\left(\left\lceil \frac{d_0^2 L^2}{\rho^2} \right\rceil\right)$$

iterations, an analogous remark also holding for Item (c).

3.2 On the inertial under-relaxed forward-backward method

Similarly to Subsection 3.1, in this subsection, we consider the monotone inclusion problem (63) but now assume the following: (C2) and (C3) as in Subsection 3.1 and instead of (C1):

(C1′) $F : \mathcal{H} \to \mathcal{H}$ is $(1/L)-$cocoercive, i.e., there exists $L > 0$ such that

$$\langle z - z', F(z) - F(z')\rangle \geq \frac{1}{L}\|F(z) - F(z')\|^2 \qquad \forall z, z' \in \mathcal{H}. \tag{67}$$

We observe that it follows from (67) that F is, in particular, L–Lipschitz continuous.

Algorithm 4. Inertial under-relaxed forward-backward method for solving (63)

Input: $z_0 = z_{-1} \in \mathcal{H}$, $0 \leq \alpha < 1$, $0 < \sigma < 1$ and $0 < \tau \leq 1$.

1: **for** $k = 1, 2, \ldots,$ **do**

2: Choose $\alpha_{k-1} \in [0, \alpha]$ and define

$$w_{k-1} := z_{k-1} + \alpha_{k-1}(z_{k-1} - z_{k-2}).$$

3: Choose $\lambda_k \in]0, 2\sigma^2/L]$ and compute

$$\tilde{z}_k = (\lambda_k B + I)^{-1}(w_{k-1} - \lambda_k F(w_{k-1})).$$

4: Define

$$z_k := (1 - \tau)w_{k-1} + \tau\tilde{z}_k.$$

Remarks.

(i) If $\alpha = 0$ and $\tau = 1$, then it follows that Algorithm 4 reduces to the forward-backward [22, 33] method for solving (63).

(ii) Inertial versions of the forward-backward method were previously proposed and studied in [32], [23] and [3]. Asymptotic convergence of the forward-backward method proposed in [23] was proved in the latter reference, in particular, under the assumption: $0 \leq \alpha_{k-1} \leq \alpha_k \leq \alpha < 1$, for all $k \geq 1$, and

$$\alpha = \alpha(\gamma) := 1 + \frac{\sqrt{9 - 4\gamma - 2\varepsilon\gamma} - 3}{\gamma},$$

for some $\varepsilon \in]0, (9 - 4\gamma)/(2\gamma)[$, where $\gamma \in (0, 2)$ and $\lambda_k \equiv \lambda := \gamma/L$ ($\gamma = 2\sigma^2$ in the notation of the present paper). The apparent limitation of this approach is that $\alpha \to 0$ if $\gamma \to 2$, i.e., the inertial effect degenerates for large values of the stepsize (see Fig. 1 in [23]). This contrasts to the approach proposed in this paper, where the under-relaxation parameter $\tau \in [0, 1[$ is crucial to allowing α sufficiently close to 1, even for large stepsize values, i.e., when $\sigma \approx 1$ (see Assumption **(A)** and part of the discussion in the third remark following it).

(iii) Algorithm 4 is a special instance (with constant relaxation) of the RIFB algorithm in [3]. We refer the reader to [3] (see, e.g., Theorems 3.8 and 3.15, and Remark 3.13) for a comprehensive discussion of the interplay and benefits of inertia and relaxation.

Next proposition shows that Algorithm 4 is also a special instance of Algorithm 1 for solving (63). Since the proof follows the same outline of [39, Proposition 5.3], we omit it here too.

Proposition 3.3. *Let $\{\tilde{z}_k\}$, $\{z_k\}$, $\{w_k\}$ and $\{\lambda_k\}$ be generated by* Algorithm 4*, let $T = F + B$ be as in* (63) *and define, for all $k \geq 1$,*

$$\varepsilon_k := \frac{\|\tilde{z}_k - w_{k-1}\|^2}{4L^{-1}} \quad \textit{and} \quad v_k := \frac{w_{k-1} - \tilde{z}_k}{\lambda_k}. \tag{68}$$

Then, the following hold for all $k \geq 1$:

$$\begin{aligned} & v_k \in (F^{\varepsilon_k} + B)(\tilde{z}_k) \subset T^{\varepsilon_k}(\tilde{z}_k), \\ & \lambda_k v_k + \tilde{z}_k - w_{k-1} = 0, \quad 2\lambda_k \varepsilon_k \leq \sigma^2 \|\tilde{z}_k - w_{k-1}\|^2, \quad z_k = w_{k-1} - \tau\lambda_k v_k. \end{aligned} \tag{69}$$

As a consequence of (69) *and* Algorithm 4*'s definition, it follows that* Algorithm 4 *is a special instance of* Algorithm 1 *for solving* (63).

We finish this section by presenting the convergence and iteration-complexity of Algorithm 4, which are a direct consequence of Proposition 3.3 and Theorems 2.5, 2.7 and 2.8. We mention that analogous remarks to those made in the Remarks following Theorem 3.2 also apply here.

Theorem 3.4 (convergence and iteration-complexity of Algorithm 4)**.** *Under the* Assumption (**A**) *on $(\alpha, \sigma, \tau) \in [0, 1[\times]0, 1[\times]0, 1]$ and $\{\alpha_k\}$, let $\{z_k\}$, $\{\tilde{z}_k\}$ and $\{\lambda_k\}$ be generated by* Algorithm 4*, let $\{v_k\}$ and $\{\varepsilon_k\}$ be as in* (68) *and let the ergodic sequences $\{\tilde{z}_k^a\}$, $\{v_k^a\}$ and $\{\varepsilon_k^a\}$ be as in* (45)*. Let also $\eta > 0$ and $q(\cdot)$ be as in* (19) *and* (33)*, respectively, let d_0 denote the distance of z_0 to $(F + B)^{-1}(0)$ and assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$. Then, the following statements hold:*

(a) *The sequence $\{z_k\}$ converges weakly to a solution of the monotone inclusion problem* (63)*.*

(b) *For all $k \geq 1$, there exists $i \in \{1, \ldots, k\}$ such that*

$$\begin{aligned} & v_i \in (F^{\varepsilon_i} + B)(\tilde{z}_i), \\ & \|v_i\| \leq \frac{d_0}{\underline{\lambda}\tau\sqrt{k}} \sqrt{\eta^{-1}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)}, \\ & \varepsilon_i \leq \frac{\sigma d_0^2}{2(1-\sigma^2)\underline{\lambda}\tau\, k}\left(1 + \frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right). \end{aligned} \tag{70}$$

(c) *If, additionally, $\alpha_k \equiv \alpha$, then, for all $k \geq 1$,*

$$\begin{aligned} & v_k^a \in (F+B)^{\varepsilon_k^a}(\tilde{z}_k^a), \\ & \|v_k^a\| \leq \frac{2(1+\alpha)d_0}{\underline{\lambda}\tau\, k}\sqrt{1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}}, \\ & \varepsilon_k^a \leq \frac{2\sqrt{2}d_0^2}{\underline{\lambda}\tau\, k}\left(1+\frac{2\alpha(1+\alpha)}{(1-\alpha)^2 q(\alpha)}\right)\left(1+\frac{\sigma}{\sqrt{(1-\sigma^2)\tau}}+\sqrt{4+\frac{(1-\tau)^2}{\eta\tau^2}}\right). \end{aligned} \tag{71}$$

4 Concluding remarks

In this paper, we proposed and studied the asymptotic convergence and iteration-complexity of an inertial under-relaxed HPE-type method. As applications, we proposed and/or studied inertial (under-relaxed) versions of the Tseng's modified forward-backward and forward-backward methods for solving structured monotone inclusion problems with either Lipschitz continuous or cocoercive operators. All the proposed and/or studied algorithms, namely Algorithms 1, 2, 3 and 4 potentially benefit from a specific policy for choosing the upper bound on the sequence of extrapolation parameters, in which case (under) relaxation plays a central role (see Assumption (**A**) and Theorems 2.7, 2.8, 3.2 and 3.4); see also the recent work [3] of Attouch and Cabot. We also emphasize that, up to the authors knowledge, this is the first time in the literature that nonasymptotic global convergence rates (iteration-complexity) are provided for inertial HPE-type methods, in particular for the proposed inertial Tseng's modified forward-backward method.

A Auxiliary results

Next lemma was proved in [40, Lemma 2.1]. Here, we present a short and direct proof for the convenience of the reader.

Lemma A.1 (Svaiter)**.** *Let $\tilde{z}, v, w \in \mathcal{H}$ and $\lambda > 0$, $\varepsilon \geq 0$ and $\sigma \in [0,1[$ be such that*

$$v \in T^{\varepsilon}(\tilde{z}), \quad \|\lambda v + \tilde{z} - w\|^2 + 2\lambda\varepsilon \leq \sigma^2\|\tilde{z} - w\|^2. \tag{72}$$

Let $\tau \in [0,1]$ and define $z_+ := w - \tau\lambda v$. Then, the following hold:

(a) *For any $z \in \mathcal{H}$,*

$$\|w - z\|^2 - \|z_+ - z\|^2 \geq (1-\sigma)^2\tau\|\tilde{z} - w\|^2 + 2\tau\lambda\left(\varepsilon + \langle \tilde{z} - z, v\rangle\right) + \tau(1-\tau)\|\lambda v\|^2.$$

(b) *For any $z^* \in T^{-1}(0)$,*

$$\|w - z^*\|^2 - \|z_+ - z^*\|^2 \geq (1-\sigma^2)\tau\|\tilde{z} - w\|^2 + \tau(1-\tau)\|\lambda v\|^2.$$

Proof. (a) Using the inequality in (72) and some algebraic manipulations we find, for any $z \in \mathcal{H}$,

$$\begin{aligned} \|w - z\|^2 - \|(w - \lambda v) - z\|^2 &= \|\tilde{z} - w\|^2 - \|\lambda v + \tilde{z} - w\|^2 + 2\lambda\langle \tilde{z} - z, v\rangle \\ &\geq (1-\sigma^2)\|\tilde{z} - w\|^2 + 2\lambda\left(\varepsilon + \langle \tilde{z} - z, v\rangle\right). \end{aligned} \tag{73}$$

The fact that $z_+ = (1-\tau)w + \tau(w - \lambda v)$ and (12) yield

$$\begin{aligned}\|z_+ - z\|^2 &= (1-\tau)\|w - z\|^2 + \tau\|(w - \lambda v) - z\|^2 - \tau(1-\tau)\|\lambda v\|^2 \\ &= \|w - z\|^2 - \tau\left(\|w - z\|^2 - \|(w - \lambda v) - z\|^2\right) - \tau(1-\tau)\|\lambda v\|^2.\end{aligned}$$

Multiplying (73) by $\tau \in [0,1]$ and using the latter identity we obtain the desired inequality in (a).

(b) This is a direct consequence of Item (a), (11), the inclusion in (72) and the fact that $0 \in T(z^*)$. $\square$

Lemma A.2. *For any $\sigma \in [0,1[$, the inverse function of the scalar map*

$$A :=]0, 1+\sigma] \ni t \mapsto \frac{4-2t}{4-t+\sqrt{16t-7t^2}} \in \left[\frac{2(1-\sigma)}{3-\sigma+\sqrt{9+2\sigma-7\sigma^2}}, 1\right[=: B$$

is given by

$$B \ni \beta \mapsto \frac{2(\beta-1)^2}{2(\beta-1)^2 + 3\beta - 1} \in A.$$

Lemma A.3 (Opial)**.** *Let $\emptyset \neq \Omega \subset \mathcal{H}$ and $\{z_k\}$ be a sequence in $\mathcal{H}$ such that $\lim_{k\to\infty} \|z_k - z^*\|$ exist for every $z^* \in \Omega$. If every (sequential) weak cluster point of $\{z_k\}$ belongs to Ω, then $\{z_k\}$ converges weakly to a point in Ω.*

The following lemma was essentially proved by Alvarez and Attouch in [2, Theorem 2.1].

Lemma A.4. *Let the sequences $\{\varphi_k\}$, $\{s_k\}$, $\{\alpha_k\}$ and $\{\delta_k\}$ in $[0,+\infty[$ and $\alpha \in \mathbb{R}$ be such that $\varphi_0 = \varphi_{-1}$, $0 \le \alpha_{k-1} \le \alpha < 1$ and*

$$\varphi_k - \varphi_{k-1} + s_k \le \alpha_{k-1}(\varphi_{k-1} - \varphi_{k-2}) + \delta_k \qquad \forall k \ge 1. \tag{74}$$

The following hold:

(a) *For all $k \ge 1$,*

$$\varphi_k + \sum_{j=1}^{k} s_j \le \varphi_0 + \frac{1}{1-\alpha}\sum_{j=1}^{k} \delta_j. \tag{75}$$

(b) *If $\sum_{k=1}^{\infty} \delta_k < +\infty$, then $\lim_{k\to\infty} \varphi_k$ exist, i.e., the sequence $\{\varphi_k\}$ converges to some element in $[0,\infty[$.*

Proof. It was proved in [2, Theorem 2.1] that $\mathcal{M} := (1-\alpha)^{-1}\sum_{j=1}^{k} \delta_j \ge \sum_{j=1}^{k} [\varphi_j - \varphi_{j-1}]_+$, where $[\cdot]_+ = \max\{\cdot, 0\}$. Using this, the assumptions $\varphi_0 = \varphi_{-1}$, $0 \le \alpha_{k-1} \le \alpha < 1$ and (74), and some algebraic manipulations we find

$$\begin{aligned}\varphi_k + \sum_{j=1}^{k} s_j &\le \varphi_0 + \alpha\sum_{j=1}^{k-1}[\varphi_j - \varphi_{j-1}]_+ + \sum_{j=1}^{k} \delta_j \\ &\le \varphi_0 + \alpha\mathcal{M} + (1-\alpha)\mathcal{M} = \varphi_0 + \mathcal{M},\end{aligned}$$

which proves (a). To finish the proof of the lemma, note that (b) was proved inside the proof of [2, Theorem 2.1]. $\square$

References

[1] 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, 2003.

[2] 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).

[3] H. Attouch and A. Cabot. Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. Preprint hal-01708216, 2018.

[4] H. Attouch and A. Cabot. Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Preprint hal-01708905, 2018.

[5] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.*, 168(1-2, Ser. B):123–175, 2018.

[6] H. Attouch and J. Peypouquet. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. *SIAM J. Optim.*, 26(3):1824–1834, 2016.

[7] H. Attouch, J. Peypouquet, and P. Redont. Fast convex optimization via inertial dynamics with Hessian driven damping. *J. Differential Equations*, 261(10):5734–5783, 2016.

[8] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. With a foreword by Hédy Attouch.

[9] R. I. Boţ and E. R. Csetnek. A hybrid proximal-extragradient algorithm with inertial effects. *Numer. Funct. Anal. Optim.*, 36(8):951–963, 2015.

[10] R. I. Boţ, E. R. Csetnek, and C. Hendrich. Inertial Douglas-Rachford splitting for monotone inclusion problems. *Appl. Math. Comput.*, 256:472–487, 2015.

[11] R. S. Burachik, A. N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Anal.*, 5(2):159–180, 1997.

[12] R. S. Burachik, C. A. Sagastizábal, and B. F. Svaiter. ϵ-enlargements of maximal monotone operators: theory and applications. In *Reformulation: nonsmooth, piecewise smooth, semismooth and smoothing methods (Lausanne, 1997)*, volume 22 of *Appl. Optim.*, pages 25–43. Kluwer Acad. Publ., Dordrecht, 1999.

[13] R. S. Burachik and B. F. Svaiter. ϵ-enlargements of maximal monotone operators in Banach spaces. *Set-Valued Anal.*, 7(2):117–132, 1999.

[14] L. C. Ceng, B. S. Mordukhovich, and J. C. Yao. Hybrid approximate proximal method with auxiliary variational inequality for vector optimization. *J. Optim. Theory Appl.*, 146(2):267–303, 2010.

[15] C. Chen, R. H. Chan, S. Ma, and J. Yang. Inertial proximal ADMM for linearly constrained separable convex optimization. *SIAM J. Imaging Sci.*, 8(4):2239–2267, 2015.

[16] C. Chen, S. Ma, and J. Yang. A general inertial proximal point algorithm for mixed variational inequality problem. *SIAM J. Optim.*, 25(4):2120–2142, 2015.

[17] J. Eckstein and P. J. S. Silva. A practical relative error criterion for augmented Lagrangians. *Math. Program.*, 141(1-2, Ser. A):319–348, 2013.

[18] J. Eckstein and W. Yao. Relative-error approximate versions of Douglas-Rachford splitting and special cases of the ADMM. *Mathematical Programming*, 2017.

[19] R. Glowinski, S. J. Osher, and W. Yin, editors. *Splitting methods in communication, imaging, science, and engineering*. Scientific Computation. Springer, Cham, 2016.

[20] Y. He and R. D. C. Monteiro. An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM J. Optim.*, 26(1):29–56, 2016.

[21] A. N. Iusem and W. Sosa. On the proximal point method for equilibrium problems in Hilbert spaces. *Optimization*, 59(8):1259–1274, 2010.

[22] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[23] D. A. Lorenz and T. Pock. An inertial forward-backward algorithm for monotone inclusions. *J. Math. Imaging Vision*, 51(2):311–325, 2015.

[24] P. A. Lotito, L. A. Parente, and M. V. Solodov. A class of variable metric decomposition methods for monotone variational inclusions. *J. Convex Anal.*, 16(3-4):857–880, 2009.

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

[26] R. D. C. Monteiro, C. Ortiz, and B. F. Svaiter. A first-order block-decomposition method for solving two-easy-block structured semidefinite programs. *Math. Program. Comput.*, 6(2):103–150, 2014.

[27] R. D. C. Monteiro, C. Ortiz, and B. F. F. Svaiter. Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems. *Comput. Optim. Appl.*, 57(1):45–69, 2014.

[28] R. D. C. Monteiro, C. Ortiz, and B. F. F. Svaiter. An adaptive accelerated first-order method for convex optimization. *Comput. Optim. Appl.*, 64(1):31–73, 2016.

[29] R. D. C. Monteiro and B. F. Svaiter. 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 Journal on Optimization*, 21:1688–1720, 2010.

[30] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20:2755–2787, 2010.

[31] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *SIAM J. Optim.*, 23(1):475–507, 2013.

[32] A. Moudafi and M. Oliny. Convergence of a splitting inertial proximal method for monotone operators. *J. Comput. Appl. Math.*, 155(2):447–454, 2003.

[33] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.*, 72(2):383–390, 1979.

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

[35] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[36] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[37] M. V. Solodov and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. *Math. Oper. Res.*, 25(2):214–230, 2000.

[38] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[39] B. F. Svaiter. A class of Fejér convergent algorithms, approximate resolvents and the hybrid proximal-extragradient method. *J. Optim. Theory Appl.*, 162(1):133–153, 2014.

[40] B. F. Svaiter. Complexity of the relaxed hybrid proximal-extragradient method under the large-step condition. Preprint A766/2015, IMPA - Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ Brasil 22460-320, 2015.

[41] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446 (electronic), 2000.