

Regularized HPE-type methods for solving monotone inclusions with improved pointwise iteration-complexity bounds

M. Marques Alves ^{*} R.D.C. Monteiro [†] B. F. Svaiter[‡]

September 3, 2015 (Revised: April 13, 2016 and September 24, 2016)

Abstract

This paper studies the iteration-complexity of new regularized hybrid proximal extragradient (HPE)-type methods for solving monotone inclusion problems (MIPs). The new (regularized HPE-type) methods essentially consist of instances of the standard HPE method applied to regularizations of the original MIP. It is shown that its pointwise iteration-complexity considerably improves the one of the HPE method while approaches (up to a logarithmic factor) the ergodic iteration-complexity of the latter method.

2000 Mathematics Subject Classification: 47H05, 47J20, 90C060, 90C33, 65K10.

Key words: proximal point methods, hybrid proximal extragradient method, pointwise iteration-complexity, ergodic iteration-complexity, Tseng's MFBS method, Korpelevich's extragradient method.

1 Introduction

We consider the *monotone inclusion problem* (MIP) of finding x such that

$$0 \in B(x) \tag{1}$$

where B is a point-to-set maximal monotone operator. Special cases of problem (1) include convex optimization, saddle-point, equilibrium and variational inequality problems (see, e.g., [13, 14]) which arise in the context of many applications (see, e.g., [3, Chapter 1]). One of the most important schemes for solving MIPs is the proximal point method (PPM), proposed by Martinet [5] and further developed by Rockafellar [14]. It is an iterative scheme which, in its exact version, generates a sequence $\{x_k\}$ according to $x_k = (I + \lambda_k B)^{-1}x_{k-1}$ (where $\lambda_k > 0$), or equivalently, x_k as the unique

^{*}Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Brazil, 88040-900 (maicon.alves@ufsc.br). This work was done while this author was a postdoc at the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. The work of this author was partially supported by CNPq grants no. 406250/2013-8, 237068/2013-3 and 306317/2014-1

[†]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grant CMMI-1300221.

[‡]IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (benar@impa.br). The work of this author was partially supported by CNPq grants no. 306247/2015-1, 474996/2013-1, and FAPERJ grant E-21/201.584/2014 (Cientista de Nosso Estado).

solution of the MIP: $0 \in \lambda_k B(x) + x - x_{k-1}$. Among other results, Rockafellar [14] proposed inexact versions of the PPM based on a summable absolute error criterion and subsequently Solodov and Svaiter [16, 17] proposed new inexact variants based on a hybrid proximal extragradient (HPE) relative error criterion. In each step, the variants proposed and studied in [16], namely the HPE method, computes $\lambda = \lambda_k > 0$ and a triple $(y, b, \varepsilon) = (y_k, b_k, \varepsilon_k)$ satisfying

$$b \in B^{[\varepsilon]}(y), \quad \|\lambda b + y - x\|^2 + 2\lambda\varepsilon \leq \sigma^2 \|y - x\|^2, \quad (2)$$

where $x = x_{k-1}$ is the current iterate, $\sigma \in [0, 1)$ is a relative error tolerance and $B^{[\varepsilon]}$ denotes the ε -enlargement [2] of B . However, instead of choosing y as the next iterate, the HPE method computes $x_+ = x_k$ by means of an extragradient step $x_+ = x - \lambda b$.

The iteration-complexity of the HPE method was established in [7] with regards to the following termination criterion in terms of precisions $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$: find a triple (y, b, ε) such that

$$b \in B^{[\varepsilon]}(y), \quad \|b\| \leq \bar{\rho}, \quad \varepsilon \leq \bar{\varepsilon}. \quad (3)$$

Assuming that the sequence of stepsizes $\{\lambda_k\}$ in the HPE method is bounded below by some constant $\underline{\lambda} > 0$, the pointwise iteration-complexity result of [7] guarantees that the most recent triple (y, b, ε) satisfying (2) will eventually satisfy the termination criterion given in (3) in at most $\mathcal{O}(\max\{d_0^2/\underline{\lambda}^2\bar{\rho}^2, d_0^2/\underline{\lambda}\bar{\varepsilon}\})$ iterations where d_0 denotes the distance of the initial iterate x_0 to the solution set of (1). Moreover, under the same condition on the sequence of stepsizes $\{\lambda_k\}$, an ergodic iteration-complexity result of [7] shows that an ergodic triple constructed from all previous generated triples satisfying (2) will eventually satisfy (3) in at most $\mathcal{O}(\max\{d_0/\underline{\lambda}\bar{\rho}, d_0^2/\underline{\lambda}\bar{\varepsilon}\})$ iterations. Clearly, the ergodic iteration-complexity is better than the pointwise one by a factor of $\mathcal{O}(\max\{1, d_0/\underline{\lambda}\bar{\rho}\})$.

Our main goal in this paper is to present regularized HPE-type methods for solving (1) which essentially consists of instances of the HPE method applied to the regularized MIP

$$0 \in B(x) + \mu(x - x_0) \quad (4)$$

where $\mu > 0$ and x_0 is an initial point. In particular, it is shown that a certain version of the regularized HPE method which dynamically adjusts $\mu > 0$ solves (1) in at most

$$\mathcal{O}\left(\left(\frac{d_0}{\underline{\lambda}\bar{\rho}} + 1\right) \left[1 + \max\left\{\log^+\left(\frac{d_0}{\underline{\lambda}\bar{\rho}}\right), \log^+\left(\frac{d_0}{\underline{\lambda}\bar{\varepsilon}}\right)\right\}\right]\right) \quad (5)$$

iterations where $\log^+(t) = \max\{\log(t), 0\}$ for every $t > 0$. This pointwise iteration-complexity bound considerably improves the one for the original HPE method. Also, note that it differs from the ergodic one for the original HPE method by only a logarithmic factor. Finally, we discuss specific instances of the regularized HPE method which are based on Tseng's modified forward-backward splitting (MFBS) method [19] and Korpelevich's extragradient method [4].

Previous most related works. In the context of variational inequalities (VIs), Nemirovski [11] has established the ergodic iteration-complexity of an extension of Korpelevich's method, namely, the mirror-prox algorithm, under the assumption that the feasible set of the problem is bounded. Nesterov [12] proposed a new dual extrapolation algorithm for solving VIs whose termination depends on the guess of a ball centered at the initial iterate. Applications of the HPE method to the iteration-complexity analysis of several zeroth-order (or, in the context of optimization, first-order) methods for solving monotone VIs, MIPs and saddle-point problems were discussed by Monteiro and Svaiter

in [7] and in the subsequent papers [8, 10]. The HPE method was also used to study the iteration-complexities of first-order (or, in the context of optimization, second-order) methods for solving either a monotone nonlinear equation (see Section 7 of [7]) and, more generally, a monotone VI (see [9]).

Organization of the paper. Section 2 contains two subsections. Subsection 2.1 presents the notation as well as some basic concepts about convexity and maximal monotone operators. Subsection 2.2 is devoted to the study of a specialization of the HPE method for solving inclusions whose underlying operator is written as a sum of a (maximal) monotone and a strongly monotone operator. Section 3 presents the main contributions of the paper, namely, the presentation of two new regularized HPE methods (a static one and a dynamic one) as well as its complexity analysis. Section 4 discusses two specific instances of the dynamic regularized HPE method of Section 3 based on Tseng’s MFBS method and Korpelevich’s extragradient method. Finally, the appendix presents the proofs of some results in Subsection 2.2.

2 Preliminaries

This section discusses some preliminary results which will be used throughout the paper. Subsection 2.1 presents the general notation and some basic concepts about convexity, maximal monotone operators, and related issues. Subsection 2.2 describes a special version of the HPE method introduced in [16] for solving monotone inclusions whose underlying operators consist of the sum of a (maximal) monotone and a strongly (maximal) monotone operator. Our motivation for considering this class of inclusions is due to (4).

2.1 Basic concepts and notation

Let X be a finite-dimensional real vector space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \| := \sqrt{\langle \cdot, \cdot \rangle}$. Given a set-valued operator $A : X \rightrightarrows X$, its *graph* and *domain* are, respectively, $\text{Gr}(A) := \{(x, v) \in X \times X : v \in A(x)\}$ and $\text{Dom}(A) := \{x \in X : A(x) \neq \emptyset\}$. The *inverse* of $A : X \rightrightarrows X$ is $A^{-1} : X \rightrightarrows X$, $A^{-1}(v) := \{x : v \in A(x)\}$. The *sum* of two set-valued operators $A, B : X \rightrightarrows X$ is defined by $A + B : X \rightrightarrows X$, $(A + B)(x) := \{a + b \in X : a \in A(x), b \in B(x)\}$.

An operator $A : X \rightrightarrows X$ is μ -strongly monotone if $\mu \geq 0$ and

$$\langle v - v', x - x' \rangle \geq \mu \|x - x'\|^2 \quad \forall (x, v), (x', v') \in \text{Gr}(A). \quad (6)$$

If $\mu = 0$ in the above inequality, then A is said to be a *monotone* operator. Moreover, $A : X \rightrightarrows X$ is *maximal monotone* if it is monotone and maximal in the following sense: if $B : X \rightrightarrows X$ is monotone and $\text{Gr}(A) \subset \text{Gr}(B)$, then $A = B$. The *resolvent* of a maximal monotone operator $A : X \rightrightarrows X$ with parameter $\lambda > 0$ is $(I + \lambda A)^{-1}$. It follows directly from this definition that $y = (I + \lambda A)^{-1}x$ if and only if $(x - y)/\lambda \in A(y)$. It is easy to see that if $A : X \rightrightarrows X$ is μ -strongly monotone and $B : X \rightrightarrows X$ is monotone, then the sum $A + B$ is also μ -strongly monotone. In particular, the sum of two monotone operators is also a monotone operator.

For $\varepsilon \geq 0$, the ε -enlargement [2] of a maximal monotone operator $B : X \rightrightarrows X$ is defined by $B^{[\varepsilon]} : X \rightrightarrows X$,

$$B^{[\varepsilon]}(x) := \{v \in X : \langle v - v', x - x' \rangle \geq -\varepsilon, \quad \forall (x', v') \in \text{Gr}(B)\}. \quad (7)$$

The following summarizes some useful properties of $B^{[\varepsilon]}$.

Proposition 2.1. *Let $A, B : X \rightrightarrows X$ be maximal monotone operators. Then,*

- (a) *if $\varepsilon_1 \leq \varepsilon_2$, then $A^{[\varepsilon_1]}(x) \subseteq A^{[\varepsilon_2]}(x)$ for every $x \in X$;*
- (b) *$A^{[\varepsilon']}(x) + (B)^{[\varepsilon]}(x) \subseteq (A + B)^{[\varepsilon' + \varepsilon]}(x)$ for every $x \in X$ and $\varepsilon, \varepsilon' \geq 0$;*
- (c) *A is monotone if, and only if, $A \subseteq A^{[0]}$;*
- (d) *A is maximal monotone if, and only if, $A = A^{[0]}$;*

Recall that for $\varepsilon \geq 0$, the ε -subdifferential of a proper closed convex function $f : X \rightarrow \overline{\mathbb{R}}$ is defined at $x \in X$ by $\partial_\varepsilon f(x) := \{v \in X : f(x') \geq f(x) + \langle v, x' - x \rangle - \varepsilon \ \forall x' \in X\}$. When $\varepsilon = 0$, then $\partial_0 f(x)$ is denoted by $\partial f(x)$ and is called the *subdifferential* of f at x . The simplest example of subdifferential is given by considering indicator functions of closed convex sets. Given a closed convex set $\mathcal{X} \subset X$ its *indicator function* is denoted by $\delta_{\mathcal{X}}$ and is defined by $\delta_{\mathcal{X}}(x) := 0$ if $x \in \mathcal{X}$ and $\delta_{\mathcal{X}}(x) := \infty$ otherwise. The *normal cone* of \mathcal{X} is defined by $N_{\mathcal{X}} := \partial \delta_{\mathcal{X}}$. We also define the *projection* on \mathcal{X} by $P_{\mathcal{X}} := (I + N_{\mathcal{X}})^{-1}$.

2.2 Solving inclusions with strongly monotone operators

In this subsection, we consider the MIP

$$0 \in (A + B)(x) \tag{8}$$

where the following assumptions hold:

- A.1) $A : X \rightrightarrows X$ is a μ -strongly maximal monotone operator for some $\mu > 0$ (see (6));
- A.2) $B : X \rightrightarrows X$ is maximal monotone;
- A.3) the solution set of (8), i.e., $(A + B)^{-1}(0)$, is nonempty.

We now make a few remarks. First, (4) is a special case of (8) in which A is the μ -strongly monotone point-to-point operator $A(\cdot) = \mu(\cdot - x_0)$. Second, A.1, A.2 and A.3 imply that $(A + B)^{-1}(0)$ is a singleton. Third, A.1 and A.2 imply that $A + B$ is μ -strongly monotone. If, in addition, $A + B$ is maximal monotone, then A.3 holds in view of Proposition 12.54 of [15].

We next state a specialized HPE method for solving (8) under the assumptions stated above. It will be used later on in Section 3 to describe regularized HPE methods for general MIPs whose pointwise iteration-complexities improve the ones for the original HPE method (see [7]).

Algorithm 1: A specialized HPE method for solving strongly monotone inclusions

(0) Let $x_0 \in X$ and $\sigma \in [0, 1)$ be given and set $k = 1$;

(1) choose $\lambda_k > 0$ and find $y_k, v_k \in X$, $\sigma_k \in [0, \sigma]$, and $\varepsilon_k \geq 0$ such that

$$v_k \in A(y_k) + B^{[\varepsilon_k]}(y_k), \quad \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma_k^2 \|y_k - x_{k-1}\|^2; \tag{9}$$

(2) set

$$x_k = x_{k-1} - \lambda_k v_k, \tag{10}$$

let $k \leftarrow k + 1$ and go to step 1.

end

We now make some remarks about Algorithm 1. First, it can be easily checked that if $\sigma = 0$ then $\sigma_k = 0$, $\varepsilon_k = 0$ and $x_k = y_k$ for every $k \geq 1$ and Algorithm 1 reduces to the exact proximal point method (PPM) for solving (8), i.e.,

$$x_k = (\lambda_k(A + B) + I)^{-1}x_{k-1} \quad \forall k \geq 1.$$

Second, since $A(y) + B^{[\varepsilon]}(y) \subset (A + B)^{[\varepsilon]}(y)$ for every y in view of Proposition 2.1(b), it follows that Algorithm 1 is a special instance of the HPE method studied in [7]. Third, like in the HPE method, step 1 of Algorithm 1 does not specify how to compute the stepsize λ_k and the triple $(y_k, v_k, \varepsilon_k)$. Their computation will depend on the instance of the method under consideration. For instance, Proposition 4.1 illustrates how these quantities can be computed for the case in which $A = \mu(\cdot - x_0)$ and B is the sum of a Lipschitz continuous monotone operator and a maximal monotone operator with an easily computable resolvent. Other results along this direction can be found in [7, 8, 9, 10].

The next result derives convergence rates for the sequences $\{v_k\}$ and $\{\varepsilon_k\}$ generated by Algorithm 1 under the assumption that the sequence of stepsizes $\{\lambda_k\}$ is bounded away from zero. Its proof is given in Appendix A.

Proposition 2.2. *Define $x^* := (A + B)^{-1}(0)$, $d_0 := \|x_0 - x^*\|$ and*

$$\theta := \left(\frac{1}{2\lambda\mu} + \frac{1}{1 - \sigma^2} \right)^{-1} \in (0, 1). \quad (11)$$

Assume that $\lambda_k \geq \lambda > 0$ for every $k \geq 1$. Then, for every $k \geq 1$, $v_k \in A(y_k) + B^{[\varepsilon_k]}(y_k)$,

$$\|v_k\| \leq \sqrt{\frac{1 + \sigma}{1 - \sigma}} \left(\frac{(1 - \theta)^{(k-1)/2}}{\lambda} \right) d_0, \quad \varepsilon_k \leq \frac{\sigma^2}{2(1 - \sigma^2)} \left(\frac{(1 - \theta)^{k-1}}{\lambda} \right) d_0^2.$$

$$\|x^* - x_k\| \leq (1 - \theta)^{k/2} d_0, \quad \|x^* - y_k\| \leq \min \left\{ \frac{1}{\sqrt{2\lambda\mu}}, 1 + \frac{1}{\sqrt{1 - \sigma^2}} \right\} (1 - \theta)^{(k-1)/2} d_0.$$

Observe that the assumption that $\sigma \in [0, 1)$ implies that θ is well-defined and lies in $(0, 1)$.

3 Regularized HPE methods for solving MIPs

This section presents regularized HPE-type methods for solving MIPs whose pointwise iteration-complexity is superior to the pointwise iteration-complexity of the original HPE method (see [7]). More specifically, it is shown that the pointwise iteration-complexity bound for the new regularized HPE method is worse than the ergodic bound for the original HPE method by only a logarithmic factor.

This section considers the MIP (1) where $B : X \rightrightarrows X$ is a point-to-set maximal monotone operator such that $B^{-1}(0) \neq \emptyset$, and discusses regularized HPE-type methods which, for a given point $x_0 \in X$, consist of solving MIPs parametrized by a scalar $\mu > 0$ as in (4). Observe that (4) is a regularized version of (1). Its operator is μ -strongly monotone and approaches the one of (1) as $\mu > 0$ approaches zero. Clearly, (4) is a special case of (8) with $A(x) = \mu(x - x_0)$ and its solution set is a singleton by Minty's theorem [6].

We denote the distance of x_0 to the solution sets of (1) and (4) by d_0 and d_μ , respectively. Clearly,

$$d_\mu = \|x_\mu^* - x_0\| \quad (12)$$

where x_μ^* denotes the unique solution of (4), i.e., $x_\mu^* = (\mu^{-1}B + I)^{-1}(x_0)$.

The following simple technical result relates d_μ with d_0 .

Lemma 3.1. *For every $\mu > 0$, $d_\mu \leq d_0$.*

Proof. Let x^* be the projection of x_0 onto $B^{-1}(0)$. Since $0 \in B(x^*)$ and $\mu(x_0 - x_\mu^*) \in B(x_\mu^*)$, the monotonicity of B and the fact that $\mu > 0$ imply that $\langle x^* - x_\mu^*, x_\mu^* - x_0 \rangle \geq 0$. Therefore,

$$d_0^2 = \|x^* - x_0\|^2 = \|x^* - x_\mu^*\|^2 + 2\langle x^* - x_\mu^*, x_\mu^* - x_0 \rangle + \|x_\mu^* - x_0\|^2 \geq \|x^* - x_\mu^*\|^2 + d_\mu^2$$

and the conclusion follows. \square

We now state a μ -regularized HPE method for solving (1) which is simply Algorithm 1 (with $A(\cdot) = \mu(\cdot - x_0)$) applied to MIP (4) but with a termination criterion added.

Algorithm 2: A static μ -regularized HPE method for solving (1).

Input: $(x_0, \sigma, \mu, \rho, \varepsilon) \in X \times [0, 1) \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$;

(0) set $k = 1$;

(1) choose $\lambda_k > 0$ and find $(y_k, b_k, \varepsilon_k) \in X \times X \times \mathbb{R}_+$ such that

$$b_k \in B^{[\varepsilon_k]}(y_k), \quad \|\lambda_k [b_k + \mu(y_k - x_0)] + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|y_k - x_{k-1}\|^2; \quad (13)$$

(2) if $\|b_k + \mu(y_k - x_0)\| > \rho$ or $\varepsilon_k > \varepsilon$, then set

$$x_k = x_{k-1} - \lambda_k [b_k + \mu(y_k - x_0)], \quad (14)$$

and $k \leftarrow k + 1$, and go to step 1; otherwise, stop the algorithm and output $(y_k, b_k, \varepsilon_k)$.

end

We now make some remarks about Algorithm 2. First, it is the special case of Algorithm 1 in which $A(\cdot) = \mu(\cdot - x_0)$, and hence solves the MIP (4). Second, since Subsection 2.2 only deals with convergence rate bounds, a stopping criterion was not added to Algorithm 1. In contrast, Algorithm 2 incorporates a stopping criterion (see step 2 above) based on which its iteration-complexity bound is derived in Proposition 3.2 and Theorem 3.3 below. Third, it is shown in Theorem 3.3(b) that Algorithm 2 solves MIP (1) if μ is chosen sufficiently small. Fourth, if no stopping criterion is added to Algorithm 2 (see its step 2), then it follows from Proposition 2.2 that $\max\{\|x_k - x_\mu^*\|, \|y_k - x_\mu^*\|\}$ converges to zero geometrically where $x_\mu^* = (\mu^{-1}B + I)^{-1}x_0$ is the unique solution of MIP (4). This contrasts to the sequence generated by a variant of the DR-HPE method stated later in this section which asymptotically converges to a solution of (1) (see the fourth remark following the statement of the DR-HPE method).

Proposition 3.2. *Assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$ and let d_μ be as in (12). Then, Algorithm 2 with input $(x_0, \sigma, \mu, \rho, \varepsilon)$ terminates in at most*

$$\left(\frac{1}{2\underline{\lambda}\mu} + \frac{1}{1 - \sigma^2} \right) \left[2 + \max \left\{ \log^+ \left(\left[\frac{1 + \sigma}{1 - \sigma} \right] \frac{d_\mu^2}{\underline{\lambda}^2 \rho^2} \right), \log^+ \left(\frac{\sigma^2 d_\mu^2}{2(1 - \sigma^2)\underline{\lambda}\varepsilon} \right) \right\} \right] \quad (15)$$

iterations with a triple $(y_k, b_k, \varepsilon_k)$ such that

$$b_k \in B^{[\varepsilon_k]}(y_k), \quad \|b_k + \mu(y_k - x_0)\| \leq \rho, \quad \varepsilon_k \leq \varepsilon \quad (16)$$

and

$$\|y_k - x_0\| \leq \left(1 + \frac{1}{\sqrt{1 - \sigma^2}}\right) d_\mu \leq \left(1 + \frac{1}{\sqrt{1 - \sigma^2}}\right) d_0, \quad (17)$$

$$\|b_k\| \leq \rho + \mu \left(1 + \frac{1}{\sqrt{1 - \sigma^2}}\right) d_\mu \leq \rho + \mu \left(1 + \frac{1}{\sqrt{1 - \sigma^2}}\right) d_0. \quad (18)$$

Proof. To prove that the number of iterations of Algorithm 2 is bounded by (15), assume that it has not terminated at the k -th iteration and define $v_k = b_k + \mu(y_k - x_0)$. Then, either $\|v_k\| > \rho$ or $\varepsilon_k > \varepsilon$. Assume first that $\|v_k\| > \rho$. Since Algorithm 2 is a special case of Algorithm 1 applied to MIP (4) with $A(x) = \mu(x - x_0)$ and v_k as above, the latter assumption and Corollary 2.2 imply that

$$\rho < \|v_k\| \leq \sqrt{\frac{1 + \sigma}{1 - \sigma}} \left(\frac{(1 - \theta)^{(k-1)/2}}{\underline{\lambda}} \right) d_\mu$$

where θ is defined in (11). Rearranging this inequality, taking logarithms of both sides of the resulting inequality and using the fact that $\log(1 - \theta) \leq -\theta$, we conclude that

$$k < 1 + \theta^{-1} \log \left(\left[\frac{1 + \sigma}{1 - \sigma} \right] \frac{d_\mu^2}{\underline{\lambda}^2 \rho^2} \right).$$

If, on the other hand, $\varepsilon_k > \varepsilon$, we conclude by using a similar reasoning that

$$k < 1 + \theta^{-1} \log \left(\frac{\sigma^2 d_\mu^2}{2(1 - \sigma^2) \underline{\lambda} \varepsilon} \right).$$

From the above two observations and the fact that $\theta < 1$ in view of (11), we conclude that the number of iterations performed by Algorithm 2 is bounded by (15). Conditions (16) follow from (13) and the testing condition in step 2 of Algorithm 2.

To prove (17), note that Lemma 2.1(5) of [18], Corollary 2.2 and (12) imply that

$$\|y_k - x_\mu^*\| \leq \frac{\|x_{k-1} - x_\mu^*\|}{\sqrt{1 - \sigma^2}} \leq \frac{(1 - \theta)^{(k-1)/2}}{\sqrt{1 - \sigma^2}} d_\mu \leq \frac{1}{\sqrt{1 - \sigma^2}} d_\mu,$$

and hence that

$$\|y_k - x_0\| \leq \|y_k - x_\mu^*\| + \|x_\mu^* - x_0\| \leq \left(1 + \frac{1}{\sqrt{1 - \sigma^2}}\right) d_\mu.$$

The latter conclusion and Lemma 3.1 yield (17). To finish the proof, note that (18) follows from the first inequality in (16), the triangle inequality and (17). \square

The complexity results presented in this paper will consist in establishing bounds in the number of iterations to obtain a triple (y, b, ε) satisfying (3), for given precisions $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$.

The following result shows that Algorithm 2 solves the MIP (1) when $\mu > 0$ is chosen sufficiently small.

Theorem 3.3. Assume that $\lambda_k \geq \underline{\lambda} > 0$ for all $k \geq 1$ and let a tolerance pair $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ be given. Then, the following statements hold:

(a) for any $\rho \in (0, \bar{\rho})$ and $\mathcal{D}_0 > 0$, Algorithm 2 with input $(x_0, \sigma, \mu, \rho, \varepsilon)$ where

$$\mu = \mu(\mathcal{D}_0, \rho) := \frac{\bar{\rho} - \rho}{\left[1 + \frac{1}{\sqrt{1 - \sigma^2}}\right] \mathcal{D}_0}, \quad \varepsilon = \bar{\varepsilon} \quad (19)$$

terminates in at most

$$\left(\frac{\left[1 + 1/\sqrt{1 - \sigma^2}\right] \mathcal{D}_0}{2\underline{\lambda}(\bar{\rho} - \rho)} + \frac{1}{1 - \sigma^2} \right) \left[2 + \max \left\{ \log^+ \left(\left[\frac{1 + \sigma}{1 - \sigma} \right] \frac{d_0^2}{\underline{\lambda}^2 \rho^2} \right), \log^+ \left(\frac{\sigma^2 d_0^2}{2(1 - \sigma^2) \underline{\lambda} \bar{\varepsilon}} \right) \right\} \right] \quad (20)$$

iterations with a triple $(y_k, b_k, \varepsilon_k)$ such that $b_k \in B^{[\varepsilon_k]}(y_k)$ and $\varepsilon_k \leq \bar{\varepsilon}$;

(b) in addition, if $\mathcal{D}_0 \geq d_0$, then the above triple also satisfies

$$\|b_k\| \leq \bar{\rho}, \quad \mu \|y_k - x_0\| \leq \bar{\rho} - \rho, \quad (21)$$

and hence satisfies (3).

Proof. (a) This statement follows from (15), Lemma 3.1 and Proposition 3.2 with μ and ε given by (19).

(b) Using the second inequalities in (17) and (18), and the first identity in (19), we conclude that

$$\max \{ \|b_k\| - \rho, \mu \|y_k - x_0\| \} \leq \frac{d_0}{\mathcal{D}_0} (\bar{\rho} - \rho).$$

Hence, if $\mathcal{D}_0 \geq d_0$, the latter inequality yields (21). \square

We now make two remarks about Theorem 3.3. First, if $\sigma \in [0, 1)$ is such that $(1 - \sigma)^{-1} = \mathcal{O}(1)$, an upper bound $\mathcal{D}_0 \geq d_0$ such that $\mathcal{D}_0 = \mathcal{O}(d_0)$ is known, and ρ is set to $\bar{\rho}/2$, then the complexity bound (20) becomes

$$\mathcal{O} \left(\left(\frac{d_0}{\underline{\lambda} \bar{\rho}} + 1 \right) \left[1 + \max \left\{ \log^+ \left(\frac{d_0}{\underline{\lambda} \bar{\rho}} \right), \log^+ \left(\frac{d_0}{\underline{\lambda} \bar{\varepsilon}} \right) \right\} \right] \right). \quad (22)$$

Second, in general an upper bound \mathcal{D}_0 as in the first remark is not known and in this case the bound (20) can be much worse than the one above whenever $\mathcal{D}_0 \gg d_0$.

In the remaining part of this section, we consider the case where an upper bound $\mathcal{D}_0 \geq d_0$ such that $\mathcal{D}_0 = \mathcal{O}(d_0)$ is not known and describe a scheme based on Algorithm 2 whose iteration-complexity order is equal to (22).

DR-HPE: A dynamic regularized HPE method for solving (1).

- (0) Let $x_0 \in X$, $\sigma \in [0, 1)$, $\bar{\lambda} > 0$ and a tolerance pair $(\bar{\rho}, \bar{\varepsilon}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ be given and choose $\rho \in (0, \bar{\rho})$; set

$$\mathcal{D}_0 = \bar{\mathcal{D}}_0 := \frac{2\bar{\lambda}(\bar{\rho} - \rho)}{(1 - \sigma^2) \left(1 + 1/\sqrt{1 - \sigma^2}\right)}; \quad (23)$$

- (1) set $\mu = \mu(\mathcal{D}_0, \rho)$ where $\mu(\cdot, \cdot)$ is defined in (19) and call Algorithm 2 with input $(x_0, \sigma, \mu, \rho, \bar{\varepsilon})$ to obtain as output (y, b, ε) ;
- (2) if $\|b\| \leq \bar{\rho}$, then stop and output (y, b, ε) ; else, set $\mathcal{D}_0 \leftarrow 2\mathcal{D}_0$ and go to step 1.

end

We now make some remarks about DR-HPE algorithm. First, each iteration of DR-HPE (referred to as an outer iteration) invokes Algorithm 2, and hence performs a certain number of iterations of the latter method (called inner iterations) which is bounded by (20). Second, due to the fact that \mathcal{D}_0 is doubled at every iteration and the definition of $\mu(\cdot, \cdot)$ in (19), it follows that μ is halved at every iteration, and hence μ converges to zero as the number of outer iterations grows. Third, it is easy to see that the output y generated in the step 1 of the DR-HPE method satisfies

$$\|y - x_\mu^*\| \leq \frac{\tilde{\rho} + \sqrt{\tilde{\rho}^2 + 4\tilde{\varepsilon}}}{2}$$

where $\tilde{\rho} = \rho/\mu$ and $\tilde{\varepsilon} = \bar{\varepsilon}/\mu$. Fourth, if one considers the version of the DR-HPE method in which no stopping criteria is added to its step 2 and the tolerances ρ and $\bar{\varepsilon}$ in its step 1 are allowed to be chosen in such a way that $\tilde{\rho} = \rho/\mu$ and $\tilde{\varepsilon} = \bar{\varepsilon}/\mu$ converge to zero as the number of outer iteration increases, then it follows that the output y of its step 1 converges to the point $x^* \in B^{-1}(0)$ closest to x_0 , in view of the above inequality and the fact that x_μ^* converges to such x^* (see [1, Lemma 1]). Fifth, our version of DR-HPE stated above keeps the tolerances ρ and $\bar{\varepsilon}$ fixed and hence the fourth remark just described does not apply to it.

The following result describes the overall inner-iteration-complexity of DR-HPE in terms of d_0 , $\bar{\lambda}$, ρ , $\bar{\rho}$ and $\bar{\varepsilon}$.

Theorem 3.4. *Let d_0 denote the distance of x_0 to the solution set of (1) and assume that the proximal stepsize in every inner iteration of DR-HPE is bounded below by a constant $\underline{\lambda} > 0$. Then, DR-HPE with input $(x_0, \sigma, \bar{\lambda}, (\bar{\rho}, \bar{\varepsilon}), \rho) \in X \times [0, 1) \times \mathbb{R}_{++} \times \mathbb{R}_{++}^2 \times \mathbb{R}_{++}$ such that $\rho \in (0, \bar{\rho})$ and $(1 - \sigma)^{-1} = O(1)$ finds a triple (y, b, ε) satisfying*

$$b \in B^{[\varepsilon]}(y), \quad \|b\| \leq \bar{\rho}, \quad \varepsilon \leq \bar{\varepsilon}$$

in at most

$$O\left(\left(1 + \frac{\bar{\lambda}}{\underline{\lambda}}\right) \left(\frac{d_0}{\bar{\lambda}(\bar{\rho} - \rho)} + 1\right) \left[1 + \max\left\{\log^+\left(\frac{d_0}{\underline{\lambda}\rho}\right), \log^+\left(\frac{d_0}{\underline{\lambda}\bar{\varepsilon}}\right)\right\}\right]\right) \quad (24)$$

iterations.

Proof. Note that at the k -th outer iteration of DR-HPE, we have $\mathcal{D}_0 = 2^{k-1}\bar{\mathcal{D}}_0$. Moreover, in view of Theorem 3.3(b), DR-HPE terminates in at most K outer iterations where K is the smallest integer $k \geq 1$ satisfying $2^{k-1}\bar{\mathcal{D}}_0 \geq d_0$, i.e.,

$$K = 1 + \left\lceil \log^+ \left(\frac{d_0}{\bar{\mathcal{D}}_0} \right) \right\rceil.$$

Define

$$\beta_1 := 2 + \max \left\{ \log^+ \left(\left[\frac{1+\sigma}{1-\sigma} \right] \frac{d_0^2}{\lambda^2 \rho^2} \right), \log^+ \left(\frac{\sigma^2 d_0^2}{2(1-\sigma^2)\lambda\bar{\varepsilon}} \right) \right\} \quad (25)$$

$$\beta_0 := \frac{\beta_1}{1-\sigma^2} = \frac{(1+1/\sqrt{1-\sigma^2})\bar{\mathcal{D}}_0}{2\bar{\lambda}(\bar{\rho}-\rho)} \beta_1 \quad (26)$$

where the identity in (26) follows from (23). In view of Theorem 3.3(a), and relations (25) and (26), we then conclude that the overall number of inner iterations of DR-HPE is bounded by

$$\tilde{K} := \beta_0 \sum_{k=1}^K \left(1 + \frac{\bar{\lambda}}{\underline{\lambda}} 2^{k-1} \right) = \beta_0 \left[K + \frac{\bar{\lambda}}{\underline{\lambda}} (2^K - 1) \right] \leq \beta_0 \left(1 + \frac{\bar{\lambda}}{\underline{\lambda}} \right) 2^K. \quad (27)$$

To prove the theorem, it suffices to show that \tilde{K} is bounded by (24). Indeed, we consider two cases, namely, whether $K = 1$ or $K > 1$. If $K = 1$, then (27) implies that $\tilde{K} \leq 2\beta_0(1 + \bar{\lambda}/\underline{\lambda})$, and hence that the order of \tilde{K} is bounded by (24) in view of the definition of β_0 in (26). Assume now that $K > 1$ and note that the definition of K implies that $k = K - 1$ violates the inequality $2^{k-1}\bar{\mathcal{D}}_0 \geq d_0$, and hence that $2^K < 4d_0/\bar{\mathcal{D}}_0$. The latter conclusion and inequality (27) then imply that $\tilde{K} < 4\beta_0 d_0(1 + \bar{\lambda}/\underline{\lambda})/\bar{\mathcal{D}}_0$, which together with (25) and (26) then imply that \tilde{K} is bounded by (24). \square

We now make a few remarks before ending this section. First, if the lower bound $\underline{\lambda} > 0$ for the sequence of proximal stepsizes is known, and $\bar{\lambda} = \underline{\lambda}$ and $\rho = \bar{\rho}/2$ are chosen as input for DR-HPE, then the iteration-complexity bound (24) reduces to bound (22). This observation justifies our claim preceding DR-HPE. Second, it is possible to prove that the ergodic complexity of the DR-HPE method is $\mathcal{O}(\max\{1/\bar{\rho}, 1/\bar{\varepsilon}\})$, which is the same (up to a multiplicative factor) as the one for the original HPE method.

4 Specific instances of the DR-HPE method

In this section, we briefly discuss specific ways of implementing step 1 of Algorithm 2.

More specifically, we assume that operator B has the structure

$$B(x) := F(x) + C(x) \quad (28)$$

where the following conditions hold:

B.1) $F : \text{Dom}(F) \subset X \rightarrow X$ is a (single-valued) monotone operator on $\text{Dom}(C) \subset \text{Dom}(F)$, i.e.,

$$\langle F(x) - F(x'), x - x' \rangle \geq 0, \quad \forall x, x' \in \text{Dom}(C); \quad (29)$$

B.2) F is L -Lipschitz continuous on a closed convex set Ω such that $\text{Dom}(C) \subset \Omega \subset \text{Dom}(F)$, i.e., there exists $L > 0$ such that

$$\|F(x) - F(x')\| \leq L\|x - x'\| \quad \forall x, x' \in \Omega; \quad (30)$$

B.3) $C : X \rightrightarrows X$ is maximal monotone.

Our goal in this section is to discuss a Tseng's modified forward-backward splitting (MFBS) type scheme [19] for implementing step 1 of Algorithm 2 for an operator B with the above structure where two evaluations of F and a single resolvent evaluation of C , i.e., an operator of the form $(I + \lambda C)^{-1}$ for some $\lambda > 0$, are made.

Let (x_0, σ, μ) be the first three entities of the input for Algorithm 2 and assume here that $\sigma \in (0, 1)$. Consider the MIP

$$0 \in B_\mu(x) := F(x) + C_\mu(x) \quad (31)$$

where $C_\mu : X \rightrightarrows X$ is defined as

$$C_\mu(x) := C(x) + \mu(x - x_0) \quad \forall x \in \text{Dom}(C). \quad (32)$$

Given $x_{k-1} \in X$, the following two relations describes an iteration of a variant of Tseng's MFBS algorithm studied in [8] (see also [7]) for the above MIP:

$$y_k = (I + \lambda C_\mu)^{-1}(x_{k-1} - \lambda F(P_\Omega(x_{k-1}))), \quad (33)$$

$$x_k = y_k - \lambda(F(y_k) - F(P_\Omega(x_{k-1}))) \quad (34)$$

where $\lambda := \sigma/L$. Since by assumption B.2 we have $\text{Dom}(C) \subset \Omega \subset \text{Dom}(F)$, and $\text{Dom}(C_\mu) = \text{Dom}(C)$, it follows that $P_\Omega(x_{k-1})$ and y_k belong to $\text{Dom}(F)$, and hence that the iteration defined in (33)–(34) is well-defined. Moreover, the assumption that the resolvent of C is computable makes the resolvent $(I + \lambda C_\mu)^{-1}$ also computable since

$$(I + \lambda C_\mu)^{-1} x = \left(I + \frac{\lambda}{1 + \lambda\mu} C \right)^{-1} \left(\frac{x + \lambda\mu x_0}{1 + \lambda\mu} \right) \quad x \in X.$$

The following proposition was essentially proved in [8, Proposition 4.5] with a different notation.

Proposition 4.1. *The points y_k and x_k defined by (33) and (34) with $\lambda = \sigma/L$ and the vector*

$$c_k = \frac{1}{\lambda}(x_{k-1} - y_k) - F(P_\Omega(x_{k-1})) - \mu(y_k - x_0)$$

satisfy

$$c_k \in C(y_k), \quad \|\lambda[F(y_k) + c_k + \mu(y_k - x_0)] + y_k - x_{k-1}\| \leq \sigma\|y_k - x_{k-1}\|, \quad (35)$$

and hence $b_k := F(y_k) + c_k$, $\lambda_k := \lambda$ and $\varepsilon_k := 0$ satisfy (13).

Proof. The inclusion in (35) follows directly from (33), (32) and the definition of c_k . On the other hand, using items (a) and (c) of [8, Proposition 4.5] (with a different notation), we obtain the inequality in (35). The last statement of the proposition follows from the definition of b_k , (28), (35) and Proposition 2.1(d). \square

In the next theorem we show the iteration-complexity of DR-HPE for solving (28) under the assumption that the iteration of the variant of Tseng's MFBS method described in (33)–(34) is used as an implementation of step 1 of Algorithm 2.

Theorem 4.2. *If $\max\{\sigma^{-1}, (1 - \sigma)^{-1}\} = \mathcal{O}(1)$, then DR-HPE in which step 1 of Algorithm 2 is implemented according to the recipe described in Proposition 4.1 terminates with a pair (y, b) satisfying*

$$b \in (F + C)(y), \quad \|b\| \leq \bar{\rho} \quad (36)$$

in at most

$$\mathcal{O}\left(\left(1 + \frac{Ld_0}{\bar{\rho} - \rho}\right) \left[1 + \log^+\left(\frac{Ld_0}{\rho}\right)\right]\right) \quad (37)$$

iterations where $\bar{\rho}$ and ρ are as in step 0 of DR-HPE.

Proof. The result is a direct consequence of Theorem 3.4 and Proposition 4.1 where $\underline{\lambda} = \bar{\lambda} = \lambda := \sigma/L$. We note that since by Proposition 4.1 we have $\varepsilon_k = 0$ for all $k \geq 1$ the complexity bound on (24) is independent of the precision $\bar{\varepsilon} > 0$. \square

We now make some comments about the special instance of DR-HPE described in Theorem 4.2 in light of a previous variant of Tseng's MFBS algorithm studied in [8] for solving MIP (28). First, the cost of an inner iteration of the above two methods are identical. Second, if $\rho = \bar{\rho}/2$, then the complexity bound (37) reduces to

$$\mathcal{O}\left(\left(1 + \frac{Ld_0}{\bar{\rho}}\right) \left[1 + \log^+\left(\frac{Ld_0}{\bar{\rho}}\right)\right]\right), \quad (38)$$

which improves the pointwise iteration-complexity bound $\mathcal{O}((Ld_0/\bar{\rho})^2)$ for the variant of Tseng's MFBS algorithm (see [8, Theorem 4.6]). Third, it is proved in [7, Theorem 6.2(b)] that the Tseng's MFBS variant finds an ergodic pair (b, y) satisfying $b \in (F + C)^\varepsilon(y)$, $\|y\| \leq \bar{\rho}$ and $\varepsilon \leq \bar{\varepsilon}$ in at most $\mathcal{O}(\max[Ld_0/\bar{\rho}, Ld_0^2/\bar{\varepsilon}])$ iterations. Note that the dependence of the latter bound on $\bar{\rho}$ differs from the one in (38) only by a logarithmic term. Moreover, in contrast to the latter bound, (38) does not depend on $\bar{\varepsilon}$. Also, the error criterion implied by the latter ergodic result is weaker than the one in (36). In summary, Theorem 4.2 establishes a pointwise iteration-complexity bound which closely approaches the latter ergodic bound while guaranteeing at the same time an error criterion stronger than the one for the aforementioned ergodic result.

We finish this section by noting that, if $C = \partial g$ where $g : X \rightarrow (-\infty, \infty]$ is a proper closed convex function, then an iteration of Korpelevich's extragradient algorithm (see for example Section 4 of [8]) can also be used to implement step 1 of Algorithm 2 and, as a consequence, yields a different instance of DR-HPE. Clearly, it is possible to derive a result for the new variant similar to Theorem 4.2 in which the error criterion becomes $b \in (F + \partial_\varepsilon g)(y)$, $\|b\| \leq \bar{\rho}$, $\varepsilon \leq \bar{\varepsilon}$ and the complexity bound is given by (24) (and hence depends on ε) with $\bar{\lambda} = \underline{\lambda} = \lambda := \sigma/L$. Note that the latter error criterion, while weaker than the one in (36), is still stronger than the one of the ergodic result for the Tseng's MBFS variant (see, for instance, [7, Corollary 5.3(b)]).

5 Acknowledgments

The authors would like to thank the two anonymous referees for their insightful comments on the first draft of the paper.

A Proof of Proposition 2.2

Throughout this section, $\{x_k\}$, $\{y_k\}$, $\{v_k\}$, $\{\lambda_k\}$, $\{\sigma_k\}$ and $\{\varepsilon_k\}$ denote the sequences generated by Algorithm 1. Also, for every $k \geq 1$, define

$$\gamma_k : X \rightarrow \mathbb{R}, \quad \gamma_k(x) := \langle v_k, x - y_k \rangle - \varepsilon_k \quad \forall x \in X \quad (39)$$

and

$$\theta_k := \left(\frac{1}{2\lambda_k\mu} + \frac{1}{1-\sigma^2} \right)^{-1} \in (0, 1). \quad (40)$$

Proposition A.1. *Let $x^* := (A + B)^{-1}(0)$ and let $\gamma_k(\cdot)$ and θ_k be as in (39) and (40), respectively. Then, for every $k \geq 1$:*

- (a) $x_k = \arg \min \lambda_k \gamma_k(x) + \|x - x_{k-1}\|^2/2$;
- (b) $\min \lambda_k \gamma_k(x) + \|x - x_{k-1}\|^2/2 \geq (1 - \sigma^2)\|y_k - x_{k-1}\|^2/2$;
- (c) $\gamma_k(x^*) \leq -\mu\|x^* - y_k\|^2$;
- (d) $\|x^* - x_{k-1}\|^2 \geq 2\lambda_k\mu\|x^* - y_k\|^2 + (1 - \sigma^2)\|y_k - x_{k-1}\|^2 + \|x^* - x_k\|^2$;
- (e) $(1 - \theta_k)\|x^* - x_{k-1}\|^2 \geq \|x^* - x_k\|^2$.

Proof. (a) This statement follows trivially from (10) and (39).

(b) Direct use of (10) and (39) yields after trivial algebraic manipulations that

$$\lambda_k \gamma_k(x_k) + \frac{1}{2}\|x_k - x_{k-1}\|^2 = \frac{1}{2} \left[\|y_k - x_{k-1}\|^2 - (\|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k) \right],$$

which, combined with (a) and (9), imply (b).

(c) Since $0 \in (A + B)(x^*)$, there exists $a^* \in A(x^*)$ such that $-a^* \in B(x^*)$. Also, the inclusion in (9) implies that there exist $a_k \in A(y_k)$ and $b_k \in B^{[\varepsilon_k]}(y_k)$ such that $v_k = a_k + b_k$. These inclusions, assumption A.1 and (7) then imply that

$$\langle a^* - a_k, x^* - y_k \rangle \geq \mu\|x^* - y_k\|^2, \quad \langle b_k + a^*, y_k - x^* \rangle \geq -\varepsilon_k.$$

To end the proof of (c), add these inequalities, observe that $a_k + b_k = v_k$, and use (39).

(d) It follows from (39), and statements (a) and (b), that

$$\begin{aligned} \lambda_k \gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 &= \left(\min \lambda_k \gamma_k(x) + \frac{1}{2}\|x - x_{k-1}\|^2 \right) + \frac{1}{2}\|x - x_k\|^2 \\ &\geq \frac{1}{2} \left((1 - \sigma^2)\|y_k - x_{k-1}\|^2 + \|x - x_k\|^2 \right) \quad \forall x \in X. \end{aligned}$$

Statement (d) now follows from the above inequality with $x = x^*$ and statement (c).

(e) This statement follows from $\|x^* - y_k\| + \|y_k - x_{k-1}\| \geq \|x^* - x_{k-1}\|$, the fact that

$$\min\{(1 - \sigma^2)r^2 + 2\mu\lambda_k s^2 \mid r, s \geq 0, r + s \geq \|x^* - x_{k-1}\|\} = \theta_k\|x^* - x_{k-1}\|^2,$$

and statement (d). □

The following Lemma follows trivially from the inequality in (9), the use of the triangle inequality and the fact that $\varepsilon_k \geq 0$.

Lemma A.2. *For every $k \geq 1$, we have $(1 - \sigma_k) \|y_k - x_{k-1}\| \leq \|\lambda_k v_k\| \leq (1 + \sigma_k) \|y_k - x_{k-1}\|$.*

In the next proposition, we establish rates of convergence for the sequences $\{x_k\}$, $\{v_k\}$ and $\{\varepsilon_k\}$ generated by Algorithm 1.

Proposition A.3. *Define $x^* := (A + B)^{-1}(0)$, $d_0 := \|x_0 - x^*\|$ and*

$$\Gamma_k := \left[\prod_{j=1}^k (1 - \theta_j) \right]^{1/2} \quad \forall k \geq 1. \quad (41)$$

Then, for every $k \geq 1$, $v_k \in A(y_k) + B^{[\varepsilon_k]}(y_k)$ and

$$\|v_k\| \leq \sqrt{\frac{1 + \sigma}{1 - \sigma}} \left(\frac{\Gamma_{k-1}}{\lambda_k} \right) d_0, \quad \varepsilon_k \leq \frac{\sigma^2}{2(1 - \sigma^2)} \left(\frac{\Gamma_{k-1}^2}{\lambda_k} \right) d_0^2, \quad (42)$$

$$\|x^* - x_k\| \leq \Gamma_k d_0, \quad \|x^* - y_k\| \leq \min \left\{ \frac{1}{\sqrt{2\lambda_k \mu}}, 1 + \frac{1}{\sqrt{1 - \sigma^2}} \right\} \Gamma_{k-1} d_0. \quad (43)$$

Proof. First note that the first inequality in (43) follows from Proposition A.1(e) and (41). Also, the second inequality in (43) follows from the inequality $\|y_k - x^*\| \leq \|y_k - x_{k-1}\| + \|x_{k-1} - x^*\|$, Proposition A.1(d) and the first inequality in (43) with $k = k - 1$. Using Proposition A.1(d) and the first inequality in (43), we conclude that

$$(1 - \sigma^2) \|y_k - x_{k-1}\|^2 \leq \Gamma_{k-1}^2 d_0^2 \quad \forall k \geq 1.$$

Note now that (42) follows from the latter inequality and the relations

$$\varepsilon_k \leq \frac{\sigma^2 \|y_k - x_{k-1}\|^2}{2\lambda_k}, \quad \|v_k\| \leq \frac{(1 + \sigma) \|y_k - x_{k-1}\|}{\lambda_k},$$

which are due to (9) and the second inequality in Lemma A.2. □

Proof of Proposition 2.2. The assumption $\lambda_k \geq \underline{\lambda} > 0$ for every $k \geq 1$ and the fact that the scalar function

$$t > 0 \mapsto \left(\frac{1}{2t\mu} + \frac{1}{1 - \sigma^2} \right)^{-1}$$

is nondecreasing, combined with (40) and (11), imply that $\theta_k \geq \theta$ for all $k \geq 1$. From the latter inequality and (41), we obtain $\Gamma_k \leq (1 - \theta)^{k/2}$ for every $k \geq 1$, which, in turn combined with Proposition A.3, completes the proof.

References

- [1] R. E. Bruck, Jr. A strongly convergent iterative solution of $0 \in U(x)$ for a maximal monotone operator U in Hilbert space. *J. Math. Anal. Appl.*, 48:114–126, 1974.
- [2] 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.
- [3] F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems, Volume I*. Springer-Verlag, New York, 2003.
- [4] G. M. Korpelevič. An extragradient method for finding saddle points and for other problems. *Èkonom. i Mat. Metody*, 12(4):747–756, 1976.
- [5] B. Martinet. Régularisation d’inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Ser. R-3):154–158, 1970.
- [6] G. J. Minty. Monotone (nonlinear) operators in Hilbert spaces. *Duke Math. J.*, 29(3):341–346, 1962.
- [7] 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 J. Optim.*, 20(6):2755–2787, 2010.
- [8] 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 J. Optim.*, 21(4):1688–1720, 2011.
- [9] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.*, 22(3):914–935, 2012.
- [10] 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.
- [11] A. Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15:229–251, 2005.
- [12] Y. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Math. Program.*, 109(2-3, Ser. B):319–344, 2007.
- [13] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.
- [14] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.
- [15] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

- [16] 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.
- [17] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.
- [18] 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.
- [19] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446 (electronic), 2000.