SIAM J. OPTIM.
Vol. 22, No. 3, pp. 914–935

ITERATION-COMPLEXITY OF A NEWTON PROXIMAL EXTRAGRADIENT METHOD FOR MONOTONE VARIATIONAL INEQUALITIES AND INCLUSION PROBLEMS*

RENATO D. C. MONTEIRO† AND BENAR F. SVAITER‡

Abstract. In a recent paper by Monteiro and Svaiter, a hybrid proximal extragradient (HPE) framework was used to study the iteration-complexity of a first-order (or, in the context of optimization, second-order) method for solving monotone nonlinear equations. The purpose of this paper is to extend this analysis to study a prox-type first-order method for monotone smooth variational inequalities and inclusion problems consisting of the sum of a smooth monotone map and a maximal monotone point-to-set operator. Each iteration of the method computes an approximate solution of a proximal subproblem obtained by linearizing the smooth part of the operator in the corresponding proximal equation for the original problem, which is then used to perform an extragradient step as prescribed by the HPE framework. Both pointwise and ergodic iteration-complexity results are derived for the aforementioned first-order method using corresponding results obtained here for a subfamily of the HPE framework.

Key words. complexity, extragradient, variational inequality, maximal monotone operator, proximal point, ergodic convergence, hybrid, Newton methods

AMS subject classifications. 90C60, 90C53, 90C33, 90C25, 47H05, 65K05, 65K10, 47J20

DOI. 10.1137/11083085X

1. Introduction.

A broad class of optimization, saddle point, equilibrium, and variational inequality (VI) problems can be posed as the *monotone inclusion problem*, namely, finding x such that $0 \in T(x)$, where T is a maximal monotone point-to-set operator. The proximal point method, proposed by Martinet [9] and further studied by Rockafellar [20, 21], is a classical iterative scheme for solving the monotone inclusion problem which generates a sequence $\{x_k\}$ according to

$$x_k = (\lambda_k T + I)^{-1}(x_{k-1}),$$

where $\lambda_k > 0$. It has been used as a generic framework for the design and analysis of several implementable algorithms. The classical inexact version of the proximal point method allows for the presence of a sequence of summable errors in the above iteration, i.e.,

$$\|x_k - (\lambda_k T + I)^{-1}(x_{k-1})\| \le e_k, \qquad \sum_{k=1}^{\infty} e_k < \infty.$$

Convergence results under the above error condition have been established in [21] and have been used in the convergence analysis of other methods that can be recast in the above framework [20].

*Received by the editors April 14, 2011; accepted for publication (in revised form) May 25, 2012; published electronically July 26, 2012.
http://www.siam.org/journals/siopt/22-3/83085.html
†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205 (monteiro@isye.gatech.edu). The work of this author was partially supported by NSF grants CCF-0808863 and CMMI-0900094 and ONR grant ONR N00014-11-1-0062.
‡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 474944/2010-7 and 303583/2008-8 and FAPERJ grant E-26/110.821/2008.

New inexact versions of the proximal point method with relative error tolerance were proposed by Solodov and Svaiter [22, 23, 25, 26]. Iteration complexity results for one of these inexact versions of the proximal point method introduced in [22], namely, the hybrid proximal extragradient (HPE) method, were established in [12]. Application of this framework in the iteration-complexity analysis of several zero-order (or, in the context of optimization, first-order) methods for solving monotone variational inequalities and monotone inclusion and saddle point problems are discussed in [12] and in the subsequent papers [13, 11].

The HPE framework was also used to study the iteration-complexity of a first-order (or, in the context of optimization, second-order) method for solving monotone nonlinear equations (see section 7 of [12]). The purpose of this paper is to extend this analysis to the more general context of monotone smooth variational inequalities and inclusion problems consisting of the sum of a smooth monotone map and a maximal monotone point-to-set operator. Each iteration of the first-order method computes an approximate solution of a proximal subproblem obtained by linearizing the smooth part of the operator in the corresponding proximal equation for the original problem, which is then used to perform an extragradient step as prescribed by the HPE framework. Both pointwise and ergodic iteration-complexity results are derived for the aforementioned first-order method using corresponding results obtained here for a subfamily of the HPE framework.

There have been other Newton-type methods in the context of degenerate unconstrained convex optimization problems for which complexity results have been derived. In [16], a Newton-type method for unconstrained convex optimization based on subproblems with cubic regularization terms is proposed and iteration-complexity results are obtained. An accelerated version of this method is studied in [15]. It should be mentioned that these methods are specifically designed for (unconstrained) convex optimization and hence do not apply to the problems studied in this paper.

This paper is organized as follows. Section 2, introduces a subfamily of the HPE framework, namely, the large-step HPE method, which will play an important role on the analysis of a first-order (or Newton-type) proximal method, namely, the Newton proximal extragradient (NPE) method. Section 3 describes the inexact NPE method and corresponding iteration-complexity results. Section 4.2 discusses a line search procedure to compute a stepsize satisfying a certain large-step condition required by the NPE method.

Notation. Throughout this paper, $\mathbb{E}$ will denote a finite dimensional inner product real vector space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. For a nonempty closed convex set $\Omega \subseteq \mathbb{E}$, we denote the orthogonal projection operator onto Ω by P_Ω. We let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote for the set of nonnegative and positive real numbers, respectively. The domain of definition of a point-to-point function F is denoted by $\operatorname{Dom} F$. Similar notation is used for the domain of a point-to-set map $F : \mathbb{E} \rightrightarrows \mathbb{E}$, which is defined in section 2.1.

2. The large-step HPE method. This section introduces a subfamily of the HPE method, namely, the large-step HPE method, which will play an important role on the analysis of the Newton-type proximal methods discussed on section 3. It consists of two subsections as follows. The first one introduces some basic concepts and the definition and basic facts about the ε-enlargement of a monotone multivalued map. The second subsection introduces the large-step HPE method and gives specialized iteration-complexity results for it, which are stronger than the corresponding ones applicable for the general HPE framework.

2.1. The ε-subdifferential and ε-enlargement of monotone operators. A point-to-set operator $T : \mathbb{E} \rightrightarrows \mathbb{E}$ is a relation $T \subseteq \mathbb{E} \times \mathbb{E}$ and

$$T(z) = \{v \in \mathbb{E} \mid (z, v) \in T\}.$$

Alternatively, one can consider T as a multivalued function of $\mathbb{E}$ into the family $\wp(\mathbb{E}) = 2^{(\mathbb{E})}$ of subsets of $\mathbb{E}$. Regardless of the approach, it is usual to identify T with its graph defined as

$$Gr(T) = \{(z, v) \in \mathbb{E} \times \mathbb{E} \mid v \in T(z)\}.$$

The domain of T, denoted by $\operatorname{Dom} T$, is defined as

$$\operatorname{Dom} T := \{z \in \mathbb{E} : T(z) \neq \emptyset\}.$$

The operator $T : \mathbb{E} \rightrightarrows \mathbb{E}$ is *monotone* if

$$\langle v - \tilde{v}, z - \tilde{z} \rangle \geq 0 \qquad \forall (z, v), (\tilde{z}, \tilde{v}) \in Gr(T),$$

and T is *maximal monotone* if it is monotone and maximal in the family of monotone operators with respect to the partial order of inclusion, i.e., $S : \mathbb{E} \rightrightarrows \mathbb{E}$ monotone and $Gr(S) \supset Gr(T)$ implies that $S = T$.

In [4], Burachik, Iusem, and Svaiter introduced the ε-enlargement of maximal monotone operators. In [12] this concept was extended to a generic point-to-set operator in $\mathbb{E}$ as follows. Given $T : \mathbb{E} \rightrightarrows \mathbb{E}$ and a scalar ε, define $T^\varepsilon : \mathbb{E} \rightrightarrows \mathbb{E}$ as

$$T^\varepsilon(z) = \{v \in \mathbb{E} \mid \langle z - \tilde{z}, v - \tilde{v} \rangle \geq -\varepsilon \quad \forall \tilde{z} \in \mathbb{E}\, \forall \tilde{v} \in T(\tilde{z})\} \quad \forall z \in \mathbb{E}. \tag{1}$$

We now state a few useful properties of the operator T^ε that will be needed in our presentation.

PROPOSITION 2.1. *Let $T, T' : \mathbb{E} \rightrightarrows \mathbb{E}$. Then,*

(a) *if $\varepsilon_1 \leq \varepsilon_2$, then $T^{\varepsilon_1}(z) \subseteq T^{\varepsilon_2}(z)$ for every $z \in \mathbb{E}$;*

(b) *$T^\varepsilon(z) + (T')^{\varepsilon'}(z) \subseteq (T + T')^{\varepsilon + \varepsilon'}(z)$ for every $z \in \mathbb{E}$ and $\varepsilon, \varepsilon' \in \mathbb{R}$;*

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

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

PROPOSITION 2.2 (see [7, Corollary 3.8(ii)]). *Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ be a maximal monotone operator. Then, $\operatorname{Dom} T^\varepsilon \subseteq \operatorname{cl}(\operatorname{Dom} T)$ for any $\varepsilon \geq 0$.*

The ε-subdifferential [3] of a function $f : \mathbb{E} \to \bar{\mathbb{R}}$ is defined as

$$\partial_\varepsilon f(z) = \{v \in \mathbb{E} \mid f(\tilde{z}) \geq f(z) + \langle \tilde{z} - z, v \rangle - \varepsilon \ \forall \tilde{z} \in \mathbb{E}\} \quad \forall z \in \mathbb{E}.$$

When $\varepsilon = 0$, the above operator is simply denoted by ∂f and is referred to as the subdifferential of f. If f is a closed convex function, it is well-known that ∂f is maximal monotone (see [19]); moreover, we have

$$\partial_\varepsilon f(z) \subseteq (\partial f)^\varepsilon(z) \qquad \forall z \in \mathbb{E} \tag{2}$$

and for many closed convex functions f this inclusion is proper.

Let X be a nonempty set in $\mathbb{E}$. Its *indicator function* $\delta_X : \mathbb{E} \to \bar{\mathbb{R}}$ is defined as

$$\delta_X(x) = \begin{cases} 0, & x \in X, \\ \infty & \text{otherwise,} \end{cases}$$

and its *normal cone operator* is the point-to-set map $N_X : \mathbb{E} \rightrightarrows \mathbb{E}$ given by

$$N_X(x) = \begin{cases} \emptyset, & x \notin X, \\ \{v \in \mathbb{E} \mid \langle y - x, v \rangle \le 0 \ \forall y \in X\}, & x \in X. \end{cases} \tag{3}$$

Clearly, the normal cone operator N_X of X can be expressed in terms of δ_X as $N_X = \partial \delta_X$.

Given T maximal monotone, a natural question is how to obtain points in the graph of T^ε. We mention two practical ways of obtaining points in the graph of T^ε. First, if $T = S + N_C$, where C is a nonempty closed convex set and S is a maximal monotone operator, then it follows from Proposition 2.1(b) and (2) that $(S + \partial_\varepsilon \delta_C)(x) \subseteq T^\varepsilon(x)$ for every $x \in \mathbb{E}$. This observation is used in the second paragraph following Definition 3.1 . Second, as a consequence of the weak transportation formula developed in [6], it follows that convex combinations of elements in the graph of T yield an element in the graph of T^ε for some $\varepsilon \ge 0$. Based on this construction, a bundle-like method for finding a zero of a bounded maximal monotone operator T is developed in [6, 5] under the assumption that a black-box is available to supply an element in $T(x)$ for any given $x \in \mathbb{E}$.

2.2. The large-step HPE method. This subsection introduces the large-step HPE method and gives specialized iteration-complexity results for it.

Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ be a maximal monotone operator. The monotone inclusion problem for T consists of finding $x \in \mathbb{E}$ such that

$$0 \in T(x)\,. \tag{4}$$

We also assume throughout this section that this problem has a solution, that is, $T^{-1}(0) \neq \emptyset$.

For the monotone inclusion problem (4), the exact proximal point iteration from x with stepsize $\lambda > 0$ is the unique solution y of the inclusion

$$0 \in (\lambda T + I)(y) - x = \lambda T(y) + y - x, \tag{5}$$

or equivalently, the y-component of the unique solution (y, v) of the inclusion and equation

$$v \in T(y), \qquad \lambda v + y - x = 0. \tag{6}$$

The method we are interested in studying in this section is based on the following notion of approximate solution of (6) introduced in [22].

DEFINITION 2.3. *Given $\hat{\sigma} \ge 0$, the triple (y, v, ε) is said to be a $\hat{\sigma}$-approximate solution of* (6) *at* (λ, x) *if*

$$v \in T^\varepsilon(y), \quad \|\lambda v + y - x\|^2 + 2\lambda\varepsilon \le \hat{\sigma}^2 \|y - x\|^2. \tag{7}$$

Observe that the error criterion (7) relaxes the inclusion in (6) to $v \in T^\varepsilon(y)$ and relaxes the equation in (6) by requiring its residual $r = \lambda v + y - x$ and the tolerance ε to be small relative to $\|y - x\|$.

We will now state the method that will be the main subject of study in this section.

Large-step HPE Method:

(0) Let $x_0 \in \mathbb{E}$, $\theta > 0$, and $0 \le \sigma < 1$ be given and set $k = 1$.

(1) If $0 \in T(x_{k-1})$, then **stop**. Otherwise, choose stepsize $\lambda_k > 0$, tolerance $\sigma_k \in [0, \sigma]$, and σ_k-approximate solution $(y_k, v_k, \varepsilon_k)$ of (6) at (λ_k, x_{k-1}), i.e.,

$$v_k \in T^{\varepsilon_k}(y_k), \quad \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \le \sigma_k^2 \|y_k - x_{k-1}\|^2, \tag{8}$$

such that

$$\lambda_k \|y_k - x_{k-1}\| \ge \theta. \tag{9}$$

(2) Define $x_k = x_{k-1} - \lambda_k v_k$, set $k \leftarrow k+1$, and go to step 1.

end

We now make several remarks about the large-step HPE method. First, the large-step HPE method is a special case of the HPE method which was introduced in [22] and whose corresponding iteration-complexity results were studied in [12]. Indeed, deleting condition (9) from the above method and the stopping criterion in step 1, one obtains the HPE method as stated in [22, 12]. Second, since the condition that $0 \in T(x_{k-1})$ in the HPE method implies that $x_l = x_{k-1} \in T^{-1}(0)$ for every $l \ge k$, the HPE method can also be alternatively stated with the stopping criterion $0 \in T(x_{k-1})$ as above. Third, step 1 of the large-step HPE method does not specify how to choose λ_k and $(y_k, v_k, \varepsilon_k)$ satisfying the prescribed conditions. The particular choice of λ_k and $(y_k, v_k, \varepsilon_k)$ will depend on the particular implementation of the method and the properties of the operator T. Fourth, condition (9) forces the stepsize λ_k to be large in some sense. Existence of λ_k and $(y_k, v_k, \varepsilon_k)$ as in step 1 can be shown using Lemma 4.3(b) as will be discussed later on. Finally, as we will see in the next section, the large-step HPE method provides a useful framework for designing and analyzing Newton-type methods in the context of the monotone variational inequality problem and its generalizations.

Clearly, the HPE method heavily relies on the notion of approximate solutions of (6) as in Definition 2.3. The use of this general notion have allowed us to show that several well-known methods for variational inequalities and convex optimization can in fact be viewed as special cases of the HPE method, and hence as inexact proximal point methods. These include, for example, the following: (i) a variant of Tseng's modified forward-backward splitting method; (ii) Korpelevich's method; (iii) the classical forward-backward splitting method for convex optimization; (iv) the alternating direction method of multipliers; and (v) the Douglas–Rachford's splitting method (see [13, 10, 11, 12, 22]). In addition, the use of the notion of approximate solution as in Definition 2.3 have allowed us to obtain new block-decomposition methods (see [11]). In particular, one such block-decomposition method for solving conic semidefinite programming has been shown in [14] to outperform the two most competitive codes for large-scale conic semidefinite programs, namely, the boundary point method introduced by Povh, Rendl, and Wiegele [18] and the Newton-CG augmented Lagrangian method by Zhao, Sun, and Toh [27]. Moreover, the use of the tolerances ε_k in the definition of approximate solutions of (6) has played an important role in the analysis of Korpelevich's method given in [13, 12] as well as the forward-backward method for convex optimization presented in [10].

Our main goal in the remaining part of this section will be to study the iteration-complexity of the large-step HPE method. For simplicity of exposition, the convergence rate results presented below implicitly assume that $0 \notin T(x_{k-1})$ for every $k \ge 1$. However, they can easily be restated without assuming such condition by saying that either the conclusion stated below holds or x_{k-1} is a solution of (4).

The proof of the following result can be found in Lemma 4.2 of [12].

PROPOSITION 2.4. *For any $x^* \in T^{-1}(0)$, the sequence $\{\|x^* - x_k\|\}$ is nonincreasing and*

$$\|x^* - x_0\|^2 \geq \sum_{k=1}^{\infty}(1-\sigma_k^2)\,\|y_k - x_{k-1}\|^2 \geq (1-\sigma^2)\sum_{k=1}^{\infty}\|y_k - x_{k-1}\|^2 . \tag{10}$$

While the general complexity results for the HPE framework obtained in [12] are expressed in terms of the sequence of stepsizes λ_k, the ones derived below for the large-step HPE are expressed in terms of the iteration count k and heavily depend on the large-step condition (9).

The first complexity result estimates the quality of the best among the iterates $y_1, \ldots, y_k$ generated by the large-step HPE method. We will refer to these estimates as *pointwise* iteration-complexity bounds.

THEOREM 2.5. *Consider the sequences $\{x_k\}$, $\{y_k\}$, $\{v_k\}$, and $\{\varepsilon_k\}$ generated by the large-step HPE method. Then, for every $k \geq 1$, $v_k \in T^{\varepsilon_k}(y_k)$ and there exists an index $i \leq k$ such that*

$$\|v_i\| \leq \frac{d_0^2}{\theta(1-\sigma)\,k}, \qquad \varepsilon_i \leq \frac{\sigma^2 d_0^3}{2\theta(1-\sigma^2)^{3/2}k^{3/2}},$$

where d_0 is the distance of x_0 to $T^{-1}(0)$.

Proof. Let x^* be such that $d_0 = \|x_0 - x^*\|$. First observe that Proposition 2.4 implies that for every $k \in \mathbb{N}$ there exists $i \leq k$ such that

$$\|y_i - x_{i-1}\|^2 \leq \frac{d_0^2}{k(1-\sigma^2)}. \tag{11}$$

Using the fact that $(y_i, v_i, \varepsilon_i)$ is a σ_i-approximate solution of (6) at (λ_i, x_{i-1}), i.e., that (8) holds with $i = k$, and the fact that $\sigma_i \leq \sigma$, we conclude that

$$\lambda_i\|v_i\| \leq \|\lambda_i v_i + y_i - x_{i-1}\| + \|y_i - x_{i-1}\| \leq (1+\sigma)\|y_i - x_{i-1}\|$$

and

$$2\lambda_i\varepsilon_i \leq \sigma^2\|y_i - x_{i-1}\|^2.$$

Multiplying the above two inequalities by $\|y_i - x_{i-1}\|$ and using (9), we conclude that

$$\theta\|v_i\| \leq (1+\sigma)\|y_i - x_{i-1}\|^2, \qquad 2\theta\varepsilon_i \leq \sigma^2\|y_i - x_{i-1}\|^3.$$

The conclusion now follows immediately from (11) and the above two inequalities. ◻

We will now describe alternative estimates for the large-step HPE method which we refer to as the *ergodic* iteration-complexity bounds. The sequence of ergodic means $\{\bar{y}_k\}$ associated with $\{y_k\}$ is

$$\bar{y}_k := \frac{1}{\Lambda_k}\sum_{i=1}^{k}\lambda_i y_i, \qquad \text{where} \qquad \Lambda_k := \sum_{i=1}^{k}\lambda_i. \tag{12}$$

Define also

$$\bar{v}_k := \frac{1}{\Lambda_k}\sum_{i=1}^{k}\lambda_i v_i, \qquad \bar{\varepsilon}_k := \frac{1}{\Lambda_k}\sum_{i=1}^{k}\lambda_i(\varepsilon_i + \langle y_i - \bar{y}_k, v_i - \bar{v}_k\rangle). \tag{13}$$

The following result describes convergence rate bounds for the ergodic sequence $\{\bar{y}_k\}$ generated by an arbitrary HPE method. Its proof follows immediately from Proposition 4.6 and the proof of Theorem 4.7 of [12].

PROPOSITION 2.6. *For every $k \in \mathbb{N}$,*

$$0 \le \bar{\varepsilon}_k \le \frac{1}{2\Lambda_k}\left[2\langle \bar{y}_k - x_0, x_k - x_0\rangle - \|x_k - x_0\|^2\right] \le \frac{2\eta_k d_0^2}{\Lambda_k} \tag{14}$$

and

$$\bar{v}_k = \frac{1}{\Lambda_k}(x_0 - x_k) \in T^{\bar{\varepsilon}_k}(\bar{y}_k), \qquad \|\bar{v}_k\| \le \frac{2d_0}{\Lambda_k},$$

where d_0 is the distance of x_0 to $T^{-1}(0)$ and

$$\eta_k := 1 + \frac{\sigma\sqrt{\tau_k}}{\sqrt{(1-\sigma^2)}}, \qquad \tau_k = \max_{i=1,\dots,k} \frac{\lambda_i}{\Lambda_k} \le 1. \tag{15}$$

The following result refines the bounds of Proposition 2.6 in the particular context of the large-step HPE method.

THEOREM 2.7. *Let $\{\lambda_k\}$, $\{\varepsilon_k\}$, $\{y_k\}$, $\{v_k\}$, and $\{x_k\}$ be the sequences generated by the large-step HPE method and consider the ergodic sequences $\{\bar{y}_k\}$, $\{\bar{v}_k\}$, and $\{\bar{\varepsilon}_k\}$ defined according to* (12) *and* (13). *Then, for every $k \ge 1$, $\bar{v}_k \in T^{\bar{\varepsilon}_k}(\bar{y}_k)$ and*

$$\|\bar{v}_k\| \le \frac{2d_0^2}{\theta\sqrt{1-\sigma^2}\,k^{3/2}}, \qquad 0 \le \bar{\varepsilon}_k \le \frac{2\eta_k d_0^3}{\theta\sqrt{1-\sigma^2}\,k^{3/2}},$$

where $\eta_k \le 1 + \sigma/\sqrt{1-\sigma^2}$ is defined in Proposition 2.6.

Proof. First note that (9) and Proposition 2.4 imply that

$$\sum_{i=1}^{\infty} \frac{1}{\lambda_i^2} \le \frac{1}{\theta^2}\sum_{i=1}^{\infty} \|y_i - x_{i-1}\|^2 \le \frac{d_0^2}{\theta^2(1-\sigma^2)}. \tag{16}$$

Noting that the minimum value of $\sum_{i=1}^k t_i$ subject to the condition that $\sum_{i=1}^k t_i^{-2} \le C$ and $t_i > 0$ for $i = 1, \dots, k$ is equal to $k^{3/2}/\sqrt{C}$, we then conclude that

$$\sum_{i=1}^{k} \lambda_i \ge \frac{\theta\sqrt{1-\sigma^2}}{d_0}\,k^{3/2}.$$

The conclusion of the theorem now follows immediately from the above inequality and Proposition 2.6. ∎

3. An inexact Newton proximal extragradient method. In this section, we consider the monotone inclusion problem

$$0 \in T(x) = (F + H)(x), \tag{17}$$

where $F : \operatorname{Dom} F \subseteq \mathbb{E} \to \mathbb{E}$ and $H : \mathbb{E} \rightrightarrows \mathbb{E}$ satisfy the following:

(C.1) H is a maximal monotone operator;

(C.2) F is monotone and differentiable on a closed convex set Ω such that $\operatorname{Dom} H \subseteq \Omega \subseteq \operatorname{Dom} F$;

(C.3) F' is L-Lipschitz continuous on Ω, i.e.,

$$\|F'(\tilde{x}) - F'(x)\| \leq L\|\tilde{x} - x\| \quad \forall x, \tilde{x} \in \Omega, \tag{18}$$

where the norm on the left-hand side is the operator norm.

Recall that for the monotone inclusion problem (17) the exact proximal iteration from x with stepsize $\lambda > 0$ is the unique solution y of the inclusion

$$0 \in \lambda(F + H)(y) + y - x, \tag{19}$$

or equivalently, the y-component of the unique solution (y, v) of the inclusion/equation

$$v \in (F + H)(y), \qquad \lambda v + y - x = 0. \tag{20}$$

Our inexact NPE method is based on the computation of an inexact solution of a linearized approximation of the above inclusion/equation which is described next.

For $\bar{x} \in \Omega$, define the first-order approximation $T_{\bar{x}} : \mathbb{E} \rightrightarrows \mathbb{E}$ of T at $\bar{x}$ as

$$T_{\bar{x}}(x) = F_{\bar{x}}(x) + H(x),$$

where $F_{\bar{x}} : \mathbb{E} \to \mathbb{E}$ is the usual first-order approximation of F given by

$$F_{\bar{x}}(x) = F(\bar{x}) + F'(\bar{x})(x - \bar{x}).$$

The following definition gives the precise notion of approximate solution used by the inexact NPE method.

Definition 3.1. *Given $(\lambda, x) \in \mathbb{R}_{++} \times \mathbb{E}$ and $\hat{\sigma} \geq 0$, the triple $(y, u, \varepsilon) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ is called a $\hat{\sigma}$-approximate Newton solution of* (20) *at (λ, x) if*

$$u \in (F_{x'} + H^{\varepsilon})(y), \qquad \|\lambda u + y - x\|^2 + 2\lambda\varepsilon \leq \hat{\sigma}^2\|y - x\|^2,$$

where $x' := P_\Omega(x)$.

We now make a few observation regarding the above definition. First, observe that any $\hat{\sigma}$-approximate Newton solution of (20) at (λ, x) is a $\hat{\sigma}$-approximate solution of (6) with $T = T_{x'}$ at (λ, x). Second, (y, u, ε) is a 0-approximate Newton solution of (20) at (λ, x) if and only if y is the (unique) solution the Newton proximal inclusion

$$0 \in \lambda(F_{x'} + H)(y) + y - x = (\lambda T_{x'} + I)(y) - x \tag{21}$$

obtained by linearizing (19), $u = (x - y)/\lambda$, and $\varepsilon = 0$. Hence, the *exact* solution y of the Newton proximal inclusion (21) trivially supplies a $\hat{\sigma}$-approximate Newton solution of (20) at (λ, x) for any $\sigma \geq 0$. Third, the y-component of a $\hat{\sigma}$-approximate Newton solution always belongs to $\operatorname{Dom} H^{\varepsilon} \subseteq \operatorname{cl}(\operatorname{Dom} H) \subseteq \Omega$, where the two inclusions are due to Proposition 2.2 and assumption C.2.

As mentioned previously, the inexact NPE method stated below assumes that for any given pair $(\lambda, x) \in \mathbb{R}_{++} \times \mathbb{E}$ one is able to compute a $\hat{\sigma}$-approximate Newton solution of (20) at (λ, x). Before stating the NPE method, we discuss how this can be accomplished in the particular case where H is the normal cone operator N_C of a nonempty closed convex set $C \subseteq \mathbb{E}$. First note that in this case (21) is equivalent to the affine (strongly monotone) variational inequality $VI(G_x; C)$, i.e., the problem of finding $y \in \mathbb{E}$ such that

$$y \in C, \quad \langle G_x(y), z - y\rangle \geq 0 \quad \forall z \in C, \tag{22}$$

where $G_x : \mathbb{E} \to \mathbb{E}$ is the affine and strongly monotone map defined as

$$G_x(y) := \lambda F_{x'}(y) + y - x \quad \forall y \in \mathbb{E}.$$

It turns out that the regularized gap function [1, 8] for $VI(G_x; C)$ defined as

$$f_{\alpha,G_x,C}(y) = \langle G_x(y), y - P_C(y - \alpha G_x(y))\rangle - \frac{1}{2\alpha}\|y - P_C(y - \alpha G_x(y))\|^2$$

with $\alpha = 1$ provides a computable sufficient criterion for checking whether y yields a $\hat{\sigma}$-approximate Newton solution of (20) at (λ, x). More specifically, due to [24, Lemma 4], if $y \in C$ satisfies

$$2f_{1,G_x,C}(y) \le \hat{\sigma}^2\|y - x\|^2, \tag{23}$$

then the triple (y, u, ε), where

$$u = \frac{1}{\lambda}[x - P_C(y - G_x(y))], \qquad \varepsilon = \frac{1}{\lambda}\langle r, G_x(y) - r\rangle, \qquad r = y - P_C(y - G_x(y))$$

is a $\hat{\sigma}$-approximate Newton solution of (20) at (λ, x). Indeed, it has be shown in [24, Lemma 4] that

$$u \in (F_{x'} + \partial_\varepsilon \delta_C)(y), \qquad r = \lambda u + y - x, \qquad 2f_{1,G_x,C}(y) = \|r\|^2 + 2\lambda\varepsilon,$$

which together with (23) implies the desired conclusion. Any method for solving affine (strongly) monotone variational inequalities, e.g., interior point methods whenever C is endowed with a computable self-concordant function [17, Chapter 7], may be used to obtain approximate solutions of (22) according to (23).

We are now ready to state the inexact NPE method.

Inexact NPE Method:

(0) Let $x_0 \in \mathbb{E}$, $\hat{\sigma} \ge 0$, and $0 < \sigma_\ell < \sigma_u$ such that

$$\sigma := \hat{\sigma} + \sigma_u < 1, \qquad \sigma_\ell(1 + \hat{\sigma}) < \sigma_u(1 - \hat{\sigma}) \tag{24}$$

be given and set $k = 1$.

(1) Compute $x'_{k-1} = P_\Omega(x_{k-1})$. If $0 \in T(x'_{k-1})$, then STOP.

(2) Otherwise, compute stepsize $\lambda_k > 0$ and a $\hat{\sigma}$-approximate Newton solution $(y_k, u_k, \varepsilon_k)$ of (20) at (λ_k, x_{k-1}), i.e.,

$$u_k \in (F_{x'_{k-1}} + H^{\varepsilon_k})(y_k), \quad \|\lambda_k u_k + y_k - x_{k-1}\|^2 + 2\lambda_k\varepsilon_k \le \hat{\sigma}^2\|y_k - x_{k-1}\|^2 \tag{25}$$

such that

$$\frac{2}{L}\sigma_\ell \le \lambda_k\|y_k - x_{k-1}\| \le \frac{2}{L}\sigma_u. \tag{26}$$

(3) Set

$$v_k := F(y_k) + u_k - F_{x'_{k-1}}(y_k), \qquad x_k := x_{k-1} - \lambda_k v_k, \tag{27}$$

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

end

Our goal in the remaining part of this section will be to study the iteration-complexity of the inexact NPE method. Our first step will be to show that the inexact NPE method is a special case of the large-step HPE method. As a consequence, iteration-complexity results for the first method will be derived from the ones obtained for the latter method in section 2.

LEMMA 3.2. *For every $y \in \Omega$ and $x \in \mathbb{E}$,*

$$\|F(y) - F_{x'}(y)\| \leq \frac{L}{2}\|y - x\|^2,$$

where $x' := P_\Omega(x)$.

Proof. The following result follows as an immediate consequence of assumptions C.2 and C.3 and the fact that $x' = P_\Omega(x)$ and $y \in \Omega$ and of a well-known property of the projection operator P_Ω. ∎

LEMMA 3.3. *Let $(\lambda, x) \in \mathbb{R}_{++} \times \mathbb{E}$ and a $\hat{\sigma}$-approximate Newton solution (y, u, ε) of (20) at (λ, x) be given, and define $v := F(y) + u - F_{x'}(y)$. Then,*

$$v \in (F+H^\varepsilon)(y) \subseteq T^\varepsilon(y), \qquad \|\lambda v+y-x\|^2+2\lambda\varepsilon \leq \left(\hat{\sigma} + \frac{L\lambda}{2}\|y - x\|\right)^2 \|y-x\|^2. \tag{28}$$

Proof. The assumption on (y, u, ε) and Definition 3.1 imply that $u - F_{x'}(y) \in H^\varepsilon(y)$. This together with the definition of v imply the first inclusion in (28), while Proposition 2.1 implies the second inclusion in (28). To simplify the proof of the inequality in (28), define

$$r := \lambda u + y - x, \qquad r' := \lambda[F(y) - F_{x'}(y)],$$

and note that the definition of v implies that

$$\lambda v + y - x = r + r'. \tag{29}$$

Lemma 3.2, the assumption that (y, u, ε) is a $\hat{\sigma}$-approximate Newton solution of (20) at (λ, x), and the definition of r and r' imply that

$$\|r\|^2 + 2\lambda\varepsilon \leq \hat{\sigma}^2\|y - x\|^2, \qquad \|r'\| \leq \frac{\lambda L}{2}\|y - x\|^2.$$

Using the three last relations, we then conclude that

$$\begin{aligned} \|\lambda v + y - x\|^2 + 2\lambda\varepsilon &= \|r + r'\|^2 + 2\lambda\varepsilon \leq \|r\|^2 + \|r'\|^2 + 2\|r\|\,\|r'\| + 2\lambda\varepsilon \\ &\leq \left[\hat{\sigma}^2 + \left(\frac{L\lambda}{2}\|y - x\|\right)^2 + 2\hat{\sigma}\left(\frac{L\lambda}{2}\|y - x\|\right)\right]\|y - x\|^2, \end{aligned}$$

and hence that the inequality in (28) holds. ∎

Define for each k

$$\sigma_k := \hat{\sigma} + \frac{L}{2}\lambda_k\|y_k - x_{k-1}\|. \tag{30}$$

We will now establish that the inexact NPE method can be viewed as a special case of the large-step HPE method.

LEMMA 3.4. *Let σ be defined as in (24). Then, for each k, $\sigma_k \leq \sigma$ and*

$$v_k \in \left(F + H^{\varepsilon_k}\right)(y_k) \subseteq T^{\varepsilon_k}(y_k), \qquad \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k\varepsilon_k \leq \sigma_k^2\|y_k - x_{k-1}\|^2. \tag{31}$$

As a consequence of (26) *and* (27), *it follows that the inexact NPE method is a special case of the large-step HPE method stated in section* 2 *with* $\theta = 2\sigma_\ell/L$.

Proof. The bound on σ_k follows from (24), (26), and (30). The relations in (31) follow immediately from (30) and Lemma 3.3 with $(\lambda, x) = (\lambda_k, x_{k-1})$ and $(y, u, \varepsilon) = (y_k, u_k, \varepsilon_k)$. ∎

The next result gives the pointwise iteration-complexity bound for the inexact NPE method.

THEOREM 3.5. *Consider the sequences* $\{x_k\}$ *and* $\{y_k\}$ *generated by the inexact NPE method and the sequence* $\{v_k\}$ *defined according to* (27). *Then, for every* $k \geq 1$, $v_k \in (F + H^{\varepsilon_k})(y_k)$ *and there exists an index* $i \leq k$ *such that*

$$\|v_i\| \leq \frac{Ld_0^2}{2\sigma_\ell(1-\sigma)\,k}, \qquad \varepsilon_i \leq \frac{\sigma^2 L d_0^3}{4\sigma_\ell(1-\sigma)^{3/2}\,k^{3/2}},$$

where σ *is given by* (24).

Proof. This result follows immediately from Theorem 2.5 and Lemma 3.4. ∎

It is worth comparing the above complexity bounds for the first-order NPE method to corresponding bounds for zero-order methods, such as Korpelevich's or Tseng's modified forward-backward splitting methods, for solving variational inequalities and inclusion problems of the type (17) obtained in [13, 12]. Indeed, the above bounds for v_i and ε_i are both better than the corresponding ones for the zero-order methods by a factor of $k^{-1/2}$.

The next result describes the ergodic iteration-complexity bound for the inexact NPE method.

THEOREM 3.6. *Let* $\{\lambda_k\}$, $\{y_k\}$, *and* $\{x_k\}$ *be the sequences generated by the inexact NPE method and consider the sequence* $\{v_k\}$ *defined by* (27) *and the ergodic sequences* $\{\bar{y}_k\}$, $\{\bar{v}_k\}$, *and* $\{\bar{\varepsilon}_k\}$ *defined according to* (12) *and* (13). *Then, for every* $k \geq 1$, $\bar{v}_k \in T^{\bar{\varepsilon}_k}(\bar{y}_k)$ *and*

$$\|\bar{v}_k\| \leq \frac{Ld_0^2}{k^{3/2}\sigma_\ell\sqrt{1-\sigma^2}}, \qquad 0 \leq \bar{\varepsilon}_k \leq \frac{\eta_k L d_0^3}{k^{3/2}\sigma_\ell\sqrt{1-\sigma^2}},$$

where σ *is given by* (24) *and* $\eta_k \leq 1 + \sigma/\sqrt{1-\sigma^2}$ *is defined in Proposition* 2.6.

Proof. This result follows immediately from Lemma 3.4 and Theorem 2.7 with $\theta = 2\sigma_\ell/L$. ∎

Finally, we mention that the above bounds for $\bar{v}_k$ and $\bar{\varepsilon}_k$ are both better than the corresponding ones for the zero-order methods obtained in [13, 12] by a factor of $k^{-1/2}$.

4. Search procedure for the stepsize $\boldsymbol{\lambda}$. The main goal of this section is to present a search procedure for computing a stepsize λ_k and an $\hat{\sigma}$-approximate Newton solution $(y_k, u_k, \varepsilon_k)$ of (20) at (λ_k, x_{k-1}) as in step 2 of the inexact NPE method.

More specifically, to simplify notation, let $x = x_{k-1}$, $x' = x'_{k-1}$, $\alpha_- := 2\sigma_\ell/L$, and $\alpha_+ := 2\sigma_u/L$. In terms of this notation, our main goal is to compute a stepsize $\lambda > 0$ and a triple $(y, u, \varepsilon) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ such that

$$u \in (F_{x'} + H^{\varepsilon})(y), \qquad \|\lambda u + y - x\|^2 + 2\lambda\varepsilon \leq \hat{\sigma}^2\|y - x\|^2, \tag{32}$$

and

$$\alpha_- \leq \lambda\|y - x\| \leq \alpha_+. \tag{33}$$

Assuming that we have at our disposal a black-box, which for any given $(\lambda, x) \in \mathbb{R}_{++} \times \mathbb{E}$ finds $(y, u, \varepsilon) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ satisfying (32), our main goal in this section is to present an iterative procedure that will solve the problem posed above.

It turns out that such a goal can be posed in a more general setting. Noting that $(F_{x'} + H^\varepsilon)(y) \subseteq T^\varepsilon_{x'}(y)$, it follows that (32) implies in particular that (y, u, ε) is a $\hat{\sigma}$-approximate solution of

$$u \in A(y), \qquad \lambda u + y - x = 0 \tag{34}$$

with $A = T_{x'}$. Hence, the above goal is a special case of the following more general goal with respect to an arbitrary maximal monotone operator $A : \mathbb{E} \rightrightarrows \mathbb{E}$, point $x \in \mathbb{E}$, and bounds $0 < \alpha_- < \alpha_+$.

General goal. Assuming that we have at our disposal a black-box, which for any given $\lambda > 0$ finds a $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x), the goal is to find a specific $\lambda > 0$ and an associated $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x) (see Definition 2.3) such that (33) holds as well.

This section contains three subsections. Subsection 4.2 presents an iterative procedure which solves the general goal posed above. Along the way, we present some preliminary results in subsection 4.1. Subsection 4.3 specializes the complexity results obtained in subsection 4.2 to the special context of the main goal of this section, i.e., that of finding a stepsize $\lambda_k > 0$ and an $\hat{\sigma}$-approximate Newton solution $(y_k, u_k, \varepsilon_k)$ of (20) at (λ_k, x_{k-1}) as in step 2 of the inexact NPE.

4.1. Preliminary results. Let a maximal monotone operator $A : \mathbb{E} \rightrightarrows \mathbb{E}$ and $x \in \mathbb{E}$ be given and define for each $\lambda > 0$,

$$y_A(\lambda; x) := (I + \lambda A)^{-1}(x), \qquad \varphi_A(\lambda; x) := \lambda \|y_A(\lambda; x) - x\|. \tag{35}$$

In this subsection, we study several basic properties of $y_A(\lambda, x)$ and φ_A. The main motivation for that is provided by the following result which connects $\varphi_A(\lambda; x)$ with the quantity $\lambda\|y - x\|$ used in the criterion (33).

LEMMA 4.1. *Let $x \in \mathbb{E}$, $\lambda > 0$, and $\hat{\sigma} \geq 0$ be given. If (y, u, ε) is a $\hat{\sigma}$-approximate solution of* (34) *at (λ, x), then*

$$(1 - \hat{\sigma})\lambda\|y - x\| \leq \varphi_A(\lambda; x) \leq (1 + \hat{\sigma})\lambda\|y - x\|. \tag{36}$$

Proof. To simplify notation, let $y_A = y_A(\lambda; x)$ and observe that (35) implies that

$$x \in \lambda A(y_A) + y_A. \tag{37}$$

Define $r := \lambda u + y - x$ and note that the assumption that (y, u, ε) is a $\hat{\sigma}$-approximate solution of (34) implies that

$$\|r\|^2 + 2\lambda\varepsilon \leq \hat{\sigma}^2\|y - x\|^2, \qquad x + r \in \lambda A^\varepsilon(y) + y.$$

Hence, it follows from Lemma A.1 with $(y, v) = (y, x + r)$ and $(\tilde{y}, \tilde{v}) = (y_A, x)$ that

$$\|y_A - y\|^2 \leq \|x - (x + r)\|^2 + 2\lambda\varepsilon = \|r\|^2 + 2\lambda\varepsilon \leq \hat{\sigma}^2\|y - x\|^2$$

and hence that $\|y_A - y\| \leq \hat{\sigma}\|y - x\|$. The conclusion of the lemma now follows from this inequality and the triangle inequality for norms. ∎

Given $x \in \mathbb{E}$ and $\hat{\sigma} \geq 0$, the following result gives a sufficient condition on $\lambda > 0$ for *any* $\hat{\sigma}$-approximate solution of (34) to satisfy (33).

Lemma 4.2. *Let $x \in \mathbb{E}$, $0 < \alpha_- < \alpha_+$, and $\hat{\sigma} \geq 0$ be given. If $\lambda > 0$ is such that*

$$(1+\hat{\sigma})\alpha_- \leq \varphi_A(\lambda; x) \leq (1-\hat{\sigma})\alpha_+, \tag{38}$$

then any $\hat{\sigma}$-approximate solution (y, u, ε) of (34) *at (λ, x) satisfies* (33).

Proof. This result follows immediately from Lemma 4.1. ∎

In view of the above result, the search for a stepsize $\lambda > 0$ satisfying (33) can be guided by the goal of finding a $\lambda > 0$ satisfying (38). In view of this observation, we now study basic properties of the function φ_A which will play important roles on the design and complexity analysis of the search procedure of subsection 4.2.

Lemma 4.3. *For every $x \in \mathbb{E}$, the following statements hold:*

(a) *$\lambda > 0 \to \varphi_A(\lambda; x)$ is a continuous function;*

(b) *for every $0 < \tilde{\lambda} < \lambda$,*

$$\frac{\lambda}{\tilde{\lambda}}\varphi_A(\tilde{\lambda}; x) \leq \varphi_A(\lambda; x) \leq \left(\frac{\lambda}{\tilde{\lambda}}\right)^2 \varphi_A(\tilde{\lambda}; x). \tag{39}$$

Proof. To prove the first inequality in (b), assume that $\tilde{\lambda} < \lambda$ and let

$$y = y_A(\lambda; x), \qquad \tilde{y} = y_A(\tilde{\lambda}; x).$$

Since $\lambda^{-1}(x - y) \in A(y)$ and $\tilde{\lambda}^{-1}(x - \tilde{y}) \in A(\tilde{y})$, it follows from the monotonicity of A that

$$\langle y - \tilde{y}, \lambda^{-1}(x - y) - \tilde{\lambda}^{-1}(x - \tilde{y})\rangle \geq 0.$$

Multiplying the above inequality by $\lambda\tilde{\lambda}$ and adding $\tilde{\lambda}\|y - \tilde{y}\|^2$ to both sides of the resulting inequality, we conclude that

$$(\lambda - \tilde{\lambda})\langle y - \tilde{y}, \tilde{y} - x\rangle \geq \tilde{\lambda}\|y - \tilde{y}\|^2.$$

Since $\tilde{\lambda} < \lambda$ by assumption, it then follows that the inner product on the left-hand side of the above inequality is nonnegative and hence that

$$\|y - x\|^2 = \|y - \tilde{y}\|^2 + 2\langle y - \tilde{y}, \tilde{y} - x\rangle + \|\tilde{y} - x\|^2 \geq \|y - \tilde{y}\|^2.$$

The first inequality in (b) now follows from the above inequality and the definition of $\varphi_A(\tilde{\lambda}; x)$ in (35). To prove the second inequality in (b) we use the fact that the approximation of A_λ of index $\lambda > 0$ of A defined as

$$A_\lambda = \lambda^{-1}\big(I - (I + \lambda A)^{-1}\big)$$

has the property that $\lambda \mapsto \|A_\lambda\|$ is nonincreasing (see [2, Proposition 2.6]). Since $\varphi_A(\lambda; x) = \lambda^2\|A_\lambda(x)\|$ in view of (35), the latter property implies that

$$\varphi_A(\lambda; x) \leq \lambda^2\|A_{\tilde{\lambda}}\| = \left(\frac{\lambda}{\tilde{\lambda}}\right)^2 \varphi_A(\tilde{\lambda}; x)$$

and hence that the second inequality in (b) holds. Finally, note that (a) follows trivially (b). ∎

Lemma 4.4. *Let $x \in \mathbb{E}$ be such that $0 \notin A(x)$. Then, the following statements hold:*

(a) *$\varphi_A(\lambda; x) > 0$ for every $\lambda > 0$;*

(b) $\lambda > 0 \to \varphi_A(\lambda; x)$ *is a strictly increasing and continuous function, which converges to* 0 *or* ∞ *as* λ *tends to* 0 *or* ∞*, respectively;*

(c) *for any* $0 < \alpha_- < \alpha_+$*, the set of all scalars* $\lambda > 0$ *satisfying*

$$\alpha_- \leq \varphi_A(\lambda; x) \leq \alpha_+ \tag{40}$$

is an closed interval $[\lambda_-, \lambda_+]$ *such that*

$$\frac{\lambda_+}{\lambda_-} \geq \sqrt{\frac{\alpha_+}{\alpha_-}}. \tag{41}$$

Proof. To prove (a), fix $\lambda > 0$. Noting that the definition of $y_A(\lambda; x)$ in (35) implies that $[x - y_A(\lambda; x)]/\lambda \in A(x)$, we conclude that $y_A(\lambda; x) \neq x$ in view of the assumption that $0 \notin A(x)$. Hence, in view of (35) we have $\varphi_A(\lambda; x) > 0$. Statement (b) follows immediately from (a) and Lemma 4.3. To prove (c), first note that (b) implies the existence of unique scalars $0 < \lambda_- < \lambda_+$ such that

$$\varphi_A(\lambda_-; x) = \alpha_-, \quad \varphi_A(\lambda_+; x) = \alpha_+,$$

and that the set of $\lambda > 0$ such that (40) holds is the interval $[\lambda_-, \lambda_+]$. Since the second inequality in (39) with $\tilde{\lambda} = \lambda_-$ and $\lambda = \lambda_+$ implies

$$\alpha_+ = \varphi_A(\lambda_+; x) \leq \left(\frac{\lambda_+}{\lambda_-}\right)^2 \varphi_A(\lambda_-; x) = \left(\frac{\lambda_+}{\lambda_-}\right)^2 \alpha_-,$$

(41) follows. ∎

The following well-known technical result will be used in the proof of Proposition 4.9. For the sake of completeness, its proof is given in the appendix.

PROPOSITION 4.5. *For every* $\lambda > 0$ *and* $x, \tilde{x} \in \mathbb{E}$*, it holds that*

$$\|y_A(\lambda; x) - y_A(\lambda; \tilde{x})\| \leq \|x - \tilde{x}\|, \quad \|[x - y_A(\lambda; x)] - [\tilde{x} - y_A(\lambda; \tilde{x})]\| \leq \|x - \tilde{x}\|. \tag{42}$$

As a consequence, if $x^* \in A^{-1}(0)$*, then*

$$\max\{\|y_A(\lambda; x) - x^*\|, \|y_A(\lambda; x) - x\|\} \leq \|x - x^*\|. \tag{43}$$

4.2. A generic line search procedure. This subsection describes an iterative procedure for solving the problem described in the general goal at the beginning of this section.

The procedure consists of two stages. The first one, namely, the *bracketing stage*, either computes a stepsize $\lambda > 0$ and a $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x) satisfying (33) or finds an interval $[t_-, t_+]$ which contains all λ's satisfying (38). The second one, namely, the *bisection stage*, iteratively performs a geometric bisection scheme on the interval $[t_-, t_+]$ until a stepsize $\lambda > 0$ and $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x) satisfying (33) is found.

We start by describing the first stage.

Bracketing stage:

Input: $x \in \mathbb{E}$ such that $0 \notin A(x)$, $\hat{\sigma} \geq 0$, initial guess $\lambda^0 > 0$, and $0 < \alpha_- < \alpha_+$ such that

$$\alpha_-(1 + \hat{\sigma}) < \alpha_+(1 - \hat{\sigma});$$

Output: either a $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ^0, x) such that (33) holds or an interval $[t_-, t_+]$ containing all λ's satisfying (38).

(1) use the black-box to compute a $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ^0, x);

(2) if $\lambda^0\|y - x\| \in [\alpha_-, \alpha_+]$, then output the $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ^0, x);
if $\lambda^0\|y - x\| < \alpha_-$, then output $t_- := \lambda^0$ and $t_+ := \alpha_+/\|y - x\|$;
if $\lambda^0\|y - x\| > \alpha_+$, then output $t_- := \alpha_-/\|y - x\|$ and $t_+ := \lambda^0$.

end

Observe that the complexity of the above scheme is equivalent to one call of the subroutine for computing a $\hat{\sigma}$-approximate solution of (34).

The justification of the above stage is based on the following result.

LEMMA 4.6. *Let $x \in \mathbb{E}$ such that $0 \notin A(x)$, $\lambda^0 > 0$, and $\hat{\sigma} \geq 0$ be given and let $[\lambda_-, \lambda_+]$ denote the interval of all λ's satisfying* (38). *Then, for any $\hat{\sigma}$-approximate solution (y, u, ε) of* (34) *at (λ^0, x), the following statements hold:*

(a) *if $\lambda^0\|y - x\| < \alpha_-$, then $\lambda^0 < \lambda_-$ and $\lambda_+ \leq \alpha_+/\|y - x\|$;*

(b) *if $\lambda^0\|y - x\| > \alpha_+$, then $\lambda_+ < \lambda^0$ and $\alpha_-/\|y - x\| \leq \lambda_-$.*

Proof. To prove statement (a), assume that $\lambda^0 > 0$ is such that $\lambda^0\|y - x\| < \alpha_-$. Lemma 4.1 then implies that $\varphi_A(\lambda^0; x) \leq (1 + \hat{\sigma})\alpha_- = \varphi_A(\lambda_-; x)$, where the last equality is due to the definition of λ_-. The claim that $\lambda^0 < \lambda_-$ now follows immediately from Lemma 4.4(b). Multiplying the first inequality in (36) with $\lambda = \lambda^0$ by λ_+/λ^0 and using Lemma 4.3 with $\tilde{\lambda} = \lambda^0$ and $\lambda = \lambda_+$, we conclude that

$$\lambda_+(1 - \hat{\sigma})\|y - x\| \leq \frac{\lambda_+}{\lambda^0}\varphi_A(\lambda^0; x) \leq \varphi_A(\lambda_+; x) = (1 - \sigma)\alpha_+,$$

where the last equality follows from the definition of λ_+. Hence, the second inequality in (a) follows. The proof of (b) follows in a similar manner. ∎

We next describe the bisection stage which is the one that accounts for the overall complexity of the procedure for computing $\lambda > 0$ and $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x) satisfying (33). As mentioned earlier, it searches for a stepsize $\lambda > 0$ satisfying the sufficient condition (38) but terminates whenever it detects a stepsize with a corresponding $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x) satisfying (33).

Bisection stage:

Input: $x \in \mathbb{E}$ such that $0 \notin A(x)$ and interval $[t_-, t_+]$ containing all λ's satisfying (38);

Output: a scalar λ and a $\hat{\sigma}$-approximate solution (y, u, ε) of (34) such that (33) holds.

(1) set $\lambda = \sqrt{t_- t_+}$ and use the black-box to compute a $\hat{\sigma}$-approximate solution (y, u, ε) of (34) at (λ, x);

(2) if $\lambda\|y - x\| \in [\alpha_-, \alpha_+]$, then output λ and (y, u, ε) and **stop**;

(3) if $\lambda\|y - x\| > \alpha_+$, then set $t_+ \leftarrow \lambda$; else set $t_- \leftarrow \lambda$.

(4) go to step 1.

end

LEMMA 4.7. *If the bracketing stage outputs an interval $[t_-, t_+]$ containing all λ's satisfying* (38)*, which is then input to the bisection stage, then the number of black-box calls during the bisection stage is bounded by*

$$2 + \log\left(\frac{\log \tau}{\log[(1 - \hat{\sigma})\alpha_+/((1 + \hat{\sigma})\alpha_-)]}\right),$$

where

$$\tau := \max\left\{\frac{\alpha_+}{\lambda^0\|y-x\|}, \frac{\lambda^0\|y-x\|}{\alpha_-}\right\}$$

and (y,u,ε) *denotes the* $\hat\sigma$*-approximate solution of* (34) *at* (λ^0,x) *computed at step* 1 *of the bracketing stage.*

Proof. Since $\log\lambda = (\log t_- + \log t_+)/2$, it follows that after j bisection iterations the scalars t_- and t_+ computed at step 3 of the bisection stage satisfy

$$\log\frac{t_+}{t_-} = \frac{1}{2^j}\log\tau, \tag{44}$$

since, in view of step 2 of the bracketing stage, τ is the value of the ratio t_+/t_- at the beginning of the bisection stage. Assume now that the method does not stop at the jth bisection iteration. Then, the values of t_- and t_+ in step 2 of this iteration satisfy $t_- \le \lambda_- < \lambda_+ \le t_+$, where $[\lambda_-,\lambda_+]$ is the interval of all λ's satisfying (38). Hence, in view of Lemma 4.4, we have

$$\frac{t_+}{t_-} \ge \frac{\lambda_+}{\lambda_-} \ge \sqrt{\frac{(1-\hat\sigma)\alpha_+}{(1+\hat\sigma)\alpha_-}}.$$

This together with (44) implies that

$$\frac{1}{2^j}\log\tau \ge \frac{1}{2}\log\frac{(1-\hat\sigma)\alpha_+}{(1+\hat\sigma)\alpha_-}$$

and hence that

$$j \le 1 + \log\left(\frac{\log\tau}{\log[(1-\hat\sigma)\alpha_+/((1+\hat\sigma)\alpha_-)]}\right).$$

Hence, the result follows. ∎

4.3. Computation of the stepsize λ_k. In this subsection, our main goal is to describe an algorithm for computing a stepsize $\lambda_k > 0$ and a σ-approximate Newton solution (y_k,u_k,ε_k) of (20) at $(x,\lambda) = (x_{k-1},\lambda_k)$ satisfying (26), as required by step 2 of the inexact NPE method.

We assume that we have at our disposal the following black-box.

Newton black-box. For any given $\lambda > 0$ and $x \in \mathbb{E}$, it computes a $\hat\sigma$-approximate Newton solution (y,u,ε) of (20) at (λ,x).

As already mentioned in the beginning of this section, the Newton black-box provides a specific black-box for finding $\hat\sigma$-approximate solution (y,u,ε) of (34) with $A = T_{x'}$ and $x' = P_\Omega(x)$ at an arbitrary $(\lambda,x) \in \mathbb{R}_{++}\times\mathbb{E}$. In view of this observation, we can use the algorithm described in subsection 4.2 with $A = T_{x'_{k-1}}$, $x = x_{k-1}$, $\alpha_- = 2\sigma_\ell/L$, $\alpha_+ = 2\sigma_u/L$ to compute λ_k, (y_k,u_k,ε_k) as in step 2 of the inexact NPE method.

We now state the whole procedure for computing λ_k and (y_k,u_k,ε_k).

Bracketing/Bisection Procedure:

Input: $x_{k-1} \in \mathbb{E}$ such that $0 \notin (F+H)(x_{k-1})$, initial guess $\lambda_k^0 > 0$, tolerance $\hat\sigma \ge 0$, and constants $0 < \sigma_\ell < \sigma_u$ satisfying

$$\sigma_\ell(1+\hat\sigma) < \sigma_u(1-\hat\sigma);$$

Output: stepsize $\lambda_k > 0$ and a $\hat{\sigma}$-approximate Newton solution $(y_k, u_k, \varepsilon_k)$ of (20) at (x_{k-1}, λ_k) satisfying (26).

(0) set $\alpha_- := 2\sigma_\ell/L$ and $\sigma_u := 2\sigma_u/L$;

(1) (**Bracketing stage)** use the Newton black-box to compute a $\hat{\sigma}$-approximate Newton solution $(y_k^0, u_k^0, \varepsilon_k^0)$ of (20) at (x_{k-1}, λ_k^0);

- (1.a) if $\lambda_k^0 \|y_k^0 - x_{k-1}\| \in [\alpha_-, \alpha_+]$, then output $\lambda_k := \lambda_k^0$ and $(y_k, u_k, \varepsilon_k) := (y_k^0, u_k^0, \varepsilon_k^0)$ and **stop**;
- (1.b) if $\lambda_k^0 \|y_k^0 - x_{k-1}\| < \alpha_-$, then set $t_- = \lambda_k^0$ and $t_+ = \alpha_+/\|y_k^0 - x_{k-1}\|$;
- (1.c) if $\lambda_k^0 \|y_k^0 - x_{k-1}\| > \alpha_+$, then set $t_- = \alpha_-/\|y_k^0 - x_{k-1}\|$ and $t_+ = \lambda_k^0$;

(2) (**Bisection stage)**

- (2.a) set $\lambda = \sqrt{t_- t_+}$ and use the Newton black-box to compute a $\hat{\sigma}$-approximate Newton solution (y, u, ε) of (20) at (x_{k-1}, λ);
- (2.b) if $\lambda\|y - x\| \in [\alpha_-, \alpha_+]$, then output $\lambda_k := \lambda$ and $(y_k, u_k, \varepsilon_k) := (y, u, \varepsilon)$, and **stop**;
- (2.c) if $\lambda\|y - x\| > \alpha_+$, then set $t_+ \leftarrow \lambda$; else set $t_- \leftarrow \lambda$;
- (2.d) go to step 2.a.

end

The following result is a specialization of Lemma 4.7 to the particular context of the above procedure.

LEMMA 4.8. *The number of Newton black-box calls in the bracketing/bisection procedure is bounded by*

$$3 + \log\left(\frac{\log \tau_k}{\log[(1-\hat{\sigma})\sigma_u/((1+\hat{\sigma})\sigma_l)]}\right),$$

where

$$\tau_k := \max\left\{\frac{2\sigma_u}{L\lambda_k^0\|y_k^0 - x_{k-1}\|}, \frac{L\lambda_k^0\|y_k^0 - x_{k-1}\|}{2\sigma_\ell}\right\}.$$

Proof. This result follows immediately from Lemma 4.7 with $\alpha_- = 2\sigma_\ell/L$ and $\alpha_+ = 2\sigma_u/L$. ☐

Our goal from now on will be to bound τ_k in terms of L, d_0, and λ_k^0 as well as the parameters $\hat{\sigma}$, σ_ℓ, and σ_u. We first state the following technical result.

PROPOSITION 4.9. *Assume that $x^* \in (F+H)^{-1}(0)$ and let $\bar{x} \in \Omega$ and $x \in \mathbb{E}$ be given. Then,*

$$\|x - (I + \lambda T_{\bar{x}})^{-1}(x)\| \leq \|x - x^*\| + \frac{\lambda L}{2}\|\bar{x} - x^*\|^2.$$

Proof. Let

$$r := F(x^*) - F_{\bar{x}}(x^*), \qquad \tilde{T}_{\bar{x}} := T_{\bar{x}} + r.$$

Using the assumption that $0 \in T(x^*)$ and the definition of $T_{\bar{x}}$ and r, we easily see that $0 \in \tilde{T}_{\bar{x}}(x^*)$. Hence, by relation (43) of Proposition 4.5 with $A = \tilde{T}_{\bar{x}}$, we have

$$\|x - (I + \lambda\tilde{T}_{\bar{x}})^{-1}(x)\| \leq \|x - x^*\|. \tag{45}$$

Using the observation that

$$(I + \lambda\tilde{T}_{\bar{x}})^{-1}(x + \lambda r) = [I + \lambda(T_{\bar{x}} + r)]^{-1}(x + \lambda r) = (I + \lambda T_{\bar{x}})^{-1}(x),$$

the triangle inequality for norms, relation (45), and the fact that by Proposition 4.5 the resolvent is nonexpansive, we then conclude that

$$\begin{aligned}\|x - (I + \lambda T_{\bar{x}})^{-1}(x)\| &= \|x - (I + \lambda \tilde{T}_{\bar{x}})^{-1}(x + \lambda r)\| \\ &\leq \|x - (I + \lambda \tilde{T}_{\bar{x}})^{-1}(x)\| \\ &\quad + \|(I + \lambda \tilde{T}_{\bar{x}})^{-1}(x) - (I + \lambda \tilde{T}_{\bar{x}})^{-1}(x + \lambda r)\| \\ &\leq \|x - x^*\| + \lambda\|r\| \leq \|x - x^*\| + \frac{\lambda L}{2}\|\bar{x} - x^*\|^2,\end{aligned}$$

where the last inequality is due to the definition of r and Lemma 3.2 with $y = x^*$ and $x = \bar{x}$. ∎

LEMMA 4.10. *Let $\varphi_k : (0, \infty) \to \mathbb{R}$ denote the function $\lambda \mapsto \varphi_A(\lambda; x_{k-1})$ with $A = T_{x'_{k-1}}$. Then, for every $\lambda > 0$, we have*

$$\varphi_k(\lambda) \leq \lambda d_0 + \frac{L}{2}(\lambda d_0)^2,$$

where d_0 is the distance of x_0 to $T^{-1}(0)$. As a consequence,

$$\lambda_k^0\|y_k^0 - x_{k-1}\| \leq \frac{1}{1 - \hat{\sigma}}\left[\lambda_k^0 d_0 + \frac{L}{2}(\lambda_k^0 d_0)^2\right].$$

Proof. Let $x^* \in T^{-1}(0)$ be such that $\|x^* - x_0\| = d_0$. Using Proposition 4.9 with $x = x_{k-1}$ and $\bar{x} = x'_{k-1}$, we conclude that

$$\begin{aligned}\varphi_k(\lambda) &= \lambda\|x_{k-1} - (I + \lambda T_{x'_{k-1}})^{-1}(x_{k-1})\| \leq \lambda\|x_{k-1} - x^*\| + \lambda^2\frac{L}{2}\|x'_{k-1} - x^*\|^2 \\ &\leq \lambda\|x_{k-1} - x^*\| + \lambda^2\frac{L}{2}\|x_{k-1} - x^*\|^2 \leq \lambda d_0 + \frac{L}{2}(\lambda d_0)^2,\end{aligned}$$

where the second last inequality follows from the fact that $x'_{k-1} = P_\Omega(x_{k-1})$ and $x^* \in \Omega$ and from a well-known property of the projection operator P_Ω, and the last inequality follows from the fact that $\|x_{k-1} - x^*\| \leq \|x_0 - x^*\| = d_0$, in view of Proposition 2.4.

The last conclusion of the lemma follows from the first one with $\lambda = \lambda_k^0$, the first remark following Definition 3.1, and Lemma 4.1 with $x = x_{k-1}$ and $(y, u, \varepsilon) = (y_k^0, u_k^0, \varepsilon_k^0)$ and $A = T_{x'_{k-1}}$. ∎

LEMMA 4.11. *Let $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$ be given and define*

$$\zeta_k := \min\left\{\frac{2\bar{\rho}(\lambda_k^0)^2}{(1 + \hat{\sigma}) + [(1 + \hat{\sigma})^2 + 2L(\lambda_k^0)^2\bar{\rho}]^{1/2}}, \frac{\sqrt{2\bar{\varepsilon}}(\lambda_k^0)^{3/2}}{\hat{\sigma}}\right\}. \tag{46}$$

Also, consider the $\hat{\sigma}$-approximate Newton solution $(y_k^0, u_k^0, \varepsilon_k^0)$ computed at step 1 of the bracketing/bisection procedure and define

$$v_k^0 := F(u_k^0) + u_k^0 - F_{x'_{k-1}}(y_k^0).$$

Then, either one of the following statements holds:

(a) $\lambda_k^0\|y_k^0 - x_{k-1}\| > \zeta_k$;

(b) $v_k^0 \in (F + H^{\varepsilon_k^0})(y_k^0) \subseteq T^{\varepsilon_k^0}(y_k^0)$ *and the following error bounds hold:*

$$\|v_k^0\| \leq \bar{\rho}, \qquad \varepsilon_k^0 \leq \bar{\varepsilon}. \tag{47}$$

Proof. Assume that (a) does not hold, i.e.,

$$\lambda_k^0 \|y_k^0 - x_{k-1}\| \le \zeta_k. \tag{48}$$

Using (48) together with Lemma 3.3 with $(\lambda, x) = (\lambda_k^0, x_{k-1})$ and $(y, u, \varepsilon) = (y_k^0, u_k^0, \varepsilon_k^0)$, we conclude that $v_k^0 \in (F + H^{\varepsilon_k^0})(y_k^0)$ and

$$\|\lambda_k^0 v_k^0 + y_k^0 - x_{k-1}\| \le \left(\hat{\sigma} + \frac{L\zeta_k}{2}\right) \|y_k^0 - x_{k-1}\|,$$

and hence that

$$\|\lambda_k^0 v_k^0\| \le \left(1 + \hat{\sigma} + \frac{L\zeta_k}{2}\right) \|y_k^0 - x_{k-1}\|.$$

Multiplying both sides of the above inequality by λ_k^0 and using (48), we obtain

$$(\lambda_k^0)^2 \|v_k^0\| \le \frac{L}{2}(\zeta_k)^2 + (1 + \hat{\sigma})\zeta_k \le (\lambda_k^0)^2 \bar{\rho},$$

where the last inequality is due to (46). Hence, the first bound in (47) follows.

Moreover, the fact that $(y_k^0, u_k^0, \varepsilon_k^0)$ is a $\hat{\sigma}$-approximate Newton solution $(y_k^0, u_k^0, \varepsilon_k^0)$ of (6) at (λ_k^0, x_{k-1}) and relations (46) and (48) imply that

$$\varepsilon_k^0 \le \frac{\hat{\sigma}^2 \|y_k^0 - x_{k-1}\|^2}{2\lambda_k^0} \le \frac{\hat{\sigma}^2 \zeta_k^2}{2(\lambda_k^0)^3} \le \bar{\varepsilon},$$

and hence that the second bound in (47) holds. ∎

THEOREM 4.12. *Consider the inexact NPE method in which the bracketing/bisection procedure discussed above with $A = T_{x'_{k-1}}$ and $x = x_{k-1}$ is used to find the stepsize λ_k and a $\hat{\sigma}$-approximate Newton solution $(y_k, u_k, \varepsilon_k)$ at (λ_k, x_{k-1}) satisfying* (26)*, and let $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$ be given. Then, for every iteration k, one of the following statements holds:*

(a) *The bracketing/bisection procedure makes at most*

$$3 + \log\log\left\{\max\left(\frac{L\lambda_k^0 d_0 + (L\lambda_k^0 d_0)^2/2}{2(1-\hat{\sigma})\sigma_\ell}, \frac{2\sigma_u}{L\zeta_k}\right)\right\} - \log\log\frac{(1-\hat{\sigma})\sigma_u}{(1+\hat{\sigma})\sigma_\ell}$$

Newton black-box calls, where ζ_k is defined in (46).

(b) *Statement* (b) *of Lemma* 4.11 *holds.*

Proof. This result follows immediately from Lemmas 4.8, 4.10, and 4.11. ∎

For prespecified tolerances $\bar{\rho} > 0$ and $\bar{\varepsilon} > 0$, assume that our goal is to find $(y, v, \varepsilon) \in \mathbb{E} \times \mathbb{E} \times \mathbb{R}_+$ such that

$$v \in F(y) + H^\varepsilon(y), \qquad \|v\| \le \bar{\rho}, \qquad \varepsilon \le \bar{\varepsilon}.$$

It follows from Theorems 3.6 and 4.12 that the inexact NPE method with $\lambda_k^0 = 1$ for every $k \ge 1$ will find such triple (y, v, ε) in at most

$$\mathcal{O}\left(\max\left\{\left(\frac{Ld_0^2}{\bar{\rho}}\right)^{2/3}, \left(\frac{Ld_0^3}{\bar{\varepsilon}}\right)^{2/3}\right\}\right)$$

iterations with each one making at most

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

Newton black-box calls.

5. Final remark. The analysis of sections 3 and 4 holds for any L satisfying (18) and hence for the Lipschitz constant $L_{F'}$ of F', which by the definition is the smallest L satisfying (18). Clearly, the best iteration-complexity bound and most likely computational performance of the NPE is obtained with $L = L_{F'}$. However, for many instances of (17), it is difficult to compute $L_{F'}$. Nevertheless, for any fixed constant $\gamma > 1$, the following variant of the inexact NPE method with an adaptive sequence of constants $\{L_k\}$ has the feature that, within a small number of iterations k_0, we have $L_k \leq \gamma L_{F'}$ for all $k \geq k_0$, and the method still has the same complexity as the original one with $L = L_{F'}$. To describe the variant, we start with an arbitrary $L_1 > 0$. Generally, at the kth iteration, we have a constant L_k and the iterate x_{k-1}. A $\hat{\sigma}$-approximate Newton solution $(y_k, u_k, \varepsilon_k)$ satisfying (26) and the pair (v_k, x_k) as in (27) are computed and the criterion

$$\|v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|y_k - x_{k-1}\|^2 \tag{49}$$

is checked. If (49) holds, x_k is accepted as the next iterate and L_{k+1} is updated as $L_{k+1} = L_k/\gamma$ if $L_1 > L_2 > \cdots > L_k$, or $L_{k+1} = L_k$, otherwise. If (49) does not hold, then set $L_k = \gamma L_k$ and repeat the k-iteration.

Appendix. In this section, we provide the proof of Proposition 4.5. We also establish a simple technical lemma that is used both in the proofs of Lemma 4.1 and Proposition 4.5. The next result was essentially proved in [24].

LEMMA A.1. *Assume that A is a maximal monotone operator and that $\varepsilon \geq 0$, $\lambda > 0$, and $y, \tilde{y}, v, \tilde{v} \in \mathbb{E}$ satisfy*

$$v \in \lambda A^{\varepsilon}(y) + y, \qquad \tilde{v} \in \lambda A(\tilde{y}) + \tilde{y}. \tag{50}$$

Then,

$$\|\tilde{v} - v\|^2 + 2\lambda\varepsilon \geq \|\tilde{y} - y\|^2 + \|(\tilde{v} - v) - (\tilde{y} - y)\|^2. \tag{51}$$

Proof. Assumption (50) is equivalent to

$$\lambda^{-1}(v - y) \in A^{\varepsilon}(y), \qquad \lambda^{-1}(\tilde{v} - \tilde{y}) \in A(\tilde{y}).$$

Hence, by the definition of A^{ε} in (1), we conclude that

$$\langle \lambda^{-1}[(\tilde{v} - \tilde{y}) - (v - y)], \tilde{y} - y \rangle \geq -\varepsilon,$$

which is easily seen to be equivalent to (51). ∎

We are now ready to give the proof of Proposition 4.5.

Proof of Proposition 4.5. To simplify notation, $y = y_A(\lambda; x)$ and $\tilde{y} = y_A(\lambda; \tilde{x})$ and note that (35) implies that

$$x \in \lambda A(y) + y, \qquad \tilde{x} \in \lambda A(\tilde{y}) + \tilde{y}.$$

Hence, it follows from Lemma A.1 with $\varepsilon = 0$, $(y, v) = (y, x)$, and $(\tilde{y}, \tilde{v}) = (\tilde{y}, \tilde{x})$ that

$$\|\tilde{y} - y\| + \|(\tilde{x} - x) - (\tilde{y} - y)\|^2 \leq \|\tilde{x} - x\|^2$$

and hence that both inequalities in (42) hold. Since the assumption that $x^* \in A^{-1}(0)$

and (35) trivially imply that $y_A(\lambda; x^*) = x^*$ and hence that

$$\|x - y_A(\lambda; x)\| = \|(x - y_A(\lambda; x)) - (x^* - y_A(\lambda; x^*))\|,$$

it follows that inequality (43) follows from the second inequality in (42) with $\tilde{x} = x^*$. ∎

REFERENCES

[1] G. Auchmuty, *Variational principles for variational inequalities*, Numer. Funct. Anal. Optim., 10 (1989), pp. 863–874.

[2] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Math. Stud. 5, North-Holland, Amsterdam, 1973.

[3] A. Brøndsted and R. T. Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc., 16 (1965), pp. 605–611.

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

[5] R. S. Burachik, C. Sagastizábal, and B. F. Svaiter, *Bundle methods for maximal monotone operators*, in Ill-Posed Variational Problems and Regularization Techniques, Lecture Notes in Econom. and Math. Systems 477, Springer, Berlin, 1999, pp. 49–64.

[6] 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, Appl. Optim. 22, Kluwer Academy, Dordrecht, 1999, pp. 25–43.

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

[8] M. Fukushima, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Math. Program., 53 (1992), pp. 99–110.

[9] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Rev. Française Informat. Recherche Opérationnelle, 4 (1970), pp. 154–158.

[10] R. D. C. Monteiro and B. F. Svaiter, *Convergence Rate of Inexact Proximal Point Methods with Relative Error Criteria for Convex Optimization*, Technical report, Optimization Online, 2010.

[11] R. D. C. Monteiro and B. F. Svaiter, *Iteration-Complexity of Block-Decomposition Algorithms and the Alternating Direction Method of Multipliers*, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, 2010.

[12] 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 (2010), pp. 2755–2787.

[13] 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 (2011), pp. 1688–1720.

[14] R. D. C. Monteiro and B. F. Svaiter, *Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems*, Technical report, Optimization Online, 2011.

[15] Y. Nesterov, *Accelerating the cubic regularization of Newton's method on convex problems*, Math. Program., 112 (2008), pp. 159–181.

[16] Y. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Math. Program., 108 (2006), pp. 177–205.

[17] Y. E. Nesterov and A. S. Nemirovski, *Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1994.

[18] J. Povh, F. Rendl, and A. Wiegele, *A boundary point method to solve semidefinite programs*, Computing, 78 (2006), pp. 277–286.

[19] R. T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific J. Math., 33 (1970), pp. 209–216.

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

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

[22] 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 (1999), pp. 323–345.

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

[24] M. V. Solodov and B. F. Svaiter, *Error bounds for proximal point subproblems and associated inexact proximal point algorithms*, Math. Program., 88 (2000), pp. 371–389.

[25] 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 (2000), pp. 214–230.

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

[27] X.-Y. Zhao, D. Sun, and K.-C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim., 20 (2010), pp. 1737–1765.