SIAM J. OPTIM.
Vol. 20, No. 6, pp. 2755–2787

ON THE COMPLEXITY OF THE HYBRID PROXIMAL EXTRAGRADIENT METHOD FOR THE ITERATES AND THE ERGODIC MEAN*

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

Dedicated to Paul Tseng's life and career

Abstract. In this paper we analyze the iteration complexity of the hybrid proximal extragradient (HPE) method for finding a zero of a maximal monotone operator recently proposed by Solodov and Svaiter. One of the key points of our analysis is the use of new termination criteria based on the ε-enlargement of a maximal monotone operator. The advantage of using these termination criteria is that their definition do not depend on the boundedness of the domain of the operator. We then show that Korpelevich's extragradient method for solving monotone variational inequalities falls in the framework of the HPE method. As a consequence, using the complexity analysis of the HPE method, we obtain new complexity bounds for Korpelevich's extragradient method which do not require the feasible set to be bounded, as assumed in a recent paper by Nemirovski. Another feature of our analysis is that the derived iteration-complexity bounds are proportional to the distance of the initial point to the solution set. The HPE framework is also used to obtain the first iteration-complexity result for Tseng's modified forward-backward splitting method for finding a zero of the sum of a monotone Lipschitz continuous map with an arbitrary maximal monotone operator whose resolvent is assumed to be easily computable. Also using the framework of the HPE method, we study the complexity of a variant of a Newton-type extragradient algorithm proposed by Solodov and Svaiter for finding a zero of a smooth monotone function with Lipschitz continuous Jacobian.

Key words. extragradient, variational inequality, maximal monotone operator, complexity, complementarity problems, Korpelevich and Newton methods

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

DOI. 10.1137/090753127

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 Rockafellar [21], is a classical iterative scheme for solving the monotone inclusion problem which generates a sequence $\{x_k\}$ according to

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

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.,

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

*Received by the editors March 17, 2009; accepted for publication (in revised form) June 7, 2010; published electronically August 17, 2010.
http://www.siam.org/journals/siopt/20-6/75312.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 N00014-08-1-0033.
‡IMPA, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brazil (benar@impa.br). The work of this author was partially supported by FAPERJ grants E-26/152.512/2006 and E-26/102.821/2008 and CNPq grants 303583/2008-8 and 480101/2008-6.

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

New inexact versions of the proximal point method, with relative error criteria were proposed by Solodov and Svaiter [22, 23, 25, 24]. In this paper, we will concern ourselves with one of these inexact versions of the proximal point method introduced in [22], namely the hybrid proximal extragradient (HPE) method. In contrast to [22], which studies only global convergence of the HPE method, we establish in this paper its iteration complexity. One of the key points of our analysis is the use of a new termination criterion based on the ε-enlargement of T introduced in [2]. More specifically, given $\varepsilon > 0$, the algorithm terminates whenever it finds a point $\bar{y}$ and a pair $(\bar{v}, \bar{\varepsilon})$ such that

$$\bar{v} \in T^{\bar{\varepsilon}}(\bar{y}), \qquad \max\{\|\bar{v}\|, \bar{\varepsilon}\} \leq \varepsilon. \tag{1}$$

For each x, $T^{\varepsilon}(x)$ is an outer approximation of $T(x)$ which coincides with $T(x)$ when $\varepsilon = 0$. Hence, for $\varepsilon = 0$ the above termination criterion reduces to the condition that $0 \in T(x)$. The ε-enlargement of maximal monotone operators is a generalization of the ε-subgradient enlargement of the subdifferential of a convex function. The advantage of using this termination criterion is that it does not require boundedness of the domain of T. Another feature of our analysis is that the derived iteration complexity bounds are proportional to the distance of the initial point to the solution set. Results of this kind are known for minimization of convex functions but, to the best our knowledge, are new in the context of monotone VI problems (see, for example, [19]).

We then establish a new result showing that Korpelevich's extragradient method for solving VI problems is a special case of the HPE method. This allows us to obtain an $\mathcal{O}(d_0/\varepsilon)$ iteration complexity for termination criterion (1), where d_0 is the distance of the initial iterate to the solution set. Since, together with every iterate $\bar{y}_k$, the method also generates a pair $(\bar{v}_k, \bar{\varepsilon}_k)$ so that (1) can be checked, there is no need to assume the feasible set to be bounded or to estimate d_0. We also translate (and sometimes strengthen) the above complexity results to the context of monotone VI problems with linear operators and/or bounded feasible sets, and monotone complementarity problems.

The HPE framework is also used to obtain the first iteration-complexity result for Tseng's modified forward-backward splitting (MF-BS) method [27] for finding a zero of the sum of a monotone Lipschitz continuous map with an arbitrary maximal monotone operator whose resolvent is assumed to be easily computable.

Also using the framework of the HPE method, we study the complexity of a variant of a Newton-type extragradient algorithm proposed in [22] for finding a zero of a smooth monotone function with Lipschitz continuous Jacobian.

Previous papers dealing with iteration-complexity analysis of methods for VIs are as follows. In [14], a unifying geometric framework based on the ellipsoid method ideas is presented for VIs with bounded feasible sets and co-coercive (also know as strong-f-monotone) maps. Bundle-type methods for solving VIs with bounded feasible sets and/or bounded variation maps are studied in [6, 12]. Nemirovski [15] studies the complexity of Korpelevich's extragradient method under the assumption that the feasible set is bounded and an upper bound of its diameter is known. Nesterov [18] proposes a new dual extrapolation algorithm for solving VI problems whose termination depends on the guess of a ball centered at the initial iterate. Finally, asymptotic convergence rate results for extragradient-type methods are thoroughly discussed in [7, 10, 26].

This paper is organized as follows. In section 2, we review the definition and some of the basic properties of the ε-enlargement of a point-to-set operator and state some new results about the ε-enlargement of a monotone Lipschitz continuous map. Section 3 introduces two notions of approximate solutions for the VI problem. It also discusses how the monotone VI problem can be viewed as a special instance of the monotone inclusion problem and interprets the above two notions of approximate solutions in terms of criterion (1). The HPE method is reviewed in section 4, where its general iteration complexity is also derived. In section 5, we derive iteration-complexity results for Korpelevich's extragradient method to obtain different types of approximate solutions, even for the case of unbounded feasible sets. In subsections 5.1 and 5.2, we obtain other complexity results for Korpelevich's method under different assumptions on the function (e.g., linearity) and the feasible set (e.g., conic and/or bounded set) of the VI problem. In section 6, we consider a particular version of Tseng's (MF-BS) method [27], review the result of [22] that it can be viewed as a particular case of the HPE method, and use this fact to derive, for the first time, its iteration complexities for finding different types of approximate solutions. In section 7, we study the iteration complexity of a Newton-type proximal extragradient method for solving a monotone smooth nonlinear equation. In section 8, we conclude our main presentation by providing some concluding remarks. Finally, we review in Appendix A other notions of error measures and discuss their relationship with the error measures used in the main presentation of the paper.

Notation. Throughout this paper, we let $\mathbb{R}^n$ denote an n-dimensional space with inner product and induced norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively.

2. The ε-enlargement. Since our complexity analysis is based on the ε-enlargement of a monotone operator, in this section we give its definition and review some of its properties. We also derive new results for the ε-enlargement of Lipschitz continuous monotone operators.

A point-to-set operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a relation $T \subset \mathbb{R}^n \times \mathbb{R}^n$ and

$$T(x) = \{v \in \mathbb{R}^n \mid (x, v) \in T\}.$$

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

$$\mathrm{Gr}(T) = \{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n \mid v \in T(x)\}.$$

An operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *monotone* if

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

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., $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ monotone and $\mathrm{Gr}(S) \supset \mathrm{Gr}(T)$ imply that $S = T$.

In [2], Burachik, Iusem, and Svaiter introduced the ε-enlargement of maximal monotone operators. Here, we extend this concept to a generic point-to-set operator in $\mathbb{R}^n$. Given $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and a scalar ε, define $T^\varepsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ as

$$T^\varepsilon(x) = \{v \in \mathbb{R}^n \mid \langle x - \tilde{x}, v - \tilde{v} \rangle \geq -\varepsilon \quad \forall \tilde{x} \in \mathbb{R}^n,\ \forall \tilde{v} \in T(\tilde{x})\} \quad \forall x \in \mathbb{R}^n. \tag{2}$$

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

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

(a) *if $\varepsilon_1 \le \varepsilon_2$, then $T^{\varepsilon_1}(x) \subset T^{\varepsilon_2}(x)$ for every $x \in \mathbb{R}^n$;*
(b) *$T^\varepsilon(x) + (T')^{\varepsilon'}(x) \subset (T + T')^{\varepsilon+\varepsilon'}(x)$ for every $x \in \mathbb{R}^n$ and $\varepsilon, \varepsilon' \in \mathbb{R}$;*
(c) *T is monotone if and only if $T \subset T^0$;*
(d) *T is maximal monotone if and only if $T = T^0$;*
(e) *if T is maximal monotone, $\{(x_k, v_k, \varepsilon_k)\} \subset \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ converges to $(\bar{x}, \bar{v}, \bar{\varepsilon})$, and $v_k \in T^{\varepsilon_k}(x_k)$ for every k, then $\bar{v} \in T^{\bar{\varepsilon}}(\bar{x})$.*

Proof. Statements (a), (b), (c), and (d) follow directly from definition (2) and the definition of (maximal) monotonicity. For a proof of statement (e), see [4]. ∎

We now make two remarks about Proposition 2.1. If T is a monotone operator and $\varepsilon \ge 0$, it follows from (a) and (d) that $T^\varepsilon(x) \supset T(x)$ for every $x \in \mathbb{R}^n$, and hence that T^ε is really an enlargement of T. Moreover, if T is maximal monotone, then (e) says that T and T^ε coincide when $\varepsilon = 0$.

The ε-enlargement of monotone operators is a generalization of the ε-subdifferential of convex functions. Recall that for a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ and scalar $\varepsilon \ge 0$, the ε-subdifferential of f is the operator $\partial_\varepsilon f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined as

$$\partial_\varepsilon f(x) = \{v \mid f(y) \ge f(x) + \langle y - x, v\rangle - \varepsilon \ \forall y \in \mathbb{R}^n\} \quad \forall x \in \mathbb{R}^n.$$

When $\varepsilon = 0$, the operator $\partial_\varepsilon f$ is simply denoted by ∂f and is referred to as the subdifferential of f. The operator ∂f is trivially monotone if f is proper. If f is a proper lower semicontinuous convex function, then ∂f is maximal monotone [20]. The conjugate of f is the function $f^* : \mathbb{R}^n \to \bar{\mathbb{R}}$ defined as

$$f^*(s) = \sup_{x\in\mathbb{R}^n} \langle s, x\rangle - f(x) \quad \forall s \in \mathbb{R}^n.$$

The following result lists some useful properties about the ε-subdifferential of a proper convex function.

PROPOSITION 2.2. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a proper convex function. Then,*

(a) *$\partial_\varepsilon f(x) \subset (\partial f)^\varepsilon(x)$ for any $\varepsilon \ge 0$ and $x \in \mathbb{R}^n$;*
(b) *$\partial_\varepsilon f(x) = \{v \mid f(x) + f^*(v) \le \langle x, v\rangle + \varepsilon\}$ for any $\varepsilon \ge 0$ and $x \in \mathbb{R}^n$;*
(c) *if $v \in \partial f(x)$ and $f(y) < \infty$, then $v \in \partial_\varepsilon f(y)$, where $\varepsilon := f(y) - [f(x) + \langle y - x, v\rangle]$.*

Note that, due to the definition of T^ε, the verification of the inclusion $v \in T^\varepsilon(x)$ requires checking an infinite number of inequalities. This verification is feasible only for specially-structured instances of operators T. However, it is possible to compute points in the graph of T^ε using the following *weak transportation formula* [3].

THEOREM 2.3 (see [3, Theorem 2.3]). *Suppose that $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone. Let $x_i, v_i \in \mathbb{R}^n$ and $\varepsilon_i, \alpha_i \in \mathbb{R}_+$, for $i = 1, \dots, k$, be such that*

$$v_i \in T^{\varepsilon_i}(x_i), \quad i = 1, \dots, k, \qquad \sum_{i=1}^k \alpha_i = 1,$$

and define

$$\bar{x} = \sum_{i=1}^k \alpha_i x_i, \quad \bar{v} = \sum_{i=1}^k \alpha_i v_i, \quad \bar{\varepsilon} = \sum_{i=1}^k \alpha_i \varepsilon_i + \alpha_i \langle x_i - \bar{x}, v_i - \bar{v}\rangle. \tag{3}$$

Then, the following statements hold:

(a) $\bar{\varepsilon} \geq 0$ *and* $\bar{v} \in T^{\bar{\varepsilon}}(\bar{x})$.

(b) *If, in addition,* $T = \partial f$ *for some proper lower semicontinuous convex function* f *and* $v_i \in \partial_{\varepsilon_i} f(x_i)$ *for* $i = 1, \ldots, k$, *then* $\bar{v} \in \partial_{\bar{\varepsilon}} f(\bar{x})$.

Whenever necessary, we will identify a map $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ with the point-to-set operator $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$,

$$F(x) = \begin{cases} \{F(x)\}, & x \in \Omega\,, \\ \emptyset & \text{otherwise.} \end{cases}$$

The following result is an immediate consequence of Theorem 2.3.

Corollary 2.4. *Let a monotone map* $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$, *points* $x_1, \ldots, x_k \in \mathbb{R}^n$, *and nonnegative scalars* $\alpha_1, \ldots, \alpha_k$ *such that* $\sum_{i=1}^k \alpha_i = 1$ *be given. Define*

$$\bar{x} := \sum_{i=1}^k \alpha_i x_i, \quad \bar{F} := \sum_{i=1}^k \alpha_i F(x_i), \quad \bar{\varepsilon} := \sum_{i=1}^k \alpha_i \langle x_i - \bar{x}, F(x_i) - \bar{F} \rangle. \tag{4}$$

Then, $\bar{\varepsilon} \geq 0$ *and* $\bar{F} \in F^{\bar{\varepsilon}}(\bar{x})$.

Proof. First use Zorn's lemma to conclude that there exist a maximal monotone $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ which extends F, that is, $F \subset T$. To end the proof, apply Theorem 2.3 to T and use the assumption that it extends F. ∎

Definition 1. *For a constant* $L \geq 0$, *the map* $F : \Omega \subset \mathbb{R}^m \to \mathbb{R}^n$ *is said to be* L*-Lipschitz continuous on* Ω *if* $\|F(x) - F(\tilde{x})\| \leq L\|x - \tilde{x}\|$ *for any* $x, \tilde{x} \in \Omega$.

Now we are ready to prove that, for a monotone Lipschitz continuous map F defined on the whole space $\mathbb{R}^n$, the distance between any vector in $F^{\varepsilon}(x)$ and $F(x)$ is proportional to $\sqrt{\varepsilon}$.

Proposition 2.5. *If* $F : \mathbb{R}^n \to \mathbb{R}^n$ *is monotone and* L*-Lipschitz continuous on* $\mathbb{R}^n$, *then for every* $x \in \mathbb{R}^n$, $\varepsilon \geq 0$, *and* $v \in F^{\varepsilon}(x)$,

$$\|F(x) - v\| \leq 2\sqrt{L\varepsilon}.$$

Proof. Let $v \in F^{\varepsilon}(x)$ be given. Then, for any $\tilde{x} \in \mathbb{R}^n$, we have

$$\begin{aligned} \langle F(x) - v, \tilde{x} - x \rangle &= \langle F(\tilde{x}) - v, \tilde{x} - x \rangle - \langle F(\tilde{x}) - F(x), \tilde{x} - x \rangle \\ &\geq -\varepsilon - \|F(\tilde{x}) - F(x)\| \|\tilde{x} - x\| \geq -\varepsilon - L\|\tilde{x} - x\|^2, \end{aligned}$$

where the first inequality follows from the definition of F^{ε} and the Cauchy–Schwarz inequality, and the second one from the assumption that $F : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous. Specializing this inequality for $\tilde{x} = x + (2L)^{-1}p$, where $p = v - F(x)$, we obtain $\|p\| \leq 2\sqrt{L\varepsilon}$. ∎

Corollary 2.6. *Let a monotone map* $F : \mathbb{R}^n \to \mathbb{R}^n$, *points* $x_1, \ldots, x_k \in \mathbb{R}^n$, *and nonnegative scalars* $\alpha_1, \ldots, \alpha_k$ *such that* $\sum_{i=1}^k \alpha_i = 1$ *be given and define* $\bar{x}$, $\bar{F}$, *and* $\bar{\varepsilon}$ *as in Corollary* 2.4. *Then,* $\bar{\varepsilon} \geq 0$ *and*

$$\|F(\bar{x}) - \bar{F}\| \leq 2\sqrt{\bar{\varepsilon}\, L}. \tag{5}$$

Proof. By Corollary 2.4, we have $\bar{F} \in F^{\bar{\varepsilon}}(\bar{x})$, which together with Proposition 2.5 implies that (5) holds. ∎

We observe that when F is an affine (monotone) map, the left-hand side of (5) is zero in view of (4). Hence, in this case, the right-hand side of (5) is a poor estimate of the error $F(\bar{x}) - \bar{F}$. We will now develop a better estimate of this error which depends on a certain constant which measures the nonlinearity of a monotone map F.

DEFINITION 2. *For a monotone map $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$, let* $\mathrm{Nonl}(F;\Omega)$ *be the infimum of all $L \geq 0$ such that there exist an L-Lipschitz map G and an affine map $\mathcal{A}$ such that*

$$F = G + \mathcal{A},$$

with G and $\mathcal{A}$ monotone.

Clearly, if F is a monotone affine map, then $\mathrm{Nonl}(F;\mathbb{R}^n) = 0$. Note also that if F is monotone and L-Lipschitz on Ω, then $\mathrm{Nonl}(F;\Omega) \leq L$. We note however that $\mathrm{Nonl}(F;\Omega)$ can be much smaller than L for many relevant instances. For example, if $F = G + \mu\mathcal{A}$, where $\mu \geq 0$, $\mathcal{A}$ is a monotone affine map and the map G is monotone and L-Lipschitz on Ω, then we have $\mathrm{Nonl}(F;\Omega) \leq L$. Hence, in the latter case, $\mathrm{Nonl}(F;\Omega)$ is bounded by a constant which does not depend on μ while the Lipschitz constant of F converges to ∞ as $\mu \to \infty$ if $\mathcal{A}$ is not constant.

PROPOSITION 2.7. *Let a monotone map $F : \mathbb{R}^n \to \mathbb{R}^n$, points $x_1, \ldots, x_k \in \mathbb{R}^n$, and nonnegative scalars $\alpha_1, \ldots, \alpha_k$ such that $\sum_{i=1}^k \alpha_i = 1$ be given and define $\bar{x}$, $\bar{F}$, and $\bar{\varepsilon}$ as in Corollary* 2.4. *Then, $\bar{\varepsilon} \geq 0$ and*

$$\|F(\bar{x}) - \bar{F}\| \leq 2\sqrt{\bar{\varepsilon}\mathcal{N}_F}, \tag{6}$$

where $\mathcal{N}_F := \mathrm{Nonl}(F;\mathbb{R}^n)$.

Proof. Suppose that $F = G + \mathcal{A}$, where G is an L-Lipschitz monotone map and $\mathcal{A}$ is an affine monotone map. Define

$$\begin{aligned} \bar{G} &= \textstyle\sum_{i=1}^k \alpha_i G(x_i), \quad \bar{\varepsilon}^g := \textstyle\sum_{i=1}^k \alpha_i \langle G(x_i) - \bar{G}, x_i - \bar{x}\rangle, \\ \bar{a} &= \textstyle\sum_{i=1}^k \alpha_i \mathcal{A}(x_i), \quad \bar{\varepsilon}^a := \textstyle\sum_{i=1}^k \alpha_i \langle \mathcal{A}(x_i) - \bar{a}, x_i - \bar{x}\rangle. \end{aligned} \tag{7}$$

Since G is a monotone map, it follows from Corollary 2.6 that $\bar{\varepsilon}^g \geq 0$ and $\|G(\bar{x}) - \bar{G}\| \leq 2\sqrt{\bar{\varepsilon}^g L}$. Moreover, since $\mathcal{A}$ is affine and monotone, we conclude that $\bar{a} = \mathcal{A}(\bar{x})$ and $\bar{\varepsilon}^a \geq 0$. Also, noting that $\bar{F} = \bar{G} + \bar{a}$ and $\bar{\varepsilon} = \bar{\varepsilon}^a + \bar{\varepsilon}^g \geq \bar{\varepsilon}^g$, we conclude that

$$\|F(\bar{x}) - \bar{F}\| = \big\|[G(\bar{x}) + \mathcal{A}(\bar{x})] - [\bar{G} + \bar{a}]\big\| = \big\|G(\bar{x}) - \bar{G}\big\| \leq 2\sqrt{\bar{\varepsilon}^g L} \leq 2\sqrt{\bar{\varepsilon} L}.$$

Bound (6) now follows by noting that $\mathcal{N}_F$ is the infimum of all L for G and $\mathcal{A}$ as above. ∎

3. Approximate solutions of the VI problem. In this section, we introduce two notions of approximate solutions of the VI problem. We then discuss how the monotone VI problem can be viewed as a special instance of the monotone inclusion problem and interpret these two notions of approximate solutions for the VI problem in terms of criterion (1).

We assume throughout this section that the following assumptions hold:

(A.1) $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ is a continuous monotone map.

(A.2) $X \subset \Omega$ is a nonempty closed convex set.

The monotone VI problem with respect to the pair (F, X), denoted by $VIP(F, X)$, consists of finding x^* such that

$$x^* \in X, \qquad \langle x - x^*, F(x^*)\rangle \geq 0 \quad \forall x \in X. \tag{8}$$

It is well known that, under the above assumptions, condition (8) is equivalent to

$$x^* \in X, \qquad \langle x - x^*, F(x)\rangle \geq 0 \quad \forall x \in X. \tag{9}$$

We will now introduce two notions of approximate solutions of the *VIP*(F, X), which are essentially relaxations of the characterizations (8) and (9) of an exact solution of the VI problem.

DEFINITION 3. *A point* $x \in X$ *is an* ε-strong solution *of VIP*(F, X) *if*

$$\theta^s(x; F) := \sup_{y \in X} \langle F(x), x - y \rangle \le \varepsilon, \tag{10}$$

and is an ε-weak solution *of VIP*(F, X) *if*

$$\theta^w(x; F) := \sup_{y \in X} \langle F(y), x - y \rangle \le \varepsilon. \tag{11}$$

The functions θ^s and $-\theta^w$ are referred to as the *gap function* and the *dual gap function*, respectively, in [5]. Note that, due to the monotonicity of F, we have $0 \le \theta^w(\cdot; F) \le \theta^s(\cdot; F)$, and hence every ε-strong solution is also an ε-weak solution.

For VI problems with unbounded feasible sets, the two above notions of approximate solutions are too strong. For example, if $X = \mathbb{R}^n$, the set of ε-strong solutions agree with the solution set. The following definition relaxes the above notions.

DEFINITION 4. *A point* $x \in X$ *is an* (ρ, ε)-strong solution *(resp.,* (ρ, ε)-weak solution*) of VIP*(F, X) *if, for some* $r \in \mathbb{R}^n$ *such that* $\|r\| \le \rho$*,* x *is an* ε*-strong (resp.,* ε*-weak) solution of VIP*$(F - r, X)$*; that is,*

$$\theta^s(x; F - r) = \sup_{y \in X} \langle F(x) - r, x - y \rangle \le \varepsilon, \tag{12}$$

$$\left(\textit{resp.,}\;\; \theta^w(x; F - r) = \sup_{y \in X} \langle F(y) - r, x - y \rangle \le \varepsilon \right).$$

Moreover, any such pair (r, ε) *will be called a* strong *(resp.,* weak*)* residual *of* x *for VIP*(F, X)*.*

We will provide some discussion about a (ρ, ε)-strong solution. First, a (ρ, ε)-strong solution is also a (ρ, ε)-weak solution. Second, it will be shown in Appendix A that, if F is L-Lipschitz continuous, then every (ρ, ε)-weak solution of *VIP*(F, X) is also a $(\rho + 2\sqrt{L\varepsilon},\ \varepsilon)$-strong solution. Third, as opposed to other notions of approximate solutions based on some gap function, it uses two tolerances which have very natural interpretations in the context of (monotone) complementarity problems. Indeed, if x is a (ρ, ε)-strong solution, then the following result, whose proof is postponed until the end of this section, shows that ρ measures the infeasibility of $F(x)$ with respect to the dual cone, while ε measures the size of the complementarity slackness.

PROPOSITION 3.1. *Assume that* $X = K$*, where* K *is a nonempty closed convex cone. Then,* $\bar{x} \in K$ *is a* (ρ, ε)*-strong solution if and only if there exists* $\bar{q} \in K^*$ *such that*

$$\|F(\bar{x}) - \bar{q}\| \le \rho, \qquad \langle \bar{x}, \bar{q} \rangle \le \varepsilon.$$

We will now characterize the above notions of approximate solutions for the *VIP*(F, X) in terms of ε-enlargements of certain maximal monotone operators. Recall that the *normal cone operator* of X is the point-to-set map $N_X : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$,

$$N_X(x) = \begin{cases} \emptyset, & x \notin X, \\ \{v \in \mathbb{R}^n, \mid \langle y - x, v \rangle \le 0 \;\forall y \in X\}, & x \in X. \end{cases} \tag{13}$$

From the above definition, it follows that (8) is equivalent to $-F(x^*) \in N_X(x^*)$, and hence to the monotone inclusion problem

$$0 \in (F + N_X)(x^*). \tag{14}$$

Note that the assumption on F and X guarantees maximal monotonicity of $F + N_X$ (see, for example, Proposition 12.3.6 of [5]).

It turns out that approximate solutions of $VIP(F, X)$ with weak, or strong, residual (r, ε) are related with certain approximate solutions of problem (14), as described by the following result.

PROPOSITION 3.2. *Let $\bar{x} \in X$ and pair $(r, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}_+$ be given. Then, the following equivalences hold:*

(a) *(r, ε) is a weak residual of $\bar{x}$ for $VIP(F, X)$ if and only if $r \in (F + N_X)^\varepsilon(\bar{x})$.*

(b) *(r, ε) is a strong residual of $\bar{x}$ for $VIP(F, X)$ if and only if $r \in (F + N_X^\varepsilon)(\bar{x})$.*

Proof. (a) Using definition (2), we have that $r \in (F + N_X)^\varepsilon(\bar{x})$ is equivalent to

$$\langle x - \bar{x}, F(x) + q - r\rangle \geq -\varepsilon \quad \forall x \in \mathbb{R}^n,\ q \in N_X(x).$$

Taking the infimum of the left-hand side for $q \in N_X(x)$ and using the fact that the domain of N_X is X and the assumption that $\bar{x} \in X$, we conclude that this condition is equivalent to $\theta^w(x; F - r) \leq \varepsilon$, i.e., to (r, ε) being a weak residual of $\bar{x}$ for $VIP(F, X)$.

(b) This equivalence follows from (a) with $r = r - F(\bar{x})$ and $F \equiv 0$. ∎

We will now provide a different characterization for $\bar{x}$ to be an approximate solution of $VIP(F, X)$ with strong residual (r, ε) for the case when the feasible set is a closed convex cone. We will first review a few well-known concepts.

The *indicator function* of X is the function $\delta_X : \mathbb{R}^n \to \bar{\mathbb{R}}$ defined as

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

The normal cone operator N_X of X can be expressed in terms of δ_X as $N_X = \partial \delta_X$. Direct use of definition (2) and the definition of ε-subdifferential shows that for $f = \delta_X$, inclusion on Proposition 2.2(a) holds as equality, i.e.,

$$(N_X)^\varepsilon = \partial_\varepsilon \delta_X. \tag{15}$$

We now state the following technical result which characterizes membership in $(N_X)^\varepsilon$ in terms of certain ε-complementarity conditions.

LEMMA 3.3. *If K is a nonempty closed convex cone and K^* is its dual cone, i.e.,*

$$K^* = \{v \in \mathbb{R}^n \mid \langle x, v\rangle \geq 0\ \forall x \in K\},$$

then, for every $x \in K$, we have

$$-q \in (N_K)^\varepsilon(x) \iff q \in K^*,\ \langle x, q\rangle \leq \varepsilon\,.$$

Proof. Since $(N_K)^\varepsilon = \partial_\varepsilon \delta_K$, it follows from Proposition 2.2(b) that the condition $-q \in (N_K)^\varepsilon(x)$ is equivalent to

$$(\delta_K)^*(-q) = \delta_K(x) + (\delta_K)^*(-q) \leq \langle x, -q\rangle + \varepsilon,$$

where the first equality follows from the assumption that $x \in K$. To end the proof, it suffices to use the fact that $(\delta_K)^* = \delta_{-K^*}$. ∎

With the aid of the above lemma, we can now give a characterization of strong residuals of feasible points for monotone complementarity problems.

PROPOSITION 3.4. *Assume that $X = K$, where K is a nonempty closed convex cone. Then, (r, ε) is a strong residual of $\bar{x} \in K$ for VIP(F, K) if and only if*

$$\bar{q} := F(\bar{x}) - r \in K^*, \quad \langle \bar{x}, \bar{q} \rangle \le \varepsilon.$$

Proof. This result follows as an immediate consequence of Proposition 3.2(b) and Lemma 3.3. □

Finally, note that Proposition 3.1 follows as an immediate consequence of Proposition 3.4.

4. The hybrid proximal extragradient method. Throughout this section, we assume that $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a maximal monotone operator. The monotone inclusion problem for T is to find x such that

$$0 \in T(x)\,. \tag{16}$$

We also assume throughout this section that this problem has a solution, that is, $T^{-1}(0) \neq \emptyset$. In this section we study the iteration complexity of the hybrid proximal extragradient method introduced in [22] for solving the above problem.

We start by stating the HPE method.

Hybrid proximal extragradient method:

0) Let $x_0 \in \mathbb{R}^n$ and $0 \le \sigma < 1$ be given and set $k = 1$.

1) Choose $\lambda_k > 0$ and find $y_k, v_k \in \mathbb{R}^n$, $\sigma_k \in [0, \sigma]$, and $\varepsilon_k \ge 0$ such that

$$v_k \in T^{\varepsilon_k}(y_k), \|\lambda_k v_k + y_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \le \sigma_k^2 \|y_k - x_{k-1}\|^2\,. \tag{17}$$

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 HPE method. First, the HPE method does not specify how to choose λ_k and how to find y_k, v_k, and ε_k as in (17). The particular choice of λ_k and the algorithm used to compute y_k, v_k, and ε_k will depend on the particular implementation of the method and the properties of the operator T. Second, if $y := (\lambda_k T + I)^{-1} x_{k-1}$ is the *exact* proximal point iterate, or equivalently,

$$v \in T(y), \tag{18}$$

$$\lambda_k v + y - x_{k-1} = 0 \tag{19}$$

for some $v \in \mathbb{R}^n$, then $(y_k, v_k) = (y, v)$ and $\varepsilon_k = 0$ satisfies (17). Therefore, the error criterion (17) relaxes the inclusion (18) to $v \in T^{\varepsilon}(y)$ and relaxes (19) by allowing a small error relative to $\|y_k - x_{k-1}\|$. Third, note also that due to step 2 of the HPE method and the error criterion in (17), we have

$$\lambda_k v_k + y_k - x_{k-1} = y_k - x_k, \qquad \|y_k - x_k\|^2 + 2\lambda_k \varepsilon_k \le \sigma_k^2 \|y_k - x_{k-1}\|^2. \tag{20}$$

Before establishing the iteration-complexity results for the HPE method, we need some technical results.

LEMMA 4.1. *For every $k \in \mathbb{N}$,*

$$(1 - \sigma)\|y_k - x_{k-1}\| \le \|\lambda_k v_k\| \le (1 + \sigma)\|y_k - x_{k-1}\|\,. \tag{21}$$

Proof. In view of (17) and the triangle inequality for norms, we have

$$| \, \|\lambda_k v_k\| - \|y_k - x_{k-1}\| \, | \le \|\lambda_k v_k + y_k - x_{k-1}\| \le \sigma_k \|y_k - x_{k-1}\| \quad \forall k \in \mathbb{N},$$

which clearly implies (21). ∎

LEMMA 4.2. *The following statements hold:*

(a) *For any $x \in \mathbb{R}^n$ and $i \in \mathbb{N}$,*

$$\|x - x_{i-1}\|^2 = \|x - x_i\|^2 + 2\lambda_i \langle y_i - x, v_i\rangle + \|y_i - x_{i-1}\|^2 - \|x_i - y_i\|^2 . \tag{22}$$

(b) *For any $x^* \in T^{-1}(0)$ and $i \in \mathbb{N}$,*

$$\|x^* - x_{i-1}\|^2 \ge \|x^* - x_i\|^2 + (1 - \sigma_i^2)\|y_i - x_{i-1}\|^2.$$

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

$$\|x^* - x_0\|^2 \ge \sum_{k=1}^{\infty} (1 - \sigma_k^2) \, \|y_k - x_{k-1}\|^2 \ge (1 - \sigma^2) \sum_{k=1}^{\infty} \|y_k - x_{k-1}\|^2 . \tag{23}$$

Proof. To prove (a), let $x \in \mathbb{R}^n$ and $i \in \mathbb{N}$. Then,

$$\begin{aligned}
\|x - x_{i-1}\|^2 &= \|x - x_i\|^2 + 2\langle x - x_i, x_i - x_{i-1}\rangle + \|x_i - x_{i-1}\|^2 \\
&= \|x - x_i\|^2 + 2\langle x - y_i, x_i - x_{i-1}\rangle + 2\langle y_i - x_i, x_i - x_{i-1}\rangle + \|x_i - x_{i-1}\|^2 \\
&= \|x - x_i\|^2 + 2\langle x - y_i, x_i - x_{i-1}\rangle + \|y_i - x_{i-1}\|^2 - \|y_i - x_i\|^2.
\end{aligned}$$

Statement (a) now follows by noting that $x_i - x_{i-1} = -\lambda_i v_i$, in view of step 2 of the HPE method.

If $0 \in T(x^*)$, then, since $v_i \in T^{\varepsilon_i}(y_i)$, the definition of T^{ε_k} implies that

$$\langle y_i - x^*, v_i\rangle = \langle y_i - x^*, v_i - 0\rangle \ge -\varepsilon_i.$$

Using statement (a) with $x = x^*$, the above inequality, and (17), we have

$$\begin{aligned}
\|x^* - x_{i-1}\|^2 &\ge \|x^* - x_i\|^2 - 2\lambda_i\varepsilon_i + \|y_i - x_{i-1}\|^2 - \|x_i - y_i\|^2 \\
&\ge \|x^* - x_i\|^2 + (1 - \sigma_i^2)\|y_i - x_{i-1}\|^2,
\end{aligned}$$

which proves (b). Statement (c) follows immediately from (b) and the assumptions $0 \le \sigma_i \le \sigma < 1$ (see steps 0 and 1 of the HPE method). ∎

LEMMA 4.3. *Let d_0 be the distance of x_0 to $T^{-1}(0)$. For every $\alpha \in \mathbb{R}$ and every k, there exists an $i \le k$ such that*

$$\|v_i\| \le d_0 \sqrt{\frac{(1+\sigma)}{(1-\sigma)} \left(\frac{\lambda_i^{\alpha-2}}{\sum_{j=1}^k \lambda_j^\alpha}\right)}, \qquad \varepsilon_i \le \frac{d_0^2 \sigma^2}{2(1-\sigma^2)} \left(\frac{\lambda_i^{\alpha-1}}{\sum_{j=1}^k \lambda_j^\alpha}\right). \tag{24}$$

Proof. Define, for each $k \in \mathbb{N}$,

$$\tau_k := \max\left\{\frac{2\varepsilon_k \lambda_k^{1-\alpha}}{\sigma^2}, \frac{\|v_k\|^2 \lambda_k^{2-\alpha}}{(1+\sigma)^2}\right\}.$$

Then, in view of the assumption that $\sigma_k \le \sigma$ for all $k \in \mathbb{N}$ and relations (17) and (21), we have

$$\tau_k \lambda_k^\alpha = \max\left\{\frac{2\varepsilon_k \lambda_k}{\sigma^2}, \frac{\lambda_k^2 \|v_k\|^2}{(1+\sigma)^2}\right\} \le \|y_k - x_{k-1}\|^2.$$

Letting $x^* \in T^{-1}(0)$ be such that $d_0 = \|x_0 - x^*\|$, the latter inequality together with (23) then implies that

$$\sum_{j=1}^{k} \tau_j \lambda_j^\alpha \le \sum_{j=1}^{k} \|y_j - x_{j-1}\|^2 \le \frac{\|x_0 - x^*\|^2}{(1-\sigma^2)} = \frac{d_0^2}{(1-\sigma^2)},$$

and hence that

$$\left(\min_{j=1,\ldots,k} \tau_j\right) \sum_{j=1}^{k} \lambda_j^\alpha \le \frac{d_0^2}{(1-\sigma^2)}.$$

The conclusion of the proposition now follows immediately from the latter inequality and the definition of τ_k. $\square$

THEOREM 4.4. *Let d_0 be the distance of x_0 to $T^{-1}(0)$. The following statements hold:*

(a) *If $\underline{\lambda} := \inf \lambda_k > 0$, then for every $k \in \mathbb{N}$ there exists $i \le k$ such that*

$$\|v_i\| \le d_0 \sqrt{\frac{1+\sigma}{1-\sigma} \left(\frac{\underline{\lambda}^{-1}}{\sum_{j=1}^k \lambda_j}\right)} \le \frac{d_0}{\underline{\lambda}\sqrt{k}} \sqrt{\frac{1+\sigma}{1-\sigma}},$$

$$\varepsilon_i \le \frac{\sigma^2 d_0^2}{2(1-\sigma^2)} \frac{1}{\sum_{i=1}^k \lambda_i} \le \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\underline{\lambda} k}.$$

(b) *For every $k \in \mathbb{N}$, there exists an index $i \le k$ such that*

$$\|v_i\| \le d_0 \sqrt{\frac{1+\sigma}{1-\sigma} \left(\frac{1}{\sum_{j=1}^k \lambda_j^2}\right)}, \qquad \varepsilon_i \le \frac{\sigma^2 d_0^2 \lambda_i}{2(1-\sigma^2)\sum_{j=1}^k \lambda_j^2}. \tag{25}$$

(c) *If $\sum_{k=1}^\infty \lambda_k^2 = \infty$, then the sequences $\{y_k\}$ and $\{x_k\}$ converge to some point in $T^{-1}(0)$.*

Proof. Statements (a) and (b) follow from Lemma 4.3 with α equal to 1 and 2, respectively.

To prove (c), first note that if $v_i = 0$ and $\varepsilon_i = 0$ for some i, then $x_i = y_i \in T^{-1}(0)$ and $x_k = x_i$ for all $k \ge i$. We may then assume that

$$a_i := \max\{\|v_i\|, \varepsilon_i\} > 0 \qquad \forall i.$$

The assumption that $\sum_{i=1}^\infty \lambda_i^2 = +\infty$ implies that

$$\lim_{k\to\infty} \max_{j=1,\ldots,k} \frac{\lambda_j}{\sum_{i=1}^k \lambda_i^2} = 0.$$

Therefore, using also (25) we conclude that

$$\lim_{k\to\infty} \min_{i=1,\ldots,k} a_i = 0.$$

Therefore, there exists a subsequence $\{a_i\}_{i\in\mathcal{K}}$ which converges to 0. Lemma 4.2(c) implies that $\{x_k\}$ is bounded and

$$\lim_{k\to\infty} \|y_k - x_{k-1}\| = 0. \tag{26}$$

Hence, $\{y_k\}$ is also bounded and this implies that there exists a subsequence $\{y_i\}_{i\in\mathcal{K}'}$, with $\mathcal{K}' \subset \mathcal{K}$, which converges to some y^*. Since $\lim_{i\in\mathcal{K}'} a_i = 0$, using Proposition 2.1(e) we conclude that $y^* \in T^{-1}(0)$. Now use (26) to conclude that y^* is an accumulation point of $\{x_k\}$. Since $\|x_k - y^*\|$ is nonincreasing in view of Lemma 4.2(c), we conclude that $\lim_{k\to\infty} \|x_k - y^*\| = 0$. In view of (26), the sequence $\{y_k\}$ also converges to y^*. ∎

Both Lemma 4.3 and Theorem 4.4 estimate the quality of the best among the iterates $y_1, \ldots, y_k$. We will refer to these estimates as the *pointwise* complexity bounds for the HPE algorithm.

We will now develop alternative estimates for the HPE method which we refer to as the *ergodic* complexity bounds. The idea of considering averages of the iterates in the analysis of gradient-type and/or proximal point–based methods for convex minimization and monotone VIs goes back to at least the mid-seventies (see [1, 13, 17, 16]) and perhaps even earlier.

The sequence of ergodic means $\{\bar{y}_k\}$ associated with $\{y_k\}$ is defined as

$$\bar{y}_k := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i y_i, \qquad \text{where} \qquad \Lambda_k := \sum_{i=1}^{k} \lambda_i. \tag{27}$$

The next result, which is a straightforward application of the transportation formula, shows that the ergodic iterate is related to the ε-enlargement of T, even when $\varepsilon_i \equiv 0$. Thus, it provides a computable residual pair for $\bar{y}_k$.

LEMMA 4.5. *For every $k \in \mathbb{N}$, define*

$$\bar{v}_k := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i v_i, \qquad \bar{\varepsilon}_k := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i (\varepsilon_i + \langle y_i - \bar{y}_k, v_i - \bar{v}_k \rangle). \tag{28}$$

Then, $\bar{\varepsilon}_k \geq 0$ and $\bar{v}_k \in T^{\bar{\varepsilon}_k}(\bar{y}_k)$.

Proof. The inequality $\bar{\varepsilon}_k \geq 0$ and the inclusion $\bar{v}_k \in T^{\bar{\varepsilon}_k}(\bar{y}_k)$ follow from Theorem 2.3 and the inclusions $v_i \in T^{\varepsilon_i}(y_i)$. ∎

The following result gives alternative expressions for the residual pair $(\bar{v}_k, \bar{\varepsilon}_k)$, which will be used for obtaining bounds on its size.

PROPOSITION 4.6. *For any k we have*

$$\bar{v}_k = \frac{1}{\Lambda_k}(x_0 - x_k), \tag{29}$$

$$\bar{\varepsilon}_k = \frac{1}{2\Lambda_k} \left[2\langle \bar{y}_k - x_0, x_k - x_0 \rangle - \|x_k - x_0\|^2 + \beta_k \right], \tag{30}$$

where

$$\beta_k := \sum_{i=1}^{k} \left(2\lambda_i \varepsilon_i + \|x_i - y_i\|^2 - \|y_i - x_{i-1}\|^2 \right) \leq 0\,. \tag{31}$$

Proof. The definitions of Λ_k and $\bar{v}_k$ in (27) and (28) and the update rule in step 2 of the HEP method imply that

$$x_k = x_0 - \sum_{i=1}^{k} \lambda_i v_i = x_0 - \Lambda_k \bar{v}_k,$$

from which (29) follows.

Direct use of the definition of $\bar{y}_k$ yields

$$\begin{aligned}\sum_{i=1}^{k}\lambda_i\langle y_i-\bar{y}_k, v_i-\bar{v}_k\rangle &= \sum_{i=1}^{k}\lambda_i\langle y_i-\bar{y}_k, v_i\rangle - \sum_{i=1}^{k}\lambda_i\langle y_i-\bar{y}_k, \bar{v}_k\rangle \\ &= \sum_{i=1}^{k}\lambda_i\langle y_i-\bar{y}_k, v_i\rangle - \left\langle \sum_{i=1}^{k}\lambda_i(y_i-\bar{y}_k), \bar{v}_k\right\rangle \\ &= \sum_{i=1}^{k}\lambda_i\langle y_i-\bar{y}_k, v_i\rangle\,.\end{aligned}$$

Adding (22) with $x=\bar{y}_k$ from $i=1$ to k, we have

$$\|\bar{y}_k-x_0\|^2 = \|\bar{y}_k-x_k\|^2 + \sum_{i=1}^{k}\left(2\lambda_i\langle y_i-\bar{y}_k, v_i\rangle + \|y_i-x_{i-1}\|^2 - \|x_i-y_i\|^2\right)\,.$$

Combining the above two equations with the definitions of $\bar{\varepsilon}_k$ in (28) and β_k in (31) we obtain

$$\bar{\varepsilon}_k = \frac{1}{2\Lambda_k}\left[\|\bar{y}_k-x_0\|^2 - \|\bar{y}_k-x_k\|^2 + \beta_k\right].$$

Relation (30) now follows from the above equation and the identity

$$\|\bar{y}_k-x_k\|^2 = \|\bar{y}_k-x_0\|^2 + 2\langle\bar{y}_k-x_0, x_0-x_k\rangle + \|x_0-x_k\|^2.$$

Finally, the inequality in (31) is due to (20) and the assumption $0\le\sigma_i\le\sigma<1$ (see steps 0 and 1 of the HPE method). ∎

The next result provides estimates on the quality measure of the ergodic mean $\bar{y}_k$. It essentially shows that the quantities $\bar{v}_k$ and $\bar{\varepsilon}_k$ appearing in (28) are $\mathcal{O}(1/\Lambda_k)$.

THEOREM 4.7. *For every $k\in\mathbb{N}$, let Λ_k, $\bar{y}_k$, $\bar{v}_k$, and $\bar{\varepsilon}_k$ be as* (27), (28). *Then, for every $k\in\mathbb{N}$, we have*

$$\|\bar{v}_k\| \le \frac{2d_0}{\Lambda_k}, \qquad \bar{\varepsilon}_k \le \frac{2\theta_k d_0^2}{\Lambda_k}, \tag{32}$$

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

$$\theta_k := 1 + \frac{\sigma\sqrt{\tau_k}}{\sqrt{(1-\sigma^2)}}, \qquad \tau_k = \max_{i=1,\ldots,k}\frac{\lambda_i}{\Lambda_k} \le 1. \tag{33}$$

Proof. Let $x^*\in T^{-1}(0)$ be such that $\|x_0-x^*\|=d_0$. Using Lemma 4.2(c), we have $\|x_k-x^*\|\le d_0$, and hence

$$\|x_k-x_0\| \le \|x_k-x^*\| + \|x^*-x_0\| \le 2d_0 \tag{34}$$

for every $k\in\mathbb{N}$. This, together with (29), yields the first bound in (32). Defining

$$\bar{x}_k = \frac{1}{\Lambda_k}\sum_{i=1}^{k}\lambda_i x_i,$$

and noting (20), (23), (27), (33), and (34), we have

$$\|\bar{x}_k - x_0\| \le \left\| \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i (x_i - x_0) \right\| \le \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i \|x_i - x_0\| \le 2d_0, \tag{35}$$

$$\|\bar{y}_k - \bar{x}_k\|^2 \le \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i \|y_i - x_i\|^2 \le \tau_k \sum_{i=1}^{k} \|y_i - x_i\|^2 \le \sigma^2 \tau_k \sum_{i=1}^{k} \|y_i - x_{i-1}\|^2 \le \frac{\sigma^2 \tau_k d_0^2}{(1-\sigma^2)}, \tag{36}$$

where the first inequalities in the above two relations are due to the convexity of $\|\cdot\|$ and $\|\cdot\|^2$, respectively. The above two relations together with (33) and the triangular inequality for norms yield

$$\|\bar{y}_k - x_0\| \le \|\bar{y}_k - \bar{x}_k\| + \|\bar{x}_k - x_0\| \le \sigma d_0 \sqrt{\frac{\tau_k}{1-\sigma^2}} + 2d_0 = (1+\theta_k)\, d_0.$$

Expressions (30) and (31), the Cauchy–Schwarz inequality, and the above relation then imply

$$\bar{\varepsilon}_k \le \frac{1}{2\Lambda_k} \left[-\|x_k - x_0\|^2 + 2\|\bar{y}_k - x_0\|\, \|x_k - x_0\| \right] \le \frac{1}{2\Lambda_k} \left[-t_k^2 + 2\,(1+\theta_k)\, d_0\, t_k \right],$$

where $t_k := \|x_k - x_0\|$. Since $0 \le t_k \le 2d_0$ and $\theta_k > 1$ by (34) and (33), respectively, it follows that the maximum of the right-hand side of the above relation, with respect to t_k, is attained at $2d_0$. This clearly implies the second inequality in (32). ∎

5. Korpelevich's extragradient method for the monotone VIP. Our main goal in this section is to establish the complexity analysis of Korpelevich's extragradient method for solving the monotone VI problem over an unbounded feasible set. First, we state Korpelevich's extragradient algorithm and show that it can be interpreted as a particular case of the HPE method. This allows us to use the results of section 4 to derive its iteration complexities for computing different notions of approximate solutions. In subsection 5.1, we obtain additional iteration-complexity results under the assumption that F is defined in the whole space $\mathbb{R}^n$ (e.g., when F is linear) and/or X is a closed convex cone. In subsection 5.2, we state the consequences of the aforementioned iteration-complexity results for the case when X is bounded.

Korpelevich's method, as well as its global convergence proof, was presented for the first time in [11]. A unifying global convergence analysis of the proximal point method and Korpelevich's method for solving $VIP(F, \mathbb{R}^n)$ is presented in [7] using the concept of modified monotone maps. Results showing that Korpelevich's method converges at a linear rate under strong assumptions on the VIP are given in [26].

Throughout this section, unless otherwise explicitly mentioned, we assume that the set $X \subset \mathbb{R}^n$ and the map $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ satisfy the following assumptions:

(B.1) $X \subset \Omega$ is a nonempty closed convex set.
(B.2) F is monotone and L-Lipschitz continuous (on Ω).
(B.3) The set of solutions of $VIP(F, X)$ is nonempty.

We start by stating Korpelevich's extragradient algorithm. The notation P_X denotes the projection operator onto the set X.

Korpelevich's extragradient algorithm:

0) Let $x_0 \in X$ and $0 < \sigma < 1$ be given and set $\lambda = \sigma/L$ and $k = 1$.
1) Compute

$$y_k = P_X(x_{k-1} - \lambda F(x_{k-1})), \qquad x_k = P_X(x_{k-1} - \lambda F(y_k)). \tag{37}$$

2) Set $k \leftarrow k+1$ and go to step 1.

end

We observe that assumptions (B.1) and (B.2) imply that the operator $T = F + N_X$ is maximal monotone (see, for example, Proposition 12.3.6 of [5]). Also, recall that solving $VIP(F,X)$ is equivalent to solving the monotone inclusion problem $0 \in T(x)$, where $T = F + N_X$ (see the discussion on the paragraph following Proposition 3.1).

We will now show that Korpelevich's extragradient algorithm for solving VIP (F,X) can be viewed as a particular case of the HEP method for solving the monotone inclusion problem $0 \in T(x)$, and this will allow us to obtain iteration-complexity bounds for Korpelevich's extragradient algorithm without assuming boundedness of the feasible set X.

THEOREM 5.1. *Let $\{y_k\}$ and $\{x_k\}$ be the sequences generated by Korpelevich's extragradient algorithm and, for each $k \in \mathbb{N}$, define*

$$q_k = \frac{1}{\lambda}\left[x_{k-1} - \lambda F(y_k) - x_k\right], \quad \varepsilon_k = \langle q_k, x_k - y_k\rangle, \quad v_k = F(y_k) + q_k. \tag{38}$$

Then,

(a) $x_k = x_{k-1} - \lambda v_k$;
(b) $q_k \in \partial_{\varepsilon_k}\delta_X(y_k)$ *and* $v_k \in (F + N_X^{\varepsilon_k})(y_k) \subset (F + N_X)^{\varepsilon_k}(y_k)$;
(c) $\|\lambda v_k + y_k - x_{k-1}\|^2 + 2\lambda\varepsilon_k \le \sigma^2\|y_k - x_{k-1}\|^2$.

As a consequence of the above statements, it follows that Korpelevich's algorithm is a special case of the HPE method.

Proof. Statement (a) follows immediately from the definition of q_k and v_k in (38). Recall that the projection map P_X has the property that

$$z - P_X(z) \in N_X(P_X(z)) \qquad \forall z \in \mathbb{R}^n. \tag{39}$$

Using this fact together with the definition of x_k and q_k in (37) and (38), respectively, it follows that

$$q_k \in N_X(x_k) = \partial\delta_X(x_k). \tag{40}$$

The first inclusion in statement (b) follows from (40) and Proposition 2.2(c). Using this inclusion, the definition of v_k, and identity (15), we obtain

$$\begin{aligned} v_k = F(y_k) + q_k \in F(y_k) + \partial_{\varepsilon_k}\delta_X(y_k) &= F(y_k) + (N_X)^{\varepsilon_k}(y_k) \\ \subset F^0(y_k) + (N_X)^{\varepsilon_k}(y_k) &\subset (F + N_X)^{\varepsilon_k}(y_k), \end{aligned}$$

where the two last inclusions follow from the monotonicity of F and statement (c) and statement (b) (with $\varepsilon' = 0$) of Proposition 2.1.

To prove statement (c), define

$$p_k = \frac{1}{\lambda}\left[x_{k-1} - \lambda F(x_{k-1}) - y_k\right]. \tag{41}$$

Using this definition, the definition of y_k in (37), and (39), we conclude that $p_k \in N_X(y_k)$. This fact, together with (38), gives the estimate

$$\varepsilon_k = \langle q_k - p_k, x_k - y_k \rangle + \langle p_k, x_k - y_k \rangle \leq \langle q_k - p_k, x_k - y_k \rangle.$$

This, together with (a), implies that

$$\begin{aligned} \|\lambda v_k + y_k - x_{k-1}\|^2 + 2\lambda\varepsilon_k &= \|y_k - x_k\|^2 + 2\lambda\varepsilon_k \\ &\leq \|x_k - y_k\|^2 + 2\lambda\langle q_k - p_k, x_k - y_k \rangle \\ &= \|\lambda(q_k - p_k) + x_k - y_k\|^2 - \lambda^2\|q_k - p_k\|^2 \\ &\leq \|\lambda(q_k - p_k) + x_k - y_k\|^2 = \|\lambda(F(x_{k-1}) - F(y_k))\|^2, \end{aligned}$$

where the last equality follows from the definition of q_k and p_k in (38) and (41). Statement (c) follows from the previous inequality, the assumption that $\lambda = \sigma/L$, and our global assumption that F is L-Lipschitz continuous on X. ∎

The following result establishes the global convergence rate of Korpelevich's extragradient algorithm in terms of the criterion (1) with $T = F + N_X$.

THEOREM 5.2. *Let $\{y_k\}$ and $\{x_k\}$ be the sequences generated by Korpelevich's extragradient algorithm and let $\{v_k\}$ and $\{\varepsilon_k\}$ be the sequences given by* (38). *For every $k \in \mathbb{N}$, define*

$$\bar{v}_k = \frac{1}{k}\sum_{i=1}^{k} v_i, \quad \bar{y}_k = \frac{1}{k}\sum_{i=1}^{k} y_i, \tag{42}$$

$$\bar{\varepsilon}_k = \frac{1}{k}\sum_{i=1}^{k} \left[\varepsilon_i + \langle y_i - \bar{y}_k, v_i - \bar{v}_k \rangle\right]. \tag{43}$$

Then, for every $k \in \mathbb{N}$, the following statements hold:

(a) *y_k is an approximate solution of VIP(F, X) with strong residual (v_k, ε_k), and there exists $i \leq k$ such that*

$$\|v_i\| \leq \frac{Ld_0}{\sigma}\sqrt{\frac{(1+\sigma)}{k(1-\sigma)}}, \quad \varepsilon_i \leq \frac{\sigma L d_0^2}{2(1-\sigma^2)k}.$$

(b) *$\bar{y}_k$ is an approximate solution of VIP(F, X) with weak residual $(\bar{v}_k, \bar{\varepsilon}_k)$, and*

$$\|\bar{v}_k\| \leq \frac{2Ld_0}{k\sigma}, \qquad \bar{\varepsilon}_k \leq \frac{2Ld_0^2\bar{\theta}_k}{k\sigma}, \tag{44}$$

where d_0 is the distance of x_0 to the solution set of VIP(F, X) and

$$\bar{\theta}_k := 1 + \frac{\sigma}{\sqrt{k(1-\sigma^2)}}. \tag{45}$$

Proof. (a) By Theorem 5.1(b), we have that $v_k \in (F + N_X^{\varepsilon_k})(y_k)$. Hence, by Proposition 3.2(b), we conclude that y_k is an approximate solution of *VIP*(F, X) with strong residual (v_k, ε_k). Also, Theorem 5.1 implies that Korpelevich's extragradient algorithm is a special case of the general HPE method of section 4, where $\lambda_k = \sigma/L$ for all $k \in \mathbb{N}$. Hence, the remaining claim in (a) follows from Theorem 4.4(a) with $\underline{\lambda} = \sigma/L$.

(b) By Theorem 5.1(b) and Lemma 4.5 with $T = F + N_X$, we conclude that $\bar{v}_k \in (F+N_X)^{\bar{\varepsilon}_k}(\bar{y}_k)$. In view of Proposition 3.2(a), this implies that $\bar{y}_k$ is an approximate solution of $VIP(F,X)$ with weak residual $(\bar{v}_k, \bar{\varepsilon}_k)$. The bounds in (44) follow from Theorem 4.7 with $T = F + N_X$ and $\lambda_k = \sigma/L$, and the fact that Λ_k, τ_k, and θ_k defined in (27) and (33) are, in this case, equal to $k\lambda/L$, $1/k$, and $\bar{\theta}_k$, respectively, where $\bar{\theta}_k$ is given by (45). ☐

Observe that the derived bounds obtained in (b) are asymptotically better than the ones obtained in (a). Indeed, while the bounds for ε_k and $\bar{\varepsilon}_k$ are $\mathcal{O}(1/k)$, the ones for v_k and $\bar{v}_k$ are $\mathcal{O}(1/\sqrt{k})$ and $\mathcal{O}(1/k)$, respectively. However, it should be emphasized that (a) describes the quality of the *strong* residual of some point among the iterates $y_1, \ldots, y_k$ while (b) describes the quality of the *weak* residual of $\bar{y}_k$.

The following result, which is an immediate consequence of Theorem 5.2, presents iteration-complexity bounds for Korpelevich's extragradient algorithm to obtain (ρ,ε)-weak and -strong solutions of $VIP(F,X)$. For simplicity, we ignore the dependence of these bounds on the parameter σ and other universal constants and express them only in terms of L, d_0, and the tolerances ρ and ε.

COROLLARY 5.3. *Consider the sequence $\{y_k\}$ generated by Korpelevich's extragradient algorithm and the sequence $\{\bar{y}_k\}$ defined as in* (42)*. Then, for every pair of positive scalars (ρ,ε), the following statements hold:*

(a) *There exists an index*

$$i = \mathcal{O}\left(\max\left[\frac{Ld_0^2}{\varepsilon}, \frac{L^2 d_0^2}{\rho^2}\right]\right) \tag{46}$$

such that the iterate y_i is a (ρ,ε)-strong solution of $VIP(F,X)$.

(b) *There exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{Ld_0^2}{\varepsilon}, \frac{Ld_0}{\rho}\right]\right) \tag{47}$$

such that, for any $k \geq k_0$, the point $\bar{y}_k$ is a (ρ,ε)-weak solution of $VIP(F,X)$.

5.1. Specialized complexity results for computing strong solutions. In this subsection we will obtain additional complexity results assuming that F is defined in all R^n, and/or that the feasible set is a cone.

We first establish the following preliminary result.

LEMMA 5.4. *Let $\{y_k\}$ and $\{x_k\}$ be the sequences generated by Korpelevich's extragradient algorithm and let $\{v_k\}$, $\{q_k\}$, and $\{\varepsilon_k\}$ be the sequences given by* (38)*. For every $k \in \mathbb{N}$, define*

$$\bar{F}_k = \frac{1}{k}\sum_{i=1}^{k} F(y_i), \quad \bar{q}_k = \frac{1}{k}\sum_{i=1}^{k} q_i, \tag{48}$$

$$\varepsilon'_k = \frac{1}{k}\sum_{i=1}^{k}\langle y_i - \bar{y}_k, F(y_i) - \bar{F}_k\rangle, \qquad \varepsilon''_k = \frac{1}{k}\sum_{i=1}^{k}\left[\varepsilon_i + \langle y_i - \bar{y}_k, q_i - \bar{q}_k\rangle\right]. \tag{49}$$

Then, for every $k \in \mathbb{N}$, we have

$$\bar{F}_k \in F^{\varepsilon'_k}(\bar{y}_k), \quad \bar{q}_k \in (N_X)^{\varepsilon''_k}(\bar{y}_k), \quad \bar{v}_k = \bar{F}_k + \bar{q}_k, \tag{50}$$

$$\bar{\varepsilon}_k = \varepsilon'_k + \varepsilon''_k, \quad \varepsilon'_k,\, \varepsilon''_k \geq 0. \tag{51}$$

Proof. Applying Theorem 2.3 with $T = F$, $x_i = y_i$, $v_i = F(x_i)$, $\varepsilon_i = 0$, and $\alpha_i = 1/k$ for $i = 1, \ldots, k$, we conclude that $\bar{F}_k \in F^{\varepsilon'_k}(\bar{y}_k)$ and $\varepsilon'_k \geq 0$. Also, it follows from Theorems 5.1(b) and 2.3 with $T = N_X$, $x_i = y_i$, $v_i = q_i$, and $\alpha_i = 1/k$ for $i = 1, \ldots, k$ that $\bar{q}_k \in (N_X)^{\varepsilon''_k}(\bar{y}_k)$ and $\varepsilon''_k \geq 0$. The identity $\bar{v}_k = \bar{F}_k + \bar{q}_k$ follows from (42), (48), and the fact that $v_i = F(y_i) + q_i$, $i = 1, \ldots, k$. The other identity $\bar{\varepsilon}_k = \varepsilon'_k + \varepsilon''_k$ now follows from (49) and the fact that $\bar{v}_k = \bar{F}_k + \bar{q}_k$ and $v_i = F(y_i) + q_i$, $i = 1, \ldots, k$. ∎

The following result shows that, if F is defined in all $\mathbb{R}^n$, satisfies assumptions (B.1)–(B.3) (and hence is L-Lipschitz continuous on the whole $\mathbb{R}^n$), and $\mathcal{N}_F < L$, then the iteration complexity for the ergodic point $\bar{y}_k$ to be a (ρ, ε)-*strong* solution is better than the iteration complexity for the best of the iterates among $y_1, \ldots, y_k$ to be a (ρ, ε)-*strong* solution. Moreover, when F is affine, it is shown that the dependence on the tolerance ρ of the first complexity is $\mathcal{O}(1/\rho)$ while that of the second complexity is $\mathcal{O}(1/\rho^2)$.

THEOREM 5.5 (F defined in all $\mathbb{R}^n$). *In addition to assumptions* (B.1)–(B.3), *assume that* $\Omega = \mathbb{R}^n$. *Let* $\{y_k\}$ *and* $\{x_k\}$ *be the sequences generated by Korpelevich's extragradient algorithm, let* $\{\bar{y}_k\}$, $\{\bar{v}_k\}$, $\{\varepsilon''_k\}$, *and* $\{\bar{q}_k\}$ *be the sequences defined as in* (42), (43), *and* (48), *and define*

$$\hat{v}_k := F(\bar{y}_k) + \bar{q}_k \quad \forall k \in \mathbb{N}. \tag{52}$$

Then, for every $k \in \mathbb{N}$, $\bar{y}_k$ *is an approximate solution of VIP*(F, X) *with strong residual* $(\hat{v}_k, \varepsilon''_k)$, *and the following bounds on* $\|\hat{v}_k\|$ *and* ε''_k *hold:*

$$\varepsilon''_k \leq \frac{2Ld_0^2\bar{\theta}_k}{k\sigma}, \qquad \|\hat{v}_k\| \leq \frac{d_0\sqrt{8\bar{\theta}_k L\mathcal{N}_F}}{\sqrt{k\sigma}} + \frac{2Ld_0}{k\sigma}, \tag{53}$$

where $\mathcal{N}_F := \mathrm{Nonl}(F; \mathbb{R}^n)$, $\bar{\theta}_k$ *is defined in* (45), *and* d_0 *is the distance of* x_0 *to the solution set of VIP*(F, X). *As a consequence, the following statements hold:*

(a) *For every pair of positive scalars* (ρ, ε), *there exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{Ld_0^2}{\varepsilon}, \frac{Ld_0}{\rho} + \frac{d_0^2 L\mathcal{N}_F}{\rho^2}\right]\right) \tag{54}$$

such that, for any $k \geq k_0$, *the point* $\bar{y}_k$ *is a* (ρ, ε)*-strong solution of VIP*(F, X).

(b) *If F is also affine, then, for every pair of positive scalars* (ρ, ε), *there exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{Ld_0^2}{\varepsilon}, \frac{Ld_0}{\rho}\right]\right) \tag{55}$$

such that, for any $k \geq k_0$, *the point* $\bar{y}_k$ *is a* (ρ, ε)*-strong solution of VIP*(F, X).

Proof. (a) First note that (52) and (50) imply that $\hat{v}_k = F(\bar{y}_k) + \bar{q}_k \in (F + N_X^{\varepsilon''_k})(\bar{y}_k)$, from which we conclude that $\bar{y}_k$ is an approximate solution of *VIP*(F, X) with strong residual $(\hat{v}_k, \varepsilon''_k)$, in view of Proposition 3.2(b). The first bound in (53) follows immediately from the second bound in (44) and the fact that $\varepsilon''_k \leq \bar{\varepsilon}_k$ in view of (51).

Now, (50), (52), Proposition 2.7 with $x_i = y_i$, and the triangle inequality for norms yield

$$\|\hat{v}_k\| \leq \|\bar{v}_k\| + \|\hat{v}_k - \bar{v}_k\| = \|\bar{v}_k\| + \|F(\bar{y}_k) - \bar{F}_k\| \leq \|\bar{v}_k\| + 2\sqrt{\bar{\varepsilon}_k\mathcal{N}_F}.$$

The second bound in (53) now follows from the above inequality and the bounds in (44).

Statement (a) follows from the bounds in (53), the definition of (ρ, ε)-strong solution, and some straightforward arguments. Statement (b) is a special case of (a), where $\mathcal{N}_F = 0$. ∎

In many important instances of F (e.g., see the discussion after Definition 2), the constant $\mathcal{N}_F = \mathrm{Nonl}(F; \mathbb{R}^n)$ is much smaller than L, and hence bound (54) can be much smaller than (46) on Corollary 5.3.

In the following result, we consider the situation where $X = K$ is a closed convex cone and $VIP(F, K)$ becomes equivalent to the following monotone complementarity problem:

$$0 = F(y) - s, \quad \langle y, s \rangle = 0, \quad (y, s) \in K \times K^*.$$

Using Proposition 3.4, we can translate the conclusions of Corollary 5.3(a) and Theorem 5.5 to the context of the above problem as follows.

COROLLARY 5.6 (monotone complementarity problems). *In addition to assumptions* (B.2)–(B.3)*, assume that $X = K$, where K is a nonempty closed convex cone. Consider the sequences $\{x_k\}$ and $\{y_k\}$ generated by Korpelevich's extragradient algorithm applied to $VIP(F, K)$ and the sequences $\{q_k\}$, $\{\bar{y}_k\}$, and $\{\bar{q}_k\}$ determined according to* (38), (42), *and* (48), *respectively. Then, for any pair of positive scalars (ρ, ε), the following statements hold:*

(a) *There exists an index*

$$k = \mathcal{O}\left(\max \left[\frac{L d_0^2}{\varepsilon}, \frac{L^2 d_0^2}{\rho^2} \right] \right)$$

such that the pair $(y, s) = (y_k, -q_k)$ satisfies

$$\|F(y) - s\| \le \rho, \quad \langle y, s \rangle \le \varepsilon, \quad (y, s) \in K \times K^*. \tag{56}$$

(b) *If $\Omega = \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^n$ is monotone and L-Lipschitz continuous on $\mathbb{R}^n$, then there exists an index*

$$k_0 = \mathcal{O}\left(\max \left[\frac{L d_0^2}{\varepsilon}, \frac{L d_0}{\rho} + \frac{d_0^2 L \mathcal{N}_F}{\rho^2} \right] \right), \tag{57}$$

where $\mathcal{N}_F := \mathrm{Nonl}(F; \mathbb{R}^n)$, such that, for any $k \ge k_0$, the pair $(y, s) = (\bar{y}_k, -\bar{q}_k)$ satisfies (56). *In particular, if F is affine, then the iteration-complexity bound* (57) *reduces to the one in* (55).

5.2. Bounded feasible set. So far, we have obtained complexity results for Korpelevich's method that *do not require* boundedness of the feasible set. In this subsection we will give some consequences of these results for the case where the feasible set X is bounded.

The following simple result shows that, when X is bounded, every approximate solution of $VIP(F, X)$ with weak (resp., strong) residual (r, ε) is an ε'-weak (resp., ε'-strong) solution of $VIP(F, X)$ for some ε'. Recall that the diameter $\mathcal{D}_X$ of a set $X \subset \mathbb{R}$ is defined as

$$\mathcal{D}_X := \sup\{\|x_1 - x_2\| : x_1, x_2 \in X\}. \tag{58}$$

LEMMA 5.7. *Assume that X has finite diameter $\mathcal{D}_X$. For any $\bar{x} \in X$, if $(r, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}_+$ is a weak (resp., strong) residual of $\bar{x}$ for VIP(F, X), then $\bar{x}$ is an ε'-weak (resp., ε'-strong) solution of VIP(F, X), where*

$$\varepsilon' := \varepsilon + \sup_{x \in X} \langle r, \bar{x} - x \rangle \leq \varepsilon + \|r\| \mathcal{D}_X.$$

Proof. This result follows from Definitions 3 and 4, the definition of $\mathcal{D}_X$, and the Cauchy–Schwarz inequality. ∎

The following result derives alternative global convergence rate bounds for Korpelevich's extragradient algorithm in the case where the feasible set X of $VIP(F, X)$ is bounded.

THEOREM 5.8 (bounded feasible sets). *Assume that conditions* (B.1)–(B.3) *hold and that the set X has finite diameter*

$$\mathcal{D}_X < \infty.$$

Consider the sequences $\{x_k\}$ and $\{y_k\}$ generated by Korpelevich's extragradient algorithm applied to VIP(F, X) and the sequences $\{\bar{y}_k\}$ and $\{\bar{\varepsilon}_k\}$ determined according to (42) *and* (43)*, respectively. Then, for every $k \in \mathbb{N}$, the following statements hold:*

(a) *$\bar{y}_k$ is an $\tilde{\varepsilon}_k$-weak solution of VIP(F, X), or equivalently,*

$$\max_{x \in X} \langle F(x), \bar{y}_k - x \rangle \leq \hat{\varepsilon}_k,$$

where

$$\tilde{\varepsilon}_k := \frac{2Ld_0}{k\sigma} \left(\mathcal{D}_X + d_0 \bar{\theta}_k \right).$$

(b) *If $\Omega = \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^n$ is monotone and L-Lipschitz continuous on $\mathbb{R}^n$, then $\bar{y}_k$ is an $\hat{\varepsilon}_k$-strong solution of VIP(F, X), or equivalently,*

$$\max_{x \in X} \langle F(\bar{y}_k), \bar{y}_k - x \rangle \leq \tilde{\varepsilon}_k,$$

where

$$\hat{\varepsilon}_k := \frac{d_0 \mathcal{D}_X \sqrt{8 \bar{\theta}_k L \mathcal{N}_F}}{\sqrt{k\sigma}} + \frac{2Ld_0(\mathcal{D}_X + \bar{\theta}_k d_0)}{k\sigma} \tag{59}$$

and $\mathcal{N}_F :=$ $\mathrm{Nonl}(F; \mathbb{R}^n)$.

Proof. Both statements follow immediately from Lemma 5.7 and Theorems 5.2(b) and 5.5(a). ∎

The following result, which is an immediate consequence of Theorem 5.8, presents iteration-complexity bounds for Korpelevich's extragradient algorithm to obtain ε-weak and -strong solutions of $VIP(F, X)$. We again ignore the dependence of these bounds on the parameter σ and other universal constants and express them only in terms of L, d_0, and the tolerance ε.

COROLLARY 5.9. *Assume that conditions* (B.1)–(B.3) *hold and that the set X has finite diameter $\mathcal{D}_X$. Consider the sequence $\{y_k\}$ generated by Korpelevich's extragradient algorithm applied to VIP(F, X) and the sequence $\{\bar{y}_k\}$ determined according to* (42)*. Then, for every $\varepsilon > 0$, the following statements hold:*

(a) *There exists an index*

$$k_0 = \mathcal{O}\left(\frac{L\mathcal{D}_X d_0}{\varepsilon}\right) \tag{60}$$

such that, for any $k \geq k_0$*, the point* $\bar{y}_k$ *is an* ε*-weak solution of* $VIP(F, X)$*.*

(b) *If* $\Omega = \mathbb{R}^n$ *and* $F : \mathbb{R}^n \to \mathbb{R}^n$ *is monotone and* L*-Lipschitz continuous on* $\mathbb{R}^n$*, then there exists an index*

$$k_0' = \mathcal{O}\left(\frac{L\mathcal{D}_X d_0}{\varepsilon} + \frac{L\mathcal{N}_F \mathcal{D}_X^2 d_0^2}{\varepsilon^2}\right),$$

where $\mathcal{N}_F := \mathrm{Nonl}(F; \mathbb{R}^n)$*, such that, for any* $k \geq k_0'$*, the point* $\bar{y}_k$ *is an* ε*-strong solution of* $VIP(F, X)$*.*

It is interesting to compare the iteration-complexity bound obtained in Corollary 5.9(a) for finding an ε-weak solution of $VIP(F, X)$ with the corresponding ones obtained in Nemirovski [15] and Nesterov [18]. Indeed, their analyses both yield an $\mathcal{O}(D_X^2 L/\varepsilon)$ iteration-complexity bound to obtain an ε-weak solution of $VIP(F, X)$. Hence, in contrast to their bounds, our bound $\mathcal{O}(d_0 D_X L/\varepsilon)$ is proportional to d_0 and formally shows for the first time that Korpelevich's extragradient method benefits from warm-start.

6. Tseng's modified forward-backward splitting method.

In this section, we analyze a special case of Tseng's MF-BS method [27] for solving the inclusion problem

$$0 \in T(x), \qquad T = F + B, \tag{61}$$

for the particular case where the following conditions hold:

(C.1) $F : \mathbb{R}^n \to \mathbb{R}^n$ is monotone and L-Lipschitz continuous.

(C.2) $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone.

(C.3) The solution set of $(F + B)^{-1}(0) \neq \emptyset$.

For this problem, an iteration of the general Tseng's MF-BS method is as follows:

$$y_k = (I + \lambda_k B)^{-1}(I - \lambda_k F)(x_{k-1}), \qquad x_k = P_Y[y_k - \lambda_k(F(y_k) - F(x_{k-1}))],$$

where Y is a closed convex set such that $(F + B)^{-1}(0) \cap Y \neq \emptyset$ and $\lambda_k > 0$ is such that

$$\lambda_k \|F(y_k) - F(x_{k-1})\| \leq \sigma \|y_k - x_{k-1}\| \tag{62}$$

with $\sigma \in (0, 1)$. Note that since we are assuming that F is L-Lipschitz continuous, the above inequality is satisfied for $\lambda_k = \sigma/L$. In the following, we will study the iteration complexity of the special case of Tseng's method, where $Y = \mathbb{R}^n$ and $\lambda_k = \sigma/L$ for every k.

We now formally state the special case of the MF-BS method studied in this section.

Tseng's MF-BS method:

0) Let $x_0 \in \mathbb{R}^n$ and $0 < \sigma < 1$ be given and set $\lambda = \sigma/L$ and $k = 1$.

1) Compute

$$y_k = (I + \lambda B)^{-1}(I - \lambda F)(x_{k-1}), \qquad x_k = y_k - \lambda(F(y_k) - F(x_{k-1})). \tag{63}$$

2) Set $k \leftarrow k + 1$ and go to step 1.

end

The following result was first observed in [22] for the case $\mathrm{Dom}(B) = \mathbb{R}^n$. Although the proof of its extension for an operator B with arbitrary domain is essentially the same, we will include it here for the sake of completeness.

PROPOSITION 6.1. *Let $\{y_k\}$ and $\{x_k\}$ be the sequences generated by Tseng's MF-BS method. For each k, define*

$$q_k = \frac{1}{\lambda}(x_{k-1} - y_k) - F(x_{k-1}), \tag{64}$$

$$v_k = q_k + F(y_k). \tag{65}$$

Then,

(a) $x_k = x_{k-1} - \lambda v_k$;

(b) $q_k \in B(y_k)$ *and* $v_k \in T(y_k) = (F + B)(y_k)$;

(c) $\|\lambda v_k + y_k - x_{k-1}\|^2 \le \sigma^2 \|y_k - x_{k-1}\|^2$.

As a consequence of the above statements, it follows that the special case of Tseng's MF-BS method described above is a special case of the HPE method.

Proof. Statement (a) follows directly from (64), (65), and the second equation in (63). The first inclusion in (b) follows from (64) and the first equation in (63), while the second inclusion follows from the first one and (65). To prove statement (c), first use (64), (65) to obtain

$$\lambda v_k + y_k - x_{k-1} = \lambda(F(y_k) - F(x_{k-1})), \tag{66}$$

which, together with the assumption of F being L-Lipschitz continuous and the definition of λ, yields the desired result. ∎

Note that, in view of (66), criterion (62) (with $\lambda_k = \lambda$) is equivalent to statement (c) of the above theorem. Also, as a consequence of Proposition 6.1, Theorem 4.4, Lemma 4.5, and Theorem 4.7, we have the following iteration-complexity result about Tseng's MF-BS method.

THEOREM 6.2. *Consider the sequences $\{y_k\}$ and $\{x_k\}$ generated by Tseng's MF-BS method, and the sequences $\{v_k\}$, $\{\bar{y}_k\}$, $\{\bar{v}_k\}$, and $\{\bar{\varepsilon}_k\}$ defined as in* (64), (65), (42), *and* (43) *with $\varepsilon_i \equiv 0$. Then, the following hold:*

(a) *For every $\rho > 0$, there exists an index*

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

such that $v_i \in (F + B)(y_i)$ and $\|v_i\| \le \rho$.

(b) *For every $\rho, \varepsilon > 0$, there exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{L d_0^2}{\varepsilon}, \frac{L d_0}{\rho}\right]\right)$$

such that $\bar{v}_k \in (F + B)^{\bar{\varepsilon}_k}(\bar{y}_k)$, $\|\bar{v}_k\| \le \rho$, and $\bar{\varepsilon}_k \le \varepsilon$ for any $k \ge k_0$.

Since $VI(F, X)$ with $X \subset \mathbb{R}^n$ closed and convex is equivalent to the inclusion problem (61) with $B = N_X$, Tseng's MF-BS method with $B = N_X$ can be used to solve $VIP(F, X)$. In this case, the iteration formula (63) reduces to

$$y_k = P_X[x_{k-1} - \lambda F(x_{k-1})], \qquad x_k = y_k - \lambda(F(y_k) - F(x_{k-1})). \tag{67}$$

Note that, in contrast to Korpelevich's method, the above algorithm for solving *VIP* (F, X) requires *just one* projection per iteration. Moreover, in addition to the trivial specialization of Theorem 6.2 to the context of monotone VIs, all the iteration-complexity results of section 5 hold for Tseng's MF-BS method with $B = N_X$.

7. Hybrid proximal extragradient for smooth monotone equations. In this section we consider the problem of solving

$$F(x) = 0, \tag{68}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the following conditions:

(D.1) F is monotone;

(D.2) F is differentiable and its Jacobian F' is L_1-Lipschitz continuous.

Note that this problem is a special case of $VIP(F, X)$, where $X = \mathbb{R}^n$.

Newton's method, applied to problem (68), under assumption (D.2), has excellent local convergence properties, provided the starting point is close to a solution x^* *and* F' is nonsingular at x^*. Global well-definedness of Newton's method requires F' to be nonsingular everywhere on $\mathbb{R}^n$. However, this additional assumption does not guarantee the global convergence of Newton's method. A globally convergent extrapolation Newton-type method for solving (68) was proposed in [8, 9] (see also [10]).

Solodov and Svaiter [22] proposed the use of Newton's method for approximately solving the proximal subproblem at each iteration of the HPE method for problem (68). In this section, we will analyze the complexity of a variant of this method. Specifically, we will consider the following special case of the HPE method.

Newton proximal extragradient (NPE) method:

0) Let $x_0 \in \mathbb{R}^n$ and $0 < \sigma_\ell < \sigma_u < 1$ be given and set $k = 1$.
1) If $F(x_{k_1}) = 0$, STOP. Otherwise
2) compute $\lambda_k \in \mathbb{R}$ and $s_k \in \mathbb{R}^n$ satisfying

$$(\lambda_k F'(x_{k-1}) + I)s_k = -\lambda_k F(x_{k-1}), \tag{69}$$

$$\frac{2}{L_1}\,\sigma_\ell \le \lambda_k \|s_k\| \le \frac{2}{L_1}\,\sigma_u. \tag{70}$$

3) Define $y_k = x_{k-1} + s_k$ and $x_k = x_{k-1} - \lambda_k F(y_k)$, set $k \leftarrow k + 1$, and go to step 1.

end

We will assume that $F(x_k) \neq 0$ for $k = 0, 1, \dots$. The complexity analysis will not be affected by this assumption, because any assertion about the results after k iterations will be valid, adding the alternative "or the algorithm finds a zero."

For practical computations, given λ_k, the step s_k shall be computed solving the linear equation

$$(F'(x_k) + \lambda_k^{-1} I)s = -F(x_k).$$

Note that the direction s_k in step 1 of the NPE method is the Newton direction with respect to the proximal point equation $\lambda F(x) + \lambda_k(x - x_{k-1}) = 0$ at the point x_{k-1}. Define, for each k,

$$\sigma_k := \frac{L_1}{2}\lambda_k \|s_k\|. \tag{71}$$

We need the following well-known result about differentiable maps with Lipschitz continuous Jacobian.

LEMMA 7.1. *Suppose that $G : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable and its Jacobian G' is L-Lipschitz continuous. Then, for every $x, s \in \mathbb{R}^n$, we have*

$$\|G(x + s) - G(x) - G'(x)s\| \le \frac{L}{2}\|s\|^2.$$

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

LEMMA 7.2. *For each k, $\sigma_\ell \le \sigma_k \le \sigma_u$ and*

$$\|\lambda_k F(y_k) + y_k - x_{k-1}\| \le \sigma_k \|y_k - x_{k-1}\|. \tag{72}$$

As a consequence, the NPE method is a special case of the HPE method stated in section 4 with $\varepsilon_k = 0$, $v_k = F(y_k)$, and $\sigma = \sigma_u$.

Proof. The bounds on σ_k follow directly from (70). Define for each k

$$G_k(x) := \lambda_k F(x) + x - x_{k-1}.$$

Then, G_k' is $\lambda_k L_1$-Lipschitz continuous and in view of (69)

$$G_k'(x_{k-1}) s_k + G_k(x_{k-1}) = 0.$$

Therefore, using Lemma 7.1 and (71) we have

$$\|G_k(y_k)\| = \|G_k(y_k) - [G(x_{k-1}) + G_k'(x_{k-1}) s_k]\| \le \frac{\lambda_k L_1}{2} \|s_k\|^2 = \sigma_k \|s_k\|,$$

which, due to the fact that $s_k = y_k - x_{k-1}$, proves (72).

Now we shall prove that the NPE method is a special case of the HPE method. Using inequality (72) and the facts that $\sigma_k \le \sigma_u$, $\varepsilon_k = 0$, $v_k = F(y_k)$, and $F = F^0 = F^{\varepsilon_k}$, we conclude that condition (17) is satisfied with $T = F$. To end the proof, note that $x_k = x_{k-1} - \lambda_k v_k$. ∎

We have seen in Theorems 4.4 and 4.7 that the performance of the HPE method depends on the sums $\sum \lambda_i$ and $\sum \lambda_i^2$. We now give lower bounds for these quantities in the context of the NPE method.

LEMMA 7.3. *The sequence $\{\lambda_k\}$ of the NPE method satisfies*

$$\frac{4\beta}{L_1^2} \sum_{i=1}^{\infty} \frac{1}{\lambda_i^2} \le d_0^2\,, \tag{73}$$

where d_0 is the distance of x_0 to $F^{-1}(0)$ and β is the minimum value of the function $(1-t^2)t^2$ over the interval $[\sigma_\ell, \sigma_u]$, i.e.,

$$\beta := \min\{(1-\sigma_\ell^2)\sigma_\ell^2, (1-\sigma_u^2)\sigma_u^2\}. \tag{74}$$

As a consequence, for any k,

$$\sum_{i=1}^{k} \lambda_i \ge \frac{2\sqrt{\beta}}{L_1 d_0} k^{3/2}, \qquad \sum_{i=1}^{k} \lambda_i^2 \ge \frac{4\beta}{L_1^2 d_0^2} k^2\,. \tag{75}$$

Proof. Using the definition of y_i and (71) we have

$$\|y_i - x_{i-1}\|^2 = \frac{1}{\lambda_i^2} (\lambda_i^2 \|s_i\|)^2 = \frac{1}{\lambda_i^2} \left(\frac{2}{L_1}\, \sigma_i\right)^2.$$

Since the NPE method is a particular case of the HPE method (Lemma 7.2), we can combine the above equation with the first inequality in (23) to conclude that

$$d_0^2 \ge \frac{4}{L_1^2} \sum_{i=1}^{\infty} (1-\sigma_i^2)\sigma_i^2 \frac{1}{\lambda_i^2}.$$

To end the proof of (73), use the inclusion $\sigma_i \in [\sigma_\ell, \sigma_u]$ and the definition of β.

The inequalities in (75) now follow by noting that the minimum value of the functions $\sum_{i=1}^k t_i$ and $\sum_{i=1}^k t_i^2$ subject to the condition that $\sum_{i=1}^k t_i^{-2} \le C$ and $t_i > 0$ for $i = 1, \dots, k$ are $k^{3/2}/\sqrt{C}$ and k^2/C, respectively. ∎

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

PROPOSITION 7.4. *Consider the sequence $\{y_k\}$ generated by the NPE method. Then, for every $k \ge 1$, there exists an index $i \le k$ such that*

$$\|F(y_i)\| \le \sqrt{\frac{1+\sigma_u}{1-\sigma_u}} \frac{L_1 d_0^2}{2k\sqrt{\beta}},$$

where β is the constant defined in Lemma 7.3. As a consequence, for any scalar $\varepsilon > 0$, there exists an index

$$i = \mathcal{O}\left(\frac{L_1 d_0^2}{\varepsilon}\right)$$

such that the iterate y_i satisfies $\|F(y_i)\| \le \varepsilon$.

Proof. By Lemma 7.2, the NPE method is a special case of the HPE method, where $\sigma = \sigma_u$ and the sequences $\{v_k\}$ and $\{\varepsilon_k\}$ are given by $v_k = F(y_k)$ and $\varepsilon_k = 0$ for every k. Hence, it follows from Theorem 4.4(b) and Lemma 7.3 that, for every $k \in \mathbb{N}$, there exists $i \le k$ such that

$$\|F(y_i)\| = \|v_i\| \le \sqrt{\frac{(1+\sigma_u)}{(1-\sigma_u)}} \frac{d_0}{\left(\sum_{i=1}^k \lambda_i^2\right)^{1/2}} \le \sqrt{\frac{1+\sigma_u}{1-\sigma_u}} \frac{L_1 d_0^2}{2k\sqrt{\beta}}.$$

The last part of the theorem follows immediately from the first part. ∎

We now state a global convergence rate result for the NPE method about the ergodic pair $(\bar{y}_k, \bar{v}_k)$ defined as in (27) and (28).

PROPOSITION 7.5. *Let $\{\lambda_k\}$, $\{y_k\}$, and $\{x_k\}$ be the sequences generated by the NPE method and consider the sequences $\{v_k\}$ and $\{\varepsilon_k\}$ defined as $v_k = F(y_k)$ and $\varepsilon_k = 0$ for every $k \ge 1$. Then, the sequences $\{\bar{y}_k\}$, $\{\bar{v}_k\}$, and $\{\bar{\varepsilon}_k\}$ defined according to* (27) *and* (28) *satisfy the following conditions:*

(a) *For every $k \ge 1$, $\bar{\varepsilon}_k \ge 0$ and $\bar{v}_k \in F^{\bar{\varepsilon}_k}(\bar{y}_k)$, or equivalently, $\langle x - \bar{y}_k, F(x) - \bar{v}_k\rangle \ge -\bar{\varepsilon}_k$ for all $x \in \mathbb{R}^n$.*

(b) *For every $k \ge 1$, we have*

$$\|\bar{v}_k\| \le \frac{L_1 d_0^2}{k^{3/2}\sqrt{\beta}}, \qquad \bar{\varepsilon}_k \le \frac{\theta L_1 d_0^3}{k^{3/2}\sqrt{\beta}}, \tag{76}$$

where

$$\theta := 1 + \frac{\sigma_u}{\sqrt{1-\sigma_u^2}}.$$

Proof. This result follows as an immediate consequence of Theorem 4.7 with $\sigma = \sigma_u$ and Lemmas 7.2 and 7.3. ∎

Our goal in the remaining part of this section will be to discuss the work involved in finding the pair (s_k, λ_k) in step 1 of the NPE method. First note that, due to the monotonicity of F, $F'(x)$ is positive semidefinite and for every $x \in \mathbb{R}^n$ and $\lambda > 0$, the system

$$(\lambda F'(x) + I)s = -\lambda F(x) \tag{77}$$

has a unique solution $s \in \mathbb{R}^n$, which we denote by $s(\lambda; x)$. Clearly,

$$s(\lambda; x) = -(F'(x) + \lambda^{-1} I)F(x) .$$

Observe that, given $x \in \mathbb{R}^n$, step 1 of the NPE method can be rephrased as the problem of computing $\lambda > 0$ such that

$$2\sigma_\ell / L_1 \le \lambda \| s(\lambda; x) \| \le 2\sigma_u / L_1. \tag{78}$$

Computation of the scalar λ will done by performing a logarithmic-type bisection scheme on λ, which will be discussed shortly. Each iteration of this bisection scheme requires computation of $s(\lambda; x)$ and evaluation of the quantity $\lambda \| s(\lambda; x) \|$ to check whether λ satisfies the aforementioned condition. Without any use of previous information, computation of $s(\lambda; x)$ requires $\mathcal{O}(n^3)$ arithmetic operations for each λ. Hence, if the aforementioned bisection scheme requires k_x evaluations of $s(\cdot; x)$, then the total cost of the bisection scheme will be $\mathcal{O}(n^3 k_x)$ arithmetic operations. However, it is possible to do better than that by using the following steps. First, compute a Hessenberg factorization of $F'(x)$, i.e., factor $F'(x)$ as $F'(x) = QHQ^T$, where Q is an orthonormal matrix and H is an upper Hessenberg matrix (namely, $H_{ij} = 0$ for every $j < i$). Then, $s = s(\lambda; x)$ can be computed by solving the system $(H + \lambda^{-1} I)\tilde{s} = Q^T F(x)$ for $\tilde{s}$ and letting $s = Q^T \tilde{s}$. Clearly, the first step of the modified bisection scheme can be performed in $\mathcal{O}(n^3)$ arithmetic operations and subsequent steps in $\mathcal{O}(n^2)$ arithmetic operations. Hence, the modified bisection scheme can be carried out in $\mathcal{O}(n^3 + n^2 k_x)$ arithmetic operations.

Given a fixed $x \in \mathbb{R}^n$, our goal now will be to describe the aforementioned bisection scheme to compute λ and to estimate its number of iterations k_x.

PROPOSITION 7.6. *For any $x \in \mathbb{R}^n$, the mapping $\lambda \to s(\lambda; x)$ is continuous on $(0, \infty)$ and*

$$\frac{\lambda \| F(x) \|}{\lambda \| F'(x) \| + 1} \le \| s(\lambda; x) \| \le \lambda \| F(x) \| \qquad \forall \lambda > 0, \tag{79}$$

where $\| F'(x) \|$ is the operator norm of $F'(x)$.

Proof. Continuity $s(\lambda)$ for $\lambda \in (0, \infty)$ follows from the fact that $F'(x)$ is positive semidefinite. To simplify the proof, let $s = s(\lambda; x)$. Using (77), the triangle inequality, and the definition of operator norm, we conclude that

$$\lambda \| F(x) \| \le \| \lambda F'(x) s \| + \| s \| \le (\| \lambda F'(x) \| + 1) \| s \|,$$

and hence that the first inequality in (79) holds. To prove the second inequality, we multiply both sides of (77) by s and use the Cauchy–Schwarz inequality and the positive definiteness of $F'(x)$ to obtain

$$\| s \|^2 \le \lambda \langle s, F'(x) s \rangle + \| s \|^2 = -\lambda \langle s, F(x) \rangle \le \lambda \| s \| \| F(x) \|,$$

and hence the second inequality in (79) holds. $\square$

The following result establishes the existence of scalars λ satisfying (78) under the mild assumption that $F(x) \neq 0$. It also provides an explicit closed interval in $\mathbb{R}_{++}$ which contains all the solutions of (78). This interval will be used as an initial bracketing interval on a log-type bisection scheme (see Routine Step 1 below) for computing a solution of (78).

PROPOSITION 7.7. *Suppose that $x \in \mathbb{R}^n$ is such that $F(x) \neq 0$ and let $\mathcal{I}_x$ be the set of all $\lambda > 0$ satisfying* (78). *Then, $\mathcal{I}_x$ is nonempty and $\mathcal{I}_x \subset [a_x, b_x]$, where*

$$a_x := \sqrt{\frac{2\sigma_\ell}{L_1\|F(x)\|}}, \qquad b_x := \frac{2\sigma_u}{L_1\|F(x)\|}\|F'(x)\| + \sqrt{\frac{2\sigma_u}{L_1\|F(x)\|}}. \tag{80}$$

Proof. Suppose that $x \in \mathbb{R}^n$ is such that $F(x) \neq 0$. By (79) and the fact that $F(x) \neq 0$, we easily see that $\lim_{\lambda\to 0} \lambda\|s(\lambda;x)\| = 0$ and $\lim_{\lambda\to\infty} \lambda\|s(\lambda;x)\| = \infty$. This conclusion together with the continuity of the function $\lambda \to \lambda\|s(\lambda;x)\|$ and the assumption that $\sigma_\ell < \sigma_u$ implies that there exists λ satisfying (78), i.e., $\mathcal{I}_x \neq \emptyset$.

Assume now that $\lambda \in \mathcal{I}_x$, and hence that (78) holds. The latter relation together with (79) implies that

$$\frac{2\sigma_\ell}{L_1} \le \lambda^2\|F(x)\|, \qquad \frac{\lambda^2\|F(x)\|}{\lambda\|F'(x)\|+1} \le \frac{2\sigma_u}{L_1}.$$

The first inequality clearly implies $a_x \le \lambda$. Multiplying the second inequality by $\lambda\|F'(x)\|+1$, we obtain a quadratic inequality in λ which, together with the inequality $(\alpha_1^2+\alpha_2^2)^{1/2} \le \alpha_1+\alpha_2$ for $\alpha_1, \alpha_2 > 0$, implies that $\lambda \le b_x$. We have thus proved that $\mathcal{I}_x \subseteq [a_x, b_x]$. ∎

LEMMA 7.8. *For every $x \in \mathbb{R}^n$ such that $F(x) \neq 0$ and $0 < \lambda < \tilde\lambda$, we have*

$$\|s(\lambda;x)\| < \|s(\tilde\lambda;x)\| \le \frac{\tilde\lambda}{\lambda}\|s(\lambda;x)\|. \tag{81}$$

Proof. To simplify the notation, let $s = s(\lambda;x)$ and $\tilde s = s(\tilde\lambda;x)$. Since by definition $s(\lambda;x)$ is the unique solution of (77), we easily see that

$$F'(x)(\tilde s - s) = \lambda^{-1}s - \tilde\lambda^{-1}\tilde s. \tag{82}$$

Noting that $s \neq 0$ due to Proposition 7.6 and the assumption that $F(x) \neq 0$, it follows from (82) that $s \neq \tilde s$. Moreover, (82) and the monotonicity of $F'(x)$ imply that

$$\langle \lambda^{-1}s - \tilde\lambda^{-1}\tilde s, \tilde s - s\rangle = \langle F'(x)(\tilde s - s), \tilde s - s\rangle \ge 0.$$

Multiplying this inequality by $\lambda\tilde\lambda$, we obtain after some straightforward algebraic manipulations that

$$(\tilde\lambda - \lambda)\langle s, \tilde s - s\rangle \ge \lambda\|\tilde s - s\|^2.$$

Since $\tilde\lambda - \lambda > 0$ by assumption and $s \neq \tilde s$, the above relation implies that $\langle s, \tilde s - s\rangle > 0$, and hence that $\|\tilde s\|^2 > \|s\|^2$. This proves the first inequality in (81). Also, the above inequality, the Cauchy–Schwarz inequality, and the fact that $\|\tilde s - s\| > 0$ imply that

$$(\tilde\lambda - \lambda)\|s\| \ge \lambda\|\tilde s - s\|.$$

Adding $\lambda\|s\|$ to both sides of this inequality and using the triangle inequality for norms, we obtain the second inequality in (81). ∎

COROLLARY 7.9. *For every $x \in \mathbb{R}^n$ such that $F(x) \neq 0$, the set $\mathcal{I}_x$ is a closed interval $[\lambda_{x,\ell}, \lambda_{x,u}]$ such that*

$$\frac{\lambda_{x,u}}{\lambda_{x,\ell}} \ge \sqrt{\frac{\sigma_u}{\sigma_\ell}}.$$

Proof. Note that $\lambda \to \lambda\|s(\lambda,x)\|$ is a strictly increasing continuous function over $\mathbb{R}_{++}$ due to the first inequality in (81). Also, by Proposition 7.6, we have $\lim_{\lambda\to 0} \lambda\|s(\lambda;x)\| = 0$ and $\lim_{\lambda\to\infty} \lambda\|s(\lambda;x)\| = \infty$. The above two observations clearly imply that $\mathcal{I}_x$ is a closed interval, say $\mathcal{I}_x = [\lambda_{x,\ell}, \lambda_{x,u}]$, and

$$\lambda_{x,\ell}\|s(\lambda_{x,\ell};x)\| = \frac{2\sigma_\ell}{L_1}, \qquad \lambda_{x,u}\|s(\lambda_{x,u};x)\| = \frac{2\sigma_u}{L_1}.$$

Moreover, the second inequality in (81) implies that

$$\lambda_{x,u}\|s(\lambda_{x,u};x)\| \leq \left(\frac{\lambda_{x,u}}{\lambda_{x,\ell}}\right)^2 \lambda_{x,\ell}\|s(\lambda_{x,\ell};x)\|.$$

Combining the above relations, we obtain the desired conclusion. ∎

With the aid of the above results, we can now show how the pair (λ_k, s_k) can be computed at step 1 of the NPE method. Using x_{k-1} as input on the following routine, it then outputs the pair (λ_k, s_k).

Routine Step 1:

Input: $x \in \mathbb{R}^n$ such that $F(x) \neq 0$ and $0 < \sigma_\ell < \sigma_u < 1$.

0) Compute a_x and b_x as in (80) and set $a \leftarrow a_x$ and $b \leftarrow b_x$.
1) Set $\lambda = \sqrt{ab}$ and compute $s = -(F'(x) + \lambda^{-1}I)F(x)$.
2) If $2\sigma_\ell/L_1 \leq \lambda\|s\| \leq 2\sigma_u/L_1$, then output the pair (λ, s) and STOP.
3) If $\lambda\|s\| > 2\sigma_u/L_1$, then set $b \leftarrow \lambda$; otherwise, set $a \leftarrow \gamma$.
4) Go to step 1.

end

Let $\varepsilon > 0$ be given. We have seen in Proposition 7.4 that the NPE method finds an iterate y_i satisfying the termination criterion $\|F(x)\| \leq \varepsilon$ in at most $\mathcal{O}(L_1 d_0^2/\varepsilon)$. Clearly, while computing such an iterate we may assume that $\|F(x_{k-1})\| > \epsilon$ every time step 1 of the NPE method is executed since otherwise x_{k-1} itself would satisfy the termination criterion $\|F(x)\| \leq \varepsilon$. Moreover, (34) implies that $\|x_{k-1} - x_0\| \leq 2d_0$. Hence, it is sufficient to analyze the complexity of Routine Step 1 under the assumption that its input $x \in \mathbb{R}^n$ satisfies $\|F(x)\| > \varepsilon$ and $\|x - x_0\| \leq 2d_0$. In the following result, we also assume that the constants σ_ℓ and σ_u are such that $(\log \sigma_u/\sigma_\ell)^{-1} = \mathcal{O}(1)$, which allows us to express the complexities only in terms of the quantities L_1, ε, d_0, and $\|F'(x_0)\|$.

PROPOSITION 7.10. *Let $\varepsilon > 0$ be given. Then, for every $x \in \mathbb{R}^n$ such that $\|F(x)\| > \varepsilon$ and $\|x - x_0\| \leq 2d_0$, the number of iterations and arithmetic operations performed by Routine Step* 1 *are bounded respectively by*

$$\mathcal{O}\left(\log\log\left(\frac{\|F'(x_0)\|}{\sqrt{L_1\varepsilon}} + \frac{\sqrt{L_1}}{\sqrt{\varepsilon}}d_0\right)\right). \tag{83}$$

and

$$\mathcal{O}\left(n^3 + n^2\log\log\left(\frac{\|F'(x_0)\|}{\sqrt{L_1\varepsilon}} + \frac{\sqrt{L_1}}{\sqrt{\varepsilon}}d_0\right)\right).$$

Proof. Since $\log\lambda = (\log a + \log b)/2$, it follows that, after k inner iterations of the above implementation of step 1 of the NPE method, the scalars a and b at step 3 satisfy

$$\log\frac{b}{a} = \frac{1}{2^k}\log\frac{b_x}{a_x}. \tag{84}$$

Assume now that the method does not stop at the kth iteration. Then, the values of a and b in step 3 of this iteration satisfy $\mathcal{I}_x = [\lambda_{x,\ell}, \lambda_{x,u}] \subseteq [a, b]$, and hence $b/a \geq \sqrt{\sigma_u/\sigma_\ell}$, due to Corollary 7.9. This together with (84) implies that

$$\frac{1}{2^k} \log \frac{b_x}{a_x} \geq \frac{1}{2} \log \frac{\sigma_u}{\sigma_\ell},$$

and hence that

$$k \leq 1 + \log \left(\frac{\log(b_x/a_x)}{\log(\sigma_u/\sigma_\ell)} \right).$$

Since (80), assumption (D.2), and the assumptions that $\|F(x)\| > \varepsilon$ and $\|x - x_0\| \leq 2d_0$ imply that

$$\frac{b_x}{a_x} = \sqrt{\frac{\sigma_u}{\sigma_\ell}} \left(1 + \sqrt{\frac{2\sigma_u}{L_1 \|F(x)\|}} \; \|F'(x)\| \right) \leq \sqrt{\frac{\sigma_u}{\sigma_\ell}} \left(1 + \sqrt{\frac{2}{L_1 \varepsilon}} \; (\|F'(x_0)\| + 2L_1 d_0) \right),$$

we conclude that the number of iterations performed by Routine Step 1 is bounded by (83). The bound on the number of arithmetic operations is due to the discussion following Proposition 7.5. ◻

8. Concluding remarks. We provide in this section some concluding remarks.

We start by summarizing the contribution of this paper as compared to [22], where the HPE method was proposed. First, iteration-complexity analysis of the HPE method is given here for the first time. Second, in contrast to our analysis, [22] does not study the properties of the ergodic mean for the HPE method. Third, the result that Korpelevich's method is a special case of the HPE method is new. Fourth, in contrast to rule (70), [22] chooses λ_k proportional to $\|F(x_k)\|^{-1/2}$ in the NPE method, and does not study the iteration complexity of the resulting method.

It is important to emphasize that all complexity bounds in section 5, with the exception of the ones obtained in subsection 5.2, do not require boundedness of the feasible set X. The bounds on these results are expressed in terms of the distance d_0 of the initial point to the solution set, instead of the diameter of X. In the usual situation where d_0 is unknown, these complexity bounds have only theoretical value in the sense that they cannot be used to terminate the method. Instead, our analysis provides practical stopping rules to obtain (ρ, ε)-weak (resp., -strong) solutions by monitoring the size of the pair (v_k, ε_k) or the ergodic pair $(\bar{v}_k, \bar{\varepsilon}_k)$ (or $(\hat{v}_k, \bar{\varepsilon}_k)$), both of which can be easily computed. It should be observed that, to obtain an ε-weak/-strong solution, it is necessary to have a (finite) bound on the diameter of the feasible set. But as we observed after Definition 4, computation of (ρ, ε)-weak/-strong solutions is a natural goal in the context of nonlinear complementarity and most likely VI problems as well.

The iteration-complexity analysis developed for the HPE method in this paper was used to obtain iteration-complexity results for two specific algorithms, namely, the Korpelevich and NPE methods. It would be interesting to derive iteration-complexity results for other algorithms by viewing them as special cases of the HPE method.

Appendix A. Relationship between error measures. In this appendix, we review other notions of error measures and discuss their relationship with the error measures introduced in the main presentation of the paper.

It is well known that x is a solution of $VIP(F, X)$ if and only if the quantity

$$\mathrm{r}_c(x; F) := c \left[x - P_X \left(x - \frac{1}{c} F(x) \right) \right] \tag{85}$$

vanishes, where $c > 0$ is a scaling factor. The norm of $\mathrm{r}_c(x;F)$ is commonly used to measure the quality of x as an approximate solution of $VIP(F,X)$. Another measure of the quality of x as an approximate solution of $VIP(F,X)$ is the regularized gap function (see [5]) defined as

$$\theta_c(x;F) := \sup_{y\in X} \langle F(x), x-y\rangle - \frac{c}{2}\|y-x\|^2. \tag{86}$$

It is easy to see that θ_c is a nonnegative function whose set of zeros coincides with the solution set of $VIP(F,X)$ and that the unique optimal solution of (86) is

$$y = P_X\left(x - \frac{1}{c}F(x)\right).$$

Hence, in view of (85) we have

$$\theta_c(x,F) = \frac{1}{c}\langle F(x), \mathrm{r}_c(x;F)\rangle - \frac{1}{2c}\|\mathrm{r}_c(x;F)\|^2. \tag{87}$$

The following result provides a connection between the regularized gap function and the notion of (ρ,ε)-strong solutions.

THEOREM A.1. *Let $x \in X$ be given. If (r,ε) is a strong residual of x for $VIP(F,X)$, then*

$$\theta_c(x;F) \le \frac{1}{2c}\|r\|^2 + \varepsilon. \tag{88}$$

Moreover, there exists a unique strong residual (r,ε) of x for which equality holds in (88)*, namely,*

$$r = \mathrm{r}_c(x;F), \qquad \varepsilon = \frac{1}{c}\left(\langle F(x), \mathrm{r}_c(x;F)\rangle - \|\mathrm{r}_c(x;F)\|^2\right). \tag{89}$$

Proof. Fix $x \in X$. To prove the theorem, it suffices to show, in view of Proposition 3.2(b) and (87), that the solution of the problem

$$\begin{array}{ll} \min & \dfrac{1}{2c}\|r\|^2 + \varepsilon \\ \text{s.t.} & r \in (F + N_X^{\varepsilon})(x) \end{array} \tag{90}$$

is unique and is given by (89). For any r, the smaller ε for which $r \in F(x) + N^{\varepsilon}(x)$ is given by

$$\sup_{y\in X} \langle r - F(x), y - x\rangle.$$

Hence, problem (90) is equivalent to

$$\min_{r} \sup_{y\in Y} \frac{1}{2c}\|r\|^2 + \langle r - F(x), y-x\rangle.$$

Since the above objective function is strongly convex, we conclude that the above problem, and as a by-product problem (90), has a unique optimal solution. Note that (88) follows from the fact that

$$\theta_c(x;F) = \sup_{y\in Y}\inf_{r} \frac{1}{2c}\|r\|^2 + \langle r - F(x), y-x\rangle \le \inf_{r}\sup_{y\in Y} \frac{1}{2c}\|r\|^2 + \langle r - F(x), y-x\rangle$$

and from the equivalence of the latter problem with (90). Clearly, in view of (89) and (87), it follows that the objective function of (90) evaluated at the pair (r, ε) given by (89) is equal to $\theta_c(x; F)$. To conclude the proof, it suffices to show that this pair also satisfies $r \in (F + N_X^\varepsilon)(x)$. Indeed, using (85), we obtain

$$P_X\left(x - \frac{1}{c}F(x)\right) = x - \frac{1}{c}\mathrm{r}_c(x; F).$$

Subtracting $F(x)$ from both sides of (85) and using (39) and the above equation, we conclude that

$$\mathrm{r}_c(x; F) - F(x) \in N_X\left(x - \frac{1}{c}\mathrm{r}_c(x; F)\right).$$

Using the above relation, Proposition 2.2(c) with $f = \delta_X$, $v = \mathrm{r}_c(x; F) - F(x)$, and $y = x - c^{-1}\mathrm{r}_c(x; F)$, and relation (15), we then conclude that $\mathrm{r}_c(x; F) - F(x) \in N_X^\varepsilon(x)$ with

$$\varepsilon = \left\langle \mathrm{r}_c(x; F) - F(x)\,,\, x - \frac{1}{c}\mathrm{r}_c(x; F) - x\right\rangle = \frac{1}{c}\left(\langle F(x), \mathrm{r}_c(x; F)\rangle - \|\mathrm{r}_c(x; F)\|^2\right). \qquad \square$$

The following result follows as an immediate consequence of the Theorem A.1.

COROLLARY A.2. *Let $x \in X$ be given. If x is a (ρ, ε)-strong solution of VIP(F, X), then for any $c > 0$,*

$$\theta_c(x; F) \le \frac{\rho^2}{2c} + \varepsilon, \qquad \|\mathrm{r}_c(x; F)\| \le \sqrt{\rho^2 + 2c\varepsilon}.$$

We now relate the notion of a weak solution with that of a strong solution.

LEMMA A.3. *If F is L-Lipschitz continuous in X, then for any $x \in X$,*

$$\theta_{2L}(x; F) \le \theta^w(x; F).$$

Proof. Using Definition 3 and the Lipschitz continuity of F, we have

$$\begin{aligned}\theta^w(x; F) &= \sup_{y\in X}\langle F(y), x - y\rangle = \sup_{y\in X}\langle F(x), x - y\rangle + \langle F(y) - F(x), x - y\rangle \\ &\ge \sup_{y\in X}\langle F(x), x - y\rangle - L\|x - y\|^2 = \theta_{2L}(x; F). \qquad \square\end{aligned}$$

THEOREM A.4. *If F is L-Lipschitz continuous in X and $x \in X$ is a (ρ, ε)-weak solution of VIP(F, X), then x is $(\tilde\rho, \varepsilon)$-strong solution, where $\tilde\rho := \rho + 2\sqrt{L\varepsilon}$.*

Proof. By the assumptions of the theorem, there exists $r \in \mathbb{R}^n$ such that $\|r\| \le \rho$ and $\theta^w(x, F_r) \le \varepsilon$, where $F_r(\cdot) = F(\cdot) - r$. Using Lemma A.3, we conclude that $\theta_{2L}(x; F_r) \le \varepsilon$. Moreover, by Theorem A.1, there exists a strong residual (r', ε') of x for VIP$(x; F_r)$ such that

$$\frac{1}{4L}\|r'\|^2 + \varepsilon' = \theta_{2L}(x; F_r) \le \varepsilon.$$

This implies that $\theta^s(x; F - r - r') = \theta^s(x; F_r - r') \le \varepsilon' \le \varepsilon$ and

$$\|r + r'\| \le \|r\| + \|r'\| \le \rho + 2\sqrt{L\varepsilon} = \tilde\rho,$$

and hence that x is $(\tilde\rho, \varepsilon)$-strong solution of VIP$(F, X)$. $\square$

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

REFERENCES

[1] R. E. Bruck, Jr., *On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space*, J. Math. Anal. Appl., 61 (1977), pp. 159–164.

[2] 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.

[3] R. S. Burachik, C. A. Sagastizábal, and B. F. Svaiter, *ϵ-enlargements of maximal monotone operators: Theory and applications*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods (Lausanne, 1997), Appl. Optim. 22, Kluwer Academic Publishers, Dordrecht, The Netherlands 1999, pp. 25–43.

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

[5] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems,* Vol. II, Springer-Verlag, New York, 2003.

[6] E. G. Gol′shteĭn, *A method for solving variational inequalities defined by monotone mappings*, Zh. Vychisl. Mat. Mat. Fiz., 42 (2002), pp. 958–968.

[7] E. G. Golshtein and N. V. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*, Wiley-Intersci. Ser. Discrete Math. Optim., John Wiley & Sons, New York, 1996 (translated from the 1989 Russian original by Tretyakov).

[8] I. V. Konnov, *Combined relaxation methods for the search for equilibrium points and solutions of related problems*, Izv. Vyssh. Uchebn. Zaved. Mat., (1993), pp. 46–53.

[9] I. V. Konnov, *A class of combined iterative methods for solving variational inequalities*, J. Optim. Theory Appl., 94 (1997), pp. 677–693.

[10] I. V. Konnov, *Combined relaxation methods for generalized monotone variational inequalities*, in Generalized Convexity and Related Topics, Lecture Notes in Econom. and Math. Systems 583, Springer-Verlag, Berlin, 2007, pp. 3–31.

[11] G. M. Korpelevič, *An extragradient method for finding saddle points and for other problems*, Èkonom. i Mat. Metody, 12 (1976), pp. 747–756.

[12] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov, *New variants of bundle methods*, Math. Program., 69 (1995), pp. 111–147.

[13] P.-L. Lions, *Une méthode itérative de résolution d'une inéquation variationnelle*, Israel J. Math., 31 (1978), pp. 204–208.

[14] T. L. Magnanti and G. Perakis, *A unifying geometric solution framework and complexity analysis for variational inequalities*, Math. Program., 71 (1995), pp. 327–351.

[15] 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 J. Optim., 15 (2004), pp. 229–251.

[16] A. S. Nemirovskiĭ, *Effective iterative methods for solving equations with monotone operators*, Èkonom. i Mat. Metody, 17 (1981), pp. 344–359.

[17] A. S. Nemirovskiĭ and D. B. Judin, *Cesàro convergence of the gradient method for the approximation of saddle points of convex-concave functions*, Dokl. Akad. Nauk SSSR, 239 (1978), pp. 1056–1059.

[18] Y. Nesterov, *Dual extrapolation and its applications to solving variational inequalities and related problems*, Math. Program., 109 (2007), pp. 319–344.

[19] Y. E. Nesterov, *Smooth minimization of nonsmooth functions*, Math. Program., 103 (2005), pp. 127–152.

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

[21] R. Tyrrell 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, *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.

[25] 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.

[26] P. Tseng, *On linear convergence of iterative methods for the variational inequality problem*, J. Comput. Appl. Math., 60 (1995), pp. 237–252.

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