SIAM J. OPTIM.
Vol. 21, No. 4, pp. 1688–1720
© 2011 Society for Industrial and Applied Mathematics

COMPLEXITY OF VARIANTS OF TSENG'S MODIFIED F-B SPLITTING AND KORPELEVICH'S METHODS FOR HEMIVARIATIONAL INEQUALITIES WITH APPLICATIONS TO SADDLE-POINT AND CONVEX OPTIMIZATION PROBLEMS*

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

Abstract. In this paper, we consider both a variant of Tseng's modified forward-backward splitting method and an extension of Korpelevich's method for solving hemivariational inequalities with Lipschitz continuous operators. By showing that these methods are special cases of the hybrid proximal extragradient method introduced by Solodov and Svaiter, we derive iteration-complexity bounds for them to obtain different types of approximate solutions. In the context of saddle-point problems, we also derive complexity bounds for these methods to obtain another type of an approximate solution, namely, that of an approximate saddle point. Finally, we illustrate the usefulness of the above results by applying them to a large class of linearly constrained convex programming problems, including, for example, cone programming and problems whose objective functions converge to infinity as the boundaries of their effective domains are approached.

Key words. extragradient, hemivariational inequality, saddle point, maximal monotone operator, complexity, Korpelevich's method, forward-backward splitting methods, convex optimization

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

DOI. 10.1137/100801652

1. Introduction.

A broad class of optimization, saddle-point (SP), 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 [18], 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.,

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

Convergence results under the above error condition have been established in [18] and have been used in the convergence analysis of other methods that can be recast in the above framework.

*Received by the editors July 12, 2010; accepted for publication (in revised form) September 20, 2011; published electronically December 22, 2011.
http://www.siam.org/journals/siopt/21-4/80165.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 grants N00014-08-1-0033 and N00014-11-1-0062.

‡IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (benar@impa.br). The work of this author was partially supported by CNPq grants 474944/2010-7, 303583/2008-8, FAPERJ grant E-26/102.821/2008 and PRONEX-Optimization.

New inexact versions of the proximal point method with relative error tolerance were proposed by Solodov and Svaiter [19, 20, 22, 21]. Iteration-complexity results for one of these inexact versions of the proximal point method introduced in [19], namely, the hybrid proximal extragradient (HPE) method, were established in [9]. As a consequence, iteration-complexity results for Korpelevich's extragradient method for VIs with Lipschitz continuous monotone operators, and a variant of Tseng's modified forward-backward splitting (MF-BS) method (see [24]) 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, were also derived by showing that both methods are special cases of the HPE method. A nice feature of the analysis in [9] is that, by working with some suitable termination criteria, it is shown that its complexity results, as opposed to those in [10], also apply to VI and/or monotone inclusion problems with unbounded feasible sets.

In this paper we continue along the same line of investigation as in our previous paper [9], which is to use the HPE method as a general framework to derive iteration-complexity results for specific algorithms for solving various types of structured monotone inclusion problems. More specifically, we will derive iteration-complexity results for an extension of Korpelevich's extragradient method for hemivariational inequality (HVI) problems and a variant of Tseng's MF-FB method for the problem of finding a zero of the sum of a maximal monotone operator and a monotone Lipschitz continuous map whose domain is not necessarily the whole space $\mathbb{R}^n$, thereby relaxing the conditions assumed in our first paper [9]. We also derive iteration-complexity results for these two methods in the context of the SP problem using an error measure specifically tailored to it. In addition, we discuss applications, as well as the complexity, of these two methods to the problem of minimizing the sum of a convex function with Lipschitz continuous gradient and a closed convex (not necessarily differentiable) function in an affine manifold. Finally, we point out how these methods can be used to solve particular instances of the above problem, including one whose objective function converges to infinity as the boundary of its domain is approached.

Previous papers dealing with iteration-complexity analysis of methods for VIs are as follows. Nemirovski [10] studies the complexity of Korpelevich's extragradient method under the assumption that the feasible set is bounded and an upper bound on its diameter is known. Nesterov [13] proposes a new dual extrapolation algorithm for solving VI problems whose termination depends on the guess of a ball centered at the initial iterate.

Asymptotic convergence rate results for extragradient-type methods are thoroughly discussed in [5, 7, 23]. The generalized Korpelevich's method discussed in this paper is well known (see, for example, Noor [14]), but to the best of our knowledge its iteration complexity has not been studied so far. Konnov [6] also has proposed a combined forward-backward splitting and hyperplane projection iteration method closely related to Tseng's MF-BS method in that they differ only in the stepsize used in the extragradient step. In addition, Konnov has established linear convergence rates under some strong regularity assumptions on the data functions.

This paper is organized as follows. Section 2 contains two subsections. Subsection 2.1 reviews some basic definitions and facts on convex functions and the definition and some basic properties of the ε-enlargement of a point-to-set operator. Subsection 2.2 reviews the HPE method and its complexity results. Section 3 contains two subsections. Subsection 3.1 reviews the HVI problem and associated error measures. Subsection 3.2 discusses the generalized SP problem, an associated error measure, and its relationship with the error measures of subsection 3.1. Section 4 contains two subsections. Subsection 4.1 discusses a generalization of Korpelevich's method to the context of the HVI problem and presents corresponding pointwise and ergodic complexity results. Subsection 4.2 discusses a variant of Tseng's MF-BS method and derives complexity bounds for it. Section 5 discusses the specialization of Tseng's MF-BS algorithm and the generalized Korpelevich's method to SP problems and derives iteration complexity based on the error criterion for SP introduced in subsection 3.2. Section 6 contains three subsections. The first two discuss ways of applying Tseng's MF-BS method to two possible reformulations of a certain structured convex optimization problem and presents corresponding iteration-complexity results. Subsection 6.3 briefly discusses applications to some specific but important convex optimization problems.

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. The domain of a function F is denoted by $\operatorname{Dom} F$. The effective domain of a function $f : \mathbb{R}^n \to [-\infty, \infty]$ is defined as $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < \infty\}$.

2. Technical background.

This section contains two subsections. In the first one, we review some basic definitions and facts about convex functions and ε-enlargement of monotone multivalued maps. This subsection also reviews the weak transportation formula for the ε-subdifferentials of closed convex functions and the ε-enlargements of maximal monotone operators, and establishes a weak transportation formula for convex-concave saddle functions. The second subsection reviews the HPE method and the basic complexity results obtained for it in [9].

2.1. The ε-subdifferential and ε-enlargement of monotone operators.

A point-to-set operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a relation $T \subseteq \mathbb{R}^n \times \mathbb{R}^n$, 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 usual to identify T with its graph defined as

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

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

$$\operatorname{Dom} T := \{x \in \mathbb{R}^n : T(x) \neq \emptyset\}.$$

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

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

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 $Gr(S) \supseteq Gr(T)$ implies that $S = T$.

For a scalar $\varepsilon \geq 0$, the ε-subdifferential of a function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is the operator $\partial_\varepsilon f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined as

$$\partial_\varepsilon f(x) = \{v \mid f(y) \geq f(x) + \langle y - x, v \rangle - \varepsilon \ \forall y \in \mathbb{R}^n\} \quad \forall x \in \mathbb{R}^n. \tag{1}$$

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 [16].

The conjugate f^* of f is the function $f^* : \mathbb{R}^n \to [-\infty, \infty]$ 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 2.1. *Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be a proper convex function. Then,*

(a) $\partial_\varepsilon f(x) \subseteq (\partial f)^\varepsilon(x)$ *for any* $\varepsilon \geq 0$ *and* $x \in \mathbb{R}^n$*;*

(b) *if* $v \in \partial f(x)$ *and* $f(y) < \infty$*, then* $v \in \partial_\varepsilon f(y)$*, where* $\varepsilon := f(y) - [f(x) + \langle y - x, v\rangle]$.

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

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

and the *normal cone operator* of X is the point-to-set map $N_X : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ given by

$$N_X(x) = \begin{cases} \emptyset, & x \notin X, \\ \{v \in \mathbb{R}^n \mid \langle y - x, v\rangle \leq 0 \ \forall y \in X\}, & x \in X. \end{cases} \tag{2}$$

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

In [3], Burachik, Iusem and Svaiter introduced the ε-enlargement of maximal monotone operators. In [9] this concept was extended to a generic point-to-set operator in $\mathbb{R}^n$ as follows. Given $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and a scalar ε, define $T^\varepsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ as

$$T^\varepsilon(x) = \{v \in \mathbb{R}^n \mid \langle x - \tilde{x}, v - \tilde{v}\rangle \geq -\varepsilon \quad \forall \tilde{x} \in \mathbb{R}^n,\ \forall \tilde{v} \in T(\tilde{x})\} \quad \forall x \in \mathbb{R}^n. \tag{3}$$

The following result states two useful properties of the operator T^ε that will be needed in our presentation.

PROPOSITION 2.2. *Let $T, T' : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$. Then,*

(a) $T^\varepsilon(x) + (T')^{\varepsilon'}(x) \subseteq (T + T')^{\varepsilon+\varepsilon'}(x)$ *for every* $x \in \mathbb{R}^n$ *and* $\varepsilon, \varepsilon' \in \mathbb{R}$*;*

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

2.2. The hybrid proximal extragradient method. This subsection reviews the HPE method and the basic complexity results obtained for it in [9].

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

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

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

We next review the HPE method introduced in [19] for solving the above problem and state the iteration-complexity results obtained for it in [9].

Hybrid Proximal Extragradient (HPE) Method:

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

(1) Choose $\lambda_k > 0$, and find $\tilde{x}_k, \tilde{v}_k \in \mathbb{R}^n$, $\sigma_k \in [0, \sigma]$, and $\varepsilon_k \geq 0$ such that

$$\tilde{v}_k \in T^{\varepsilon_k}(\tilde{x}_k), \qquad \|\lambda_k\tilde{v}_k + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda_k\varepsilon_k \leq \sigma_k^2\|\tilde{x}_k - x_{k-1}\|^2. \tag{5}$$

(2) Define $x_k = x_{k-1} - \lambda_k\tilde{v}_k$, 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 $\tilde{x}_k$, $\tilde{v}_k$, and ε_k as in (5). The particular choice of λ_k and the algorithm used to compute $\tilde{x}_k, \tilde{v}_k$, and ε_k will depend on the particular implementation of the method and the properties of the operator T. Second, if $\tilde{x} := (\lambda_k T + I)^{-1} x_{k-1}$ is the *exact* proximal point iterate or, equivalently,

$$\tilde{v} \in T(\tilde{x}), \tag{6}$$

$$\lambda_k \tilde{v} + \tilde{x} - x_{k-1} = 0 \tag{7}$$

for some $\tilde{v} \in \mathbb{R}^n$, then $(\tilde{x}_k, \tilde{v}_k) = (\tilde{x}, \tilde{v})$ and $\varepsilon_k = 0$ satisfies (5). Therefore, the error criterion (5) relaxes the inclusion (6) to $\tilde{v} \in T^{\varepsilon}(\tilde{x})$ and relaxes (7) by allowing a small error relative to $\|\tilde{x}_k - x_{k-1}\|$.

We now state a few results about the convergence behavior of the HPE method. The proof of the following result can be found in Lemma 4.2 of [9].

Proposition 2.3. *For any $x^* \in T^{-1}(0)$, the sequence $\{\|x^* - x_k\|\}$ is nonincreasing and*

$$\|x^* - x_0\|^2 \geq (1 - \sigma^2) \sum_{k=1}^{\infty} \|\tilde{x}_k - x_{k-1}\|^2 . \tag{8}$$

The proof of the following result, which establishes the convergence rate of the residual $(\tilde{v}_k, \varepsilon_k)$ of x_k, can be found in Theorem 4.4 of [9].

Theorem 2.4. *Let d_0 be the distance of x_0 to $T^{-1}(0)$. Then, for every $k \in \mathbb{N}$, $\tilde{v}_k \in T^{\varepsilon_k}(\tilde{x}_k)$ and there exists an index $i \leq k$ such that*

$$\|\tilde{v}_i\| \leq d_0 \sqrt{\frac{1+\sigma}{1-\sigma} \left(\frac{1}{\sum_{j=1}^{k} \lambda_j^2} \right)}, \qquad \varepsilon_i \leq \frac{\sigma^2 d_0^2 \lambda_i}{2(1-\sigma^2) \sum_{j=1}^{k} \lambda_j^2}. \tag{9}$$

Theorem 2.4 estimates the quality of the best among the iterates $\tilde{x}_1, \ldots, \tilde{x}_k$. We will refer to these estimates as the *pointwise* complexity bounds for the HPE algorithm.

We will now describe 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 1970s (see [2, 8, 12, 11]) and perhaps even earlier.

The sequence of ergodic means $\{\tilde{x}_k^a\}$ associated with $\{\tilde{x}_k\}$ is

$$\tilde{x}_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i \tilde{x}_i, \qquad \text{where} \qquad \Lambda_k := \sum_{i=1}^{k} \lambda_i. \tag{10}$$

Define also

$$\tilde{v}_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i \tilde{v}_i, \qquad \varepsilon_k^a := \frac{1}{\Lambda_k} \sum_{i=1}^{k} \lambda_i (\varepsilon_i + \langle \tilde{x}_i - \tilde{x}_k^a, \tilde{v}_i - \tilde{v}_k^a \rangle). \tag{11}$$

The following result describes the convergence properties of the ergodic sequence $\{\tilde{x}^a\}$.

THEOREM 2.5. *For every $k \in \mathbb{N}$,*

$$0 \le \varepsilon_k^a \le \frac{1}{2\Lambda_k}\left[2\langle \tilde{x}_k^a - x_0, x_k - x_0\rangle - \|x_k - x_0\|^2\right] \le \frac{2\eta_k d_0^2}{\Lambda_k}, \tag{12}$$

and

$$\tilde{v}_k^a = \frac{1}{\Lambda_k}(x_0 - x_k) \in T^{\varepsilon_k^a}(\tilde{x}_k^a), \qquad \|\tilde{v}_k^a\| \le \frac{2d_0}{\Lambda_k},$$

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

$$\eta_k := 1 + \frac{\sigma\sqrt{\tau_k}}{\sqrt{(1-\sigma^2)}}, \qquad \tau_k = \max_{i=1,\dots,k} \frac{\lambda_i}{\Lambda_k} \le 1. \tag{13}$$

Proof. This result follows immediately from Proposition 4.6 and the proof of Theorem 4.7 of [9]. ∎

3. HVI and SP problems.

In this section, we describe the two classes of problems that we will deal with in this paper, namely, the HVI problem and the generalized SP problem. We will also discuss error measures in the context of these problems that will be used later in the complexity results of sections 4 and 5.

We first give two preliminary definitions.

DEFINITION 3.1. *For a constant $L \ge 0$, the map $F : \operatorname{Dom} F \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is said to be L-Lipschitz continuous on $\Omega \subseteq \operatorname{Dom} F$ if $\|F(x) - F(\tilde{x})\| \le L\|x - \tilde{x}\|$ for every $x, \tilde{x} \in \Omega$. When $\Omega = \operatorname{Dom} F$ we simply say that F is L-Lipschitz continuous.*

DEFINITION 3.2. *$F : \operatorname{Dom} F \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone on $\Omega \subseteq \operatorname{Dom} F$ if $F|_\Omega$ is monotone in the sense of subsection 2.1, i.e., $\langle F(x) - F(\tilde{x}), x - \tilde{x}\rangle \ge 0$ for every $x, \tilde{x} \in \Omega$. When $\Omega = \operatorname{Dom} F$ we simply say that F is monotone.*

3.1. The HVI problem and associated error bounds.

The HVI problem consists of the inclusion problem

$$0 \in T(x) := (F + \partial g)(x), \tag{14}$$

where the following conditions are assumed to hold:

(K.1) g is a closed proper convex function $g : \mathbb{R}^n \to [-\infty, \infty]$;
(K.2) $F : \operatorname{Dom}(F) \to \mathbb{R}^n$ is monotone on $\operatorname{cl}(\operatorname{dom} g) \subseteq \operatorname{Dom} F$;
(K.3) F is L-Lipschitz continuous on $\operatorname{cl}(\operatorname{dom} g)$;
(K.4) the solution set of (14) is nonempty.

We now make a few observations about (14). First, the above assumptions together with Proposition A.1 imply that $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a maximal monotone operator. Second, $x \in \mathbb{R}^n$ is solution of (14) if and only if $x \in \operatorname{dom} g$ and

$$g(y) - g(x) + \langle y - x, F(x)\rangle \ge 0 \quad \forall y \in \mathbb{R}^n. \tag{15}$$

Due to the above interpretation, the inclusion problem (14) is also known as the HVI problem associated with the pair (F, g). In the following, we will denote it by

$HVI(F,g)$. Third, under condition (K.2) and the condition that F is continuous on its domain, (15) is known to be equivalent to

$$g(y) - g(x) + \langle y - x, F(y)\rangle \geq 0 \quad \forall y \in \mathbb{R}^n. \tag{16}$$

Fourth, when $g = \delta_X$ for some nonempty closed convex set $X \subseteq \mathbb{R}^n$, the above problem reduces to the monotone VI problem $VI(F;X)$, i.e., the problem of finding $x \in X$ such that

$$\langle y - x, F(x)\rangle \geq 0 \quad \forall y \in X.$$

We will now discuss different notions of error measures for the above problem. First, we introduce two notions of approximate solutions for problem (14) which are essentially relaxations of the characterizations (15) and (16) of a solution of (14).

DEFINITION 3.3. *A point $x \in \operatorname{dom} g$ is an* ε-strong solution *of* (14) *if*

$$g(y) - g(x) + \langle y - x, F(x)\rangle \geq -\varepsilon \quad \forall y \in \operatorname{dom} g \tag{17}$$

and is an ε-weak solution *of* (14) *if*

$$g(y) - g(x) + \langle y - x, F(y)\rangle \geq -\varepsilon \quad \forall y \in \operatorname{dom} g. \tag{18}$$

Define also $\theta^s(x;F,g)$ and $\theta^w(x;F,g)$ as the smallest ε satisfying (17) *and* (18), *respectively, namely,*

$$\theta^s(x;F,g) := \sup_{y \in \operatorname{dom} g} g(x) - g(y) + \langle x - y, F(x)\rangle, \tag{19}$$

$$\theta^w(x;F,g) := \sup_{y \in \operatorname{dom} g} g(x) - g(y) + \langle x - y, F(y)\rangle. \tag{20}$$

Observe that if $g = \delta_X$, then the above functions reduce to the gap function and the absolute value of the dual gap function mentioned in [4]. Clearly, $\theta^s(x;F,g)$ and $\theta^w(x;F,g)$ are nonnegative for every $x \in \operatorname{dom}(g)$. Note also that, under assumption (K.2), we have $0 \leq \theta^w(x;F,g) \leq \theta^s(x;F,g)$ for every $x \in \operatorname{dom} g$, and hence every ε-strong solution is also an ε-weak solution. For a detailed discussion on error measures for HVI, we refer the reader to Patriksson [15].

For variational inequalities $VI(F,X)$, i.e., problem (14) with $g = \delta_X$, with unbounded feasible sets X, the above notions of approximate solutions are too strong. For example, if $X = \mathbb{R}^n$, the set of ε-strong solutions agrees with the solution set. The following definition relaxes the above notions.

DEFINITION 3.4. *A point $x \in \operatorname{dom} g$ is a* (ρ,ε)-strong solution *(resp.,* (ρ,ε)-weak solution*) of* (14) *if there exists $r \in \mathbb{R}^n$ such that $\|r\| \leq \rho$ and x is an ε-strong (resp., ε-weak) solution of $HVI(F-r;g)$, that is,*

$$\theta^s(x;F-r,g) \leq \varepsilon, \quad (\textit{resp.,}\ \ \theta^w(x;F-r,g) \leq \varepsilon)\,. \tag{21}$$

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

Given $x \in \operatorname{dom} g$ and $c > 0$, define

$$r_c(x;F,g) := c\left[x - \left(I + \frac{1}{c}\partial g\right)^{-1}\left(x - \frac{1}{c}F(x)\right)\right], \tag{22}$$

$$\theta_c(x;F,g) := \sup_{y \in \mathbb{R}^n} g(x) - g(y) + \langle x - y, F(x)\rangle - \frac{c}{2}\|y - x\|^2. \tag{23}$$

It is well known that $x \in \mathbb{R}^n$ is a solution of (14) if and only if $\theta_c(x; F, g) = 0$, both of which are also equivalent to $r_c(x; F, g) = 0$. Hence, $\theta_c(x; F, g)$, or the size of $r_c(x; F, g)$, can be used as an error measure for x. Clearly, since $g(y) = \infty$ for all $y \notin \operatorname{dom} g$, the above supremum can be equivalently taken with respect to $y \in \operatorname{dom} g$.

The following result describes some important relationships between the different error measures introduced in this section, as well as the ε-subdifferential of g and/or the ε-enlargement of $F + \partial g$.

PROPOSITION 3.5. *Let $x \in \operatorname{dom} g$. Then,*

(a) *(r, ε) is a strong residual of x for $HVI(F, g)$ if and only if $r \in F(x) + \partial_\varepsilon g(x)$;*

(b) *if (r, ε) is a weak residual of x for $HVI(F, g)$, then $r \in (F + \partial g)^\varepsilon(x)$;*

(c) *if $r \in (F^{\varepsilon'} + \partial g_{\varepsilon''})(x)$ and $\varepsilon' + \varepsilon'' \leq \varepsilon$, then (r, ε) is a weak residual of x for $HVI(F, g)$.*

(d) *if $c \geq 2L$, then $\theta_c(x; F, g) \leq \theta^w(x; F, g)$.*

(e) *if (r, ε) is a strong residual of x for $HVI(F, g)$, then, for any $c > 0$,*

$$\theta_c(x; F, g) \leq \frac{1}{2c}\|r\|^2 + \varepsilon; \tag{24}$$

moreover, for any fixed $c > 0$, there exists a unique strong residual (r, ε) of x for $HVI(F, g)$ for which equality holds in (24)*, namely, the pair $(r, \varepsilon) = (r_c(x; F, g), \varepsilon_c(x; F, g))$, where*

$$\varepsilon_c(x; F, g) := g(x) - g(y_c) - \langle x - y_c, r_c - F(x)\rangle \geq 0, \qquad y_c := x - c^{-1} r_c(x; F, g).$$

Proof. See Propositions C.1 and C.2 and Theorem C.3 in Appendix C for a proof of this result. □

The following result shows that if one knows that x is a $(\rho/\sqrt{2}, \varepsilon/2)$-strong solution of $HVI(F, g)$ without an explicit certificate (r, ε) to back up this knowledge, then it is possible to explicitly construct such a certificate for a slightly larger tolerance, i.e., (ρ, ε).

PROPOSITION 3.6. *If $x \in \operatorname{dom} g$ is a $(\rho/\sqrt{2}, \varepsilon/2)$-strong solution of $HVI(F, g)$ and $\bar{c} := \rho^2/(2\varepsilon)$, then the pair $(r_{\bar{c}}(x; F, g), \varepsilon_{\bar{c}}(x; F, g))$ is a strong residual of x for $HVI(F, g)$ satisfying the estimates*

$$\|r_{\bar{c}}(x; F, g)\| \leq \rho, \qquad \varepsilon_{\bar{c}}(x; F, g) \leq \varepsilon.$$

Proof. See Proposition C.4 in Appendix C for a proof of this result. □

It follows from the observation in the paragraph following (20) that, under assumption (K.2), every strong residual (r, ε) of x for $HVI(F, g)$ is also a weak residual of x for $HVI(F, g)$. We will now state a sort of a converse of this claim whose proof is given in Proposition C.5 in Appendix C.

PROPOSITION 3.7. *If condition* (K.3) *holds and (r, ε) is a weak residual of x for $HVI(F, g)$, then, for any positive scalar $c \geq 2L$, the vector*

$$r_c := r_c(x; F - r, g)$$

satisfies $\|r_c\| \leq \sqrt{2c\varepsilon}$, and the pair $(r + r_c, \varepsilon)$ is a strong residual of x for $HVI(F, g)$.

We will now present a result which will be useful in obtaining sharper iteration-complexity bounds for the sequence of ergodic means generated by the algorithms discussed in section 4. First, we introduce the following constant associated with a Lipschitz continuous map.

DEFINITION 3.8. *For a map $F : \mathrm{Dom}\, F \subseteq \mathbb{R}^n \to \mathbb{R}^n$ which is monotone on $X \subseteq \mathrm{Dom}\, F$, let $\mathcal{N}(F; X)$ be the infimum of all $L \geq 0$ such that there exist an L-Lipschitz monotone map $G : X \to \mathbb{R}^n$ and a monotone affine map $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$F(x) = G(x) + \mathcal{A}(x) \quad \forall x \in X.$$

We now make a few observations about the above definition. Clearly, if F is a monotone affine map, then $\mathcal{N}(F; X) = 0$ for any $X \subseteq \mathbb{R}^n$. Note also that if F is monotone and L-Lipschitz on X, then $\mathcal{N}(F; X) \leq L$. We note, however, that $\mathcal{N}(F; X)$ 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 $G : \mathrm{Dom}\, G \to \mathbb{R}^n$ is monotone and L-Lipschitz on $X \subseteq \mathrm{Dom}\, G$, then we have $\mathcal{N}(F; X) \leq L$. Hence, in the latter case, $\mathcal{N}(F; X)$ is bounded by a constant which does not depend on μ, while the Lipschitz constant of F with respect to X converges to ∞ as $\mu \to \infty$, as long as $\mathcal{A}$ is not constant.

We are now ready to state the aforementioned result.

THEOREM 3.9. *Assume that conditions* (K.1)–(K.3) *listed in subsection* 4.1 *hold. Let $x_i, v_i \in \mathbb{R}^n$ and $\varepsilon_i, \alpha_i \in \mathbb{R}_+$, for $i = 1, \ldots, k$, be such that*

$$v_i \in (F + \partial_{\varepsilon_i} g)(x_i), \quad i = 1, \ldots, k, \qquad \sum_{i=1}^{k} \alpha_i = 1, \tag{25}$$

and define

$$x^a = \sum_{i=1}^{k} \alpha_i x_i, \quad v^a = \sum_{i=1}^{k} \alpha_i v_i, \quad \varepsilon^a = \sum_{i=1}^{k} \alpha_i [\varepsilon_i + \langle x_i - x^a, v_i \rangle], \quad F^a := \sum_{i=1}^{k} \alpha_i F(x_i). \tag{26}$$

Then, the following statements hold:

(a) *$\varepsilon^a \geq 0$, and (v^a, ε^a) is a weak residual of x^a for $HVI(F, g)$;*

(b) *for every $c \geq 2\mathcal{N}(F; \mathrm{dom}\, g)$, the vector $r_c := r_c(x^a; F - v^a, g)$ satisfies $\|r_c\| \leq \sqrt{2\varepsilon^a c}$, and the pair $(v^a + r_c, \varepsilon^a)$ is a strong residual of x^a for $HVI(F, g)$;*

(c) *if Ω is a closed convex set such that $\mathrm{dom}\, g \subseteq \Omega \subseteq \mathrm{Dom}\, F$, then, for every $c \geq 2\mathcal{N}(F; \Omega)$, the vector*

$$\hat{r}_c := r_c(x^a; F - F^a; \delta_\Omega) = c[x^a - P_\Omega(x^a - c^{-1}(F(x^a) - F^a))] \tag{27}$$

satisfies $\|\hat{r}_c\| \leq \sqrt{2\varepsilon^a c}$, and the pair $(v^a + \hat{r}_c, \varepsilon^a)$ is a strong residual of x^a for $HVI(F, g)$.

In particular, there exists $r \in \mathbb{R}^n$ such that $v^a + r \in (F + \partial_{\varepsilon^a} g)(x^a)$ and $\|r\| \leq 2\sqrt{\varepsilon^a \mathcal{N}(F; \mathrm{dom}\, g)}$.

3.2. The generalized SP problem and associated error bounds. In this subsection, we describe the generalized SP problem and its reformulation as a problem as in (14). Hence, the generalized Korpelevich's extragradient method or the variant of Tseng's MF-BS algorithm can be used to approximately solve this problem and iteration-complexity results similar to those derived at the beginning of this section apply. In this subsection, we also describe a different notion of an approximate solution for the generalized SP problem, i.e., that of an approximate saddle point, and establish an iteration-complexity result to obtain such solution.

We first introduce a few definitions. Let $\Psi : \mathrm{dom}\,\Psi \subseteq \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and let two proper closed convex functions $g_x : \mathbb{R}^n \to [-\infty, \infty]$ and $g_y : \mathbb{R}^m \to [-\infty, \infty]$ be such that $\mathrm{dom}\, g_x \times \mathrm{dom}\, g_y \subseteq \mathrm{dom}\,\Psi$ be given. Also, define

$$X := \mathrm{dom}\, g_x, \quad Y := \mathrm{dom}\, g_y,$$

and the function $\widehat{\Psi} : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, \infty]$ as

$$\widehat{\Psi}(x, y) = \begin{cases} \Psi(x, y) + g_x(x) - g_y(y), & (x, y) \in X \times Y, \\ \infty, & x \notin X, \\ -\infty, & x \in X, y \notin Y. \end{cases} \tag{28}$$

The generalized SP problem determined by the triple $(\Psi; g_x, g_y)$, which we denote by $GSP(\Psi; g_x, g_y)$, consists of finding a pair $(x, y) \in X \times Y$ such that

$$\widehat{\Psi}(x, y') \le \widehat{\Psi}(x, y) \le \widehat{\Psi}(x', y) \quad \forall (x', y') \in X \times Y.$$

For a fixed map Ψ, each pair (g_x, g_y) determines a different SP problem. The case where $g_x = \delta_X$ and $g_y = \delta_Y$ yields the standard SP simply by $SP(\Psi; X \times Y)$.

Definition 3.10. *The pair $(x, y) \in X \times Y$ is called an ε-saddle point of $GSP(\Psi; g_x, g_y)$ if*

$$\theta^{sp}((x, y); \Psi, g_x, g_y) := \sup\{\widehat{\Psi}(x, y') - \widehat{\Psi}(x', y) : (x', y') \in X \times Y\} \le \varepsilon.$$

The function $\theta^{sp}(\cdot; \Psi, g_x, g_y)$ is also known as the gap function associated with $SP(\Psi; g_x, g_y)$ in that it can be viewed as the difference between a primal function $p(\cdot) = p(\cdot; \Psi, g_x, g_y) : X \to \mathbb{R}$ and a dual function $d(\cdot) = d(\cdot, \Psi; g_x, g_y) : Y \to \mathbb{R}$ defined as

$$p(x) = \sup_{y' \in Y} \Psi(x, y') + g_x(x) - g_y(y') \quad \forall x \in X,$$
$$d(y) = \inf_{x' \in X} \Psi(x', y) + g_x(x') - g_y(y) \quad \forall y \in Y.$$

Clearly, (x, y) is an ε-saddle point of $GSP(\Psi; g_x, g_y)$ if and only if $(x, y) \in X \times Y$ and $(0, 0) \in \partial_\varepsilon[\widehat{\Psi}(\cdot, y) - \widehat{\Psi}(x, \cdot)](x, y)$. Moreover, $\theta^{sp}((x, y); \Psi, g_x, g_y)$ is the smallest $\varepsilon \ge 0$ satisfying one of these two equivalent conditions. More generally, we can introduce the following more general definition of an approximate saddle point for $GSP(\Psi; g_x, g_y)$.

Definition 3.11. *The pair $(x, y) \in X \times Y$ is called a (ρ, ε)-saddle point of $GSP(\Psi; g_x, g_y)$ if there exist a pair $r = (r_x, r_y) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $\|r\| \le \rho$ and (x, y) is an ε-saddle point of $GSP(\Psi_r; g_x, g_y)$, where $\Psi_r : \mathrm{dom}\,\Psi \to \mathbb{R}$ is defined as*

$$\Psi_r(x', y') = \Psi(x', y') + \big\langle(-r_x, r_y), (x', y')\big\rangle \quad \forall (x', y') \in \mathrm{dom}\,\Psi. \tag{29}$$

Moreover, any such pair (r, ε) will be called an SP-residual of (x, y) for $GSP(\Psi_r; g_x, g_y)$.

For the sake of future reference, we state the following simple result without proof.

Proposition 3.12. *For a point $(x, y) \in X \times Y$, the pair $(r, \varepsilon) = ((r_x, r_y), \varepsilon)$ is an SP-residual of (x, y) for $GSP(\Psi_r; g_x, g_y)$ if and only if $(r_x, r_y) \in \partial_\varepsilon[\widehat{\Psi}(\cdot, y) - \widehat{\Psi}(x, \cdot)](x, y)$.*

In the following, we will discuss the close connection between $GSP(\Psi; g_x, g_y)$ and a related HVI problem and, as a by-product, the specializations of the algorithms discussed in section 4 to $GSP(\Psi; g_x, g_y)$. We first need to make some assumptions on Ψ:

(S.1) $\operatorname{dom}\Psi$ is open, Ψ is differentiable, and $\operatorname{dom}\Psi \supseteq \operatorname{cl}(X \times Y)$;
(S.2) the function $\Psi(\cdot, y) - \Psi(x, \cdot) : \operatorname{cl}(X \times Y) \to \mathbb{R}$ is convex for every $(x, y) \in \operatorname{cl}(X \times Y)$;
(S.3) $\nabla\Psi$ is L-Lipschitz continuous.

Define the functions $F : \operatorname{dom}\Psi \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, \infty]$ as

$$F(x, y) := (\nabla_x \Psi(x, y), -\nabla_y \Psi(x, y)), \qquad g(x, y) = g_x(x) + g_y(y). \tag{30}$$

PROPOSITION 3.13. *Assume that Ψ satisfies conditions* (S.1) *and* (S.2)*, and consider the functions F and g defined according to* (30)*. Then,*

$$\theta^w((x, y); F, g) \le \theta^{sp}((x, y); \Psi, g_x, g_y)) \le \theta^s((x, y); F, g) \qquad \forall (x, y) \in X \times Y. \tag{31}$$

Equivalently, every ε-strong solution $(x, y) \in X \times Y$ of $HVI(F; g)$ is an ε-saddle point of $GSP(\Psi; g_x, g_y)$, and the latter is always an ε-weak solution of $HVI(F; g)$.

Proof. To prove the first inequality in (31), assume that (x, y) is an ε-saddle point of $GSP(\Psi; g_x, g_y)$. Then, using relations (28) and (30) and the assumptions (S.1) and (S.2), we conclude that for every $(x', y') \in X \times Y$, we have

$$\begin{aligned}
-\varepsilon \le \widehat{\Psi}(x', y) - \widehat{\Psi}(x, y') &= [\Psi(x', y) + g_x(x') - g_y(y)] - [\Psi(x, y') + g_x(x) - g_y(y')] \\
&= [\Psi(x', y) - \Psi(x, y')] + g(x', y') - g(x, y) \\
&= [\Psi(x', y') - \Psi(x, y')] + [(-\Psi)(x', y') - (-\Psi)(x', y)] + g(x', y') - g(x, y) \\
&\le \big\langle \nabla_x \Psi(x', y'), x' - x \big\rangle + \big\langle -\nabla_y \Psi(x', y'), y' - y \big\rangle + g(x', y') - g(x, y) \\
&= \big\langle F(x', y'), (x', y') - (x, y) \big\rangle + g(x', y') - g(x, y),
\end{aligned}$$

which clearly implies that (x, y) is an ε-weak solution of $HVI(F; g)$ in view of Definition 3.3.

To show the second inequality in (31), set $\varepsilon = \theta^s((x, y); F, g)$ and observe that $(0, 0) \in F(x, y) + \partial_\varepsilon g(x, y)$. Using the fact that

$$F(x, y) \in \partial[\Psi|_{X\times Y}(\cdot, y) - \Psi|_{X\times Y}(x, \cdot)](x, y),$$

we then conclude that

$$\begin{aligned}
(0, 0) &\in \partial[\Psi|_{X\times Y}(\cdot, y) - \Psi|_{X\times Y}(x, \cdot)](x, y) + \partial_\varepsilon g(x, y) \\
&\subseteq \partial_\varepsilon[\Psi(\cdot, y) - \Psi(x, \cdot) + g](x, y) = \partial_\varepsilon[\widehat{\Psi}(\cdot, y) - \widehat{\Psi}(x, \cdot)](x, y),
\end{aligned}$$

and hence that $\theta^{sp}((x, y); \Psi, g_x, g_y) \le \varepsilon = \theta^s((x, y); F, g)$. ◻

4. Variants of Korpelevich's and Tseng's MF-BS methods. In this section, we present two algorithms for solving special types of monotone inclusion problems. The first, discussed in subsection 4.1, is an extension of Korpelevich's method for solving the inclusion problem (4), where T is the sum of a Lipschitz continuous map and the subdifferential of a closed convex function. The second, discussed in subsection 4.2, is a variant of Tseng's MF-BS method for solving the inclusion problem (4), where T is the sum of a Lipschitz continuous map and a maximal monotone operator. We show that both methods are special cases of the HPE method and, as a consequence, derive both pointwise and ergodic iteration-complexity results for them that follow naturally from the general convergence theory outlined in subsection 2.2 for the HPE method and the results derived in section 3.

4.1. Generalized Korpelevich's method for solving HVI problems. In this subsection, we analyze a generalized version of Korpelevich's method for solving HVI (14). We note that this section closely parallels section 5 of [9]. However, the technical tools developed in this paper to derive some of the results of this section are much more sophisticated due not only to the generalization to HVIs but also to the more general assumptions imposed.

We start by stating the method.

Generalized Korpelevich's extragradient algorithm:

(0) Let $x_0 \in \mathrm{Dom}(\partial g)$ and $\sigma \in (0,1)$ be given, and set $\lambda = \sigma/L$ and $k = 1$.

(1) Compute

(32)
$$\tilde{x}_k = (I + \lambda\partial g)^{-1}(x_{k-1} - \lambda F(x_{k-1})), \quad x_k = (I + \lambda\partial g)^{-1}(x_{k-1} - \lambda F(\tilde{x}_k)).$$

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

Note that if x_{k-1} is in $\mathrm{Dom}(\partial g)$, then, in view of Assumptions K.1 and K.2 and the fact that $\mathrm{Dom}(\partial g) \subseteq \mathrm{dom}\, g$, the quantities $F(x_{k-1})$ and $\tilde{x}_k$ are well defined, $\tilde{x}_k \in \mathrm{Dom}(\partial g)$, and the same holds for x_k. Hence, the algorithm is well defined and both sequences $\{x_k\}$ and $\{\tilde{x}_k\}$ are in $\mathrm{Dom}(\partial g)$. Moreover, when $g = \delta_X$ for some nonempty closed convex set $X \subseteq \mathbb{R}^n$, the above algorithm reduces to Korpelevich's method for solving the monotone variational inequality problem $VI(F;X)$. We also observe that the iterates $\tilde{x}_k$ and x_k in (32) are also characterized as

$$\tilde{x}_k = \mathrm{argmin}\langle F(x_{k-1}), x\rangle + g(x) + \frac{1}{2\lambda}\|x - x_{k-1}\|^2,$$
$$x_k = \mathrm{argmin}\langle F(\tilde{x}_k), x\rangle + g(x) + \frac{1}{2\lambda}\|x - x_{k-1}\|^2.$$

The next result establishes that the generalized Korpelevich's extragradient algorithm is a special case of the HPE method. Its proof is quite similar to the one given for Theorem 5.1 in [9], but, for sake of completeness, we include its proof in Appendix B.

PROPOSITION 4.1. *Let $\{\tilde{x}_k\}$ be the sequence generated by the generalized Korpelevich's extragradient algorithm, and, for each k, define*

$$q_k = \frac{1}{\lambda}(x_{k-1} - x_k) - F(\tilde{x}_k), \qquad p_k = \frac{1}{\lambda}(x_{k-1} - \tilde{x}_k) - F(x_{k-1}), \tag{33}$$

$$\varepsilon_k = g(\tilde{x}_k) - g(x_k) - \langle \tilde{x}_k - x_k, q_k\rangle, \quad \tilde{v}_k = F(\tilde{x}_k) + q_k. \tag{34}$$

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

(a) $q_k \in \partial_{\varepsilon_k} g(\tilde{x}_k)$ *and* $\tilde{v}_k \in [F + \partial_{\varepsilon_k} g](\tilde{x}_k) \subseteq [F + \partial g]^{\varepsilon_k}(\tilde{x}_k)$*;*

(b) $x_k = x_{k-1} - \lambda\tilde{v}_k$*;*

(c) $\|\lambda\tilde{v}_k + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda\varepsilon_k \le \sigma^2\|\tilde{x}_k - x_{k-1}\|^2$*;*

(d) $p_k \in \partial g(\tilde{x}_k)$ *and*

$$\|F(\tilde{x}_k) + p_k\| \le \frac{(1+\sigma)L}{\sigma}\|\tilde{x}_k - x_{k-1}\|. \tag{35}$$

As a consequence of (a)–(c)*, it follows that the generalized Korpelevich's extragradient algorithm is a special case of the HPE method.*

Note that in view of Proposition 4.1(a), we have $\tilde{v}_k \in F(\tilde{x}_k) + \partial_{\varepsilon_k} g(\tilde{x}_k)$ and, due to the fact that the generalized Korpelevich's algorithm is a special case of the HPE method together with Theorem 2.4, we also have $\max\{\|\tilde{v}_i\|, \varepsilon_i\} = \mathcal{O}(1/\sqrt{k})$ for some $i \le k$. The following theorem provides a variant of this result where a vector close to $\tilde{v}_k$ satisfies the above conclusions with $\varepsilon_i = 0$.

THEOREM 4.2. *Let $\{\tilde{x}_k\}$ and $\{x_k\}$ be the sequences generated by the generalized Korpelevich's extragradient algorithm, and let $\{p_k\}$ be the sequence defined in* (33). *Then, for every $k \in \mathbb{N}$, $F(\tilde{x}_k) + p_k \in [F + \partial g](\tilde{x}_k)$, and there exists $i \le k$ such that*

$$\|F(\tilde{x}_i) + p_i\| \le \frac{Ld_0}{\sigma}\sqrt{\frac{1+\sigma}{k(1-\sigma)}},$$

where d_0 is the distance of x_0 to the solution set of $HVI(F,g)$.

Proof. The inclusion $F(\tilde{x}_k)+p_k \in [F+\partial g](\tilde{x}_k)$ follows immediately from the first part of Proposition 4.1(d). Also, by Proposition 4.1, we know that the generalized Korpelevich's extragradient method is a special case of the HPE framework and that, for every $k \in \mathbb{N}$, (35) holds. Hence, by Proposition 2.3, (8) holds and, as a consequence,

$$\min_{1\le i\le k} \|\tilde{x}_i - x_{i-1}\|^2 \le \frac{1}{k}\sum_{i=1}^{k} \|\tilde{x}_i - x_{i-1}\|^2 \le \frac{d_0^2}{k(1-\sigma^2)}.$$

The result now follows from the above inequality and (35). $\square$

We will now present ergodic complexity results for the generalized Korpelevich's method. These results use not only the general complexity results for the HPE method but also some of the new results derived in subsection 3.1.

THEOREM 4.3. *Let $\{\tilde{x}_k\}$ and $\{x_k\}$ be the sequences generated either by the variant of Tseng's MF-BS method or by the generalized Korpelevich's extragradient algorithm, and for every $k \in \mathbb{N}$, define*

$$\tilde{x}_k^a = \frac{1}{k}\sum_{i=1}^{k}\tilde{x}_i, \qquad \tilde{v}_k^a = \frac{1}{k\lambda}(x_0 - x_k), \qquad \tilde{F}_k^a = \sum_{i=1}^{k} F(\tilde{x}_i), \tag{36}$$

$$\tilde{\varepsilon}_k^a = \frac{1}{2\Lambda_k}\left[2\langle \tilde{x}_k^a - x_0, x_k - x_0\rangle - \|x_k - x_0\|^2\right]. \tag{37}$$

Let d_0 be the distance of x_0 to the solution set of $HVI(F,g)$. Then, for every $k \in \mathbb{N}$, $(\tilde{v}_k^a, \tilde{\varepsilon}_k^a)$ is a weak residual of $\tilde{x}_k^a$ for $HVI(F,g)$, and

$$\|\tilde{v}_k^a\| \le \frac{2Ld_0}{k\sigma}, \qquad \tilde{\varepsilon}_k^a \le \frac{2Ld_0^2\bar{\eta}_k}{k\sigma}, \tag{38}$$

where

$$\bar{\eta}_k := 1 + \frac{\sigma}{\sqrt{k(1-\sigma^2)}}. \tag{39}$$

As a consequence, given tolerances $\rho > 0$ and $\varepsilon > 0$, an ergodic iterate $\tilde{x}_k^a$ with a (easily computable) weak residual $(\tilde{v}_k^a, \tilde{\varepsilon}_k^a)$ satisfying $\|\tilde{v}_k^a\| \le \rho$ and $\tilde{\varepsilon}_k^a \le \varepsilon$ will be found in at most

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

iterations.

Proof. By Proposition 4.1, the generalized Korpelevich's method is a special case of the HPE method with $T = F+\partial g$. Hence, the first conclusion of the theorem follows immediately from Theorem 2.5 with $\lambda_k = \sigma/L$ for every k and Theorem 3.9(a) with $x_i = \tilde{x}_i$, $v_i = \tilde{v}_i$, and $\alpha_i = 1/k$ for $i = 1, \ldots, k$. The last part of the theorem follows immediately from the first one. $\square$

Letting $r_k := \tilde{v}_k^a + r_{2L}(\tilde{x}_k^a; F - \tilde{v}_k^a, g)$, we conclude from Proposition 3.7 that $(r_k, \tilde{\varepsilon}_k^a)$ is a strong residual of $\tilde{x}_k^a$ such that

$$\|r_k\| \leq \|\tilde{v}_k^a\| + 2\sqrt{L\tilde{\varepsilon}_k^a} = \mathcal{O}\left(\frac{Ld_0}{\sqrt{k}}\right).$$

This argument would yield a complexity bound of

$$\mathcal{O}\left(\max\left\{\frac{L^2 d_0^2}{\rho^2}, \frac{Ld_0^2}{\varepsilon}\right\}\right)$$

iterations to find an ergodic iterate with a (ρ,ε)-strong residual $(r_k, \tilde{\varepsilon}_k^a)$.

The following result states that the factor L^2 on the first term inside the maximand can actually be improved to $L\mathcal{N}(F; \operatorname{dom} g)$ if a different residual is used in place of r_k. Due to its level of difficulty, its proof will be given in Appendix E. We note that for the particular case in which F is assumed to be monotone and Lipschitz everywhere on $\Re^n$, an easier version of this result is given in Theorem 5.5 of [9] in the context of VI. Under the more general assumptions on (14) imposed here, the current version heavily makes use of Theorem 3.9, which is derived for the first time in this paper.

THEOREM 4.4. *Let $(\rho,\varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, and define $\bar{c} := \rho^2/(2\varepsilon)$. Consider the sequence of ergodic iterates $\{\tilde{x}_k^a\}$ generated by the generalized Korpelevich's method. Then, there exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{Ld_0^2}{\varepsilon}, \frac{Ld_0}{\rho} + \frac{d_0^2 L\mathcal{N}(F;\operatorname{dom} g)}{\rho^2}\right]\right) \tag{40}$$

such that, for any $k \geq k_0$, the pair $(r_{\bar{c}}(\tilde{x}_k^a; F, g), \varepsilon_{\bar{c}}(\tilde{x}_k^a; F, g))$ is a strong residual of $\{\tilde{x}_k^a\}$ satisfying

$$\|r_{\bar{c}}(\tilde{x}_k^a; F, g)\| \leq \rho, \qquad \varepsilon_{\bar{c}}(\tilde{x}_k^a; F, g) \leq \varepsilon. \tag{41}$$

As a consequence, any such ergodic iterate $\tilde{x}_k^a$ is a (ρ,ε)-strong solution of $HVI(F,g)$.

4.2. A variant of Tseng's MF-BS method. In this section, we analyze a variant of Tseng's MF-BS method [24] for solving the inclusion problem

$$0 \in T(x) := (F+B)(x), \tag{42}$$

where the following assumptions hold:

(T.1) $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone;
(T.2) $F : \operatorname{Dom}(F) \to \mathbb{R}^n$ is monotone on $\operatorname{Dom} B \subseteq \operatorname{Dom} F$;
(T.3) F is L-Lipschitz continuous on a closed convex set Ω such that $\operatorname{Dom}(B) \subseteq \Omega \subseteq \operatorname{Dom}(F)$;
(T.4) the solution set of (42) is nonempty.

We observe that Tseng's original assumptions (see [24]) are slightly more general than the above assumptions in that the set Ω does not have to include $\operatorname{Dom} B$ but only a solution of (42). However, the above assumptions are more general than those imposed in section 6 of [9], where the complexity of a special case of Tseng's MF-BS method is studied under the condition that $\Omega = \Re^n$.

We note also that, under the above assumptions, $T = F + B$ is a maximal monotone operator such that $\operatorname{Dom} T = \operatorname{Dom} B \subseteq \Omega$ (see Proposition A.1).

We now state the variant of Tseng's MF-BS method studied in this paper.

A variant of Tseng's MF-BS method:

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

(1) Compute

$$x'_{k-1} = P_\Omega(x_{k-1}), \tag{43}$$

$$\tilde{x}_k = (I + \lambda B)^{-1}(x_{k-1} - \lambda F(x'_{k-1})), \tag{44}$$

$$x_k = \tilde{x}_k - \lambda(F(\tilde{x}_k) - F(x'_{k-1})). \tag{45}$$

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

Note that if $\Omega = \mathbb{R}^n$, then $x'_{k-1} = x_{k-1}$, and hence the above algorithm reduces to a special case of Tseng's MF-BS method, whose iteration complexity was studied in [9]. We also note that, when $\Omega \neq \mathbb{R}^n$, the above algorithm is different from Tseng's MF-BS method.

The next result establishes that the above algorithm is a special case of the HPE method in which $\varepsilon_k = 0$ for all $k \in \mathbb{N}$.

PROPOSITION 4.5. *Let $\{x_k\}$, $\{x'_k\}$, and $\{\tilde{x}_k\}$ be the sequences generated by the variant of Tseng's MF-BS method, and, for each k, define*

$$b_k = \frac{1}{\lambda}(x_{k-1} - \tilde{x}_k) - F(x'_{k-1}), \tag{46}$$

$$\tilde{v}_k = F(\tilde{x}_k) + b_k. \tag{47}$$

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

(a) *$b_k \in B(\tilde{x}_k)$ and $\tilde{v}_k \in [F+B](\tilde{x}_k)$;*

(b) *$x_k = x_{k-1} - \lambda \tilde{v}_k$;*

(c) *$\|\lambda \tilde{v}_k + \tilde{x}_k - x_{k-1}\| \leq \sigma \|\tilde{x}_k - x_{k-1}\|$.*

As a consequence of (a)–(c), *it follows that the new variant of Tseng's MF-BS method is a special case of the HPE method in which $\varepsilon_k = 0$ for all $k \in \mathbb{N}$.*

Proof. The first inclusion in (a) follows from (44) and (46), while the second inclusion follows from the first one and (47). Statement (b) follows from (45), (46), and (47). For (c), note that relations (43), (46), and (47), the definition of λ, assumption (T.3), and the fact that $\tilde{x}_k \in \operatorname{Dom} B \subseteq \Omega$ and P_Ω is a nonexpansive operator imply

$$\begin{aligned}\|\lambda \tilde{v}_k + \tilde{x}_k - x_{k-1}\| &= \|\lambda(F(\tilde{x}_k) + b_k) + \tilde{x}_k - x_{k-1}\| = \|\lambda(F(\tilde{x}_k) - F(x'_{k-1})\| \\ &\leq \lambda L \|\tilde{x}_k - x'_{k-1}\| = \sigma \|P_\Omega(\tilde{x}_k) - P_\Omega(x_{k-1})\| \leq \sigma \|\tilde{x}_k - x_{k-1}\|. \quad \square\end{aligned}$$

We will now state a result that follows as an immediate consequence of the previous proposition and Theorem 2.4.

THEOREM 4.6. *Let $\{\tilde{x}_k\}$ and $\{x_k\}$ be the sequences generated by the variant of Tseng's MF-BS algorithm, and let $\{b_k\}$ be the sequence defined in* (46). *Then, for every $k \in \mathbb{N}$, $F(\tilde{x}_k) + b_k \in [F+B](\tilde{x}_k)$, and there exists $i \leq k$ such that*

$$\|F(\tilde{x}_i) + b_i\| \leq \frac{Ld_0}{\sigma}\sqrt{\frac{1+\sigma}{k(1-\sigma)}},$$

where d_0 is the distance of x_0 to the solution set of (42).

We end this section by making an observation about the ergodic behavior of the variant of Tseng's MF-BS algorithm. Assume that $B = \partial g$, where g is a closed proper convex function. In this case, all the ergodic results stated in subsection 4.1, namely, Theorems 4.3 and 4.4, hold for the variant of Tseng's MF-BS algorithm as well.

5. Applications to SP problems. In this section, we present specializations of the generalized Korpelevich's method and the variant of Tseng's MF-BS method for solving $GSP(\Psi; g_x, g_y)$. They are essentially the generalized Korpelevich's method and the variant of Tseng's MF-BS method described in subsection 4.1 applied to $HVI(F; g)$ with F and g given by (30).

We start by stating the specializations of two methods discussed in section 4 in the context of $GSP(\Psi; g_x, g_y)$.

Generalized Korpelevich's extragradient algorithm for $GSP(\Psi; g_x, g_y)$:

(0) Let $(x_0, y_0) \in \mathrm{Dom}(\partial g_x) \times \mathrm{Dom}(\partial g_y)$ and $\sigma \in (0,1)$ be given, and set $\lambda = \sigma/L$ and $k = 1$.

(1) Compute

$$(\tilde{x}_k, \tilde{y}_k) = \operatorname{argmin} \left\{ \begin{array}{l} \langle F(x_{k-1}, y_{k-1}), (x,y) \rangle + g_x(x) + g_y(y) \\ + \frac{1}{2\lambda} \|(x - x_{k-1}, y - y_{k-1})\|^2 \end{array} \right\},$$

$$(x_k, y_k) = \operatorname{argmin} \left\{ \begin{array}{l} \langle F(\tilde{x}_k, \tilde{y}_k), (x,y) \rangle + g_x(x) + g_y(y) \\ + \frac{1}{2\lambda} \|(x - x_{k-1}, y - y_{k-1})\|^2 \end{array} \right\}.$$

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

To state the specialization of the variant of Tseng's MF-BS method to the context of $GSP(\Psi; g_x, g_y)$, we first introduce one more assumption.

(S.4) There exist closed convex sets Ω_x and Ω_y such that $X \times Y \subseteq \Omega_x \times \Omega_y \subseteq \mathrm{dom}\,\Psi$.

Variant of Tseng's MF-BS method for $GSP(\Psi; g_x, g_y)$:

(0) Let $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and $\sigma \in (0,1)$ be given, and set $\lambda = \sigma/L$ and $k = 1$.

(1) Compute

$$(x'_{k-1}, y'_{k-1}) = (P_{\Omega_x}(x_{k-1}), P_{\Omega_y}(y_{k-1})),$$

$$(\tilde{x}_k, \tilde{y}_k) = \operatorname{argmin} \left\{ \begin{array}{l} \langle F(x'_{k-1}, y'_{k-1}), (x,y) \rangle + g_x(x) + g_y(y) \\ + \frac{1}{2\lambda} \|(x - x_{k-1}, y - y_{k-1})\|^2 \end{array} \right\},$$

$$(x_k, y_k) = (\tilde{x}_k, \tilde{y}_k) - \lambda[F(\tilde{x}_k, \tilde{y}_k) - F(x'_{k-1}, y'_{k-1})].$$

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

Clearly, all the results derived in section 4 apply to the above two algorithms. For the sake of brevity, we will not translate their statements to the context of $GSP(\Psi; g_x, g_y)$. However, we will show that Theorem 4.3 can be strengthened by replacing the error measure θ^w with θ^{sp} (see inequality (31)).

Before doing so, we state the following technical result.

PROPOSITION 5.1. *Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be given convex sets, and let $\Gamma : X \times Y \to \mathbb{R}$ be a function such that, for each pair $(x,y) \in X \times Y$, the function $\Gamma(\cdot, y) - \Gamma(x, \cdot) : X \times Y \to \mathbb{R}$ is convex. Suppose that, for $i = 1, \ldots, k$, $(x_i, y_i) \in X \times Y$ and $(v_i, w_i) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfy*

$$(v_i, w_i) \in \partial_{\varepsilon_i} \left(\Gamma(\cdot, y_i) - \Gamma(x_i, \cdot) \right)(x_i, y_i). \tag{48}$$

Let $\alpha_1, \ldots, \alpha_k \geq 0$ be such that $\sum_{i=1}^k \alpha_i = 1$, and define

$$(x^a, y^a) = \sum_{i=1}^k \alpha_i (x_i, y_i), \quad (v^a, w^a) = \sum_{i=1}^k \alpha_i (v_i, w_i), \tag{49}$$

$$\varepsilon^a := \sum_{i=1}^k \alpha_i [\varepsilon_i + \langle x_i - x^a, v_i \rangle + \langle y_i - y^a, w_i \rangle]. \tag{50}$$

Then, $\varepsilon^a \geq 0$ *and*

$$(v^a, w^a) \in \partial_{\varepsilon^a} \left(\Gamma(\cdot, y^a) - \Gamma(x^a, \cdot)\right)(x^a, y^a). \tag{51}$$

Proof. By (48), we have

$$\Gamma(x, y_i) - \Gamma(x_i, y) \geq \langle v_i, x - x_i\rangle + \langle w_i, y - y_i\rangle - \varepsilon_i \quad \forall (x, y) \in X \times Y.$$

Using the assumption that $\Gamma(x, \cdot)$ is concave and $\Gamma(\cdot, y)$ is convex for every $(x, y) \in X \times Y$, the assumption that $\sum_{i=1}^k \alpha_i = 1$ and $\alpha_i \geq 0$ for $i = 1, \ldots, k$, and relations (49) and (50), we conclude that

$$\begin{aligned}
\Gamma(x, y^a) - \Gamma(x^a, y) &\geq \sum_{i=1}^k \alpha_i [\Gamma(x, y_i) - \Gamma(x_i, y)] \\
&\geq \sum_{i=1}^k \alpha_i \left(\langle v_i, x - x_i\rangle + \langle w_i, y - y_i\rangle - \varepsilon_i\right) \\
&= \sum_{i=1}^k \alpha_i \left(\langle v_i, x - x^a\rangle + \langle w_i, y - y^a\rangle\right) \\
&\quad - \sum_{i=1}^k \alpha_i \left(\langle v_i, x_i - x^a\rangle + \langle w_i, y_i - y^a\rangle + \varepsilon_i\right) \\
&= \langle v^a, x - x^a\rangle + \langle w^a, y - y^a\rangle - \varepsilon^a
\end{aligned}$$

for every $(x, y) \in X \times Y$. We have thus shown that (51) holds. The nonnegativity of ε^a follows from the above relation with $(x, y) = (x^a, y^a)$. ∎

We are now ready to give a stronger version of Theorem 4.3 based on the notion of SP-residuals.

THEOREM 5.2. *Consider the sequences* $\{(x_k, y_k)\}$ *and* $\{(\tilde{x}_k, \tilde{y}_k)\}$ *generated by either the variant of Tseng's MF-BS method or the generalized Korpelevich's method for* $GSP(\Psi; g_x, g_y)$*, and define*

$$(\tilde{x}_k^a, \tilde{y}_k^a) := \frac{1}{k} \sum_{i=1}^k (\tilde{x}_i, \tilde{y}_i), \qquad \tilde{v}_k^a = (\tilde{v}_{x,k}^a, \tilde{v}_{y,k}^a) := \frac{1}{k\lambda} \left[(x_0, y_0) - (x_k, y_k)\right] \tag{52}$$

and

$$\tilde{\varepsilon}_k^a := \frac{1}{2k\lambda} [2\langle \tilde{x}_k^a - x_0, x_k - x_0\rangle + 2\langle \tilde{y}_k^a - y_0, y_k - y_0\rangle - \|x_k - x_0\|^2 - \|y_k - y_0\|^2]. \tag{53}$$

Then, $(\tilde{v}_k^a, \tilde{\varepsilon}_k^a)$ *is an SP-residual of* $(\tilde{x}_k^a, \tilde{y}_k^a)$ *for* $GSP(\Psi; g_x, g_y)$*, or, equivalently,*

$$\tilde{v}_k^a \in \partial_{\tilde{\varepsilon}_k^a} [\widehat{\Psi}(\cdot, \tilde{y}_k^a) - \widehat{\Psi}(\tilde{x}_k^a, \cdot)](\tilde{x}_k^a, \tilde{y}_k^a)$$

and

$$\|\tilde{v}_k^a\| \leq \frac{2Ld_0}{k\sigma}, \qquad \tilde{\varepsilon}_k^a \leq \frac{2Ld_0^2 \bar{\eta}_k}{k\sigma}, \tag{54}$$

where $\bar{\eta}_k$ *are defined in* (39) *and* d_0 *is the distance of* (x_0, y_0) *to the set of saddle-points of* $GSP(\Psi; g_x, g_y)$*. As a consequence, for every pair of positive scalars* (ρ, ε)*,* *there exists an index*

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

such that, for any $k \geq k_0$*, the pair* $(\tilde{v}_k^a, \tilde{\varepsilon}_k^a)$ *is an easily computable certificate that the point* $(\tilde{x}_k^a, \tilde{y}_k^a)$ *is a* (ρ, ε)*-saddle point of* $GSP(\Psi; g_x, g_y)$*.*

Proof. Clearly, the above two methods are the generalized Korpelevich's method and the variant of Tseng's MF-BS method applied to $HVI(F; g)$, respectively, with F and g given by (30). Hence, all the results derived earlier for these two methods apply here as well. In particular, by Propositions 4.5 and 4.1, we conclude that there exists $\varepsilon_k \geq 0$ such that

$$\tilde{v}_k = \frac{1}{\lambda}(x_{k-1} - x_k, y_{k-1} - y_k) \in (F + \partial_{\varepsilon_k} g)(\tilde{x}_k, \tilde{y}_k), \tag{55}$$

$$\|\lambda \tilde{v}_k + (\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\|^2 + 2\lambda\varepsilon_k \leq \sigma^2 \|(\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\|^2.$$

Hence, using also definition (30), we conclude that

$$\begin{aligned}
\tilde{v}_k \in F(\tilde{x}_k, \tilde{y}_k) + \partial_{\varepsilon_k} g(\tilde{x}_k, \tilde{y}_k) &= (\nabla_x \Psi(\tilde{x}_k, \tilde{y}_k), -\nabla_y \Psi(\tilde{x}_k, \tilde{y}_k)) + \partial_{\varepsilon_k} g(\tilde{x}_k, \tilde{y}_k) \\
&\subseteq \partial[\Psi|_{X\times Y}(\cdot, \tilde{y}_k) - \Psi|_{X\times Y}(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k) + \partial_{\varepsilon_k} g(\tilde{x}_k, \tilde{y}_k) \\
&\subseteq \partial_{\varepsilon_k}[\widehat{\Psi}(\cdot, \tilde{y}_k) - \widehat{\Psi}(\tilde{x}_k, \cdot)](\tilde{x}_k, \tilde{y}_k),
\end{aligned}$$

where the latter identity follows from the definition of $\widehat{\Psi}$ in (28). Hence, it follows from (12), (52), (55), and Proposition 5.1 that

$$\tilde{v}_k^a = (\tilde{v}_{x,k}^a, \tilde{v}_{y,k}^a) \in \partial_{\tilde{\varepsilon}_k^a}[\widehat{\Psi}(\cdot, \tilde{y}_k^a) - \widehat{\Psi}(\tilde{x}_k^a, \cdot)](\tilde{x}_k^a, \tilde{y}_k^a),$$

or, equivalently, $\theta^{sp}((\tilde{x}_k^a, \tilde{y}_k^a); \widehat{\Psi}_{\tilde{v}_k^a}, g_x, g_y) \leq \tilde{\varepsilon}_k^a$, in view of Proposition 3.12. Moreover, the bounds (54) follow directly from Theorem 4.3. ∎

An important observation about Theorem 5.2 is that the variant of Tseng's MF-BS method or the generalized Korpelevich's algorithm for $GSP(\Psi; g_x, g_y)$ can be used to solve problems for which the gap function $\theta^{sp}(\cdot; \Psi_v, g_x, g_y)$ cannot be easily evaluated for any perturbed function Ψ_v (see (29)), since the method can be terminated whenever the computable quantities $\tilde{v}_k^a$ and ε_k^a defined in (52) and (53) are sufficiently small. Moreover, this termination criterion does not depend on any knowledge of d_0, which is used only in the theoretical complexity bound for the algorithm.

The last result of this section considers the special case where X and Y are both bounded sets and an explicit bound on the diameter of $X \times Y$ is given.

PROPOSITION 5.3. *Assume that* X, Y *are bounded sets, and let* D_{XY} *denote the diameter of* $X \times Y$*. Then, for every* $k \in \mathbb{N}$*, the point* $(\tilde{x}_k^a, \tilde{y}_k^a)$ *defined in* (52) *is an* $\hat{\varepsilon}_k$*-saddle point of* $GSP(\Psi, g_x, g_y)$*, where*

$$\hat{\varepsilon}_k := D_{XY}\, \|\tilde{v}_k^a\| + \tilde{\varepsilon}_k^a \leq \frac{2Ld_0}{k\sigma}(D_{XY} + d_0 \bar{\eta}_k);$$

$\bar{\eta}_k$*,* $\tilde{v}_k^a$*, and* $\tilde{\varepsilon}_k^a$ *are defined in* (39), (52) *and* (53)*, respectively; and* d_0 *is the distance of* (x_0, y_0) *to the set of saddle points of* $GSP(\Psi; g_x, g_y)$*. As a consequence, for every* $\varepsilon > 0$*, there exists an index*

$$k_0 = \mathcal{O}\left(\frac{Ld_0 D_{XY}}{\varepsilon}\right)$$

such that, for any $k \geq k_0$*, the point* $(\tilde{x}_k^a, \tilde{y}_k^a)$ *is an* ε*-saddle point of* $GSP(\Psi; g_x, g_y)$*.*

Proof. Use Theorem 5.2, Definitions 3.10 and 3.11, and the Cauchy–Schwarz inequality. □

6. Applications to convex optimization problems. In this section, we consider applications of the theory developed in the previous section to the problem

$$\min\{f(x) + h(x) : \mathcal{A}x = b\}, \tag{56}$$

where the following assumptions are made:

(O.1) $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map and $b \in \mathbb{R}^m$;
(O.2) $f, h : \mathbb{R}^n \to [-\infty, \infty]$ are proper closed convex functions;
(O.3) $\mathrm{dom}(g) \subseteq \mathrm{dom}(f)$, and there exists a point $\hat{x} \in \mathrm{ri}(\mathrm{dom}\, g) \cap \mathrm{ri}(\mathrm{dom}\, f)$ such that $\mathcal{A}\hat{x} = b$;
(O.4) the solution set of (56) is nonempty.

We now make some observations. First, under the above assumptions, x^* is an optimal solution if and only if it satisfies the condition

$$0 \in \partial f(x) + \partial h(x) + N_{\mathcal{M}}(x), \tag{57}$$

where $\mathcal{M} := \{x \in \mathbb{R}^n : \mathcal{A}x = b\}$. Second, the above assumptions also guarantee that $\partial f + \partial g + N_{\mathcal{M}}$ is maximal monotone.

Clearly, when f has Lipschitz continuous gradient and the resolvent of the sum of $\partial(h+\delta_{\mathcal{M}})$ is easy to compute, we could apply the methods of section 4 directly to (57) with $F = \nabla f$ and $g = h + \delta_{\mathcal{M}}$. However, for most practical problems, the resolvent of $\partial(h + \delta_{\mathcal{M}})$ is hard to compute, but the resolvent of ∂h can be easily computed. In this section, we will consider specific reformulations of (56) which can be solved by the algorithms of section 4 using only the resolvent of ∂h.

The following result motivates the aforementioned reformulations of (56).

PROPOSITION 6.1. *For a point $x^* \in \mathbb{R}^n$, the following conditions are all equivalent:*

(a) *x^* is a solution of* (57)*;*
(b) *there exist $y^* \in \mathbb{R}^m$ and $s^* \in \partial h(x^*)$ such that $0 \in \partial f(x^*) + \mathcal{A}^* y^* + s^*$ and $\mathcal{A}x^* = b$;*
(c) *$Ax^* = b$, and there exist $w^* \in \mathcal{R}(\mathcal{A}^*)$ and $s^* \in \mathbb{R}^n$ such that $x^* \in \partial h^*(s^*)$ and $0 \in \partial f(x^*) + w^* + s^*$.*

6.1. Dualization approaches with respect to the affine constraint. In this subsection we make the following additional assumption:

(O.5) f is differentiable on a closed convex set $\Omega_x \supseteq \mathrm{dom}(h)$, and ∇f is L-Lipschitz continuous on Ω_x.

Using the characterization of $N_{\mathcal{M}}(x)$ given by

$$N_{\mathcal{M}}(x) = \begin{cases} \{\mathcal{A}^* y : y \in \mathbb{R}^m\} & \text{if } x \in \mathcal{M}, \\ \emptyset & \text{otherwise}, \end{cases}$$

we obtain the following primal-dual reformulation of (57):

$$0 \in \nabla f(x) + \mathcal{A}^* y + \partial h(x), \quad 0 = b - \mathcal{A}x. \tag{58}$$

Given a pair of positive scalars (ρ, ε), we will examine in this subsection the complexity of finding a pair $(x, y) \in \mathrm{dom}\, h \times \mathbb{R}^m$ such that

$$\|Ax - b\| \le \rho, \quad 0 \in \nabla f(x) + \mathcal{A}^* y + \partial h_\varepsilon(x) + B(\rho), \tag{59}$$

or, equivalently, a triple $(x, y, s) \in \mathrm{dom}\, h \times \mathbb{R}^m \times \mathbb{R}^n$ such that

$$\|Ax - b\| \leq \rho, \quad \|\nabla f(x) + \mathcal{A}^* y + s\| \leq \rho, \quad s \in \partial_\varepsilon h(x). \tag{60}$$

Alternatively, we are also interested in the complexity of finding a pair $(x, y) \in \mathrm{dom}\, h \times \mathbb{R}^m$ satisfying

$$\|Ax - b\| \leq \rho, \quad 0 \in \partial_\varepsilon (f + h)(x) + \mathcal{A}^* y + B(\rho). \tag{61}$$

Observe that if (x, y) satisfies (59), then it also satisfies (61).

In this subsection, we view (58) as being equivalent to the HVI problem

$$0 \in F(x, y) + \partial g(x, y), \tag{62}$$

where

$$F(x, y) := \begin{pmatrix} \nabla f(x) + \mathcal{A}^* y \\ b - \mathcal{A} x \end{pmatrix}, \quad g(x, y) = h(x). \tag{63}$$

In order to apply the variant of Tseng's MF-BS (and/or Korpelevich's extragradient) method to the above HVI problem, we need an upper bound on the Lipschitz constant of F. The tighter this bound is, the larger will be the stepsize λ and hence, the smaller the complexity bound. In the following result, we derive such an upper bound.

LEMMA 6.2. *If $G : \Omega \to \mathbb{R}^n$ is monotone and L-Lipschitz continuous, $b \in \mathbb{R}^m$, and $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then the operator*

$$T(x, y) = (G(x) + \mathcal{A}^* y, b - \mathcal{A} x)$$

is monotone and $\tilde{L}$-Lipschitz continuous in $\Omega \times \mathbb{R}^m$, where

$$\tilde{L} := \frac{L + \sqrt{L^2 + 4\|\mathcal{A}\|^2}}{2} \leq L + \|\mathcal{A}\|. \tag{64}$$

Proof. Let $(u, v) = T(x, y) - T(x', y')$. Then,

$$\|u\| \leq L\|x - x'\| + \|\mathcal{A}\|\|y - y'\|, \qquad \|v\| \leq \|\mathcal{A}\|\|x - x'\|.$$

Therefore,

$$\left(\|u\|^2 + \|v\|^2\right)^{\frac{1}{2}} \leq \left\| \begin{bmatrix} L & \|\mathcal{A}\| \\ \|\mathcal{A}\| & 0 \end{bmatrix} \begin{pmatrix} \|x - x'\| \\ \|y - y'\| \end{pmatrix} \right\|.$$

To end the proof, note that $\tilde{L}$ is the spectral radius of the 2×2 matrix on the right-hand side of the above inequality. ∎

Variant of Tseng's MF-BS method for (62)–(63):

(0) Let $(x_0, y_0) \in \Re^n \times \mathbb{R}^m$ and $0 < \sigma < 1$ be given, and set $\lambda = \sigma/\tilde{L}$ and $k = 1$, where $\tilde{L}$ is given by (64).

(1) Compute $x'_{k-1} = P_{\Omega_x}(x_{k-1})$,

$$\tilde{x}_k = (I + \lambda \partial h)^{-1}(x_{k-1} - \lambda(\nabla f(x'_{k-1}) + \mathcal{A}^* y_{k-1})), \quad \tilde{y}_k = y_{k-1} + \lambda(\mathcal{A} x'_{k-1} - b), \tag{65}$$

and

$$x_k = \tilde{x}_k - \lambda(\nabla f(\tilde{x}_k) - \nabla f(x'_{k-1}) + \mathcal{A}^*(\tilde{y}_k - y_{k-1})), \quad y_k = \tilde{y}_k + \lambda \mathcal{A}(\tilde{x}_k - x'_{k-1}). \tag{66}$$

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

end

Due to the definition of the function F in (63), relations (65) and (66) are equivalent to

$$(\tilde{x}_k, \tilde{y}_k) = (I + \lambda \partial g)^{-1} [\, (x_{k-1}, y_{k-1}) - \lambda F \circ P_\Omega (x_{k-1}, y_{k-1}) \,],$$
$$(x_k, y_k) = (\tilde{x}_k, \tilde{y}_k) - \lambda [F(\tilde{x}_k, \tilde{y}_k) - F \circ P_\Omega (x_{k-1}, y_{k-1})],$$

where $\Omega := \Omega_x \times \mathbb{R}^m$. Hence, the above algorithm is Tseng's MF-BS method applied to $HVI(F, g)$ with F and g given by (63). Note that the sequence $\{\tilde{x}_k\}$ remains in $\operatorname{dom} h$, while the sequence $\{x_k\}$ does not have to be in $\operatorname{dom} h$.

THEOREM 6.3. *Consider the sequences* $\{(x_k, y_k)\}$ *and* $\{(\tilde{x}_k, \tilde{y}_k)\}$ *generated by Tseng's MF-BS method for solving* (62)–(63)*, and define for every* $k \in \mathbb{N}$

$$p_k^x = \frac{1}{\lambda} [x_{k-1} - \tilde{x}_k] - [\nabla f(x'_{k-1}) + \mathcal{A}^* y_{k-1}]. \tag{67}$$

Then, for every $k \in \mathbb{N}$, $p_k^x \in \partial h(\tilde{x}_k)$ *and there exists* $i \le k$ *such that*

$$\left\| \begin{pmatrix} \nabla f(\tilde{x}_i) + \mathcal{A}^* \tilde{y}_i + p_i^x \\ b - \mathcal{A} \tilde{x}_i \end{pmatrix} \right\| \le \frac{(L + \|\mathcal{A}\|) d_0}{\sigma} \sqrt{\frac{1+\sigma}{k(1-\sigma)}},$$

where d_0 *is the distance of* (x_0, y_0) *to the solution set of* (58)*. As a consequence, for any* $\rho > 0$*, there exists an index*

$$k = \mathcal{O}\left(\frac{(L + \|\mathcal{A}\|)^2 d_0^2}{\rho^2} \right)$$

such that the triple $(x, y, s) = (\tilde{x}_k, \tilde{y}_k, p_k^x)$ *satisfies* (60) *with* $\varepsilon = 0$.

Proof. This result follows immediately from Theorem 4.6 and the fact that for F and g as in (63) and $B = \partial g$, the vector b_k in (46) is equal to the vector $(p_k^x, 0)$. ∎

THEOREM 6.4. *Consider the sequences* $\{(x_k, y_k)\}$ *and* $\{(\tilde{x}_k, \tilde{y}_k)\}$ *generated by Tseng's MF-BS method for solving* (62)–(63)*. Define the sequences* $\{(\tilde{x}_k^a, \tilde{y}_k^a)\}$, $\{\tilde{\varepsilon}_k^a\}$, *and* $\tilde{v}_k^a = \{(\tilde{v}_{x,k}^a, \tilde{v}_{y,k}^a)\}$ *as in Theorem* 5.2*. Then, for every* $k \in \mathbb{N}$,

$$\tilde{v}_{x,k}^a \in \partial_{\tilde{\varepsilon}_k^a} (f + h)(\tilde{x}_k) + \mathcal{A}^* \tilde{y}_k^a, \qquad \tilde{v}_{y,k}^a = b - \mathcal{A} \tilde{x}_k^a$$

and

$$\|\tilde{v}_k^a\| \le \frac{2(L + \|\mathcal{A}\|) d_0}{k\sigma}, \qquad \tilde{\varepsilon}_k^a \le \frac{2(L + \|\mathcal{A}\|) d_0^2 \bar{\eta}_k}{k\sigma},$$

where $\bar{\eta}_k$ *is given by* (39) *and* d_0 *is the distance of* (x_0, y_0) *to the solution set of* (58)*. As a consequence, for any pair of positive scalars* (ρ, ε)*, there exists an index*

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

such that $(\tilde{x}_k^a, \tilde{y}_k^a)$ *satisfies* (61) *for any* $k \ge k_0$.

Proof. Define $g_x := h$, $g_y \equiv 0$, and $\Psi : \operatorname{dom} f \times \Re^m \to \Re$ as

$$\Psi(x, y) := f(x) + \langle y, \mathcal{A}x - b \rangle, \qquad (x, y) \in \operatorname{dom} f \times \mathbb{R}^m.$$

Clearly, the above algorithm corresponds to the variant of Tseng's MF-BS method for $GSP(\Psi; g_x, g_y)$ with Ω_x as in assumption (O.5) and $\Omega_y = \mathbb{R}^m$. The result now follows from Theorem 5.2 and elementary rules of subdifferential calculus. ∎

PROPOSITION 6.5. *Let* $(\rho, \varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, *and set* $\bar{c} := \rho^2/(2\varepsilon)$. *Consider the sequence* $\{(\tilde{x}_k^a, \tilde{y}_k^a)\}$ *of ergodic iterates defined in* (52), *where* $\{(\tilde{x}_k, \tilde{y}_k)\}$ *is the sequence generated by Tseng's MF-BS method for solving* (62)–(63). *Moreover, for every* $k \in \mathbb{N}$, *define*

$$u_k := \tilde{x}^a - \frac{1}{\bar{c}}\left[\nabla f(\tilde{x}_k^a) + \mathcal{A}^* \tilde{y}_k^a\right], \qquad q_k^x := \bar{c}\left[u_k - \left(I + \frac{1}{\bar{c}}\partial h\right)^{-1}(u_k)\right].$$

Then, there exists an index

$$k_0 = \mathcal{O}\left(\max\left[\frac{(L+\|\mathcal{A}\|)d_0^2}{\varepsilon}, \frac{(L+\|\mathcal{A}\|)d_0}{\rho} + \frac{d_0^2(L+\|\mathcal{A}\|)\,\mathcal{N}(\nabla f; \operatorname{dom} h)}{\rho^2}\right]\right)$$

such that, for any $k \geq k_0$, *the triple* $(x, y, s) = (\tilde{x}_k^a, \tilde{y}_k^a, q_k^x) \in \operatorname{dom} f \times \mathbb{R}^m \times \mathbb{R}^n$ *satisfies* (60), *where* d_0 *is the distance of* (x_0, y_0) *to the solution set of* (58).

Proof. This result follows immediately from (22), Definition 3.4, and Theorem 4.4 applied to $HVI(F, g)$ with F and g given by (63). □

6.2. Dualization approach with respect to *h*. In this subsection, in addition to assumptions (O.1)–(O.4), we further assume that

$(O.5')$ f is differentiable on $\mathcal{M}$ and ∇f is L-Lipschitz continuous on $\mathcal{M}$.

Using the fact that $s \in \partial h(x)$ if and only if $x \in \partial h^*(s)$, it follows that (57) is equivalent to

$$0 \in \nabla f(x) + s + N_{\mathcal{M}}(x), \qquad 0 \in -x + \partial h^*(s). \tag{68}$$

Given a pair of positive scalars (ρ, ε), we will examine in this subsection the complexity of finding a pair $(x, s) \in \mathcal{M} \times \mathbb{R}^n$ such that

$$x \in \partial_\varepsilon h^*(s) + B(\rho), \qquad 0 \in \nabla f(x) + s + N_{\mathcal{M}}(x) + B(\rho), \tag{69}$$

or equivalently, a quadruple $(x, x', w, s) \in \mathcal{M} \times \mathbb{R}^n \times \mathcal{R}(\mathcal{A}^*) \times \mathbb{R}^n$ such that

$$x' \in \partial_\varepsilon h^*(s), \qquad \|x' - x\| \leq \rho, \qquad \|\nabla f(x) + w + s\| \leq \rho. \tag{70}$$

Alternatively, we are also interested in the complexity of finding a pair $(x, s) \in \mathcal{M} \times \Re^n$ satisfying

$$x \in \partial_\varepsilon h^*(s) + B(\rho), \qquad 0 \in \partial_\varepsilon f(x) + s + \mathcal{R}(\mathcal{A}^*) + B(\rho). \tag{71}$$

Observe that if (x, s) satisfies (69), then it also satisfies (71). Observe also that for any pair (x, s) satisfying the above conditions, we must have $s \in \operatorname{dom} h^*$.

In this subsection, we view (68) as being equivalent to the HVI problem

$$0 \in F(x, s) + \partial g(x, y), \tag{72}$$

where

$$F(x, s) := \begin{pmatrix} \nabla f(x) + s \\ -x \end{pmatrix}, \qquad g(x, s) = \delta_{\mathcal{M}}(x) + h^*(s). \tag{73}$$

The following result follows as an immediate consequence of Lemma 6.2.

PROPOSITION 6.6. *The map* F *defined as above is* $(L+1)$*-Lipschitz continuous.*

Variant of Tseng's MF-BS method for (72)–(73):

(0) Let $(x_0, s_0) \in \Re^n \times \Re^n$ and $0 < \sigma < 1$ be given, and set $\lambda = \sigma/(L+1)$ and $k = 1$.

(1) Compute $x'_{k-1} = P_{\mathcal{M}}(x_{k-1})$,

(74)
$$\tilde{x}_k = P_{\mathcal{M}}\left(x_{k-1} - \lambda(\nabla f(x'_{k-1}) + s_{k-1})\right), \quad \tilde{s}_k = (I+\lambda\partial h^*)^{-1}(s_{k-1}+\lambda x'_{k-1}),$$

and

$$(75)\quad x_k = \tilde{x}_k - \lambda[\nabla f(\tilde{x}_k) - \nabla f(x'_{k-1}) + \tilde{s}_k - s_{k-1}]. \quad s_k = \tilde{s}_k + \lambda(\tilde{x}_k - x'_{k-1}).$$

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

end

We now make a few observations regarding the above algorithm. First, we note that the sequence $\{(\tilde{x}_k, \tilde{s}_k)\}$ is in $\mathcal{M} \times \operatorname{dom} h^*$. Second, due to the definition of the function F and g in (73), relations (74) and (75) are equivalent to

$$(\tilde{x}_k, \tilde{s}_k) = (I + \lambda\partial g)^{-1}(\,(x_{k-1}, s_{k-1}) - \lambda F \circ P_\Omega(x_{k-1}, s_{k-1})\,),$$
$$(x_k, s_k) = (\tilde{x}_k, \tilde{s}_k) - \lambda[F(\tilde{x}_k, \tilde{s}_k) - F \circ P_\Omega(x_{k-1}, s_{k-1})\,],$$

where $\Omega := \mathcal{M} \times \Re^n$. Hence, the above algorithm is Tseng's MF-BS method applied to the inclusion problem (72)–(73). Third, observe that the above method requires two projections onto $\mathcal{M}$ and exactly one evaluation of the resolvent of h^* per iteration. Fourth, we also observe that the resolvent of h^* can also be computed using the resolvent of h according to

$$(I + \lambda\partial h^*)^{-1}(x) = x - \lambda[I + \lambda^{-1}\partial h]^{-1}(\lambda^{-1}x) \qquad \forall x \in \mathbb{R}^n.$$

The following result follows as an immediate consequence of Theorem 4.2.

THEOREM 6.7. *Consider the sequences $\{(x_k, s_k)\}$, $\{(\tilde{x}_k, \tilde{s}_k)\}$, and $\{x'_k\}$ generated by Tseng's MF-BS method for solving* (72)–(73)*, and define for every $k \in \mathbb{N}$*

$$(76)\qquad p_k^x = \frac{1}{\lambda}[x_{k-1} - \tilde{x}_k] - [\nabla f(x'_{k-1}) + s_{k-1}], \qquad p_k^s = \frac{1}{\lambda}[s_{k-1} - \tilde{s}_k] + x'_{k-1}.$$

Then, for every $k \in \mathbb{N}$, $\tilde{x}_k \in \mathcal{M}$ and $(p_k^x, p_k^s) \in \mathcal{R}(\mathcal{A}^*) \times \partial h^*(\tilde{s}_k)$, and there exists $i \le k$ such that

$$\left\| \begin{pmatrix} \nabla f(\tilde{x}_i) + \tilde{s}_i + p_i^x \\ -\tilde{x}_i + p_i^s \end{pmatrix} \right\| \le \frac{(L+1)d_0}{\sigma}\sqrt{\frac{1+\sigma}{k(1-\sigma)}},$$

where d_0 is the distance of (x_0, s_0) to the solution set of (68)*. As a consequence, for any $\rho > 0$, there exists an index*

$$k = \mathcal{O}\left(\frac{(L+1)^2 d_0^2}{\rho^2}\right)$$

such that the quadruple $(x, x', w, s) = (\tilde{x}_k, p_k^s, p_k^x, \tilde{s}_k)$ satisfies (70) *with $\varepsilon = 0$.*

THEOREM 6.8. *Consider the sequences $\{(x_k, s_k)\}$ and $\{(\tilde{x}_k, \tilde{s}_k)\}$ generated by Tseng's MF-BS method for solving* (72)–(73)*. Define the sequences $\{(\tilde{x}_k^a, \tilde{s}_k^a)\}$, $\{\tilde{\varepsilon}_k^a\}$, and $\tilde{v}_k^a = \{(\tilde{v}_{x,k}^a, \tilde{v}_{s,k}^a)\}$ as in Theorem 5.2 with y_k and $\tilde{y}_k$ replaced by s_k and $\tilde{s}_k$, respectively. Then, for every $k \in \mathbb{N}$, $\tilde{x}_k^a \in \mathcal{M}$,*

$$\tilde{v}_{x,k}^a \in \partial_{\tilde{\varepsilon}_k^a} f(\tilde{x}_k^a) + \tilde{s}_k^a + \mathcal{R}(\mathcal{A}^*), \qquad \tilde{v}_{s,k}^a \in -\tilde{x}_k^a + \partial_{\tilde{\varepsilon}_k^a} h^*(\tilde{s}_k^a),$$

and

$$\|\tilde{v}_k^a\| \le \frac{2(L+1)d_0}{k\sigma}, \qquad \tilde{\varepsilon}_k^a \le \frac{2(L+1)d_0^2\bar{\eta}_k}{k\sigma},$$

where $\bar{\eta}_k$ *is given by* (39) *and* d_0 *is the distance of* (x_0, s_0) *to the solution set of* (68). *As a consequence, for any pair of positive scalars* (ρ, ε)*, there exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{(L+1)d_0^2}{\varepsilon}, \frac{(L+1)d_0}{\rho}\right]\right)$$

such that $(\tilde{x}_k^a, \tilde{s}_k^a)$ *satisfies* (71) *for any* $k \ge k_0$.

Proof. Define $g_x := \delta_{\mathcal{M}}$, $g_s := h^*$, and $\Psi : \operatorname{dom} f \times \mathbb{R}^n \to \mathbb{R}$ as $\Psi(x,s) = f(x) + \langle x, s\rangle$. Clearly, the above algorithm corresponds to the variant of Tseng's MF-BS method for $GSP(\Psi; g_x, g_y)$. The result now follows from Theorem 5.2 and elementary rules of subdifferential calculus. ∎

PROPOSITION 6.9. *Let* $(\rho, \varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$*, and set* $\bar{c} := \rho^2/(2\varepsilon)$*. Consider the sequence of ergodic iterates* $\{(\tilde{x}_k^a, \tilde{s}_k^a\}$ *generated by the variant of Tseng's MF-BS method for solving* (72)–(73)*. Moreover, for every* $k \in \mathbb{N}$*, define*

$$q_k^x := -P_{\mathcal{R}(\mathcal{A}^*)}(\nabla f(\tilde{x}_k^a) + \tilde{s}_k^a), \qquad q_k^s := (\bar{c}\tilde{s}_k^a + \tilde{x}_k^a) - \bar{c}\left(I + \frac{1}{\bar{c}}\partial h^*\right)^{-1}\left(\tilde{s}_k^a + \frac{1}{\bar{c}}\tilde{x}_k^a\right).$$

Let d_0 *be the distance of* (x_0, s_0) *to the solution set of* (68)*. Then, there exists an index*

$$k_0 = \mathcal{O}\left(\max\left[\frac{(L+1)d_0^2}{\varepsilon}, \frac{(L+1)d_0}{\rho} + \frac{d_0^2(L+1)\mathcal{N}(\nabla f; \mathcal{M})}{\rho^2}\right]\right)$$

such that, for any $k \ge k_0$*, the quadruple* $(x, x', w, s) = (\tilde{x}_k^a, q_k^s, q_k^x, \tilde{s}_k^a)$ *satisfies* (70).

6.3. Examples. In this subsection, we give two specific instances of optimization problems which can be solved by the methods discussed in the previous two subsections.

Consider first the case where $h = \delta_X$ for some closed convex set in $\mathbb{R}^n$. In this case, (56) becomes

$$\min\{f(x) : \mathcal{A}x = b,\, x \in X\}.$$

Note that in this case the resolvent $(I + \lambda\partial h)^{-1}$, which needs to be evaluated at each step of the methods described in subsections 6.1 and 6.2, reduces to the projection map with respect to X for any $\lambda > 0$. Moreover, when X is a cone, the termination criterion (60) reduces to

$$\begin{aligned} &\|\mathcal{A}x - b\| \le \rho, \quad \|\nabla f(x) + \mathcal{A}^*y + s\| \le \rho, \\ &x \in X, \quad -s \in X^*, \quad \langle x, -s\rangle \le \varepsilon, \end{aligned}$$

where $X^* := \{s \in \mathbb{R}^n : \langle x, s\rangle \ge 0\, \forall x \in X\}$ is the dual cone of X.

We now consider the second case, where h is given by

$$h(x) = \begin{cases} -\sum_{i=1}^n \log x_i & \text{if } x > 0, \\ \infty & \text{otherwise.} \end{cases}$$

In this case, it is easy to see that evaluation of the resolvent of ∂h amounts to solving n single-variable quadratic equations. Moreover, it is easy to verify that the termination criteria (60) reduces to

$$\|\mathcal{A}x - b\| \le \rho, \quad \|\nabla f(x) + \mathcal{A}^* y + s\| \le \rho,$$
$$x > 0, \quad -s > 0, \quad \langle x, -s\rangle - \sum_{i=1}^{n} \log(-x_i s_i) + n \le \varepsilon.$$

It should be noted that the latter condition is approximately enforcing the condition $-x_i s_i = 1$ for every $i = 1, \dots, n$ and that it is a well-known interiority condition in the theory of interior-point methods.

We also note that we can apply the latter idea to the case when $\mathbb{R}^n$ is the set of $p \times p$ symmetric matrices (and hence $n = p(+1)/2$), $h(X) = -\log\det X$ when X is positive definite, and $+\infty$ otherwise. In this case, evaluation of the resolvent of h amounts to computing a symmetric eigenvalue decomposition and the solution of p single-variable quadratic equations.

Appendix A. Maximal monotonicity of the HVI problem.

PROPOSITION A.1. *If $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone and $F : \mathrm{Dom}(F) \to \mathbb{R}^n$ is a map such that $\mathrm{Dom}\, F \supseteq \mathrm{cl}(\mathrm{Dom}\, B)$ and F restricted to $\mathrm{cl}(\mathrm{Dom}\, B)$ is monotone and continuous, then $F + B$ is maximal monotone.*

Proof. Let $C = \mathrm{cl}(\mathrm{Dom}\, B)$. Since $\overline{F} := F|_C$ is monotone and C is convex (see Proposition 6.4.1 of [1]), $\overline{F} + N_C$ is maximal monotone (see, for example, Proposition 12.3.6 of [4]). Using the fact that the sum of two maximal monotone operators is also maximal monotone as long as the relative interiors of their domains intersect (see Theorem 2 of [17]) and the relative interiors of the domains of B and $\overline{F} + N_C$ are identical, then it follows that $\overline{F} + N_C + B$ is maximal monotone. Moreover, we clearly have that $B \subseteq N_C + B$. Since B is maximal monotone, we actually have that $B = N_C + B$, and hence $\overline{F} + N_C + B = \overline{F} + B = F + B$. We have thus shown that $F + B$ is maximal monotone. ∎

Appendix B. Proof of Proposition 4.1.

Proof. First observe that, in view of the definitions of x_k and q_k in (32) and (33), respectively, we have $q_k \in \partial g(x_k)$. Hence, it follows from Proposition 2.1(c) and the definition of ε_k in (34) that the first inclusion in Proposition 4.1(a) holds. Letting $\bar{F} := F|_{\mathrm{cl}(\mathrm{dom}\, g)}$, the latter inclusion and the definition of $\tilde{v}_k$ in (34) then imply that

$$\begin{aligned}\tilde{v}_k = F(\tilde{x}_k) + q_k &\in [\bar{F} + \partial_{\varepsilon_k} g](\tilde{x}_k) \subseteq [(\bar{F})^0 + (\partial g)^{\varepsilon_k}](\tilde{x}_k)\\ &\subseteq [\bar{F} + \partial g]^{\varepsilon_k}(\tilde{x}_k) = [F + \partial g]^{\varepsilon_k}(\tilde{x}_k),\end{aligned}$$

where the two last inclusions follow from parts (a) and (b) of Proposition 2.2 and Proposition 2.1(a). We have thus shown that Proposition 4.1(a) holds.

Statement (b) of Proposition 4.1 follows immediately from the definition of q_k and $\tilde{v}_k$ in (33) and (34), respectively.

We now show (c) and (d). First note that the definitions of $\tilde{x}_k$ and p_k in (32) and (33), respectively, imply that $p_k \in \partial g(\tilde{x}_k)$. This fact, together with the definition of ε_k in (33), yields the estimate

$$\varepsilon_k = -[g(x_k) - g(\tilde{x}_k) - \langle p_k, x_k - \tilde{x}_k\rangle] + \langle p_k - q_k, \tilde{x}_k - x_k\rangle \le \langle p_k - q_k, \tilde{x}_k - x_k\rangle,$$

which, together with statement (b), then implies that

$$\begin{aligned}\|\lambda\tilde{v}_k + \tilde{x}_k - x_{k-1}\|^2 + 2\lambda\varepsilon_k &= \|\tilde{x}_k - x_k\|^2 + 2\lambda\varepsilon_k \le \|\tilde{x}_k - x_k\|^2 + 2\lambda\langle p_k - q_k, \tilde{x}_k - x_k\rangle \\ &= \|\lambda(p_k - q_k) + \tilde{x}_k - x_k\|^2 - \lambda^2\|p_k - q_k\|^2 \\ &\le \|\lambda(p_k - q_k) + \tilde{x}_k - x_k\|^2 \\ &= \|\lambda(F(x_{k-1}) - F(\tilde{x}_k))\|^2 \le (\lambda L\|x_{k-1} - \tilde{x}_k\|) \\ &\le \sigma^2\|x_{k-1} - \tilde{x}_k\|^2,\end{aligned}$$

where the last equality follows from the definitions of p_k and q_k in (33), and the last two inequalities are due to assumption (K.3) and the assumption that $\lambda \le \sigma/L$. It remains to show that (35) holds. Indeed, the definition of p_k in (33), the triangle inequality for norms, assumption (K.3), and the identity $\lambda = \sigma/L$ imply that

$$\begin{aligned}\|F(\tilde{x}_k) + p_k\| &= \left\|F(\tilde{x}_k) - F(x_{k-1}) + \frac{1}{\lambda}(x_{k-1} - \tilde{x}_k)\right\| \\ &\le \|F(\tilde{x}_k) - F(x_{k-1}\| + \frac{1}{\lambda}\|x_{k-1} - \tilde{x}_k\| \\ &\le \left(L + \frac{1}{\lambda}\right)\|x_{k-1} - \tilde{x}_k\| = \frac{(1+\sigma)L}{\sigma}\|\tilde{x}_k - x_{k-1}\|. \qquad \square\end{aligned}$$

Appendix C. Proofs of the results in subsection 3.1.

In this section we assume that assumptions (K.1)–(K.3) of section 4.1 hold. The proof of Proposition 3.5 will be divided into three parts.

PROPOSITION C.1. *Let $x \in \operatorname{dom} g$. Then*

(a) *(r, ε) is a strong residual of x for $HVI(F, g)$ if and only if $r \in F(x) + \partial_\varepsilon g(x)$;*

(b) *if (r, ε) is a weak residual of x for $HVI(F, g)$, then $r \in (F + \partial g)^\varepsilon(x)$;*

(c) *if $r \in (F^{\varepsilon'} + \partial g_{\varepsilon''})(x)$ and $\varepsilon' + \varepsilon'' \le \varepsilon$, then (r, ε) is a weak residual of x for $HVI(F, g)$.*

Proof. (a) Using the definition (1) of the ε-subdifferential and relations (19) and (21), we conclude that (r, ε) is a strong residual of x for $HVI(F, g)$ if and only if $r - F(x) \in \partial_\varepsilon g(x)$.

(b) Suppose that (r, ε) is a weak residual of x for $HVI(F, g)$. Then, using (21), (20), and the fact that the domain of ∂g is contained in $\operatorname{dom} g$, we conclude that, for every $y \in \operatorname{Dom} \partial g$ and $u \in \partial g(y)$,

$$\begin{aligned}\varepsilon \;&\ge\; g(x) - g(y) + \langle x - y, F(y) - r\rangle \\ &= g(x) - g(y) - \langle x - y, u\rangle + \langle x - y, F(y) + u - r\rangle \ge \langle x - y, F(y) + u - r\rangle.\end{aligned}$$

Hence, it follows from definition (3) that $r \in (F + \partial g)^\varepsilon(x)$.

(c) Suppose now that $r \in (F^{\varepsilon'} + \partial g_{\varepsilon''})(x)$ and $\varepsilon' + \varepsilon'' \le \varepsilon$. Hence, there exist $r' \in F^{\varepsilon'}(x)$ and $r'' \in \partial_{\varepsilon''} g(x)$ such that $r' + r'' = r$. Hence, for any $y \in \operatorname{dom}(g) \subseteq \operatorname{Dom} F$, we have

$$\langle y - x, F(y) - r'\rangle \ge -\varepsilon', \quad g(y) \ge g(x) + \langle y - x, r''\rangle - \varepsilon''.$$

Adding these two inequalities and using the fact that $r = r' + r''$ and $\varepsilon' + \varepsilon'' \le \varepsilon$, we then conclude that

$$g(y) \ge g(x) + \langle y - x, r - F(x)\rangle - \varepsilon \quad \forall y \in \operatorname{dom} g$$

and hence that (r, ε) is a weak residual of x for $HVI(F, g)$, due to (20) and (21). $\square$

PROPOSITION C.2. *If $x \in \operatorname{dom} g$ and $c \geq 2L$, then $\theta_c(x;F,g) \leq \theta^w(x;F,g)$.*

Proof. Let $x \in \operatorname{dom} g$ and $c \geq 2L$ be given. Then, in view of (20), (23), the assumption that $c \geq 2L$, and assumption (K.3), we have

$$\begin{aligned}\theta^w(x;F,g) &= \sup_{y\in\operatorname{dom} g} g(x)-g(y)+\langle x-y,F(y)\rangle \\ &\geq \sup_{y\in\operatorname{dom} g} g(x)-g(y)-\langle y-x,F(x)\rangle-\|F(x)-F(y)\|\,\|y-x\| \\ &\geq \sup_{y\in\mathbb{R}^n} g(x)-g(y)-\langle y-x,F(x)\rangle-L\|y-x\|^2 \geq \theta_c(x;F,g). \qquad \square\end{aligned}$$

The following result establishes a relationship between θ_c and strong residuals.

THEOREM C.3. *Let $x \in \operatorname{dom} g$ and $c > 0$ be given. If (r,ε) is a strong residual of x for $HVI(F,g)$, then, for any $c > 0$,*

$$\theta_c(x;F,g) \leq \frac{1}{2c}\|r\|^2+\varepsilon; \tag{77}$$

moreover, for any fixed $c > 0$, there exists a unique strong residual (r,ε) of x for $HVI(F,g)$ for which equality holds in (77), namely, $(r,\varepsilon) = (r_c(x;F,g),\varepsilon_c(x;F,g))$, where

$$\varepsilon_c(x;F,g) := g(x)-g(y_c)-\langle x-y_c,r_c-F(x)\rangle \geq 0, \qquad y_c := x-c^{-1}r_c(x;F,g). \tag{78}$$

Proof. To simplify notation, let $(r_c,\varepsilon_c) := (r_c(x;F,g),\varepsilon_c(x;F,g))$. By Definition 3.4, (r,ε) is a strong residual of x for $HVI(F,g)$ if and only if $\theta^s(x;F-r,g) := \sup_y g(x)-g(y)+\langle x-y,F(x)-r\rangle \leq \varepsilon$. This observation together with definition (23) of $\theta_c(\cdot;F,g)$ then implies that

$$\begin{aligned}&\inf\left\{\varepsilon+\frac{1}{2c}\|r\|^2 : (r,\varepsilon) \text{ strong residual of } x \text{ for } HVI(F,g)\right\}\\ &= \inf_r \sup_y g(x)-g(y)+\langle x-y,F(x)-r\rangle+\frac{1}{2c}\|r\|^2\\ &\geq \sup_y \inf_r g(x)-g(y)+\langle x-y,F(x)-r\rangle+\frac{1}{2c}\|r\|^2\\ &= \sup_y g(x)-g(y)+\langle x-y,F(x)\rangle-\frac{c}{2}\|x-y\|^2 = \theta_c(x;F,g),\end{aligned}$$

which proves the first claim of the theorem.

We now prove the second claim. Using (78) and (22) we conclude that

$$y_c = \left(I+\frac{1}{c}\partial g\right)^{-1}\left(x-\frac{1}{c}F(x)\right).$$

Therefore, from the optimality conditions for the maximization problem (23), we conclude that its maximizer is y_c and

$$\theta_c(x;F,g) = g(x)-g(y_c)+\langle x-y_c,F(x)\rangle-\frac{c}{2}\|y_c-x\|^2, \tag{79}$$

$$r_c-F(x) = c(x-y_c)-F(x) \in \partial g(y_c). \tag{80}$$

Moreover, (80) and Proposition 2.1(c) imply that $\varepsilon_c \geq 0$ and $r_c-F(x) \in \partial_{\varepsilon_c} g(x)$. Hence, in view of Proposition C.1(a), (r_c,ε_c) is a strong residual of x for $HVI(F,g)$.

To end the proof, use (78) and (79) to conclude that (77) holds as an equality for (r_c, ε_c). ◻

The following result follows as a consequence of Theorem C.3.

PROPOSITION C.4. *If $x \in \operatorname{dom} g$ is a $(\rho/\sqrt{2}, \varepsilon/2)$-strong solution of $HVI(F, g)$ and $\bar{c} := \rho^2/(2\varepsilon)$, then the pair $(r_{\bar{c}}(x; F, g), \varepsilon_{\bar{c}}(x; F, g))$ is a strong residual of x for $HVI(F, g)$ satisfying the estimates*

$$\|r_{\bar{c}}(x; F, g)\| \leq \rho, \qquad \varepsilon_{\bar{c}}(x; F, g) \leq \varepsilon.$$

Proof. To simplify notation, denote $(r_{\bar{c}}(x; F, g), \varepsilon_{\bar{c}}(x; F, g))$ simply by $(r_{\bar{c}}, \varepsilon_{\bar{c}})$. Since x is a $(\rho/\sqrt{2}, \varepsilon/2)$-strong solution of $HVI(F, g)$, there exists $r \in \mathbb{R}^n$ such that $\|r\| \leq \rho/\sqrt{2}$ and $(r, \varepsilon/2)$ is a strong residual of x for $HVI(F, g)$. Hence, it follows from Theorem C.3 with $c = \bar{c} := \rho^2/(2\varepsilon)$ that

$$\frac{1}{2\bar{c}}\|r_{\bar{c}}\|^2 + \varepsilon_{\bar{c}} = \theta_{\bar{c}}(x; F, g) \leq \frac{1}{2\bar{c}}\|r\|^2 + \frac{\varepsilon}{2} \leq \frac{\rho^2}{4\bar{c}} + \frac{\varepsilon}{2} = \varepsilon,$$

which clearly implies that $\varepsilon_{\bar{c}} \leq \varepsilon$ and $\|r_{\bar{c}}\| \leq \sqrt{2\bar{c}\,\varepsilon} = \rho$. ◻

The following result also follows as a consequence of Theorem C.3.

PROPOSITION C.5. *If condition* (K.3) *holds and (r, ε) is a weak residual of x for $HVI(F, g)$, then, for any positive scalar $c \geq 2L$, the vector*

$$r_c := r_c(x; F - r, g) \tag{81}$$

satisfies $\|r_c\| \leq \sqrt{2c\varepsilon}$, and the pair $(r + r_c, \varepsilon)$ is a strong residual of x for $HVI(F, g)$.

Proof. Assume that $c \geq 2L$ is given. By definition, the second assumption of the proposition means that $\theta^w(x; F - r, g) \leq \varepsilon$. This together with condition (K.3) and Proposition C.2 implies that $\theta_c(x; F - r, g) \leq \varepsilon$. Hence, considering the pair (r_c, ε_c) with r_c defined by (81) and $\varepsilon_c := \varepsilon_c(x; F - r, g)$ (see (78)), it follows from Theorem C.3 that (r_c, ε_c) is a strong residual of x for $HVI(x; F - r, g)$ and

$$\frac{1}{2c}\|r_c\|^2 + \varepsilon_c = \theta_c(x; F - r, g) \leq \varepsilon,$$

from which the conclusion of the proposition immediately follows. ◻

The following result shows that near a point $x \in \operatorname{dom} g$, it is always possible to construct a point $y(x)$ with strong residual $(r, \varepsilon) = (r, 0)$ such that $\|y(x) - x\| \leq \|r_c(x; F, g)\|/c$ and $\|r\| \leq \|r_c\|(1 + L/c)$.

PROPOSITION C.6. *Assume that $F : \operatorname{Dom} F \subseteq \mathbb{R}^n$ satisfies assumption* (K.3), *and let $x \in \operatorname{dom} g$ be given. Then, for every $c > 0$, the vectors y_c defined in* (78) *and*

$$\hat{r}_c := F(y_c) - F(x) + r_c \tag{82}$$

satisfy

$$\|y_c - x\| = \frac{1}{c}\|r_c\|, \quad \hat{r}_c \in (F + \partial g)(y_c), \quad \|\hat{r}_c\| \leq \left(1 + \frac{L}{c}\right)\|r_c\|, \tag{83}$$

where $r_c := r_c(x; F, g)$. Moreover, if, in addition, x is a (ρ, ε)-strong solution, then

$$\|y_c - x\| \leq \frac{\sqrt{\rho^2 + 2c\varepsilon}}{c}, \quad \|\hat{r}_c\| \leq \left(1 + \frac{L}{c}\right)\sqrt{\rho^2 + 2c\varepsilon} \quad \forall c > 0.$$

Proof. The equality in (83) follows immediately from the definition of y_c in (78), and the inclusion in (83) follows from (80) and (82). Moreover, using (82) and assumption (K.3), we have

$$\|\hat{r}_c\| \le \|F(y_c) - F(x)\| + \|r_c\| \le L\|y_c - x\| + \|r_c\| \le \left(1 + \frac{L}{c}\right)\|r_c\|.$$

The last claim of Proposition C.6 follows immediately from the first one and Proposition C.4. □

Appendix D. Proof of Theorem 3.9.

In this appendix, we provide the proof of Theorem 3.9. We first state the following well-known technical result.

PROPOSITION D.1. *Assume that $g : \Re^n \to [-\infty, \infty]$ is a proper closed convex function. Let $x_i, v_i \in \mathbb{R}^n$ and $\varepsilon_i, \alpha_i \in \mathbb{R}_+$, for $i = 1, \ldots, k$, be such that*

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

and define

$$x^a := \sum_{i=1}^{k} \alpha_i x_i, \qquad v^a := \sum_{i=1}^{k} \alpha_i v_i,$$

$$\varepsilon^a := \sum_{i=1}^{k} \alpha_i [\varepsilon_i + \langle x_i - x^a, v_i - v^a \rangle] = \sum_{i=1}^{k} \alpha_i [\varepsilon_i + \langle x_i - x^a, v_i \rangle].$$

Then, $\varepsilon^a \ge 0$ and $v^a \in \partial_{\varepsilon^a} g(x^a)$.

We are now ready to give the proof of Theorem 3.9.

Proof of Theorem 3.9. (a) Note that, by (25), we have $v_i - F(x_i) \in \partial_{\varepsilon_i} g(x_i)$ for every $i = 1, \ldots, k$. Defining

$$\varepsilon_0^a := \sum_{i=1}^{k} \alpha_i \left[\varepsilon_i + \left\langle x_i - x^a, v_i - F(x_i) \right\rangle\right], \tag{84}$$

it then follows from the definitions of v^a and F^a in (26) and Proposition D.1 that $\varepsilon_0^a \ge 0$ and $v^a - F^a \in \partial_{\varepsilon_0^a} g(x^a)$. Moreover, defining

$$\varepsilon_1^a := \sum_{i=1}^{k} \alpha_i \langle x_i - x^a, F(x_i) - F^a \rangle = \sum_{i=1}^{k} \alpha_i \langle x_i - x^a, F(x_i) \rangle,$$

it follows from Corollary 2.4 of [9] that $\varepsilon_1^a \ge 0$ and $F^a \in F^{\varepsilon_1^a}(x^a)$. Hence, it follows that $v^a \in [F^{\varepsilon_1^a} + \partial_{\varepsilon_0^a} g](x^a)$. Noting that $\varepsilon^a = \varepsilon_0^a + \varepsilon_1^a$, we then conclude from Proposition 3.5(c) that (v^a, ε^a) is a weak residual of x^a for $HVI(F, g)$.

(b) The function F can be decomposed as $F = G + \mathcal{A}$, where G is an $\mathcal{N}(F; \operatorname{dom} g)$-Lipschitz monotone map and $\mathcal{A}$ is an affine monotone map. Define

$$\mathcal{A}^a = \sum_{i=1}^{k} \alpha_i \mathcal{A}(x_i), \qquad \hat{\varepsilon}_1^a := \sum_{i=1}^{k} \alpha_i \langle x_i - x^a, \mathcal{A}(x_i) \rangle, \tag{85}$$

$$G^a = \sum_{i=1}^{k} \alpha_i G(x_i), \qquad \hat{\varepsilon}_2^a := \sum_{i=1}^{k} \alpha_i [\varepsilon_i + \langle\, x_i - x^a\,,\, v_i - \mathcal{A}(x_i)\,\rangle\,, \tag{86}$$

and note that $\varepsilon^a = \hat{\varepsilon}_1^a + \hat{\varepsilon}_2^a$ in view of (26). Since $\mathcal{A}$ is an affine monotone map, we have $\mathcal{A}^a = \mathcal{A}(x^a)$ and $\hat{\varepsilon}_1^a \geq 0$, and hence $\varepsilon^a \geq \hat{\varepsilon}_2^a$. By (25) and the fact that $F = G + \mathcal{A}$, we have

$$v_i - \mathcal{A}(x_i) \in (G + \partial_{\varepsilon_i} g)(x_i), \quad i = 1, \ldots, k.$$

Using the definitions of v^a, $\mathcal{A}^a$, and $\hat{\varepsilon}_2^a$ in (26), (85), and (86), respectively, and statement (a) with $F = G$ and $v_i = v_i - \mathcal{A}(x_i)$, we then conclude that the pair $(v^a - \mathcal{A}^a, \hat{\varepsilon}_2^a)$, and hence $(v^a - \mathcal{A}(x^a), \varepsilon^a)$, is a weak residual of x^a for $HVI(G, g)$. Now, the identity

$$r_c := r_c(x^a; F - v^a) = r_c(x^a; G - (v^a - \mathcal{A}^a)),$$

which is due to definition (22), and the identity $F(x^a) = G(x^a) + \mathcal{A}(x^a) = G(x^a) + \mathcal{A}^a$, the fact that G is $\mathcal{N}(F; \operatorname{dom} g)$-Lipschitz continuous, and Proposition 3.7 imply that, for any $c \geq 2\mathcal{N}(F; \operatorname{dom} g)$, $\|r_c\| \leq \sqrt{2c\varepsilon^a}$ and the pair $(v^a - \mathcal{A}^a + r_c, \varepsilon^a)$ is a strong residual of x^a for $HVI(G, g)$. Since by definition this means that

$$v^a - \mathcal{A}^a + r_c \in G(x^a) + \partial_{\varepsilon^a} g(x^a)$$

or, equivalently, $v^a + r_c \in F(x^a) + \partial_{\varepsilon^a}(x^a)$, we conclude that $(v^a + r, \varepsilon^a)$ is a strong residual of x^a for $HVI(F, g)$.

(c) Since $0 \in N_\Omega(x_i)$, and hence $F(x_i) \in F(x_i) + N_\Omega(x_i)$, for every $i = 1, \ldots, k$, it follows from statement (b) with $g = \delta_\Omega$, $v_i = F(x_i)$, and $\varepsilon_i = 0$ that, for every $c \geq \mathcal{N}(F; \Omega)$, the vector $\hat{r}_c$ defined in (27) satisfies

$$\hat{r}_c + F^a \in F(x^a) + N_\Omega^{\varepsilon_1^a}(x^a), \qquad \|\hat{r}_c\| \leq \sqrt{2\varepsilon_1^a c}, \tag{87}$$

where F^a is defined in (26) and

$$\varepsilon_1^a := \sum_{i=1}^{k} \alpha_i \langle x_i - x^a, F(x_i) \rangle \geq 0. \tag{88}$$

On the other hand, we know from the proof of statement (a) that $v^a - F^a \in \partial_{\varepsilon_0^a} g(x^a)$, where $\varepsilon_0^a \geq 0$ is defined in (84). Combining this last observation with (87), we then conclude that

$$\hat{r}_c + v^a - F(x^a) \in \partial_{\varepsilon_0^a} g(x^a) + N_\Omega^{\varepsilon_1^a}(x^a) \subseteq \partial_{\varepsilon_0^a + \varepsilon_1^a}(g + \delta_\Omega)(x^a) = \partial_{\varepsilon^a} g(x^a),$$

where the last identity follows from the definition of ε^a, ε_0^a, and ε_1^a in (26), (84), and (88), respectively, and the fact that $\operatorname{dom} g \subseteq \Omega$ by assumption. Statement (c) now immediately follows from the latter conclusion, Proposition 3.5(a), the inequality in (87), and the fact that $\varepsilon_1^a \leq \varepsilon^a$.

The last claim of the theorem follows from (b) with $c = 2\mathcal{N}(F; \operatorname{dom} g)$. ∎

Appendix E. Proof of Theorem 4.4.

Before giving the proof of Theorem 4.4, we first establish the following result.

THEOREM E.1. *Let $\{\tilde{x}_k\}$ and $\{x_k\}$ be the sequences generated by the generalized Korpelevich's extragradient algorithm, and consider the ergodic sequences $\{\tilde{x}_k^a\}$, $\{\tilde{v}_k^a\}$, and $\{\tilde{\varepsilon}_k^a\}$ defined according to* (36) *and* (37)*, and the sequence $\{\tilde{F}_k^a\}$ defined as*

$$\tilde{F}_k^a = \sum_{i=1}^{k} F(\tilde{x}_i). \tag{89}$$

Let d_0 be the distance of x_0 to the solution set of $HVI(F, g)$. Then, for every $k \in \mathbb{N}$, the following statements hold:

(a) *If a positive scalar $c \geq 2\mathcal{N}(F; \operatorname{dom} g)$ is known and we define*

$$\hat{v}_k^a := \tilde{v}_k^a + r_c(\tilde{x}_k^a; F - \tilde{v}_k^a, g), \tag{90}$$

then $(\hat{v}_k^a, \tilde{\varepsilon}_k^a)$ is a strong residual of $\tilde{x}_k^a$ for $HVI(F, g)$ and

$$\|\hat{v}_k^a\| \leq \frac{2Ld_0}{k\sigma} + 2d_0\sqrt{\frac{cL\bar{\eta}_k}{k\sigma}}. \tag{91}$$

(b) *Alternatively, if a closed convex set $\Omega \subseteq \mathbb{R}^n$ such that $\operatorname{dom} g \subseteq \Omega \subseteq \operatorname{Dom} F$ and a positive constant $c \geq 2\mathcal{N}(F; \Omega)$ are known and we define*

$$\check{v}_k^a := \tilde{v}_k^a + r_c(\tilde{x}_k^a; F - \tilde{F}_k^a, \delta_\Omega), \tag{92}$$

then $(\check{v}_k^a, \tilde{\varepsilon}_k^a)$ is a strong residual of $\tilde{x}_k^a$ for $HVI(F, g)$, and $\check{v}_k^a$ also satisfies (91) with $\hat{v}_k^a$ replaced by $\check{v}_k^a$.

As a consequence, if a constant $c > 0$ satisfying $2\mathcal{N}(F; \operatorname{dom} g) \leq c = \mathcal{O}(\mathcal{N}(F; \operatorname{dom} g))$ is known, then for every pair of positive scalars (ρ, ε), there exists an index

$$k_0 = \mathcal{O}\left(\max\left[\frac{Ld_0^2}{\varepsilon}, \frac{Ld_0}{\rho} + \frac{d_0^2 L\mathcal{N}(F; \operatorname{dom} g)}{\rho^2}\right]\right) \tag{93}$$

such that, for any $k \geq k_0$, the pair $(\hat{v}_k^a, \tilde{\varepsilon}_k^a)$ is an easily computable certificate that the point $\tilde{x}_k^a$ is a (ρ, ε)-strong solution of $HVI(F, g)$.

Proof. By Proposition 4.5 (resp., Proposition 4.1), the variant of Tseng's MF-BS method (resp., the generalized Korpelevich's method) is a special case of the HPE method with $T = F + \partial g$, and, for every $k \in \mathbb{N}$, $\tilde{v}_k \in [F + \partial_{\varepsilon_k} g](\tilde{x}_k)$, where $\varepsilon_k = 0$ (resp., ε_k is given by (34)). Hence, statements Theorem E.1(a), (b), and (c) follow immediately from Theorem 2.5 with $\lambda_k = \sigma/L$ for every k and Theorem 3.9 with $x_i = \tilde{x}_i$, $v_i = \tilde{v}_i$, and $\alpha_i = 1/k$ for $i = 1, \ldots, k$. The last part of Theorem E.1 follows from statement (b) and some straightforward arguments. ∎

We now make a few observations regarding the last result. First, note that the complexity bound in Theorem E.1(a) can be significantly better than that obtained in Theorem 4.2, namely, when the constant $c \ll L$. Second, in the case where $\Omega = \mathbb{R}^n$, the vector $\check{v}_k^a$ reduces to

$$\check{v}_k^a = \tilde{v}_k^a + F(\tilde{x}_k^a) - \tilde{F}_k^a,$$

which does not depend on c. Hence, in this case, knowledge of a constant $c \geq \mathcal{N}(F; \operatorname{dom} g)$ is not required. Moreover, when F is affine and we choose $\Omega = \mathbb{R}^n$, (91) holds for any $c > 0 = \mathcal{N}(F; \operatorname{dom} g)$, from which we conclude that

$$\|\check{v}_k^a\| \leq \frac{2Ld_0}{k\sigma},$$

and hence that (93) also holds also when $\mathcal{N}(F; \operatorname{dom} g) = 0$. Third, Theorem E.1(a) with $\Omega = \mathbb{R}^n$ and g being an indicator function reduces to Theorem 5.5 of [9]. Fourth, the main drawback about the last statement of Theorem E.1 is the fact that a constant $c \geq \mathcal{N}(F; \operatorname{dom} g)$ such that $c = \mathcal{O}(\mathcal{N}(F; \operatorname{dom} g))$ must be known. The natural question then arises as to whether it is possible to compute a certificate that $\tilde{x}_k^a$ is a (ρ, ε)-strong solution of $VI(F, X)$ within a number of iterations bounded by (93), without any knowledge of a constant c as above.

An affirmative answer to the latter question is given by Theorem 4.4, whose proof we now give.

Proof of Theorem 4.4. Consider the sequences $\{\tilde{x}_k^a\}$ and $\{\tilde{\varepsilon}_k^a\}$ defined in (36) and (37), and the sequence $\{\hat{v}_k^a\}$ defined in (90) with $c = 2\mathcal{N}(F; \operatorname{dom} g)$. Then, by Theorem E.1 the pair $(\hat{v}_k^a, \tilde{\varepsilon}_k^a)$ is a strong residual of $\tilde{x}_k^a$ for $HVI(F,g)$ satisfying the estimates (38) and (91) with $c = 2\mathcal{N}(F; \operatorname{dom} g)$. Hence, we easily conclude that there exists k_0 such that (40) holds and $\tilde{x}_k^a$ is a $(\rho/\sqrt{2}, \varepsilon/2)$-strong solution of $HVI(F,g)$ for any $k \geq k_0$. This conclusion together with Proposition 3.6(b) then implies that the pair $(r_{\bar{c}}(\tilde{x}_k^a; F, g), \varepsilon_{\bar{c}}(\tilde{x}_k^a; F, g))$ is a strong residual of $\tilde{x}_k^a$ for $HVI(F,g)$ satisfying (41). ∎

Acknowledgment. The authors would also like to thank the two anonymous referees for their insightful comments on the first few drafts of this paper.

REFERENCES

[1] A. Auslender and M. Teboulle, *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer Monogr. Math., Springer-Verlag, New York, 2003.

[2] 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.

[3] 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.

[4] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Volume* II, Springer-Verlag, New York, 2003.

[5] E. G. Golshtein and N. V. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*, Wiley-Intersci. Ser. Discrete Math. Optim., John Wiley & Sons Inc., New York, 1996.

[6] I. V. Konnov, *A combined method for solving variational inequalities with monotone operators*, Zh. Vychisl. Mat. Mat. Fiz., 39 (1999), pp. 1091–1097.

[7] 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, Berlin, 2007, pp. 3–31.

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

[9] R. D. C. Monteiro and B. F. Svaiter, *On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean*, SIAM J. Optim., 20 (2010), pp. 2755–2787.

[10] 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.

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

[12] 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.

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

[14] M. A. Noor, *An extraresolvent method for monotone mixed variational inequalities*, Math. Comput. Modelling, 29 (1999), pp. 95–100.

[15] M. Patriksson, *Nonlinear Programming and Variational Inequality Problems. A Unified Approach*, Appl. Optim. 23, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.

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

[17] R. T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc., 149 (1970), pp. 75–88.

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

[19] 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.

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

[21] 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.

[22] 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.

[23] P. Tseng, *On linear convergence of iterative methods for the variational inequality problem. Linear/nonlinear iterative methods and verification of solution*, (Matsuyama, 1993), J. Comput. Appl. Math., 60 (1995), pp. 237–252.

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