ON THE CONVERGENCE PROPERTIES OF NON-EUCLIDEAN EXTRAGRADIENT METHODS FOR VARIATIONAL INEQUALITIES WITH GENERALIZED MONOTONE OPERATORS *

CONG D. DANG † AND GUANGHUI LAN ‡

Abstract. In this paper, we study a class of generalized monotone variational inequality (GMVI) problems whose operators are not necessarily monotone (e.g., pseudo-monotone). We present non-Euclidean extragradient (N-EG) methods for computing approximate strong solutions of these problems, and demonstrate how their iteration complexities depend on the global Lipschitz or Hölder continuity properties for their operators and the smoothness properties for the distance generating function used in the N-EG algorithms. We also introduce a variant of this algorithm by incorporating a simple line-search procedure to deal with problems with more general continuous operators. Numerical studies are conducted to illustrate the significant advantages of the developed algorithms over the existing ones for solving large-scale GMVI problems.

Keywords: Complexity, Monotone variational inequality, Pseudo-monotone variational inequality, Extragradient methods, Non-Euclidean methods, Prox-mapping

1. Introduction.

Variational inequality (VI) has been widely studied in the literature due to its encompassing power of describing a wide range of optimization, equilibrium and complementarity problems (see [7] and references therein). Given an nonempty closed convex set $X \subseteq \mathbb{R}^n$ and a continuous mapping $F : X \to \mathbb{R}^n$, the variational inequality problem, denoted by $\mathrm{VI}(X, F)$, is to find $x^* \in X$ satisfying

$$\langle F(x^*), x - x^*\rangle \geq 0 \quad \forall\, x \in X. \tag{1.1}$$

Such a point x^* is often called a *strong solution* of $\mathrm{VI}(X, F)$.

The extragradient method, initially proposed by Korpelevich [14], is a classical method for solving VI problems. It improves the usual gradient projection method (e.g., [27, 5]) by performing an additional metric projection step at each iteration. While earlier studies on extragradient methods were focused on their asymptotical convergence analysis (see, e.g., [29, 30, 32]), much recent effort has been directed to the complexity analysis of these types of methods. In particular, Nemirovski [21] presented a generalized version of Korpelevich's extragradient method and analyzed its iteration complexity in terms of the computation of a *weak solution*, i.e., a point $x^* \in X$ such that

$$\langle F(x), x - x^*\rangle \geq 0 \quad \forall\, x \in X. \tag{1.2}$$

Note that if $F(\cdot)$ is monotone and continuous, a weak solution of $\mathrm{VI}(X, F)$ must be a strong solution and vise versa. Moreover, he showed that one can possibly improve the complexity results by employing the non-Euclidean projection (prox-mapping) steps (see (3.3)) in place of the two metric projection steps in Korpelevich's extragradient method. These types of methods are referred to as *non-Euclidean extragradient (N-EG)* methods in this paper. Similar results have also been developed by Auslender and Teboulle [1] for their interior projection methods applied to monotone VI problems. More recently, Monteiro and Svaiter [19] established the complexity for a class of hybrid proximal extragradient methods [28] which covers Korpelevich's extragradient method as a special case. Other approaches and their associated rate of convergence for solving VI problems have also been studied (see, e.g., [25, 12]). Note that in all these previous studies in [21, 1, 19, 25, 12], the operator $F(\cdot)$ is assumed to be monotone.

In this paper, we consider a more general class of VI problems for which the operator $F(\cdot)$ is not necessarily monotone. In particular, we make the following much weaker assumption about the monotonicity of $\mathrm{VI}(X, F)$, i.e., relation (1.2) holds for any strong solution $x^* \in X$ (e.g., [10]). This class of VI problems, refereed to as *generalized monotone variational inequalities* (GMVI), cover both monotone and pseudo-monotone VI problems.

*The author of this paper was partially supported by NSF grants CMMI-1000347 and DMS-1319050, ONR grant N00014-13-1-0036, and NSF CAREER Award CMMI-1254446.

†Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, 32611. (email: congdd@ufl.edu).

‡Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, 32611. (email: glan@ise.ufl.edu).

It is also not difficult to construct GMVI problems whose operators are neither monotone nor pseudo-monotone (c.f., (2.4)). However, to the best of our knowledge, there does not exist any complexity results for the extragradient methods applied to GMVI problems in the literature. In particular, the previous complexity studies conducted for VI [21, 1, 19, 25, 12] relies on the monotonicity assumption of the operator $F(\cdot)$ and hence are not applicable for the possibly non-monotone GMVI problems. Moreover, most of the previous complexity analysis has been conducted for computing a weak solution approximately satisfying (1.2), and there exists very few complexity results for computing approximate strong solutions (see [19]). In particular, if our goal is to compute an approximate strong solution, it was unclear, even for monotone VI problems, how the N-EG method will be more advantageous over Korpelevich's extragradient method.

The main goal of this paper is to present a generalization of the N-EG method in [21] for solving GMVI problems and discuss its convergence properties in terms of the computation of approximate strong solutions. Our major contributions are summarized as follows. Firstly, we present a new termination criterion based on the residual function associated with the prox-mapping, and discuss its relations with a few other possible notions of approximate strong solutions to $\text{VI}(X, F)$. In particular, we show that under certain conditions, if a point $x \in X$ has a small residual, it must be associated with an approximate strong solution $y \in X$ with a small optimality gap $g(y)$, where

$$g(y) = \max_{z \in X} \langle F(y), y - z \rangle.$$

We also show how this termination criterion is related to the notion of an approximate strong solution recently proposed by Monteiro and Svaiter [19] to deal with VI problems with unbounded X.

Secondly, we study the complexity of the N-EG method for solving GMVI problems whose operator $F(\cdot)$ satisfies certain global continuity assumptions. In particular, by employing a novel analysis, we show that, if $F(\cdot)$ is Lipschitz continuous, then the N-EG method applied to GMVI problems can generate a solution $y_k \in X$ with $g(y_k)$ bounded by

$$\mathcal{O}(1) \left(\frac{L(\alpha + \mathcal{Q})\Omega^2_{\omega,X}}{\alpha\sqrt{k}} \right).$$

Here, $\mathcal{O}(1)$ denotes an absolute constant, L is the Lipschitz constant of F, α and $\mathcal{Q}$ are certain constants of the distance generating function $\omega(\cdot)$ used to define the prox-mapping, and $\Omega^2_{\omega,X}$ is a characteristic constant depending on $\omega(\cdot)$ and X. We also consider GMVI problem with Hölder continuous operators and show that the N-EG method possesses an $\mathcal{O}(1/k^{\nu/2})$ rate of convergence for solving this class of problems, where $\nu \in (0, 1]$ denotes the level of continuity. Our development also improves an existing result for computing an approximate strong solution for Lipschitz and monotone VI with unbounded X by removing a purification procedure introduced by Monteiro and Svaiter in [20].

Thirdly, in order to deal with more general GMVI problems whose operators are not necessarily Hölder continuous, we present a variant of N-EG method by incorporating a simple line-search procedure (N-EG-LS) and show that it can generate a sequence of solutions converging to a strong solution of $\text{VI}(X, F)$. It should be noted that, while earlier extragradient type methods for GMVI problems with a general continuous operator (e.g., Solodov and Svaiter [29], Sun [30]) rely on a certain monotonicity property of the metric projection (e.g., Gafni and Bertsekas [8]), such a property is not assumed by the prox-mapping in general. We present certain sufficient conditions on the prox-mapping which can guarantee the convergence of the N-EG-LS algorithm. More specifically, we show that these conditions are satisfied by the prox-mapping induced by distance generating functions with Lipschitz continuous gradients.

Finally, we present promising numerical results for the developed N-EG methods for solving GMVI problems. In particular, we demonstrate that the N-EG-LS method can be more advantageous over N-EG method if the Lipschitz constant is big or unknown. Moreover, we show that the N-EG methods with a properly chosen distance generating function $\omega(\cdot)$ can outperform the Euclidean methods especially when the dimension n is big.

This paper is organized as follows. In Section 2, we describe in more details the GMVI problems. We discuss the prox-mapping and the termination criterion associated with the prox-mapping in Section 3. Then we present the N-EG method for solving GMVI problems with Lipschitz or Hölder continuous operators in Section 4. Section 5 is devoted to the N-EG-LS method applied to GMVI problems with general continuous operators. Finally, numerical results are presented in Section 6.

1.1. Notation and terminology. let $\mathbb{R}^n$ be an arbitrary finite dimensional vector space endowed with the inner product $\langle \cdot, \cdot \rangle$, $\|\cdot\|$ denote a norm in $\mathbb{R}^n$ (not necessarily the one associated with the inner product), and $\|\cdot\|_*$ denotes its conjugate norm. Let $X \subseteq \mathbb{R}^n$ be closed and convex. A function $f : X \to \mathbb{R}$ is said to have L-Lipschitz continuous gradient if it is differentiable and

$$\|\nabla f(x) - \nabla f(y)\|_* \le L\|x - y\|, \forall x, y \in X.$$

For a given $m \times n$ real-valued matrix A, letting $\|A\|_2$ be the spectral norm and $\|A\|_{\max} = \max_{ij}\{|A_{ij}|\}$, we have

$$\|A\|_{\max} \le \|A\|_2 \le \sqrt{mn}\|A\|_{\max}. \tag{1.3}$$

We use $\mathbb{N}$ to denote the set of natural numbers.

2. The problem of interest. Given an nonempty closed convex set $X \subseteq \mathbb{R}^n$ and a continuous mapping $F : X \to \mathbb{R}^n$, the problem of interest in this paper is find a strong solution x^* of $\mathrm{VI}(X, F)$, i.e., a vector $x^* \in X$ such that (1.1) holds. In this paper, we assume that the solution set X^* of $\mathrm{VI}(X, F)$ is nonempty. Moreover, the following assumption is made throughout the paper.

A1 For any $x^* \in X^*$ we have

$$\langle F(x), x - x^* \rangle \ge 0 \quad \forall\, x \in X. \tag{2.1}$$

Clearly, Assumption A1 is satisfied if $F(\cdot)$ is monotone, i.e.,

$$\langle F(x) - F(y), x - y \rangle \ge 0 \quad \forall\, x, y \in X. \tag{2.2}$$

Moreover, this assumption holds if $F(\cdot)$ is pseudo-monotone, i.e.,

$$\langle F(y), x - y \rangle \ge 0 \implies \langle F(x), x - y \rangle \ge 0. \tag{2.3}$$

As an example, $F(\cdot)$ is pseudo-monotone if it is the gradient of a real-valued differentiable pseudo-convex function. It is also not difficult to construct VI problems that satisfy (2.1), but their operator $F(\cdot)$ is neither monotone nor pseudo-monotone anywhere (see [29] and references therein). One set of simple examples are given by all the functions $F : \mathbb{R} \to \mathbb{R}$ satisfying

$$F(x) \begin{cases} = 0, & x = x_0; \\ \ge 0, & x \ge x_0; \\ \le 0, & x \le x_0. \end{cases} \tag{2.4}$$

These problems, although satisfying Assumption A1 with $x^* = x_0$, can be neither monotone nor pseudo-monotone.

For future reference, we say that $\mathrm{VI}(X, F)$ is a generalized monotone VI (GMVI) problem whenever Assumption A1 is satisfied.

The condition given by (1.1) is the standard definition of a *strong solution* to $\mathrm{VI}(X, F)$. Recall that a solution x^* satisfying (2.1) is usually called is a *weak solution* to $\mathrm{VI}(X, F)$. Clearly, under our assumption, a strong solution must be a weak solution for the GMVI problems. The inverse is also true if $F(\cdot)$ is continuous and monotone. However, such a relation does not necessarily hold when $F(\cdot)$ is not monotone and hence the computation of an approximate weak solution is not particularly useful in these cases. In addition, as pointed out by Monteiro and Svaiter [19], a strong solution to $\mathrm{VI}(X, F)$ admits certain natural explanations for some important classes of monotone VI problems, e.g., the complementarity problems. This paper focuses on the computation of approximate strong solutions to VI(X, F) (see Section 3).

Depending on the continuity properties of $F(\cdot)$, we consider the following four different classes of VI problems.

i) $F(\cdot)$ is Lipschitz continuous:

$$\|F(x) - F(y)\|_* \leq L\|x - y\|, \quad \forall\, x, y \in X; \tag{2.5}$$

ii) $F(\cdot)$ is Hölder continuous: for some $\nu \in (0, 1]$,

$$\|F(x) - F(y)\|_* \leq L\|x - y\|^\nu, \quad \forall\, x, y \in X; \tag{2.6}$$

iii) $F(\cdot)$ is locally Lipschitz continuous: for every $x \in X$, there exists a neighborhood B_x of x, such that

$$\|F(x) - F(y)\|_* \leq L\|x - y\|, \quad \forall\, x, y \in B_x \subset X; \tag{2.7}$$

iv) $F(\cdot)$ is continuous:

$$\lim_{y \to x} \|F(x) - F(y)\|_* = 0, \quad \forall x \in X. \tag{2.8}$$

Clearly, if $F(\cdot)$ is Lipschitz continuous, then it is Hölder continuous with $\nu = 1$. Moreover, in view of the assumption that $\nu > 0$, a Hölder continuous $F(\cdot)$ must be continuous but not vise versa. In addition, a locally Lipschitz continuous $F(\cdot)$ must be continuous but the inverse is not necessarily true. In Sections 4 and 5, we will present algorithms for solving different classes of VI problems and show how their convergence properties depend on the continuity assumption of $F(\cdot)$.

3. Prox-mapping and termination criteria. In this section, we discuss the main computational construct, i.e., the prox-mapping, that will be used in the non-Euclidean extragradient methods. We also present a termination criterion based on the prox-mapping and show how it relates to some other termination criteria for solving VI problems. It is worth noting that the results in this section does not require Assumption A1.

3.1. Distance generating function and prox-mapping. We review the concept of prox-mapping (e.g., [21, 2, 22]) in this subsection.

A function $\omega : X \to \mathbb{R}$ is called a *distance generating function* modulus $\alpha > 0$ with respect to $\|\cdot\|$, if the following conditions hold: i) $\omega(\cdot)$ is convex and continuous on X; ii) the set

$$X^o = \{x \in X : \partial w(x) \neq \emptyset\}$$

is convex (note that X^o always contains the relative interior of X); and iii) restricted to X^o, $\omega(\cdot)$ is continuously differentiable and strongly convex with parameter α with respect to $\|\cdot\|$, i.e.,

$$\langle \nabla\omega(x') - \nabla\omega(x), x' - x\rangle \geq \alpha\|x' - x\|^2, \ \ \forall x', x \in X^o. \tag{3.1}$$

Given a distance generating function ω, the *prox-function* $V : X^o \times X \to \mathbb{R}_+$ is defined by

$$V(x, z) = \omega(z) - [\omega(x) + \langle \nabla\omega(x), z - x\rangle]. \tag{3.2}$$

The function $V(\cdot, \cdot)$ is also called the Bregman's distance, which was initially studied by Bregman [6] and later by many others (see [2, 3, 13, 31] and references therein). In this paper, we assume that the prox-function $V(x, z)$ is chosen such that, for a given $x \in X^o$, the *prox-mapping* $P_x : \mathbb{R}^n \to \mathbb{R}^n$ defined as

$$P_x(\phi) = \arg\min_{z \in X} \left\{\langle \phi, z\rangle + V(x, z)\right\} \tag{3.3}$$

is easily computable. It can be seen from the strong convexity of $\omega(\cdot)$ and (3.2) that

$$V(x, z) \geq \frac{\alpha}{2}\|x - z\|^2, \quad \forall x, z \in X. \tag{3.4}$$

In some cases, we assume that the distance generating function $\omega(\cdot)$ satisfies

$$\|\nabla\omega(x) - \nabla\omega(z)\|_* \le \mathcal{Q}\|x - z\|, \quad \forall x, z \in X, \tag{3.5}$$

for some $\mathcal{Q} \in (0, \infty)$. Under this assumption, it can be easily seen that (see, e.g., Lemma 1.2.3 of [24])

$$V(x, z) \le \frac{\mathcal{Q}}{2}\|x - z\|^2, \quad \forall x, z \in X. \tag{3.6}$$

We say that the prox-function $V(\cdot, \cdot)$ is growing quadratically whenever condition (3.6) holds.

Proposition 3.1 below provides a few examples for the selection of $\|\cdot\|$ and distance generating function $\omega(\cdot)$. More such examples can be found, for example, in [2, 4, 23].

PROPOSITION 3.1.

a) If $X = \mathbb{R}^n$, $\|\cdot\| = \|\cdot\|_2$ and $\omega(x) = \|x\|_2^2/2$, then we have $\alpha = \mathcal{Q} = 1$.

b) If $X = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i \ge 0, i = 1, \ldots, n\}$, $\|\cdot\| = \|\cdot\|_1$ and $\omega(x) = \sum_{i=1}^n (x_i + \delta/n)\log(x_i + \delta/n)$ with $\delta = 10^{-16}$, then we have $\alpha = \mathcal{O}(1)$ and $\mathcal{Q} = 1 + n/\delta$. Here $\mathcal{O}(1)$ denotes an absolute constant.

c) If $X = \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}$, where $\|x\|_1 = \sum_{i=1}^n |x_i|$, $\|\cdot\| = \|\cdot\|_1$, and

$$\omega(x) = \frac{1}{2}\|x\|_p^2 = \frac{1}{2}\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{2}{p}} \tag{3.7}$$

with $p = 1 + 1/\ln n$, then we have $\alpha = \mathcal{O}(1)(1/\ln n)$.

Proof. Part a) is obvious and part b) has been shown in Chapter 5 of [4]. Moreover, the strong convexity of $\omega(x)$ in (3.7) and the estimation of its modulus (with $p = 1 + 1/\ln n$) is shown in [23]. ∎

The distance generating function $\omega(\cdot)$ also gives rise to the following characteristic entity that will be used frequently in our convergence analysis:

$$D_{\omega,X} := \sqrt{\max_{x\in X} \omega(x) - \min_{x \in X} \omega(x)}. \tag{3.8}$$

Let x_1 be the minimizer of ω over X. Observe that $x_1 \in X^o$, whence $\nabla w(x_1)$ is well defined and satisfies $\langle \nabla\omega(x_1), x - x_1\rangle \ge 0$ for all $x \in X$, which combined with the strong convexity of ω implies that

$$\frac{\alpha}{2}\|x - x_1\|^2 \le V(x_1, x) \le \omega(x) - \omega(x_1) \le D_{\omega,X}^2, \ \ \forall x \in X, \tag{3.9}$$

and hence

$$\|x - x_1\| \le \Omega_{\omega,X} := \sqrt{\frac{2}{\alpha}} D_{\omega,X} \text{ and } \|x - x'\| \le 2\Omega_{\omega,X}, \ \forall x, x' \in X. \tag{3.10}$$

3.2. A termination criterion based on the prox-mapping. In this subsection, we introduce a termination criterion for solving VI associated with the prox-mapping.

We first provide a simple characterization of a strong solution to $\mathrm{VI}(X, F)$.

LEMMA 3.2. *A point $x \in X$ is a strong solution of $\mathrm{VI}(X, F)$ if and only if*

$$x = P_x(\gamma F(x)) \tag{3.11}$$

for some $\gamma > 0$.

Proof. If (3.11) holds, then by the optimality condition of (3.3), i.e.,

$$\langle \gamma F(x) + \nabla\omega(P_x(\gamma F(x))) - \nabla\omega(x), z - P_x(\gamma F(x))\rangle \ge 0, \quad \forall z \in X, \tag{3.12}$$

we have $\langle \gamma F(x), z - x\rangle \geq 0$ for any $z \in X$, which, in view of the fact that $\gamma > 0$ and definition (1.1), implies that x is a strong solution of VI(X, F). The "only if" part of the statement easily follows from the optimality condition of (3.3). ∎

Motivated by Lemma 3.2, we can define the residual function for a given $x \in X$ as follows.

Definition 3.3. *Let $\|\cdot\|$ be a given norm in $\mathbb{R}^n$, $\omega(\cdot)$ be a distance generating function modulus $\alpha > 0$ w.r.t. $\|\cdot\|$ and $P_x(\cdot)$ be the prox-mapping defined in (3.3). Then, for some positive constant γ, we define the residual $R_\gamma(\cdot)$ at the point $x \in X$ as*

$$R_\gamma(x) := \frac{1}{\gamma}\left[x - P_x(\gamma F(x))\right]. \tag{3.13}$$

Observe that in the Euclidean setup where $\|\cdot\| = \|\cdot\|_2$ and $\omega(x) = \|x\|_2^2/2$, the residual $R_\gamma(\cdot)$ in (3.13) reduces to

$$R_\gamma(x) = \frac{1}{\gamma}\left[x - \Pi_X(x - \gamma F(x))\right], \tag{3.14}$$

where $\Pi_X(\cdot)$ denotes the metric projection over X. Such a residual function in (3.14) has been used in the asymptotic analysis of different algorithms for solving VI problems (e.g., [29, 30]). In particular, if $F(\cdot)$ is the gradient of a real-valued differentiable function $f(\cdot)$, the residual $R_\gamma(\cdot)$ in (3.14) corresponds to the well-known projected gradient of $f(\cdot)$ at x (see, e.g., [17, 18, 24]).

The following two results are immediate consequences of Lemma 3.2 and Definition 3.3.

Lemma 3.4. *A point $x \in X$ is a strong solution of* VI(X, F) *if and only if $\|R_\gamma(x)\| = 0$ for some $\gamma > 0$;*

Lemma 3.5. *Suppose that $x_k \in X$ and $\gamma_k \in (0, \infty)$, $k = 1, 2, \ldots$, satisfy the following conditions:*

i) $\lim_{k\to\infty} V(x_k, P_{x_k}(\gamma_k F(x_k))) = 0$;

ii) There exists $K \in \mathbb{N}$ and $\gamma^* > 0$ such that $\gamma_k \geq \gamma^*$ for any $k \geq K$.

Then we have $\lim_{k\to\infty} \|R_{\gamma_k}(x_k)\| = 0$. *If in addition, the sequence $\{x_k\}$ is bounded, there exists an accumulation point $\tilde{x}$ of $\{x_k\}$ such that $\tilde{x} \in X^*$, where X^* denotes the solution set of* VI(X, F).

Proof. Denote $y_k = P_{x_k}(\gamma_k F(x_k))$. It follows from (3.4) and condition i) that $\lim_{k\to\infty} \|x_k - y_k\| = 0$. This observation, in view of Condition ii) and Definition 3.3, then implies that $\lim_{k\to\infty} \|R_{\gamma_k}(x_k)\| = 0$. Moreover, if $\{x_k\}$ is bounded, there exist a subsequence $\{\tilde{x}_i\}$ of $\{x_k\}$ obtained by setting $\tilde{x}_i = x_{n_i}$ for $n_1 \leq n_2 \leq \ldots$, such that $\lim_{i\to\infty} \|\tilde{x}_i - \tilde{x}\| = 0$. Let $\{\tilde{y}_i\}$ be the corresponding subsequence in $\{y_k\}$, i.e., $y_i = P_{x_{n_i}}(\gamma_{n_i} F(x_{n_i}))$, and $\tilde{\gamma}_i = \gamma_{n_i}$. We have $\lim_{i\to\infty} \|\tilde{x}_i - \tilde{y}_i\| = 0$. Moreover, by (3.12), we have

$$\langle F(\tilde{x}_i) + \frac{1}{\tilde{\gamma}_i}\left[\nabla\omega(\tilde{y}_i) - \nabla\omega(\tilde{x}_i)\right], z - \tilde{y}_i\rangle \geq 0, \quad \forall z \in X, \forall i \geq 1.$$

Tending i to $+\infty$ in the above inequality, and using the continuity of $F(\cdot)$ and $\nabla\omega(\cdot)$, and condition ii), we conclude that $\langle F(\tilde{x}), z - \tilde{x}\rangle \geq 0$ for any $z \in X$. ∎

In the remaining part of this section, we relate the residual $R_\gamma(\cdot)$ to a few other possible termination criteria for solving VI(X, F).

Observe that, if the set X is bounded, then in view of definition (1.1), one can measure the inaccuracy of a solution $x \in X$ by the gap function (see [11, 10] and references therein):

$$g(x) := \sup_{z\in X} \langle F(x), x - z\rangle. \tag{3.15}$$

It can be easily seen that $g(x) \geq 0$ for any $x \in X$ and that the point $x^* \in X$ is a strong solution of VI(X, F) if and only if $g(x^*) = 0$. Note also that the gap function in (3.15) does not depend on any algorithmic parameters, while the definition of the residual function $R_\gamma(\cdot)$ in (3.13) depends on the selection of $\|\cdot\|$ and $\omega(\cdot)$.

However, if X is unbounded, then the gap function $g(\cdot)$ in (3.15) may not be well-defined. To address this issue, Monteiro and Svaiter [19] suggested a generalization of the gap function $g(\cdot)$ so as to deal with the unbounded feasible sets. More specifically, they define a new gap function $\tilde{g}(\cdot, \phi)$ as

$$\tilde{g}(x, \phi) := \sup_{z \in X} \langle F(x) + \phi, x - z \rangle. \tag{3.16}$$

Observe that there always exists a point ϕ (e.g., $\phi = -F(x)$) such that $\tilde{g}(x, \phi)$ is well-defined for any $x \in X$. Accordingly, an approximate strong solution of $\mathrm{VI}(X, F)$ can be defined as follows.

Definition 3.6. *A point $x \in X$ is called an (ϵ, δ)-strong solution of $\mathrm{VI}(X, F)$, if there exists some ϕ such that $\|\phi\|_* \le \epsilon$ and $\tilde{g}(x, \phi) \le \delta$.*

Observe that the above definition of an (ϵ, δ)-strong solution of $\mathrm{VI}(X, F)$ relies on the selection of the norm $\|\cdot\|_*$ (and hence $\|\cdot\|$). In [19], Monteiro and Svaiter focused on the Euclidean setup where $\|\cdot\| = \|\cdot\|_2$ and considered VI problems with monotone and Lipschitz continuous operator $F(\cdot)$. Moreover, as pointed out in [19], the two tolerances, namely ϵ and δ, used in the definition of an (ϵ, δ)-strong solution possess natural interpretations in the context of complementarity problems. In particular, the following result has been shown in Proposition 3.1 of [19].

Proposition 3.7. *Assume that $X = K$, where K is a nonempty closed convex cone (and K^* be its dual cone). Then, $x \in K$ is an (ϵ, δ)-strong solution of $\mathrm{VI}(X, F)$ if and only if there exists $\phi \in K^*$ such that*

$$\|F(x) - \phi\|_* \le \epsilon, \quad \langle x, \phi \rangle \le \delta.$$

In other words, the first tolerance ϵ measures the infeasibility of $F(x)$ with respect to the dual cone while the second tolerance δ measures the size of the complementarity slackness.

In the following result we show some relations between the the gap functions defined in (3.15) and (3.16) and the residual function $R_\gamma(\cdot)$ defined in (3.13).

Proposition 3.8. *Let $x \in X$ be given. Also, for some $\gamma > 0$, let us denote*

$$x^+ := P_x(\gamma F(x)), \tag{3.17}$$

$$\phi_\gamma(x) := F(x) - F(x^+) + \frac{1}{\gamma}\left[\nabla\omega(x^+) - \nabla\omega(x)\right]. \tag{3.18}$$

a) We have $\tilde{g}(x^+, \phi_\gamma(x)) \le 0$ for any $x \in X$ and $\gamma > 0$;

b) If $\omega(\cdot)$ has $\mathcal{Q}$-Lipschitz continuous gradients w.r.t. $\|\cdot\|$ and $F(\cdot)$ is Hölder continuous, i.e., condition (2.6) holds for some $\nu \in (0, 1]$, then

$$\|\phi_\gamma(x)\|_* \le L\left[\gamma\|R_\gamma(x)\|\right]^\nu + \mathcal{Q}\|R_\gamma(x)\|; \tag{3.19}$$

c) If, in addition, the set X is bounded, then

$$g(x^+) \le 2\Omega_{\omega,X}\left[L\gamma^\nu\|R_\gamma(x)\|^\nu + \mathcal{Q}\|R_\gamma(x)\|\right], \tag{3.20}$$

where $\Omega_{\omega,X}$ is defined in (3.10).

Proof. Using the definition of x^+ in (3.17), the optimality condition of (3.3) and the fact that $\nabla V(x, z) = \nabla\omega(z) - \nabla\omega(x)$, we have

$$\langle F(x) + \frac{1}{\gamma}\left[\nabla\omega(x^+) - \nabla\omega(x)\right], z - x^+ \rangle \ge 0, \quad \forall\, z \in X,$$

which together with (3.16) and (3.18) then imply that

$$\begin{aligned} \tilde{g}(x^+, \phi_\gamma(x)) &= \sup_{z \in X}\langle F(x^+) + \phi_\gamma(x), x^+ - z\rangle \\ &= \sup_{z \in X}\langle F(x) + \frac{1}{\gamma}\left[\nabla\omega(x^+) - \nabla\omega(x)\right], x^+ - z\rangle \le 0. \end{aligned} \tag{3.21}$$

We have thus shown part a). Now it follows from the triangular inequality, (3.18), (3.5), (2.6) and (3.13) that

$$\begin{aligned}\|\phi_\gamma(x)\|_* &\le \|F(x) - F(x^+)\|_* + \frac{1}{\gamma}\|\nabla\omega(x^+) - \nabla\omega(x)\|_* \\ &\le L\|x - x^+\|^\nu + \frac{\mathcal{Q}}{\gamma}\|x^+ - x\| = L\left[\gamma\|R_\gamma(x)\|\right]^\nu + \mathcal{Q}\|R_\gamma(x)\|,\end{aligned}$$

which implies (3.19). By using the above conclusion, (3.15) and (3.21), we have

$$\begin{aligned}g(x^+) &= \sup_{z\in X}\langle[F(x^+) + \phi_\gamma(x)] - \phi_\gamma(x), x^+ - z\rangle \\ &\le \sup_{z\in X}\langle -\phi_\gamma(x), x^+ - z\rangle \\ &\le \left[L\left(\gamma\|R_\gamma(x)\|\right)^\nu + \mathcal{Q}\|R_\gamma(x)\|\right]\sup_{z\in X}\|x^+ - z\|,\end{aligned}$$

The above inequality together with (3.10) then imply (3.20). ∎

By using Proposition 3.8, we can easily see the relation between the residual function $R_\gamma(\cdot)$ and the notion of an (ϵ, δ)-strong solution under the Euclidean setup.

Corollary 3.9. *Suppose that $\|\cdot\| = \|\cdot\|_2$ and $\omega(x) = \|\cdot\|_2^2/2$, and also assume that $F(\cdot)$ is Lipschitz continuous (i.e., condition (2.5) holds). Then, if, for some $x \in X$ and $\gamma > 0$, the point x^+ given by (3.17) satisfies $\|R_\gamma(x)\| \le \epsilon$, it must be an $((L\gamma + 1)\epsilon, 0)$-strong solution. In particular, if $\gamma \le 1/L$, then the point x^+ must be an $(2\epsilon, 0)$-strong solution of* $\mathrm{VI}(X, F)$.

Proof. The result directly follows from Proposition 3.8.b) and the facts that $\nu = 1$ and $\mathcal{Q} = 1$. ∎

It is interesting to note that under the conditions given in Corollary 3.9, if $\gamma < 1/L$ and $\|R_\gamma(x)\| \le \epsilon/2$ then the solution given (3.17) must be an $(\epsilon, 0)$-strong solution. Such a solution is stronger than an (ϵ, δ)-strong solution with $\delta > 0$ as it does not depend on the second tolerance δ. For example, in the context of complementarity problems, the complementarity slackness constraint will be satisfied exactly by an $(\epsilon, 0)$-strong solution.

4. VI problems with Lipschitz or Hölder continuous operators.

Our main goal in this section is to establish the complexity of a non-Euclidean extragradient (N-EG) method for solving the GMVI problems discussed in Section 2. We assume throughout this section that the operator $F(\cdot)$ is either Lipschitz or, more generally, Hölder continuous.

The extragradient method is a classical method for solving VI problems that was initially proposed by Korpelevich [14]. While earlier studies on Korpelevich's extragradient method or its variants were focused on their asymptotical convergence behaviour (see, e.g., [29, 30, 32]), the complexity analysis of these types of methods has only appeared recently in the literature [21, 19]. More specifically, Nemirovski [21] established the complexity of a generalized version of Korpelevich's extragradient method for computing a weak solution of $\mathrm{VI}(X, F)$ and showed that one can possibly improve its performance by replacing the projection step with the prox-mapping defined in (3.3). Some of these results were generalized by Auslender and Teboulle [1] in their interior projection methods for monotone variational inequalities. More recently, Monteiro and Svaiter [19] studied the complexity of the original Korpelevich's extragradient method (under a more general framework). Most of these previous studies need to assume the operator $F(\cdot)$ to be monotone and Lipschitz continuous. To the best of our knowledge, the complexity of extragradient-type methods for solving more general VI problems (e.g., the operator $F(\cdot)$ is pseudo-monotone and has different levels of continuity) has never been studied in the literature. It is worth noting that under this general setting, the notion of a weak solution is not useful any more and one has to resort to the notions of approximate strong solutions as discussed in Section 3. Moreover, it is unclear how one can benefit from taking the prox-mapping (rather than metric projection) in the extragradient method for the computation of strong solutions to $\mathrm{VI}(X, F)$.

The non-Euclidean extragradient (N-EG) method for GMVI:

Input: Initial point $x_1 \in X$ and stepsizes $\{\gamma_k\}_{k\ge1}$.

0) Set $k = 1$.
1) Compute

$$y_k = P_{x_k}(\gamma_k F(x_k)), \tag{4.1}$$
$$x_{k+1} = P_{x_k}(\gamma_k F(y_k)). \tag{4.2}$$

2) Set $k = k + 1$ and go to Step 1.

We now add a few remarks about the above N-EG method. Firstly, observe that under the Euclidean case when $\|\cdot\| = \|\cdot\|_2$ and $\omega(x) = \|x\|^2/2$, the computation of (y_t, x_t), $t \geq 1$, is the same as Korpelevich's extragradient or Euclidean extragradient (E-EG) method. Secondly, it should be noted that, while the above N-EG method is similar to Nemirovski's mirror-prox method for solving monotone VI problems, its convergence analysis and the specification of the algorithmic parameters (e.g., $\|\cdot\|$, $\omega(\cdot)$ and γ_k) differ significantly from those in [21], since we are dealing with a much wider class of problems and using different termination criteria.

In order to establish the convergence properties of the above N-EG method, we first need to show a few technical results.

Let $p(u)$ be a convex function over a convex set $X \in \mathbb{R}^n$. Assume that u^* is an optimal solution of the problem $\min\{p(u) + \|u - \tilde{x}\|^2 : u \in X\}$ for some $\tilde{x} \in X$. Due to the well-known fact that the sum of a convex and a strongly convex function is also strongly convex, one can easily see that

$$p(u) + \|u - \tilde{x}\|^2 \geq \min\{p(v) + \|v - \tilde{x}\|^2 : v \in X\} + \|u - u^*\|^2.$$

The next lemma generalizes this result to the case where the function $\|u - \tilde{x}\|^2$ is replaced with the prox-function $V(\tilde{x}, u)$ associated with a convex function ω. It can be viewed as a Bregman version of "growth formula" for strongly convex functions and is based on a Pythagora like formula for Bregman distances. The proof of this result can be found, e.g., in Lemma 1 of [15] and Lemma 6 of [16].

Lemma 4.1. *Let X be a convex set in $\mathbb{R}^n$ and $p, \omega : X \to \mathbb{R}$ be differentiable convex functions. Assume that u^* is an optimal solution of* $\min\{p(u) + V(\tilde{x}, u) : u \in X\}$. *Then,*

$$p(u^*) + V(\tilde{x}, u^*) + V(u^*, u) \leq p(u) + V(\tilde{x}, u), \quad \forall u \in X.$$

With this result, we can show an important recursion of the N-EG method for VI(X, F). More specifically, let $x^* \in X^*$ be an optimal solution, the next result describes how the the distance $V(x_k, x^*)$ decreases at each iteration of the N-EG method.

Lemma 4.2. *Let $x_1 \in X$ be given and the pair $(y_k, x_{k+1}) \in X \times X$ be computed according to (4.1)-(4.2). Also let X^* denote the solution set of* VI(X, F). *Then, the following statements hold:*

a) There exists $x^* \in X^*$ such that

$$-\frac{\gamma_k^2}{2\alpha}\|F(x_k) - F(y_k)\|_*^2 + V(x_k, y_k) \leq V(x_k, x^*) - V(x_{k+1}, x^*); \tag{4.3}$$

b) If $F(\cdot)$ is Hölder continuous (i.e., condition (2.6) holds), then there exists $x^* \in X^*$ such that for any $\nu \in (0, 1]$,

$$V(x_k, y_k) - 2^{\nu-1} L^2 \gamma_k^2 \alpha^{-(1+\nu)} \left[V(x_k, y_k)\right]^\nu \leq V(x_k, x^*) - V(x_{k+1}, x^*). \tag{4.4}$$

In particular, if $F(\cdot)$ is Lipschitz continuous (i.e., condition (2.5) holds), then we have

$$\left(1 - L^2\gamma_k^2\alpha^{-2}\right) V(x_k, y_k) \leq V(x_k, x^*) - V(x_{k+1}, x^*). \tag{4.5}$$

Proof. We first show part a). By (4.1) and Lemma 4.1 (with $p(\cdot) = \gamma_k\langle F(x_k), \cdot\rangle$, $\tilde{x} = x_k$ and $u^* = y_k$), we have

$$\gamma_k\langle F(x_k), y_k - x\rangle + V(x_k, y_k) + V(y_k, x) \leq V(x_k, x), \ \forall x \in X.$$

Letting $x = x_{k+1}$ in the above inequality, we obtain

$$\gamma_k \langle F(x_k), y_k - x_{k+1} \rangle + V(x_k, y_k) + V(y_k, x_{k+1}) \le V(x_k, x_{k+1}) \tag{4.6}$$

Moreover, by (4.2) and Lemma 4.1 (with $p(\cdot) = \gamma_k \langle F(y_k), \cdot \rangle$, $\tilde{x} = x_k$ and $u^* = x_{k+1}$), we have

$$\gamma_k \langle F(y_k), x_{k+1} - x \rangle + V(x_k, x_{k+1}) + V(x_{k+1}, x) \le V(x_k, x), \ \forall x \in X.$$

Replacing $V(x_k, x_{k+1})$ in the above inequality with the bound in (4.6) and noting that $\langle F(y_k), x_{k+1} - x \rangle = \langle F(y_k), y_k - x \rangle - \langle F(y_k), y_k - x_{k+1} \rangle$, we have

$$\gamma_k \langle F(y_k), y_k - x \rangle + \gamma_k \langle F(x_k) - F(y_k), y_k - x_{k+1} \rangle + V(x_k, y_k) + V(y_k, x_{k+1}) + V(x_{k+1}, x) \le V(x_k, x),$$

which, in view of Assumption A1, then implies that

$$\gamma_k \langle F(x_k) - F(y_k), y_k - x_{k+1} \rangle + V(x_k, y_k) + V(y_k, x_{k+1}) + V(x_{k+1}, x^*) \le V(x_k, x^*), \tag{4.7}$$

In order to show (4.3), we only need to bound the left hand side of (4.7). By using Cauchy Schwarz inequality and (3.4), we have

$$\begin{aligned}
&\gamma_k \langle F(x_k) - F(y_k), y_k - x_{k+1} \rangle + V(x_k, y_k) + V(y_k, x_{k+1}) \\
&\ge -\gamma_k \|F(x_k) - F(y_k)\|_* \|y_k - x_{k+1}\| + V(x_k, y_k) + V(y_k, x_{k+1}) \\
&\ge -\gamma_k \|F(x_k) - F(y_k)\|_* \left[\frac{2}{\alpha} V(y_k, x_{k+1}) \right]^{\frac{1}{2}} + V(x_k, y_k) + V(y_k, x_{k+1}) \\
&\ge -\frac{\gamma_k^2}{2\alpha} \|F(x_k) - F(y_k)\|_*^2 + V(x_k, y_k),
\end{aligned}$$

where the last inequality follows from Young's inequality. Combining the above observation with (4.7), we arrive at relation (4.3). Now, it follows from the assumption (2.6) and (3.4) that

$$\|F(x_k) - F(y_k)\|_*^2 \le L^2 \|x_k - y_k\|^{2\nu} \le L^2 \left[\frac{2}{\alpha} V(y_k, x_k) \right]^{\nu}.$$

Combining the previous observation with (4.3), we obtain (4.4). Relation (4.5) immediately follows from (4.4) with $\nu = 1$. ■

We are now ready to establish the complexity of the N-EG method for solving GMVI problems. We start with the relatively easier case when $F(\cdot)$ is Lipschitz continuous.

THEOREM 4.3. *Suppose that $F(\cdot)$ is Lipschitz continuous (i.e., condition (2.5) holds) and that the stepsizes γ_k are set to*

$$\gamma_k = \frac{\alpha}{\sqrt{2}L}, \quad k \ge 1. \tag{4.8}$$

Also let $R_\gamma(\cdot)$, $g(\cdot)$, $\tilde{g}(\cdot, \cdot)$ and $\phi_\gamma(\cdot)$ be defined in (3.13), (3.15), (3.16) and (3.18), respectively.

a) For any $k \in \mathbb{N}$, there exists $i \le k$ such that

$$\|R_{\gamma_i}(x_i)\|^2 \le \frac{8L^2}{\alpha^3 k} V(x_1, x^*), \quad k \ge 1; \tag{4.9}$$

b) If $\omega(\cdot)$ has $\mathcal{Q}$-Lipschitz continuous gradients w.r.t. $\|\cdot\|$, then for every $k \in \mathbb{N}$, there exists $i \le k$ such that

$$\tilde{g}(y_i, \phi_{\gamma_i}(x_i)) \le 0 \quad \textit{and} \quad \|\phi_{\gamma_i}(x_i)\|_* \le \frac{2L(\alpha + \sqrt{2}\mathcal{Q})}{\alpha^{3/2}} \sqrt{\frac{V(x_1, x^*)}{k}}; \tag{4.10}$$

c) If, in addition, the set X is bounded, then for every $k \in \mathbb{N}$, there exists $i \le k$ such that

$$g(y_i) \le \frac{2\sqrt{2}L\,(\alpha + \sqrt{2}\mathcal{Q})\,\Omega^2_{\omega,X}}{\alpha\sqrt{k}}, \tag{4.11}$$

where $\Omega_{\omega,X}$ is defined in (3.10).

Proof. Using (4.5) and (4.8), we have

$$\frac{1}{2}V(x_k, y_k) \le V(x_k, x^*) - V(x_{k+1}, x^*), \quad k \ge 1.$$

Also it follows from (3.4) and definition (3.13) that

$$V(x_k, y_k) \ge \frac{\alpha}{2}\|x_k - y_k\|^2 = \frac{\alpha\gamma_k^2}{2}\|R_{\gamma_k}(x_k)\|^2. \tag{4.12}$$

Combining the above two observations, we obtain

$$\gamma_k^2\|R_{\gamma_k}(x_k)\|^2 \le \frac{4}{\alpha}\left[V(x_k, x^*) - V(x_{k+1}, x^*)\right], k \ge 1.$$

By summing up these inequalities we arrive at

$$\sum_{i=1}^{k}\gamma_i^2 \min_{i=1,\ldots,k}\|R_{\gamma_i}(x_i)\|^2 \le \sum_{i=1}^{k}\gamma_i^2\|R_{\gamma_i}(x_i)\|^2 \le \frac{4}{\alpha}V(x_1, x^*), \quad k \ge 1,$$

which implies that

$$\min_{i=1,\ldots,k}\|R_{\gamma_i}(x_i)\|^2 \le \frac{4}{\alpha\sum_{i=1}^{k}\gamma_i^2}V(x_1, x^*). \tag{4.13}$$

Using the above inequality and (4.8), we obtain the bound in (4.9). Part b) directly follows from Proposition 3.8.a) and b) and bound (4.9). Moreover, Part c) follows from Proposition.c), bound (4.9) and the definition of $\Omega_{X,\omega}$ in (3.10). ∎

We now add a few comments about the results obtained in Theorem 4.3. Firstly, in view of Theorem 4.3.b), under the Euclidean setup where $\|\cdot\| = \|\cdot\|_2$ and $\omega(x) = \|x\|_2^2/2$ (hence $\alpha = 1$), the error bounds obtained in (4.10) is stronger than the corresponding ones in Theorem 5.2 of [19] in the following sense: (a) the bound is applicable to a more general class of VI problems, i.e., the GMVI problems; and (b) the second residual δ used in the notion of an (ϵ, δ)-strong solution associated with the sequence $\{y_k\}$ always vanishes as opposed to the one obtained in [19]. Secondly, our results apply also to the non-Euclidean setup where $\omega(\cdot)$ is not necessarily $\|\cdot\|_2^2/2$. This can lead to more efficient variants of the N-EG method for solving large-scale GMVI problems as demonstrated in Section 6.

In the next result, we discuss the convergence properties of the N-EG method for solving GMVI problems with Hölder continuous operators. For the sake of simplicity, we assume here that the number of iterations k is fixed a priori.

THEOREM 4.4. *Suppose that $F(\cdot)$ is Hölder continuous (i.e., condition (2.6) holds) and that the stepsizes $\gamma_i, i = 1, \ldots, k$, in the N-EG method are set to*

$$\gamma_i = \frac{\alpha^{\frac{1+\nu}{2}}}{L(2\nu)^{\frac{\nu}{2}}}\left(\frac{1}{k}\right)^{\frac{1-\nu}{2}}, \tag{4.14}$$

where k be the number of iterations given a priori. Also let $R_\gamma(\cdot)$, $g(\cdot)$, $\tilde{g}(\cdot,\cdot)$ and $\phi_\gamma(\cdot)$ be defined in (3.13), (3.15), (3.16) and (3.18), respectively.

a) There exists $i \le k$ such that

$$\|R_{\gamma_i}(x_i)\|^2 \le \frac{8L^2}{\alpha^{2+\nu} k^\nu}[1 + V(x_1, x^*)]; \tag{4.15}$$

b) If $\omega(\cdot)$ has $\mathcal{Q}$-Lipschitz continuous gradients w.r.t. $\|\cdot\|$, then there exists $i \le k$ such that

$$\tilde{g}(y_i, \phi_{\gamma_i}(x_i)) \le 0 \quad \text{and} \quad \|\phi_{\gamma_i}(x_i)\|_* \le \frac{2\sqrt{2} C_\alpha L}{(\alpha k)^{\frac{\nu}{2}}}[1 + V(x_1, x^*)]^{\frac{1}{2}}, \tag{4.16}$$

where $C_\alpha = 4 + \mathcal{Q}/\alpha$;

c) If, in addition, the set X is bounded, then there exists $i \le k$ such that

$$g(y_i) \le \frac{4\sqrt{2} C_\alpha L \Omega_{\omega,X}}{(\alpha k)^{\frac{\nu}{2}}}[1 + V(x_1, x^*)]^{\frac{1}{2}}, \tag{4.17}$$

where $\Omega_{\omega,X}$ is defined in (3.10).

Proof. Since the case when $\nu = 1$ has been shown in Theorem 4.3, we assume that $\nu \in (0, 1)$. After rearranging the terms in (4.4), we obtain, $\forall i = 1, \ldots, k$,

$$\frac{1}{2} V(x_i, y_i) \le V(x_i, x^*) - V(x_{i+1}, x^*) + \underbrace{2^{\nu-1} L^2 \gamma_i^2 \alpha^{-(1+\nu)} [V(x_i, y_i)]^\nu - \frac{1}{2} V(x_i, y_i)}_{\Delta_i}. \tag{4.18}$$

Note that, in view of the observation that

$$\max_{d \ge 0} \{a\, d^\nu - d/2\} \le a(2\nu a)^{-\frac{\nu}{\nu-1}} = \alpha^{-\frac{1}{\nu-1}}(2\nu)^{-\frac{\nu}{\nu-1}}, \quad \forall\, a > 0,\, \nu \in (0, 1),$$

and relation (4.14), we have

$$\Delta_i \le \frac{1}{2}\left[L^2 \gamma_i^2 \alpha^{-(1+\nu)}\right]^{-\frac{1}{\nu-1}} (2\nu)^{-\frac{\nu}{\nu-1}} = \frac{1}{2k} \le \frac{1}{k}, \quad i = 1, \ldots, k. \tag{4.19}$$

Moreover, it follows from (4.12) and (4.14) that

$$V(x_i, y_i) \ge \frac{\alpha \gamma_i^2}{2} \|R_{\gamma_i}(x_i)\|^2 = \frac{\alpha^{2+\nu} k^{\nu-1}}{2L^2 (2\nu)^\nu} \|R_{\gamma_i}(x_i)\|^2.$$

Using the previous two bounds in (4.18), we conclude

$$\frac{\alpha^{2+\nu} k^{\nu-1}}{4L^2 (2\nu)^\nu} \|R_{\gamma_i}(x_i)\|^2 \le V(x_i, x^*) - V(x_{i+1}, x^*) + \frac{1}{k}, \quad i = 1, \ldots, k. \tag{4.20}$$

Summing up the above inequalities and using the fact that $k \min_{i=1,\ldots,k} \|R_{\gamma_i}(x_i)\|^2 \le \sum_{i=1}^k \|R_{\gamma_i}(x_i)\|^2$, we have

$$\frac{\alpha^{2+\nu} k^\nu}{4L^2 (2\nu)^\nu} \min_{i=1,\ldots,k} \|R_{\gamma_i}(x_i)\|^2 \le V(x_1, x^*) - V(x_{N+1}, x^*) + 1 \le V(x_1, x^*) + 1,$$

which clearly implies (4.15) in view of the fact that $(2\nu)^\nu \le 2$.

We now show part b). The first inequality of (4.16) follows directly from Proposition 3.8.a) and the definitions of x_i and y_i. Note that by (3.19) and the fact that $\gamma_1 = \gamma_2 = \ldots = \gamma_k$ due to (4.14), we have

$$\begin{aligned}\min_{i=1,\ldots,k} \|\phi_{\gamma_i}(x_i)\| &\le \min_{i=1,\ldots,k} \{L\,(\gamma_i \|R_{\gamma_i}(x_i)\|)^\nu + \mathcal{Q}\|R_{\gamma_i}(x_i)\|\} \\ &\le L\left(\gamma_k \min_{i=1,\ldots,k} \|R_{\gamma_i}(x_i)\|\right)^\nu + \mathcal{Q} \min_{i=1,\ldots,k} \|R_{\gamma_i}(x_i)\|,\end{aligned}$$

which together with (4.14) and (4.15) then imply that

$$\begin{aligned}\min_{i=1,\ldots,k} \|\phi_{\gamma_i}(x_i)\| &\le \frac{L}{k^{\frac{\nu}{2}}}\left[\left(\frac{2\sqrt{2}(1+V(x_1,x^*))^{\frac{1}{2}}}{(2\nu)^{\frac{\nu}{2}}\alpha^{\frac{1}{2}}}\right)^{\nu} + \frac{2\sqrt{2}\mathcal{Q}(1+V(x_1,x^*))^{\frac{1}{2}}}{\alpha^{\frac{2+\nu}{2}}}\right] \\ &\le \frac{2\sqrt{2}L(1+V(x_1,x^*))^{\frac{1}{2}}}{k^{\frac{\nu}{2}}}\left[\frac{1}{(2\nu)^{\frac{\nu^2}{2}}\alpha^{\frac{\nu}{2}}} + \frac{\mathcal{Q}}{\alpha^{\frac{2+\nu}{2}}}\right] \\ &\le \frac{2\sqrt{2}L(1+V(x_1,x^*))^{\frac{1}{2}}}{(\alpha k)^{\frac{\nu}{2}}}\left(4+\frac{\mathcal{Q}}{\alpha}\right),\end{aligned}$$

where the last two inequalities follow from the facts that $\nu \in (0,1]$ and that

$$(2\nu)^{\frac{\nu^2}{2}} \ge \nu^{\frac{\nu^2}{2}} \ge \nu^{\frac{\nu}{2}} \ge e^{-\frac{1}{2e}} \ge \frac{1}{4}.$$

Part c) follows from (3.20) and an argument similar to the one used in the proof of part b). ■

A few remarks about the results obtained in Theorem 4.4 are in place. Firstly, it seems possible to relax the assumption that the number of iterations k is given a priori. For example, if we set

$$\gamma_i = \frac{\alpha^{\frac{1+\nu}{2}}}{L(2\nu)^{\frac{\nu}{2}}}\left(\frac{1}{i}\right)^{\frac{1-\nu}{2}}, \quad i = 1,2,\ldots, \tag{4.21}$$

we can show essentially the same rate of convergence as the one obtained in (4.17) when the set X is bounded, and slightly worse (with an additional logarithmic factor $\log(1/k)$) convergence rate than those stated in (4.15) and (4.16) and (4.17) if the set X is unbounded.

Secondly, according to Theorem 4.4.b), if $F(\cdot)$ is Hölder continuous, an $(\epsilon, 0)$-strong solution of VI(X,F) can be computed in at most

$$\mathcal{O}\left\{\frac{V(x_1,x^*)^{\frac{1}{\nu}}}{\alpha^{\frac{\nu+2}{\nu}}}\left(\frac{\mathcal{Q}L}{\epsilon}\right)^{\frac{2}{\nu}}\right\}$$

iterations. This complexity result appears to be new also for the case when $F(\cdot)$ is monotone and the distance generating function $\omega(\cdot) = \|\cdot\|_2^2/2$.

Thirdly, the results in Theorem 4.4 indicate how the rates of convergence of the N-EG method depend on the continuity assumption about $F(\cdot)$. In view of Theorem 4.4, if $F(\cdot)$ is continuous but not necessarily Hölder continuous, i.e., $\nu = 0$, the sequence $\{y_k\}$ generated by the N-EG method will not converge to any solutions of VI(X,F). We will address this issue in next section to deal with GMVI problems with much weaker continuity assumptions on their operators.

5. VI problems with general continuous operators. In this section, we consider GMVI problems where the operator $F(\cdot)$ is continuous but not necessarily Lipschitz or Hölder continuous. Our goal is to show that the N-EG method, after incorporating a simple line search procedure, can generate a sequence of solutions converging to the optimal one of VI(X,F) under this more general setting. More specifically, we study the conditions on the continuity assumption of $F(\cdot)$ and those on the distance generating function $\omega(\cdot)$ in order to guarantee the convergence of these types of N-EG methods applied to more general GMVI problems.

It should be noted that there exist a few earlier developments (e.g., Sun [30], Solodov and Svaiter [29]) that generalized Korpelevich's extragradient method for solving GMVI problems. However, to the best of our knowledge, there does not exist any non-Euclidean extragradient type methods for solving GMVI problems. It should be noted that, while earlier extragradient type methods for GMVI problems (e.g., [29, 30]) rely on a certain monotonicity property of the metric projection (e.g., Gafni and Bertsekas [8]), in general such a property is not assumed by the prox-mapping.

A related but different work to ours is due to Auslender and Teboulle [1]. In [1], Auslender and Teboulle studied the generalization of the mirror-prox method for solving monotone and locally Lipschitz continuous VI problems by incorporating a line search procedure. More specifically, they show the convergence of their methods under the assumption that the prox-function $V(\cdot,\cdot)$ is quadratic. Our development significantly differs from [1] in the following aspects: i) we deal with VI problems with generalized monotone (e.g., pseudo-monotone) operators; ii) we use more general prox-functions which are not necessarily quadratic; and iii) we consider VI problems with general continuous operators, in addition to those with locally Lipschitz continuous operators.

We are now ready to describe a variant of the N-EG method obtained by incorporating a simple linear search procedure for solving VI problems with general continuous operators.

The N-EG method with line search (N-EG-LS):

Input: Initial point $x_1 \in X$, initial stepsize $\gamma_0 \in (0,1)$ and $\lambda \in (0,1)$.

0) Set $k = 1$;
1) Compute $R_{\gamma_0}(x_k)$. If $\|R_{\gamma_0}(x_k)\| = 0$ **terminate** the algorithm. Otherwise, choose the γ_k with the largest value from the list $\{\gamma_0, \gamma_0\lambda, \gamma_0\lambda^2, \ldots\}$ such that

$$\|F(x_k) - F(y_k)\|_*^2 \le \frac{\alpha}{\gamma_k^2} V(x_k, y_k), \tag{5.1}$$

where $y_k = P_{x_k}(\gamma_k F(x_k))$;

3) Compute

$$x_{k+1} = P_{x_k}(\gamma_k F(y_k)); \tag{5.2}$$

4) Set $k = k+1$ and go to Step 1.

In order to establish the convergence of the above N-EG-LS method, we first need to show certain properties of the line-search procedure used in Step 1 of this algorithm. In the remaining part of this section, we say that *the line search procedure is well-defined* if the following two conditions hold:

a) The line search procedure will terminate (i.e., condition (5.1) will be satisfied) after a finite number of steps in choosing γ_k from the list $\{\gamma_0, \gamma_0\lambda, \ldots\}$ for any $k \ge 1$;
b) There exists $K \in \mathbb{N}$ and $\gamma^* > 0$ such that

$$\gamma_k \ge \gamma^* \quad \forall k \ge K. \tag{5.3}$$

Note that condition b) is used to show the convergence of the sequence $\{x_k\}$ (c.f., Lemma 3.5).

Traditionally, the well-definedness of the line search procedure was established by using certain important properties of the metric projection. Specifically, let us denote, for any $x \in \mathbb{R}^n$ and $d \in \mathbb{R}^n$,

$$\theta(\beta) := \frac{\|\Pi_X(x+\beta d) - x\|_2}{\beta}, \quad \beta > 0. \tag{5.4}$$

Then, it is shown by Gafni and Bertsekas [8] that the function $\theta(\beta)$ is monotonically nonincreasing with respect to β. In Proposition 5.2, we show that, if $F(\cdot)$ is a general continuous operator, the line-search procedure is well-defined under a much weaker assumption than the aforementioned the monotonicity of $\theta(\beta)$. Moreover, we present a sufficient condition on $\omega(\cdot)$ which can guarantee that the above assumption is satisfied. In addition, we show in Proposition 5.3 that we do not need any conditions similar to (5.4) when $F(\cdot)$ is locally Lipschitz continuous. The following well-known property of the prox-mapping (e.g., Lemma 2.1 in [21]) is used in the proof of Proposition 5.2.

Lemma 5.1. *Let $x \in X$ be given. we have*

$$\|P_x(\phi_1) - P_x(\phi_2)\| \le \alpha^{-1}\|\phi_1 - \phi_2\|_*, \quad \forall\, \phi_1, \phi_2 \in \mathbb{R}^n. \tag{5.5}$$

Proposition 5.2. *Suppose that $F(\cdot)$ is continuous and also assume that there exists $q > 0$ such that, for any $x \in X$ and $\phi \in \mathbb{R}^n$,*

$$\frac{V(x, P_x(\gamma\phi))}{\gamma^2} \leq \frac{qV(x, P_x(\beta\phi))}{\beta^2}, \quad \gamma \geq \beta > 0. \tag{5.6}$$

Then, the line search procedure in Step 1 of the N-EG-LS method is well-defined. In particular, if the distance generating function $\omega(\cdot)$ has $\mathcal{Q}$-Lipschitz continuous gradients w.r.t. $\|\cdot\|$, then relation (5.6) holds with $q = 1 + \mathcal{Q}^2/\alpha^2$, where α is the modulus of $\omega(\cdot)$.

Proof. Suppose first that condition (5.6) holds. Consider an arbitrary iteration k, $k \geq 1$. Let us denote $\gamma_{kj} := \gamma_0\lambda^j$ and $y_{kj} := P_{x_k}(\gamma_{kj}F(x_k))$, $j \geq 0$. Observe that $\|R_{\gamma^0}(x_k)\| > 0$ whenever the line search procedure occurs. Using this observation, (3.4) and (3.13), we have

$$\frac{V(x_k, y_{k0})}{(\gamma_0)^2} \geq \frac{\alpha}{2(\gamma_0)^2}\|x_k - y_{k0}\|^2 = \frac{\alpha}{2}\|R_{\gamma_0}(x_k)\|^2 > 0. \tag{5.7}$$

The above inequality together with (5.6) then imply that

$$\frac{V(x_k, y_{kj})}{\gamma_{kj}^2} \geq \frac{V(x_k, y_{k0})}{q(\gamma_0)^2} \geq \frac{\alpha}{2q}\|R_{\gamma_0}(x_k)\|^2 > 0, \quad \forall\, j \geq 1.$$

Assume for contradiction that the line search procedure does not terminate in a finite number of steps. Then, we have

$$\|F(x_k) - F(y_{kj})\|_*^2 > \frac{\alpha V(x_k, y_{kj})}{\gamma_{kj}^2}, \forall j \geq 1.$$

It then follows from the above two inequalities that

$$\|F(x_k) - F(y_{kj})\|_*^2 > \frac{\alpha^2}{2q}\|R_{\gamma_0}(x_k)\|^2 > 0, \forall j \geq 1. \tag{5.8}$$

On the other hand, using the Lipschitz continuity of the prox-mapping (see (5.5)), and the fact that $\lim_{j\to+\infty}\gamma_{kj} = 0$, we have $\lim_{j\to+\infty}\|x_k - y_{kj}\| = 0$. This observation, in view of the fact that $F(\cdot)$ is continuous, then imply that $\lim_{j\to+\infty}\|F(x_k) - F(y_{kj})\|_*^2 = 0$, which clearly contradicts with (5.8). Hence, the line search procedure must terminate in a finite number of steps.

We now show that there exists $K \in \mathbb{N}$ and $\gamma^* > 0$ such that (5.3) holds. Assume for contradiction that $\lim_{k\to+\infty}\gamma_k = 0$. Let us denote $\hat{x}_k := P_{x_k}(\beta^{-1}\gamma_k F(x_k))$. By the choice of γ_k, we know that (5.1) is not satisfied for $y_k = \hat{x}_k$, hence we have

$$\|F(x_k) - F(\hat{x}_k)\|_*^2 > \frac{\alpha}{(\beta^{-1}\gamma_k)^2}V(x_k, \hat{x}_k) \geq \frac{\alpha}{q(\gamma_0)^2}V(x_k, \hat{x}_k) \geq \frac{\alpha^2}{2q(\gamma_0)^2}\|R_{\gamma_0}(x_k)\|^2 > 0, \quad k \geq 1. \tag{5.9}$$

where the second inequality is due to (5.6). Using the Lipschitz continuity of the prox-mapping (see (5.5)), and the assumption that $\lim_{k\to+\infty}\gamma_k = 0$, we have $\lim_{k\to+\infty}\|x_k - \hat{x}_k\| = 0$. This observation, in view of the fact that $F(\cdot)$ is continuous, then imply that $\lim_{k\to+\infty}\|F(x_k) - F(\hat{x}_k)\|_*^2 = 0$, which clearly contradicts with (5.9).

We now show that relation (5.6) holds if $\omega(\cdot)$ has $\mathcal{Q}$-Lipschitz continuous gradients. Denote $x_\gamma^+ \equiv P_x(\gamma\phi)$, $x_\beta^+ \equiv P_x(\beta\phi)$. It follows from Lemma 4.1 (with $p(\cdot) = \gamma\langle\phi, \cdot\rangle$, $\tilde{x} = x$ and $u^* = x_\gamma^+$) that

$$\gamma\langle\phi, x_\gamma^+ - z\rangle + V(x, x_\gamma^+) + V(x_\gamma^+, z) \leq V(x, z), \ \forall z \in X.$$

Letting $z = x_\beta^+$ in the above relation, we have

$$V(x, x_\beta^+) - V(x, x_\gamma^+) \geq \gamma\langle\phi, x_\gamma^+ - x_\beta^+\rangle + V(x_\gamma^+, x_\beta^+),$$

which implies that

$$
\begin{aligned}
q\gamma^2 V(x, x_\beta^+) - \beta^2 V(x, x_\gamma^+) &= (q\gamma^2 - \beta^2)V(x, x_\beta^+) + \beta^2[V(x, x_\beta^+) - V(x, x_\gamma^+)] \\
&\geq (q\gamma^2 - \beta^2)V(x, x_\beta^+) + \beta^2\gamma\langle \phi, x_\gamma^+ - x_\beta^+\rangle + \beta^2 V(x_\gamma^+, x_\beta^+) \\
&\geq \left(\frac{\gamma\mathcal{Q}}{\alpha}\right)^2 V(x, x_\beta^+) + \beta^2\gamma\langle \phi, x_\gamma^+ - x_\beta^+\rangle + \beta^2 V(x_\gamma^+, x_\beta^+),
\end{aligned}
$$

where the last inequality follows from the definition of q and the fact that $\gamma \geq \beta$. Also note that by the optimality condition of (3.3), we have

$$
\langle \beta\phi + \nabla\omega(x_\beta^+) - \nabla\omega(x), x_\gamma^+ - x_\beta^+\rangle \geq 0.
$$

Combining the above two conclusions, relations (3.4) and (3.5), we obtain

$$
\begin{aligned}
q\gamma^2 V(x, x_\beta^+) - \beta^2 V(x, x_\gamma^+) &\geq \left(\frac{\gamma\mathcal{Q}}{\alpha}\right)^2 V(x, x_\beta^+) + \beta\gamma\langle \nabla\omega(x_\beta^+) - \nabla\omega(x), x_\beta^+ - x_\gamma^+\rangle + \beta^2 V(x_\gamma^+, x_\beta^+) \\
&\geq \left(\frac{\gamma\mathcal{Q}}{\alpha}\right)^2 V(x, x_\beta^+) - \beta\gamma\|\nabla\omega(x_\beta^+) - \nabla\omega(x)\|_*\|x_\beta^+ - x_\gamma^+\| + \beta^2 V(x_\gamma^+, x_\beta^+) \\
&\geq \frac{\alpha}{2}\left[\left(\frac{\gamma\mathcal{Q}}{\alpha}\right)^2 \|x - x_\beta^+\|^2 - 2\beta\gamma\frac{\mathcal{Q}}{\alpha}\|x_\beta^+ - x\|\|x_\beta^+ - x_\gamma^+\| + \beta^2\|x_\beta^+ - x_\gamma^+\|^2\right] \\
&= \frac{\alpha}{2}\left(\frac{\gamma\mathcal{Q}}{\alpha}\|x - x_\beta^+\| - \beta\|x_\beta^+ - x_\gamma^+\|\right)^2 \geq 0,
\end{aligned}
$$

from which (5.6) immediately follows. ■

The next result identifies certain special cases where we do not need any additional assumptions on the $\omega(\cdot)$ in order to guarantee the well-definedness of the linear search procedure.

PROPOSITION 5.3. *Suppose that $F(\cdot)$ is locally Lipschitz continuous. Then, regardless the choice of $\omega(\cdot)$, the line search procedure in the N-EG-LS method is well-defined. In particular, if $F(\cdot)$ is Lipschitz continuous, then the line search procedure will terminate in at most*

$$
\max\left\{1, \log_{\frac{1}{\lambda}} \frac{\alpha}{\sqrt{2}\gamma_0 L}\right\}
$$

steps. Moreover, in the latter case we have

$$
\gamma_k \geq \min\left\{\frac{\lambda\alpha}{\sqrt{2}L}, \gamma_0\right\}, \ \forall k \geq 1. \tag{5.10}
$$

Proof. Consider the locally Lipschitz continuous case first. Suppose for contradiction that the line search procedure is not well-defined. Then, by (3.4) and (5.1), we must have

$$
\|F(x_k) - F(y_{kj})\|_*^2 > \frac{\alpha V(x_k, y_{kj})}{\gamma_{kj}^2} \geq \frac{\alpha^2\|x_k - y_{kj}\|^2}{2\gamma_{kj}^2}, \forall j \geq 1,
$$

which, in view of (2.7), then implies that $L^2 > \alpha^2/(2\gamma_{kj}^2), j \geq 1$. Tending j to $+\infty$, we have arrived at a contradiction. In order to show that there exists $K \in \mathbb{N}$ and $\gamma^* > 0$ such that (5.3) holds, suppose for contradiction that $\lim_{k\to+\infty}\gamma_k = 0$. Let us denote $\hat{x}_k := P_{x_k}(\beta^{-1}\gamma_k F(x_k))$. By the choice of γ_k, (5.1) and (3.4), we have

$$
\|F(x_k) - F(\hat{x}_k)\|_*^2 > \frac{\alpha}{(\beta^{-1}\gamma_k)^2} V(x_k, \hat{x}_k) \geq \frac{\alpha^2}{2(\beta^{-1}\gamma_k)^2}\|x_k - \hat{x}_k\|^2, \quad k \geq 1, \tag{5.11}
$$

which, in view of (2.7), then implies that $L^2 > \alpha^2/(2(\beta^{-1}\gamma_k)^2)$. Tending k to $+\infty$, we have arrived at a contradiction.

Now consider the Lipschitz continuous case. By (2.5) and (3.4), we have

$$\|F(x_k) - F(y_k)\|_*^2 \le L^2 \|x_k - y_k\|^2 \le \frac{2L^2 V(x_k, y_k)}{\alpha}.$$

Comparing the above inequality with (5.1), we can easily show the last part of the result. ∎

We are now ready to establish the main convergence properties of the above N-EG-LS method applied to GMVI problems with a general continuous operator $F(\cdot)$.

THEOREM 5.4. *Suppose that the line-search procedure is well-defined. Then the sequences $\{x_k\}_{k\ge1}$ and $\{y_k\}_{k\ge1}$ generated by the N-EG-LS method converge to a strong solution of* $\mathrm{VI}(X, F)$*.*

Proof. First note that relation (4.3) still holds for the variant of N-EG method due to our definitions of x_k, y_k and γ_k, $k \ge 1$. Moreover, it follows from the well-definedness of the line search step, relation (5.1) must hold for some $\gamma_k > 0$. Using relations (4.3) and (5.1), we can easily see that, for some $x^* \in X^*$,

$$\frac{1}{2} V(x_k, y_k) \le V(x_k, x^*) - V(x_{k+1}, x^*), \quad k \ge 1. \tag{5.12}$$

Clearly (5.12) implies that the sequence $V(x_k, x^*)$ is nonincreasing. Therefore, it converges. Moreover, the sequence $\{x_k\}$ is bounded. Summing up the inequalities in (5.12), we obtain

$$\frac{1}{2} \sum_{k=1}^{\infty} V(x_k, y_k) \le V(x_1, x^*),$$

which then implies that

$$\lim_{k\to+\infty} V(x_k, y_k) = 0. \tag{5.13}$$

Using these observations, the fact that condition (5.3) holds for some $K \in \mathbb{N}$ and γ^*, and Lemma 3.5, there exists an accumulation point $\tilde{x}$ of $\{x_k\}$ such that $\tilde{x} \in X^*$. We can replace x^* in (5.12) by $\tilde{x}$. Thus the sequence $\{V(x_k, \tilde{x})\}$ converges. Since $\tilde{x}$ is an accumulation point of $\{x_k\}$, it easily follows that $\{V(x_k, \tilde{x})\}$ converges to zero, i.e., $\{x^k\}$ converges to $\tilde{x} \in X^*$. The previous conclusion together with (5.13) then imply the convergence of $\{y^k\}$. ∎

Remark. Observe that we can estimate the rate of convergence of the N-EG-LS method applied to GMVI problems with Lipschitz continuous operators. Indeed, by using (4.13) and (5.10), we have

$$\min_{i=1,\dots,k} \|R_{\gamma_i}(x_i)\|^2 \le \frac{4V(x_1, x^*)}{\alpha k \min\{\lambda^2\alpha^2/2L, \gamma_0^2\}}, k \ge 1.$$

The above bound is slightly worse than the one in (4.9). However, one potential advantage of the N-EG-LS method over the N-EG method is that it does not require the explicit input of the Lipschitz constant L.

6. Numerical Results. In this section, we report preliminary results of our computational experiments where we compare the performance of different variants of the N-EG method discussed in this paper.

6.1. Problem instances. We focus on an important class of VI problems $\mathrm{VI}(X, F)$, where X is the standard simplex given by $X = \{x \in \mathbb{R}^n | \sum_i x_i = 1, x_i \ge 0, i = 1, \dots, n\}$ and F is a continuous. These problems are chosen for the following reasons: i) they have been extensively studied in the literature (e.g., [4, 26, 30, 33, 9]); ii) A set of complexity results for these problems have been developed in this paper; and iii) it is expected that the study on these VI problems with relatively simple feasible set X can shed some light on problems with more complicated feasible set X.

In particular, the following instances have been used in our numerical experiments. Note that for most of these problems, the operator F is not necessarily monotone.

a. Kojima-Shindo (KS) problem
This problem was studied in [26]. The operator $F : \mathbb{R}^4 \to \mathbb{R}^4$ is defined as:

$$F(x) = \begin{bmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{bmatrix}.$$

b. Watson (WAT) problem
The operator $F : \mathbb{R}^{10} \to \mathbb{R}^{10}$ is given by $F(x) = Ax + b$, where

$$A = \begin{pmatrix} 0 & 0 & -1 & -1 & -1 & 1 & 1 & 0 & 1 & 1 \\ -2 & -1 & 0 & 1 & 1 & 2 & 2 & 0 & -1 & 0 \\ 1 & 0 & 1 & -2 & -1 & -1 & 0 & 2 & 0 & 0 \\ 2 & 1 & -1 & 0 & 1 & 0 & -1 & -1 & -1 & 1 \\ -2 & 0 & 1 & 1 & 0 & 2 & 2 & -1 & 1 & 0 \\ -1 & 0 & 1 & 1 & 1 & 0 & -1 & 2 & 0 & 1 \\ 0 & -1 & 1 & 0 & 2 & -1 & 0 & 0 & 1 & -1 \\ 0 & -2 & 2 & 0 & 0 & 1 & 2 & 2 & -1 & 0 \\ 0 & -1 & 0 & 2 & 2 & 1 & 1 & 1 & -1 & 0 \\ 2 & -1 & -1 & 0 & 1 & 0 & 0 & -1 & 2 & 2 \end{pmatrix},$$

$b = e_i$ and e_i is the unit vector. Hence, we have 10 different instances of Watson problem, i.e., WAT1,WAT2, ..., WAT10, obtained by setting $q = e_1, e_2, \ldots, e_{10}$. This problem was studied by Watson in [33].

c. Sun problem
This problem was discussed by Sun in [30] and we consider problems possibly in larger dimension. The operator $F : \mathbb{R}^n \to \mathbb{R}^n$ is given by $F(x) = Ax + b$, where

$$A = \begin{pmatrix} 1 & 2 & 2 & \cdot & \cdot & \cdot & 2 \\ 0 & 1 & 2 & \cdot & \cdot & \cdot & 2 \\ 0 & 0 & 1 & \cdot & \cdot & \cdot & 2 \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & 2 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

and $b = (-1, ..., -1)$. We consider problem instances with dimension n ranging from $8,000$ to $30,000$.

d. Modified HP Hard (MHPH) problem
We modify the Harker's procedure ([9]) to build an affine function $F(x) = Ax + b$, where the positive definite matrix A is randomly generated as $A = MM^T$ (hence the VI problems are monotone). Each entry of the $n \times n$ matrix M is uniformly generated in $(-15, -12)$ and vector b has been uniformly generated in $(-500, 0)$. We generated instances with dimension n ranging from $1,000$ to $8,000$.

e. Randomly generated (RG) instances
We consider an affine function $F(x) = Ax + b$, where each entry of the $n \times n$ matrix A is uniformly generated in $(-50, 150)$ and q is uniformly generated in $(-200, 300)$. We do not know if these VI problems are monotone or not. The dimension n of these problem instances ranges from $1,000$ to $3,000$.

Parameters	Number of projection calls np			
(γ_0, λ)	$n = 1,000$	$n = 3,000$	$n = 5,000$	Total np
(0.2,0.2)	1,746	3,074	9,350	14,170
(0.2,0.4)	2,397	3,069	4,595	10,061
(0.2,0.8)	4,693	7,291	13,293	25,277
(0.4,0.2)	3,307	3,987	5,174	12,468
(0.4,0.4)	1,924	3,457	5,811	11,192
(0.4,0.8)	5,020	7,794	14,529	27,343
(0.8,0.2)	2,056	3,891	4,456	10,403
(0.8,0.4)	2,384	3,349	7,476	13,209
(0.8,0.8)	5,342	8,783	15,285	29,410

Table 6.1

Fine-tuning procedure of E-EG-LS for modified HP Hard problem

	Algorithms								
	E-EG			E-EG-LS†			Algorithm C		
n	k	np	CPU time	k	np	CPU time	k	np	CPU time
20	49,502	99,004	33.650	578	5,198	2.0592	2,622	12,198	4.5396
40	9,069	18,138	6.833	23	205	0.1092	118	515	0.2184
50	124,275	248,550	91.011	85	676	0.2808	73	298	0.1404
70	12,911	25,822	10.203	20	193	0.0936	59	264	0.1404
100	71,321	142,642	285.48	41	384	0.7956	61	287	0.7332
150	43,486	86,972	198.79	29	303	0.7488	89	405	0.9204
200	757,758	1,515,516	3,622.1	504	5,493	14.383	1262	5,967	16.443

†: We use parameters $\gamma_0 = 0.4$, $\lambda = 0.4$ for E-EG-LS.

Table 6.2

E-EG vs. E-EG-LS for modified HP Hard problems

6.2. Euclidean algorithms for GMVI problems. Our first experiments are carried out to compare the two Euclidean extragradient methods, i.e., the E-EG and E-EG-LS method presented in this paper. We also compare these methods with a different method for solving pseudo-monotone VI problems developed by Sun (Algorithm C in [30]).

Note that two parameters $\gamma_0 \in (0, 1)$ and $\lambda \in (0, 1)$ are required for the line search procedure in the N-EG-LS or E-EG-LS methods. We used a simple fine-tuning procedure to determine these parameters which is briefly described as follows: for each group of problems, we choose a smaller set of representative instances and run these algorithms for each pair (totally 9 pairs) of parameters (γ_0, λ) chosen from $\{0.2, 0.4, 0.8\}$. We terminate these algorithm until the value of gap function $g(\cdot)$ falls bellow 10^{-1} (our target accuracy is 10^{-3}) and report the number of calls to the projection (or prox-mapping). For each algorithm, we choose a pair (γ_0, λ) corresponding to the smallest total number of projection calls for this set of representative instances, and then use these parameters for all the instances of the same problem. For example, the results of E-EG-LS method applied to MHPH problem using the aforementioned fine-tuning procedure are reported in Table 1. And, in view of these results, we set $\gamma_0 = 0.2$ and $\lambda = 0.4$ for the E-EG-LS method applied to all instances for the MHPH problem.

Also for the Sun's algorithm, we used the parameters suggested in [30]. All these algorithms were implemented in MATLAB R2009b on a Core i5 3.1 Ghz computer.

We compare the number of iterations k, total number calls of projection np and *CPU time* for the above three algorithms whenever the gap function $g(\cdot)$ evaluated at the point x_k (see (3.15)) falls below 10^{-3}. The results are reported for the HP hard and WAT problems as shown in Table 2 and Table 3, while the results for other problems are similar. We conclude from these results that using the Euclidean setup, the performance of E-EG-LS method is comparable to Sun's method. Moreover, E-EG-LS method can significantly outperform the E-EG method especially when the Lipschitz constant L is big. In next subsection, we will demonstrate how we can improve the performance of the E-EG-LS method by incorporating the non-Euclidean setup.

	Algorithms								
	E-EG			E-EG-LS†			Algorithm C		
INST	k	np	CPU time	k	np	CPU time	k	np	CPU time
WAT1	81	162	0.0936	56	183	0.0936	43	183	0.0936
WAT2	25	50	0.0312	18	55	0.0468	24	105	0.0936
WAT3	-	-	-	-	-	-	548	2,395	0.8892
WAT4	94	188	0.0780	60	192	0.0936	54	225	0.1092
WAT5	26	52	0.0312	18	54	0.0468	21	90	0.0780
WAT6	55	110	0.0468	37	113	0.0936	64	259	0.1248
WAT7	55	110	0.0468	37	113	0.0624	95	342	0.1560
WAT8	53	104	0.0624	31	94	0.0624	29	132	0.0780
WAT9	12	24	0.0312	8	24	0.0312	9	27	0.0624
WAT10	50	100	0.0468	34	102	0.0624	28	125	0.0936

−: indicates that the algorithm diverges and the instance is not a GMVI problem.
†: We use parameters $\gamma_0 = 0.2$, $\lambda = 0.8$ for *Euclidean algorithm*.

TABLE 6.3
E-EG vs. E-EG-LS for WAT instances

	Algorithms								
	Euclidean			p-norm			entropy		
INST	k	np	CPU time	k	np	CPU time	k	np	CPU time
KS †	7	36	0.0312	7	36	0.0312	17	60	0.0312
WAT1‡	56	183	0.0936	48	149	0.0936	77	275	0.0468
WAT2	18	55	0.0468	20	60	0.0624	29	90	0.0156
WAT3	-	-	-	-	-	-	-	-	-
WAT4	60	192	0.0936	73	223	0.1248	34	102	0.0156
WAT5	18	54	0.0468	21	63	0.0624	38	114	0.0156
WAT6	37	113	0.0936	30	90	0.0624	48	144	0.0156
WAT7	37	113	0.0624	36	107	0.0780	44	132	0.0156
WAT8	31	94	0.0624	31	93	0.0624	51	153	0.0312
WAT9	8	24	0.0312	8	24	0.0312	14	42	0.0312
WAT10	34	102	0.0624	29	87	0.0936	39	117	0.0156

−: indicates that the algorithm diverges and the instance is not a GMVI problem.
†: We use parameters $\gamma_0 = 0.2$, $\lambda = 0.4$ for *Euclidean algorithm*, $\gamma_0 = 0.2$, $\lambda = 0.4$ for *p-norm algorithm*,$\gamma_0 = 0.8$, $\lambda = 0.2$ for *entropy algorithm*.
‡: We use parameters $\gamma_0 = 0.2$, $\lambda = 0.8$ for *Euclidean algorithm*, $\gamma_0 = 0.2$, $\lambda = 0.8$ for *p-norm algorithm*,$\gamma_0 = 0.8$, $\lambda = 0.8$ for *entropy algorithm*.

TABLE 6.4
Euclidean vs. non-Euclidean for smaller instances

6.3. Euclidean vs. non-Euclidean algorithms for GMVI problems. In this subsection, we conduct experiments to illustrate how one can improve the performance of the extragradient methods by considering the following different settings: the *p-norm* algorithm with $\|\cdot\| = \|\cdot\|_1$ and $\omega(x) = \|x\|_p^2/2$ with $p = 1 + 1/\ln n$, the *entropy algorithm* with $\|\cdot\| = \|\cdot\|_1$ and $\omega(x) = \sum_i (x_i + \delta/n)\log(x_i + \delta/n)$ with $\delta = 10^{-16}$, as well as the *Euclidean algorithm* with $\|\cdot\| = \|\cdot\|_2$ and $\omega(x) = \|x\|_2^2/2$.

We first compare these algorithms for a set of relatively smaller problem instances (namely the KS and WAT problem). As can be seen from Table 4, since the problem dimensions, $n = 4$ for the KS problem and $n = 10$ for the WAT problem, are very small, we do not observe significant advantages of the non-Euclidean algorithms.

We then consider problems of higher dimension. More specifically, we compare these three methods applied to SUN, MHPH and RG problems with the dimension from $1,000$ to $30,000$ and report the results in Table 5, Table 6 and Table 7 respectively. Clearly, for many instances, the *non-Euclidean algorithms* outperform the *Euclidean algorithms* in terms of the number of projection calls (np) and the total CPU time. Interestingly, *p-norm algorithm* is the fastest and the most stable one among all these algorithms. In particular, for the MHPH instances, the *p-norm algorithm* can be approximately twice faster than *Euclidean algorithm*. For the RG instances, the p-norm algorithm always outperforms the other two algorithms. In particular, it succeeds in solving all the problem instances up to accuracy 10^{-3} within our iteration limit ($100,000$ projection calls), while the other two algorithms failed for quite a few of these instances.

	Algorithms								
	Euclidean†			p-norm‡			entropy§		
n	k	np	CPU time	k	np	CPU time	k	np	CPU time
8,000	24	153	24.133	16	74	23.904	24	73	15.507
10,000	24	153	31.840	17	79	31.949	24	73	18.502
12,000	25	166	45.817	17	79	40.155	25	76	23.104
14,000	26	178	59.312	17	81	47.315	25	76	28.735
16,000	26	178	74.740	17	81	58.547	25	76	35.381
18,000	26	178	91.167	17	81	65.817	25	76	43.852
20,000	26	178	110.355	17	81	75.739	25	76	51.527
22,000	26	178	132.54	17	81	86.175	26	79	61.854
24,000	26	178	153.19	17	81	96.034	26	79	71.495
26,000	26	178	183.08	17	81	108.58	26	79	79.217
28,000	27	192	227.62	17	81	121.23	26	79	93.616
30,000	27	192	251.04	17	81	138.51	26	79	104.24

†: *Euclidean algorithm* parameters $\gamma_0 = 0.4, \lambda = 0.4$.
‡: *p-norm algorithm* parameters $\gamma_0 = 0.2, \lambda = 0.4$.
§: *entropy algorithm* parameters $\gamma_0 = 0.8, \lambda = 0.8$.

Table 6.5
Euclidean vs. non-Euclidean for Sun problem

	Algorithms								
	Euclidean †			p-norm ‡			entropy§		
n	k	np	CPU time	k	np	CPU time	k	np	CPU time
1,000	318	3,868	26.395	113	818	9.969	386	2,609	38.813
1,500	523	6,877	77.657	108	822	16.976	240	1,640	58.926
2,000	200	2,519	41.262	148	1,147	28.877	272	1,863	110.18
2,500	332	4,418	110.82	205	1,604	49.562	391	2,774	257.67
3,000	323	4,341	165.63	685	5,350	246.60	1,189	8,431	1,189.0
3,500	262	3,549	159.73	146	1,150	68.561	256	1,815	345.51
4,000	304	4,173	263.77	521	4,087	254.72	905	6,563	1,576.7
4,500	455	6,454	417.96	328	2,588	199.20	660	4,777	1,468.7
5,000	471	6,730	529.51	749	5914	531.80	1,344	9,577	3,452.8
5,500	562	8,298	745.75	386	3,099	318.62	699	5,235	2,304.4
6,000	495	7,288	729.18	463	3,723	434.43	737	5,448	2,901.1
6,500	429	6,327	704.88	472	3,837	509.09	904	6,426	3,968.6
7,000	398	5,857	779.04	829	6605	976.66	1,586	11,587	8,272.6
7,500	360	5,256	791.16	389	3,064	517.58	707	5,125	4,228.3
8,000	699	10,761	1,769.3	1,248	10,327	1,866.7	2,969	22,200	20,543.0

†: *Euclidean algorithm* parameters $\gamma_0 = 0.2, \lambda = 0.4$.
‡: *p-norm algorithm* parameters $\gamma_0 = 0.2, \lambda = 0.2$.
§: *entropy algorithm* parameters $\gamma_0 = 0.8, \lambda = 0.2$.

Table 6.6
Euclidean vs. non-Euclidean for HP Hard problem

	Algorithms								
	Euclidean†			p-norm‡			entropy§		
n	k	np	CPU time	k	np	CPU time	k	np	CPU time
1,000	250	2,001	13.993	55	482	8.7361	104	497	11.0605
1,500	4,381	35,049	561.82	1,630	14,647	352.56	♯	♯	♯
2,000	1,627	14,017	223.88	791	6,994	208.48	3,170	16,084	822.72
2,500	♯	♯	♯	223	1,963	87.454	♯	♯	♯
3,000	1,046	9,415	307.67	36	313	15.928	409	2.070	264.99

♯: indicates that the number of projection calls $np > 100,000$.
†: *Euclidean algorithm* parameters $\gamma_0 = 0.8, \lambda = 0.2$.
‡: *p-norm algorithm* parameters $\gamma_0 = 0.2, \lambda = 0.4$.
§: *entropy algorithm* parameters $\gamma_0 = 0.2, \lambda = 0.2$.

Table 6.7
Euclidean vs. non-Euclidean for randomly generated instances

7. Conclusion. This paper studies a class of generalized monotone variational inequality (GMVI) problems whose operators are not necessarily monotone (e.g., pseudo-monotone) or Lipschitz continuous. Our main constribution consists of: i) defining proper termination criterion for solving these VI problems; ii) presenting non-Euclidean extragradient (N-EG) methods for computing approximate strong solutions of these problems; iii) demonstrating how the iteration complexities of the N-EG methods depend on the global Lipschitz or Hölder continuity properties for their operators and the smoothness properties for the distance generating function used in the N-EG algorithms; and iv) introducing a variant of the N-EG algorithm by incorporating a simple line-search procedure to deal with problems with more general, not necessarily Hölder continuous operators. Moreover, numerical studies are conducted to illustrate the significant advantages of the developed algorithms over the existing ones for solving large-scale GMVI problems.

REFERENCES

[1] A. Auslender and M. Teboulle. Interior projection-like methods for monotone variational inequalities. *Mathematical Programming*, 104:39–68, 2005.

[2] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16:697–725, 2006.

[3] H.H. Bauschke, J.M. Borwein, and P.L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Controal and Optimization*, 42:596–636, 2003.

[4] A. Ben-Tal and A. S. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, Engineering Applications*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2000.

[5] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, New York, second edition, 1999.

[6] L.M. Bregman. The relaxation method of finding the common point convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Phys.*, 7:200–217, 1967.

[7] F. Facchinei and J. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I and II*. Comprehensive Study in Mathematics. Springer-Verlag, New York, 2003.

[8] E. M. Gafni and D. P. Bertsekas. Two-metric projection methods for constrained optimization. *SIAM Journal on Control and Optimization*, 22:936–964, 1984.

[9] P. T. Harker and J. Pang. A damped-Newton method for the linear complementarity problem. In *Computational solution of nonlinear systems of equations (Fort Collins, CO, 1988)*, volume 26 of *Lectures in Appl. Math.*, pages 265–284. Amer. Math. Soc., Providence, RI, 1990.

[10] P. T. Harker and J. S. Pang. Finite-dimensional variational inequality and complementarity problems: a survey of theory, algorithms, and applications. *Mathematical Programming, Series B*, 48:161–220, 1990.

[11] D.W. Hearn. The gap function of a convex program. *Operations Research Letters*, 1:67–71, 1982.

[12] A. Nedich J. Koshal and U. V. Shanbhag. Multiuser optimization: Distributed algorithms and error analysis. *SIAM Journal on Optimization*, 21:1168–1199, 2011.

[13] K.C. Kiwiel. Proximal minimization methods with generalized bregman functions. *SIAM Journal on Controal and Optimization*, 35:1142–1168, 1997.

[14] G. Korpelevich. The extragradient method for finding saddle points and other problems. *Eknomika i Matematicheskie Metody*, 12:747–756, 1976.

[15] G. Lan. An optimal method for stochastic composite optimization. *Mathematical Programming*, 133(1):365–397, 2012.

[16] G. Lan, Z. Lu, and R. D. C. Monteiro. Primal-dual first-order methods with $\mathcal{O}(1/\epsilon)$ iteration-complexity for cone programming. *Mathematical Programming*, 126:1–29, 2011.

[17] G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. Manuscript, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA, June 2008.

[18] G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order augmented lagrangian methods for convex programming. Technical report, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, USA, September 2013. *Mathematical Programming* (Under second-round review).

[19] R.D.C. Monteiro and B.F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. Manuscript, School of ISyE, Georgia Tech, Atlanta, GA, 30332, USA, March 2009.

[20] R.D.C. Monteiro and B.F. Svaiter. Complexity of variants of tsengs modified f-b splitting and korpelevich's methods for hemi-variational inequalities with applications to saddle-point and convex optimization problems. Manuscript, School of ISyE, Georgia Tech, Atlanta, GA, 30332, USA, June 2010.

[21] A. S. Nemirovski. Prox-method with rate of convergence $o(1/t)$ for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15:229–251, 2005.

[22] A. S. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19:1574–1609, 2009.

[23] A. S. Nemirovski and D. Yudin. *Problem complexity and method efficiency in optimization.* Wiley-Interscience Series in Discrete Mathematics. John Wiley, XV, 1983.

[24] Y. E. Nesterov. *Introductory Lectures on Convex Optimization: a basic course.* Kluwer Academic Publishers, Massachusetts, 2004.

[25] Y. E. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109:319–344, 2007.

[26] J. Pang and S. A. Gabriel. Ne/sqp: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60:295–337, 1993.

[27] M. Sibony. Méthodes itératives pour les équations et inéquations aux dérivées partielles nonlinéares de type monotone. *Calcolo*, 7:65–183, 1970.

[28] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323345, 1999.

[29] M. V. Solodov and B. F. Svaiter. A new projection method for variational inequality problems. *SIAM Journal on Control and Optimization*, 37:765–776, 1999.

[30] D. Sun. A new step-size skill for solving a class of nonlinear projection equations. *Journal of Computational Mathematics*, pages 357–368, 1995.

[31] M. Teboulle. Convergence of proximal-like algorithms. *SIAM Journal on Optimization*, 7:1069–1083, 1997.

[32] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431–446, 2000.

[33] L. T. Watson. Solving the nonlinear complementarity problem by a homotopy method. *SIAM J. Control Optimization*, 17:36–46, 1979.