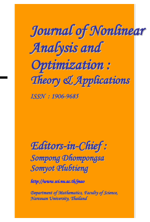
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 6, No. 1, (2015), 41-51
ISSN : 1906-9685
http://www.math.sci.nu.ac.th



A NEW NON-LIPSCHITZIAN PROJECTION METHOD FOR SOLVING VARIATIONAL INEQUALITIES IN EUCLIDEAN SPACES

AVIV GIBALI*

Department of Mathematics,ORT Braude College,21982 Karmiel, Israel

ABSTRACT. The extragradient method introduced by Korpelevich [18] and Antipin [1] is a double projection method designed for solving variational inequalities. The double projection per iteration enable to obtain convergent under monotonicity and Lipschitz continuity while other single projection methods, for example the projected gradient method requires strong monotonicity. The subgradient extragradient method [5] is a modification of the extragradient in which the second projection onto the feasible set is replaced by a projection onto a specific constructible half-space which is actually one of the subgradient half-spaces. Still, this algorithm requires Lipschitz continuity. In this work we introduce a self-adaptive subgradient extragradient method by adopting Armijo-like searches which enables to obtain convergent under the assumption of pseudo-monotonicity and continuity.

KEYWORDS : Extragradient method; Variational inequality; Nonexpansive mapping; Armijo-Goldstein rule
AMS Subject Classification: 65K15 90C25.

1. INTRODUCTION

In this paper, we are concerned with the Variational Inequality Problem (VIP) in the Euclidean space $\mathbb{R}^n$. Let $C \subseteq \mathbb{R}^n$ be a non-empty, closed and convex set and let $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$. The VIP consists in finding a point $x^* \in C$, such that

$$\langle \mathcal{F}(x^*), x - x^* \rangle \geq 0, \text{ for all } x \in C. \tag{1.1}$$

Korpelevich [18] and Antipin [1] proposed an algorithm for solving the VIP, known as the Extragradient Method, see also Facchinei and Pang [11, Chapter 12]. In each iteration in order to get the next iterate x^{k+1}, two orthogonal projections onto C are calculated, according to the following iterative step. Given the current iterate x^k, calculate

$$y^k = P_C(x^k - \tau\mathcal{F}(x^k)), \tag{1.2}$$

$$x^{k+1} = P_C(x^k - \tau\mathcal{F}(y^k)), \tag{1.3}$$

* *Corresponding author.*
Email address : avivg@braude.ac.il.
Article history : Received July 10, 2014. Accepted May 27, 2015.

where τ is some positive number and P_C denotes the Euclidean nearest point projection onto C. Although convergence was proved in [18] under the assumptions of Lipschitz continuity and pseudo-monotonicity, there is still the need to calculate two projections onto the closed convex set C which might seriously affect the efficiency of the algorithm. Censor et al. [5] (see also [6, 7]) presented the Subgradient Extragradient Method (SEM), in which the second projection (1.3) onto C is replaced by a projection onto a specific constructible half-space which is actually one of the subgradient half-spaces. In order to prove convergence the authors assume that $\mathcal{F}$ is monotone on C, Lipschitz continuous on $\mathbb{R}^n$, and the Lipschitz constant L is known, so $\tau \in (0, 1/L)$.

In this paper we present a new modification of the SEM for solving the VIP (1.1) when the mapping $\mathcal{F}$ is assumed to be only continuous instead of Lipschitz. Using an Armijo-Goldstein-type rule ([2]) the step size τ is updated and convergence of the algorithm is then guaranteed under the assumptions of pseudo-monotonicity and continuity of $\mathcal{F}$. Other step size adaptations are also presented and the convergence proof can be obtain by following similar arguments. Our convergence theorem relies on the work of Khobotov [17] and Solodov and Tseng [23].

The paper is organized as follows. In Section 2 we present some preliminaries and definitions that will be needed in the sequel. Later, in Section 3 the new algorithm is presented and its convergence is analyzed. Finally, in Section 4 we illustrate the algorithm performance.

2. PRELIMINARIES

In this section we present some useful definitions and results that will be needed for our convergence theorem.

Definition 2.1. *Let $C \subset \mathbb{R}^n$ be a non-empty, closed and convex set and $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^n$.*

(i) The mapping $\mathcal{F}$ is called `pseudo-monotone` *if for any $x, y \in \mathbb{R}^n$ it holds*

$$\langle \mathcal{F}(y), x - y \rangle \geq 0 \;\Rightarrow\; \langle \mathcal{F}(x), x - y \rangle \geq 0. \tag{2.1}$$

Observe that by substituting $y = x^*$ in (2.1) we get

$$\langle \mathcal{F}(x), x - x^* \rangle \geq 0 \text{ for all } x \in C \text{ and for all } x^* \in \mathrm{SOL}(C, \mathcal{F}) \tag{2.2}$$

where $\mathrm{SOL}(C, \mathcal{F})$ is the solution set of (1.1).

(ii) The mapping $\mathcal{F}$ is called `Lipschitz continuous on` $\mathbb{R}^n$ *if for any $x, y \in \mathbb{R}^n$ there exists an $L \geq 0$ such that*

$$\|\mathcal{F}(x) - \mathcal{F}(y)\| \leq L\|x - y\|. \tag{2.3}$$

(iii) A sequence $\left\{x^k\right\}_{k=0}^{\infty} \subset \mathbb{R}^n$ is called `Fejer-monotone` *with respect to C if for every $u \in C$*

$$\|x^{k+1} - u\| \leq \|x^k - u\| \text{ for all } k \geq 0. \tag{2.4}$$

The following lemma is due to Gafni and Bertsekas [12] and it is central in our convergence theorem. This can also be found in Toint [24] or more recently, for example in [14] and [8].

Lemma 2.2. *Let $C \subset \mathbb{R}^n$ be a non-empty, closed and convex set. For every $x \in C$, $z \in \mathbb{R}^n$ and $\alpha > 0$, the function*

$$h(\alpha) = \frac{\|P_C\left(x + \alpha z\right) - x\|}{\alpha} \tag{2.5}$$

is monotonically non-increasing.

For each point $x \in \mathbb{R}^n$, there exists a unique nearest point in C, denoted by $P_C(x)$; that is,

$$\|x - P_C(x)\| \leq \|x - y\| \text{ for all } y \in C. \tag{2.6}$$

The mapping $P_C : \mathbb{R}^n \to C$ is called the *metric projection* of $\mathbb{R}^n$ onto C. It is well known that P_C is a *non-expansive* mapping of $\mathbb{R}^n$ onto C, i.e.,

$$\|P_C(x) - P_C(y)\| \leq \|x - y\| \text{ for all } x, y \in \mathbb{R}^n. \tag{2.7}$$

The metric projection P_C is characterized [13, Section 3] by the following two properties:

$$P_C(x) \in C \tag{2.8}$$

and

$$\langle x - P_C(x), P_C(x) - y\rangle \geq 0 \text{ for all } x \in \mathbb{R}^n, \ y \in C, \tag{2.9}$$

and if C is a hyperplane, then (2.9) becomes an equality. It follows that

$$\|x - y\|^2 \geq \|x - P_C(x)\|^2 + \|y - P_C(x)\|^2 \text{ for all } x \in \mathbb{R}^n, \ y \in C. \tag{2.10}$$

The next definition of a fixed point set of a mapping T and (2.9) give an equivalent formulation for the VIP.

Definition 2.3. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a given mapping. The* `fixed point set` *of T is defined as*

$$\mathrm{Fix}(T) := \{x \in \mathbb{R}^n \mid T(x) = x\}. \tag{2.11}$$

A well-known relation between the solution set of the VIP (1.1), $\mathrm{SOL}(C, \mathcal{F})$, and the fixed point set of the operator $P_C(I - \lambda\mathcal{F})$ is: for any $\lambda \geq 0$;

$$\mathrm{SOL}(C, \mathcal{F}) = \mathrm{Fix}(P_C(I - \lambda\mathcal{F})), \tag{2.12}$$

see e.g., Eaves [9]. By converting this relation into an iterative method for solving the VIP (1.1) we can get the well-known projected gradient method. Next we show how by using similar techniques we can recover the extragradient method ((1.2)–(1.3)).

Lemma 2.4. *Let $C \subset \mathbb{R}^n$ be non-empty, closed and convex. Let $\mathcal{F} : C \to \mathbb{R}^n$ be Lipschitz continuous with constant $L > 0$. For any $\lambda \in (0, 1/L)$, we get*

$$\mathrm{SOL}(C, \mathcal{F}) = \mathrm{Fix}(P_C(I - \lambda\mathcal{F}(P_C(I - \lambda\mathcal{F})))). \tag{2.13}$$

Proof. (i) Let $x \in \mathrm{SOL}(C, \mathcal{F})$. Applying (2.12) twice, we get

$$P_C(x - \lambda\mathcal{F}(P_C(x - \lambda\mathcal{F}(x)))) = P_C(x - \lambda\mathcal{F}(x)) = x \tag{2.14}$$

which implies that $x \in \mathrm{Fix}(P_C(I - \lambda\mathcal{F}(P_C(I - \lambda\mathcal{F}))))$.

(ii) On the other hand, let $x \in \mathrm{Fix}(P_C(I - \lambda\mathcal{F}(P_C(I - \lambda\mathcal{F}))))$. Denote by $y := P_C(x - \lambda\mathcal{F}(x))$, we get $x = P_C(x - \lambda\mathcal{F}(y))$. We now show that $x = y$. Indeed, following the non-expansiveness of the metric projection P_C and the Lipschitz continuity of $\mathcal{F}$

$$\begin{aligned}\|x - y\| &= \|P_C(x - \lambda\mathcal{F}(y)) - P_C(x - \lambda\mathcal{F}(x))\| \\ &\leq \|(x - \lambda\mathcal{F}(y)) - (x - \lambda\mathcal{F}(x))\| = \lambda \|\mathcal{F}(x) - \mathcal{F}(y)\| \\ &\leq \frac{\lambda}{L} \|x - y\|.\end{aligned} \tag{2.15}$$

following the assumption on λ we get that $x = y$, meaning that $x = y = P_C(x - \lambda\mathcal{F}(x))$, i.e., $x \in \mathrm{Sol}(\mathcal{F}, C)$. □

Notation 2.5. *Any closed and convex set $C \subset \mathbb{R}^n$ can be represented as*

$$C = \{x \in \mathbb{R}^n \mid c(x) \leq 0\}, \tag{2.16}$$

where $c : \mathbb{R}^n \to \mathbb{R}$ is an appropriate convex function. Take, for example, $c(x) = \operatorname{dist}(x, C)$, where dist is the distance function; see, e.g., [15, Chapter B, Subsection 1.3(c)].

We denote the *subdifferential set* of c at a point x by

$$\partial c(x) := \{\xi \in \mathbb{R}^n \mid c(y) \geq c(x) + \langle \xi, y - x\rangle \text{ for all } y \in \mathbb{R}^n\}. \tag{2.17}$$

For $z \in \mathbb{R}^n$, take any $\xi \in \partial c(z)$ and define

$$T(z) := \{w \in \mathbb{R}^n \mid c(z) + \langle \xi, w - z\rangle \leq 0\}. \tag{2.18}$$

This is a half-space the bounding hyperplane of which separates the set C from the point z if $\xi \neq 0$; otherwise $T(z) = \mathbb{R}^n$; see, e.g., [3, Lemma 7.3].

3. THE ALGORITHM

Our new modification of the subgradient extragradient algorithm without the Lipschitz assumption is given next.

Algorithm 3.1. ***The self-adaptive subgradient extragradient algorithm***

Step 0: *Select a starting point $x^0 \in \mathbb{R}^n$. Choose $\alpha_{-1} \in (0, \infty)$, $\varepsilon \in (0, 1)$, and $\beta \in (0, 1)$.*

Step 1: *Given the current iterate x^k, choose α_k to be the largest*

$$\alpha \in \{\alpha_{k-1}, \alpha_{k-1}\beta, \alpha_{k-1}\beta^2, \ldots\} \tag{3.1}$$

satisfying

$$\alpha \left\langle x^k - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle \leq (1 - \varepsilon) \left\|x^k - y^k\right\|^2 \tag{3.2}$$

where

$$y^k = P_C(x^k - \alpha\mathcal{F}(x^k)). \tag{3.3}$$

Step 2: *If $x^k = y^k$ then stop. Otherwise, denote $a^k := \left(x^k - \alpha_k\mathcal{F}(x^k)\right) - y^k$ and construct the set T_k as follows*

$$T_k := \begin{cases} \{w \in \mathbb{R}^n \mid \left\langle a^k, w - y^k\right\rangle \leq 0\}, & \text{if } a^k \neq 0, \\ \mathbb{R}^n & \text{if } a^k = 0. \end{cases} \tag{3.4}$$

Calculate the next iterate

$$x^{k+1} = P_{T_k}(x^k - \alpha_k\mathcal{F}(y^k)), \tag{3.5}$$

set $k \leftarrow (k + 1)$ and return to ***Step 1***.

Remark 3.2. *1. Observe that (3.2) can be viewed as a local approximation of the Lipschitz constant L, and then we get $\alpha < 1/L$. Indeed, if*

$$L_k = \frac{\left\langle x^k - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle}{\left\|x^k - y^k\right\|^2} \tag{3.6}$$

then $\alpha \leq (1 - \varepsilon)/L_k$.

2. There exists many other techniques for the choice of α_k in Step 1, for example Khobotov [17]

$$\alpha_k = \min\left\{\alpha_{k-1}, \beta\frac{\left\|x^k - y^k\right\|}{\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|}\right\}, \tag{3.7}$$

and Marcotte [20]

$$\alpha_k = \min\left\{\frac{\alpha_{k-1}}{2}, \frac{\|x^k - y^k\|}{\sqrt{2}\,\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\|}\right\}. \tag{3.8}$$

It looks like any of the above choices (and other e.g., [25]*) might work as well.*

Remark 3.3. *Observe that if c is lower semi-continuous and Gâteaux differentiable at y^k, then $\{(x^k - \tau F(x^k)) - y^k\} = \partial c(y^k) = \{\nabla c(y^k)\}$; otherwise $(x^k - \tau F(x^k)) - y^k \in \partial c(y^k)$. See* [3, Facts 7.2] *and* [10] *for more details.*

For the convergence of the algorithm, the following assumptions are needed.

Condition 3.4. *The set $\mathrm{SOL}(C, \mathcal{F})$ is non-empty.*

Condition 3.5. *The mapping $\mathcal{F}$ is pseudo-monotone on C, that is (2.2).*

Condition 3.6. *The mapping $\mathcal{F}$ is continuous on $\mathbb{R}^n$.*

Remark 3.7. *Censor et al.* [5] *(see also* [6, 7]*) introduced an extension of Korpelevich's extragradient method which is the Subgradient Extragradient Method (SEM). The general idea of the SEM is close to that in Algorithm 3.1 in which, given the current iterate x^k, the next iterate x^{k+1} is calculated as the projection onto the constructible set T_k (3.4). But while the convergence of the SEM is guaranteed under strong assumptions as monotonicity, Lipschitz continuity on $\mathbb{R}^n$ and the knowing the Lipschitz constant, Algorithm 3.1 requires only pseudo-monotonicity and continuity. This advantage is not only theoretical but also plays a central role in practice when the information regarding the Lipschitz constant in missing or when the mapping is only continues mappings; see Section 4 for numerical experiments. In addition, other step size adaptations ((3.2) in **Step 1**) can be chosen, for example Khobotov's* [17] *or Marcotte's* [20].

3.1. **Convergence of the self-adaptive subgradient extragradient algorithm.**

For the convergence we first show that **Step 2** is valid.

Lemma 3.8. *If for some $k \geq 0$, $x^k = y^k$ in Algorithm 3.1, then $x^k, y^k \in \mathrm{SOL}(C, \mathcal{F})$.*

Proof. Assume that $x^k = y^k$; then $x^k = P_C(x^k - \alpha_k \mathcal{F}(x^k))$, so $x^k \in C$. By the variational characterization of the projection with respect to C (2.9), we have

$$\left\langle w - x^k, (x^k - \alpha_k \mathcal{F}(x^k)) - x^k \right\rangle \leq 0, \text{ for all } w \in C, \tag{3.9}$$

which implies that

$$\alpha_k \left\langle w - x^k, \mathcal{F}(x^k) \right\rangle \geq 0, \text{ for all } w \in C. \tag{3.10}$$

Since $\alpha_k > 0$, we have that $x^k \in \mathrm{SOL}(C, \mathcal{F})$. □

From now on we assume that the algorithm generates infinite sequences $\{x^k\}_{k=0}^\infty$ and $\{y^k\}_{k=0}^\infty$. Next we prove that α_k is well defined.

Lemma 3.9. *For all $k \geq 0$, there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$ (3.2) holds. Hence α_k is well defined.*

Proof. By Condition 3.6, $\mathcal{F}$ is continuous on $\mathbb{R}^n$. Since the metric projection is also continuous on $\mathbb{R}^n$, we obtain

$$\lim_{\alpha \to 0} P_C(x^k - \alpha \mathcal{F}(x^k)) = P_C(x^k). \tag{3.11}$$

We now examine the two cases, $x^k \in C$ and $x^k \notin C$.

(i) If $x^k \in C$, then $x^k = P_C(x^k)$. By the continuity of $\mathcal{F}$ and (3.11), we get that for sufficiently small $\alpha \in (0,1]$,

$$\left\|\mathcal{F}(x^k)\right\| \left\|\mathcal{F}(x^k) - \mathcal{F}\left(P_C(x^k - \alpha\mathcal{F}(x^k))\right)\right\| \leq (1-\varepsilon)\left\|x^k - P_C(x^k - \mathcal{F}(x^k))\right\|^2. \tag{3.12}$$

Now, let $\alpha \in (0,1]$ be sufficiently small. By the Cauchy-Schwarz inequality, the non-expansivness of the metric projection and Lemma 2.2 we get

$$\begin{aligned}
&\alpha\left\langle x^k - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle \\
&= \alpha\left\langle P_C(x^k) - P_C(x^k - \alpha\mathcal{F}(x^k)), \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle \\
&\leq \alpha\left\|P_C(x^k) - P_C(x^k - \alpha\mathcal{F}(x^k))\right\| \left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\| \\
&\leq \alpha^2\left\|\mathcal{F}(x^k)\right\| \left\|\mathcal{F}(x^k) - \mathcal{F}\left(y^k\right)\right\| \\
&\leq \alpha^2(1-\varepsilon)\left\|x^k - P_C(x^k - \mathcal{F}(x^k))\right\|^2 \\
&\leq (1-\varepsilon)\left\|x^k - y^k\right\|^2,
\end{aligned} \tag{3.13}$$

so (3.2) is valid.

(ii) If $x^k \notin C$, then

$$\lim_{\alpha\to 0} \alpha\left\langle x^k - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle = 0 \tag{3.14}$$

while

$$\lim_{\alpha\to 0}(1-\varepsilon)\left\|x^k - y^k\right\|^2 = (1-\varepsilon)\left\|x^k - P_C(x^k)\right\|^2 > 0, \tag{3.15}$$

implying the claim. □

The next Lemma is central for the convergence theorem.

Lemma 3.10. *Let $\{x^k\}_{k=0}^\infty$, $\{y^k\}_{k=0}^\infty$ be any two sequences generated by Algorithm 3.1. Assume that Conditions 3.4--3.6* hold, and let $x^* \in \mathrm{SOL}(C,\mathcal{F})$. Then for every $k \geq 0$

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \left\|x^k - y^k\right\|^2\left(1 - \alpha_k^2\frac{\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|^2}{\left\|x^k - y^k\right\|^2}\right). \tag{3.16}$$

Proof. Let $x^* \in \mathrm{SOL}(C,\mathcal{F})$. Since $y^k \in C$, we have by Condition 3.5

$$\left\langle \mathcal{F}(y^k), y^k - x^*\right\rangle \geq 0 \text{ for all } k \geq 0, \tag{3.17}$$

which implies that

$$\left\langle \mathcal{F}(y^k), x^{k+1} - x^*\right\rangle \geq \left\langle \mathcal{F}(y^k), x^{k+1} - y^k\right\rangle. \tag{3.18}$$

By the definition of T_k, we have

$$\left\langle x^{k+1} - y^k, \left(x^k - \alpha_k\mathcal{F}(x^k)\right) - y^k\right\rangle \leq 0 \text{ for all } k \geq 0, \tag{3.19}$$

then

$$\begin{aligned}
\left\langle x^{k+1} - y^k, (x^k - \alpha_k\mathcal{F}(y^k)) - y^k\right\rangle &= \left\langle x^{k+1} - y^k, x^k - \alpha_k\mathcal{F}(x^k) - y^k\right\rangle \\
&\quad + \alpha_k\left\langle x^{k+1} - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle \\
&\leq \alpha_k\left\langle x^{k+1} - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle.
\end{aligned} \tag{3.20}$$

Now by letting $z^k = x^k - \alpha_k \mathcal{F}(y^k)$ for simplicity, we obtain

$$\begin{aligned}\left\|x^{k+1} - x^*\right\|^2 &= \left\|P_{T_k}(z^k) - x^*\right\|^2 \\ &= \left\|z^k - x^*\right\|^2 + \left\|z^k - P_{T_k}(z^k)\right\|^2 + 2\left\langle P_{T_k}(z^k) - z^k, z^k - x^*\right\rangle.\end{aligned} \tag{3.21}$$

Since

$$\begin{aligned}&2\left\|z^k - P_{T_k}(z^k)\right\|^2 + 2\left\langle P_{T_k}(z^k) - z^k, z^k - x^*\right\rangle \\ &= 2\left\langle z^k - P_{T_k}(z^k), x^* - P_{T_k}(z^k)\right\rangle \le 0 \text{ for all } k \ge 0,\end{aligned} \tag{3.22}$$

we get for all $k \ge 0$

$$\left\|z^k - P_{T_k}(z^k)\right\|^2 + 2\left\langle P_{T_k}(z^k) - z^k, z^k - x^*\right\rangle \le -\left\|z^k - P_{T_k}(z^k)\right\|^2. \tag{3.23}$$

So,

$$\begin{aligned}\left\|x^{k+1} - x^*\right\|^2 &\le \left\|z^k - x^*\right\|^2 - \left\|z^k - P_{T_k}(z^k)\right\|^2 \\ &= \left\|(x^k - \alpha_k \mathcal{F}(y^k)) - x^*\right\|^2 - \left\|(x^k - \alpha_k \mathcal{F}(y^k)) - x^{k+1}\right\|^2 \\ &= \left\|x^k - x^*\right\|^2 - \left\|x^k - x^{k+1}\right\|^2 + 2\alpha_k\left\langle x^* - x^{k+1}, \mathcal{F}(y^k)\right\rangle \\ &\le \left\|x^k - x^*\right\|^2 - \left\|x^k - x^{k+1}\right\|^2 + 2\alpha_k\left\langle y^k - x^{k+1}, \mathcal{F}(y^k)\right\rangle,\end{aligned} \tag{3.24}$$

where the last inequality follows from (3.18).
So

$$\begin{aligned}\left\|x^{k+1} - x^*\right\|^2 &\le \left\|x^k - x^*\right\|^2 - \left\|x^k - x^{k+1}\right\|^2 + 2\alpha_k\left\langle y^k - x^{k+1}, \mathcal{F}(y^k)\right\rangle \\ &= \left\|x^k - x^*\right\|^2 - \left(\left\langle x^k - y^k + y^k - x^{k+1}, x^k - y^k + y^k - x^{k+1}\right\rangle\right) \\ &\quad + 2\alpha_k\left\langle y^k - x^{k+1}, \mathcal{F}(y^k)\right\rangle \\ &= \left\|x^k - x^*\right\|^2 - \left\|x^k - y^k\right\|^2 - \left\|y^k - x^{k+1}\right\|^2 \\ &\quad + 2\left\langle x^{k+1} - y^k, x^k - \alpha_k \mathcal{F}(y^k) - y^k\right\rangle.\end{aligned} \tag{3.25}$$

By (3.20)

$$\begin{aligned}\left\|x^{k+1} - x^*\right\|^2 \le \left\|x^k - x^*\right\|^2 - \left\|x^k - y^k\right\|^2 - \left\|y^k - x^{k+1}\right\|^2 \\ + 2\alpha_k\left\langle x^{k+1} - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle.\end{aligned} \tag{3.26}$$

Using Cauchy–Schwarz inequality, we have

$$2\alpha_k\left\langle x^{k+1} - y^k, \mathcal{F}(x^k) - \mathcal{F}(y^k)\right\rangle \le 2\alpha_k\left\|x^{k+1} - y^k\right\|\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|, \tag{3.27}$$

in addition

$$\begin{aligned}&0 \le \left(\left\|x^{k+1} - y^k\right\| - \alpha_k\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|\right)^2 \\ &= \left\|x^{k+1} - y^k\right\|^2 - 2\alpha_k\left\|x^{k+1} - y^k\right\|\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\| + \alpha_k^2\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|^2,\end{aligned} \tag{3.28}$$

so

$$2\alpha_k\left\|x^{k+1} - y^k\right\|\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\| \le \left\|x^{k+1} - y^k\right\|^2 + \alpha_k^2\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|^2.$$

Combining the above inequalities yields

$$\begin{aligned}\left\|x^{k+1} - x^*\right\|^2 &\le \left\|x^k - x^*\right\|^2 - \left\|x^k - y^k\right\|^2 - \left\|y^k - x^{k+1}\right\|^2 \\ &\quad + \left\|x^{k+1} - y^k\right\|^2 + \alpha_k^2\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|^2 \\ &= \left\|x^k - x^*\right\|^2 - \left\|x^k - y^k\right\|^2 + \alpha_k^2\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|^2.\end{aligned} \tag{3.29}$$

Since $x^k \neq y^k$, we obtain

$$\left\|x^{k+1} - x^*\right\|^2 \leq \left\|x^k - x^*\right\|^2 - \left\|x^k - y^k\right\|^2 \left(1 - \alpha_k^2 \frac{\left\|\mathcal{F}(x^k) - \mathcal{F}(y^k)\right\|^2}{\left\|x^k - y^k\right\|^2}\right) \quad (3.30)$$

and the desired result is obtained. □

We are now ready to present the convergence theorem of Algorithm 3.1. The outline of the proof is similar to [23, theorem 3.2].

Theorem 3.11. *Let $\{x^k\}_{k=0}^{\infty}$ and $\{y^k\}_{k=0}^{\infty}$ be any two sequences generated by Algorithm 3.1. Assume that Conditions 3.4--3.6* hold, then both sequences converge to the same point $\hat{x} \in \mathrm{SOL}(C, \mathcal{F})$.

Proof. Let $\bar{x} \in \mathrm{SOL}(C, \mathcal{F})$. According to Remark 3.2, we get from (3.30)

$$\left\|x^{k+1} - \bar{x}\right\|^2 \leq \left\|x^k - \bar{x}\right\|^2 \text{ for all } k \geq 0 \quad (3.31)$$

and $\left\|x^k - y^k\right\| \to 0$. Observe that (3.31) states that$\{x^k\}_{k=0}^{\infty}$ is Fejér-monotone with respect to $\mathrm{SOL}(C, \mathcal{F})$. So following [4, Theorem 5.11] it is sufficient to find a cluster point of $\{x^k\}_{k=0}^{\infty}$ in $\mathrm{SOL}(C, \mathcal{F})$ and obtain the desired result. According to (3.2) the sequence $\{\alpha_k\}_{k=-1}^{\infty}$ is non-increasing and therefore $\lim_{k\to\infty} \alpha_k = \widehat{\alpha}$. Now, we consider the following two cases: (i) $\widehat{\alpha} > 0$ and (ii) $\widehat{\alpha} = 0$.

(i) If $\widehat{\alpha} > 0$, by (3.31) the sequence $\{x^k\}_{k=0}^{\infty}$ is bounded, therefore there exists a subsequence $\{x^{k_j}\}_{j=0}^{\infty}$ of $\{x^k\}_{k=0}^{\infty}$ such that

$$\lim_{j\to\infty} x^{k_j} = \hat{x}, \quad (3.32)$$

since $\left\|x^k - y^k\right\| \to 0$ we also have

$$\lim_{j\to\infty} y^{k_j} = \hat{x}. \quad (3.33)$$

By the continuity of $\mathcal{F}$ (Condition 3.6) and of the metric projection

$$\hat{x} = \lim_{j\to\infty} y^{k_j} = \lim_{j\to\infty} P_C(x^{k_j} - \alpha_{k_j}\mathcal{F}(x^{k_j})) = P_C(\hat{x} - \widehat{\alpha}\mathcal{F}(\hat{x})). \quad (3.34)$$

Following similar arguments as in the proof of Lemma 3.8 it follows that $\hat{x} \in \mathrm{SOL}(C, \mathcal{F})$ and the result follows from [4, Theorem 5.11].

(ii) If $\widehat{\alpha} = 0$, we argue by contradiction by supposing that every cluster point of the sequence $\{y^k\}_{k=0}^{\infty}$ is not in $\mathrm{SOL}(C, \mathcal{F})$. Since $\widehat{\alpha} = 0$, there exists a subsequence of indices $\{k_l\}_{l=0}^{\infty}$ of $\{k\}_{k=-1}^{\infty}$ such that $\{\alpha_i\}_{i\in\{k_l\}_{l=0}^{\infty}}$ is monotonically decreasing. Taking the limit as $i \to \infty$ (passing to a subsequence if needed), we get that $\lim_{i\to\infty} y^i = \widetilde{y} \notin \mathrm{SOL}(C, \mathcal{F})$.

Since $\widetilde{y} \notin \mathrm{SOL}(C, \mathcal{F})$, then $y^i \notin \mathrm{SOL}(C, \mathcal{F})$ for all sufficiently large $i \in \{k_l\}_{l=0}^{\infty}$. Thus for these i and $\alpha > 0$ we have by (2.12) $y^i \neq P_C(y^i - \alpha\mathcal{F}(y^i))$ and moreover since $\{k\}_{k=-1}^{\infty}$ is infinite, $y^i \neq x^i$ (we did not stop at **Step 2**).

From the continuity of $\mathcal{F}$ and Lemma 3.9, we get for all of these i with $\lim_{i\to\infty} \alpha_i = 0$, that the right hand side of (3.2) goes to a positive limit while the left hand side goes to zero. Therefore inequality (3.2) holds for all sufficiently small $\alpha > 0$, in particular, it holds for $\alpha = \alpha_{i-1}$ for all $i \in \{k_l\}_{l=0}^{\infty}$ sufficiently large. But since α_i is chosen as the largest element in $\{\alpha_{i-1}, \alpha_{i-1}\beta, \alpha_{i-1}\beta^2, \ldots\}$ we get a contradiction to our hypothesis on $\{k_l\}_{l=0}^{\infty}$, that is $\alpha_i < \alpha_{i-1}$ for all $i \in \{k_l\}_{l=0}^{\infty}$.

Thus, there exists at least one cluster point of $\{y^k\}_{k=0}^{\infty}$ and also of $\{x^k\}_{k=0}^{\infty}$, say $\widetilde{x}$, that belongs to $\mathrm{SOL}(C, \mathcal{F})$ and again the desired result follows [4, Theorem 5.11]. □

4. NUMERICAL EXPERIMENTS

In this section we present several numerical examples to illustrate the performance of our algorithm. We choose the test problems from [19] (see also [25]). All computations were performed using MATLAB R2012a on an Intel Core i5-2348M 2.67GHz running 64-bit Windows. The cpu time is measured in seconds using the intrinsic MATLAB function cputime. The projection onto the feasible set C is performed using CVX version 1.22. The numerical results are presented in Table 4, we choose the termination criteria as $\left\|x^k - y^k\right\| \leq \delta$ for small $\delta > 0$.

Example 4.1. *We take $\mathcal{F}(x) = Mx + q$ with the matrix M randomly generated as suggested in [16], $M = AA^T + B + D$; where every entry of the n-square matrix A and of the n-skew-symmetric matrix B is uniformly generated from $(-5, 5)$, and every diagonal entry of the n diagonal matrix D is uniformly generated from $(0, 0.3)$, with every entry of q uniformly generated from $(500, 0)$. The feasible set is*

$$C := \{x \in \mathbb{R}^n_+ \mid \sum_{i=1}^{n} x_i = n\}.$$

Example 4.2. *Kojima-Shindo Nonlinear Complementarity Problem (NCP), see e.g., [22]. With $n = 4$, the feasible set is*

$$C := \{x \in \mathbb{R}^4_+ \mid x_1 + x_2 + x_3 + x_4 = 4\}$$

and $\mathcal{F}$ is given as follows.

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

Example 4.3. *Here the feasible set $C = \mathbb{R}^5$ and $\mathcal{F}$ is given as $\mathcal{F} = (f_1, f_2, f_3, f_4, f_5)$ where for all $i = 1, \ldots, 5$*

$$f_i = 2\left(x_i - i + 2\right) \exp\left(\sum_{i=1}^{5} \left(x_i - i + 2\right)^2\right). \tag{4.1}$$

				(1.2)–(1.3)		Algorithm 3.1	
Example	x^0	parameters	dim	iter.	time	iter.	time
4.1	$\mathbf{1}_{10}$	$\tau = \frac{0.4}{\|M\|}$, $\alpha_{-1} = 0.9$, $\varepsilon = 0.2$, $\beta = 0.5$	10	70	6.7	77	3.1
	$\mathbf{1}_{20}$		20	80	10.6	76	5.9
	$\mathbf{1}_{40}$		40	161	28	170	16
	$\mathbf{1}_{70}$		70	247	307	266	163
4.2	$\mathbf{1}_4$	$\alpha_{-1} = 0.7$, $\varepsilon = 0.2$, $\beta = 0.5$, $\tau = 0.01$	4	-	-	53	4.5
	$(\frac{1}{2}, \frac{1}{2}, 2, 1)$		4	-	-	62	6
4.3	$\mathbf{1}_5$	$\alpha_{-1} = 0.7$, $\varepsilon = 0.3$, $\beta = 0.5$, $\tau = 0.01$	5	-	-	53	2.5
	$\mathbf{0}_5$		5	-	-	62	2.1

We use the notation $\mathbf{1}_n$ and $\mathbf{0}_n$ for the unit and the zero vectors in $\mathbb{R}^n$. In Example 4.1 we generate a random data which depend on the dimension. As can be seen Korpelevich method ((1.2)–(1.3)) preforms bad compared to Algorithm 3.1

and a reasonable explanation is the fact that two projections onto C are calculated in each iteration while in Algorithm 3.1 the second projection is easily computed.

In Example 4.2 the Lipschitz constant is unknown, so one needs to guess it; but if L is very large then τ is very small and that makes the method very inefficient. So, in our experiments we stop the algorithm as we did not reached the termination criteria in a reasonable time and similarly in Example 4.3.

Acknowledgments. We thank the anonymous referees for their thorough review and highly appreciate the comments and suggestions, which significantly contributed to improving the quality of this paper.

References

[1] A. S. Antipin, On a method for convex programs using a symmetrical modification of the Lagrange function, Ekonomikai Matematicheskie Metody 12(1976), 1164–1173.

[2] L. Armijo, Minimization of functions having Lipschitz continuous first partial derivatives, Pacific Journal of Mathematics 16(1966), 1–3.

[3] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Review 38(1996), 367–426.

[4] H. H. Bauschke and P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Space, Springer, New York, 2011.

[5] Y. Censor, A. Gibali and S. Reich, Extensions of Korpelevich's extragradient method for solving the variational inequality problem in Euclidean space, Optimization 61(2012), 1119–1132.

[6] Y. Censor, A. Gibali and S. Reich, The subgradient extragradient method for solving the variational inequality problem in Hilbert space, Journal of Optimization Theory and Applications 148 (2011), 318–335.

[7] Y. Censor, A. Gibali and S. Reich, Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space, Optimization Method and Software 26(2011), 827–845.

[8] Y. Dong and X. Zhang, New step lengths in projection method for variational inequality problems, Applied Mathematics and Computation 220(2013), 239–245.

[9] B. C. Eaves, On the basic theorem of complementarity, Mathematical Programming 1(1971), 68–75.

[10] I. Ekeland and R. Témam, Convex Analysis and Variational Problems, North-Holland, Amsterdam, 1976.

[11] F. Facchinei and J. S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I and Volume II, Springer-Verlag, New York, NY, USA, 2003.

[12] E. M. Gafni and D. B. Bertsekas, Two-metric projection methods for constrained optimization, SIAM Journal on Control and Optimization, 22(1984), 936–964.

[13] K. Goebel and S. Reich, Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings, Marcel Dekker, New York and Basel, 1984.

[14] D. R. Han and W. Y. Sun, A new modified Goldstein–Levitin–Polyak projection method for variational inequality problems, Computers and Mathematics with Applications 47(2004), 1817–1825.

[15] J.-B. Hiriart-Urruty and C. Lemaréchal, Fundamentals of Convex Analysis, Springer-Verlag, Berlin, Heidelberg, Germany, 2001.

[16] P. T. Harker and J.-S. Pang, A damped-newton method for the linear complementarity problem, Lectures in Applied Mathematics 26(1990), 265–284.

[17] E. N. Khobotov, Modification of the extra-gradient method for solving variational inequalities and certain optimization problems, USSR Computational Mathematics and Mathematical Physics 27(1989), 120–127.

[18] G. M. Korpelevich, The extragradient method for finding saddle points and other problems, Ekonommika i Matematcheskie Metody, 12(1976), 747–756.

[19] Y. Malitsky, Simple projection methods for monotone variational inequality, Technical report.

[20] P. Marcotte, Application of Khobotov's algorithm to variational inequalities and network equilibrium problems, INFORM 29(1991), 258–270.

[21] J. Oleksyn, Extragradient Methods for Elliptic Inverse Problems and Image Denoising, MS.c. Thesis, School of Mathematical Sciences, Rochester Institute of Technology, June 2011.

[22] J.-S. Pang and S. A. Gabriel, NE/SQP: A robust algorithm for the nonlinear complementarity problem, Mathematical Programming 60(1993), 295–337.

[23] M. Solodov and P. Tseng, Modified Projection-Type Methods For Monotone Variational Inequalities, SIAM Journal on Control and Optimization 34(1996), 1814–1830.

[24] PH. L. Toint, Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space, IMA Journal of Numerical Analysis 8(1988), 231–252.

[25] Y. J. Wang, N. H. Xiu, C. Y. Wang, A new version of extragradient method for variational inequality problems, Computers and Mathematics with Applications 42 (2001), 969–979.