Journal of Scientific Computing (2022) 90:71
https://doi.org/10.1007/s10915-021-01751-1

Subgradient Extragradient Method with Double Inertial Steps for Variational Inequalities

Yonghong Yao[1,2] · Olaniyi S. Iyiola[3] · Yekini Shehu[4]

Received: 17 October 2021 / Revised: 4 December 2021 / Accepted: 14 December 2021 /
Published online: 5 January 2022

Abstract

In this paper, we obtain successively weak, strong and linear convergence analysis of the sequence of iterates generated by our proposed subgradient extragradient method with double inertial extrapolation steps and self-adaptive step sizes for solving variational inequalities for which the cost operator is pseudo-monotone and Lipschitz continuous in real Hilbert spaces. Our proposed method is a combination of double inertial extrapolation steps, relaxation step and subgradient extragradient method which is aimed to increase the speed of convergence of many available subgradient extragradient methods with inertia for solving variational inequalities. Several versions of subgradient extragradient methods with inertial extrapolation step serve as special cases of our proposed method and the inertia in our proposed method is more relaxed and chosen in [0, 1]. Numerical implementations of our method show that our method is efficient, implementable and the benefits gained when subgradient extragradient method with double inertial extrapolation steps are considered for variational inequalities instead of subgradient extragradient methods with one inertial extrapolation step available in the literature.

Keywords Hilbert space · Inertial step · Weak and linear convergence · Subgradient extragradient method · Variational inequality

Mathematics Subject Classification 47H05 · 47J20 · 47J25 · 65K15 · 90C25

✉ Yekini Shehu
yekini.shehu@zjnu.edu.cn

Yonghong Yao
yaoyonghong@aliyun.com

Olaniyi S. Iyiola
niyi4oau@gmail.com; oiyiola@clarkson.edu

1 School of Mathematical Sciences, Tiangong University, Tianjin 300387, China

2 The Key Laboratory of Intelligent Information and Big Data Processing of NingXia Province, and Health Big Data Research Institute, North Minzu University, Yinchuan 750021, China

3 Department of Mathematics, Clarkson University, Potsdam, NY, USA

4 College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua 321004, China

1 Introduction

Suppose C is a nonempty, closed and convex subset of a real Hilbert space H and $A : C \to H$ is a continuous operator. Consider the variational inequality problem (VI(A, C), for short): find $x \in C$ such that

$$\langle Ax, y - x\rangle \geq 0 \quad \forall y \in C. \tag{1}$$

Several real-world problems from mechanics, economics, engineering and so on, can be recast into VI(A, C)(1) (see, for example, [1,2,12,14,17–19,23,24]). We assume, throughout this paper, that S denotes the set of solutions of VI(A, C)(1).
One of the projection methods for solving VI(A, C) (1) is the subgradient extragradient method introduced in [6] by Censor et al.: $x_1 \in H$,

$$\begin{cases} y_n = P_C(x_n - \lambda_n Ax_n), \\ T_n := \{w \in H : \langle x_n - \lambda_n Ax_n - y_n, w - y_n\rangle \leq 0\}, \\ x_{n+1} = P_{T_n}(x_n - \lambda_n Ay_n), \quad n \geq 1, \end{cases} \tag{2}$$

where $0 < \inf_{n\geq 1} \lambda_n \leq \sup_{n\geq 1} \lambda_n < \frac{1}{L}$. Censor et al. [6] proved that the sequence $\{x_n\}$ generated by (2) converges weakly to a solution of VI(A, C)(1) (see also [5] for strong convergence results). The subgradient extragradient method (2) has shown numerically to improve the extragradient method of Korpelevich [20] when computing projection onto the feasible set C is computationally expensive (i.e., method (2) minimizes the number of projections onto C per iteration in the extragradient method) and has been studied by several authors when A is either monotone or pseudo-monotone in Hilbert spaces.
In [35], Yang et al. studied a modification of the subgradient extragradient method (2) with the following self-adaptive step-size procedure: Given $\lambda_1 > 0$, $x_1 \in H$ and $\mu \in (0, 1)$,

$$\begin{cases} y_n = P_C(x_n - \lambda_n Ax_n), \\ T_n := \{w \in H : \langle x_n - \lambda_n Ax_n - y_n, w - y_n\rangle \leq 0\}, \\ x_{n+1} = P_{T_n}(x_n - \lambda_n Ay_n), \end{cases} \tag{3}$$

where $\{\lambda_n\}$ is given by

$$\lambda_{n+1} = \begin{cases} \min\left\{\frac{\mu(\|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2)}{2\langle Ax_n - Ay_n, x_{n+1} - y_n\rangle}, \lambda_n\right\}, & \langle Ax_n - Ay_n, x_{n+1} - y_n\rangle > 0 \\ \lambda_n, & \text{otherwise} \end{cases} \tag{4}$$

and showed that the sequence $\{x_n\}$ generated by (3) converges weakly to a solution of VI(A, C) (1).
Furthermore, Fan et al. in [13] studied the subgradient extragradient method (2) with a single inertial extrapolation step:

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(w_n - \lambda_n Aw_n), \\ T_n := \{w \in H : \langle w_n - \lambda_n Aw_n - y_n, w - y_n\rangle \leq 0\}, \\ x_{n+1} = (1 - \alpha_n)w_n + \alpha_n P_{T_n}(w_n - \lambda_n Ay_n), \end{cases} \tag{5}$$

where $0 \leq \theta_n \leq \theta_{n+1} < 1$, $\lambda_n \to 0$ as $n \to \infty$, $\sum_{n=1}^{\infty} \lambda_n = \infty$,

$$\delta > \frac{4\theta[\theta(1 + \theta) + \sigma]}{1 - \theta^2}$$

and

$$0 < \alpha \leq \alpha_n \leq \frac{\delta - 4\theta[\theta(1+\theta) + \frac{1}{4}\theta\delta + \sigma]}{4\delta[\theta(1+\theta) + \frac{1}{4}\theta\delta + \sigma]}.$$

Fan et al. [13] gave weak convergence result of method (5). Similar recent results on solving VI(A, C)(1) using versions of method (5) are found in [31], where $0 \leq \theta_n \leq \theta_{n+1} \leq \frac{1}{10}$, $\lambda_n \to 0$ as $n \to \infty$ and $\sum_{n=1}^{\infty} \lambda_n = \infty$; in [32] where $0 \leq \theta_n \leq \theta_{n+1} \leq \frac{1}{4}$, $0 < \alpha \leq \alpha_n \leq \frac{1}{2}$ and $\{\lambda_n\}$ is generated by a self-adaptive procedure; in [34] where $0 \leq \theta_n \leq \theta \leq \frac{-2\bar{\theta}-1+\sqrt{8\bar{\theta}+1}}{2(1-\bar{\theta})}$, $\bar{\theta} := \frac{1-\mu_0}{2}$, $0 < \mu < \mu_0 < 1$ and in [30] where $\alpha_n = 1$ and θ_n is chosen as

$$\theta_n = \begin{cases} \min\left\{\frac{1}{n^2\|x_n - x_{n-1}\|^2}, \theta\right\}, & x_n \neq x_{n-1}, \\ \theta, & \text{otherwise} \end{cases}$$

and $\theta > 0$. It is observed in all the versions of inertial subgradient extragradient method mentioned above, the inertial term θ_n in the single inertial extrapolation step is chosen such that $0 \leq \theta_n \leq \theta < 1$.

Recently, Shehu et al. [27] proposed a relaxed version of inertial subgradient extragradient method to solve VI(A, C)(1):

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(w_n - \lambda_n A w_n), \\ T_n := \{w \in H : \langle w_n - \lambda_n A w_n - y_n, w - y_n \rangle \leq 0\}, \\ x_{n+1} = (1-\alpha_n)x_n + \alpha_n P_{T_n}(w_n - \lambda_n A y_n), \end{cases} \tag{6}$$

where $0 \leq \theta_n \leq 1$ and $0 < \alpha_n < \frac{1}{2}$. Shehu et al. [27] obtained weak and strong convergence results giving the numerical advantage of the relaxed choice of inertia $\theta_n \in [0, 1]$ in (6) but no linear convergence result was given.

Quite recently, Chang et al. [7] introduced the following inertial subgradient extragradient algorithm with adaptive step sizes to solve VI(A, C)(1): $x_1 = x_0 \in H$, $y_1 = P_C(x_1 - \lambda_1 A x_1)$,

$$\begin{cases} z_{n+1} = x_n + \delta(x_n - x_{n-1}), \\ y_{n+1} = P_C(z_{n+1} - \lambda_n A y_n), \\ T_n := \{w \in H : \langle z_{n+1} - \lambda_n A y_n - y_{n+1}, w - y_{n+1} \rangle \leq 0\}, \\ x_{n+1} = P_{T_n}(z_{n+1} - \lambda_n A y_{n+1}), \end{cases} \tag{7}$$

with the following conditions satisfied:

$$\begin{cases} 0 \leq \delta < \frac{5-\sqrt{13}}{6}, \\ 0 < \alpha < \frac{1}{1+\sqrt{2}+\delta}, \\ \delta(1+\delta)(1+\alpha) - \phi(\alpha)\left(\delta^2 - \frac{\delta}{2}\right) < \frac{(\phi(\alpha)-\delta\alpha)\phi(\alpha)(1-2\delta)}{2\phi(\alpha)-\delta\alpha-2\phi(\alpha)\delta}, \\ \phi(\alpha) := 1 - (1+\sqrt{2})\alpha. \end{cases} \tag{8}$$

Weak convergence of the sequence $\{x_n\}$ generated by (7) was presented alongside some numerical results. However, neither strong nor linear convergence results for (7) was given by Chang et al. [7]. Another drawback of the method in [7] is that the inertia δ is quite restrictive and chosen in $[0, \frac{1}{4})$ and moreover, the conditions imposed in (8) seem stringent on the iterative parameters. Some further recent results on inertial subgradient extragradient method are found in [8–11,16].

Our aim in this paper is to further improve the inertial subgradient extragradient method to solve VI(A, C)(1). Specifically, in the present paper we study the inertial subgradient extragradient method and improve on the results in [7,13,27,30–32,34,35] in the following ways:

- We proposed inertial subgradient extragradient method with double inertial extrapolation steps and relaxation parameter. The method considered in Shehu et al. [27] becomes a special case and so also are the inertial subgradient extragradient methods studied in [6,7,13,30–32,34,35]. One of the inertial factors in our double inertial extrapolation steps is allowed to be equal to 1 and the other inertial factor can be chosen as close as possible to 1 (see our proposed Algorithm 1 below) which is an improvement over the methods in [7,13,31,32,34] where the only single inertial extrapolation step is studied and the inertia is bounded away from 1.
- We obtain weak, strong and linear convergence of our proposed method, respectively. Weak and strong convergence results are given for inertial subgradient extragradient method with double inertial extrapolation steps and relaxation parameter with self-adaptive step size. In the special case when the constant step size is considered alongside the known Lipschitz constant and modulus of strong pseudo-monotonicity of the cost operator, we show that linear rate of convergence is achieved. This is an improvement over the results in [6,7,13,27,31,32,34,35] where no linear rate of convergence is obtained for inertial subgradient extragradient method.
- We give numerical computations of our proposed method and compare it with the methods in [7,13,27,30,34,35]. Our preliminary computational results show that our proposed method is efficient and converges faster (in terms of CPU time and number of iterations) than the inertial subgradient extragradient methods in [7,13,27,30,34,35].

Our paper is organized as follows: We present some definitions and lemmas in Sect. 2. In Sect. 3, we introduce our proposed method. Weak convergence analysis of our proposed method is given in Sect. 4 and strong convergence analysis is given in Sect. 5. Furthermore, linear convergence is obtained in Sect. 6. We present numerical implementations in Sect. 7 and give some concluding remarks in Sect. 8.

2 Preliminaries

This section gives some definitions and lemmas that would be needed in our convergence analysis of this paper.

Definition 2.1 A mapping $A : H \to H$ is called

(a) *η-strongly monotone* on H if there exists a constant $\eta > 0$ such that $\langle Ax - Ay, x - y\rangle \geq \eta\|x - y\|^2$ for all $x, y \in H$;
(b) *monotone* on H if $\langle Ax - Ay, x - y\rangle \geq 0$ for all $x, y \in H$;
(c) *δ-strongly pseudo-monotone* on H if there exists $\delta > 0$ such that $\langle Ay, x - y\rangle \geq 0 \Rightarrow \langle Ax, x - y\rangle \geq \delta\|x - y\|^2, \;\; x, y \in H$;
(d) *pseudo-monotone* on H if, for all $x, y \in H$, $\langle Ax, y - x\rangle \geq 0 \Rightarrow \langle Ay, y - x\rangle \geq 0$;
(e) *L-Lipschitz continuous* on H if there exists a constant $L > 0$ such that $\|Ax - Ay\| \leq L\|x - y\|$ for all $x, y \in H$.
(f) *sequentially weakly continuous* if for each sequence $\{x_n\}$ we have: $\{x_n\}$ converges weakly to x implies $\{Ax_n\}$ converges weakly to Ax.

Remark 2.2 Observe that (a) implies (b), (a) implies (c), (c) implies (d), and (b) implies (d) in the above definitions. If A is η-strongly pseudo-monotone and continuous mapping on finite-dimensional subspaces, it has been shown (see, e.g., [36, Theorem 4.8]) that VI(A, C) (1) has unique solution.

Definition 2.3 The mapping $P_C : H \to C$ which assigns to each point $u \in H$ the unique point $w \in C$ such that

$$\|u - w\| \leq \|u - y\| \;\; \forall y \in C$$

is called the *metric projection* of H onto C.

The metric projection P_C satisfies (see, for example, [3])

$$\langle x - y, P_C x - P_C y\rangle \geq \|P_C x - P_C y\|^2 \;\; \forall x, y \in H. \tag{9}$$

Furthermore, $P_C x$ is characterized by the properties

$$P_C x \in C \quad \text{and} \quad \langle x - P_C x, P_C x - y\rangle \geq 0 \;\; \forall y \in C. \tag{10}$$

This characterization implies that

$$\|x - y\|^2 \geq \|x - P_C x\|^2 + \|y - P_C x\|^2 \;\; \forall x \in H, \;\forall y \in C. \tag{11}$$

Lemma 2.4 *The following statements hold in H:*

(i) $\|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2$ *for all $x, y \in H$;*
(ii) $\|x + y\|^2 \leq \|x\|^2 + 2\langle y, x + y\rangle$ *for all $x, y \in H$;*
(iii) $\|\alpha x + \beta y\|^2 = \alpha(\alpha + \beta)\|x\|^2 + \beta(\alpha + \beta)\|y\|^2 - \alpha\beta\|x - y\|^2 \;\; \forall x, y \in H, \; \alpha, \beta \in \mathbb{R}$.

Lemma 2.5 (Maingé [21]) *Let $\{\varphi_n\}$, $\{\delta_n\}$ and $\{\theta_n\}$ be sequences in $[0, +\infty)$ such that*

$$\varphi_{n+1} \leq \varphi_n + \theta_n(\varphi_n - \varphi_{n-1}) + \delta_n \;\; \forall n \geq 1, \;\; \sum_{n=1}^{+\infty} \delta_n < +\infty,$$

and there exists a real number θ with $0 \leq \theta_n \leq \theta < 1$ for all $n \in \mathbb{N}$. Then the following assertions hold true:

(i) $\sum_{n=1}^{+\infty}[\varphi_n - \varphi_{n-1}]_+ < +\infty$, *where* $[t]_+ := \max\{t, 0\}$;
(ii) *there exists $\varphi^* \in [0, +\infty)$ such that* $\lim_{n\to\infty} \varphi_n = \varphi^*$.

Lemma 2.6 (Opial [25]) *Let C be a nonempty subset of H and let $\{x_n\}$ be a sequence in H such that the following two conditions hold:*

(i) *for any $x \in C$,* $\lim_{n\to\infty} \|x_n - x\|$ *exists;*
(ii) *every sequential weak cluster point of $\{x_n\}$ is in C.*

Then $\{x_n\}$ converges weakly to a point in C.

3 Proposed Method

In this section we introduce and discuss our projection-type method.

Assumption 3.1 Suppose that the following assumptions are satisfied:

(a) The feasible set C is a nonempty, closed, and convex subset of H.

(b) $A : H \to H$ is pseudo-monotone, L-Lipschitz continuous and A satisfies the following condition: whenever $\{x_n\} \subset H$ and $x_n \rightharpoonup v^*$, one has $\|Av^*\| \leq \liminf_{n\to\infty} \|Ax_n\|$.
(c) The solution set S of VI(A, C) (1) is nonempty.
(d) $0 \leq \theta_n \leq \theta_{n+1} \leq 1$.
(e) $0 \leq \delta < \min\left\{\frac{\epsilon-\sqrt{2\epsilon}}{\epsilon}, \theta_1\right\}$, $\epsilon \in (2, \infty)$.
(f) $0 < \alpha \leq \alpha_n \leq \alpha_{n+1} < \frac{1}{1+\epsilon}$, $\epsilon \in (2, \infty)$.

Remark 3.2 The condition " whenever $\{x_n\} \subset H$ and $x_n \rightharpoonup v^*$, one has $\|Av^*\| \leq \liminf_{n\to\infty} \|Ax_n\|$." in Assumption 3.1(b) is strictly weaker than the sequentially weakly continuous assumption in [7,13,30–35]. For example, take $A(v) = v\|v\|\ \forall v \in C$. Then A satisfies our assumption but not sequentially weakly continuous.

We propose our subgradient extragradient method with double inertial extrapolation step and self-adaptive step sizes.

Algorithm 1 Double Inertial Subgradient Extragradient Method with Adaptive Step Size

1: Choose the parameters $\mu \in (0, 1)$ and $\lambda_1 > 0$. Let $x_0, x_1 \in H$ be given starting points. Set $n := 1$.
2: Compute

$$\begin{cases} z_n = x_n + \delta(x_n - x_{n-1}), \\ w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(w_n - \lambda_n Aw_n), \end{cases} \tag{12}$$

where

$$\lambda_{n+1} = \begin{cases} \min\left\{\frac{\mu\|w_n-y_n\|}{\|Aw_n-Ay_n\|}, \lambda_n\right\}, & Aw_n \neq Ay_n \\ \lambda_n, & \text{otherwise.} \end{cases} \tag{13}$$

If $w_n = y_n = x_n$, STOP. Otherwise
3: Compute

$$x_{n+1} = (1-\alpha_n)z_n + \alpha_n P_{T_n}(w_n - \lambda_n Ay_n), \quad n \geq 1, \tag{14}$$

where $\{T_n\}$ is given by

$$T_n := \{w \in H : \langle w_n - \lambda_n Aw_n - y_n, w - y_n\rangle \leq 0\}$$

4: Set $n \leftarrow n+1$, and **go to 2**.

Remark 3.3 Observe that if $x_n = w_n = y_n$, then (12) implies the equality $x_n = P_C(x_n - \lambda_n Ax_n)$ and so $x_n \in S$.

Remark 3.4 Note that by (13), $\lambda_{n+1} \leq \lambda_n\ \forall n \geq 1$ and there exists $\lambda > 0$ such that $\lim_{n\to\infty} \lambda_n = \lambda \geq \min\left\{\frac{\mu}{L}, \lambda_1\right\}$.

Remark 3.5 (a) In our Algorithm 1, the inertia $\theta_n = 1$ is allowed. This is not the case for the methods in [7,13,31,32,34,35] where $\theta_n < 1$. Observe also in our proposed Algorithm 1 that as ϵ increases, so is δ approaching 1 since $\lim_{\epsilon\to\infty} \frac{\epsilon-\sqrt{2\epsilon}}{\epsilon} = 1$. Furthermore, as ϵ increases, α_n reduces. Conversely, as ϵ approaches 2, δ approaches zero and α_n approaches $\frac{1}{3}$.

(b) In the inertial subgradient extragradient methods proposed in [7,32,34], many stringent conditions are imposed on the inertial factor. For example, in [7], δ is assumed to satisfy seemingly strong condition (8) above; in [32] that $0 \leq \theta_n \leq \theta \leq \frac{1}{4}$ and $\{\theta_n\}$ was assumed to satisfy the condition $0 \leq \theta_n \leq \theta \leq \frac{-2\bar{\theta}-1+\sqrt{8\bar{\theta}+1}}{2(1-\bar{\theta})}$, $\bar{\theta} := \frac{1-\mu_0}{2}$, $0 < \mu < \mu_0 < 1$ in [34]. Our Assumption 3.1(d)–(f) are easier to implement than the conditions in [7,32,34].

(c) In the paper [13], it was assumed that $0 \leq \theta_n \leq \theta < 1$ with

$$\delta > \frac{4\theta[\theta(1+\theta)+\sigma]}{1-\theta^2}$$

and

$$0 < \alpha \leq \alpha_n \leq \frac{\delta - 4\theta[\theta(1+\theta)+\frac{1}{4}\theta\delta+\sigma]}{4\delta[\theta(1+\theta)+\frac{1}{4}\theta\delta+\sigma]}.$$

In the result of this paper, these strong assumptions are dispensed with and replaced with a more relaxed and seemingly easier conditions in Assumption 3.1 (d)-(f).

(d) A self-adaptive step size rule is incorporated in Algorithm 1. It is more efficient than the step size rule in the methods of [13,31], where it is assumed that $\lambda_n \to 0$ as $n \to \infty$ and that $\sum_{n=1}^{\infty} \lambda_n = \infty$. Furthermore, our adaptive step size rule does not require the knowledge of the Lipschitz constant of A as an input parameter. Also, Algorithm 1 does not require any line search. $\Diamond$

4 Convergence Analysis

In this section we show that under Assumption 3.1, the sequence $\{x_n\}$ generated by Algorithm 1 converges weakly to a point in S.
For the rest of this paper, we define

$$u_n := P_{T_n}(w_n - \lambda_n A y_n) \;\; \forall n \geq 1.$$

First we prove the following lemma for $\{x_n\}$ generated by Algorithm 1.

Lemma 4.1 *Suppose that $\{x_n\}$ is generated by Algorithm 1 and Assumption 3.1 holds. Then $\{x_n\}$ is bounded.*

Proof Choose a point $x^* \in S$. Then $\langle Ax^*, y_n - x^*\rangle \geq 0$ and since A is pseudo-monotone, we have $\langle Ay_n, y_n - x^*\rangle \geq 0$. Hence $\langle Ay_n, y_n - x^* + u_n - u_n\rangle \geq 0$ and thus

$$\langle Ay_n, x^* - u_n\rangle \leq \langle Ay_n, y_n - u_n\rangle. \tag{15}$$

From the definition of T_n we have $\langle w_n - \lambda_n Aw_n - y_n, u_n - y_n\rangle \leq 0$. Therefore

$$\begin{aligned}
&\langle w_n - \lambda_n Ay_n - y_n, u_n - y_n\rangle \\
&= \langle w_n - \lambda_n Aw_n - y_n, u_n - y_n\rangle + \lambda_n\langle Aw_n - Ay_n, u_n - y_n\rangle \\
&\leq \lambda_n\langle Aw_n - Ay_n, u_n - y_n\rangle.
\end{aligned} \tag{16}$$

Using (11) and (15), we now obtain

$$\begin{aligned}\|u_n - x^*\|^2 &\leq \|w_n - \lambda_n Ay_n - x^*\|^2 - \|w_n - \lambda_n Ay_n - u_n\|^2 \\ &= \|w_n - x^*\|^2 - \|w_n - u_n\|^2 + 2\lambda_n \langle Ay_n, x^* - u_n\rangle \\ &\leq \|w_n - x^*\|^2 - \|w_n - u_n\|^2 + 2\lambda_n \langle Ay_n, y_n - u_n\rangle \\ &= \|w_n - x^*\|^2 - \|w_n - u_n + y_n - y_n\|^2 + 2\lambda_n \langle Ay_n, y_n - u_n\rangle \\ &= \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad + 2\langle w_n - \lambda_n Ay_n - y_n, u_n - y_n\rangle. \end{aligned} \tag{17}$$

Next, using (16) and (17), we see that

$$\begin{aligned}\|u_n - x^*\|^2 &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad + 2\langle w_n - \lambda_n Ay_n - y_n, u_n - y_n\rangle \\ &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad + 2\lambda_n \langle Aw_n - Ay_n, u_n - y_n\rangle \\ &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad + 2\lambda_n \|Aw_n - Ay_n\|\|u_n - y_n\| \\ &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad + \frac{2\lambda_n \mu}{\lambda_{n+1}} \|w_n - y_n\|\|u_n - y_n\| \\ &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad + \frac{\lambda_n \mu}{\lambda_{n+1}} \left(\|w_n - y_n\|^2 + \|u_n - y_n\|^2\right) \\ &= \|w_n - x^*\|^2 - \left(1 - \frac{\lambda_n \mu}{\lambda_{n+1}}\right)\|w_n - y_n\|^2 \\ &\quad - \left(1 - \frac{\lambda_n \mu}{\lambda_{n+1}}\right)\|u_n - y_n\|^2. \end{aligned} \tag{18}$$

Since $\lim_{n\to\infty}\left(1 - \frac{\lambda_n \mu}{\lambda_{n+1}}\right) = 1 - \mu > 0$, there exists a natural number $N \geq 1$ such that

$$\|u_n - x^*\| \leq \|w_n - x^*\| \quad \forall n \geq N. \tag{19}$$

Now, from Algorithm 1, we have

$$\begin{aligned}\|x_{n+1} - x^*\|^2 &= \|(1 - \alpha_n)(z_n - x^*) + \alpha_n(u_n - x^*)\|^2 \\ &= (1 - \alpha_n)\|z_n - x^*\|^2 + \alpha_n\|u_n - x^*\|^2 \\ &\quad - \alpha_n(1 - \alpha_n)\|z_n - u_n\|^2 \\ &\leq (1 - \alpha_n)\|z_n - x^*\|^2 + \alpha_n\|w_n - x^*\|^2 \\ &\quad - \alpha_n(1 - \alpha_n)\|z_n - u_n\|^2, \forall n \geq N. \end{aligned} \tag{20}$$

From $x_{n+1} = (1 - \alpha_n)z_n + \alpha_n u_n$ we have

$$\|u_n - z_n\| = \frac{1}{\alpha_n}\|x_{n+1} - z_n\|, \quad \forall n \geq 1. \tag{21}$$

Combining (21) with (20) gives

$$\begin{aligned}\|x_{n+1}-x^*\|^2 &\le (1-\alpha_n)\|z_n-x^*\|^2+\alpha_n\|w_n-x^*\|^2\\&\quad-\frac{(1-\alpha_n)}{\alpha_n}\|x_{n+1}-z_n\|^2,\ \forall n\ge N.\end{aligned} \tag{22}$$

Also, by Lemma 2.4 (iii), we have

$$\begin{aligned}\|w_n-x^*\|^2 &= \|x_n+\theta_n(x_n-x_{n-1})-x^*\|^2\\&=\|(1+\theta_n)(x_n-x^*)-\theta_n(x_{n-1}-x^*)\|^2\\&=(1+\theta_n)\|x_n-x^*\|^2-\theta_n\|x_{n-1}-x^*\|^2\\&\quad+\theta_n(1+\theta_n)\|x_n-x_{n-1}\|^2,\ \forall n\ge 1.\end{aligned} \tag{23}$$

Applying Lemma 2.4 (iii) again, we have

$$\begin{aligned}\|z_n-x^*\|^2 &= \|x_n+\delta(x_n-x_{n-1})-x^*\|^2\\&=\|(1+\delta)(x_n-x^*)-\delta(x_{n-1}-x^*)\|^2\\&=(1+\delta)\|x_n-x^*\|^2-\delta\|x_{n-1}-x^*\|^2\\&\quad+\delta(1+\delta)\|x_n-x_{n-1}\|^2,\ \forall n\ge 1.\end{aligned} \tag{24}$$

Also,

$$\begin{aligned}\|x_{n+1}-z_n\|^2 &= \|x_{n+1}-(x_n+\delta(x_n-x_{n-1}))\|^2\\&=\|(x_{n+1}-x_n)-\delta(x_n-x_{n-1})\|^2\\&=\|x_{n+1}-x_n\|^2+\delta^2\|x_n-x_{n-1}\|^2-2\delta\langle x_{n+1}-x_n,x_n-x_{n-1}\rangle\\&\ge\|x_{n+1}-x_n\|^2+\delta^2\|x_n-x_{n-1}\|^2-2\delta\|x_{n+1}-x_n\|\|x_n-x_{n-1}\|\\&\ge(1-\delta)\|x_{n+1}-x_n\|^2+(\delta^2-\delta)\|x_n-x_{n-1}\|^2.\end{aligned} \tag{25}$$

Substituting (23), (24) and (25) into (22), we have

$$\begin{aligned}\|x_{n+1}-x^*\|^2 &\le (1-\alpha_n)\Big[(1+\delta)\|x_n-x^*\|^2-\delta\|x_{n-1}-x^*\|^2\\&\quad+\delta(1+\delta)\|x_n-x_{n-1}\|^2\Big]+\alpha_n\Big[(1+\theta_n)\|x_n-x^*\|^2\\&\quad-\theta_n\|x_{n-1}-x^*\|^2+\theta_n(1+\theta_n)\|x_n-x_{n-1}\|^2\Big]\\&\quad-\frac{(1-\alpha_n)}{\alpha_n}\Big[(1-\delta)\|x_{n+1}-x_n\|^2+(\delta^2-\delta)\|x_n-x_{n-1}\|^2\Big]\\&=(1-\alpha_n)(1+\delta)\|x_n-x^*\|^2-\delta(1-\alpha_n)\|x_{n-1}-x^*\|^2\\&\quad+(1-\alpha_n)\delta(1+\delta)\|x_n-x_{n-1}\|^2+\alpha_n(1+\theta_n)\|x_n-x^*\|^2\\&\quad-\alpha_n\theta_n\|x_{n-1}-x^*\|^2+\alpha_n\theta_n(1+\theta_n)\|x_n-x_{n-1}\|^2\\&\quad-\frac{(1-\alpha_n)(1-\delta)}{\alpha_n}\|x_{n+1}-x_n\|^2-\frac{(1-\alpha_n)(\delta^2-\delta)}{\alpha_n}\|x_n-x_{n-1}\|^2\\&=\Big((1-\alpha_n)(1+\delta)+\alpha_n(1+\theta_n)\Big)\|x_n-x^*\|^2\\&\quad-\Big(\delta(1-\alpha_n)+\alpha_n\theta_n\Big)\|x_{n-1}-x^*\|^2\\&\quad+\Big[(1-\alpha_n)\delta(1+\delta)+\alpha_n\theta_n(1+\theta_n)-\frac{(1-\alpha_n)(\delta^2-\delta)}{\alpha_n}\Big]\|x_n-x_{n-1}\|^2\end{aligned}$$

$$
\begin{aligned}
&- \frac{(1-\alpha_n)(1-\delta)}{\alpha_n}\|x_{n+1}-x_n\|^2\\
&= \Big(1+\alpha_n\theta_n+\delta(1-\alpha_n)\Big)\|x_n-x^*\|^2-\Big(\alpha_n\theta_n+\delta(1-\alpha_n)\Big)\|x_{n-1}-x^*\|^2\\
&\quad -\rho_n\|x_{n+1}-x_n\|^2+\sigma_n\|x_n-x_{n-1}\|^2, \qquad (26)
\end{aligned}
$$

where

$$\rho_n := \frac{(1-\alpha_n)(1-\delta)}{\alpha_n}$$

and

$$\sigma_n := (1-\alpha_n)\delta(1+\delta)+\alpha_n\theta_n(1+\theta_n)-\frac{(1-\alpha_n)(\delta^2-\delta)}{\alpha_n}.$$

Define

$$\Gamma_n := \|x_n-x^*\|^2-\Big(\alpha_n\theta_n+\delta(1-\alpha_n)\Big)\|x_{n-1}-x^*\|^2+\sigma_n\|x_n-x_{n-1}\|^2,\quad n\geq 1.$$

Then

$$
\begin{aligned}
\Gamma_{n+1}-\Gamma_n &= \|x_{n+1}-x^*\|^2-\Big(\alpha_{n+1}\theta_{n+1}+\delta(1-\alpha_{n+1})\Big)\|x_n-x^*\|^2\\
&\quad +\sigma_{n+1}\|x_{n+1}-x_n\|^2-\|x_n-x^*\|^2+\Big(\alpha_n\theta_n+\delta(1-\alpha_n)\Big)\|x_{n-1}-x^*\|^2\\
&\quad +\sigma_n\|x_n-x_{n-1}\|^2\\
&\leq \Big(\alpha_n\theta_n+\delta(1-\alpha_n)-\alpha_{n+1}\theta_{n+1}-\delta(1-\alpha_{n+1})\Big)\|x_n-x^*\|^2\\
&\quad -\rho_n\|x_{n+1}-x_n\|^2+\sigma_{n+1}\|x_{n+1}-x_n\|^2\\
&= \Big((\theta_n-\delta)\alpha_n-(\theta_{n+1}-\delta)\alpha_{n+1}\Big)\|x_n-x^*\|^2\\
&\quad -\rho_n\|x_{n+1}-x_n\|^2+\sigma_{n+1}\|x_{n+1}-x_n\|^2. \qquad (27)
\end{aligned}
$$

Since $\theta_n\leq\theta_{n+1}, \alpha_n\leq\alpha_{n+1}$ and $0\leq\delta\leq\theta_1\leq\theta_n, \forall n\geq 1$, we have that $\theta_n-\delta\geq 0$ and $\theta_{n+1}-\delta\geq 0$ so that $\theta_n-\delta\leq\theta_{n+1}-\delta$ and $(\theta_n-\delta)\alpha_n\leq(\theta_{n+1}-\delta)\alpha_{n+1}$ for all $n\geq 1$. Hence,

$$(\theta_n-\delta)\alpha_n-(\theta_{n+1}-\delta)\alpha_{n+1}\leq 0. \qquad (28)$$

Then from (27), we have

$$
\begin{aligned}
\Gamma_{n+1}-\Gamma_n &\leq -\rho_n\|x_{n+1}-x_n\|^2+\sigma_{n+1}\|x_{n+1}-x_n\|^2\\
&= -(\rho_n-\sigma_{n+1})\|x_{n+1}-x_n\|^2. \qquad (29)
\end{aligned}
$$

By Conditions (e) and (f) of Assumption 3.1, we have

$$
\begin{aligned}
\rho_n-\sigma_{n+1} &= \frac{(1-\alpha_n)(1-\delta)}{\alpha_n}-(1-\alpha_{n+1})\delta(1+\delta)-\alpha_{n+1}\theta_{n+1}(1+\theta_{n+1})\\
&\quad +\frac{(1-\alpha_{n+1})(\delta^2-\delta)}{\alpha_{n+1}}\\
&\geq \epsilon(1-\delta)+\epsilon(\delta^2-\delta)-(1-\alpha_{n+1})\delta(1+\delta)-2\alpha_{n+1}\\
&> \epsilon(1-\delta)+\epsilon(\delta^2-\delta)-2(1-\alpha_{n+1})-2\alpha_{n+1}\\
&= \epsilon(1-\delta)+\epsilon(\delta^2-\delta)-2\\
&= \epsilon\delta^2-2\epsilon\delta+\epsilon-2. \qquad (30)
\end{aligned}
$$

Since $\delta < \frac{\epsilon-\sqrt{2\epsilon}}{\epsilon}$, we have that $\epsilon\delta^2 - 2\epsilon\delta + \epsilon - 2 > 0$. Therefore, from (29) and (30), we have

$$\Gamma_{n+1} - \Gamma_n \leq -\beta \|x_{n+1} - x_n\|^2, \tag{31}$$

where $\beta := \epsilon\delta^2 - 2\epsilon\delta + \epsilon - 2 > 0$. Hence, $\{\Gamma_n\}$ is non-increasing. Furthermore,

$$\begin{aligned}\Gamma_n &= \|x_n - x^*\|^2 - \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_{n-1} - x^*\|^2 + \sigma_n\|x_n - x_{n-1}\|^2 \\ &\geq \|x_n - x^*\|^2 - (\alpha_n\theta_n + \delta(1-\alpha_n))\|x_{n-1} - x^*\|^2.\end{aligned} \tag{32}$$

So,

$$\begin{aligned}\|x_n - x^*\|^2 &\leq \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_{n-1} - x^*\|^2 + \Gamma_n \\ &\leq \Big(\alpha_n + \delta(1-\alpha_n)\Big)\|x_{n-1} - x^*\|^2 + \Gamma_n \\ &\leq \Big(\frac{1}{1+\epsilon} + \delta(1-\alpha)\Big)\|x_{n-1} - x^*\|^2 + \Gamma_n \\ &= \gamma\|x_{n-1} - x^*\|^2 + \Gamma_n \\ &\leq \gamma\|x_{n-1} - x^*\|^2 + \Gamma_1 \\ &\;\vdots \\ &\leq \gamma^n\|x_0 - x^*\|^2 + (1 + \gamma + \cdots + \gamma^{n-1})\Gamma_1 \\ &\leq \gamma^n\|x_0 - x^*\|^2 + \frac{\Gamma_1}{1-\gamma},\end{aligned} \tag{33}$$

where $\gamma := \frac{1}{1+\epsilon} + \delta(1-\alpha) \in (0,1)$ since $\delta < \frac{\epsilon-\sqrt{2\epsilon}}{\epsilon} < \frac{\epsilon}{(1+\epsilon)(1-\alpha)}$, $\alpha \in (0,1)$ by the choice of δ. Hence, $\{\|x_n - x^*\|\}$ is bounded and so is $\{x_n\}$. Also,

$$\begin{aligned}-\gamma\|x_{n-1} - x^*\|^2 &\leq \|x_n - x^*\|^2 - \gamma\|x_{n-1} - x^*\|^2 \\ &\leq \Gamma_n \leq \Gamma_1.\end{aligned}$$

Now,

$$\begin{aligned}\Gamma_{n+1} &= \|x_{n+1} - x^*\|^2 - \Big(\alpha_{n+1}\theta_{n+1} + \delta(1-\alpha_{n+1})\Big)\|x_n - x^*\|^2 \\ &\quad + \sigma_{n+1}\|x_{n+1} - x_n\|^2 \\ &\geq -\Big(\alpha_{n+1}\theta_{n+1} + \delta(1-\alpha_{n+1})\Big)\|x_n - x^*\|^2.\end{aligned} \tag{34}$$

Using (33) and (34), we have

$$\begin{aligned}-\Gamma_{n+1} &\leq \Big(\alpha_{n+1}\theta_{n+1} + \delta(1-\alpha_{n+1})\Big)\|x_n - x^*\|^2 \\ &\leq \gamma\|x_n - x^*\|^2 \\ &\leq \gamma^{n+1}\|x_0 - x^*\|^2 + \frac{\gamma\Gamma_1}{1-\gamma}.\end{aligned} \tag{35}$$

By (31) and (35), we get

$$\begin{aligned}\beta\sum_{k=1}^{n}\|x_{k+1} - x_k\|^2 &\leq \Gamma_1 - \Gamma_{n+1} \\ &\leq \gamma^{n+1}\|x_0 - x^*\|^2 + \frac{\Gamma_1}{1-\gamma}.\end{aligned} \tag{36}$$

Therefore,

$$\sum_{n=1}^{\infty} \|x_{n+1} - x_n\|^2 \leq \frac{\Gamma_1}{\beta(1-\gamma)} < +\infty. \tag{37}$$

This implies that $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$. From (26) we get

$$\begin{aligned}
\|x_{n+1} - x^*\|^2 &\leq \Big(1 + \alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_n - x^*\|^2 - \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_{n-1} - x^*\|^2 \\
&\quad - \rho_n\|x_{n+1} - x_n\|^2 + \sigma_n\|x_n - x_{n-1}\|^2 \\
&\leq \|x_n - x^*\|^2 + (\alpha_n + \delta(1-\alpha_n))(\|x_n - x^*\|^2 - \|x_{n-1} - x^*\|^2) \\
&\quad + \sigma_n\|x_n - x_{n-1}\|^2 \\
&\leq \|x_n - x^*\|^2 + (\alpha_n + \delta(1-\alpha_n))(\|x_n - x^*\|^2 - \|x_{n-1} - x^*\|^2) \\
&\quad + \Big[(1-\alpha)\delta(1+\delta) + \frac{2}{1+\epsilon} \\
&\quad + \frac{(1-\alpha)}{\alpha}(\delta - \delta^2)\Big]\|x_n - x_{n-1}\|^2,
\end{aligned} \tag{38}$$

where

$$\sigma_n \leq (1-\alpha)\delta(1+\delta) + \frac{2}{1+\epsilon} + \frac{(1-\alpha)}{\alpha}(\delta - \delta^2), \forall n \geq 1.$$

Note also that

$$\alpha_n + \delta(1-\alpha_n) < \frac{1}{1+\epsilon} + \delta(1-\alpha) < 1$$

since $\delta < \frac{\epsilon}{(1+\epsilon)(1-\alpha)}$ because $\delta < \frac{\epsilon - \sqrt{2\epsilon}}{\epsilon} < \frac{\epsilon}{(1+\epsilon)(1-\alpha)}$ for $\alpha \in (0,1)$ and $\epsilon \in (2,\infty)$. Invoking Lemma 2.5 in (38) (and noting (37)), we get

$$\lim_{n\to\infty} \|x_n - x^*\| \text{ exists.}$$

Therefore, $\{x_n\}$ is bounded. □

Theorem 4.2 *Let the sequence $\{x_n\}$ be generated by Algorithm 1. If Assumption 3.1 holds, then it converges weakly to a point in S.*

Proof From $\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0$, we have

$$\begin{aligned}
\|x_{n+1} - z_n\| &\leq \|x_{n+1} - x_n\| + \|x_n - z_n\| \\
&= \|x_{n+1} - x_n\| + \delta\|x_n - x_{n-1}\| \to 0, \ n \to \infty.
\end{aligned}$$

Also,

$$\begin{aligned}
\|x_{n+1} - w_n\| &\leq \|x_{n+1} - x_n\| + \|x_n - w_n\| \\
&\leq \|x_{n+1} - x_n\| + \|x_n - x_{n-1}\| \to 0, n \to \infty.
\end{aligned}$$

Now,

$$\|x_{n+1} - z_n\| = \alpha_n\|z_n - u_n\| \geq \ \alpha\|z_n - u_n\|,$$

which means that

$$\lim_{n\to\infty} \|z_n - u_n\| = 0. \tag{39}$$

Furthermore,

$$\begin{aligned}\|w_n - z_n\| &= \|x_n + \delta(x_n - x_{n-1}) - x_n - \theta_n(x_n - x_{n-1})\| \\ &\le \delta\|x_{n-1} - x_n\| + \theta_n\|x_n - x_{n-1}\| \\ &\le \delta\|x_{n-1} - x_n\| + \|x_n - x_{n-1}\| \to 0, n \to \infty, \end{aligned} \tag{40}$$

Now, from Algorithm 1 we have

$$\begin{aligned}\|u_n - z_n\| &= \frac{1}{\alpha_n}\|x_{n+1} - z_n\| \\ &\le \frac{1}{\alpha}\|x_{n+1} - z_n\| \to 0, n \to \infty, \end{aligned} \tag{41}$$

and similarly,

$$\|w_n - u_n\| \le \|w_n - z_n\| + \|u_n - z_n\| \to 0, n \to \infty.$$

From (18), we have for some $M > 0$,

$$\begin{aligned}\Big(1 - \frac{\lambda_n \mu}{\lambda_{n+1}}\Big)\|w_n - y_n\|^2 &\le \|w_n - x^*\|^2 - \|u_n - x^*\|^2 \\ &= \Big(\|w_n - x^*\| + \|u_n - x^*\|\Big)\Big(\|w_n - x^*\| - \|u_n - x^*\|\Big) \\ &\le M\|w_n - u_n\| \to 0, n \to \infty. \end{aligned} \tag{42}$$

By (42) (and noting that $\lim_{n\to\infty} \lambda_n = \lambda$), we see that $\lim_{n\to\infty} \|w_n - y_n\| = 0$. Also

$$\|x_n - z_n\| \le \|x_n - x_{n+1}\| + \|x_{n+1} - z_n\| \to 0, n \to \infty \tag{43}$$

and

$$\|x_n - w_n\| \le \|x_n - x_{n+1}\| + \|x_{n+1} - w_n\| \to 0, n \to \infty. \tag{44}$$

By Lemma 4.1, $\{x_n\}$ is bounded. Hence, let v^* be a weak cluster point of $\{x_n\}$. Then, we can choose a subsequence of $\{x_n\}$, denoted by $\{x_{n_k}\}$ such that $x_{n_k} \rightharpoonup v^* \in H$. Since $\lim_{n\to\infty} \|x_n - w_n\| = 0$, we have that $w_{n_k} \rightharpoonup v^* \in H$. Therefore, we obtain from (42) that $\lim_{k\to\infty} \|w_{n_k} - y_{n_k}\| = 0$. Now using Lemma [29, Lemma 3.7], we have that $v^* \in S$. Since $\lim_{n\to\infty} \|x_n - x^*\|$ exists for any $x^* \in S$ and every sequential weak cluster point of $\{x_n\}$ is in S, the two assumptions of Lemma 2.6 are verified. Therefore, $\{x_n\}$ converges weakly to a point in S. □

Remark 4.3 All the results in this section can still be obtained for the case where A is a pseudo-monotone operator, L-Lipschitz continuous on bounded subsets of H and the functional $g(x) := \|Ax\|, x \in H$ is weakly lower semi-continuous on H.

5 Strong Convergence

In this section, we give strong convergence result of our proposed Algorithm 1 when the operator A in VI(A, C) (1) is strongly pseudo-monotone and Lipschitz continuous on H. The strong convergence result is obtained without the prior knowledge of both the modulus of strong pseudo-monotonicity and the Lipschitz constant of the cost operator A.

Theorem 5.1 *Suppose Assumption 3.1(a), (e) and (f) hold. Let the sequence $\{x_n\}$ be generated by Algorithm 1. If A is strongly pseudo-monotone and Lipschitz continuous on H, then $\{x_n\}$ converges strongly to the unique solution x^* of VI(A, C) (1).*

Proof Observe that the strong pseudo-monotonicity of A implies that VI(A, C) (1) has a unique solution. Let us denote this unique solution by x^*. Hence, $\langle Ax^*, y_n - x^*\rangle \geq 0$. By the strong pseudo-monotonicity of A, we have $\langle Ay_n, y_n - x^*\rangle \geq \eta\|y_n - x^*\|^2$, where η is some positive constant. Thus,

$$\langle Ay_n, y_n - u_n + u_n - x^*\rangle \geq \eta\|y_n - x^*\|^2.$$

and

$$\langle Ay_n, x^* - u_n\rangle \leq \langle Ay_n, y_n - u_n\rangle - \eta\|y_n - x^*\|^2. \tag{45}$$

Using (11) and (53), we get

$$\begin{aligned}
\|u_n - x^*\|^2 &\leq \|w_n - \lambda_n Ay_n - x^*\|^2 - \|w_n - \lambda_n Ay_n - u_n\|^2 \\
&= \|w_n - x^*\|^2 - \|w_n - u_n\|^2 + 2\lambda_n\langle Ay_n, x^* - u_n\rangle \\
&\leq \|w_n - x^*\|^2 - \|w_n - u_n\|^2 + 2\lambda_n\langle Ay_n, y_n - u_n\rangle \\
&\quad - 2\lambda_n\eta\|y_n - x^*\|^2 \\
&= \|w_n - x^*\|^2 - \|w_n + y_n - y_n - u_n\|^2 + 2\lambda_n\langle Ay_n, y_n - u_n\rangle \\
&\quad - 2\lambda_n\eta\|y_n - x^*\|^2 \\
&= \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 - 2\lambda_n\eta\|y_n - x^*\|^2 \\
&\quad + 2\langle w_n - \lambda_n Ay_n - y_n, u_n - y_n\rangle.
\end{aligned} \tag{46}$$

Using (16), we also have

$$\langle w_n - \lambda_n Ay_n - y_n, u_n - y_n\rangle \leq \lambda_n\langle Aw_n - Ay_n, u_n - y_n\rangle.$$

By (46), we get

$$\begin{aligned}
\|u_n - x^*\|^2 &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 - 2\lambda_n\eta\|y_n - x^*\|^2 \\
&\quad + 2\langle w_n - \lambda_n Ay_n - y_n, u_n - y_n\rangle \\
&\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 - 2\lambda_n\eta\|y_n - x^*\|^2 \\
&\quad + 2\lambda_n\langle Aw_n - Ay_n, u_n - y_n\rangle \\
&\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 - 2\lambda_n\eta\|y_n - x^*\|^2 \\
&\quad + \frac{\lambda_n\mu}{\lambda_{n+1}}\|w_n - y_n\|^2 + \frac{\lambda_n\mu}{\lambda_{n+1}}\|u_n - y_n\|^2,
\end{aligned} \tag{47}$$

where

$$\begin{aligned}
2\langle Aw_n - Ay_n, u_n - y_n\rangle &\leq 2\|Aw_n - Ay_n\|\|u_n - y_n\| \\
&\leq \frac{2\mu}{\lambda_{n+1}}\|w_n - y_n\|\|u_n - y_n\| \\
&\leq \frac{\mu}{\lambda_{n+1}}\|w_n - y_n\|^2 + \frac{\mu}{\lambda_{n+1}}\|u_n - y_n\|^2.
\end{aligned}$$

Since $\lim_{n\to\infty}\lambda_n = \lambda$ and the sequence $\{\lambda_n\}$ is monotonically decreasing, we have $\lambda_n \geq \lambda \ \ \forall n \geq 1$. Let ρ be a fixed number in the interval $(\mu, 1)$. Since $\lim_{n\to\infty}\frac{\lambda_n\mu}{\lambda_{n+1}} = \mu < \rho$, there exists a natural number N such that $\frac{\lambda_n \mu}{\lambda_{n+1}} < \rho \ \ \forall n \geq N$. So, $\forall n \geq N$, we have

$$\lambda_n \geq \lambda, \quad \frac{\lambda_n \mu}{\lambda_{n+1}} < \rho. \tag{48}$$

Plugging (48) in (47), we have $\forall n \geq N$,

$$\begin{aligned} \|u_n - x^*\|^2 \leq & \|w_n - x^*\|^2 - (1-\rho)\|w_n - y_n\|^2 - (1-\rho)\|y_n - u_n\|^2 - 2\lambda\eta\|y_n - x^*\|^2 \\ = & \|(1+\theta_n)(x_n - x^*) - \theta_n(x_{n-1} - x^*)\|^2 - (1-\rho)\|w_n - y_n\|^2 \\ & - (1-\rho)\|y_n - u_n\|^2 - 2\lambda\eta\|y_n - x^*\|^2 \\ = & (1+\theta_n)\|x_n - x^*\|^2 - \theta_n\|x_{n-1} - x^*\|^2 \\ & + \theta_n(1+\theta_n)\|x_n - x_{n-1}\|^2 - (1-\rho)\|w_n - y_n\|^2 \\ & - (1-\rho)\|y_n - u_n\|^2 - 2\lambda\eta\|y_n - x^*\|^2. \end{aligned} \tag{49}$$

Repeating similar arguments from (20) to (26), we obtain

$$\begin{aligned} \|x_{n+1} - x^*\|^2 \leq & \Big(1 + \alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_n - x^*\|^2 - \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_{n-1} - x^*\|^2 \\ & - \rho_n\|x_{n+1} - x_n\|^2 + \sigma_n\|x_n - x_{n-1}\|^2 + 2\alpha_n\lambda\eta\|y_n - x^*\|^2, \forall n \geq N. \end{aligned} \tag{50}$$

Therefore,

$$\begin{aligned} 2\alpha\lambda\eta\|y_n - x^*\|^2 \leq & \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 \\ & + \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\Big(\|x_n - x^*\|^2 - \|x_{n-1} - x^*\|^2\Big) \\ & + \sigma_n\|x_n - x_{n-1}\|^2 \\ \leq & \|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2 \\ & + \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\Big(\|x_n - x^*\|^2 - \|x_{n-1} - x^*\|^2\Big) \\ & + M^*\|x_n - x_{n-1}\|^2, \end{aligned}$$

where

$$M^* := (1-\alpha)\delta(1+\delta) + \frac{2}{1+\epsilon} - \frac{(1-\alpha)(\delta^2 - \delta)}{\alpha}.$$

Hence

$$\begin{aligned} 2\alpha\lambda\eta\sum_{k=N}^{n}\|y_k - x^*\|^2 \leq & \|x_N - x^*\|^2 - \|x_{n+1} - x^*\|^2 \\ & + \Big(\alpha_n\theta_n + \delta(1-\alpha_n)\Big)\|x_n - x^*\|^2 \\ & - \Big(\alpha_{N-1}\theta_{N-1} + \delta(1-\alpha_{N-1})\Big)\|x_{N-1} - x^*\|^2 \\ & + M^*\sum_{k=N}^{n}\|x_k - x_{k-1}\|^2. \end{aligned}$$

Since the sequence $\{x_n\}$ is bounded and $\sum_{k=N}^{\infty}\|x_k - x_{k-1}\|^2 < \infty$ by (37), we have that $\sum_{k=N}^{\infty}\|y_k - x^*\|^2 < \infty$. Hence $\lim_{n\to\infty}\|y_n - x^*\| = 0$. Consequently, we get

$$\|x_n - x^*\| \leq \|x_n - w_n\| + \|w_n - y_n\| + \|y_n - x^*\| \to 0, n \to \infty.$$

This concludes the proof. □

Remark 5.2 Theorem 5.1 is obtained for the sequence $\{x_n\}$ generated by Algorithm 1, which is a subgradient extragradient method with double inertial extrapolation steps without a priori knowledge of the modulus of strong pseudo-monotonicity and the Lipschitz constant of A. As far as we know, our result is one of the few available strong convergence results for solving VI(A, C) (1) with a combination of subgradient extragradient method and double inertial extrapolation steps. The benefits of adding double inertial extrapolation step as against single inertial extrapolation step are discussed in the numerical illustrations given in Sect. 7. ♢

6 Linear Convergence

In this section, we give linear rate of convergence of sequence $\{x_n\}$ generated by our proposed Algorithm 1 to the unique solution x^* of VI(A, C)(1) under the following assumptions:

Assumption 6.1 (a) The feasible set C is a nonempty, closed, and convex subset of H.
(b) $A : H \to H$ is η strongly pseudo-monotone and L-Lipschitz continuous.
(c) Choose the constant step size $\lambda \in \left(0, \frac{1}{L}\right)$ such that $\tau := 1-\frac{1}{2}\min\{1-\lambda L, 2\lambda\eta\} \in (\frac{1}{2}, 1)$.
(d) Take $\delta = 0$ and choose $\theta_n = \theta \in [0, 1]$ such that $0 \leq \theta < \frac{1-\tau}{\tau}$.
(e) $\alpha_n = \alpha \in (0, \frac{1}{3})$.

Under Assumption 6.1, our proposed Algorithm 1 reduces to the following Algorithm.

Algorithm 2 Single Inertial Subgradient Extragradient Method

1: Let $x_0, x_1 \in H$ be given starting points. Set $n := 1$.
2: Compute

$$\begin{cases} w_n = x_n + \theta(x_n - x_{n-1}), \\ y_n = P_C(w_n - \lambda A w_n), \end{cases} \tag{51}$$

If $w_n = y_n = x_n$, STOP. Otherwise
3: Compute

$$x_{n+1} = (1-\alpha)x_n + \alpha P_{T_n}(w_n - \lambda A y_n), \quad n \geq 1, \tag{52}$$

where the sequence $\{T_n\}$ is defined by

$$T_n := \{w \in H : \langle w_n - \lambda A w_n - y_n, w - y_n\rangle \leq 0\}.$$

4: Set $n \leftarrow n + 1$, and **go to 2**.

Using Assumption 6.1, we now give the linear convergence of Algorithm 2 below.

Theorem 6.2 *Suppose that Assumption 6.1 is fulfilled. Then $\{x_n\}_{n=1}^{\infty}$ generated by Algorithm 2 converges linearly to the unique point in S.*

Proof Let $x^* \in S$ be the unique solution of VI(A, C) (1). Since A is η-strongly pseudo-monotone, we have that

$$\langle A y_n, y_n - x^*\rangle \geq \eta\|y_n - x^*\|^2.$$

Therefore,

$$\begin{aligned}\langle Ay_n, u_n - x^*\rangle &= \langle Ay_n, u_n - y_n\rangle + \langle Ay_n, y_n - x^*\rangle \\ &\geq \eta\|y_n - x^*\|^2 + \langle Ay_n, y_n - x^*\rangle.\end{aligned}$$

Hence,

$$\begin{aligned}-2\lambda\langle Ay_n, u_n - x^*\rangle \leq &-2\lambda\eta\|y_n - x^*\|^2 \\ &-2\lambda\langle Ay_n, u_n - y_n\rangle.\end{aligned} \tag{53}$$

By (11) and (53), we have

$$\begin{aligned}\|u_n - x^*\|^2 &\leq \|w_n - \lambda Ay_n - x^*\|^2 - \|w_n - \lambda Ay_n - u_n\|^2 \\ &= \|w_n - x^*\|^2 - \|w_n - u_n\|^2 - 2\lambda\langle Ay_n, u_n - x^*\rangle \\ &\leq \|w_n - x^*\|^2 - \|w_n - u_n\|^2 - 2\lambda\eta\|y_n - x^*\|^2 \\ &\quad - 2\lambda\langle Ay_n, u_n - y_n\rangle \\ &= \|w_n - x^*\|^2 - \|w_n - u_n + y_n - y_n\|^2 - 2\lambda\eta\|y_n - x^*\|^2 \\ &\quad - 2\lambda\langle Ay_n, u_n - y_n\rangle \\ &= \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad - 2\lambda\eta\|y_n - x^*\|^2 + 2\langle w_n - \lambda Ay_n - y_n, u_n - y_n\rangle.\end{aligned} \tag{54}$$

Note that from the definition of T_n, we get

$$\langle w_n - \lambda Aw_n - y_n, u_n - y_n\rangle \leq 0$$

and thus

$$\langle w_n - \lambda Ay_n - y_n, u_n - y_n\rangle \leq \lambda\langle Aw_n - Ay_n, u_n - y_n\rangle. \tag{55}$$

Hence,

$$\begin{aligned}\|u_n - x^*\|^2 &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad - 2\lambda\eta\|y_n - x^*\|^2 + 2\langle w_n - \lambda Ay_n - y_n, u_n - y_n\rangle \\ &\leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 - \|y_n - u_n\|^2 \\ &\quad - 2\lambda\eta\|y_n - x^*\|^2 + 2\lambda\langle Aw_n - Ay_n, u_n - y_n\rangle \\ &\leq \|w_n - x^*\|^2 - (1 - \lambda L)\|w_n - y_n\|^2 - (1 - \lambda L)\|y_n - u_n\|^2 \\ &\quad - 2\lambda\eta\|y_n - x^*\|^2 \\ &\leq \|w_n - x^*\|^2 - (1 - \lambda L)\|w_n - y_n\|^2 - 2\lambda\eta\|y_n - x^*\|^2 \\ &\leq \|w_n - x^*\|^2 - \frac{1}{2}\min\{1 - \lambda L, 2\lambda\eta\}\|w_n - x^*\|^2 \\ &= \left(1 - \frac{1}{2}\min\{1 - \lambda L, 2\lambda\eta\}\right)\|w_n - x^*\|^2.\end{aligned} \tag{56}$$

From Algorithm 2, we get

$$\begin{aligned}\|x_{n+1} - x^*\|^2 &= \|(1-\alpha)(x_n - x^*) + \alpha(u_n - x^*)\|^2 \\ &= (1-\alpha)\|x_n - x^*\|^2 + \alpha\|u_n - x^*\|^2 \\ &\quad - \alpha(1-\alpha)\|x_n - u_n\|^2 \\ &\leq (1-\alpha)\|x_n - x^*\|^2 + \alpha\tau\|w_n - x^*\|^2\end{aligned}$$

$$
\begin{aligned}
&\quad - \frac{(1-\alpha)}{\alpha}\|x_{n+1}-x_n\|^2 \\
&\le (1-\alpha)\|x_n - x^*\|^2 + \alpha\tau\Big[(1+\theta)\|x_n - x^*\|^2 \\
&\quad - \theta\|x_{n-1}-x^*\|^2 + \theta(1+\theta)\|x_n - x_{n-1}\|^2\Big] \\
&\quad - \frac{(1-\alpha)}{\alpha}\|x_{n+1}-x_n\|^2 \\
&= (1-\alpha(1-\tau(1+\theta)))\|x_n - x^*\|^2 - \theta\alpha\tau\|x_{n-1}-x^*\|^2 \\
&\quad + \theta(1+\theta)\alpha\tau\|x_n - x_{n-1}\|^2 - \frac{(1-\alpha)}{\alpha}\|x_{n+1}-x_n\|^2.
\end{aligned} \tag{57}
$$

This implies that

$$
\begin{aligned}
\|x_{n+1}-x^*\|^2 + \frac{(1-\alpha)}{\alpha}\|x_{n+1}-x_n\|^2 &\le (1-\alpha(1-\tau(1+\theta)))\|x_n - x^*\|^2 \\
&\quad - \theta\alpha\tau\|x_{n-1}-x^*\|^2 \\
&\quad + \theta(1+\theta)\alpha\tau\|x_n - x_{n-1}\|^2.
\end{aligned} \tag{58}
$$

Since $0 < \alpha < \frac{1}{2}$, we have from (58) that

$$
\begin{aligned}
\|x_{n+1}-x^*\|^2 + \|x_{n+1}-x_n\|^2 &\le (1-\alpha(1-\tau(1+\theta)))\|x_n - x^*\|^2 \\
&\quad - \theta\alpha\tau\|x_{n-1}-x^*\|^2 \\
&\quad + \theta(1+\theta)\alpha\tau\|x_n - x_{n-1}\|^2 \\
&\le (1-\alpha(1-\tau(1+\theta)))\|x^k - x^*\|^2 \\
&\quad + \theta(1+\theta)\alpha\tau\|x^k - x^{k-1}\|^2 \\
&= (1-\alpha(1-\beta^2(1+\theta)))\Big[\|x_n - x^*\|^2 \\
&\quad + \frac{\theta(1+\theta)\alpha\tau}{(1-\alpha(1-\tau(1+\theta)))}\|x_n - x_{n-1}\|^2\Big] \\
&\le (1-\alpha(1-\tau(1+\theta)))(\|x_n - x^*\|^2 \\
&\quad + \|x_n - x_{n-1}\|^2),
\end{aligned} \tag{59}
$$

where the last inequality follows from the fact that $\frac{\theta(1+\theta)\alpha\tau}{(1-\alpha(1-\tau(1+\theta)))} < 1$.
Now, define $a_n := \|x_n - x^*\|^2 + \|x_n - x_{n-1}\|^2$. Then we obtain from (59) that

$$
a_{n+1} \le (1-\alpha(1-\tau(1+\theta)))a_n. \tag{60}
$$

By induction, we obtain

$$
a_{n+1} \le (1-\alpha(1-\tau(1+\theta)))^n a_1. \tag{61}
$$

Therefore, by the definition of a_n, we have

$$
\|x_{n+1}-x^*\|^2 \le (1-\alpha(1-\tau(1+\theta)))^n a_1, n \ge 1.
$$

Hence, we obtain our desired conclusion. □

Remark 6.3 (a) If we choose the step size λ in Algorithm 2 such that $\lambda \in \left(\frac{1}{2\eta+L}, \frac{1}{L}\right)$, then $\tau = 1 - \frac{1}{2}\min\{1-\lambda L, 2\lambda\eta\} = 1 - \frac{1-\lambda L}{2} \in (\frac{1}{2}, 1)$. Thus Assumption 6.1(c) is fulfilled with this choice of λ.

Table 1 Methods parameters for Example 7.1

Alg. 1	$\lambda_1 = 1.1$ $\delta = 2.499 \times 10^{-4}$	$\mu = 0.1$	$\theta_n = 1$	$\alpha_n = 0.3332$
Shehu Alg.	$\lambda_1 = 1.1$	$\mu = 0.1$	$\theta_n = 1$	$\alpha_n = 0.3332$
Fan Alg.	$\lambda_n = \frac{1}{n+1}$	$\alpha_n = 0.1$	$\theta_n = 0.6$	
Yang Alg.	$\lambda_1 = 1.1$	$\mu = 0.1$		
Yang2 Alg.	$\lambda_1 = 1.1$	$\mu = 0.1$	$\alpha_n = \frac{1}{20(n+2)}$	
Chang Alg.	$\lambda_1 = 1.1$	$\mu = 0.1$	$\alpha = 0.343$	$\delta = 0.2$
	$n_0 = 1000$		$\mu_n = 1 + \frac{1}{1 + fix\left(\frac{n}{n_0}\right)}$	

(b) As τ in Assumption 6.1(c) approaches $\frac{1}{2}$, so is θ in Algorithm 2 approaches 1 by Assumption 6.1(d).

7 Numerical Examples

In this section we present many computational experiments and compare the method we proposed in Sect. 3 with some existing methods which are available in the literature. All the codes were written in MATLAB R2019a and ran on a PC Desktop Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz 2.81 GHz, RAM 16.00 GB.

In all these examples, we present numerical comparisons of our proposed Algorithm 1 with the methods in [7,13,27,30,34,35]. In all the numerical implementations, we choose $\delta \in \left[0, \min\left\{\frac{\epsilon - \sqrt{2\epsilon}}{\epsilon}, \theta_1\right\}\right)$, $\epsilon \in (2, \infty)$, $\theta_n \in [0, 1]$ and $\alpha_n \in \left(0, \frac{1}{1+\epsilon}\right)$, $\epsilon \in (2, \infty)$ in our Algorithm 1 for the numerical comparisons with the methods in [7,13,27,30,34,35].
For convenience, [7, Algorithm 3.1] is denoted by "Chang Alg.", [13, Algorithm 3.1] is denoted by "Fan Alg.", [27, Algorithm 1] is denoted by "Shehu Alg.", [30, Algorithm 3.1] is denoted by "Thong Alg.", [34, (2) Theorem 3.1] is denoted by "Yang Alg.", and [35, Algorithm 2] is denoted by "Yang2 Alg." in this section.

Example 7.1 We define $A : \mathbb{R}^m \to \mathbb{R}^m$ by

$$Ax = \left(e^{-x^T Qx} + \beta\right)(Px + q),$$

where Q is a positive definite matrix (i.e, $x^T Qx \geq \theta\|x\|^2 \ \forall x \in \mathbb{R}^m$), P is a positive semi-definite matrix, $q \in \mathbb{R}^m$ and $\beta > 0$. It can be seen that A is differentiable and by the Mean Value Theorem A is Lipschitz continuous. It is also shown in [4, Example 2.1] that A is pseudo-monotone but not monotone.

Take $C := \{x \in \mathbb{R}^m | Bx \leq b\}$, where B is a matrix of size $l^* \times m$ and $b \in \mathbb{R}^{l^*}_+$ with $l^* = 10$. Let us take $x_0 = (1, 1, ..., 1)^T$ and x_1 is generated randomly in $\mathbb{R}^m$. In this example, we use the stopping criterion $\|w_n - y_n\| < 10^{-5}$ with different dimensions m.

Example 7.2 This example is taken from [15] and has been considered by many authors for numerical experiments (see, for example, [22,26,28]). The operator A is defined by $Ax :=$

Table 2 Example 7.1 comparison for different values of m

	m = 100		m = 120	
	No. of Iter.	CPU time	No. of Iter.	CPU time
Alg. 1	16	1.5712	16	1.9640
Shehu Alg.	20	1.7986	17	2.2744
Fan Alg.	64482	1752.389	55756	1555.9497
Yang Alg	185	5.9472	213	8.9348
Yang2 Alg.	190	18.2391	221	11.0381
Chang Alg.	37	2.2902	45	4.0698

Table 3 Methods parameters for Example 7.2

Alg. 1	$\lambda_1 = 0.1$ $\delta = 0.0241$	$\mu = 0.9$	$\theta_n = 1$	$\alpha_n = 0.2903$
Shehu Alg.	$\lambda_1 = 0.1$	$\mu = 0.9$	$\theta_n = 1$	$\alpha_n = 0.2903$
Thong Alg.	$\lambda_1 = 0.1$	$\alpha = 1$	$\mu = 0.9$	
Yang Alg.	$\lambda_1 = 0.1$	$\mu = 0.1$		
Yang2 Alg.	$\lambda_1 = 0.1$	$\mu = 0.1$	$\alpha_n = \frac{1}{20(n+2)}$	
Chang Alg.	$\lambda_1 = 0.1$	$\alpha = 0.3$	$\delta = 0.2$	$n_0 = 1000$
		$\mu_n = 1 + \frac{1}{1 + fix\left(\frac{n}{n_0}\right)}$		

Table 4 Example 7.2 comparison for different k and m

	m = 10		m = 20	
	No. of Iter.	CPU time	No. of Iter.	CPU time
$k = 20$				
Alg. 1	674	12.3199	2115	57.6222
Shehu Alg.	1067	17.6383	3027	59.0911
Thong Alg.	1949	32.4864	5105	91.6340
Yang Alg	1682	28.1467	5428	121.5206
Yang2 Alg.	7104	119.3229	32196	601.5723
Chang Alg.	1019	17.9513	3506	70.1590
$k = 30$				
Alg. 1	547	11.8926	1658	37.9337
Shehu Alg.	757	13.2258	2010	45.9194
Thong Alg.	2029	44.1655	3412	74.1134
Yang Alg	1052	20.2902	4170	89.4936
Yang2 Alg.	4883	108.3688	20520	392.3700
Chang Alg.	958	22.5218	2487	70.7776

Table 5 Methods parameters for Example 7.3

Alg. 1	$\lambda_1 = 1.1$ $\delta = 0.4950$	$\mu = 0.99$	$\theta_n = 1$	$\alpha_n = 0.2250$
Shehu Alg.	$\lambda_1 = 1.1$	$\mu = 0.99$	$\theta_n = 1$	$\alpha_n = 0.2250$
Thong Alg.	$\lambda = 1.1$	$\alpha = 2$	$\mu = 0.99$	
Fan Alg.	$\lambda_n = \frac{1}{n+1}$	$\theta_n = 0.1$	$\alpha_n = 0.05$	
Yang2 Alg.	$\lambda_1 = 1.1$	$\mu = 0.99$	$\alpha_n = \frac{1}{20(n+2)}$	

Table 6 Example 7.3 comparison for different cases

	Case I		Case II	
	No. of Iter.	CPU time	No. of Iter.	CPU time
$k = 20$				
Alg. 1	416	0.058408	416	0.06156
Shehu Alg.	799	0.11686	799	0.12665
Thong Alg.	688	0.092934	10000	1.995
Fan Alg	2836	0.4066	2836	0.38609
Yang2 Alg.	1947	0.2977	1947	0.29401
$k = 30$				
Alg. 1	415	0.062968	419	0.059773
Shehu Alg.	791	0.11601	796	0.11553
Thong Alg.	9997	1.984	567	0.079136
Fan Alg	2811	0.3843	2630	0.36257
Yang2 Alg.	1947	0.30259	1991	0.30477

$Mx+q$, where $M = BB^T + S + D$, with $B, S, D \in \mathbb{R}^{m\times m}$ randomly generated matrices such that S is skew-symmetric (hence the operator does not arise from an optimization problem), D is a positive definite diagonal matrix (hence the variational inequality has a unique solution) and $q = 0$. The feasible set C is described by the linear inequality constraints $Bx \le b$ for some random matrix $B \in \mathbb{R}^{k\times m}$ and a random vector $b \in \mathbb{R}^k$ with nonnegative entries. Hence the zero vector is feasible and therefore the unique solution of the corresponding variational inequality. These projections are computed using the MATLAB solver `fmincon`. Hence, for this class of problems, the evaluation of A is relatively inexpensive, whereas projections are costly. We present the corresponding numerical results (number of iterations and CPU times in seconds) using different dimensions m and different numbers of inequality constraints k. We choose the stopping criterion as $\|e_n\|_2 := \|x_n\| \le \epsilon$, where $\epsilon = 0.001$. The sizes $k = 20, 30$ and $m = 10, 20$. The matrices B, S, D and the vector b are generated randomly.

Next, we give some examples in an infinite dimensional Hilbert space.

Example 7.3 Let $H := L^2([0,1])$ with norm and inner product given by $\|x\| := \left(\int_0^1 x(t)^2 dt\right)^{\frac{1}{2}}$ and $\langle x, y\rangle := \int_0^1 x(t)y(t)dt, \quad x, y \in H$, respectively. We define the feasible

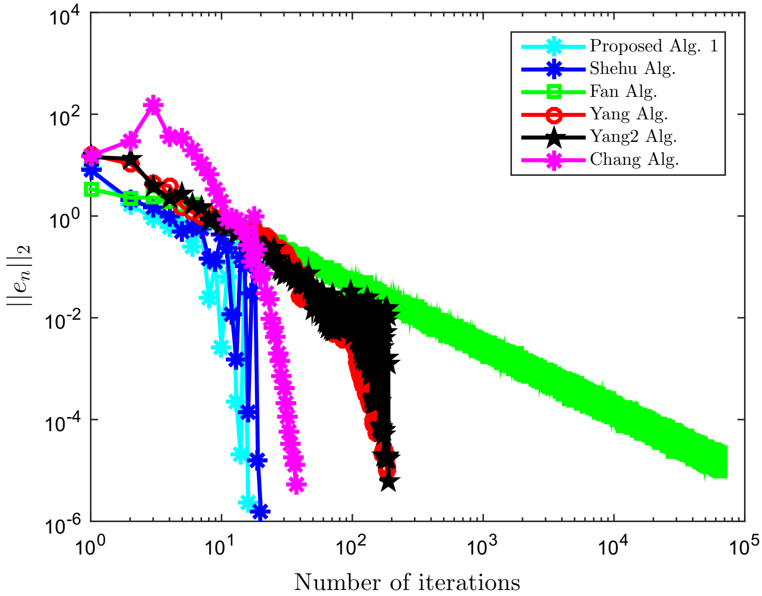

Fig. 1 Example 7.1: $m = 100$

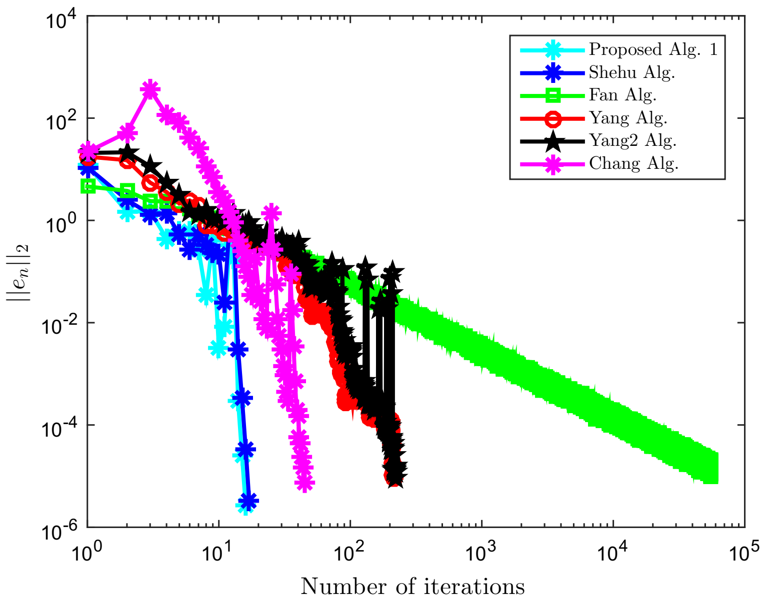

Fig. 2 Example 7.1: $m = 120$

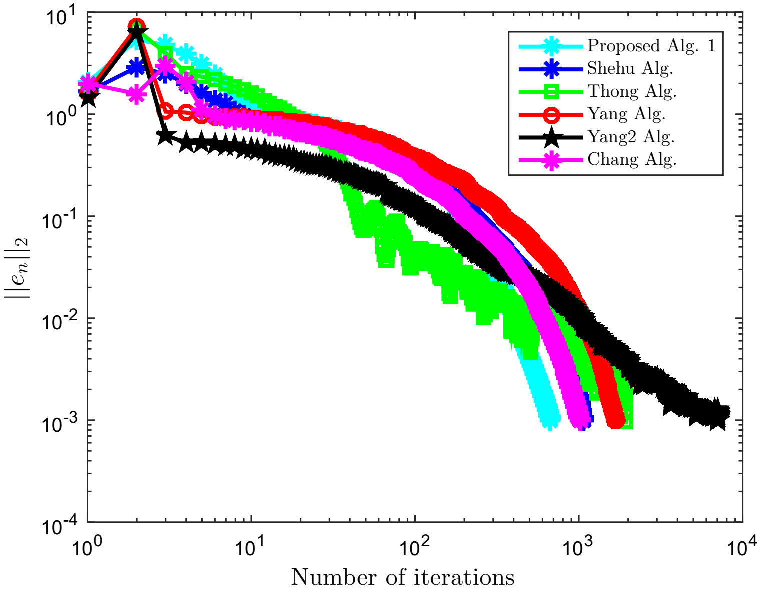

Fig. 3 Example 7.2: $(m, k) = (10, 20)$

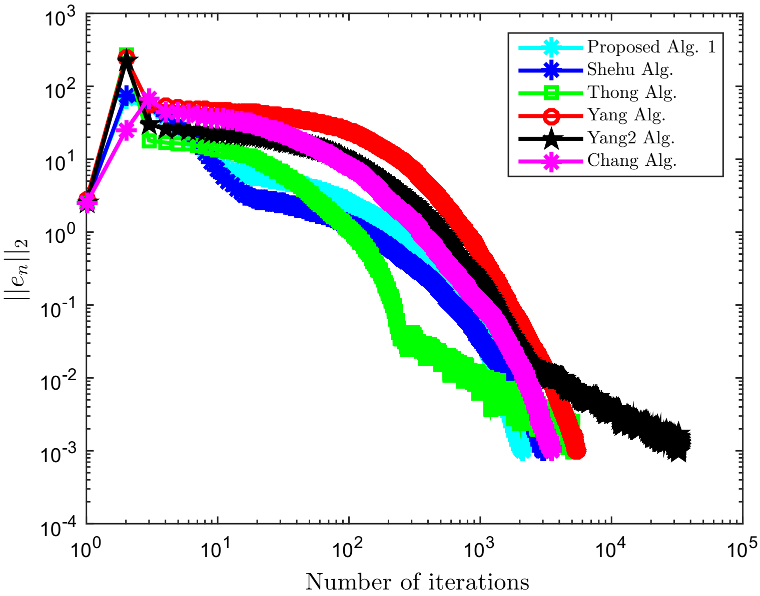

Fig. 4 Example 7.2: $(m, k) = (20, 20)$

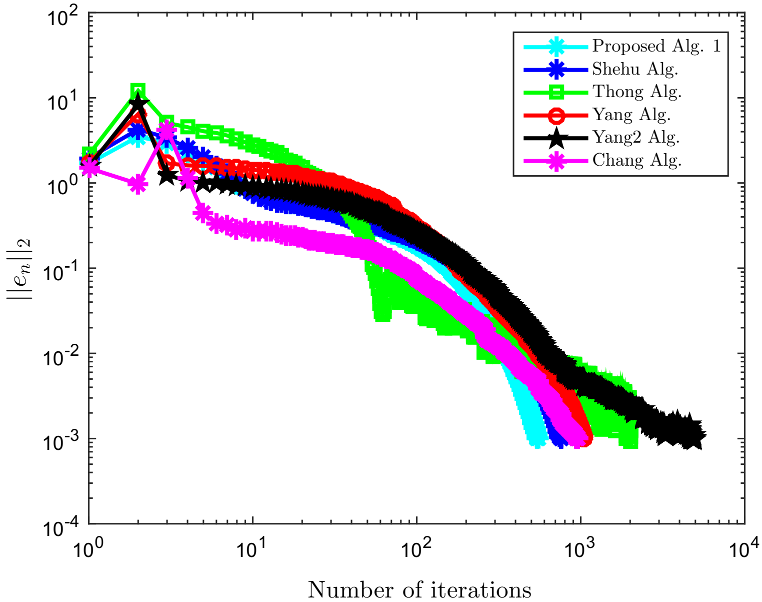

Fig. 5 Example 7.2: $(m, k) = (10, 30)$

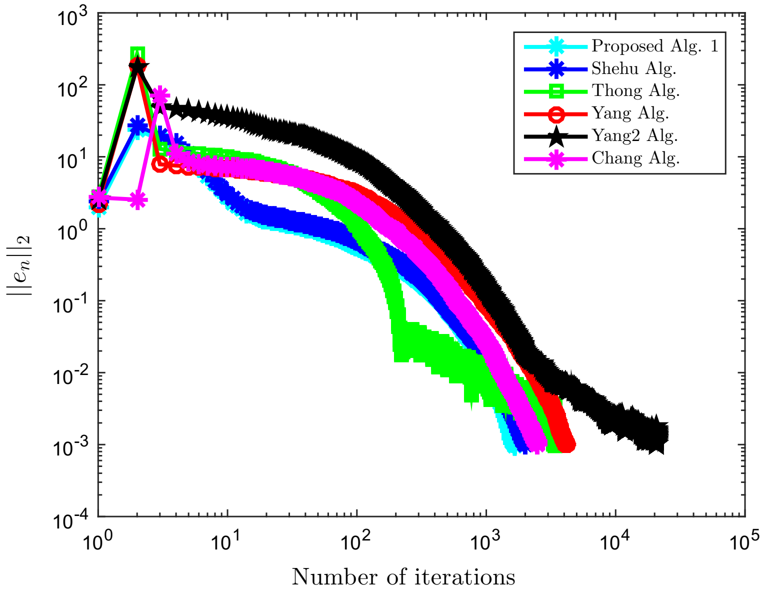

Fig. 6 Example 7.2: $(m, k) = (20, 30)$

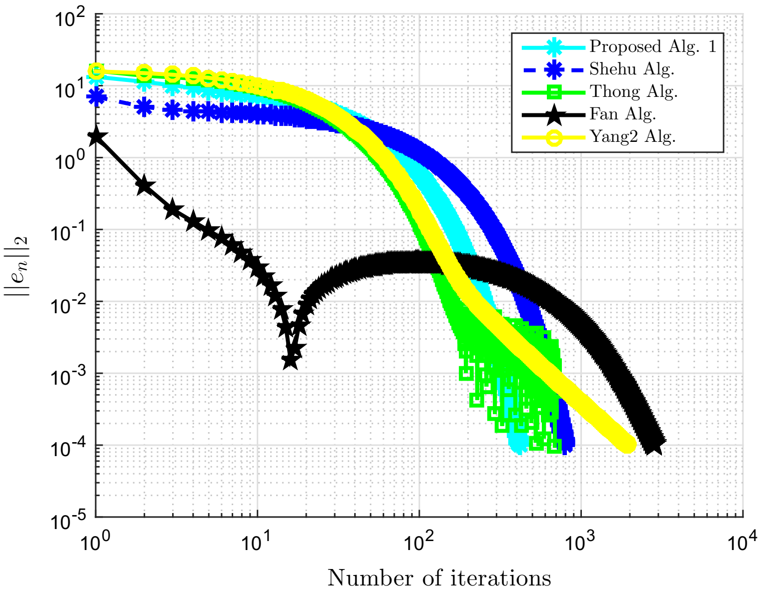

Fig. 7 Example 7.3: Case I

set C by: $C := \{x \in L^2([0,1]) : \int_0^1 tx(t)dt = 2\}$. Define $A : L^2([0,1]) \to L^2([0,1])$ by

$$Ax(t) := \max\{x(t), 0\}, \quad x \in L^2([0,1]), \quad t \in [0,1].$$

Then A is monotone and Lipschitz with $L = 1$. Observe that $S \neq \emptyset$ since $0 \in S$ and that

$$P_C(x)(t) := x(t) - \frac{\int_0^1 tx(t)dt - 2}{\int_0^1 t^2 dt} t, \quad t \in [0,1].$$

We set the stopping criterion to be $e_n := \|x_{n+1} - x_n\| < \epsilon$, where $\epsilon = 10^{-4}$ and consider four different cases as follows:

Case I $x_0 = \frac{1}{13}\left[97t^2 + 4t\right]$ and $x_1 = \frac{1}{250}\left[t^2 - e^{-7t}\right]$

Case II $x_0 = \frac{1}{13}\left[97t^2 + 4t\right]$ and $x_1 = \frac{1}{100}\left[\sin(3t) + \cos(10t)\right]$

Case III $x_0 = \frac{1}{250}\left[t^2 - e^{-7t}\right]$ and $x_1 = \frac{1}{100}\left[\sin(3t) + \cos(10t)\right]$

Case IV $x_0 = \frac{1}{100}\left[\sin(3t) + \cos(10t)\right]$ and $x_1 = \frac{1}{13}\left[97t^2 + 4t\right]$

The following remarks on the numerical implementations given above are in order.

Remark 7.4 (1). The numerical results from the above Examples (both finite and infinite dimension) show that our proposed Algorithm 1 is considerably fast, easy to implement and very efficient (Table 1).

(2). For the given Examples, our proposed Algorithm 1, which is a subgradient extragradient method with double inertial extrapolation steps, outperforms several existing subgradient extragradient methods with single inertial extrapolation step studied in [7,13,27,30,34,35]. This is evident from Tables 2, 3, 4, 5 and 6 and Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 with respect to speed and number of iterations.

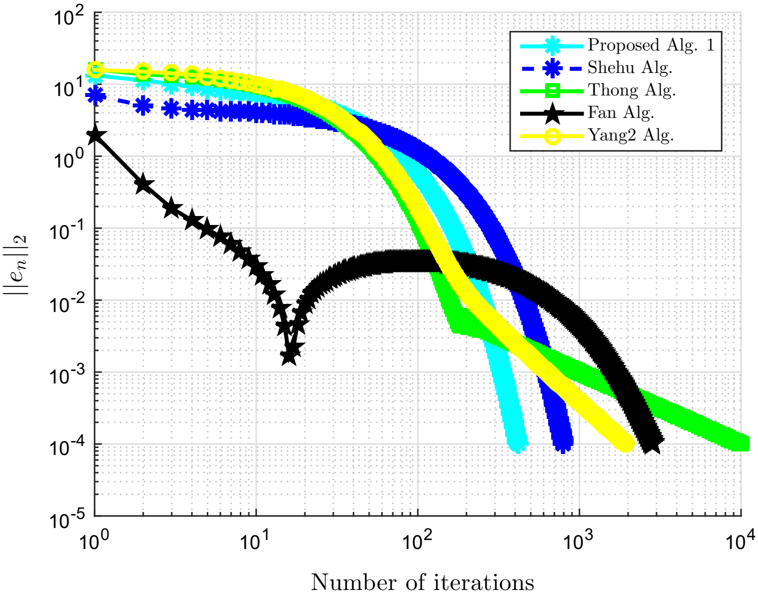

Fig. 8 Example 7.3: Case II

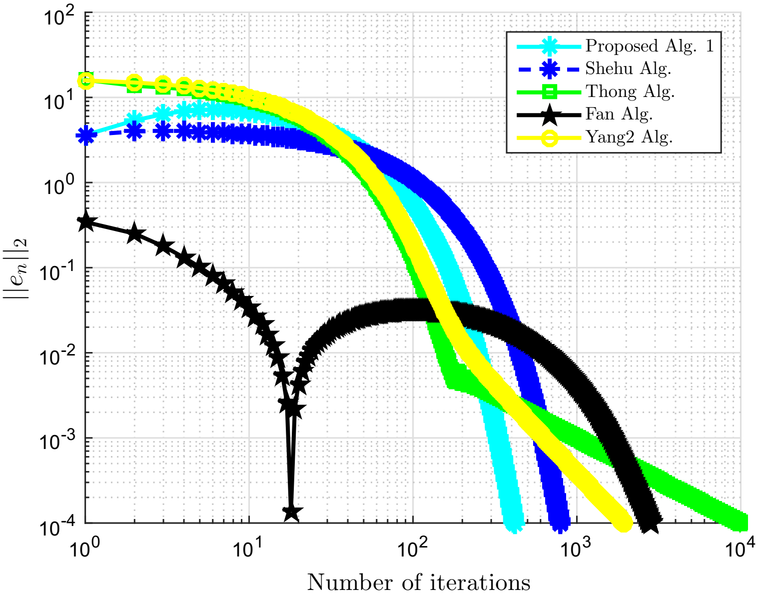

Fig. 9 Example 7.3: Case III

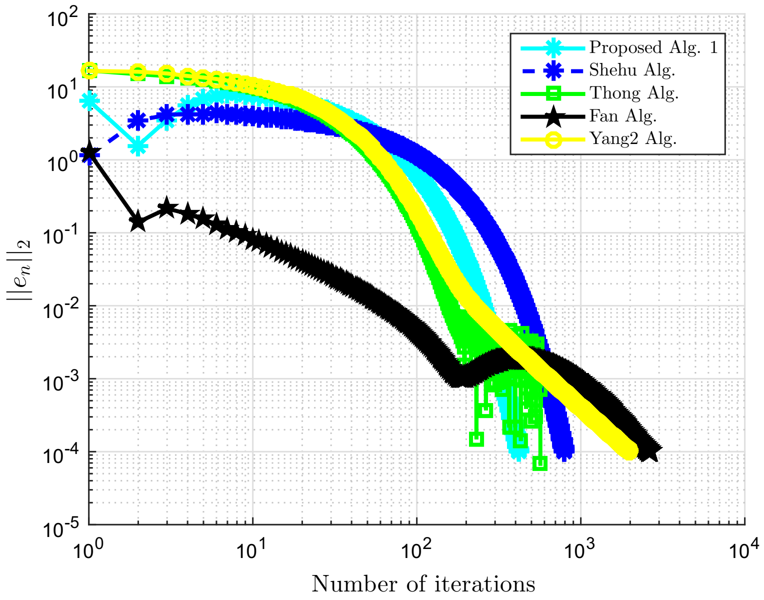

Fig. 10 Example 7.3: Case IV

8 Conclusion

We have proposed an inertial version of the subgradient extragradient method of Censor et al. [6] with the possibility $\theta_n = 1$ for the inertial factor and self-adaptive step sizes. Thus, in contrast with the methods considered so far, our proposed method extends the choices of the inertial factor to $\theta_n = 1$. We obtain weak convergence of our method in real Hilbert spaces under simpler conditions than previously assumed for other inertial subgradient extragradient methods. We also present a strong convergence result for the case where the cost function is strongly monotone with adaptive step sizes. In all our results the Lipschitz constant (or its estimate) of the cost function is not needed during implementations. Preliminary numerical experiments show that our methods are efficient and promising. One of the goals of our future projects is to estimate the rate of convergence for Algorithm 1.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. Aubin, J.-P., Ekeland, I.: Applied Nonlinear Analysis. Wiley, New York (1984)
2. Baiocchi, C., Capelo, A.: Variational and Quasivariational Inequalities; Applications to Free Boundary Problems. Wiley, New York (1984)
3. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics. Springer, New York (2011)

4. Boţ, R.I., Csetnek, E.R., Vuong, P.T.: The forward-backward-forward method from discrete and continuous perspective for pseudomonotone variational inequalities in Hilbert spaces. Eur. J. Oper. Res. **287**(1), 49–60 (2020)
5. Censor, Y., Gibali, A., Reich, S.: Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space. Optim. Methods Softw. **26**, 827–845 (2011)
6. Censor, Y., Gibali, A., Reich, S.: The subgradient extragradient method for solving variational inequalities in Hilbert space. J. Optim. Theory Appl. **148**, 318–335 (2011)
7. Chang, X., Liu, S., Deng, Z., Li, S.: An inertial subgradient extragradient algorithm with adaptive stepsizes for variational inequality problems. Optim. Methods Softw. (2021). https://doi.org/10.1080/10556788.2021.1910946
8. Ceng, L.C., Petrusel, A., Qin, X., Yao, J.C.: A modified inertial subgradient extragradient method for solving pseudomonotone variational inequalities and common fixed point problems. Fixed Point Theory **21**, 93–108 (2020)
9. Ceng, L.C., Petrusel, A., Qin, X., Yao, J.C.: Two inertial subgradient extragradient algorithms for variational inequalities with fixed-point constraints. Optimization **70**, 1337–1358 (2021)
10. Ceng, L.C., Shang, M.: Hybrid inertial subgradient extragradient methods for variational inequalities and fixed point problems involving asymptotically nonexpansive mappings. Optimization **70**, 715–740 (2021)
11. Ceng, L.C., Yuan, Q.: Composite inertial subgradient extragradient methods for variational inequalities and fixed point problems. J. Inequal. Appl. Paper No. 274 (2019)
12. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer Series in Operations Research, vol. II. Springer, New York (2003)
13. Fan, J., Liu, L., Qin, X.: A subgradient extragradient algorithm with inertial effects for solving strongly pseudomonotone variational inequalities. Optimization **69**, 2199–2215 (2020)
14. Glowinski, R., Lions, J.-L., Trémolières, R.: Numerical Analysis of Variational Inequalities. North-Holland, Amsterdam (1981)
15. Harker, P.T., Pang, J.-S.: A damped-Newton method for the linear complementarity problem. In: Allgower, G., Georg, K. (eds.) Computational Solution of Nonlinear Systems of Equations, Lectures in Applied Mathematics, vol. 26, pp. 265–284. AMS, Providence (1990)
16. He, L., Cui, Y.L., Ceng, L.C., Zhao, T.Y., Wang, D.Q., Hu, H.Y.: Strong convergence for monotone bilevel equilibria with constraints of variational inequalities and fixed points using subgradient extragradient implicit rule. J. Inequal. Appl. Paper No. 146 (2021)
17. Khobotov, E.N.: Modification of the extragradient method for solving variational inequalities and certain optimization problems. USSR Comput. Math. Math. Phys. **27**, 120–127 (1989)
18. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and Their Applications. Academic Press, New York (1980)
19. Konnov, I.V.: Combined Relaxation Methods for Variational Inequalities. Springer-Verlag, Berlin (2001)
20. Korpelevich, G.M.: The extragradient method for finding saddle points and other problems. Ekonomika i Mat. Metody **12**, 747–756 (1976)
21. Maingé, P.-E.: Convergence theorems for inertial KM-type algorithms. J. Comput. Appl. Math. **219**, 223–236 (2008)
22. Malitsky, Yu.V., Semenov, V.V.: A hybrid method without extrapolation step for solving variational inequality problems. J. Global Optim. **61**, 193–202 (2015)
23. Marcotte, P.: Applications of Khobotov's algorithm to variational and network equlibrium problems. Inf. Syst. Oper. Res. **29**, 258–270 (1991)
24. Nagurney, A.: Network Economics: A Variational Inequality Approach. Kluwer Academic Publishers, Dordrecht (1999)
25. Opial, Z.: Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull. Am. Math. Soc. **73**, 591–597 (1967)
26. Shehu, Y., Dong, Q.-L., Jiang, D.: Single projection method for pseudo-monotone variational inequality in Hilbert spaces. Optimization **68**, 385–409 (2019)
27. Shehu, Y., Iyiola, O.S., Reich, S.: A modified inertial subgradient extragradient method for solving variational inequalities. Optim. Eng. (2021). https://doi.org/10.1007/s11081-020-09593-w
28. Solodov, M.V., Svaiter, B.F.: A new projection method for variational inequality problems. SIAM J. Control Optim. **37**, 765–776 (1999)
29. Thong, D.V., Yang, J., Cho, Y.J., Rassias, T.M.: Explicit extragradient-like method with adaptive stepsizes for pseudomonotone variational inequalities. Optim. Lett. **15**, 2181–2199 (2021)
30. Thong, D.V., Van Hieu, D., Rassias, T.M.: Self adaptive inertial subgradient extragradient algorithms for solving pseudomonotone variational inequality problems. Optim. Lett. **14**, 115–144 (2020)
31. Thong, D.V., Hieu, D.V.: Inertial extragradient algorithms for strongly pseudomonotone variational inequalities. J. Comput. Appl. Math. **341**, 80–98 (2018)

32. Tian, M., Tong, M.: Self-adaptive subgradient extragradient method with inertial modification for solving monotone variational inequality problems and quasi-nonexpansive fixed point problems. J. Inequal. Appl. Paper No. 7 (2019)
33. Vuong, P.T.: On the weak convergence of the extragradient method for solving pseudo-monotone variational inequalities. J. Optim. Theory Appl. **176**, 399–409 (2018)
34. Yang, J.: Self-adaptive inertial subgradient extragradient algorithm for solving pseudomonotone variational inequalities. Appl. Anal. **100**, 1067–1078 (2021)
35. Yang, J., Liu, H., Liu, Z.: Modified subgradient extragradient algorithms for solving monotone variational inequalities. Optimization **67**, 2247–2258 (2018)
36. Yao, J.C.: Variational inequalities with generalized monotone operators. Math. Oper. Res. **19**, 691–705 (1994)