Optimization Methods and Software

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/goms20

A new inertial projected reflected gradient method with application to optimal control problems

Chinedu Izuchukwu & Yekini Shehu

To cite this article: Chinedu Izuchukwu & Yekini Shehu (2024) A new inertial projected reflected gradient method with application to optimal control problems, Optimization Methods and Software, 39:1, 197-226, DOI: 10.1080/10556788.2023.2246168

To link to this article: https://doi.org/10.1080/10556788.2023.2246168

Published online: 15 Aug 2023.

OPTIMIZATION METHODS & SOFTWARE
2024, VOL. 39, NO. 1, 197–226
https://doi.org/10.1080/10556788.2023.2246168

Taylor & Francis
Taylor & Francis Group

A new inertial projected reflected gradient method with application to optimal control problems

Chinedu Izuchukwu[a] and Yekini Shehu[b]

[a]School of Mathematics, University of the Witwatersrand, Johannesburg, South Africa; [b]School of Mathematical Sciences, Zhejiang Normal University, Jinhua, People's Republic of China

ABSTRACT
The projected reflected gradient method has been shown to be a simple and elegant method for solving variational inequalities. The method involves one projection onto the feasible set and one evaluation of the cost operator per iteration and has been shown numerically to be more efficient than most available methods for solving variational inequalities. Convergence results for methods with similar elegant structures of projected reflected gradient method are still rare. In this paper, we present weak and linear convergence of a projected reflected gradient method with an inertial extrapolation step and give some applications arising from optimal control problems. We first obtain weak convergence result for the projected reflected gradient method with an inertial extrapolation step for solving variational inequalities under standard assumptions with self-adaptive step sizes. We further obtain a linear convergence rate when the cost operator is strongly monotone and Lipschitz continuous. Finally, we give some numerical applications arising from optimal control. Preliminary results show that our method is effective and efficient when compared to other related state-of-the-art methods in the literature and show the advantage gained by incorporating inertial terms into the projected reflected gradient methods.

ARTICLE HISTORY
Received 18 December 2021
Accepted 4 August 2023

KEYWORDS
Variational inequalities; inertial projection method; weak convergence; linear convergence; optimal control; Hilbert spaces

2020 MSC CLASSIFICATIONS
47H05; 47J20; 47J25; 65K15; 90C25

1. Introduction

Let $\mathcal{C}$ be a nonempty, closed and convex subset of a real Hilbert space $\mathbb{H}$ and $\mathcal{A}:\mathbb{H}\to\mathbb{H}$ be a continuous operator. The Variational Inequality Problem (for short, VIP) is defined as:

$$\text{Find } v\in\mathcal{C} \text{ such that } \langle\mathcal{A}v, y-v\rangle \geq 0\ \forall y\in\mathcal{C}. \tag{1}$$

Let S denote the solutions set of VIP (1). Several problems in economics, engineering mechanics, mathematical programming, transportation and many more can be modelled as VIP (1) (see, for example, [2,11,18]).

Assumption 1.1. *In the convergence analysis of this paper, we assume that the following conditions are satisfied:*

CONTACT Yekini Shehu yekini.shehu@zjnu.edu.cn, deltanougt2006@yahoo.com School of Mathematical Sciences, Zhejiang Normal University, Jinhua 321004, People's Republic of China

(a) *The feasible set $\mathcal{C}$ is a nonempty, closed and convex subset of $\mathbb{H}$.*
(b) *$\mathcal{A} : \mathbb{H} \to \mathbb{H}$ is monotone and L-Lipschitz continuous.*
(c) *The solution set S of VIP (1) is nonempty.*

The extragradient method proposed in [12] is one of the popular methods to solve VIP (1) when $\mathcal{A}$ is monotone (or pseudo-monotone) and Lipschitz continuous in Hilbert spaces. The extragradient method is given by: $x_0 \in \mathcal{C}$, $\lambda_n \in (0, \frac{1}{L})$ and $L > 0$,

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} x_n) \\ x_{n+1} = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} y_n), \end{cases} \tag{2}$$

where $P_{\mathcal{C}}$ is a projection onto a set $\mathcal{C}$ and $\{x_n\}$ converges weakly to a solution of VIP (1) with $\|x_n - y_n\| \to 0, n \to \infty$. It can be seen that extragradient method (2) has two major drawbacks, in particular, it involves two evaluations of the cost operator $\mathcal{A}$ and also involves the computations of two projections onto the feasible set $\mathcal{C}$ per iteration. These make the extragradient method (2) computationally expensive in situations where $\mathcal{A}$ has a complex evaluation and the structure of $\mathcal{C}$ is complex. Therefore, the need to reduce the number of evaluations of $\mathcal{A}$ and projections onto $\mathcal{C}$ per iteration for the extragradient method (2) arises.

An attempt in this direction was initiated by Popov in [20]. Popov [20] introduced the following method: $x_0, y_0 \in \mathcal{C}$, $\lambda_n \in (0, \frac{1}{3L}]$ and $L > 0$,

$$\begin{cases} x_{n+1} = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} y_n) \\ y_{n+1} = P_{\mathcal{C}}(x_{n+1} - \lambda_n \mathcal{A} y_n). \end{cases} \tag{3}$$

Method (3) converges weakly when $\mathcal{A}$ is pseudo-monotone and Lipschitz continuous in Hilbert spaces. The Popov's extragradient method (3) requires only one evaluation of $\mathcal{A}$ per iteration but two computations of projections onto $\mathcal{C}$ per iteration.

Another modification of the extragradient method (2) is the subgradient extragradient method introduced by Censor et al. in [7], which is given by: $x_0 \in \mathbb{H}$, $\lambda_n \in (0, \frac{1}{L})$ and $L > 0$,

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} x_n) \\ D_n := \{w \in \mathbb{H} : \langle x_n - \lambda_n \mathcal{A} x_n - y_n, w - y_n \rangle \leq 0\} \\ x_{n+1} = P_{D_n}(x_n - \lambda_n \mathcal{A} y_n). \end{cases} \tag{4}$$

Several authors have established weak convergence of the subgradient extragradient method (4) when $\mathcal{A}$ is monotone (or pseudo-monotone) and Lipchitz continuous (see, for example, [5,6]).

The forward-backward-forward method (also known as Tseng's method) introduced by Tseng in [23] is another modification of the extragradient method (2), which is given by: $x_0 \in \mathbb{H}$, $\lambda_n \in (0, \frac{1}{L})$ and $L > 0$,

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} x_n) \\ x_{n+1} = y_n + \lambda_n \left(\mathcal{A} x_n - \mathcal{A} y_n\right). \end{cases} \tag{5}$$

The weak convergence of the forward-backward-forward method (5) was established in [23]. We mention also that both the subgradient extragradient method (4) and the forward-backward-forward method (5) require only one computation of projection onto $\mathcal{C}$ per iteration but two evaluations of $\mathcal{A}$ are needed per iteration. Inertial extrapolation type of the above-mentioned methods (2)–(5) have also been studied and convergence results obtained, for example, in [8,24].

In [16], Malitsky introduced the following projected reflected gradient method: choose $x_0 = y_0 \in \mathbb{H}$, $\lambda_n \in (0, \frac{\sqrt{2}-1}{L})$ and $L > 0$,

$$\begin{cases} x_{n+1} = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} y_n) \\ y_{n+1} = 2x_{n+1} - x_n, \end{cases} \tag{6}$$

and obtained weak convergence results for solving VIP (1) in real Hilbert spaces when $\mathcal{A}$ is monotone and Lipschitz continuous on $\mathbb{H}$. Linear convergence results are also obtained when $\mathcal{A}$ is strongly monotone and Lipschitz continuous. The main feature of method (6) is that it requires only one projection onto the feasible set $\mathcal{C}$ with no need for further projections either onto the half-space or feasible set $\mathcal{C}$ and one evaluation of $\mathcal{A}$ per iteration. It has been shown numerically in [16] that method (6) is computationally cheaper than the extragradient method (2), Popov's extragradient method (3), subgradient extragradient method (4) and forward-backward-forward method (5). Several versions of the projected reflected gradient method (6) have been proposed in the literature for solving VIP (1) in real Hilbert spaces and weak convergence results obtained when $\mathcal{A}$ is monotone and Lipschitz continuous on $\mathbb{H}$ (see, for example, [13–15]).

Yang and Liu [25] proposed a modification of the projected reflected gradient method (6), that does not require either the knowledge of the Lipschitz constant of the operator or additional projections in Hilbert spaces: $x_0 = y_0 \in \mathbb{H}$,

$$\begin{cases} x_{n+1} = P_{\mathcal{C}}(x_n - \lambda_n \mathcal{A} y_n) \\ y_{n+1} = x_{n+1} + \delta(x_{n+1} - x_n), \end{cases} \tag{7}$$

where $\delta \in (1, \infty)$ and for $\lambda_0 > 0$, λ_n is given self-adaptively in [25, Algorithm A]. Yang and Liu [25] obtained weak and linear convergence results for Algorithm (7) in real Hilbert spaces.

In [9], Dong et al. proposed the following general inertial projected gradient method with a self-adaptive step sizes for solving VIP (1): $x_0 \in \mathcal{C}, w_0, y_0 \in \mathbb{H}$,

$$\begin{cases} x_{n+1} = P_{\mathcal{C}}(w_n - \lambda_n \mathcal{A} y_n) \\ y_{n+1} = x_{n+1} + \delta(x_{n+1} - x_n) \\ w_{n+1} = x_{n+1} + \theta(x_{n+1} - x_n), \end{cases} \tag{8}$$

where $\delta \in (1, \infty), \theta \in [0, \frac{\delta(\delta-1)}{3\delta^2-1})$ and for $\lambda_0 > 0$, λ_n is given self-adaptively in [9, Algorithm 1]. Dong et al. [9] obtained weak and linear convergence results for Algorithm (8) in a real Hilbert space. It can be seen that the proposed method (7) is a special case of the general inertial projected gradient method (8) when $\theta = 0$ in (8).

Furthermore, it is observed that the choice $\delta = 1$ is not possible in general inertial projected gradient method (8) and in method (7) since $\delta \in (1, \infty)$. Therefore, methods (7) and (8) are not direct extensions of the projected reflected gradient method (6) and hence the results of Yang and Liu [25] and Dong et al. [9] cannot be reduced to the results of Malitsky [16].

Recently, Thong et al. [17] studied a generalized variational inequality problem in real Hilbert spaces and proposed a method for which the projected reflected gradient method

(6) is a special case. We state the abridged version of the method proposed in [17] for the VIP (1) that we are considering: $x_0 = y_0 \in \mathbb{H}$,

$$\begin{cases} x_{n+1} = P_C(x_n - \lambda_n \mathcal{A} y_n) \\ y_{n+1} = x_{n+1} + \delta(x_{n+1} - x_n), \end{cases} \tag{9}$$

where $\delta > \frac{1}{2}$ and for $\lambda_0 > 0$, λ_n is given self-adaptively in [17, Algorithm 1]. Weak convergence of method (9) is established under standard assumptions and linear convergence is presented under strong monotonicity of $\mathcal{A}$. It should be noted that when $\delta = 1$ in method (9), we have projected reflected gradient method (6). However, method (9) does not contain inertial extrapolation step in the form of (8).

The following natural question arises:

Question: Can we propose a projection type method with one evaluation of $\mathcal{A}$, one computation of P_C, inertial extrapolation step and self-adaptive step sizes to solve VIP (1) for which projected reflected gradient method (6) is a special case?

Our aim in this paper is to answer the above question in the affirmative. We propose a projection method to solve VIP (1) with the following features:

- our method involves one projection onto feasible set C per iteration;
- one evaluation of $\mathcal{A}$ is only needed per iteration;
- inertial extrapolation step is incorporated to speed up the iterations;
- self-adaptive step sizes are adopted;
- we recover projected reflected gradient method (6).

Our contributions in this paper are:

- We propose an inertial projected reflected gradient method to solve VIP (1) and obtain weak convergence results when $\mathcal{A}$ is monotone and Lipschitz continuous in real Hilbert spaces. Our method is simple and does not require either further evaluations of $\mathcal{A}$ or further projections. Our results improve on the results and methods proposed in [9,25] by showing that the choice $\delta = 1$ is possible. Furthermore, the adaptive step sizes are suitably updated at each iteration, independently of the Lipschitz constant of the cost operator $\mathcal{A}$, and without any line search procedure unlike the approach used in [13,14].
- Linear convergence results are obtained when $\mathcal{A}$ is strongly monotone and Lipschitz continuous.
- We present numerical results obtained by our proposed method to solve problems arising from optimal control and compare with the methods in [9,16,17,25]. Computational results show that the proposed method is more efficient and converges faster (in terms of CPU time and number of iterations) than the methods in [9,16,17,25].

We organize the rest of the paper as follows: Section 2 contains basic definitions and results needed in subsequent sections. In Section 3, we present and discuss the proposed method. The weak convergence results are given in Section 4 while linear convergence results are given in Section 5. We give some numerical illustrations in Section 6 and concluding remarks are given in Section 7.

2. Preliminaries

This section contains some definitions and basic results that will be used in subsequent analyses. We first state the formal definition of some classes of operators that play an important role in our analysis.

Definition 2.1: An operator $\mathcal{A} : \mathbb{H} \to \mathbb{H}$ is called

(i) *L-Lipschitz continuous* with constant $L > 0$ if

$$\|\mathcal{A}u - \mathcal{A}v\| \le L\|u - v\| \ \forall u, \quad v \in \mathbb{H},$$

(ii) *η-strongly monotone* with constant $\eta > 0$ if

$$\langle \mathcal{A}u - \mathcal{A}v, u - v\rangle \ge \eta\|u - v\|^2 \ \forall u, v \in \mathbb{H},$$

(iii) *monotone*, if

$$\left\langle \mathcal{A}u - \mathcal{A}v, u - v\right\rangle \ge 0 \ \forall u, \quad v \in \mathbb{H}.$$

We next recall some properties of the projection, cf. [3] for more details. To this end, let $\mathcal{C} \subseteq \mathbb{H}$ be a nonempty, closed and convex subset of $\mathbb{H}$. For any point $u \in \mathbb{H}$, there exists a unique point $P_{\mathcal{C}}u \in \mathcal{C}$ such that

$$\|u - P_{\mathcal{C}}u\| \le \|u - y\| \ \ \forall y \in \mathcal{C}.$$

$P_{\mathcal{C}}$ is called the *metric projection* of $\mathbb{H}$ onto $\mathcal{C}$. We know that $P_{\mathcal{C}}$ is a nonexpansive mapping of $\mathbb{H}$ onto $\mathcal{C}$. It is also known that $P_{\mathcal{C}}$ satisfies

$$\langle x - y, P_{\mathcal{C}}x - P_{\mathcal{C}}y\rangle \ge \|P_{\mathcal{C}}x - P_{\mathcal{C}}y\|^2 \ \ \forall x, \quad y \in \mathbb{H}. \tag{10}$$

Furthermore, $P_{\mathcal{C}}x$ is characterized by the properties

$$P_{\mathcal{C}}x \in \mathcal{C} \quad \text{and} \quad \langle x - P_{\mathcal{C}}x, P_{\mathcal{C}}x - y\rangle \ge 0 \ \ \forall y \in \mathcal{C}. \tag{11}$$

This characterization implies that

$$\|x - y\|^2 \ge \|x - P_{\mathcal{C}}x\|^2 + \|y - P_{\mathcal{C}}x\|^2 \ \ \forall x \in \mathbb{H}, \forall y \in \mathcal{C}. \tag{12}$$

Lemma 2.2. *The following identities hold for all $u, v, w \in \mathbb{H}$ and $\kappa \in \mathbb{R}$:*

(i) $2\langle u, v\rangle = \|u\|^2 + \|v\|^2 - \|u - v\|^2 = \|u + v\|^2 - \|u\|^2 - \|v\|^2$;
(ii) $\|\kappa u + (1 - \kappa)v\|^2 = \kappa\|u\|^2 + (1 - \kappa)\|v\|^2 - \kappa(1 - \kappa)\|u - v\|^2$.

Lemma 2.3 (([22, Lem. 7.1.7]).). *Let $\mathcal{C}$ be a nonempty, closed and convex subset of $\mathbb{H}$. Let $\mathcal{A} : \mathcal{C} \to \mathbb{H}$ be a continuous, monotone mapping and $z \in \mathcal{C}$. Then*

$$z \in S \iff \langle \mathcal{A}x, x - z\rangle \ge 0 \quad \textit{for all } x \in \mathcal{C}.$$

3. Proposed method

In this section, we present our method from the point of view of a dynamical system and later obtain the discrete version.

3.1. Derivations from dynamical systems

Assume that $\mathcal{A}$ is a monotone operator in a real Hilbert space $\mathbb{H}$ and consider the following dynamical system associated with the VIP (1):

$$\dot{x}(t) + x(t) = P_C(x(t) - \lambda(t)A(\dot{x}(t) + x(t))), \tag{13}$$

where $\lambda : [0, \infty) \to [0, \infty)$ is a Lebesgue measurable function. In order to discretize (13), we replace $x(t) \approx x_n$. Since two derivatives appear in (13), we have many combinations of possible discretizations. We consider a forward discretization of $\dot{x}(t)$ on the left-hand side and a backward discretization of $\dot{x}(t)$ on the right-hand side. Thus,

$$\dot{x}(t) \approx x_{n+1} - x_n \text{ on the left hand side}$$

and

$$\dot{x}(t) \approx x_n - x_{n-1} \text{ on the right hand side.}$$

Therefore, we obtain

$$\begin{aligned} x_{n+1} &= P_C(x_n - \lambda_n A(x_n + x_n - x_{n-1})) \\ &= P_C(x_n - \lambda_n A(2x_n - x_{n-1})). \end{aligned} \tag{14}$$

This is the projected reflected gradient method introduced by Malitsky in [16] for solving VIP (1).

In this paper, we propose, instead of (13), the following dynamical system

$$\dot{x}(t) + x(t) = P_C(\theta\dot{x}(t) + x(t) - \lambda(t)A(\dot{x}(t) + x(t))), \tag{15}$$

where $\theta \geq 0$. Again, discretizing (15) as done in (13), we obtain

$$x_{n+1} = P_C(x_n + \theta(x_n - x_{n-1}) - \lambda_n A(2x_n - x_{n-1})). \tag{16}$$

This is the projected reflected gradient method with inertial term which we intend to study in this paper. Observe that method (16) reduces to method (14) when $\theta = 0$.

3.2. Our method in discrete version

In this subsection, we introduce and discuss our projection-type method to solve VIP (1). Our proposed method is a combination of projected reflected gradient method, inertial extrapolation term and self-adaptive step sizes.

Remark 3.1: Note that if $w_n = y_n = x_{n+1}$, then (17) implies the equality $x_{n+1} = P_C(x_{n+1} - \lambda_n \mathcal{A}x_{n+1})$ and so $x_{n+1} \in S$.

Remark 3.2: Note that by (18), $\lambda_{n+1} \leq \lambda_n$. Furthermore, if $\langle \mathcal{A}y_n - \mathcal{A}y_{n-1}, y_n - x_{n+1}\rangle > 0$, then using the definition of ρ_n, we obtain

$$\begin{aligned}\lambda_{n+1} &= \frac{\mu\rho_n}{\langle \mathcal{A}y_n - \mathcal{A}y_{n-1}, y_n - x_{n+1}\rangle}\\ &\geq \frac{\mu\left(\|y_n - y_{n-1}\|^2 + 2\|y_n - x_{n+1}\|^2\right)}{\langle \mathcal{A}y_n - \mathcal{A}y_{n-1}, y_n - x_{n+1}\rangle}\\ &\geq \frac{2\sqrt{2}\mu\|y_n - y_{n-1}\|\|y_n - x_{n+1}\|}{\|\mathcal{A}y_n - \mathcal{A}y_{n-1}\|\|y_n - x_{n+1}\|}\\ &\geq \frac{2\sqrt{2}\mu}{L}.\end{aligned}$$

Therefore, by the definition of $\{\lambda_n\}$ in (18), we have

$$\lambda_n \geq \min\left\{\lambda_0, \frac{2\sqrt{2}\mu}{L}\right\}.$$

Consequently, there exists $\lambda > 0$ such that

$$\lim_{n\to\infty}\lambda_n = \lambda \geq \min\left\{\lambda_0, \frac{2\sqrt{2}\mu}{L}\right\} > 0.$$

4. Weak convergence results

We present the weak convergence result of Algorithm 1 in this Section. First, we state and prove the following important lemmas that are crucial for our convergence analysis.

Lemma 4.1. *Suppose $\{x_n\}$ is generated by Algorithm 1. Assume further that the conditions in Assumption 1.1 are satisfied. Then $\forall x^* \in S$,*

$$\begin{aligned}&2\lambda_n\langle \mathcal{A}y_{n-1} - \mathcal{A}y_n, y_n - x_{n+1}\rangle + 2\lambda_n\langle \mathcal{A}y_n, y_n - x^*\rangle + 2\langle x_{n+1} - w_n, x_{n+1} - x^*\rangle\\ &\leq (1-\theta)\frac{\lambda_n}{\lambda_{n-1}}\|x_{n+1} - x_n\|^2 + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2\\ &\quad + \theta\|x_n - y_{n-1}\|^2 + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2. \qquad (19)\end{aligned}$$

Proof: Using $x_n = P_C(w_{n-1} - \lambda_{n-1}\mathcal{A}y_{n-1})$, we obtain by (11) that $\forall x \in \mathcal{C}$,

$$\langle x_n - w_{n-1} + \lambda_{n-1}\mathcal{A}y_{n-1}, x - x_n\rangle \geq 0. \qquad (20)$$

Setting $x = x_{n+1}$ and later $x = x_{n-1}$ in (20), we have

$$\langle x_n - w_{n-1} + \lambda_{n-1}\mathcal{A}y_{n-1}, x_{n+1} - x_n\rangle \geq 0 \qquad (21)$$

and

$$\langle x_n - w_{n-1} + \lambda_{n-1}\mathcal{A}y_{n-1}, x_{n-1} - x_n\rangle \geq 0. \qquad (22)$$

Algorithm 1 Inertial Projected Reflected Gradient Method

1: Let $x_{-1}, x_0 \in \mathcal{C}$ be chosen arbitrarily. Choose $\lambda_0 > 0$ and $\theta \in [0, \frac{1}{7})$. Define $\bar{\mu} := \frac{1-7\theta}{4+2\sqrt{2}}$ and choose $\mu \in (0, \bar{\mu})$. Set $n := 0$.

2: Compute

$$\begin{cases} w_n = x_n + \theta(x_n - x_{n-1}) \\ y_n = 2x_n - x_{n-1} \\ x_{n+1} = P_{\mathcal{C}}(w_n - \lambda_n \mathcal{A} y_n). \end{cases} \tag{17}$$

If $w_n = y_n = x_{n+1}$, STOP and $x_{n+1} \in S$. Otherwise

3: Compute

$$\lambda_{n+1} = \begin{cases} \lambda_n, & \text{if } \langle \mathcal{A}y_n - \mathcal{A}y_{n-1}, y_n - x_{n+1} \rangle \le 0 \\ \min\left\{ \frac{\mu \rho_n}{\langle \mathcal{A}y_n - \mathcal{A}y_{n-1}, y_n - x_{n+1} \rangle}, \lambda_n \right\}, & \text{otherwise,} \end{cases} \tag{18}$$

where $\rho_n := \sqrt{2}\|y_{n-1} - x_n\|^2 + (2 + \sqrt{2})\|x_n - y_n\|^2 + 2\|x_{n+1} - y_n\|^2$.

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

Adding (21) and (22) gives

$$\langle x_n - w_{n-1} + \lambda_{n-1} \mathcal{A} y_{n-1}, x_{n+1} - x_n + x_{n-1} - x_n \rangle \ge 0. \tag{23}$$

Noting that $y_n = 2x_n - x_{n-1}$, we have from (23) that

$$\langle x_n - w_{n-1} + \lambda_{n-1} \mathcal{A} y_{n-1}, x_{n+1} - y_n \rangle \ge 0.$$

Therefore, we get

$$\langle \lambda_{n-1} \mathcal{A} y_{n-1}, y_n - x_{n+1} \rangle \le \langle x_n - w_{n-1}, x_{n+1} - y_n \rangle. \tag{24}$$

Now, by $w_n = x_n + \theta(x_n - x_{n-1})$, one has

$$\begin{aligned} x_n - w_{n-1} &= x_n - x_{n-1} - \theta(x_{n-1} - x_{n-2}) \\ &= x_n - x_{n-1} - \theta\left[(x_n - x_{n-1}) - (x_n - y_{n-1})\right] \\ &= (1-\theta)(x_n - x_{n-1}) + \theta(x_n - y_{n-1}) \\ &= (1-\theta)(y_n - x_n) + \theta(x_n - y_{n-1}). \end{aligned} \tag{25}$$

Using (25) in (24) gives

$$
\begin{aligned}
2\lambda_n\langle \mathcal{A}y_{n-1}, y_n - x_{n+1}\rangle &\le \frac{2\lambda_n}{\lambda_{n-1}}\langle x_n - w_{n-1}, x_{n+1} - y_n\rangle \\
&= \frac{2\lambda_n(1-\theta)}{\lambda_{n-1}}\langle y_n - x_n, x_{n+1} - y_n\rangle + \frac{2\lambda_n\theta}{\lambda_{n-1}}\langle x_n - y_{n-1}, x_{n+1} - y_n\rangle \\
&\le \frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\left[\|x_{n+1} - x_n\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - y_n\|^2\right] \\
&\quad + \frac{\lambda_n\theta}{\lambda_{n-1}}\left[\|x_n - y_{n-1}\|^2 + \|x_{n+1} - y_n\|^2\right] \\
&\le (1-\theta)\frac{\lambda_n}{\lambda_{n-1}}\|x_{n+1} - x_n\|^2 + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2 \\
&\quad + \theta\|x_n - y_{n-1}\|^2 + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2. \qquad (26)
\end{aligned}
$$

Again using $x_{n+1} = P_C(w_n - \lambda_n \mathcal{A}y_n)$, we obtain from (11) that $\forall x \in \mathcal{C}$,

$$\langle x_{n+1} - w_n + \lambda_n \mathcal{A}y_n, x - x_{n+1}\rangle \ge 0.$$

Since $x^* \in S \subset \mathcal{C}$, we have in particular that

$$\langle x_{n+1} - w_n + \lambda_n \mathcal{A}y_n, x_{n+1} - x^*\rangle \le 0. \qquad (27)$$

Addition of (26) and (27) gives

$$
\begin{aligned}
&2\lambda_n\langle \mathcal{A}y_{n-1}, y_n - x_{n+1}\rangle + 2\langle x_{n+1} - w_n + \lambda_n \mathcal{A}y_n, x_{n+1} - x^*\rangle \\
&\le (1-\theta)\frac{\lambda_n}{\lambda_{n-1}}\|x_{n+1} - x_n\|^2 + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2 \\
&\quad + \theta\|x_n - y_{n-1}\|^2 + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2.
\end{aligned}
$$

One can rewrite the last inequality above as (19). ■

Lemma 4.2. *Under the same conditions as assumed Lemma 4.1, we have*

$$
\begin{aligned}
\|x_{n+1} - x^*\|^2 \le\ & \|w_n - x^*\|^2 - \|x_{n+1} - w_n\|^2 \\
& + \frac{\lambda_n}{\lambda_{n+1}} 2\sqrt{2}\mu\left[\|y_{n-1} - x_n\|^2 + (1+\sqrt{2})\|x_n - y_n\|^2 + \sqrt{2}\|x_{n+1} - y_n\|^2\right] \\
& - 4\lambda_n\langle \mathcal{A}x^*, x_n - x^*\rangle + 2\lambda_n\langle \mathcal{A}x^*, x_{n-1} - x^*\rangle + (1-\theta)\|x_{n+1} - x_n\|^2 \\
& + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2 + \theta\|x_n - y_{n-1}\|^2 \\
& + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2. \qquad (28)
\end{aligned}
$$

Proof: Observe by Lemma 2.2 (i) we have that

$$2\langle x_{n+1}-w_n, x_{n+1}-x^*\rangle = \|x_{n+1}-w_n\|^2 + \|x_{n+1}-x^*\|^2 - \|w_n-x^*\|^2. \tag{29}$$

Therefore, (19) becomes

$$\begin{aligned}\|x_{n+1}-x^*\|^2 \le \|w_n-x^*\|^2 - \|x_{n+1}-w_n\|^2 + 2\lambda_n\langle \mathcal{A}y_{n-1}-\mathcal{A}y_n, x_{n+1}-y_n\rangle \\ - 2\lambda_n\langle \mathcal{A}y_n, y_n-x^*\rangle + (1-\theta)\frac{\lambda_n}{\lambda_{n-1}}\|x_{n+1}-x_n\|^2 \\ + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-y_n\|^2 + \theta\|x_n-y_{n-1}\|^2 \\ + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2.\end{aligned} \tag{30}$$

Observe from (18) that

$$\begin{aligned}2\lambda_n\langle \mathcal{A}y_{n-1}-\mathcal{A}y_n, x_{n+1}-y_n\rangle &= \frac{\lambda_n}{\lambda_{n+1}}2\lambda_{n+1}\langle \mathcal{A}y_{n-1}-\mathcal{A}y_n, x_{n+1}-y_n\rangle \\ &\le \frac{\lambda_n}{\lambda_{n+1}}2\sqrt{2}\mu\big[\|y_{n-1}-x_n\|^2 + (1+\sqrt{2})\|x_n-y_n\|^2 \\ &\quad + \sqrt{2}\|x_{n+1}-y_n\|^2\big].\end{aligned} \tag{31}$$

Since $\mathcal{A}$ is monotone on $\mathbb{H}$, we have

$$2\lambda_n\langle \mathcal{A}y_n, y_n-x^*\rangle \ge 2\lambda_n\langle \mathcal{A}x^*, y_n-x^*\rangle. \tag{32}$$

Now, by (17), we have

$$\begin{aligned}y_n - x^* &= 2x_n - x_{n-1} - x^* \\ &= 2(x_n-x^*) - (x_{n-1}-x^*).\end{aligned} \tag{33}$$

Using (33), we obtain

$$\begin{aligned}2\lambda_n\langle \mathcal{A}x^*, y_n-x^*\rangle &= 2\lambda_n\langle \mathcal{A}x^*, 2(x_n-x^*) - (x_{n-1}-x^*)\rangle \\ &= 4\lambda_n\langle \mathcal{A}x^*, x_n-x^*\rangle - 2\lambda_n\langle \mathcal{A}x^*, x_{n-1}-x^*\rangle.\end{aligned} \tag{34}$$

If we use (34) in (32), we then obtain

$$2\lambda_n\langle \mathcal{A}y_n, y_n-x^*\rangle \ge 4\lambda_n\langle \mathcal{A}x^*, x_n-x^*\rangle - 2\lambda_n\langle \mathcal{A}x^*, x_{n-1}-x^*\rangle. \tag{35}$$

Combining (30), (31) and (35), we have (28) (noting that $\frac{\lambda_n}{\lambda_{n-1}} \le 1$). ■

Lemma 4.3. *Suppose* $\{x_n\}$ *is generated by Algorithm 1. Assume further that the conditions in Assumption 1.1 are satisfied and* $\forall x^* \in S$*, define*

$$a_n := \|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2 + 2\theta\|x_n - x_{n-1}\|^2 + \left(\theta + \frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}\right)\|x_n - y_{n-1}\|^2 + 2\lambda_{n-1}\langle \mathcal{A}x^*, x_{n-1} - x^*\rangle$$

and

$$b_n := \left(\frac{\lambda_n}{\lambda_{n-1}} - \alpha_n - 2\theta\right)\|x_{n+1} - x_n\|^2,$$

where

$$\alpha_n := \max\left\{\theta + \frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}},\ 3\theta + \frac{4\mu\lambda_n}{\lambda_{n+1}} + \frac{2\sqrt{2}\mu\lambda_{n+1}}{\lambda_{n+2}}\right\}.$$

Then there exists $n_0 > 0$ *such that* $\forall n \geq n_0$*, we have*

$$a_{n+1} \leq a_n - b_n, \text{ and } b_n \geq 0.$$

Proof: By Cauchy–Schwarz inequality, we get

$$\begin{aligned}\|x_{n+1} - w_n\|^2 &= \|(x_{n+1} - x_n) - \theta(x_n - x_{n-1})\|^2 \\ &= \|x_{n+1} - x_n\|^2 - 2\theta\langle x_{n+1} - x_n, x_n - x_{n-1}\rangle \\ &\quad + \theta^2\|x_n - x_{n-1}\|^2 \\ &\geq (1-\theta)\|x_{n+1} - x_n\|^2 + \theta(\theta - 1)\|x_n - x_{n-1}\|^2. \end{aligned} \tag{36}$$

Using Lemma 2.2 (ii) in (17), we get

$$\begin{aligned}\|w_n - x^*\|^2 &= \|(1+\theta)(x_n - x^*) - \theta(x_{n-1} - x^*)\|^2 \\ &= (1+\theta)\|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2 \\ &\quad + \theta(1+\theta)\|x_n - x_{n-1}\|^2. \end{aligned} \tag{37}$$

Applying (36) and (37) in (28), we obtain

$$\begin{aligned}\|x_{n+1}-x^*\|^2 \leq\ & (1+\theta)\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+\theta(1+\theta)\|x_n-x_{n-1}\|^2\\&-(1-\theta)\|x_{n+1}-x_n\|^2-\theta(\theta-1)\|x_n-x_{n-1}\|^2\\&+\frac{\lambda_n}{\lambda_{n+1}}2\sqrt{2}\mu\left[\|y_{n-1}-x_n\|^2+(1+\sqrt{2})\|x_n-y_n\|^2+\sqrt{2}\|x_{n+1}-y_n\|^2\right]\\&-4\lambda_n\langle\mathcal{A}x^*,x_n-x^*\rangle+2\lambda_n\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle+(1-\theta)\|x_{n+1}-x_n\|^2\\&+\left(\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-y_n\|^2+\theta\|x_n-y_{n-1}\|^2\\&+\left(2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2 \qquad (38)\\=\ &\|x_n-x^*\|^2+\theta\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+2\theta\|x_n-x_{n-1}\|^2\\&+\frac{\lambda_n}{\lambda_{n+1}}2\sqrt{2}\mu\left[\|y_{n-1}-x_n\|^2+(1+\sqrt{2})\|x_n-y_n\|^2+\sqrt{2}\|x_{n+1}-y_n\|^2\right]\\&-4\lambda_n\langle\mathcal{A}x^*,x_n-x^*\rangle+2\lambda_n\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle\\&+\left(\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-y_n\|^2+\theta\|x_n-y_{n-1}\|^2\\&+\left(2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2. \qquad (39)\end{aligned}$$

Observe that $\langle\mathcal{A}x^*,x_n-x^*\rangle\geq 0$ since $x^*\in S$. Similarly, we also have $\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle\geq 0$. By Remark 3.2, we have that $\lambda_n\leq\lambda_{n-1}$, and from $\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle\geq 0$, we obtain from (39) that

$$\begin{aligned}&\|x_{n+1}-x^*\|^2\\&\leq\|x_n-x^*\|^2+\theta\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+2\theta\|x_n-x_{n-1}\|^2\\&\quad+\frac{\lambda_n}{\lambda_{n+1}}2\sqrt{2}\mu\left[\|y_{n-1}-x_n\|^2+(1+\sqrt{2})\|x_n-y_n\|^2+\sqrt{2}\|x_{n+1}-y_n\|^2\right]\\&\quad-4\lambda_n\langle\mathcal{A}x^*,x_n-x^*\rangle+2\lambda_{n-1}\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle\\&\quad+\left(\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-y_n\|^2+\theta\|x_n-y_{n-1}\|^2+\left(2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2\\&\leq\|x_n-x^*\|^2+\theta\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+2\theta\|x_n-x_{n-1}\|^2\\&\quad+\frac{\lambda_n}{\lambda_{n+1}}2\sqrt{2}\mu\left[\|y_{n-1}-x_n\|^2+(1+\sqrt{2})\|x_n-y_n\|^2+\sqrt{2}\|x_{n+1}-y_n\|^2\right]\\&\quad-2\lambda_n\langle\mathcal{A}x^*,x_n-x^*\rangle+2\lambda_{n-1}\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle\\&\quad+\left(\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-y_n\|^2+\theta\|x_n-y_{n-1}\|^2\\&\quad+\left(2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2. \qquad (40)\end{aligned}$$

Therefore, we obtain

$$
\begin{aligned}
&\|x_{n+1}-x^*\|^2-\theta\|x_n-x^*\|^2+2\theta\|x_{n+1}-x_n\|^2\\
&\quad+\left(\theta+\frac{2\sqrt{2}\mu\lambda_{n+1}}{\lambda_{n+2}}\right)\|y_n-x_{n+1}\|^2+2\lambda_n\langle\mathcal{A}x^*,x_n-x^*\rangle\\
&\le\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+2\theta\|x_{n+1}-x_n\|^2+2\theta\|x_n-x_{n-1}\|^2\\
&\quad+\left(\theta+\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}\right)\|y_{n-1}-x_n\|^2+2\lambda_{n-1}\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle\\
&\quad-\left(\frac{\lambda_n}{\lambda_{n-1}}-\theta-\frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}\right)\|x_n-y_n\|^2\\
&\quad-\left(\frac{\lambda_n}{\lambda_{n-1}}-2\theta-\frac{4\mu\lambda_n}{\lambda_{n+1}}-\theta-\frac{2\sqrt{2}\mu\lambda_{n+1}}{\lambda_{n+2}}\right)\|y_n-x_{n+1}\|^2\\
&=\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+2\theta\|x_n-x_{n-1}\|^2+\left(\theta+\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}\right)\|y_{n-1}-x_n\|^2\\
&\quad+2\lambda_{n-1}\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle+2\theta\|x_{n+1}-x_n\|^2\\
&\quad-\left(\frac{\lambda_n}{\lambda_{n-1}}-\theta-\frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}\right)\|x_n-y_n\|^2\\
&\quad-\left(\frac{\lambda_n}{\lambda_{n-1}}-3\theta-\frac{4\mu\lambda_n}{\lambda_{n+1}}-\frac{2\sqrt{2}\mu\lambda_{n+1}}{\lambda_{n+2}}\right)\|x_{n+1}-y_n\|^2\\
&\le\|x_n-x^*\|^2-\theta\|x_{n-1}-x^*\|^2+2\theta\|x_n-x_{n-1}\|^2+\left(\theta+\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}\right)\|y_{n-1}-x_n\|^2\\
&\quad+2\lambda_{n-1}\langle\mathcal{A}x^*,x_{n-1}-x^*\rangle+2\theta\|x_{n+1}-x_n\|^2\\
&\quad-\left(\frac{\lambda_n}{\lambda_{n-1}}-\alpha_n\right)\left(\|x_n-y_n\|^2+\|x_{n+1}-y_n\|^2\right).
\end{aligned}
\tag{41}
$$

Note that $\lim_{n\to\infty}(\frac{\lambda_n}{\lambda_{n-1}}-\alpha_n)=1-3\theta-(4+2\sqrt{2})\mu>1-3\theta-(4+2\sqrt{2})\mu\ge 0$. Therefore, there exists $n_0>0$ such that $\frac{\lambda_n}{\lambda_{n-1}}-\alpha_n>0,\ \forall n\ge n_0$.

Using Cauchy–Schwarz inequality in (41), we get $\forall n\ge n_0$,

$$
\begin{aligned}
a_{n+1}&\le a_n+2\theta\|x_{n+1}-x_n\|^2-\frac{1}{2}\left(\frac{\lambda_n}{\lambda_{n-1}}-\alpha_n\right)\|x_{n+1}-x_n\|^2\\
&=a_n-\frac{1}{2}\left(\frac{\lambda_n}{\lambda_{n-1}}-\alpha_n-4\theta\right)\|x_{n+1}-x_n\|^2.
\end{aligned}
\tag{42}
$$

Since $\lim_{n\to\infty}\lambda_n = \lambda$ and $\lim_{n\to\infty}\alpha_n = 3\theta + (4+2\sqrt{2})\mu$, we have $\lim_{n\to\infty}(\frac{\lambda_n}{\lambda_{n-1}} - \alpha_n - 4\theta) = 1 - 3\theta - (4+2\sqrt{2})\mu - 4\theta > 0$, since $\mu < \frac{1-7\theta}{4+2\sqrt{2}}$. Therefore,

$$b_n = \frac{1}{2}\left(\frac{\lambda_n}{\lambda_{n-1}} - \alpha_n - 4\theta\right)\|x_{n+1} - x_n\|^2 > 0, \ \ \forall n \geq n_0.$$

Hence, (42) becomes

$$a_{n+1} \leq a_n - b_n, \quad \forall n \geq n_0. \tag{43}$$

■

Using Lemma 4.3, we obtain the following weak convergence result for VIP (1) under the conditions stated in Assumption 1.1.

Theorem 4.4. *Suppose the conditions in Assumption 1.1 are fulfilled and let $\{x_n\}$ be a sequence generated by Algorithm 1. Then $\{x_n\}$ converges weakly to a point in S.*

Proof: First we need to show that the sequence $\{a_n\}$ defined in Lemma 4.3 is non-negative. That is, we show that $a_n \geq 0, \ \ \forall n \geq n_0$.

From the definition of a_n we observe that $\forall n \geq n_0$,

$$a_n \geq \|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2 + 2\theta\|x_n - x_{n-1}\|^2. \tag{44}$$

By Lemma 2.2 (i), we obtain

$$\begin{aligned}\|x_{n-1} - x^*\|^2 &= \|(x_{n-1} - x_n) + (x_n - x^*)\|^2 \\ &= \|x_{n-1} - x_n\|^2 + \|x_n - x^*\|^2 + 2\langle x_{n-1} - x_n, x_n - x^*\rangle \\ &\leq 2\|x_{n-1} - x_n\|^2 + 2\|x_n - x^*\|^2. \end{aligned} \tag{45}$$

Combining (44) and (45), one has

$$\begin{aligned} a_n &\geq \|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2 + 2\theta\|x_n - x_{n-1}\|^2 \\ &\geq \|x_n - x^*\|^2 - 2\theta\|x_n - x_{n-1}\|^2 - 2\theta\|x_n - x^*\|^2 \\ &\quad + 2\theta\|x_n - x_{n-1}\|^2 \\ &= (1-2\theta)\|x_n - x^*\|^2, \quad \forall n \geq n_0. \end{aligned} \tag{46}$$

Hence, from (46), we have that $a_n \geq 0, \ \ \forall n \geq n_0$. This implies that $\{a_n\}$ is convergent since $a_{n+1} \leq a_n$ and $a_n \geq 0$. Thus, $\lim_{n\to\infty} a_n$ exists. Using (46), we have that $\{x_n\}$ is bounded.

From (42), we obtain that

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0.$$

Since $y_n = 2x_n - x_{n-1}$ and $w_n = x_n + \theta(x_n - x_{n-1})$, we have

$$\lim_{n\to\infty} \|y_n - x_n\| = \lim_{n\to\infty} \|x_n - x_{n-1}\| = 0 = \lim_{n\to\infty} \|w_n - x_n\|. \tag{47}$$

Also,

$$\|x_{n+1} - y_n\| \le \|x_{n+1} - x_n\| + \|y_n - x_n\| \to 0, \quad n \to \infty \tag{48}$$

and

$$\|x_{n+1} - w_n\| \le \|x_{n+1} - x_n\| + \|x_n - w_n\| \to 0, \quad n \to \infty. \tag{49}$$

Consequently, we have from the definition of $\{a_n\}$ that

$$\lim_{n\to\infty} \left[\|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2 + 2\lambda_{n-1}\langle \mathcal{A}x^*, x_{n-1} - x^*\rangle\right] \text{ exists.} \tag{50}$$

Since $\{x_n\}$ is bounded, there exists $\{x_{n_k}\} \subset \{x_n\}$ such that $x_{n_k} \rightharpoonup p \in \mathbb{H}$. One can also see that $y_{n_k} \rightharpoonup p$ by (47).

We next show that $p \in S$. By (11), one has $\forall y \in \mathcal{C}$,

$$\langle x_{n_k+1} - w_{n_k} + \lambda_{n_k}\mathcal{A}y_{n_k}, y - x_{n_k+1}\rangle \ge 0.$$

By monotonicity of $\mathcal{A}$, we have

$$\begin{aligned} 0 &\le \langle x_{n_k+1} - w_{n_k}, y - x_{n_k+1}\rangle + \lambda_{n_k}\langle \mathcal{A}y_{n_k}, y - x_{n_k+1}\rangle \\ &= \langle x_{n_k+1} - w_{n_k}, y - x_{n_k+1}\rangle + \lambda_{n_k}\langle \mathcal{A}y_{n_k}, y - y_{n_k}\rangle + \lambda_{n_k}\langle \mathcal{A}y_{n_k}, y_{n_k} - x_{n_k+1}\rangle \\ &\le \langle x_{n_k+1} - w_{n_k}, y - x_{n_k+1}\rangle + \lambda_{n_k}\langle \mathcal{A}y, y - y_{n_k}\rangle + \lambda_{n_k}\langle \mathcal{A}y_{n_k}, y_{n_k} - x_{n_k+1}\rangle. \end{aligned} \tag{51}$$

Passing to the limit as $k \to \infty$ in (51), we obtain (noting (48) and (49)),

$$\langle Ay, y - p\rangle \ge 0 \; \forall y \in \mathcal{C}.$$

By Lemma 2.3, we get that $p \in S$.

We now show that $x_n \rightharpoonup x^* \in S$. Let us assume that there exist $\{x_{n_k}\} \subset \{x_n\}$ and $\{x_{n_j}\} \subset \{x_n\}$ such that $x_{n_k} \rightharpoonup v^*, k \to \infty$ and $x_{n_j} \rightharpoonup x^*, j \to \infty$. We show that $v^* = x^*$.

Observe that

$$2\langle x_n, x^* - v^*\rangle = \|x_n - v^*\|^2 - \|x_n - x^*\|^2 - \|v^*\|^2 + \|x^*\|^2 \tag{52}$$

and

$$\begin{aligned} 2\langle -\theta x_{n-1}, x^* - v^*\rangle &= -\theta\|x_{n-1} - v^*\|^2 + \theta\|x_{n-1} - x^*\|^2 \\ &\quad + \theta\|v^*\|^2 - \theta\|x^*\|^2. \end{aligned} \tag{53}$$

Addition of (52) and (53) gives

$$2\langle x_n - \theta x_{n-1}, x^* - v^*\rangle = \Big(\|x_n - v^*\|^2 - \theta\|x_{n-1} - v^*\|^2\Big) \\ - \Big(\|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2\Big) \\ + (1-\theta)(\|x^*\|^2 - \|v^*\|^2). \quad (54)$$

According to (50), we have

$$\lim_{n\to\infty}\Big[\|x_n - v^*\|^2 - \theta\|x_{n-1} - v^*\|^2 + 2\lambda_{n-1}\langle \mathcal{A}v^*, x_{n-1} - v^*\rangle\Big]$$

exists and

$$\lim_{n\to\infty}\Big[\|x_n - x^*\|^2 - \theta\|x_{n-1} - x^*\|^2 + 2\lambda_{n-1}\langle \mathcal{A}x^*, x_{n-1} - x^*\rangle\Big]$$

exists. Hence, we obtain from (54) that

$$\lim_{n\to\infty}\Big[\langle x_n - \theta x_{n-1}, x^* - v^*\rangle + \lambda_{n-1}\Big(\langle \mathcal{A}v^*, x_{n-1} - v^*\rangle - \langle \mathcal{A}x^*, x_{n-1} - x^*\rangle\Big)\Big]$$

exists. Now, since $\lim_{n\to\infty}\lambda_{n-1} = \lambda$, we obtain

$$\begin{aligned}
&\langle v^* - \theta v^*, x^* - v^*\rangle + \lambda\langle \mathcal{A}x^*, x^* - v^*\rangle \\
&= \lim_{k\to\infty}\Big[\langle x_{n_k} - \theta x_{n_k-1}, x^* - v^*\rangle + \lambda_{n_k-1}\Big(\langle \mathcal{A}v^*, x_{n_k-1} - v^*\rangle \quad -\langle \mathcal{A}x^*, x_{n_k-1} - x^*\rangle\Big)\Big] \\
&= \lim_{n\to\infty}\Big[\langle x_n - \theta x_{n-1}, x^* - v^*\rangle + \lambda_{n-1}\Big(\langle \mathcal{A}v^*, x_{n-1} - v^*\rangle \quad -\langle \mathcal{A}x^*, x_{n-1} - x^*\rangle\Big)\Big] \\
&= \lim_{k\to\infty}\Big[\langle x_{n_j} - \theta x_{n_j-1}, x^* - v^*\rangle + \lambda_{n_j-1}\Big(\langle \mathcal{A}v^*, x_{n_j-1} - v^*\rangle \quad -\langle \mathcal{A}x^*, x_{n_j-1} - x^*\rangle\Big)\Big] \\
&= \langle x^* - \theta x^*, x^* - v^*\rangle + \lambda\langle \mathcal{A}v^*, x^* - v^*\rangle,
\end{aligned}$$

and this yields

$$(1-\theta)\|x^* - v^*\|^2 + \lambda\langle \mathcal{A}v^* - \mathcal{A}x^*, x^* - v^*\rangle = 0. \quad (55)$$

Since $x^* \in S$ and $v^* \in S$, we have that

$$\begin{aligned}
\langle \mathcal{A}v^* - \mathcal{A}x^*, x^* - v^*\rangle &= \langle \mathcal{A}v^*, x^* - v^*\rangle + \langle \mathcal{A}x^*, v^* - x^*\rangle \\
&\geq 0. \quad (56)
\end{aligned}$$

Since $\mathcal{A}$ is monotone, we have $\langle \mathcal{A}v^* - \mathcal{A}x^*, x^* - v^*\rangle \leq 0$. This together with (56) imply that

$$\langle \mathcal{A}v^* - \mathcal{A}x^*, x^* - v^*\rangle = 0. \quad (57)$$

Since $\theta \in [0, 1)$, we obtain from (55) and (57) that $x^* = v^*$. Hence, $\{x_n\}$ converges weakly to a point in S. This completes the proof. ■

5. Linear convergence result

In this section, we obtain linear convergence result of Algorithm 1 under the condition that $\mathcal{A}$ in VIP (1) is η-strongly monotone and L-Lipschitz continuous on $\mathbb{H}$. We know that under this condition on $\mathcal{A}$, VIP (1) has a unique solution.

We first prove the following lemma.

Lemma 5.1. *Suppose $\{a_n\}$ and $\{b_n\}$ are non-negative sequences such that*

$$a_{n+1} + b_{n+1} \le (1-\rho)a_n + r\rho a_{n-1} + rb_n + \theta(a_n - a_{n-1}),$$

where $\rho \in (0,\infty), r \in (0,1)$ and $\theta \in [0, \min\{r\rho, r\})$. Then $\{a_n\}$ converges linearly to 0.

Proof: For any $\delta_1, \delta_2 > 0$, we obtain

$$\begin{aligned} a_{n+1} + \delta_1 a_n + b_{n+1} &\le (1-\rho+\delta_1)a_n + r\rho a_{n-1} + rb_n \\ &\quad + \theta(a_n - a_{n-1}) \\ &= (1-\rho+\delta_1+\theta)a_n + (r\rho-\theta)a_{n-1} + rb_n \\ &= \delta_2(a_n + \delta_1 a_{n-1}) + rb_n + (1-\rho+\theta+\delta_1-\delta_2)a_n \\ &\quad + (r\rho - \theta - \delta_1\delta_2)a_{n-1} \\ &\le \alpha(a_n + \delta_1 a_{n-1} + b_n) + (1-\rho+\theta+\delta_1-\delta_2)a_n \\ &\quad + (r\rho-\theta-\delta_1\delta_2)a_{n-1}, \end{aligned} \tag{58}$$

where $\alpha := \max\{\delta_2, r\}$. Let us choose $\delta_1, \delta_2 > 0$ such that

$$\begin{cases} 1-\rho+\theta+\delta_1-\delta_2 = 0 \\ r\rho - \theta - \delta_1\delta_2 = 0. \end{cases}$$

It suffices to choose $\delta_2 > 0$ such that

$$\delta_2^2 - (1-\rho+\theta)\delta_2 - (r\rho-\theta) = 0.$$

Solving this quadratic equation gives

$$\delta_2 = \frac{(1-\rho+\theta) + \sqrt{((\rho-\theta)-1)^2 + 4(r\rho-\theta)}}{2},$$

where $\rho \in (0,\infty), r \in (0,1)$ and $\theta \in [0,\min\{r\rho, r\})$. Then it is easy to see that $\delta_2 \in (0,1)$ since $\theta < \min\{r\rho, r\}$. Therefore, $\alpha := \max\{\delta_2, r\} < 1$. Furthermore, we obtain from (58) that

$$a_{n+1} + \delta_1 a_n + b_{n+1} \le \alpha(a_n + \delta_1 a_{n-1} + b_n) \quad \le \alpha^{n+1}(a_0 + \delta_1 a_{-1} + b_0) = \alpha^{n+1}M,$$

for some $M > 0$. ∎

Remark 5.2: Lemma 5.1 reduces to [17, Lemma 4] when $\theta = 0$.

We now give the linear convergence result.

Theorem 5.3. *Suppose $\mathcal{A}$ is η-strongly monotone and L-Lipschitz continuous on $\mathbb{H}$. Assume that $0 \leq \theta < \min\{2\eta\lambda_0, \frac{4\sqrt{2}\eta\mu}{L}, \frac{1}{7}\}$. Then $\{x_n\}$ generated by Algorithm 1 converges to a unique solution of the VIP (1) at least R-linearly.*

Proof: Let z be the unique element of S. By Lemma 2.2 (ii) (taking $\kappa = 2$), we have

$$\begin{aligned} \|y_n - z\|^2 &= \|2(x_n - z) - (x_{n-1} - z)\|^2 \\ &= 2\|x_n - z\|^2 - \|x_{n-1} - z\|^2 + 2\|x_n - x_{n-1}\| \\ &\geq 2\|x_n - z\|^2 - \|x_{n-1} - z\|^2. \end{aligned} \tag{59}$$

By the strong monotonicity of $\mathcal{A}$ and (59), we have

$$\begin{aligned} &2\lambda_n \left[\langle Ay_n - Az, y_n - z\rangle - \eta\left(2\|x_n - z\|^2 - \|x_{n-1} - z\|^2\right)\right] \\ &\geq 2\lambda_n \left[\langle Ay_n - Az, y_n - z\rangle - \eta\|y_n - z\|^2\right] \\ &\geq 0. \end{aligned} \tag{60}$$

By (30), we obtain

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq\ & \|w_n - z\|^2 - \|x_{n+1} - w_n\|^2 + 2\lambda_n\langle \mathcal{A}y_{n-1} - \mathcal{A}y_n, x_{n+1} - y_n\rangle \\ & - 2\lambda_n\langle \mathcal{A}y_n, y_n - z\rangle + \frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\|x_{n+1} - x_n\|^2 + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2 \\ & + \theta\|x_n - y_{n-1}\|^2 + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2. \end{aligned} \tag{61}$$

Adding (60) to the right hand side of (61), we get

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq\ & \|w_n - z\|^2 - \|x_{n+1} - w_n\|^2 - 4\lambda_n\eta\|x_n - z\|^2 \\ & + 2\lambda_n\eta\|x_{n-1} - z\|^2 + 2\lambda_n\langle \mathcal{A}y_{n-1} - \mathcal{A}y_n, x_{n+1} - y_n\rangle \\ & - 2\lambda_n\langle \mathcal{A}z, y_n - z\rangle + \frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\|x_{n+1} - x_n\|^2 + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2 \\ & + \theta\|x_n - y_{n-1}\|^2 + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2. \end{aligned} \tag{62}$$

Applying (36) and (37) in (62), we get

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq\ & (1 + \theta - 4\eta\lambda_n)\|x_n - z\|^2 + (2\eta\lambda_n - \theta)\|x_{n-1} - z\|^2 + 2\theta\|x_n - x_{n-1}\|^2 \\ & - (1-\theta)\|x_{n+1} - x_n\|^2 + 2\lambda_n\langle \mathcal{A}y_{n-1} - \mathcal{A}y_n, x_{n+1} - y_n\rangle \\ & - 2\lambda_n\langle \mathcal{A}z, y_n - z\rangle + \frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\|x_{n+1} - x_n\|^2 + \left(\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n - y_n\|^2 \\ & + \theta\|x_n - y_{n-1}\|^2 + \left(2\theta - \frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1} - y_n\|^2. \end{aligned} \tag{63}$$

Putting (31) and (34) into (63), we have

$$\begin{aligned}
\|x_{n+1}-z\|^2 \le\ & (1+\theta-4\eta\lambda_n)\|x_n-z\|^2+(2\eta\lambda_n-\theta)\|x_{n-1}-z\|^2+2\theta\|x_n-x_{n-1}\|^2\\
&-(1-\theta)\|x_{n+1}-x_n\|^2+\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}\|x_n-y_{n-1}\|^2\\
&+\frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}\|x_n-y_n\|^2\\
&+\frac{4\mu\lambda_n}{\lambda_{n+1}}\|x_{n+1}-y_n\|^2-4\lambda_n\langle \mathcal{A}z,x_n-z\rangle+2\lambda_n\langle \mathcal{A}z,x_{n-1}-z\rangle\\
&+\frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\|x_{n+1}-x_n\|^2+\left(\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-y_n\|^2\\
&+\theta\|x_n-y_{n-1}\|^2+\left(2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2\\
=\ & (1+\theta-4\eta\lambda_n)\|x_n-z\|^2+(2\eta\lambda_n-\theta)\|x_{n-1}-z\|^2+2\theta\|x_n-x_{n-1}\|^2\\
&+\left(\frac{(1-\theta)\lambda_n}{\lambda_{n-1}}+\theta-1\right)\|x_{n+1}-x_n\|^2\\
&+\left(\frac{4\mu\lambda_n}{\lambda_{n+1}}+2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2\\
&+\left(\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}+\theta\right)\|x_n-y_{n-1}\|^2-4\lambda_n\langle \mathcal{A}z,x_n-z\rangle\\
&+2\lambda_n\langle \mathcal{A}z,x_{n-1}-z\rangle. \qquad (64)
\end{aligned}$$

Recall that

$$y_n-x_n=x_n-x_{n-1}. \qquad (65)$$

Using (65) in (64) gives

$$\begin{aligned}
\|x_{n+1}-z\|^2 \le\ & (1+\theta-4\eta\lambda_n)\|x_n-z\|^2+(2\eta\lambda_n-\theta)\|x_{n-1}-z\|^2\\
&+\left(\frac{(1-\theta)\lambda_n}{\lambda_{n-1}}+\theta-1\right)\|x_{n+1}-x_n\|^2\\
&+\left(2\theta+\frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}+\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_n-x_{n-1}\|^2\\
&+\left(\frac{4\mu\lambda_n}{\lambda_{n+1}}+2\theta-\frac{\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-y_n\|^2+\left(\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}+\theta\right)\|x_n-y_{n-1}\|^2\\
&-4\lambda_n\langle \mathcal{A}z,x_n-z\rangle+2\lambda_n\langle \mathcal{A}z,x_{n-1}-z\rangle. \qquad (66)
\end{aligned}$$

Thus, (66) becomes

$$\begin{aligned}\|x_{n+1}-z\|^2 \le\ & (1+\theta-4\eta\lambda_n)\|x_n-z\|^2+(2\eta\lambda_n-\theta)\|x_{n-1}-z\|^2\\ & -\left(1-\theta-\frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-x_n\|^2+\left(\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}+\theta\right)\|x_n-y_{n-1}\|^2\\ & -\left(\frac{\lambda_n}{\lambda_{n-1}}-\frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}-3\theta\right)\|x_n-y_n\|^2\\ & -\left(\frac{\lambda_n}{\lambda_{n-1}}-\frac{4\mu\lambda_n}{\lambda_{n+1}}-2\theta\right)\|x_{n+1}-y_n\|^2\\ & -4\lambda_n\langle \mathcal{A}z,x_n-z\rangle+2\lambda_n\langle \mathcal{A}z,x_{n-1}-z\rangle. \end{aligned} \tag{67}$$

Therefore,

$$\begin{aligned}& \|x_{n+1}-z\|^2+4\lambda_n\langle \mathcal{A}z,x_n-z\rangle+\left(\frac{\lambda_n}{\lambda_{n-1}}-\frac{4\mu\lambda_n}{\lambda_{n+1}}-2\theta\right)\|x_{n+1}-y_n\|^2\\ & \le (1+\theta-4\eta\lambda_n)\|x_n-z\|^2+(2\eta\lambda_n-\theta)\|x_{n-1}-z\|^2\\ & \quad -\left(1-\theta-\frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\right)\|x_{n+1}-x_n\|^2+\left(\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}+\theta\right)\|x_n-y_{n-1}\|^2\\ & \quad -\left(\frac{\lambda_n}{\lambda_{n-1}}-\frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}-3\theta\right)\|x_n-y_n\|^2+2\lambda_n\langle \mathcal{A}z,x_{n-1}-z\rangle. \end{aligned} \tag{68}$$

Observe that

$$\begin{aligned}\left(1-\theta-\frac{(1-\theta)\lambda_n}{\lambda_{n-1}}\right) &= (1-\theta)\left(1-\frac{\lambda_n}{\lambda_{n-1}}\right)\\ &= (1-\theta)\left(\frac{\lambda_{n-1}-\lambda_n}{\lambda_n}\right)\\ &\ge 0. \end{aligned} \tag{69}$$

Define $\beta_n := \frac{(4+2\sqrt{2})\mu\lambda_n}{\lambda_{n+1}}+3\theta$. Then

$$\lim_{n\to\infty}\left(\frac{\lambda_n}{\lambda_{n-1}}-\beta_n\right)=1-(4+2\sqrt{2})\mu-3\theta\ge 0, \tag{70}$$

since $\mu<\frac{1-7\theta}{4+2\sqrt{2}}=\bar{\mu}$. Therefore, from (68), there exists $n_0\in\mathbb{N}$ such that $\forall n\ge n_0$, we have

$$\begin{aligned}& \|x_{n+1}-z\|^2+4\lambda_n\langle \mathcal{A}z,x_n-z\rangle+\left(\frac{\lambda_n}{\lambda_{n-1}}-\frac{4\mu\lambda_n}{\lambda_{n+1}}-2\theta\right)\|x_{n+1}-y_n\|^2\\ & \le (1+\theta-4\eta\lambda_n)\|x_n-z\|^2+(2\eta\lambda_n-\theta)\|x_{n-1}-z\|^2\\ & \quad +\left(\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}+\theta\right)\|x_n-y_{n-1}\|^2+2\lambda_n\langle \mathcal{A}z,x_{n-1}-z\rangle. \end{aligned} \tag{71}$$

Then $\forall n \geq n_0$, we get

$$\begin{aligned}\|x_{n+1}-z\|^2 &+ 4\lambda_n\langle \mathcal{A}z, x_n - z\rangle + \left(\frac{\lambda_n}{\lambda_{n-1}} - \frac{4\mu\lambda_n}{\lambda_{n+1}} - 2\theta\right)\|x_{n+1}-y_n\|^2 \\ &\leq (1+\theta-4\eta\lambda_n)\|x_n - z\|^2 + (2\eta\lambda_n - \theta)\|x_{n-1}-z\|^2 \\ &\quad + \left(\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}} + \theta\right)\|x_n - y_{n-1}\|^2 + 2\lambda_n\langle \mathcal{A}z, x_{n-1} - z\rangle. \end{aligned} \tag{72}$$

Define $\xi_n = \max\{\frac{\frac{2\sqrt{2}\mu\lambda_n}{\lambda_{n+1}}+\theta}{\frac{\lambda_{n-1}}{\lambda_{n-2}} - \frac{4\mu\lambda_{n-1}}{\lambda_n} - 2\theta}, \frac{1}{2}\frac{\lambda_n}{\lambda_{n-1}}\}$. Let $\xi = \sup \xi_n$. Then it is easy to see that $\xi \in (0,1)$ since $\theta < \frac{1-(4+2\sqrt{2})\mu}{3}$, which is automatically satisfied because $\mu < \frac{1-7\theta}{4+2\sqrt{2}}$.

Now, $\forall n \geq n_1$, let $a_n := \|x_n - z\|^2$ and $b_{n+1} := (\frac{\lambda_n}{\lambda_{n-1}} - \frac{4\mu\lambda_n}{\lambda_{n+1}} - 2\theta)\|x_{n+1} - y_n\|^2 + 4\lambda_n\langle \mathcal{A}z, x_n - z\rangle$. Then we have from (72) that $\forall n \geq n_1$,

$$\begin{aligned} a_{n+1} + b_{n+1} &\leq (1+\theta - 4\eta\lambda_n)\, a_n + (2\eta\lambda_n - \theta)a_{n-1} + \xi_n b_n \\ &\leq (1-4\eta\lambda_n)a_n + 2\eta\lambda_n a_{n-1} + \xi b_n + \theta(a_n - a_{n-1}) \\ &= (1-4\eta\lambda_n)a_n + \frac{1}{2}(4\eta\lambda_n)a_{n-1} + \xi b_n + \theta(a_n - a_{n-1}). \end{aligned} \tag{73}$$

Define $r := \max\{\xi, \frac{1}{2}\} \in (0,1)$. Since $4\eta\lambda_n \geq 4\eta\lambda =: \rho > 0$, we have from (73) that

$$a_{n+1} + b_{n+1} \leq (1-\rho)a_n + r\rho a_{n-1} + r b_n + \theta(a_n - a_{n-1}). \tag{74}$$

Observe that $\frac{1}{2} \leq r, \min\{\lambda_0, \frac{2\sqrt{2}\mu}{L}\} \leq \lambda$ and this implies $\rho = 4\eta\lambda \geq \min\{4\eta\lambda_0, \frac{8\sqrt{2}\eta\mu}{L}\}$ by Remark 3.2. This means $\min\{1, 4\eta\lambda_0, \frac{8\sqrt{2}\mu}{L}\} \leq \min\{1,\rho\}$. Therefore,

$$\begin{aligned} \min\left\{4\eta\lambda_0, \frac{8\sqrt{2}\eta\mu}{L}, \frac{2}{7}\right\} &\leq \min\left\{4\eta\lambda_0, \frac{8\sqrt{2}\eta\mu}{L}, 1\right\} \\ &\leq \min\{1,\rho\}. \end{aligned}$$

Thus,

$$\begin{aligned} \min\left\{2\eta\lambda_0, \frac{4\sqrt{2}\eta\mu}{L}, \frac{1}{7}\right\} &= \frac{1}{2}\min\left\{4\eta\lambda_0, \frac{8\sqrt{2}\eta\mu}{L}, \frac{2}{7}\right\} \\ &\leq r\min\{1,\rho\} = \min\{r\rho, r\}. \end{aligned} \tag{75}$$

Noting (75) and invoking Lemma 5.1 in (74), we obtain the desired conclusion. ■

Remark 5.4: If $\theta = 0$ in Theorem 5.3, we do not need the condition $0 \leq \theta < \min\{2\eta\lambda_0, \frac{4\sqrt{2}\eta\mu}{L}, \frac{1}{7}\}$.

6. Numerical applications to optimal control

In this section, we provide computational experiments comparing Algorithm 1 to the existing state-of-the-art methods in the literature; [9, Algorithm 1], [16, Algorithm 4.1], [17, Algorithm 1] and [25, Algorithm A] using test examples from optimal control. All the computations are performed using Matlab 2016 (b) which is running on a personal computer with an Intel(R) Core(TM) i5-2600 CPU at 2.30GHz and 8.00 Gb-RAM.

Let $0 < T \in \mathbb{R}$. We denote by $L_2([0,T],\mathbb{R}^m)$ the Hilbert space of square integrable, measurable vector functions $u : [0,T] \to \mathbb{R}^m$ with inner product $\langle u, v\rangle = \int_0^T \langle u(t), v(t)\rangle dt$, and norm $\|u\|_2 = \sqrt{\langle u, u\rangle}$.

Consider the optimal control problem:

$$u^*(t) = \operatorname{argmin}\{f(u) : u \in U\} \tag{76}$$

on the interval $[0,T]$, assuming that such a control exists. Here U is the set of admissible controls, which has the form of an m-dimensional box and consists of continuous function:

$$U = \left\{u(t) \in L_2([0,T],\mathbb{R}^m) : u_i(t) \in [u_i^-, u_i^+], i = 1,2,\dots,m\right\}.$$

Specially, the control can be bang-bang (piecewise constant function). The terminal objective has the form

$$f(u) = \phi(x(T)),$$

where ϕ is a convex and differentiable function, defined on the attainability set.

Suppose that the trajectory $x(t) \in L_2([0,T]$ satisfies constraints in the form of a system of linear differential equation:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(0) = x_0, \quad t \in [0,T],$$

where $A(t) \in \mathbb{R}^{n\times n}$, $B(t) \in \mathbb{R}^{n\times m}$ are given continuous matrices for every $t \in [0,T]$. By the Pontryagin Maximum Principle there exists a function $p^* \in L_2([0,T]$ such that the triple (x^*, p^*, u^*) solves for a.e. $t \in [0,T]$ the system

$$\begin{cases} \dot{x}^*(t) = A(t)x^*(t) + B(t)u^*(t) \\ x^*(0) = x_0, \end{cases} \tag{77}$$

$$\begin{cases} \dot{p}^*(t) = -A(t)^\top p^*(t) \\ p^*(T) = \nabla\phi(x(T)), \end{cases} \tag{78}$$

$$0 \in B(t)^\top p^*(t) + N_U(u^*(t)), \tag{79}$$

where $N_U(u)$ is the normal cone to U at u. Denoting $Gu(t) := B(t)^\top p(t)$, it is known that G is the gradient of the objective cost function f, i.e. $G = \nabla f$. We can write (79) as VIP (1)

$$\langle Gu^*, v - u^*\rangle \geq 0, \quad \forall v \in U. \tag{80}$$

Now, let us discretize the continuous functions and choose a natural number N with the *mesh size* $h := T/N$. We identify any discretized control $u^N := (u_0, u_1, \dots, u_{N-1})$ with its piece-wise constant extension:

$$u^N(t) = u_i \text{ for } t \in [t_i, t_{i+1}), \quad i = 0, 1, \ldots, N.$$

Moreover, we identify any discretized state $x^N := (x_0, x_1, \ldots, x_N)$ with its piece-wise linear interpolation:

$$x^N(t) = x_i + \frac{t - t_i}{h}(x_{i+1} - x_i), \text{ for } t \in [t_i, t_{i+1}), \ i = 0, 1, \ldots, N-1.$$

Similarly for the co-state variable $p^N := (p_0, p_1, \ldots, p_N)$.

Example 6.1 (Control of a harmonic oscillator [19, Example 7]):

$$\begin{array}{ll} \text{minimize} & x_2(3\pi) \\ \text{subject to} & \dot{x_1}(t) = x_2(t), \\ & \dot{x_2}(t) = -x_1(t) + u(t), \ \forall t \in [0, 3\pi], \\ & x(0) = 0, \\ & u(t) \in [-1, 1]. \end{array} \tag{81}$$

The exact optimal control for this problem is:

$$u^*(t) = \begin{cases} 1 & \text{if } t \in [0, \pi/2) \cup (3\pi/2, 5\pi/2) \\ -1 & \text{if } t \in (\pi/2, 3\pi/2) \cup (5\pi/2, 3\pi]. \end{cases}$$

Let us now consider examples in which the terminal function is not linear.

Example 6.2 (Rocket Car [1,21]):

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\left((x_1(5))^2 + (x_2(5))^2\right) \\ \text{subject to} & \dot{x_1}(t) = x_2(t), \\ & \dot{x_2}(t) = u(t), \ \forall t \in [0, 5], \\ & x_1(0) = 6, \ x_2(0) = 1, \\ & u(t) \in [-1, 1]. \end{array} \tag{82}$$

The exact optimal control is

$$u^* = \begin{cases} 1 & \text{if } t \in (\tau, 5] \\ -1 & \text{if } t \in (0, \tau], \end{cases}$$

where $\tau = 3.5174292$.

Example 6.3 (See [4, Example 6.3]):

$$\begin{array}{ll} \text{minimize} & -x_1(2) + (x_2(2))^2 \\ \text{subject to} & \dot{x_1}(t) = x_2(t), \\ & \dot{x_2}(t) = u(t), \ \forall t \in [0, 2], \\ & x_1(0) = 0, \ x_2(0) = 0, \\ & u(t) \in [-1, 1]. \end{array} \tag{83}$$

Table 1. Comparison of algorithms for Examples 6.1–6.3.

	Example 6.1		Example 6.2		Example 6.3	
Algorithms	CPU	Iter	CPU	Iter	CPU	Iter
Algorithm 1	4.3490	836	3.1440	1531	1.0673	324
Algorithm 1 [9]	13.3210	8642	14.6563	7317	4.1292	1209
Algorithm 4.1 [16]	14.3387	9317	16.1188	7743	5.0042	1240
Algorithm 1 [17]	11.6642	7454	5.3318	2714	1.9882	594
Algorithm A [25]	13.4731	8432	15.2751	7745	4.5184	1294

Table 2. Comparison of algorithms for Example 6.4.

Alg.1		Alg.1 [9]		Alg.4.1 [16]		Alg.1 [17]		Alg.A [25]		
*lm*CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter	
1010	10.5035	451	31.5812	1480	70.5178	9990	58.1122	7269	33.0061	2423
20	10.5385	484	20.9037	832	64.2881	4095	31.9602	1780	21.0148	846
30	20.5721	485	31.0762	939	71.4619	10008	44.6089	3956	33.1827	2349
2010	20.6319	561	31.0805	961	67.7763	7288	55.5648	5098	43.7503	2840
20	20.7575	677	30.9745	906	71.8235	11139	55.8291	5209	43.4602	2787
30	21.2526	1061	31.5295	1395	77.2462	15582	57.1335	6396	32.6229	1843
3010	21.3093	1082	31.3812	1149	64.0363	3581	33.1262	1997	33.5213	2452
20	21.3274	1198	31.4911	1386	68.4674	7849	58.3597	7284	44.2001	3245

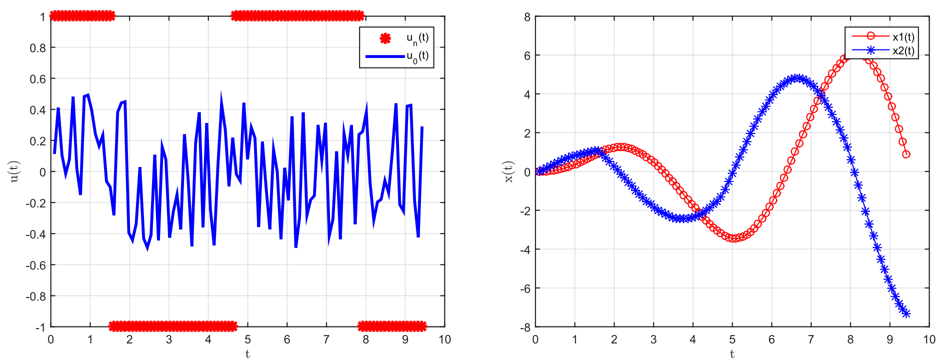

Figure 1. Example 6.1 computed by Algorithm 1: Left: random initial control (middle) and optimal control (edge); Right: optimal trajectories.

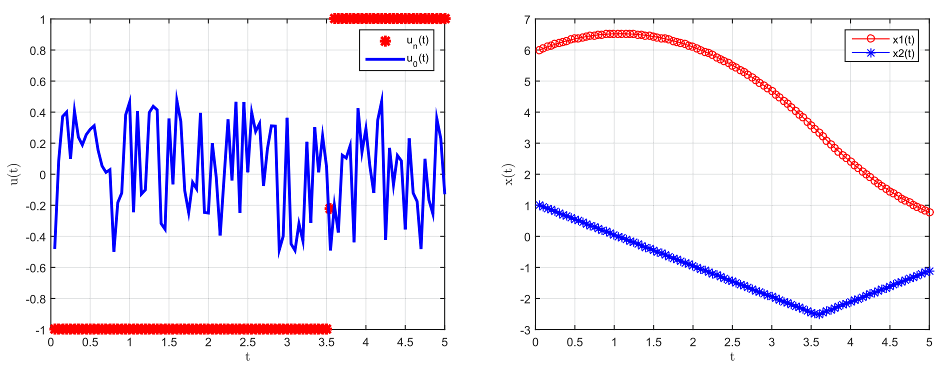

Figure 2. Example 6.2 computed by Algorithm 1: Left: random initial control (middle) and optimal control (edge); Right: optimal trajectories.

The exact optimal control is

$$u^* = \begin{cases} 1 & \text{if } t \in [0, 6/5) \\ -1 & \text{if } t \in (6/5, 2]. \end{cases}$$

In Examples 6.1–6.3, we set $N = 100$ and define

$\mathrm{TOL}_n := \max\{\|x_{n+1} - x_n\|^2, \ \|x_n - x_{n-1}\|^2\}$ for all the algorithms under comparison, with stopping criterion $\mathrm{TOL}_n \le 10^{-4}$. Note that $\mathrm{TOL}_n = 0$ implies that x_n is a solution

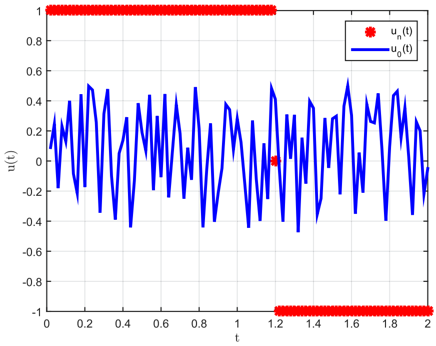

Figure 3. Example 6.3 computed by Algorithm 1: Left: random initial control (middle) and optimal control (edge); Right: optimal trajectories.

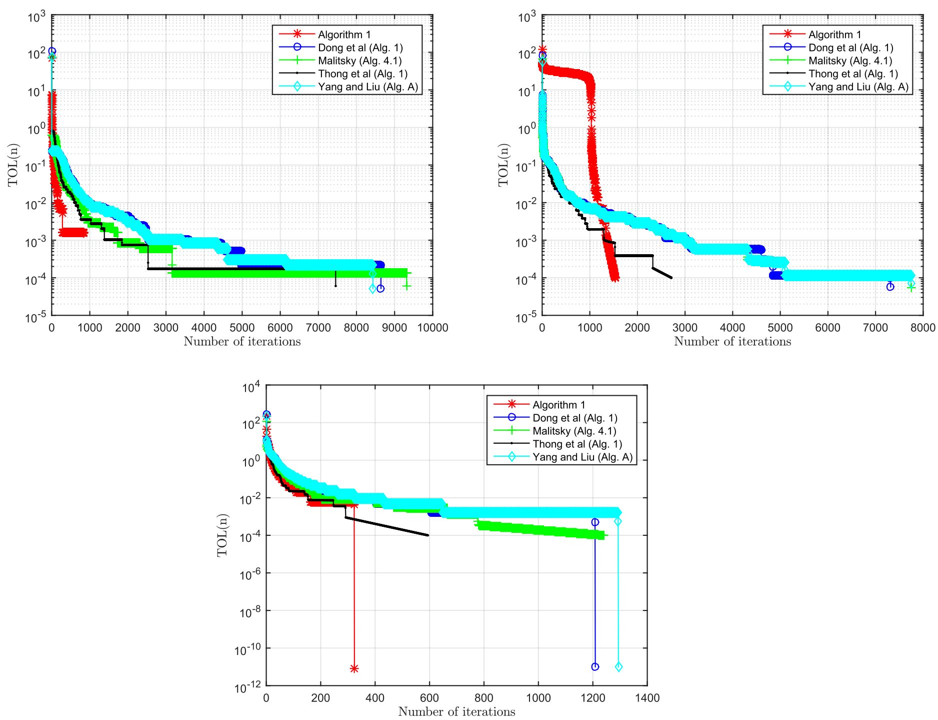

Figure 4. The behaviour of TOL_n: Top Left: Example 6.1; Top Right: Example 6.2; Bottom: Example 6.3.

of Problem (1). Also, we randomly choose the initial control $u_0(t)$ in $[-1, 1]$ for all the algorithms under comparison.

Our final example is commonly used as test problem for variational inequality in the literature ([10]).

Example 6.4: Let the operator $A(x) := Mx + q,\ x \in \mathbb{R}^m$, where

$$M = BB^T + C + D,$$

and B is an $m \times m$ matrix, C is an $m \times m$ skew-symmetric matrix, D is an $m \times m$ diagonal matrix whose diagonal entries are nonnegative (so M is positive semidefinite), q is a vector in $\mathbb{R}^m$. The feasible set $\mathcal{C} \subset \mathbb{R}^m$ is a closed and convex subset defined by $\mathcal{C} := \{x \in \mathbb{R}^m : Qx \leq b\}$, where Q is an $l \times m$ matrix and b is a nonnegative vector. It is clear that A is monotone and L-Lipschitz-continuous with $L = \|M\|$. Let $q = 0$. Then, the solution set $S = \{0\}$.

For the numerical computation, we consider different problem sizes $l = 10, 20, 30$ and $m = 10, 20, 30$. We also randomly generate the matrices B, C, D, Q, b, and the starting points for all the algorithms under comparison. Furthermore, we define $\text{TOL}_n := \|x_n\|$ for all the algorithms, with stopping criterion $\text{TOL}_n \leq 10^{-3}$.

In Examples 6.1–6.4, we choose the parameters as follows:

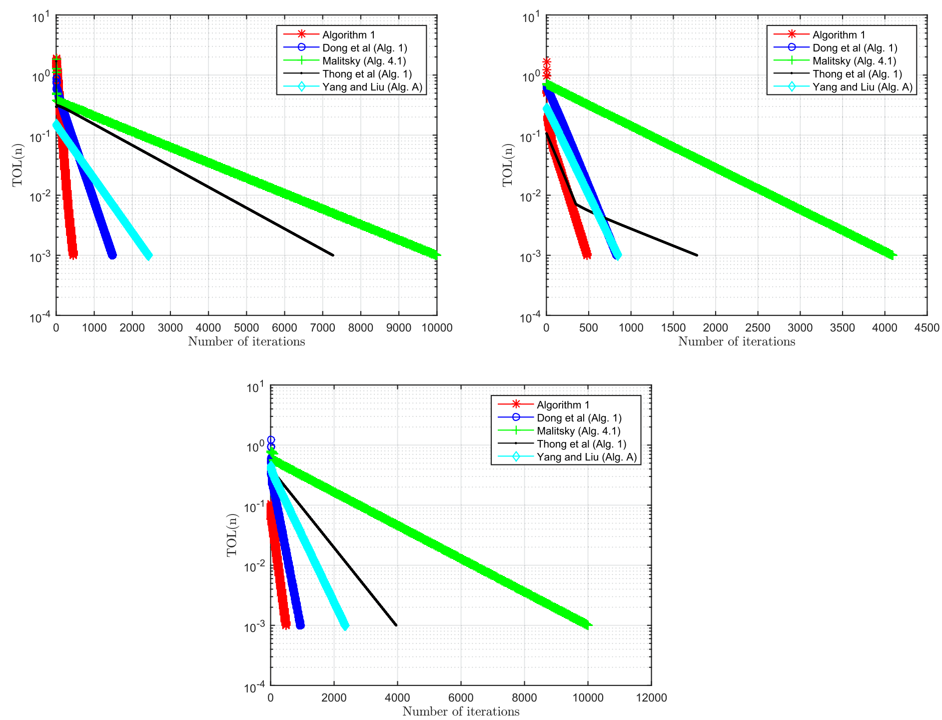

Figure 5. The behaviour of TOL_n for Example 6.4 with $l = 10$: Top Left: $\boldsymbol{m = 10}$; Top Right: $\boldsymbol{m = 20}$; Bottom: $\boldsymbol{m = 30}$.

- Algorithm 1: $\lambda_0 = 1$, $\theta = \frac{0.99}{7}$ and $\mu = 0.99\bar{\mu}$.
- Algorithm 1 [9]: $\lambda_0 = 1, \delta = 1.6, \theta = \frac{0.99\delta(\delta-1)}{3\delta^2-1}$ and $\alpha = \frac{0.99(\sqrt{2}-1)(\delta-\theta-2\theta\delta)}{\delta^2}$.
- Algorithm 4.1 [16]: $\lambda_0 = 1$ and $\alpha = 0.4$.
- Algorithm 1 [17]: $\lambda_0 = 1, \theta = 1$ and $\mu = 0.99\bar{\mu}$.
- Algorithm A [25]: $\lambda_0 = 1$, $\delta = 1.01$ and $\alpha = \frac{0.99(\sqrt{2}-1)}{\delta}$.

Note that for all the algorithms, we choose the same λ_0 as above for the sake of uniformity, while the choices of other parameters are inspired by the choices made in the papers.

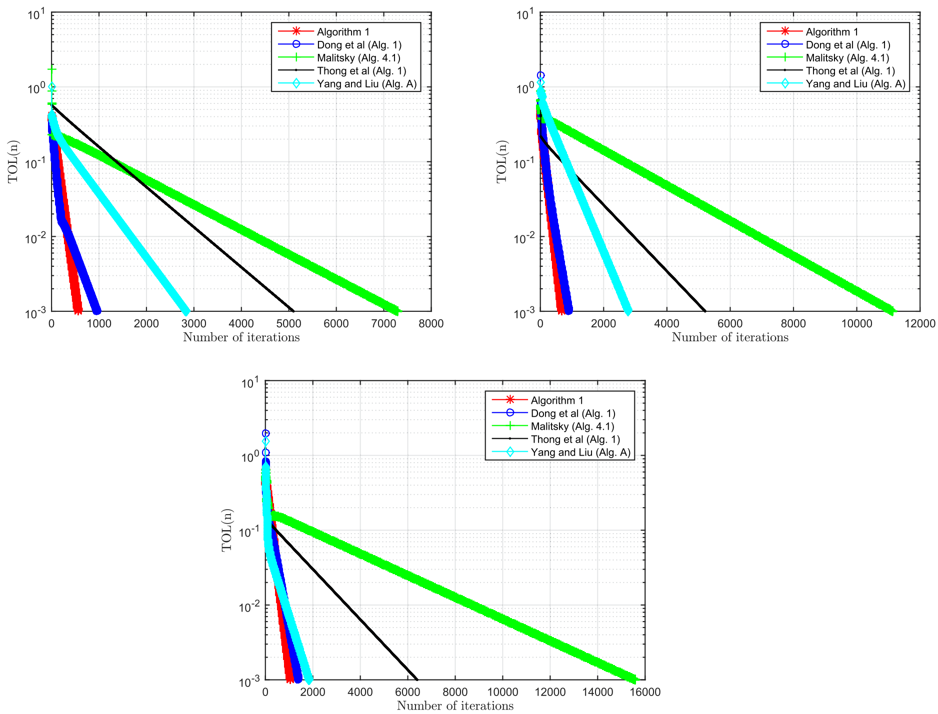

Figure 6. The behaviour of TOL_n for Example 6.4 with $l = 20$: Top Left: $\boldsymbol{m = 10}$; Top Right: $\boldsymbol{m = 20}$; Bottom: $\boldsymbol{m = 30}$.

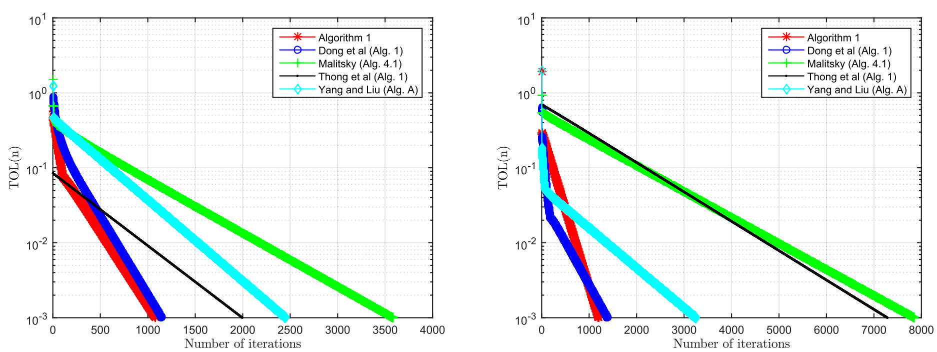

Figure 7. The behaviour of TOL_n for Example 6.4 with $l = 30$: Top Left: $\boldsymbol{m = 10}$; Top Right: $\boldsymbol{m = 20}$.

In Tables 1–2, "Iter" and "CPU" mean the CPU time in seconds and the number of iterations, respectively. Furthermore, in Table 2 and Figures 4–7, "Alg." means Algorithm.

The numerical results for Examples 6.1–6.3 are given in Table 1 and Figures 1–4 while those of Example 6.4 are given in Table 2 and Figures 5–7.

The results show that Algorithm 1 performs better than [9, Algorithm 1], [16, Algorithm 4.1], [17, Algorithm 1] and [25, Algorithm A] in terms of CPU time and number of iterations.

7. Final remarks

This paper presents weak and linear convergence results of an inertial projected reflected gradient method for monotone variational inequality problems in real Hilbert spaces. The method requires, at each iteration, one projection onto the feasible set and one evaluation of the cost operator of the variational inequality. Numerical examples drawn from optimal control are given to illustrate the theoretical analysis and comparisons with other state-of-the-art algorithms. Part of our future research concentrates on the development of suitable modifications of the inertial projected reflected gradient method where the choice of inertial factor $\theta = 1$ is possible which might improve the convergence speed of the proposed method of this paper.

Acknowledgments

We are very thankful to the referee and Senior Editor for the valuable comments and suggestions that helped to improve our manuscript.

Declarations

Code availability

The Matlab codes employed to run the numerical experiments are available upon request to the Authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Chinedu Izuchukwu is a Lecturer at the School of Mathematics, University of the Witwatersrand, South Africa. He obtained his doctoral degree from the University of KwaZulu-Natal, South Africa. He was a postdoctoral fellow at the Technion-Israel Institute of Technology, Haifa, Israel. His research interests lie in the areas of continuous and discrete optimization.

Yekini Shehu is a Professor of Mathematics at the School of Mathematical Sciences, Zhejiang Normal University, Jinhua, China. His research include Optimization & Variational Problems, Bilevel Optimization Problems, Quasi-variational Inequalities, Nonlinear Integral Equations, Monotone Inclusion Problems and Variational Inequalities. Professor Shehu is a recipient of the prestigious Alexander von Humboldt Postdoctoral Research Fellowship awarded by the Alexander von Humboldt Foundation, Bonn, Germany, from 01 May 2016 to 30 April 2019. He was also a postdoctoral

fellow at the Institute of Science and Technology (IST), Vienna, Austria from 01 June 2019 to 30 September 2021.

References

[1] W. Alt, R. Baier, M. Gerdts, and F. Lempio, *Error bounds for euler approximation of linear-quadratic control problems with bang-bang solutions*, Numer. Algebra Control Optim. 2(3) (2012), pp. 547–570.

[2] C. Baiocchi and A. Capelo, *Variational and Quasivariational Inequalities; Applications to Free Boundary Problems*, Wiley, New York, 1984.

[3] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011.

[4] A. Bressan and B. Piccoli, *Introduction to the Mathematical Theory of Control*, AIMS Series on Applied Mathematics, Springfield, 2007.

[5] Y. Censor, A. Gibali, and S. Reich, *The subgradient extragradient method for solving variational inequalities in Hilbert space*, J. Optim. Theory Appl. 148(2) (2011), pp. 318–335.

[6] Y. Censor, A. Gibali, and S. Reich, *Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space*, Optim. Meth Softw. 26(4-5) (2011), pp. 827–845.

[7] Y. Censor, A. Gibali, and S. Reich, *Extensions of Korpelevich's extragradient method for the variational inequality problem in Euclidean space*, Optimization 61(9) (2012), pp. 1119–1132.

[8] X. Chang, S. Liu, Z. Deng, and S. Li, *An inertial subgradient extragradient algorithm with adaptive stepsizes for variational inequality problems*, Optim. Meth. Soft. 37(4) (2022), pp. 1507–1526.

[9] Q.L. Dong, S. He, and L. Liu, *A general inertial projected gradient method for variational inequality problems*, Comp. Appl. Math. 40(5) (2021), pp. 168.

[10] P.T. Harker and J.-S. Pang, *A damped-Newton method for the linear complementarity problem*, in Computational Solution of Nonlinear Systems of Equations, Lectures in Appl. Math. 26, G. Allgower and K. Georg, eds., AMS, Providence, RI, 1990, pp. 265–284.

[11] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York, 1980.

[12] G.M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Ekonomika I Mat. Metody 12 (1976), pp. 747–756.

[13] P.-E. Maingé, *Numerical approach to monotone variational inequalities by a one-step projected reflected gradient method with line-search procedure*, Comp. Math. Appl. 72(3) (2016), pp. 720–728.

[14] P.-E. Maingé and M.L. Gobinddass, *Convergence of one-step projected gradient methods for variational inequalities*, J. Optim. Theory Appl. 171(1) (2016), pp. 146–168.

[15] Y. Malitsky, *Golden ratio algorithms for variational inequalities*, Math. Program. 184(1-2) (2020), pp. 383–410.

[16] Y.V. Malitsky, *Projected reflected gradient methods for monotone variational inequalities*, SIAM J. Optim. 25(1) (2015), pp. 502–520.

[17] D.V. Thong, A. Gibali, and P.T. Vuong, *An explicit algorithm for solving monotone variational inequalities*, Appl. Numer. Math. 171 (2022), pp. 408–425.

[18] A. Nagurney, *Network Economics: A Variational Inequality Approach*, Kluwer Academic Publishers, Dordrecht, 1999.

[19] A. Pietrus, T. Scarinci, and V.M. Veliov, *High order discrete approximations to Mayer's problems for linear systems*, SIAM J. Control Optim. 56(1) (2018), pp. 102–119.

[20] L.D. Popov, *A modification of the Arrow-Hurwicz method for finding saddle points*, Math. Notes 28(5) (1980), pp. 845–848.

[21] J. Preininger and P.T. Vuong, *On the convergence of the gradient projection method for convex optimal control problems with bang-bang solutions*, Comput. Optim. Appl. 70(1) (2018), pp. 221–238.

[22] W. Takahashi, *Nonlinear Functional Analysis*, Yokohama Publishers, Yokohama, 2000.

[23] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim. 38(2) (2000), pp. 431–446.
[24] D.V. Thong, X.H. Li, Q.L. Dong, Y.J. Cho, and T.M. Rassias, *An inertial Popov's method for solving pseudomonotone variational inequalities*, Optim. Lett. 15(2) (2021), pp. 757–777.
[25] J. Yang and H. Liu, *A modified projected gradient method for monotone variational inequalities*, J. Optim. Theory Appl. 179(1) (2018), pp. 197–211.