Commun Nonlinear Sci Numer Simulat 138 (2024) 108217

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns

Research paper

A Golden Ratio Algorithm With Backward Inertial Step For Variational Inequalities

Chinedu Izuchukwu [a], Yekini Shehu [b,*]

[a] School of Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg, 2050, South Africa
[b] School of Mathematical Sciences, Zhejiang Normal University, Jinhua 321004, People's Republic of China

ARTICLE INFO

MSC:
47H09
47H10
49J20
49J40

Keywords:
Variational inequalities
Quasi-monotone
Weak convergence
Strong convergence

ABSTRACT

In this paper we study the convergence analysis of a Golden Ratio Algorithm with a backward inertial step and a fully adaptive step size procedure for the purpose of approximating solutions of variational inequalities in Hilbert spaces. We present a weak convergence result when the operator is quasi-monotone and locally Lipschitz continuous, and a strong convergence result in the setting of strong pseudo-monotonicity. Our algorithm features one operator evaluation and one projection computation at the current iteration. We recover interesting algorithms in the literature, for instance, the Golden Ratio Algorithm. Numerical experiments were conducted to validate the theoretical convergence analysis.

1. Introduction

Let us consider the Variational Inequality Problem (VIP)

$$\text{find } \tilde{x} \in C \text{ such that } \langle \mathcal{P}\tilde{x}, \tilde{y} - \tilde{x} \rangle \geq 0, \ \forall \tilde{y} \in C, \tag{1}$$

where $\mathcal{P}$ is a continuous operator defined on a real Hilbert space H and C is a nonempty closed and convex subset of H. Problems of the form (1) are useful in fields like engineering, transportation, economics, mathematical programming and mechanics (see [1–4] for illustrative examples).

Several methods have been studied for approximating solutions of the VIP (1) when the operator $\mathcal{P}$ is monotone (or pseudo-monotone) and Lipschitz continuous. A common method for this purpose is the extragradient method [5]: $x_1 \in C$, $\lambda_k \in (0, \frac{1}{L})$ and $L > 0$,

$$\begin{cases} y_k = P_C(x_k - \lambda_k \mathcal{P} x_k) \\ x_{k+1} = P_C(x_k - \lambda_k \mathcal{P} y_k), k \geq 1. \end{cases} \tag{2}$$

When $\mathcal{P}$ exhibits monotonicity or pseudo-monotonicity and Lipschitz continuity conditions in a real Hilbert space, (2) is proved to converge weakly (see [6–8]). However, in this method, the operator $\mathcal{P}$ and the projection onto C, P_C are computed twice at each iteration. This leads to the computational expensiveness of this method especially when $\mathcal{P}$ and P_C are difficult to compute.

* Corresponding author.
E-mail addresses: chinedu.izuchukwu@wits.ac.za (C. Izuchukwu), yekini.shehu@zjnu.edu.cn (Y. Shehu).

https://doi.org/10.1016/j.cnsns.2024.108217
Received 10 March 2024; Received in revised form 5 June 2024; Accepted 8 July 2024
Available online 14 July 2024

Popov [9] on the one hand, proposed the following method which reduces the number of computations of $\mathcal{P}$ in the extragradient method (2) from two to one: $x_1, y_1 \in C$ and $\lambda_k \in (0, \frac{1}{3L}]$,

$$\begin{cases} x_{k+1} = P_C(x_k - \lambda_k \mathcal{P} y_k) \\ y_{k+1} = P_C(x_{k+1} - \lambda_k \mathcal{P} y_k), k \geq 1. \end{cases} \tag{3}$$

However, P_C is computed twice at each iteration. Similar to the extragradient method, the Popov's method (3) is proved to converge weakly in a real Hilbert space when $\mathcal{P}$ exhibits monotonicity or pseudo-monotonicity and Lipschitz continuity conditions (as established in [10,11]).

Censor–Gibali–Reich [12] on the other hand, proposed the following subgradient extragradient method which reduces the amount of computations of the metric projection in Algorithm (2) to one: $x_1 \in H$, $\lambda_k \in (0, \frac{1}{L})$ and $L > 0$,

$$\begin{cases} y_k = P_C(x_k - \lambda_k \mathcal{P} x_k) \\ D_k := \{w \in H : \langle x_k - \lambda_k \mathcal{P} x_k - y_k, w - y_k \rangle \leq 0\}, \\ x_{k+1} = P_{D_k}(x_k - \lambda_k \mathcal{P} y_k), k \geq 1. \end{cases} \tag{4}$$

The subgradient extragradient method (4) is also proved to converge weakly in a real Hilbert space when $\mathcal{P}$ exhibits monotonicity or pseudo-monotonicity and Lipschitz continuity conditions (as established in [13–16]).

An alternative to Algorithm (4) is the following forward–backward–forward method, which was first proposed by Tseng [17] in the context of monotone inclusion problems: $x_1 \in H$, $\lambda_k \in (0, \frac{1}{L})$ and $L > 0$,

$$\begin{cases} y_k = P_C(x_k - \lambda_k \mathcal{P} x_k) \\ x_{k+1} = y_k + \lambda_k \left(\mathcal{P} x_k - \mathcal{P} y_k\right), k \geq 1. \end{cases} \tag{5}$$

Like the subgradient extragradient method (4), in this method, P_C is computed only once while $\mathcal{P}$ is computed twice at each iteration. Similarly, the weak convergence of this method has been widely studied in real Hilbert space when $\mathcal{P}$ exhibits monotonicity or pseudo-monotonicity and Lipschitz continuity conditions. For inertial variants of methods (2)–(5), see [18–23] and the references contained there.

Related results

The available research on the convergence results for projection-type methods applied to the VIP (1) in scenarios where $\mathcal{P}$ exhibits both quasi-monotonicity and Lipschitz continuity conditions remains limited in real Hilbert spaces. This gap persists because the Minty variant of VIP (1) does not coincide with VIP (1) when $\mathcal{P}$ is quasi-monotone. Consequently, the techniques applied in establishing the convergence results for pseudo-monotone cases cannot be directly extrapolated to the quasi-monotone scenarios. While methods proposed in [24–26] address VIP (1) under the assumption of non-monotonicity and continuity of $\mathcal{P}$ in finite-dimensional spaces, those in [25,26] entail computationally expensive projection onto $(k+1)$ half-spaces, particularly burdensome as k increases. Additionally, the method presented in [24] necessitates the computation of projections onto the intersection of two half-spaces and the feasible set C, further exacerbating computational costs. Further related weak and strong convergence results on quasimonotone variational inequality problem can be found in [27,28].

In their work cited as [29], Liu and Yang demonstrated weak convergence properties for the extragradient method (2), the subgradient extragradient method (4), and the forward–backward–forward method (5), under the conditions of quasi-monotonicity, Lipschitz continuity, and sequential weak continuity of $\mathcal{P}$ within an infinite-dimensional Hilbert space. Furthermore, weak convergence results for extragradient method (2) were also outlined in [30], focusing on quasi-monotonic settings. Wang et al. [31] provided similar weak convergence outcomes for the inertial version of the projection and contraction method, incorporating Barzilai–Borwein stepsize, with A exhibiting quasi-monotonicity and Lipschitz continuity within Hilbert spaces. Importantly, the algorithms presented in [31], specifically Algorithm 3.1 and Algorithm 3.2, require two projections P_C and two evaluations of $\mathcal{P}$ during each iteration.

In designing projection-type methods for tackling variational inequalities, a key concern revolves around reducing the amount of computations of projections and functional evaluations per iteration. This arises from the fact that when dealing with a general closed and convex feasible set, the numerical implementation of projection-type methods entails solving minimal distance problems. The efficiency of the method can be significantly affected if numerous such distance problems need to be addressed. Additionally, in large-scale optimization scenarios where functional evaluation is resource-intensive, the efficiency of algorithms could also suffer from the need for multiple functional evaluations. Recent advancements in iterative techniques have aimed to address these challenges by proposing and exploring methods requiring only one projection and one functional evaluation. These developments are documented and scrutinized in contemporary literature, as evidenced by references such as [32–36].

In his work [33], Malitsky developed a projected reflected gradient method tailored for tackling variational inequalities. Within this study, Malitsky derived convergence results, including both weak and linear convergence, under assumptions of monotonicity and Lipschitz continuity, particularly in real Hilbert spaces. An intriguing aspect of the projected reflected gradient method is its simplified approach, necessitating one functional evaluation and one projection per iteration. Furthermore, Malitsky demonstrated, through numerical experimentation, that this method demonstrates superior computational efficiency compared to other established techniques, such as Korpelevich's extragradient method [5], Popov's extragradient method [9], the subgradient extragradient method [12], and Tseng's forward–backward–forward method [17]. Various iterations of Malitsky's projected reflected gradient method have since garnered attention, with studies by Chang and Bai [37], further explorations by Malitsky himself [34], and investigations by other researchers [35], yielding pertinent convergence results within the context of monotonicity and Lipschitz continuity.

Malistky [32, Algorithm 1] introduced the following Golden Ratio Algorithm with a fully self adaptive stepsizes for tackling VIP (1) in finite-dimensional spaces:

Algorithm 1 Adaptive Golden Ratio Algorithm

1: Choose $y_0, y_1 \in C, \lambda_0 > 0, \phi \in (1, \varphi], \bar{\lambda} > 0$. Set $x_0 = y_1, \theta_0 = 1, \rho := \frac{1}{\phi} + \frac{1}{\phi^2}$ and $k := 1$.

2: Compute

$$\lambda_k = \min\left\{\rho\lambda_{k-1}, \frac{\phi\theta_{k-1}}{4\lambda_{k-1}}\frac{\|y_k - y_{k-1}\|^2}{\|\mathcal{P}y_k - \mathcal{P}y_{k-1}\|^2}, \bar{\lambda}\right\} \tag{6}$$

and

3: compute

$$\begin{cases} x_k = \frac{(\phi-1)y_k + x_{k-1}}{\phi}, \\ y_{k+1} = P_C(x_k - \lambda_k \mathcal{P}y_k), \end{cases} \tag{7}$$

where

$$\theta_k = \frac{\lambda_k \phi}{\lambda_{k-1}}. \tag{8}$$

The Algorithm 1 is designed to streamline the process, requiring only a single functional evaluation of $\mathcal{P}$ and one projection P_C during each iteration. Additionally, the step sizes λ_k produced from 6 are adaptively determined. The convergence analysis provided in [32] demonstrates both an ergodic $\mathcal{O}(1/k)$ convergence rate and R-linear convergence rate for Algorithm 1.

In a recent development, Yang and Liu introduced a modification of the Golden Ratio Algorithm, as outlined in [36, Algorithm 4.1], specifically tailored for addressing the VIP (1) problem in cases where the operator A exhibits pseudo-monotonicity within infinite-dimensional Hilbert spaces:

Algorithm 2 Self-Adaptive Algorithm

1: Choose $\lambda_1 > 0, x_0, y_0, y_1 \in C, \mu \in (0,1), \alpha \in (0,1), \theta \in (0,1], \delta \in \left(\frac{\sqrt{1+4(\frac{\alpha}{2-\theta}+1-\alpha)}-1}{2}, 1\right)$. Set $k := 1$.

2: Compute

$$\begin{cases} x_k = (1-\delta)y_k + \delta x_{k-1}, \\ y_{k+1} = P_C(x_k - \lambda_k \mathcal{P}y_k), \end{cases} \tag{9}$$

where

$$\lambda_{k+1} = \begin{cases} \min\left\{\frac{\alpha\mu\theta(\|y_k - y_{k-1}\|^2 + \|y_{k+1} - y_k\|^2)}{4\delta\langle \mathcal{P}y_{k-1} - \mathcal{P}y_k, y_{k+1} - y_k\rangle}, \ \lambda_k\right\}, & \text{if } \langle \mathcal{P}y_{k-1} - \mathcal{P}y_k, y_{k+1} - y_k\rangle > 0, \\ \lambda_k, & \text{otherwise.} \end{cases} \tag{10}$$

Utilizing Algorithm 2, Yang and Liu [36] established weak and strong convergence outcomes under the conditions of pseudo-monotonicity and strong pseudo-monotonicity, respectively, within Hilbert spaces. However, it is noteworthy that the step sizes generated by Eq. (10) exhibit a restrictive monotone nonincreasing behavior, unlike the more fully adaptive step sizes outlined in 6.

Our contributions

Inspired by the findings presented in [32,36], our focus in this paper is to propose and analyze a Golden Ratio Algorithm augmented with a backward inertial step and fully adaptive step sizes for addressing variational inequalities within the framework of quasi-monotonicity and locally Lipschitz continuity in infinite-dimensional Hilbert spaces. To this end, we contribute the following to the literature:

- We propose a Golden Ratio Algorithm incorporating a backward inertial step, which entails a single computation of projection onto the feasible set and necessitates only one functional evaluation per iteration. Notably, our proposed algorithm encompasses the Golden Ratio Algorithm [32, Algorithm 1] as a special case.

- Another notable aspect of our method lies in the flexibility of the step sizes, which are not confined to being nonincreasing, as in [16,18,19,38] and related works.
- We derive weak and strong convergence results for our proposed method.
- Numerical experiments are conducted to showcase the superiority of our proposed approach over several related methods documented in the literature.

Organization

We organize this paper in the following manner:

- Section 2: The foundation for our convergence analysis is laid by introducing fundamental definitions and key results.
- Section 3: The main results are presented, wherein we introduce our algorithm along with both weak and strong convergence results.
- Section 4: Several numerical examples are presented to illustrate the effectiveness of the proposed method.
- Section 5: Concluding remarks are offered to summarize the findings and provide insights for future research directions.

2. Preliminaries

In this paper, we use the notation $x_n \rightharpoonup \tilde{v}$ to indicate the weak convergence of the sequence $\{x_n\}$ to a point $\tilde{v}$.
We call the operator $\mathcal{P}$;

(i) *L-Lipschitz continuous*, if for all $\tilde{u}, \tilde{v} \in H$, there is an $L > 0$ for which

$$\|\mathcal{P}\tilde{u} - \mathcal{P}\tilde{v}\| \le L\|\tilde{u} - \tilde{v}\|;$$

(ii) *L-strongly monotone*, if for all $\tilde{u}, \tilde{v} \in H$, there exists $L > 0$ such that

$$\langle \mathcal{P}\tilde{u} - \mathcal{P}\tilde{v}, \tilde{u} - \tilde{v}\rangle \ge L\|\tilde{u} - \tilde{v}\|^2;$$

(iii) *monotone*, if for all $\tilde{u}, \tilde{v} \in H$,

$$\left\langle \mathcal{P}\tilde{u} - \mathcal{P}\tilde{v}, \tilde{u} - \tilde{v}\right\rangle \ge 0;$$

(iv) *$\tilde{\eta}$-strongly pseudo-monotone* if for all $\tilde{u}, \tilde{v} \in H$, there exists $\tilde{\eta} > 0$ such that $\langle \mathcal{P}\tilde{v}, \tilde{u} - \tilde{v}\rangle \ge 0 \Rightarrow \langle \mathcal{P}\tilde{u}, \tilde{u} - \tilde{v}\rangle \ge \tilde{\eta}\|\tilde{u} - \tilde{v}\|^2$;

(v) *pseudo-monotone*, if for all $\tilde{u}, \tilde{v} \in H$,

$$\langle \mathcal{P}\tilde{v}, \tilde{u} - \tilde{v}\rangle \ge 0 \implies \langle \mathcal{P}\tilde{u}, \tilde{u} - \tilde{v}\rangle \ge 0;$$

(vi) quasi-monotone, if for all $\tilde{u}, \tilde{v} \in H$,

$$\langle \mathcal{P}\tilde{v}, \tilde{u} - \tilde{v}\rangle > 0 \implies \langle \mathcal{P}\tilde{u}, \tilde{u} - \tilde{v}\rangle \ge 0;$$

(vii) *sequentially weakly-strongly continuous,* if for any sequence $\{u_n\}$ converging weakly to $\tilde{v}$, we have that $\{\mathcal{P}u_n\}$ converges strongly to $\mathcal{P}\tilde{v}$,

(viii) *sequentially weakly continuous,* if for any sequence $\{u_n\}$ converging weakly to $\tilde{v}$, we have that $\{\mathcal{P}u_n\}$ converges weakly to $\mathcal{P}\tilde{v}$.

The implications (ii) implies (iii), (iv) implies (v), and (iii) implies (v) implies (vi) are valid; however, converse of these statements are not universally valid.

Consider the Minty expression of the VIP (1):

$$\text{find } \tilde{x} \in C \text{ such that } \langle \mathcal{P}\tilde{y}, \tilde{y} - \tilde{x}\rangle \ge 0, \text{ for all } \tilde{y} \in C. \tag{11}$$

The set of solutions of this problem is designated as S_D. We know that S_D is a closed and convex subset of C. Given the convexity of C and the continuity of $\mathcal{P}$, we get $S_D \subset S$.

The lemma below provides conditions under which S_D is guaranteed to be nonempty.

Lemma 2.1 (*see [26]*). *If any of the following conditions hold:*

(i) $\mathcal{P}$ is pseudo-monotone on C and $S \neq \emptyset$,
(ii) $\mathcal{P}$ represents the gradient of G, where G is a differential quasiconvex function defined on an open set $K \supset C$ and achieves its global minimum on C,
(iii) $\mathcal{P}$ is quasi-monotone and non-zero on a bounded set C,
(iv) $\mathcal{P}$ is non-zero and quasi-monotone on C and there is $\tilde{r} > 0$ for which $\|\tilde{x}\| \ge \tilde{r}$ for any $\tilde{x} \in C$, there is $\tilde{y} \in C$ satisfying $\|\tilde{y}\| \le \tilde{r}$ and $\langle \mathcal{P}\tilde{x}, \tilde{y} - \tilde{x}\rangle \le 0$,
(v) $\mathcal{P}$ is quasi-monotone on C, the interior of C is nonempty, and there is $\tilde{x}^* \in S$ for which $\mathcal{P}\tilde{x}^*$ is non-zero,

then we have that $S_D \neq \emptyset$.

The mapping P_C defined on H onto C and designates to each $\tilde{x} \in H$, the unique point in C such that

$$\|\tilde{x} - P_C\tilde{x}\| = \inf_{\tilde{y}\in C}\{\|\tilde{x} - \tilde{y}\|\},$$

is called a metric projection. The inequality below holds for all $\tilde{y} \in C$;

$$\langle \tilde{x} - P_C\tilde{x}, \tilde{y} - P_C\tilde{x}\rangle \leq 0. \tag{12}$$

Lemma 2.2. *The equalities below are valid:*
(i) $2\langle \tilde{u}, \tilde{v}\rangle = \|\tilde{u}\|^2 + \|\tilde{v}\|^2 - \|\tilde{u} - \tilde{v}\|^2 = \|\tilde{u} + \tilde{v}\|^2 - \|\tilde{u}\|^2 - \|\tilde{v}\|^2,\ \forall \tilde{u}, \tilde{v} \in H$;
(ii) $\|(1 + \tilde{\alpha})\tilde{u} - \tilde{\alpha}\tilde{v}\|^2 = (1 + \tilde{\alpha})\|\tilde{u}\|^2 - \tilde{\alpha}\|\tilde{v}\|^2 + \tilde{\alpha}(1 + \tilde{\alpha})\|\tilde{u} - \tilde{v}\|^2,\ \forall \tilde{u}, \tilde{v} \in H,\ \tilde{\alpha} \in \mathbb{R}$.

Lemma 2.3 ([39]). *Let X be a subset of H, with $X \neq \emptyset$ and let $\{y_k\}$ be a sequence in H that satisfies the following conditions:*
(i) $\lim_{k\to\infty} \|y_k - y\|$ exists for any $y \in X$;
(ii) every sequential weak limit point of $\{y_k\}$ belongs to X.
Then $\{y_k\}$ converges weakly to a point in X.

3. Main results

Our goal here is to discuss and delve into the analysis of the algorithm that we presented, both in terms of weak and strong convergence.

3.1. Proposed algorithm

We begin with the assumption that the parameters $\delta \in \left[\frac{\sqrt{5}-1}{2}, 1\right)$ and $\theta \in \left(-\frac{1}{2}, 0\right]$ satisfy the following condition:

$$\delta(1 + \delta)(1 + \theta) \geq 1. \tag{13}$$

Next, the algorithm is detailed as follows:

Algorithm 3 Fully Adaptive Method

1: Pick $\lambda_0, \bar{\lambda} > 0,\ y_1,\ y_0, x_0, w_0 \in C$. Choose $\delta \in \left[\frac{\sqrt{5}-1}{2}, 1\right)$ and $\theta \in \left(-\frac{1}{2}, 0\right]$ such that (13) is fulfilled. Set $\theta_0 = 1,\ \rho := \delta(1+\delta)(1+\theta)$ and $k := 1$.
2: Find the stepsize:

$$\lambda_k = \min\left\{\rho\lambda_{k-1}, \frac{\theta_{k-1}}{4\lambda_{k-1}\delta}\frac{\|y_k - y_{k-1}\|^2}{\|\mathcal{P}y_k - \mathcal{P}y_{k-1}\|^2}, \bar{\lambda}\right\} \tag{14}$$

and
3: compute

$$\begin{cases} x_k = (1-\delta)y_k + \delta w_{k-1}, \\ w_k = x_k + \theta(x_k - x_{k-1}), \\ y_{k+1} = P_C(w_k - \lambda_k \mathcal{P}y_k), \end{cases} \tag{15}$$

where

$$\theta_k = \frac{\lambda_k}{\lambda_{k-1}\delta}. \tag{16}$$

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

We give the following comments on our proposed algorithm.

Remark 3.1.

(i) Our proposed Algorithm 3 simplifies to the adaptive Golden Ratio Algorithm presented in [32, Algorithm 1] with $\delta = \frac{1}{\phi}$ (ϕ denoting the golden ratio, $\phi^2 = 1 + \phi, \phi = \frac{\sqrt{5}+1}{2}$) and $\theta = 0$.
(ii) Our Algorithm 3 can be seen as a modified Golden Ratio Algorithm (see [32, Algorithm 1]) with inertial extrapolation. In this case, the inertial factor θ in Algorithm 3 is non-positive; a feature which is different from many inertial algorithms in the literature where the inertial factor is usually non-negative. The non-positivity of the inertial factor θ could probably be due to the nature of Algorithm 3. Hence, our Algorithm 3 could be seen as Golden Ratio Algorithm [32, Algorithm 1] with a backward inertial step $w_k = x_k + \theta(x_k - x_{k-1})$.

(iii) The iterate w_k in Algorithm 3 can also be seen as a convex combination of x_k and x_{k-1}. Thus, we can write $w_k = (1-(-\theta))x_k + (-\theta)x_{k-1}$, where $-\theta \in \left[0, \frac{1}{2}\right)$.
(iv) In [36, Algorithm 4.1], the stepsizes that the authors defined are nonincreasing. Hence, they are restrictive. But the stepsizes we presented in (14) do not need to be nonincreasing; rather, λ_k can be bigger than λ_{k-1} (since for $\delta > \frac{\sqrt{5}-1}{2}$ and $\theta > \frac{1}{\delta(1+\delta)} - 1$, we have $\rho > 1$) and $\{\lambda_k\}$ is bounded.
(v) The iteration (15) simplifies to [36, Algorithm4.1] under the condition that $\mathcal{P}$ is a pseudo-monotone operator and $\theta = 0$. Furthermore, in [37, Algorithm 3.1], larger stepsizes are employed, but with the requirement that $\mathcal{P}$ is monotone and $\theta = 0$.
(vi) A single projection P_C onto C and a single functional evaluation of $\mathcal{P}$ per iteration are needed in Algorithm 3. This method seems simpler than the methods outlined in [29–31], which require multiple functional evaluations and projections P_C at each iteration for solving VIP (1) in the setting of quasi-monotonicity.

We make the assumptions below to ensure the weak convergence of the algorithm we presented above (Algorithm 3).

Assumption 3.2.

(a) $S_D \neq \emptyset$,
(b) $\mathcal{P}$ is locally Lipschitz continuous on C, meaning that $\mathcal{P}$ is Lipschitz continuous on any bounded subset of C,
(c) whenever $\{y_k\} \subset C$ and $y_k \rightharpoonup v^*$, we have $\|\mathcal{P}v^*\| \leq \liminf_{n\to\infty} \|\mathcal{P}y_k\|$,
(d) $\mathcal{P}$ is quasi-monotone on H.

Remark 3.3. In Assumption 3.2(c), we imposed a condition weaker than the sequentially weakly continuity requirement placed on $\mathcal{P}$ in [29,36] (refer to [7] for further details).

From (14), we can derive the following insights:

Remark 3.4. Eq. (14) in Algorithm 3 gives rise to two scenarios:

Case 1: When $\lambda_k \leq \rho\lambda_{k-1}$, it follows that

$$\theta_k \leq \frac{\rho}{\delta}. \tag{17}$$

Case 2: When $\frac{\theta_{k-1}}{\lambda_{k-1}\delta} = \frac{\theta_k\theta_{k-1}}{\lambda_k}$, we obtain

$$\lambda_k^2\|\mathcal{P}y_k - \mathcal{P}y_{k-1}\|^2 \leq \frac{\theta_k\theta_{k-1}}{4}\|y_k - y_{k-1}\|^2. \tag{18}$$

Applying similar thoughts as in [32, Lemma 2], we can demonstrate that when y_k is bounded, both sequences $\{\lambda_k\}$ and $\{\theta_k\}$ are bounded and remain separated from zero. This assertion is elaborated in the lemma below.

Lemma 3.5. *Let Assumption 3.2 (b) holds. Then $\{\lambda_k\}$ and $\{\theta_k\}$ are both bounded away from zero provided that $\{y_k\}$ is bounded.*

Proof. Using (14), we get $\lambda_k \leq \bar{\lambda}, \forall k$. Since $\{y_k\}$ is bounded, we see from Assumption 3.2 (b) that there exists $L > 0$ such that $\|\mathcal{P}y_k - \mathcal{P}y_{k-1}\| \leq L\|y_k - y_{k-1}\|$. Let us take L large enough such that $\lambda_j \geq \frac{1}{4L^2\delta^2\bar{\lambda}}$ for $j = 0, 1$. Now, suppose $\lambda_j \geq \frac{1}{4L^2\delta^2\bar{\lambda}}$, $j = 0, 1, \ldots, k-1$. Then we have either (since $\rho \geq 1$)

$$\lambda_k = \rho\lambda_{k-1} \geq \lambda_{k-1} \geq \frac{1}{4L^2\delta^2\bar{\lambda}}$$

or

$$\begin{aligned} \lambda_k &= \frac{1}{4\lambda_{k-2}\delta^2}\frac{\|y_k - y_{k-1}\|^2}{\|\mathcal{P}y_k - \mathcal{P}y_{k-1}\|^2} \text{ (from (14) and (16))} \\ &\geq \frac{1}{4\lambda_{k-2}L^2\delta^2} \geq \frac{1}{4L^2\delta^2\bar{\lambda}}. \end{aligned} \tag{19}$$

Thus, we have in both cases that

$$\lambda_k \geq \frac{1}{4L^2\delta^2\bar{\lambda}}. \tag{20}$$

Therefore, it can be concluded that $\{\theta_k\}$ is bounded away from zero. □

3.2. Weak and strong convergence

This subsection provides the analysis of the algorithm, both in terms of weak and strong convergence. We commence as follows.

Lemma 3.6. *Under the fulfillment of Assumption 3.2 (a) and (b), we have that the sequences $\{x_k\}$ and $\{y_k\}$ produced by Algorithm 3 are bounded .*

Proof. Using (15), gives

$$y_{k+1} = P_C(w_k - \lambda_k \mathcal{P} y_k),$$

which by (12), further gives

$$0 \leq \langle y_{k+1} - w_k + \lambda_k \mathcal{P} y_k, y - y_{k+1} \rangle \ \forall y \in C. \tag{21}$$

Similarly from

$$y_k = P_C(w_{k-1} - \lambda_{k-1} \mathcal{P} y_{k-1})$$

we see that

$$0 \leq \langle y_k - w_{k-1} + \lambda_{k-1} \mathcal{P} y_{k-1}, y - y_k \rangle \ \forall y \in C. \tag{22}$$

Putting $y = \bar{x} \in S_D$ in (21) and $y = y_{k+1}$ in (22), we have

$$0 \leq \langle y_{k+1} - w_k + \lambda_k \mathcal{P} y_k, \bar{x} - y_{k+1} \rangle \tag{23}$$

and

$$0 \leq \langle y_k - w_{k-1} + \lambda_{k-1} \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle. \tag{24}$$

From $x_k = (1-\delta) y_k + \delta w_{k-1}$, we get

$$y_k = \frac{1}{1-\delta} x_k - \frac{\delta}{1-\delta} w_{k-1}.$$

Therefore,

$$y_k - x_k = \frac{\delta}{1-\delta}(x_k - w_{k-1}) = \delta(y_k - w_{k-1}). \tag{25}$$

Using (25) in (24), we have

$$0 \leq \langle \frac{1}{\delta}(y_k - x_k) + \lambda_{k-1} \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle. \tag{26}$$

From (23), we have (since $\delta > 0$)

$$0 \leq \frac{1}{\delta} \langle y_{k+1} - w_k + \lambda_k \mathcal{P} y_k, \bar{x} - y_{k+1} \rangle. \tag{27}$$

Also from (26), we obtain

$$\begin{aligned} 0 &\leq \langle \frac{\theta_k}{\delta}(y_k - x_k) + \theta_k \lambda_{k-1} \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle \\ &= \langle \frac{\theta_k}{\delta}(y_k - x_k) + \frac{\lambda_k}{\delta} \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle. \end{aligned} \tag{28}$$

Adding (27) and (28), we get

$$\begin{aligned} 0 \leq\ & \frac{1}{\delta} \langle y_{k+1} - w_k + \lambda_k \mathcal{P} y_k, \bar{x} - y_{k+1} \rangle \\ & + \langle \frac{\theta_k}{\delta}(y_k - x_k) + \frac{\lambda_k}{\delta} \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle. \end{aligned} \tag{29}$$

Thus,

$$\begin{aligned} 0 \leq\ & \langle y_{k+1} - w_k, \bar{x} - y_{k+1} \rangle + \theta_k \langle y_k - x_k, y_{k+1} - y_k \rangle \\ & + \lambda_k \langle \mathcal{P} y_k, \bar{x} - y_{k+1} \rangle + \lambda_k \langle \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle. \end{aligned} \tag{30}$$

Therefore,

$$\begin{aligned} & \langle w_k - y_{k+1}, \bar{x} - y_{k+1} \rangle + \theta_k \langle x_k - y_k, y_{k+1} - y_k \rangle \\ \leq\ & \lambda_k \langle \mathcal{P} y_k, \bar{x} - y_{k+1} \rangle + \lambda_k \langle \mathcal{P} y_{k-1}, y_{k+1} - y_k \rangle. \end{aligned} \tag{31}$$

Note that

$$\begin{aligned} \langle w_k - y_{k+1}, \bar{x} - y_{k+1} \rangle = \ & \frac{1}{2} \Big[\|w_k - y_{k+1}\|^2 + \|y_{k+1} - \bar{x}\|^2 \\ & - \|w_k - \bar{x}\|^2 \Big] \end{aligned} \tag{32}$$

and

$$\begin{aligned} \theta_k \langle x_k - y_k, y_{k+1} - y_k \rangle = \ & \frac{\theta_k}{2} \Big[\|x_k - y_k\|^2 + \|y_{k+1} - y_k\|^2 \\ & - \|x_k - y_{k+1}\|^2 \Big]. \end{aligned} \tag{33}$$

Using (32) and (33) in (31), we have

$$\begin{aligned}&\frac{1}{2}\|y_{k+1}-\bar{x}\|^2-\frac{1}{2}\|w_k-\bar{x}\|^2+\frac{1}{2}\|w_k-y_{k+1}\|^2\\&+\frac{\theta_k}{2}\|x_k-y_k\|^2+\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2-\frac{\theta_k}{2}\|x_k-y_{k+1}\|^2\\&\le\lambda_k\langle\mathcal{P}y_k,\bar{x}-y_{k+1}\rangle+\lambda_k\langle\mathcal{P}y_{k-1},y_{k+1}-y_k\rangle.\end{aligned}$$

That is,

$$\begin{aligned}\frac{1}{2}\|y_{k+1}-\bar{x}\|^2 &\le \frac{1}{2}\|w_k-\bar{x}\|^2-\frac{1}{2}\|w_k-y_{k+1}\|^2\\&\quad-\frac{\theta_k}{2}\|x_k-y_k\|^2-\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2+\frac{\theta_k}{2}\|x_k-y_{k+1}\|^2\\&\quad+\lambda_k\langle\mathcal{P}y_k,\bar{x}-y_{k+1}\rangle+\lambda_k\langle\mathcal{P}y_{k-1},y_{k+1}-y_k\rangle\\&=\frac{1}{2}\|w_k-\bar{x}\|^2-\frac{1}{2}\|w_k-y_{k+1}\|^2\\&\quad-\frac{\theta_k}{2}\|x_k-y_k\|^2-\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2+\frac{\theta_k}{2}\|x_k-y_{k+1}\|^2\\&\quad+\lambda_k\langle\mathcal{P}y_k,\bar{x}-y_k\rangle+\lambda_k\langle\mathcal{P}y_k-\mathcal{P}y_{k-1},y_k-y_{k+1}\rangle \qquad (34)\\&\le\frac{1}{2}\|w_k-\bar{x}\|^2-\frac{1}{2}\|w_k-y_{k+1}\|^2\\&\quad-\frac{\theta_k}{2}\|x_k-y_k\|^2-\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2+\frac{\theta_k}{2}\|x_k-y_{k+1}\|^2\\&\quad+\lambda_k\langle\mathcal{P}y_k-\mathcal{P}y_{k-1},y_k-y_{k+1}\rangle, \qquad (35)\end{aligned}$$

where the last inequality follows since $\bar{x}\in S_D$ (hence, $\langle\mathcal{P}y_k,\bar{x}-y_k\rangle\le 0$). Now, note that

$$\begin{aligned}&2\lambda_k\langle\mathcal{P}y_k-\mathcal{P}y_{k-1},y_k-y_{k+1}\rangle\le 2\lambda_k\|\mathcal{P}y_k-\mathcal{P}y_{k-1}\|\|y_k-y_{k+1}\|\\&\le\sqrt{\theta_k\theta_{k-1}}\|y_k-y_{k-1}\|\|y_k-y_{k+1}\|\\&\le\frac{\theta_k}{2}\|y_k-y_{k+1}\|^2+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2.\end{aligned} \tag{36}$$

We have from (35) and (36) that

$$\begin{aligned}\|y_{k+1}-\bar{x}\|^2 &\le \|w_k-\bar{x}\|^2-\|w_k-y_{k+1}\|^2\\&\quad-\theta_k\left[\|x_k-y_k\|^2+\|y_{k+1}-y_k\|^2-\|x_k-y_{k+1}\|^2\right]\\&\quad+\frac{\theta_k}{2}\|y_k-y_{k+1}\|^2+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2.\end{aligned} \tag{37}$$

From $x_k=(1-\delta)y_k+\delta w_{k-1}$, one has

$$y_{k+1}=\frac{1}{1-\delta}x_{k+1}-\frac{\delta}{1-\delta}w_k. \tag{38}$$

Applying Lemma 2.2 (ii) in (38) gives

$$\begin{aligned}\|y_{k+1}-\bar{x}\|^2 &= \|\frac{1}{1-\delta}(x_{k+1}-\bar{x})-\frac{\delta}{1-\delta}(w_k-\bar{x})\|^2\\&=\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2-\frac{\delta}{1-\delta}\|w_k-\bar{x}\|^2\\&\quad+\frac{\delta}{(1-\delta)^2}\|x_{k+1}-w_k\|^2\\&=\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2-\frac{\delta}{1-\delta}\|w_k-\bar{x}\|^2\\&\quad+\delta\|y_{k+1}-w_k\|^2.\end{aligned} \tag{39}$$

Using (39) in (37), we obtain

$$\begin{aligned}&\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2-\frac{\delta}{1-\delta}\|w_k-\bar{x}\|^2+\delta\|y_{k+1}-w_k\|^2\\&\le\|w_k-\bar{x}\|^2-\|w_k-y_{k+1}\|^2-\theta_k\|x_k-y_k\|^2\\&\quad-\theta_k\|y_{k+1}-y_k\|^2+\theta_k\|x_k-y_{k+1}\|^2\\&\quad+\frac{\theta_k}{2}\|y_k-y_{k+1}\|^2+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2.\end{aligned} \tag{40}$$

Thus,

$$\begin{aligned}\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2 \leq\ & \frac{1}{1-\delta}\|w_k-\bar{x}\|^2-(1+\delta)\|w_k-y_{k+1}\|^2\\ &+\theta_k\|x_k-y_{k+1}\|^2-\theta_k\|x_k-y_k\|^2\\ &+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2-\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2. \end{aligned} \tag{41}$$

Now

$$\begin{aligned}\|w_k-\bar{x}\|^2 =\ & \|(1+\theta)(x_k-\bar{x})-\theta(x_{k-1}-\bar{x})\|^2\\ =\ & (1+\theta)\|x_k-\bar{x}\|^2-\theta\|x_{k-1}-\bar{x}\|^2\\ &+\theta(1+\theta)\|x_k-x_{k-1}\|^2. \end{aligned} \tag{42}$$

Also, observe that (since $\theta \leq 0$)

$$\begin{aligned}\|y_{k+1}-w_k\|^2 =\ & \|(1+\theta)(x_k-y_{k+1})-\theta(x_{k-1}-y_{k+1})\|^2\\ =\ & (1+\theta)\|x_k-y_{k+1}\|^2-\theta\|x_{k-1}-y_{k+1}\|^2\\ &+\theta(1+\theta)\|x_k-x_{k-1}\|^2\\ \geq\ & (1+\theta)\|x_k-y_{k+1}\|^2+\theta(1+\theta)\|x_k-x_{k-1}\|^2. \end{aligned} \tag{43}$$

Putting (42) and (43) in (41), we obtain

$$\begin{aligned}&\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2 \leq \frac{1}{1-\delta}\Big[(1+\theta)\|x_k-\bar{x}\|^2-\theta\|x_{k-1}-\bar{x}\|^2\\ &+\theta(1+\theta)\|x_k-x_{k-1}\|^2\Big]-(1+\delta)(1+\theta)\|x_k-y_{k+1}\|^2\\ &+\theta_k\|y_{k+1}-x_k\|^2-\theta_k\|y_k-x_k\|^2+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2\\ &-\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2-(1+\delta)(\theta^2+\theta)\|x_k-x_{k-1}\|^2. \end{aligned}$$

Thus,

$$\begin{aligned}&\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2 \leq \frac{1}{1-\delta}\|x_k-\bar{x}\|^2+\frac{\theta}{1-\delta}\|x_k-\bar{x}\|^2\\ &-\frac{\theta}{1-\delta}\|x_{k-1}-\bar{x}\|^2+\frac{\theta(1+\theta)}{1-\delta}\|x_k-x_{k-1}\|^2\\ &+\Big(\theta_k-(1+\delta)(1+\theta)\Big)\|y_{k+1}-x_k\|^2-(1+\delta)(\theta^2+\theta)\|x_k-x_{k-1}\|^2\\ &-\theta_k\|y_k-x_k\|^2+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2-\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2. \end{aligned} \tag{44}$$

Re-arranging (44) gives

$$\begin{aligned}&\frac{1}{1-\delta}\|x_{k+1}-\bar{x}\|^2-\frac{\theta}{1-\delta}\|x_k-\bar{x}\|^2+\frac{\theta_k}{2}\|y_{k+1}-y_k\|^2\\ \leq\ &\frac{1}{1-\delta}\|x_k-\bar{x}\|^2-\frac{\theta}{1-\delta}\|x_{k-1}-\bar{x}\|^2+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2\\ &+\Big(\theta_k-(1+\delta)(1+\theta)\Big)\|y_{k+1}-x_k\|^2-\theta_k\|y_k-x_k\|^2\\ &+\Big(\frac{\theta(1+\theta)}{1-\delta}-(1+\delta)(\theta^2+\theta)\Big)\|x_k-x_{k-1}\|^2. \end{aligned} \tag{45}$$

Define

$$\begin{aligned}a_k :=\ & \frac{1}{1-\delta}\|x_k-\bar{x}\|^2-\frac{\theta}{1-\delta}\|x_{k-1}-\bar{x}\|^2\\ &+\frac{\theta_{k-1}}{2}\|y_k-y_{k-1}\|^2,\ k\geq 1. \end{aligned}$$

Then (45) becomes

$$\begin{aligned}a_{k+1} \leq\ & a_k+\Big(\theta_k-(1+\delta)(1+\theta)\Big)\|y_{k+1}-x_k\|^2\\ &+\Big(\frac{\theta(1+\theta)}{1-\delta}-(1+\delta)(\theta^2+\theta)\Big)\|x_k-x_{k-1}\|^2-\theta_k\|y_k-x_k\|^2. \end{aligned} \tag{46}$$

Going forward, we examine the next two cases:

Case I: If $-\frac{1}{2}<\theta<0$. Then, since

$$\theta_k \leq (1+\delta)(1+\theta)$$

and

$$\frac{\theta(1+\theta)}{1-\delta} - (1+\delta)(\theta^2+\theta) < 0,$$

a condition that is satisfied when $-\frac{1}{2} < \theta < 0$, we obtain from (46) that $\{a_k\}$ is non-increasing and $\lim_{k\to\infty} a_k$ exists. This implies that $\{a_k\}$ is bounded.

Now, from (46), we obtain

$$\lim_{k\to\infty} \|y_{k+1} - x_k\| = 0 \tag{47}$$

and

$$\lim_{k\to\infty} \|x_k - x_{k-1}\| = 0. \tag{48}$$

From (15) and (48), we obtain

$$\lim_{k\to\infty} \|w_k - x_k\| = 0. \tag{49}$$

From (47) and (49), we obtain

$$\lim_{k\to\infty} \|y_{k+1} - w_k\| = 0. \tag{50}$$

From (25) and (50), we obtain

$$\lim_{k\to\infty} \|y_k - x_k\| = 0. \tag{51}$$

Now, combining (49), (50) and (51) gives

$$\lim_{k\to\infty} \|y_{k+1} - y_k\| = 0. \tag{52}$$

The existence of the limit of $\{a_k\}$ and (52), give

$$\lim_{k\to\infty} \left[\frac{1}{1-\delta}\|x_k - \bar{x}\|^2 - \frac{\theta}{1-\delta}\|x_{k-1} - \bar{x}\|^2\right]$$

exists and hence $\{\frac{1}{1-\delta}\|x_k - \bar{x}\|^2 - \frac{\theta}{1-\delta}\|x_{k-1} - \bar{x}\|^2\}$ is bounded. Noting that

$$\frac{1}{1-\delta}\|x_k - \bar{x}\|^2 \le \frac{1}{1-\delta}\|x_k - \bar{x}\|^2 - \frac{\theta}{1-\delta}\|x_{k-1} - \bar{x}\|^2,$$

we have that $\{x_k\}$ is bounded. Consequently, $\{y_k\}$ is also bounded.

Case II: Assume that $\theta = 0$. Then $w_k = x_k$ in Algorithm 3. Consequently, (15) reduces to

$$\begin{cases} x_k = (1-\delta)y_k + \delta x_{k-1}, \\ y_{k+1} = P_C(x_k - \lambda_k \mathcal{P} y_k), \end{cases}$$

and $\{a_k\}$ becomes

$$a_k = \frac{1}{1-\delta}\|x_k - \bar{x}\|^2 + \frac{\theta_{k-1}}{2}\|y_k - y_{k-1}\|^2, \; k \ge 1$$

and

$$\begin{aligned} a_{k+1} \le\ & a_k + \Big(\theta_k - (1+\delta)\Big)\|y_{k+1} - x_k\|^2 \\ & - \theta_k \|y_k - x_k\|^2. \end{aligned} \tag{53}$$

Hence, one can obtain that $\lim_{k\to\infty} a_k$ exists, $\{a_k\}$, $\{x_k\}$ and are bounded by following similar arguments as in Case 1. □

By leveraging the concepts presented in [29] and exploiting the quasi-monotonicity of $\mathcal{P}$, we derive the following lemma.

Lemma 3.7. *If v^* is a weak cluster point of $\{y_k\}$, then under the fulfillment of Assumption 3.2(a)–(d), we have either $v^* \in S_D$ or $\mathcal{P}v^* = 0$.*

Proof. Based on Lemma 3.6, both $\{x_k\}$ and $\{y_k\}$ are bounded. If v^* is a weak cluster point of $\{y_k\}$, then it is also a weak cluster point of $\{x_k\}$ since $\|x_k - y_k\| \to 0$ as $k \to \infty$. Consequently, a subsequence of $\{x_k\}$, denoted by $\{x_{k_j}\}$, can be selected such that $x_{k_j} \rightharpoonup v^* \in C$.

Now, examine the next two cases.

Case I: If $\limsup_{j\to\infty} \|\mathcal{P} y_{k_j}\| = 0$. Then, $\lim_{j\to\infty} \|\mathcal{P} y_{k_j}\| = \liminf_{j\to\infty} \|\mathcal{P} y_{k_j}\| = 0$. Hence, by Assumption 3.2(c), we arrive at

$$0 < \|\mathcal{P}v^*\| \le \liminf_{j\to\infty} \|\mathcal{P} y_{k_j}\| = 0. \tag{54}$$

Thus, we can infer that $\mathcal{P}v^* = 0$.

Case II: If $\limsup_{j\to\infty}\|\mathcal{P}y_{k_j}\|>0$. Then we select without loss of generality, a subsequence of $\{\mathcal{P}y_{k_j}\}$ still designated as $\{\mathcal{P}y_{k_j}\}$ such that $\lim_{j\to\infty}\|\mathcal{P}y_{k_j}\|=M_1>0$. By (12), we see that, for all $y\in C$,

$$\begin{aligned}0 &\le \langle y_{k_j+1}-x_{k_j}+\lambda_{k_j}\mathcal{P}y_{k_j}, y-y_{k_j+1}\rangle\\ &= \langle y_{k_j+1}-x_{k_j}, y-y_{k_j+1}\rangle+\lambda_{k_j}\langle \mathcal{P}y_{k_j}, y-y_{k_j+1}\rangle\\ &= \langle y_{k_j+1}-x_{k_j}, y-y_{k_j+1}\rangle+\lambda_{k_j}\langle \mathcal{P}y_{k_j}, y_{k_j}-y_{k_j+1}\rangle\\ &\quad+\lambda_{k_j}\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle.\end{aligned} \tag{55}$$

Since $\lim_{k\to\infty}\|y_{k+1}-y_k\|=0=\lim_{k\to\infty}\|y_{k+1}-x_k\|$, we get from (55)

$$0\le\liminf_{j\to\infty}\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle\le\limsup_{j\to\infty}\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle<\infty,\ \forall y\in C. \tag{56}$$

Based on (56), we next analyze two scenarios under **Case II**:

Case 1: For all $y\in C$, let us assume that $\limsup_{j\to\infty}\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle>0$. Then we select a subsequence of $\{y_{k_j}\}$ designated as $\{y_{k_{j_l}}\}$ such that $\lim_{l\to\infty}\langle \mathcal{P}y_{k_{j_l}}, y-y_{k_{j_l}}\rangle>0$. Thus, there exists $l_0\ge1$ such that $\langle \mathcal{P}y_{k_{j_l}}, y-y_{k_{j_l}}\rangle>0,\ \forall l\ge l_0$. Since $\mathcal{P}$ is quasi-monotone, we get $\langle \mathcal{P}y, y-y_{k_{j_l}}\rangle\ge0,\ \forall y\in C, l\ge l_0$. As $l\to\infty$, we obtain $\langle \mathcal{P}y, y-v^*\rangle\ge0,\ \forall y\in C$. That is $v^*\in S_D$.

Case 2: For all $y\in C$, let us assume that $\limsup_{j\to\infty}\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle=0$. Then, we infer from (56) that

$$\lim_{j\to\infty}\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle=0,\ \forall y\in C. \tag{57}$$

This implies

$$\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle+|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}>0,\ \forall y\in C. \tag{58}$$

Again, using $\lim_{j\to\infty}\|\mathcal{P}y_{k_j}\|=M_1>0$, we can find $j_0\ge1$ such that $\|\mathcal{P}y_{k_j}\|>\frac{M_1}{2},\ \forall j\ge j_0$. Hence, we can set $b_{k_j}=\frac{\mathcal{P}y_{k_j}}{\|\mathcal{P}y_{k_j}\|^2},\ \forall j\ge j_0$. Thus, $\langle \mathcal{P}y_{k_j}, b_{k_j}\rangle=1,\ \forall j\ge j_0$. Therefore, by (58), we get

$$\left\langle \mathcal{P}y_{k_j}, y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\right\rangle>0.$$

Combining the above inequality with the assumption that $\mathcal{P}$ is quasi-monotonicity, give

$$\left\langle \mathcal{P}\left(y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]\right), y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\right\rangle\ge0.$$

Thus,

$$\begin{aligned}&\langle \mathcal{P}y, y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\rangle\\ &\ge\langle \mathcal{P}y-\mathcal{P}(y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]), y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\rangle\\ &\ge-\|\mathcal{P}y-\mathcal{P}(y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right])\|\cdot\|y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\|\\ &\ge-L\|b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]\|\cdot\|y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\|\\ &=\frac{-L}{\|\mathcal{P}y_{k_j}\|}\left(|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right)\cdot\|y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\|\\ &\ge\frac{-2L}{M_1}\left(|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right)M_2,\end{aligned} \tag{59}$$

where M_2 is a positive number which exists due to the fact that $\{y+b_{k_j}\left[|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right]-y_{k_j}\}$ is bounded. Looking at (57), we get $\lim_{j\to\infty}\left(|\langle \mathcal{P}y_{k_j}, y-y_{k_j}\rangle|+\frac{1}{j+1}\right)=0$. As $j\to\infty$ in (59), we obtain for all $y\in C$ that $\langle \mathcal{P}y, y-v^*\rangle\ge0$. This gives that $v^*\in S_D$. □

Now, we proceed to establish our weak convergence result, which constitutes one of our key findings.

Theorem 3.8. *Under the fulfillment of Assumption 3.2(a)–(d) and $\mathcal{P}x\neq0, \forall x\in C$, we have that the sequences $\{x_k\}$ and $\{y_k\}$ produced by Algorithm 3 converge weakly.*

Proof. Let the set of weak cluster points of y_k be designated as $w_\omega(y_k)$. Our aim is to establish that

$$w_\omega(y_k)\subset S_D.$$

Let $v^*\in w_\omega(y_k)$. We can find a subsequence $\{y_{k_j}\}\subset\{y_k\}$ for which $y_{k_j}\rightharpoonup v^*$ as $j\to\infty$. By the weakly closedness of C, we arrive at $v^*\in C$. Again, since for all $x\in C$, $\mathcal{P}x$ is nonzero, we see that $\mathcal{P}v^*$ is also nonzero. Utilizing Lemma 3.7, we conclude that $v^*\in S_D$. Thus, $w_\omega(x_n)\subset S_D$.

Define

$$b_k^* = 2\langle x_{k-1} - x_k, x_k - \bar{x}\rangle + \|x_{k-1} - x_k\|^2,$$
$$d_k^* = -\frac{\theta_{k-1}}{2}\|y_k - y_{k-1}\|^2.$$

Observe that

$$\|x_{k-1} - \bar{x}\|^2 = \|x_{k-1} - x_k\|^2 + 2\langle x_{k-1} - x_k, x_k - \bar{x}\rangle + \|x_k - \bar{x}\|^2$$

implies that

$$b_k^* = \|x_{k-1} - \bar{x}\|^2 - \|x_k - \bar{x}\|^2.$$

Consequently,

$$a_k + d_k^* + \frac{\theta}{1-\delta} b_k^* = \frac{(1-\theta)}{1-\delta}\|x_k - \bar{x}\|^2. \tag{60}$$

Since $\lim_{k\to\infty}\|x_{k-1} - x_k\| = 0$, $\lim_{k\to\infty}\|y_{k-1} - y_k\| = 0$, and both $\{x_k\}$ and $\{\theta_k\}$ are bounded, we have that $d_k^* \to 0$, and $b_k^* \to 0$. Since also $\lim_{k\to\infty} a_k$ exists, it follows from (60) that

$$\lim_{k\to\infty}\|x_k - \bar{x}\| \quad \text{exists for each } \bar{x} \in S_D. \tag{61}$$

Moreover, since $\lim_{k\to\infty}\|x_k - \bar{x}\|$ exists, by Lemma 2.3, we deduce that $\{x_k\}$ weakly converges to a point in S_D. Additionally, due to $\|x_k - y_k\| \to 0$ as $k \to \infty$, we see that $\{y_k\}$ weakly converges to a point in S_D. Therefore, the proof is finished. □

Remark 3.9. We give some nonasymptotic $\mathcal{O}(1/k)$ convergence rate for our Algorithm 3. One can show from (46) that (when $-\frac{1}{2} < \theta < 0$)

$$\begin{aligned}\left((1+\delta)(\theta^2+\theta) - \frac{\theta(1+\theta)}{1-\delta}\right)\sum_{j=1}^{k}\|x_j - x_{j-1}\|^2 &\le \sum_{j=1}^{k}(a_j - a_{j+1})\\ &\le a_1 - a_{k+1}.\end{aligned} \tag{62}$$

Therefore,

$$\min_{1\le j\le k}\|x_{j+1} - x_j\|^2 \le \frac{a_1}{\left((1+\delta)(\theta^2+\theta) - \frac{\theta(1+\theta)}{1-\delta}\right)k}, -\frac{1}{2} < \theta < 0.$$

Thus

$$\min_{1\le j\le k}\|x_{j+1} - x_j\|^2 \le \mathcal{O}(1/k), -\frac{1}{2} < \theta < 0.$$

Furthermore, we can show that if $\theta = 0$,

$$\min_{1\le j\le k}\|x_{j+1} - x_j\|^2 \le \mathcal{O}(1/k).$$

That is,

$$\min_{1\le j\le k}\|x_{j+1} - x_j\|^2 \le \mathcal{O}(1/k), -\frac{1}{2} < \theta < 0.$$

Similarly, we can obtain

$$\min_{1\le j\le k}\|y_j - x_j\|^2 \le \mathcal{O}(1/k), -\frac{1}{2} < \theta < 0,$$

and

$$\min_{1\le j\le k}\|y_{j+1} - y_j\|^2 \le \mathcal{O}(1/k), -\frac{1}{2} < \theta < 0.$$

Remark 3.10. Suppose we define the set of trivial solutions to VIP (1) as

$$S_T := \{\tilde{x} \in C : \langle \mathcal{P}\tilde{x}, \tilde{y} - \tilde{x}\rangle = 0, \forall \tilde{y} \in C\}.$$

If $\mathcal{P}\tilde{x} = 0$ in VIP (1), then $\langle \mathcal{P}\tilde{x}, \tilde{y} - \tilde{x}\rangle = 0, \forall \tilde{y} \in C$ and $\tilde{x} \in S_T$ if $\tilde{x} \in C$. As part of our future investigation, we shall consider Theorem 3.8 without assuming the condition "$\mathcal{P}x \neq 0, \forall x \in C$".

In the scenario where $\mathcal{P}$ is η-strongly pseudo-monotone and Lipschitz continuous, we establish the strong convergence result below.

Theorem 3.11. *The sequences $\{x_k\}$ and $\{y_k\}$ produced from Algorithm 3 converge strongly to a unique point in S provided that $\mathcal{P}$ is η-strongly pseudo-monotone and Lipschitz continuous on C.*

Proof. If $\mathcal{P}$ is η-strongly pseudo-monotone and Lipschitz continuous, then VIP (1) admits a unique solution, as demonstrated in, for example, [40]. Let this unique solution be $\bar{x}$. Then, $\bar{x} \in S$ means $\langle \mathcal{P}\bar{x}, y_k - \bar{x}\rangle \geq 0$. Utilizing the η-strongly pseudo-monotonicity of $\mathcal{P}$, gives

$$\langle \mathcal{P}y_k, y_k - \bar{x}\rangle \geq \eta\|y_k - \bar{x}\|^2.$$

Thus,

$$\lambda_k\langle \mathcal{P}y_k, \bar{x} - y_k\rangle \leq -\lambda_k\eta\|y_k - \bar{x}\|^2.$$

As a result, Eq. (34) gives

$$\begin{aligned}
\frac{1}{2}\|y_{k+1} - \bar{x}\|^2 \leq\ & \frac{1}{2}\|w_k - \bar{x}\|^2 - \frac{1}{2}\|w_k - y_{k+1}\|^2 \\
& - \frac{\theta_k}{2}\|x_k - y_k\|^2 - \frac{\theta_k}{2}\|y_{k+1} - y_k\|^2 + \frac{\theta_k}{2}\|x_k - y_{k+1}\|^2 \\
& + \lambda_k\langle \mathcal{P}y_k, \bar{x} - y_k\rangle + \lambda_k\langle \mathcal{P}y_k - \mathcal{P}y_{k-1}, y_k - y_{k+1}\rangle \\
\leq\ & \frac{1}{2}\|w_k - \bar{x}\|^2 - \frac{1}{2}\|w_k - y_{k+1}\|^2 \\
& - \frac{\theta_k}{2}\|x_k - y_k\|^2 - \frac{\theta_k}{2}\|y_{k+1} - y_k\|^2 + \frac{\theta_k}{2}\|x_k - y_{k+1}\|^2 \\
& + \lambda_k\langle \mathcal{P}y_k - \mathcal{P}y_{k-1}, y_k - y_{k+1}\rangle - \lambda_k\eta\|y_k - \bar{x}\|^2.
\end{aligned} \tag{63}$$

Following similar arguments as in (36)–(45), we get

$$\begin{aligned}
& \frac{1}{1-\delta}\|x_{k+1} - \bar{x}\|^2 - \frac{\theta}{1-\delta}\|x_k - \bar{x}\|^2 + \frac{\theta_k}{2}\|y_{k+1} - y_k\|^2 \\
\leq\ & \frac{1}{1-\delta}\|x_k - \bar{x}\|^2 - \frac{\theta}{1-\delta}\|x_{k-1} - \bar{x}\|^2 + \frac{\theta_{k-1}}{2}\|y_k - y_{k-1}\|^2 \\
& + \Big(\theta_k - (1+\delta)(1+\theta)\Big)\|y_{k+1} - x_k\|^2 - \theta_k\|y_k - x_k\|^2 \\
& + \Big(\frac{\theta(1+\theta)}{1-\delta} - (1+\delta)(\theta^2+\theta)\Big)\|x_k - x_{k-1}\|^2 \\
& - \frac{\eta}{2L^2\delta^2\bar{\lambda}}\|y_k - \bar{x}\|^2,
\end{aligned} \tag{64}$$

where the last term follows from (20).

Define

$$\begin{aligned}
a_k :=\ & \frac{1}{1-\delta}\|x_k - \bar{x}\|^2 - \frac{\theta}{1-\delta}\|x_{k-1} - \bar{x}\|^2 \\
& + \frac{\theta_{k-1}}{2}\|y_k - y_{k-1}\|^2,\ k \geq 1.
\end{aligned}$$

Then we obtain from (64) that

$$a_{k+1} \leq a_k - \frac{\eta}{2L^2\delta^2\bar{\lambda}}\|y_k - \bar{x}\|^2. \tag{65}$$

Given that $\lim_{k\to\infty} a_k$ exists, we can infer from (65) that

$$\lim_{k\to\infty}\|y_k - \bar{x}\| = 0.$$

Combining this with (51), gives

$$\lim_{k\to\infty}\|x_k - \bar{x}\| = 0.$$

Therefore the proof is finished. □

4. Numerical experiments

Here, we undertake a comparison between our Algorithm 3 and other methods from the literature, namely, Malitsky Algorithm [32, Algorithm 1] and Yang & Liu Algorithm [36, Algorithm 4.1]. We evaluate their performances using the following test examples.

4.1. Examples in finite dimensions

Example 4.1. Consider the example, also analyzed in [29], where $C = [-1, 1]$ and $\mathcal{P}$ is defined as:

$$\mathcal{P}v = \begin{cases} 2v - 1, & v > 1, \\ v^2, & v \in [-1, 1], \\ -2v - 1, & v < -1. \end{cases}$$

In this case, $\mathcal{P}$ is shown to be quasi-monotone and Lipschitz continuous, yet neither pseudo-monotone nor monotone. Additionally, it is noted that $S_D = \{-1\}$ and $S = \{-1, 0\}$.

Example 4.2 ([29]). Consider the set $C = \{v \in \mathbb{R}^2 : v_1^2 + v_2^2 \leq 1,\ 0 \leq v_1\}$ and the mapping $\mathcal{P}(v_1, v_2) = (-v_1 e^{v_2}, v_2)$. It can be observed that $\mathcal{P}$ is not quasi-monotone, and consequently, $\mathcal{P}$ is not pseudo-monotone. Additionally, it can be verified that $(1,0) \in S_D$ and $S = (1,0), (0,0)$.

Example 4.3. Consider the mapping $\mathcal{P} : \mathbb{R}^m \to \mathbb{R}^m$ defined by:

$$\mathcal{P}v = \left(e^{-v^T Q v} + \beta\right)(Pv + q),$$

where Q, P and β are positive definite matrix ($v^T Q v \geq \theta |v|^2\ \forall v \in \mathbb{R}^m$), positive semi-definite matrix and any positive number, respectively. Also, $q \in \mathbb{R}^m$. It is noteworthy that $\mathcal{P}$ is differentiable, and there exists a positive number M for which $\|\nabla \mathcal{P} v\| \leq M$ for all $v \in \mathbb{R}^m$. Thus, applying the Mean Value Theorem, we see that $\mathcal{P}$ is Lipschitz continuous (as established in [41]). Additionally, $\mathcal{P}$ is shown to be a pseudo-monotone operator which is not monotone.

Take $C := \{v \in \mathbb{R}^m | Bv \leq b\}$, where B is a matrix of size $l^* \times m$ and $b \in \mathbb{R}_+^{l^*}$ with $l^* = 10$.

4.2. Example in infinite dimensional space

For any $0 < T \in \mathbb{R}$, let $L_2([0,T], \mathbb{R}^m)$ represent the Hilbert space of square integrable, measurable vector functions $z : [0,T] \to \mathbb{R}^m$. This space is equipped with an inner product defined as:

$$\langle z, v\rangle = \int_0^T \langle z(t), v(t)\rangle dt,$$

and the corresponding norm:

$$\|z\|_2 = \sqrt{\langle z, z\rangle} < \infty.$$

Let us examine the optimal control problem

$$z^*(t) = \operatorname{argmin}\{f(z) : z \in Z\} \tag{66}$$

under the assumption that such a control exists. In this case, Z represents the set of admissible controls:

$$Z = \left\{z(t) \in L_2([0,T], \mathbb{R}^m) : z_i(t) \in [z_i^-, z_i^+], i = 1, 2, \ldots, m\right\}.$$

Thus, Z takes the shape of an m-dimensional box and it constitutes the piecewise continuous functions. In particular, the control can be of the bang–bang type, characterized by a piecewise constant function.

The terminal objective can be expressed as:

$$f(z) = g(w(T)),$$

where g is a differentiable and convex function which is defined on the attainability set.

Consider a scenario where the trajectory $w(t) \in L_2([0,T])$ satisfies constraints represented by a system of differential equations:

$$\dot{w}(t) = D(t)w(t) + E(t)z(t), \quad w(0) = w_0, \quad t \in [0,T],$$

where $D(t) \in \mathbb{R}^{n\times n}$ and $E(t) \in \mathbb{R}^{n\times m}$ are continuous matrices given for each $t \in [0,T]$. According to the Pontryagin maximum principle, we can find a function $p^* \in L_2([0,T])$ for which (w', p', z') solves, for almost every $t \in [0,T]$, the system of equations below:

$$\begin{cases} \dot{w}^*(t) = D(t)w^*(t) + E(t)z^*(t) \\ w^*(0) = w_0, \end{cases} \tag{67}$$

$$\begin{cases} \dot{p}^*(t) = -D(t)^\top p^*(t) \\ p^*(T) = \nabla g(w(T)), \end{cases} \tag{68}$$

$$0 \in E(t)^\top p^*(t) + N_Z(z^*(t)), \tag{69}$$

where $N_Z(z)$ denotes the normal cone to Z at z, given as:

$$N_Z(z) := \begin{cases} \emptyset & \text{if } z \notin Z \\ \{\ell \in H : \langle \ell, v - z\rangle \leq 0,\ \forall v \in Z\} & \text{if } z \in Z. \end{cases}$$

Denoting $Gz(t) := E(t)^\top p(t)$, it is well-established that Gz represents the gradient of the objective cost function f [42–44]. We can transform (69) into the following VIP: find $z \in Z$ such that

$$\langle Gz, v - z\rangle \geq 0,\ \forall v \in Z. \tag{70}$$

Let us discretize the continuous functions by selecting a natural number N with a *mesh size* $h := T/N$. We can represent any discretized control $z^N := (z_0, z_1, \dots, z_{N-1})$ as its piece-wise constant extension:

$$z^N(t) = z_i \text{ for } t \in [t_i, t_{i+1}), \ i = 0, 1, \dots, N.$$

Additionally, we will denote any discretized state $w^N := (w_0, w_1, \dots, w_N)$ with its piece-wise linear interpolation

$$w^N(t) = w_i + \frac{t - t_i}{h}\left(w_{i+1} - w_i\right), \text{ for } t \in [t_i, t_{i+1}), \ i = 0, 1, \dots, N-1.$$

We also denote the discretized co-state variable as $p^N := (p_0, p_1, \dots, p_N)$.

Numerical ordinary differential equations (ODEs) offer various discretization techniques, including Euler and Runge–Kutta methods [45]. In this study, we opt for the Euler method. This entails solving the system of ODEs (67) and (68) at each iteration using the Euler method.

$$\begin{cases} w_{i+1}^N = w_i^N + h\left[D(t_i)w_i^N + E(t_i)z_i^N\right] \\ w(0) = w_0, \end{cases} \tag{71}$$

$$\begin{cases} p_i^N = p_{i+1}^N + hD(t_i)^T p_{i+1}^N \\ p(N) = \nabla g(w_N). \end{cases} \tag{72}$$

The Euler discretization is well-known for its error estimate of $O(h)$ [46,47]. This indicates that the discrepancy between the discretized solution $z^N(t)$ and the original solution $z^*(t)$ is proportional to the mesh size h. In other words, we can find a positive constant C for which

$$\|z^N - z^*\| \le Ch.$$

The subsequent example is adapted from [48, Example 7].

Example 4.4 (*Control of a Harmonic Oscillator*).

$$\begin{array}{ll} \text{minimize} & w_2(3\pi) \\ \text{subject to} & \dot{w}_1(t) = w_2(t), \\ & \dot{w}_2(t) = -w_1(t) + z(t), \ \forall t \in [0, 3\pi], \\ & w(0) = 0, \\ & z(t) \in [-1, 1]. \end{array} \tag{73}$$

We know the precise optimal control of this problem to be:

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

Let us explore examples where the terminal function is not linear. One such instance is the following Rocket Car problem described in [46,49].

Example 4.5 (*Rocket Car*).

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\left(\left(w_1(5)\right)^2 + \left(w_2(5)\right)^2\right) \\ \text{subject to} & \dot{w}_1(t) = w_2(t), \\ & \dot{w}_2(t) = z(t), \ \forall t \in [0, 5], \\ & w_1(0) = 6, \ w_2(0) = 1, \\ & z(t) \in [-1, 1]. \end{array} \tag{74}$$

The exact optimal control is

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

where $\tau = 3.5174292$.

The following example is taken from [50].

Example 4.6 (*See [50, Example 6.3]*).

$$\begin{array}{ll} \text{minimize} & -w_1(2) + \left(w_2(2)\right)^2 \\ \text{subject to} & \dot{w}_1(t) = w_2(t), \\ & \dot{w}_2(t) = z(t), \ \forall t \in [0, 2], \\ & w_1(0) = 0, \ w_2(0) = 0, \\ & z(t) \in [-1, 1]. \end{array} \tag{75}$$

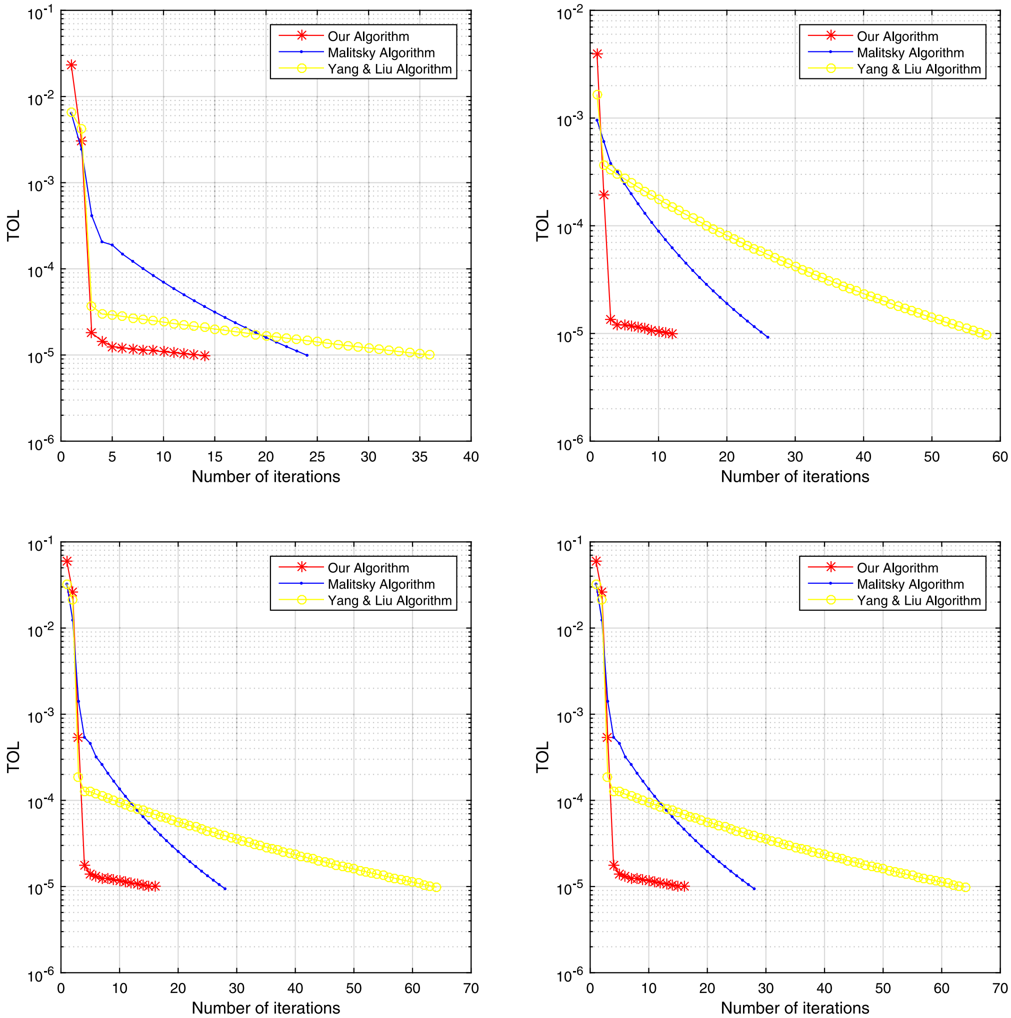

Fig. 1. The performance of TOL_n for Example 4.1 with $\epsilon = 10^{-5}$ is illustrated as follows: Top Left: **Case 1**; Top Right: **Case 2**; Bottom left: **Case 3**; Bottom Right: **Case 4**.

The exact optimal control is

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

During the computation, we make use of the following:

- Algorithm 3 (Our Algorithm): $\delta = 0.9, \theta = -0.1, \lambda_0 = 2, \bar{\lambda} = 2$.
- Algorithm 1 (Malitsky Algorithm): $\phi = \varphi = \frac{\sqrt{5}+1}{2}$, $\lambda_0 = 2, \bar{\lambda} = 2$.
- Algorithm 2 (Yang & Liu Algorithm): $\delta = 0.5\left(\frac{\sqrt{1+4(\frac{\alpha}{2-\theta}+1-\alpha)}-1}{2} + 1\right), \theta = 1, \alpha = 0.5, \mu = 0.5, \lambda_1 = 2$.

We then use the stopping criterion; $\mathrm{TOL}_n := \max\{\|y_{k+1} - x_k\|^2, \|y_k - x_k\|^2\} < \epsilon$ for all algorithms, where ϵ is the predetermined error. Furthermore, we choose y_0, y_1, x_0 and w_0 as follows:

For Example 4.1: Case 1: $y_1 = 0.2, y_0 = 0.1, x_0 = 0.1, w_0 = 0.03$; Case 2: $y_1 = 0.1, y_0 = 0.2, x_0 = 0.15, w_0 = 0.03$; Case 3: $y_1 = 0.3, y_0 = -0.1, x_0 = 0.15, w_0 = 0.03$ and Case 4: $y_1 = 0.3, y_0 = 0.3, x_0 = 0.15, w_0 = 0.03$.

For Example 4.2: Case 1: $y_1 = (0.1, 0.5)$, $y_0 = (0.2, 0.1)$, $x_0 = (0.2, 0.1)$, $w_0 = (0.8, 0.01)$; Case 2: $y_1 = (0.1, 0.2)$, $y_0 = (0.1, 0.2)$, $x_0 = (0.2, 0.1)$, $w_0 = (0.8, 0.01)$; Case 3: $y_1 = (0.3, 0.6)$, $y_0 = (0.5, 0.5)$, $x_0 = (0.2, 0.1)$, $w_0 = (0.8, 0.01)$ and Case 4: $y_1 = (0.5, 0.5)$, $y_0 = (0.3, 0.6)$, $x_0 = (0.2, 0.1)$, $w_0 = (0.8, 0.01)$.

For Example 4.3: $m \in \{4, 10, 20, 30\}$ while $y_0 = (1, 1, \ldots, 1)^T$, $x_0 = w_0 = (0, 0, \ldots, 0)^T$ and y_1 is generated randomly in $\mathbb{R}^m$.

For Examples 4.4–4.6: we select $N = 100$ and randomly select the initial controls from the interval $[-1, 1]$.

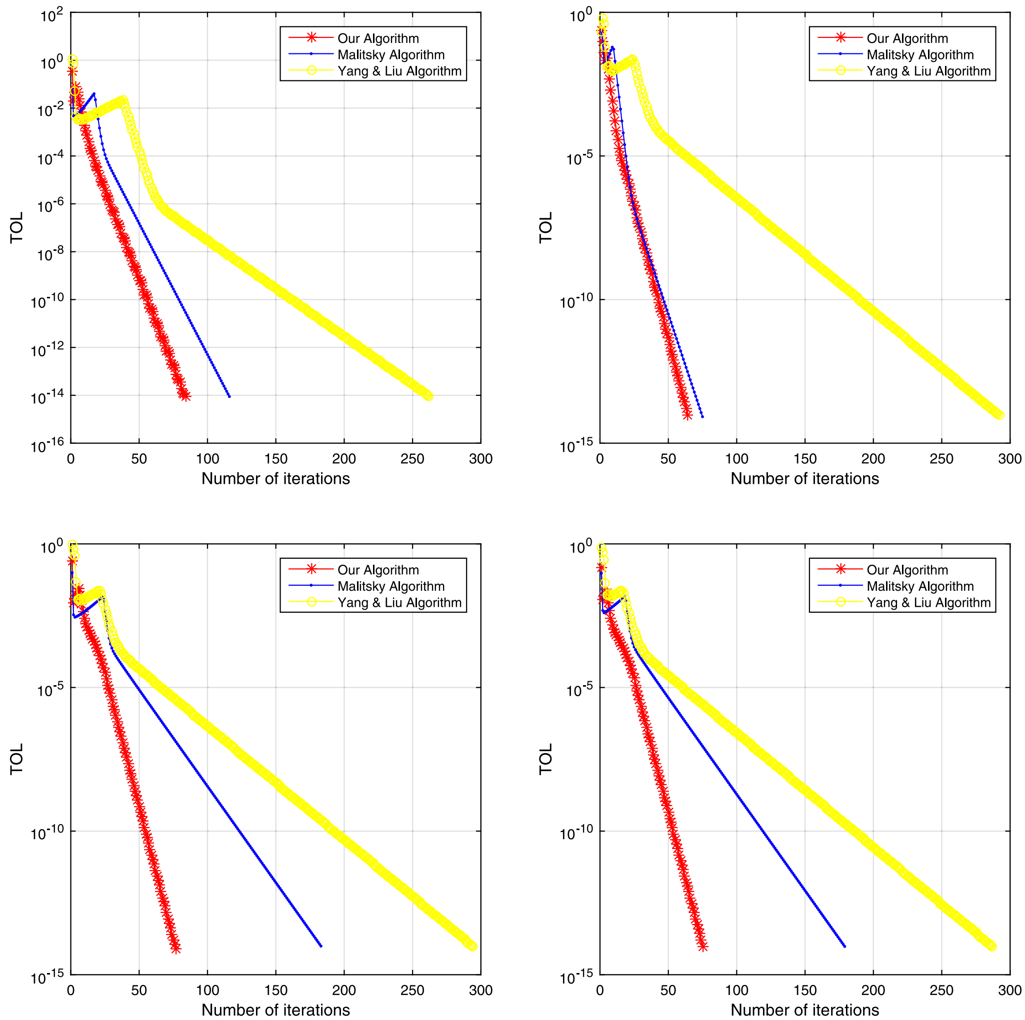

Fig. 2. The performance of TOL_n for Example 4.2 with $\epsilon = 10^{-14}$ is illustrated as follows: Top Left: **Case 1**; Top Right: **Case 2**; Bottom left: **Case 3**; Bottom Right: **Case 4**.

Table 1
Example 4.1: Comparative analysis of algorithms with a tolerance level $\epsilon = 10^{-5}$.

Algorithms	Case 1		Case 2		Case 3		Case 4	
	CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter
Our Algorithm	0.0068	14	0.0064	12	0.0072	16	0.0064	16
Malitsky Algorithm	0.0169	24	0.0160	26	0.0186	28	0.0163	28
Yang & Liu Algorithm	0.1104	36	0.1118	58	0.1155	64	0.1126	64

Table 2
Example 4.2: Comparative analysis of algorithms with a tolerance level $\epsilon = 10^{-14}$.

Algorithms	Case 1		Case 2		Case 3		Case 4	
	CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter
Our Algorithm	0.0114	84	0.0117	64	0.0126	77	0.0114	75
Malitsky Algorithm	0.1134	116	0.1114	75	0.1124	183	0.1135	179
Yang & Liu Algorithm	0.3306	262	1.0244	292	1.0282	294	1.0274	287

The calculations were conducted on a personal computer featuring an Intel(R) Core(TM) i5-2600 CPU clocked at 2.30 GHz and 8.00 GB of RAM, using Matlab 2016 (b). Tables 1–4 list the number of iterations under "Iter" and the CPU time in seconds under "CPU". (see Figs. 1–7).

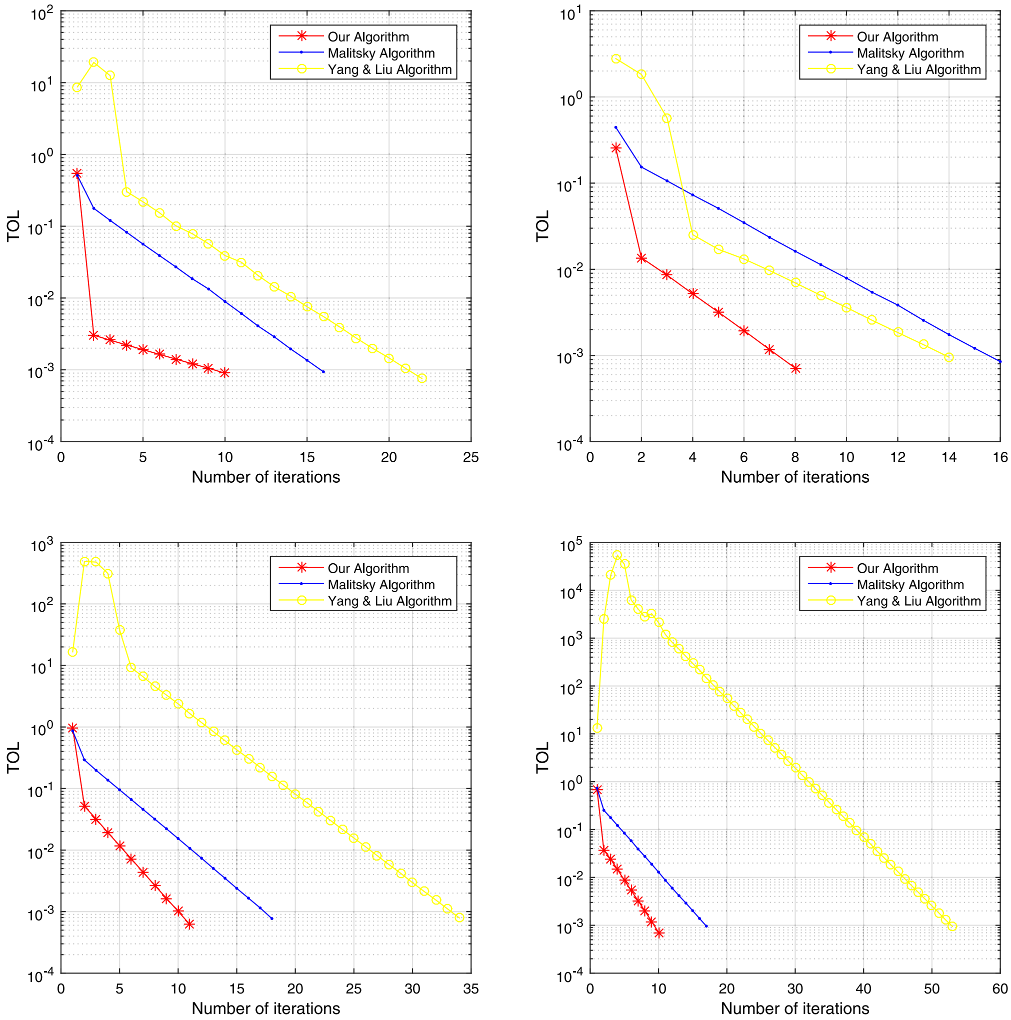

Fig. 3. The performance of TOL_n for Example 4.3 with $\epsilon = 10^{-3}$ is illustrated as follows: Top Left: $m = 4$; Top Right: $m = 10$; Bottom left: $m = 20$; Bottom Right: $m = 30$.

Table 3
Example 4.3: Comparative analysis of algorithms with a tolerance level $\epsilon = 10^{-3}$.

Algorithms	$m = 4$		$m = 10$		$m = 20$		$m = 30$	
	CPU	Iter	CPU	Iter	CPU	Iter	CPU	Iter
Our Algorithm	0.0213	10	0.0138	8	0.0192	11	0.0137	10
Malitsky Algorithm	0.1269	16	0.1135	16	0.1138	18	0.1142	17
Yang & Liu Algorithm	1.0318	22	1.0176	14	1.0204	34	1.0250	53

Table 4
Comparative analysis of algorithms for Examples 4.4–4.6 with a tolerance level $\epsilon = 10^{-4}$.

Algorithms	Example 4.4		Example 4.5		Example 4.6	
	CPU	Iter	CPU	Iter	CPU	Iter
Our Algorithm	0.1315	75	0.0566	29	0.2269	129
Malitsky Algorithm	1.2092	138	1.8203	501	1.8963	496
Yang & Liu Algorithm	1.5233	262	5.0218	1170	4.5964	954

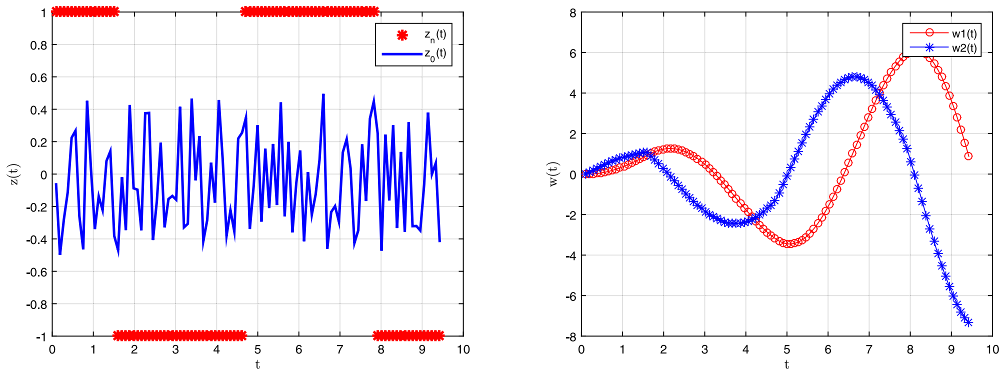

Fig. 4. Example 4.4 computed by our Algorithm (Algorithm 3): On the left: The blue line represents the random initial control, while the red line represents the optimal control. On the right: Displays the optimal trajectories.

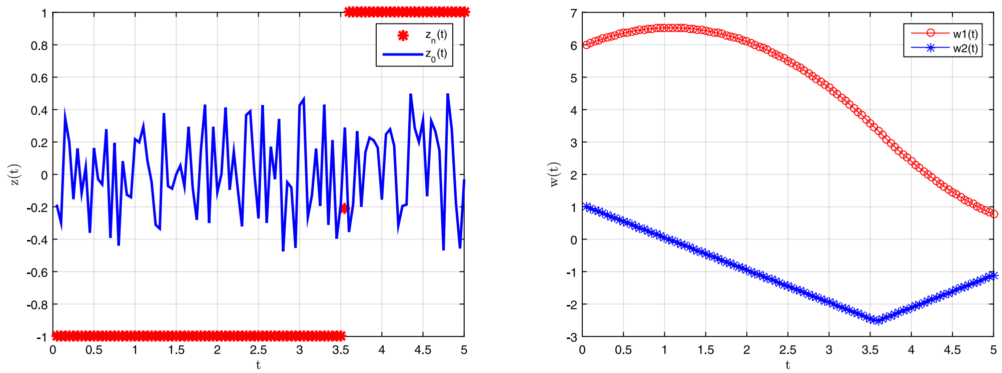

Fig. 5. Example 4.5 computed by our Algorithm (Algorithm 3): On the left: The blue line represents the random initial control, while the red line represents the optimal control. On the right: Displays the optimal trajectories.

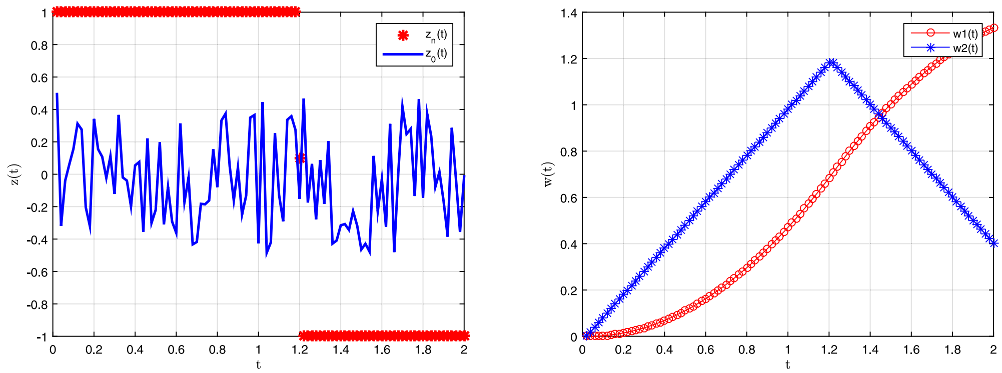

Fig. 6. Example 4.6 computed by our Algorithm (Algorithm 3): On the left: The blue line represents the random initial control, while the red line represents the optimal control. On the right: Displays the optimal trajectories.

5. Conclusion

We have developed a fully adaptive Golden Ratio Algorithm with a backward inertial step, demonstrating weak convergence for variational inequalities when dealing with a locally Lipschitz continuous quasi-monotone operator. Additionally, we have established strong convergence when dealing with Lipschitz continuous strongly pseudo-monotone operators. A distinguishing feature of our method is its flexibility regarding stepsizes, as it does not mandate them to be nonincreasing. Furthermore, our approach demands only one operator evaluation and one projection computation per iteration during numerical implementations, streamlining the computational process. Notably, our method encompasses the Golden Ratio Algorithm as a special case. Numerical experiments

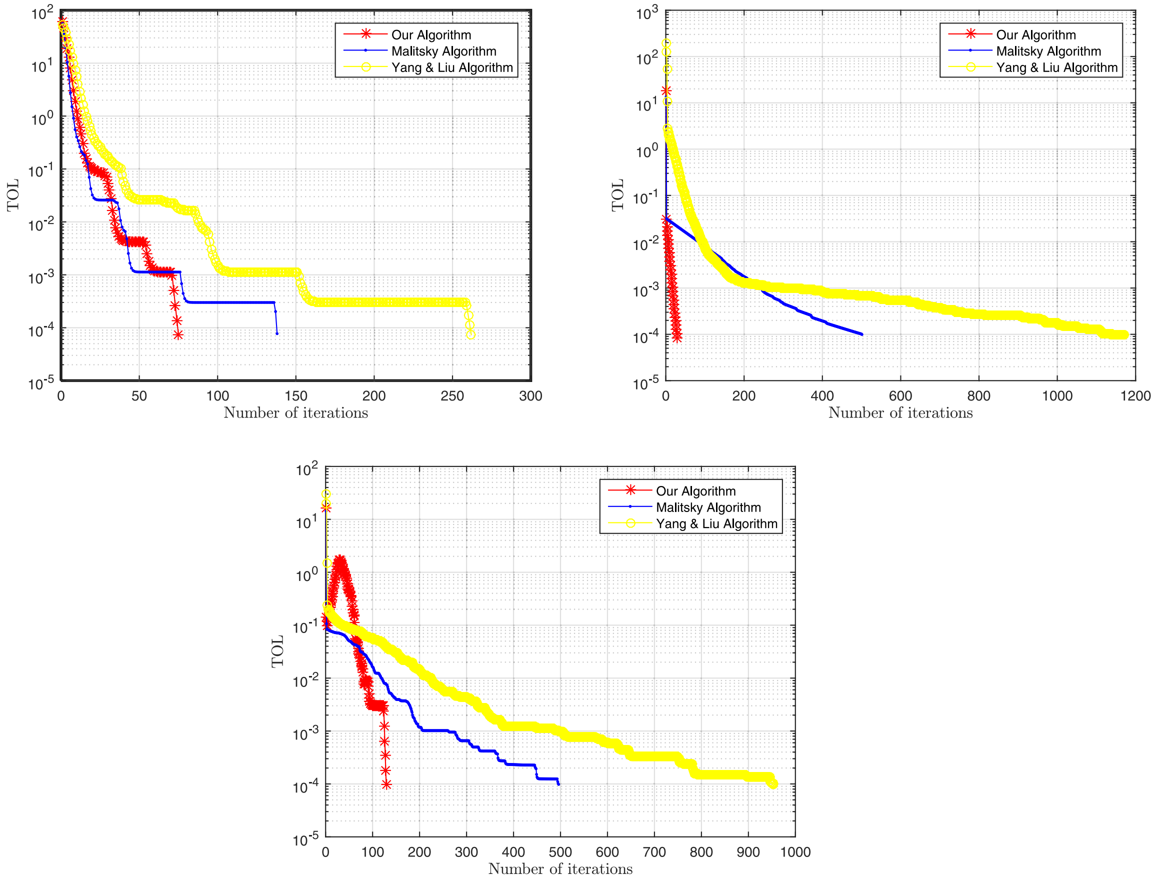

Fig. 7. The behavior of TOL_n: Top Left: Example 4.4; Top Right: Example 4.5; Bottom: Example 4.6.

showcased the competitiveness of our algorithm compared to existing methods in the literature. In the future, we aim to explore the application of our algorithm in solving nonconvex optimization problems, thus extending its utility and relevance.

Declarations

Ethical approval and consent to participate

All the authors of this paper have given their ethical approval and provided their consent to participate in the research presented herein.

Consent for publication

All authors have granted permission to publish identifiable information in both the journal and the article.

Code availability

The MATLAB scripts used for running the numerical simulations can be provided upon request.

Availability of supporting data

This study does not involve the generation or analysis of datasets; therefore, there is no data to share.

Funding

Not Applicable.

CRediT authorship contribution statement

Chinedu Izuchukwu: Writing – review & editing, Visualization, Validation, Software, Methodology, Data curation. **Yekini Shehu:** Writing – original draft, Supervision, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

We are extremely grateful to the two anonymous referees for their suggestions and comments which have improved on the earlier version of our manuscript.

References

[1] Baiocchi C, Capelo A. Variational and quasivariational inequalities; Applications to free boundary problems. New York: Wiley; 1984.
[2] Fichera G. Sul pproblem elastostatico di signorini con ambigue condizioni al contorno. Atti Accad Naz Lincei Rend Cl Sci Fis Mat Natur 1963;34:138–42.
[3] Kinderlehrer D, Stampacchia G. An introduction to variational inequalities and their applications. New York: Academic Press; 1980.
[4] Stampacchia G. Variational inequalities. In: Theory and Applications of Monotone Operators, Proceedings of the NATO Advanced Study Institute. Venice, Italy , Edizioni Odersi, Gubbio, Italy; 1968, p. 102–92.
[5] Korpelevich GM. An extragradient method for finding sadlle points and for other problems. Ekon Mat Metody 1976;12:747–56.
[6] Noor M. Extragradient methods for pseudomonotone variational inequalities. J Optim Theory Appl 2003;117:475–88.
[7] Thong DV, Yang J, Cho YJ, Rassias TM. Explicit extragradient-like method with adaptive stepsizes for pseudomonotone variational inequalities. Optim Lett 2021;15:2181–99.
[8] Vuong PT. On the weak convergence of the extragradient method for solving pseudomonotone variational inequalities. J Optim Theory Appl 2018;176:399–409.
[9] Popov LD. A modification of the Arrow-Hurwicz method for finding saddle points. Math Notes 1980;28:845–8.
[10] Malitsky YV, Semenov VV. An extragradient algorithm for monotone variational inequalities. Cybernet Systems Anal 2014;50:271–7.
[11] Gibali A, Thong DV. A new low-cost double projection method for solving variational inequalities. Optim Eng 2020;21:1613–34.
[12] Censor Y, Gibali A, Reich S. Extensions of Korpelevich's extragradient method for the variational inequality problem in euclidean space. Optimization 2011;61:1119–32.
[13] Censor Y, Gibali A, Reich S. The subgradient extragradient method for solving variational inequalities in Hilbert space. J Optim Theory Appl 2011;148:318–35.
[14] Censor Y, Gibali A, Reich S. Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space. Optim Meth Softw 2011;26:827–45.
[15] Kraikaew R, Saejung S. Strong convergence of the halpern subgradient extragradient method for solving variational inequalities in Hilbert spaces. J Optim Theory Appl 2014;163:399–412.
[16] Yang J. Self-adaptive inertial subgradient extragradient algorithm for solving pseudomonotone variational inequalities. Appl Anal 2021;100:1067–78.
[17] Tseng P. A modified forward–backward splitting method for maximal monotone mappings. SIAM J Control Optim 2000;38:431–46.
[18] Chang X, Liu S, Deng Z, Li S. An inertial subgradient extragradient algorithm with adaptive stepsizes for variational inequality problems. Optim Meth Soft 2022;37:1507–26.
[19] Thong DV, Hieu DV. Inertial extragradient algorithms for strongly pseudomonotone variational inequalities. J Comput Appl Math 2018;341:80–98.
[20] Thong DV, Van Hieu D, Rassias TM. Self adaptive inertial subgradient extragradient algorithms for solving pseudomonotone variational inequality problems. Optim Lett 2020;14:115–44.
[21] Thong DV, Li XH, Dong QL, Cho YJ, Rassias TM. An inertial Popov's method for solving pseudomonotone variational inequalities. Optim Lett 2021;15:757–77.
[22] Adamu A, Kumam P, Kitkuan D, Padcharoen A. Relaxed modified Tseng algorithm for solving variational inclusion problems in real Banach spaces with applications. Carpathian J Math 2023;39:01–26.
[23] Wang Z-b, Sunthrayuth P, Adamu A, Cholamjiak P. Modified accelerated Bregman projection methods for solving quasi-monotone variational inequalities. Optimization 2023. http://dx.doi.org/10.1080/02331934.2023.2187663.
[24] Tang G-j, Wan Z, Huang N-J. Strong convergence of a projection-type method for mixed variational inequalities in Hilbert spaces. Numer Funct Anal Optim 2018;39:1103–19.
[25] Wang Z-b, Chen Z-y, Xiao Y-b, Zhang C. A new projection-type method for solving multi-valued mixed variational inequalities without monotonicity. Appl Anal 2020;99:1453–66.
[26] Ye M, He Y. A double projection method for solving variational inequalities without monotonicity. Comput Optim Appl 2015;60(1):141–50.
[27] Yu Y, Zhao Y, Yin TC. Convergence of extragradient-type methods for fixed point problems and quasimonotone variational inequalities. J Nonlinear Convex Anal 2023;24:2225–37.
[28] Yu Y, Yin TC. Strong convergence theorems for a nonmonotone equilibrium problem and a quasi-variational inclusion problem. J Nonlinear Convex Anal 2024;25:503–12.
[29] Liu H, Yang J. Weak convergence of iterative methods for solving quasimonotone variational inequalities. Comput Optim Appl 2020;77:491–508.
[30] Salahuddin. The extragradient method for quasi-monotone variational inequalities. Optimization 2022;71:2519–28.
[31] Wang Z-b, Chen X, Yi J, Chen Z-y. Inertial projection and contraction algorithms with larger step sizes for solving quasimonotone variational inequalities. J Global Optim 2022;82:499–522.
[32] Malitsky Y. Golden ratio algorithms for variational inequalities. Math Program 2020;184:383–410.

[33] Malitsky Y. Projected reflected gradient methods for monotone variational inequalities. SIAM J Optim 2015;25:502–20.
[34] Malitsky Y, Tam MK. A forward–backward splitting method for monotone inclusions without cocoercivity. SIAM J Optim 2020;30:1451–72.
[35] Mainge PE, Gobinddass ML. Convergence of one-step projected gradient methods for variational inequalities. J Optim Theory Appl 2016;171:146–68.
[36] Yang J, Liu H. A self-adaptive method for pseudomonotone equilibrium problems and variational inequalities. Comput Optim Appl 2020;75:423–40.
[37] Chang X, Bai J. A projected extrapolated gradient method with larger step size for monotone variational inequalities. J Optim Theory Appl 2021;190:602–27.
[38] Yang J, Liu H, Liu Z. Modified subgradient extragradient algorithms for solving monotone variational inequalities. Optimization 2018;67:2247–58.
[39] Opial Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull Amer Math Soc 1967;73:591–7.
[40] Kim DS, Vuong PT, Khanh PD. Qualitative properties of strongly pseudomonotone variational inequalities. Optim Lett 2016;10:1669–79.
[41] Boţ RI, Csetnek ER, Vuong PT. The forward–backward-forward method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces. European J Oper Res 2020;287:49–60.
[42] Khoroshilova EV. Extragradient-type method for optimal control problem with linear constraints and convex objective function. Optim Lett 2013;7:1193–214.
[43] Vuong PT, Shehu Y. Convergence of an extragradient-type method for variational inequality with applications to optimal control problems. Numer Algorithms 2019;81:269–91.
[44] Nikol'skii MS. Convergence of the gradient projection method in optimal control problems. Comp Math Model 2007;18:148–56.
[45] Butcher JC. Numerical methods for ordinary differential equations. John Wiley; 2003.
[46] Alt W, Baier R, Gerdts M, Lempio F. Error bounds for Euler approximation of linear-quadratic control problems with bang–bang solutions. Numer Algebra Control Optim 2012;2:547–70.
[47] Bonnans JF, Festa A. Error estimates for the Euler discretization of an optimal control problem with first-order state constraints. SIAM J Numer Anal 2017;55:445–71.
[48] Pietrus A, Scarinci T, Veliov VM. High order discrete approximations to Mayer's problems for linear systems. SIAM J Control Optim 2018;56:102–19.
[49] Preininger J, Vuong PT. On the convergence of the gradient projection method for convex optimal control problems with bang–bang solutions. Comput Optim Appl 2018;70:221–38.
[50] Bressan B, Piccoli B. Introduction to the mathematical theory of control. AIMS Ser Appl Math 2007.