A Strongly Convergent Projection and Contraction Algorithm with Extrapolations from the Past

Yekini Shehu* Jianchao Bai†

Abstract

This paper introduces a projection and contraction algorithm with extrapolation from the past for solving variational inequalities in Hilbert spaces. The proposed method reduces the number of operator evaluations per iteration from two to one by reusing past iterations. Under standard assumptions, we prove strong convergence of the generated sequence. Numerical experiments on solving large-scale Harker-Pang problems demonstrate its superior efficiency and robustness compared to several well-established methods.

Keywords: extrapolation from the past, variational inequality, strong convergence, projection and contraction algorithm

AMS subject classifications. 65K10, 65Y20, 90C25

1 Introduction

Given a nonempty closed convex subset $\mathcal{C}$ of the real Hilbert space $\mathbb{H}$ (equipped with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$) and a continuous operator $F : \mathbb{H} \to \mathbb{H}$, the variational inequality problem is to find a $x \in \mathcal{C}$ such that

$$\langle Fx, z - x \rangle \geq 0, \quad \forall z \in \mathcal{C}. \tag{1.1}$$

It is common knowledge from the references [1, 2, 15, 40, 21, 38] that problems arising from economics, engineering mechanics, mathematical programming, transportation, and other applied sciences can be rewritten in the form of (1.1). The problem (1.1) is also equivalent to $\min_{x \in \mathcal{C}} f(x)$, where $f : \mathbb{H} \to (-\infty, +\infty)$ is a continuously differentiable convex function whose gradient is denoted by $\nabla f = F$. Throughout this paper, the solution set of (1.1), denoted by S, is assumed to be nonempty.

A popular approach to solving (1.1) is the projection-type algorithm that generally enjoys weak convergence in terms of the resulting sequence to a solution obtained with various conditions imposed on F. Some of which include the following extragradient algorithm [22]:

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n F x_n), \\ x_{n+1} = P_{\mathcal{C}}(x_n - \lambda_n F y_n), \end{cases} \tag{1.2}$$

with $\lambda_n > 0$ being the step size and $P_{\mathcal{C}}(x)$ being the *metric projection* of x onto $\mathcal{C}$, Popov extragradient algorithm [35], subgradient extragradient algorithm [9], forward-backward-forward

*School of Mathematical Sciences, Zhejiang Normal University, Jinhua 321004, China (yekini.shehu@zjnu.edu.cn).

†Corresponding author. School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China; Center for Applied Mathematics of Fujian Province, Fuzhou University, Fuzhou 350116, China (jianchaobai@nwpu.edu.cn).

algorithm [42, 4], projection and contraction method [19], projected reflected gradient algorithm [30], forward-reflected-backward algorithm [28] and the golden ratio algorithm [29]. These methods, except the last three, feature two evaluations of F at each iteration, resulting in a situation that may affect their efficiency and optimal numerical performance. The forward-backward-forward algorithm, designed by [42], reads

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n F x_n), \\ x_{n+1} = y_n + \lambda_n \left(F x_n - F y_n \right). \end{cases} \tag{1.3}$$

Clearly, it still needs two evaluations Fx_n and Fy_n at the nth iteration.

Recently, reference [16] introduced the concept *extrapolation from the past*, which is to "store and re-use the extrapolated gradient for the extrapolation", and then presented a modified version of (1.2):

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda F y_{n-1}), \\ x_{n+1} = P_{\mathcal{C}}(x_n - \lambda F y_n). \end{cases} \tag{1.4}$$

The above scheme was proved to be convergent for any $\lambda \in (0, 1/\sqrt{3}L]$, where L is the Lipschitz constant of F. Later, more and more researchers focused on developing various algorithms by using the concept extrapolation from the past. For example, by combining the first update of (1.4) and the second update of (1.3), Bot, et al. [7] introduced the following weakly convergent variant of the forward-backward-forward method for any $\lambda_n \leq 1/(2L)$:

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n F y_{n-1}), \\ x_{n+1} = y_n + \lambda_n \left(F y_{n-1} - F y_n \right). \end{cases} \tag{1.5}$$

It is worth noting that the above stepsize λ_n still depends on the Lipschitz constant L, but the Lipschitz constant of a mapping is often unknown or difficult to approximate. Recently, a novel dual inertial Tseng's extragradient method was proposed in [32], which does not need the prior knowledge of Lipschitz constant and any line-search procedure. The projection algorithm with extrapolation in [34] and the inertial forward-backward-forward algorithm in [48] not only used the concept extrapolation from the past, but also provided similar adaptive updates for the stepsize λ_n. Tongnoi further studied the convergence of (1.5) for solving variational inequality problem (1.1) in [43] and monotone inclusion problems in [44]. As observed in [7, 43], one can obtain the forward-reflected-backward algorithm [28]:

$$y_{n+1} = P_{\mathcal{C}}(y_n - 2\lambda F y_n + \lambda F y_{n-1}), \tag{1.6}$$

when x_{n+1} is substituted into the first step of (1.5) at y_{n+1} with constant step size. If $\mathcal{C} = \mathbb{H}$, then both (1.5) and (1.6) become

$$y_{n+1} = y_n - 2\lambda F y_n + \lambda F y_{n-1}.$$

Another classical approach to solving (1.1) is the following Projection and Contraction Method (PCM, [14]), that is, a modification of the extragradient algorithm (1.2):

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n F x_n), \\ d_n = x_n - y_n - \lambda_n (F x_n - F y_n), \\ x_{n+1} = x_n - \gamma \beta_n d_n, \end{cases} \tag{1.7}$$

where $\gamma \in (0, 2), \lambda_n \in (0, 1/L)$ and

$$\beta_n = \begin{cases} \frac{\langle x_n - y_n, d_n \rangle}{\|d_n\|^2}, & d_n \neq 0, \\ 0, & d_n = 0. \end{cases} \tag{1.8}$$

This PCM iteratively updates $\{x_n\}$ using the projection step y_n and the contraction step with a descent direction d_n, which makes the PCM (1.7) tolerates errors in the contraction step better than the forward-backward-forward algorithm (1.3). Just like the forward-backward-forward algorithm (1.3), the PCM (1.7) improves on the computational complexity of the extragradient algorithm (1.2) in terms of $P_{\mathcal{C}}$. Nonetheless, it still requires two values of F at each iteration, which can be computationally expensive if F is complex. Therefore, it would be of interest to further reduce the computational complexity. Thus, we need to investigate "the projection and contraction method (1.7) with extrapolation from the past" and obtain strong convergence results. It should be noted that different authors have extensively studied the PCM-type method; see, for example, [8, 10, 12, 13, 31, 32, 33, 37, 39, 41, 49]. For other types of methods related to PCM for solving (1.1) and its reformulation problem, we refer to the alternating direction method of multipliers [20, 24, 5], the operator splitting methods [17, 25], and the proximal point algorithm-based descent method [23].

Several works (see, for example, [12, 39, 41, 48]) have been done on the strongly convergent version of the scheme (1.7) as follows:

$$\begin{cases} y_n = P_{\mathcal{C}}(x_n - \lambda_n F x_n), \\ d_n = x_n - y_n - \lambda_n (F x_n - F y_n), \\ x_{n+1} = \alpha_n x_0 + (1-\alpha_n)(x_n - \gamma \beta_n d_n), \end{cases} \tag{1.9}$$

where $\{\alpha_n\} \subset (0,1)$ and β_n is given in (1.8). However, *the setbacks of two evaluations of F per iteration evident in* (1.7) *are still in place in strongly convergent* (1.9).

Recently, along the line of projection-type methods, Anh [45] introduced a new relaxed projection approach for solving variational inequality problems in Hilbert spaces by constructing a solution mapping with strongly quasi-nonexpansive properties. Subsequently, Thanh and Anh [46] developed Halpern relaxed projection methods that achieve strong convergence for partially pseudomonotone and Lipschitz continuous operators. Very recently, Trang et al. [47] proposed viscosity-projection methods combining viscosity approximation with projection steps for solving variational inequalities, with successful applications to adaptive image restoration. While these methods have enlarged the toolkit of projection algorithms, they either rely on relaxed projection mappings that require additional solution-map evaluations, or employ Halpern/viscosity averaging that needs two operator evaluations per iteration. In contrast, the present work focuses on integrating extrapolation from the past into a projection and contraction framework, thereby reducing the per-iteration cost to a single operator evaluation while maintaining strong convergence.

Our Contributions. Building on the forward-backward-forward method with past extrapolation (1.5) and projection-contraction methods (1.7) and (1.9), we propose a strongly convergent projection and contraction algorithm with extrapolation from the past. Our contributions are:

- **Strong convergence analysis:** We establish strong convergence under standard conditions. The algorithm generalizes method (1.5) when $\alpha_n = 0$ and $\beta_n \equiv 0$, offering greater flexibility.
- **Adaptive parameter rule:** We provide an adaptive updating rule for the stepsize parameter λ_n, relaxing reliance on prior knowledge of the Lipschitz constant.
- **Numerical validation:** We conduct extensive comparative experiments on large-scale variational inequality problems to illustrate the practical performance of the algorithm.

Outline. Preliminary definitions and lemmas are presented in Section 2. Our main results for the proposed new algorithm are discussed in Section 3, where we provide strong convergence results and certain connections between Algorithm 3.1 and an inertial forward-backward-forward algorithm. Section 4 conducts parameter tuning experiments and comparison experiments to evaluate the performance of our algorithm. We conclude with a summary of what we have done and an overview of future plans in Section 5.

2 Preliminaries

The following definitions are prepared before showing several technical lemmas.

Definition 2.1 *Let $F : \mathbb{H} \to \mathbb{H}$ be an operator. We say that F is*

(a) monotone *on $\mathbb{H}$ if $\langle Fu - Fv, u - v\rangle \geq 0$ for all $u, v \in \mathbb{H}$;*

(b) η-strongly pseudo-monotone *on $\mathbb{H}$ if there exists $\eta > 0$ such that $\langle Fv, u - v\rangle \geq 0 \Rightarrow \langle Fu, u - v\rangle \geq \eta\|u - v\|^2$ for all $u, v \in \mathbb{H}$;*

(c) pseudo-monotone *on $\mathbb{H}$ if for each $u, v \in \mathbb{H}$, $\langle Fu, v - u\rangle \geq 0 \Rightarrow \langle Fv, v - u\rangle \geq 0$;*

(d) L-Lipschitz continuous *on $\mathbb{H}$ if there exists a constant $L > 0$ such that $\|Fu - Fv\| \leq L\|u - v\|$ for all $x, y \in \mathbb{H}$.*

Given $x \in \mathbb{H}$, there exists a unique point $P_{\mathcal{C}}(x) \in \mathcal{C}$ (cf. [3]) such that

$$\|x - P_{\mathcal{C}}(x)\| \leq \|x - v\|, \qquad \forall v \in \mathcal{C}.$$

It is known that $P_{\mathcal{C}}$ satisfies

$$\langle u - v, P_{\mathcal{C}}(u) - P_{\mathcal{C}}(v)\rangle \geq \|P_{\mathcal{C}}(u) - P_{\mathcal{C}}(v)\|^2, \qquad \forall u, v \in \mathbb{H}, \tag{2.1}$$

and

$$\langle u - P_{\mathcal{C}}(u), P_{\mathcal{C}}(u) - v\rangle \geq 0 \qquad \forall v \in \mathcal{C}. \tag{2.2}$$

Thus,

$$\|u - v\|^2 \geq \|u - P_{\mathcal{C}}(u)\|^2 + \|v - P_{\mathcal{C}}(v)\|^2, \qquad \forall u \in \mathbb{H}, v \in \mathcal{C}. \tag{2.3}$$

Lemma 2.1 *Let $u, v \in \mathbb{H}$. Then, we have*
(i) $2\langle u, v\rangle = \|u\|^2 + \|v\|^2 - \|u - v\|^2 = \|u + v\|^2 - \|u\|^2 - \|v\|^2$;
(ii) $\|u - v\|^2 + \|u + v\|^2 = 2\|u\|^2 + 2\|v\|^2$.

Lemma 2.2 *[27, Lemma 3.1] Suppose $\{t_n\}$ and $\{r_n\}$ are sequences of non-negative real numbers such that*

$$t_{n+1} \leq (1 - \alpha_n)t_n + s_n + r_n,$$

where $\alpha_n \in (0, 1)$ and $s_n \in \mathbb{R}$. If $\sum_{n=1}^{\infty} r_n < \infty$ and $s_n \leq \alpha_n M$ for some $M \geq 0$. Then, $\{t_n\}$ is bounded.

Lemma 2.3 *[36] Suppose $\{a_n\}$ is a sequence of nonnegative real numbers, $\{c_n\}$ is a sequence in $(0, 1)$ with $\sum_{n=1}^{\infty} c_n = \infty$, and $\{b_n\}$ is a sequence of real numbers. Assume that*

$$a_{n+1} \leq (1 - c_n)a_n + c_n b_n.$$

If $\limsup_{k\to\infty} b_{n_k} \leq 0$ for every subsequence $\{a_{n_k}\}$ of $\{a_n\}$ satisfying $\liminf_{k\to\infty}(a_{n_k+1} - a_{n_k}) \geq 0$, then $\lim_{n\to\infty} a_n = 0$.

Lemma 2.4 *[11] Suppose $\mathcal{C}$ is a nonempty, closed, convex subset of a real Hilbert space $\mathbb{H}$ and $F : \mathcal{C} \to \mathbb{H}$ is pseudo-monotone and continuous. Then, $\widehat{x}$ is a solution to the variational inequality problem* (1.1) *if and only if*

$$\langle Fx, x - \widehat{x}\rangle \geq 0, \qquad \widehat{x} \in \mathcal{C}.$$

3 Main Results

This section presents a new projection and contraction algorithm with extrapolation from the past, along with its strong convergence results.

3.1 The Proposed Algorithm

Before presenting our proposed algorithm, which is motivated by the projection and contraction algorithm (1.7), we state a basic assumption as follows. If F is monotone and L-Lipschitz continuous, then Assumption 3.1 holds. Also, Assumption 3.1 is fulfilled when $\mathbb{H}$ is finite-dimensional, and F is a continuous pseudo-monotone operator.

Assumption 3.1 *Let $F : \mathcal{C} \to \mathbb{H}$ be a pseudo-monotone operator which is also L-Lipschitz continuous ($L > 0$) satisfying*

$$\text{if } y_n \rightharpoonup y \text{ and } \liminf_{n\to\infty} \|Fy_n\| = 0, \text{ then } Fy = 0.$$

Now, we present our strongly convergent algorithm, called PCM-ep, as follows:

Algorithm 3.1 *(This algorithm attempts to solve problem* (1.1)*)*

Initialization. *Choose parameters $\mu \in \left(0, \sqrt{\frac{\epsilon}{2(2\epsilon+1)}}\right)$ and $\gamma \in \left(0, \frac{2}{2+\epsilon}\right)$ for any $\epsilon > 0$. Let $\alpha_n \in [0, 1)$ and $\tau_n \in (0, \infty)$ be a sequence such that $\sum_{n=0}^{\infty} \tau_n << +\infty$ and initialize $x_0, y_{-1} \in \mathbb{H}, \lambda_0 > 0$. Set $n = 0$.*

Iterative Step. *Given λ_n, x_n and y_{n-1}, compute*

$$\begin{cases} w_n = \alpha_n x_0 + (1 - \alpha_n) x_n, \\ y_n = P_{\mathcal{C}}(w_n - \lambda_n F y_{n-1}), \\ u_n = y_n + \lambda_n (F y_{n-1} - F y_n), \\ x_{n+1} = u_n - \gamma \beta_n d_n, \end{cases} \tag{3.1}$$

where

$$d_n = w_n - y_n - \lambda_n (F y_{n-1} - F y_n), \quad \beta_n = \begin{cases} \dfrac{\max\{\langle u_n - y_n, d_n \rangle, 0\}}{\|d_n\|^2}, & d_n \neq 0, \\ 0, & d_n = 0. \end{cases}$$

Update

$$\lambda_{n+1} = \begin{cases} \min\left\{ \frac{\mu \|y_{n-1} - y_n\|}{\|F y_{n-1} - F y_n\|}, \lambda_n + \tau_n \right\}, & F y_{n-1} \neq F y_n, \\ \lambda_n + \tau_n, & \textit{otherwise}. \end{cases} \tag{3.2}$$

Remark 3.1 *Here, we give some remarks about Algorithm 3.1.*

(i) The contraction step for the update x_{n+1} features u_n and a descent direction d_n instead of x_n and a descent direction d_n presented in the previous algorithms (1.7) *and* (1.9)*.*

(ii) At the current iteration, the two values of F, i.e., Fx_n and Fy_n, in the projection and contraction algorithm (1.7) *are replaced with Fy_{n-1} and Fy_n in the proposed Algorithm 3.1. Note that Fy_{n-1} is previously known from the previous iteration and is being recalled in the current iteration. Therefore, only Fy_n is computed at the current iteration in Algorithm 3.1.*

(iii) Our proposed Algorithm 3.1 improves on the strongly convergence projection and contraction algorithm (1.9) *by reducing the two values of F in* (1.9) *to just only one value of F at each iteration.*

Lemma 3.1 *Assume that F is L-Lipschitz continuous on $\mathbb{H}$. Let $\{\lambda_n\}$ be the sequence generated by* (3.2)*. Then*

$$\lim_{n\to\infty} \lambda_n = \lambda \in \Big[\min\Big\{\lambda_0, \frac{\mu}{L}\Big\}, \lambda_0 + \tau\Big],$$

where $\tau = \sum_{n=1}^{\infty} \tau_n$.

Proof. Since F is Lipschitz continuous with constant $L > 0$, in the case that $Fy_{n-1} \neq Fy_n$, we obtain

$$\frac{\mu\|y_{n-1} - y_n\|}{\|Fy_{n-1} - Fy_n\|} \geq \frac{\mu\|y_{n-1} - y_n\|}{L\|y_{n-1} - y_n\|} = \frac{\mu}{L}. \tag{3.3}$$

By the definition of λ_{n+1} and mathematical induction, the sequence $\{\lambda_n\}$ has upper bound $\lambda_0 + \tau$ and lower bound $\min\{\lambda_0, \frac{\mu}{L}\}$. Following a similar line to the proof of Lemma 3.1 [26], we obtain the desired result. □

3.2 Strong Convergence Analysis

In this section, we present the strong convergence analysis of the sequence generated by our proposed Algorithm 3.1. We first get from (3.2) that

$$\|Fy_{n-1} - Fy_n\| \leq \frac{\mu}{\lambda_{n+1}}\|y_{n-1} - y_n\|. \tag{3.4}$$

Lemma 3.2 *Let Assumption 3.1 be fulfilled. Then, for any $x^* \in \mathcal{S}$, the sequences $\{u_n\}$ and $\{y_n\}$ generated by Algorithm 3.1 satisfy*

$$\|u_n - x^*\|^2 \leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 + \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2. \tag{3.5}$$

Proof. By the update of $y_n = P_{\mathcal{C}}(w_n - \lambda_n Fy_{n-1})$ and the properties of $P_{\mathcal{C}}$, we obtain

$$\langle y_n - w_n + \lambda_n Fy_{n-1}, x^* - y_n\rangle \geq 0. \tag{3.6}$$

The pseudomonotonicity of F implies

$$\langle \lambda_n Fy_n, y_n - x^*\rangle \geq 0. \tag{3.7}$$

Combine (3.7) and (3.6) together with the update $u_n = y_n + \lambda_n(Fy_{n-1} - Fy_n)$ to obtain

$$\langle u_n - w_n, x^* - y_n\rangle \geq 0.$$

Simple algebra shows

$$\begin{aligned}
& \langle u_n - w_n, u_n - x^*\rangle \leq \langle u_n - w_n, u_n - y_n\rangle \\
= \; & \langle u_n - w_n, u_n - w_n\rangle + \langle u_n - w_n, w_n - y_n\rangle \\
= \; & \|u_n - w_n\|^2 + \langle u_n - w_n, w_n - y_n\rangle \\
= \; & \|u_n - w_n\|^2 + \langle y_n + \lambda_n(Fy_{n-1} - Fy_n) - w_n, w_n - y_n\rangle \\
= \; & \|u_n - w_n\|^2 + \langle y_n - w_n, w_n - y_n\rangle + \langle \lambda_n(Fy_{n-1} - Fy_n), w_n - y_n\rangle \\
= \; & \|u_n - w_n\|^2 - \|w_n - y_n\|^2 + \lambda_n\langle w_n - y_n, Fy_{n-1} - Fy_n\rangle.
\end{aligned}$$

According to the first item of Lemma 2.1, we have

$$2\langle u_n - w_n, u_n - x^*\rangle = \|u_n - w_n\|^2 - \|w_n - x^*\|^2 + \|u_n - x^*\|^2.$$

Substituting this identity into (3.8), it follows that

$$\|u_n - x^*\|^2 \le \|w_n - x^*\|^2 + \|u_n - w_n\|^2 - 2\|w_n - y_n\|^2 + 2\lambda_n\langle w_n - y_n, Fy_{n-1} - Fy_n\rangle. \tag{3.8}$$

Now, combine the update of u_n and (3.4) to have

$$\begin{aligned}\|u_n - w_n\|^2 &= \|y_n + \lambda_n(Fy_{n-1} - Fy_n) - w_n\|^2 \\ &= \|y_n - w_n\|^2 + 2\lambda_n\langle y_n - w_n, Fy_{n-1} - Fy_n\rangle + \lambda_n^2\|Fy_{n-1} - Fy_n\|^2 \\ &\le \|y_n - w_n\|^2 + 2\lambda_n\langle y_n - w_n, Fy_{n-1} - Fy_n\rangle + \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2.\end{aligned}$$

Substituting the last relationship into (3.8) results in

$$\begin{aligned}\|u_n - x^*\|^2 \le\ & \|w_n - x^*\|^2 + \|y_n - w_n\|^2 + 2\lambda_n\langle y_n - w_n, Fy_{n-1} - Fy_n\rangle \\ & + \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2 - 2\|w_n - y_n\|^2 + 2\lambda_n\langle w_n - y_n, Fy_{n-1} - Fy_n\rangle \\ =\ & \|w_n - x^*\|^2 - \|w_n - y_n\|^2 + \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2.\end{aligned}$$

The proof is completed. □

Theorem 3.1 *Let Assumption 3.1 be fulfilled and $\{x_n\}$ be generated by Algorithm 3.1. For any $x^* \in \mathcal{S}$, we have*

$$\|x_{n+1} - x^*\|^2 \le \|u_n - x^*\|^2 - \frac{(2-\gamma)}{\gamma}\|x_{n+1} - u_n\|^2. \tag{3.9}$$

Proof. First of all, the way of generating x_{n+1} implies

$$\|x_{n+1} - x^*\|^2 = \|(u_n - x^*) - \gamma\beta_n d_n\|^2 = \|u_n - x^*\|^2 - 2\gamma\beta_n\langle u_n - x^*, d_n\rangle + \gamma^2\beta_n^2\|d_n\|^2.$$

Since $y_n = P_{\mathcal{C}}(w_n - \lambda_n Fy_{n-1})$, we have by (2.2) that

$$\langle y_n - x^*, w_n - y_n - \lambda_n Fy_{n-1}\rangle \ge 0. \tag{3.10}$$

Adding the previous (3.7) to (3.10) together with the definition of d_n leads to

$$\langle y_n - x^*, w_n - y_n - \lambda_n Fy_{n-1} + \lambda_n Fy_n\rangle = \langle y_n - x^*, d_n\rangle \ge 0,$$

which makes $\langle u_n - x^*, d_n\rangle = \langle u_n - y_n, d_n\rangle + \langle y_n - x^*, d_n\rangle$ become

$$\langle u_n - x^*, d_n\rangle \ge \langle u_n - y_n, d_n\rangle.$$

By the update of β_n in Algorithm 3.1, we have

$$\beta_n\|d_n\|^2 = \max\{\langle u_n - y_n, d_n\rangle, 0\} \tag{3.11}$$

(Thus $\beta_n \ge 0$ always holds), and hence

$$\begin{aligned}\|x_{n+1} - x^*\|^2 &= \|u_n - x^*\|^2 - 2\gamma\beta_n\langle u_n - x^*, d_n\rangle + \gamma^2\beta_n^2\|d_n\|^2 \\ &\le \|u_n - x^*\|^2 - 2\gamma\beta_n\langle u_n - y_n, d_n\rangle + \gamma^2\beta_n\langle u_n - y_n, d_n\rangle \\ &= \|u_n - x^*\|^2 - \gamma(2-\gamma)\beta_n\langle u_n - y_n, d_n\rangle. \end{aligned} \tag{3.12}$$

Noting that we have from (3.11) and the update of x_{n+1} that

$$\beta_n\langle u_n - y_n, d_n\rangle = \|\beta_n d_n\|^2 = \frac{1}{\gamma^2}\|x_{n+1} - u_n\|^2.$$

So, plugging the last equality into (3.12) ensures the conclusion (3.9). □

The last two lemmas can not directly ensure the convergence of $\{x_n\}$ and $\{y_n\}$. However, their convergence can be proved with the aid of the following potential sequence

$$a_n := \|x_n - x^*\|^2 + (1+\epsilon)\|x_n - u_{n-1}\|^2 + \Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2, \tag{3.13}$$

where $\epsilon > 0$ and $x^* \in \mathcal{S}$.

Lemma 3.3 *Let Assumption 3.1 be fulfilled and a_n be defined in* (3.13)*. Suppose $\alpha_n \in [0, 1)$ with $\lim\limits_{n\to\infty} \alpha_n = 0$. Then, for any $\epsilon > 0$, it holds*

$$\begin{aligned} a_{n+1} \quad \leq \quad & (1-\alpha_n)a_n + \alpha_n\|x_0 - x^*\|^2 - \Big(\frac{(2-\gamma)}{\gamma} - (1+\epsilon)\Big)\|x_{n+1} - u_n\|^2 \\ & - \left[\frac{(1-\alpha_n)}{2} - \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2} - \Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\right]\|y_n - y_{n-1}\|^2. \end{aligned} \tag{3.14}$$

In addition, both $\{x_n\}$ and $\{y_n\}$ are bounded.

Proof. Plug (3.5) into (3.9) to obtain

$$\|x_{n+1} - x^*\|^2 \leq \|w_n - x^*\|^2 - \|w_n - y_n\|^2 + \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2 - \frac{(2-\gamma)}{\gamma}\|x_{n+1} - u_n\|^2. \tag{3.15}$$

By the definition of w_n, the first term in the right-hand-side of (3.15) amounts to

$$\begin{aligned} \|w_n - x^*\|^2 &= \|\alpha_n x_0 + (1-\alpha_n)x_n - x^*\|^2 \\ &= \|(x_n - x^*) - \alpha_n(x_n - x_0)\|^2 \\ &= \|x_n - x^*\|^2 + \alpha_n^2\|x_n - x_0\|^2 - 2\alpha_n\langle x_n - x^*, x_n - x_0\rangle \\ &= \|x_n - x^*\|^2 + \alpha_n^2\|x_n - x_0\|^2 - \alpha_n\|x_n - x^*\|^2 - \alpha_n\|x_n - x_0\|^2 + \alpha_n\|x_0 - x^*\|^2, \end{aligned}$$

where the last equality uses Lemma 2.1 (i). Replacing x^* with y_n in the last equality gives

$$\|w_n - y_n\|^2 = \|x_n - y_n\|^2 + \alpha_n^2\|x_n - x_0\|^2 - \alpha_n\|x_n - y_n\|^2 - \alpha_n\|x_n - x_0\|^2 + \alpha_n\|x_0 - y_n\|^2.$$

From the last two equations, we have

$$\begin{aligned} & \|w_n - x^*\|^2 - \|w_n - y_n\|^2 \\ = \; & (1-\alpha_n)\|x_n - x^*\|^2 + \alpha_n^2\|x_n - x_0\|^2 - \alpha_n\|x_n - x_0\|^2 + \alpha_n\|x_0 - x^*\|^2 \\ & - (1-\alpha_n)\|x_n - y_n\|^2 - \alpha_n^2\|x_n - x_0\|^2 + \alpha_n\|x_n - x_0\|^2 - \alpha_n\|x_0 - y_n\|^2 \\ = \; & (1-\alpha_n)\|x_n - x^*\|^2 - (1-\alpha_n)\|x_n - y_n\|^2 + \alpha_n\|x_0 - x^*\|^2 - \alpha_n\|x_0 - y_n\|^2. \end{aligned}$$

Substituting this equality into (3.15) with a relaxation gives

$$\begin{aligned} \|x_{n+1} - x^*\|^2 \quad \leq \quad & (1-\alpha_n)\|x_n - x^*\|^2 - (1-\alpha_n)\|x_n - y_n\|^2 \\ & + \alpha_n\|x_0 - x^*\|^2 + \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2 - \frac{(2-\gamma)}{\gamma}\|x_{n+1} - u_n\|^2. \end{aligned} \tag{3.16}$$

Notice that,

$$\begin{aligned}\|x_n - y_n\|^2 &\geq -\|x_n - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2 \\ &= -\|x_n - u_{n-1} + u_{n-1} - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2 \\ &= -\|(x_n - u_{n-1}) + \lambda_{n-1}(Fy_{n-2} - Fy_{n-1})\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2,\end{aligned}$$

where the last equality follows from the update $u_{n-1} = y_{n-1} + \lambda_{n-1}F(y_{n-2} - y_{n-1})$. Now, for any $\epsilon > 0$, we expand the squared norm

$$\begin{aligned}\|(x_n - u_{n-1}) + \lambda_{n-1}(Fy_{n-2} - Fy_{n-1})\|^2 &= \lambda_{n-1}^2\|Fy_{n-2} - Fy_{n-1}\|^2 \\ &+ \|x_n - u_{n-1}\|^2 + 2\big\langle x_n - u_{n-1}, \lambda_{n-1}(Fy_{n-2} - Fy_{n-1})\big\rangle \\ &\leq \|x_n - u_{n-1}\|^2 + \lambda_{n-1}^2\|Fy_{n-2} - Fy_{n-1}\|^2 + 2\lambda_{n-1}\|x_n - u_{n-1}\|\|Fy_{n-2} - Fy_{n-1}\| \\ &\leq \|x_n - u_{n-1}\|^2 + \frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 + 2\frac{\lambda_{n-1}\mu}{\lambda_n}\|x_n - u_{n-1}\|\|y_{n-2} - y_{n-1}\| \\ &\leq \|x_n - u_{n-1}\|^2 + \frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 + \epsilon\|x_n - u_{n-1}\|^2 + \frac{1}{\epsilon}\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 \\ &= (1+\epsilon)\|x_n - u_{n-1}\|^2 + \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2,\end{aligned}$$

where the first inequality uses the Cauchy-Schwarz inequality, the second inequality follows from (3.4), and the last inequality uses Young's inequality. Substituting this upper bound back into the first inequality yields the final estimate

$$\|x_n - y_n\|^2 \geq -(1+\epsilon)\|x_n - u_{n-1}\|^2 - \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 + \frac{1}{2}\|y_n - y_{n-1}\|^2. \tag{3.17}$$

Now, plug (3.17) into (3.16) to obtain

$$\begin{aligned}\|x_{n+1} - x^*\|^2 \leq (1-\alpha_n)\Big[&\|x_n - x^*\|^2 + (1+\epsilon)\|x_n - u_{n-1}\|^2 \\ &+ \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2\Big] + \alpha_n\|x_0 - x^*\|^2 \\ &- \Big[\frac{(1-\alpha_n)}{2} - \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2}\Big]\|y_n - y_{n-1}\|^2 - \frac{(2-\gamma)}{\gamma}\|x_{n+1} - u_n\|^2.\end{aligned}$$

Thus,

$$\begin{aligned}&\|x_{n+1} - x^*\|^2 + (1+\epsilon)\|x_{n+1} - u_n\|^2 + \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_n - y_{n-1}\|^2 \\ &\leq (1-\alpha_n)\Big[\|x_n - x^*\|^2 + (1+\epsilon)\|x_n - u_{n-1}\|^2 + \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2\Big] + \alpha_n\|x_0 - x^*\|^2 \\ &- \Big[\frac{(1-\alpha_n)}{2} - \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2} - \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\Big]\|y_n - y_{n-1}\|^2 - \Big(\frac{(2-\gamma)}{\gamma} - (1+\epsilon)\Big)\|x_{n+1} - u_n\|^2,\end{aligned}$$

which, by the definition of a_n in (3.13), confirms the inequality in (3.14).

Observe that

$$\liminf_{n\to\infty}\Big[\frac{(1-\alpha_n)}{2} - \frac{\mu^2\lambda_n^2}{\lambda_{n+1}^2} - \Big(1 + \frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\Big] = \frac{1}{2} - \Big(2 + \frac{1}{\epsilon}\Big)\mu^2 > 0, \tag{3.18}$$

since $\mu \in \left(0, \sqrt{\frac{\epsilon}{2(2\epsilon+1)}}\right)$ and $\alpha_n \in [0,1)$. Also, since $0 < \gamma < \frac{2}{2+\epsilon}$, we have $\frac{(2-\gamma)}{\gamma} - (1+\epsilon) > 0$. Consequently, we have from Lemma 2.2 and (3.14) that $\{a_n\}$ is bounded. Therefore, both $\{x_n\}$ and $\{y_n\}$ are bounded. □

Theorem 3.2 *Assume that Assumption 3.1 is fulfilled and $\{(x_n, y_n)\}$ is generated by Algorithm 3.1. Suppose $\lim_{n\to\infty} \alpha_n = 0$ and $\sum_{n=0}^{\infty} \alpha_n = \infty$. Then, the sequences $\{x_n\}$ and $\{y_n\}$ converge strongly to an element $x^* \in \mathcal{S}$, where $x^* = P_{\mathcal{S}}(x_0)$.*

Proof. It follows from the update of w_n in (3.1) that

$$\begin{aligned} \|w_n - x^*\|^2 &= \|\alpha_n(x_0 - x^*) + (1-\alpha_n)(x_n - x^*)\|^2 \\ &= \alpha_n^2\|x_0 - x^*\|^2 + (1-\alpha_n)^2\|x_n - x^*\|^2 + 2\alpha_n(1-\alpha_n)\langle x_0 - x^*, x_n - x^*\rangle \end{aligned} \tag{3.19}$$

and

$$\begin{aligned} \|w_n - y_n\|^2 &= \|\alpha_n(x_0 - y_n) + (1-\alpha_n)(x_n - y_n)\|^2 \\ &= \alpha_n^2\|x_0 - y_n\|^2 + (1-\alpha_n)^2\|x_n - y_n\|^2 + 2\alpha_n(1-\alpha_n)\langle x_0 - y_n, x_n - y_n\rangle \\ &\geq \alpha_n^2\|x_0 - y_n\|^2 + (1-\alpha_n)^2\|x_n - y_n\|^2 - 2\alpha_n(1-\alpha_n)\|x_0 - y_n\|\|x_n - y_n\| \\ &\geq \alpha_n^2\|x_0 - y_n\|^2 + (1-\alpha_n)^2\|x_n - y_n\|^2 - 2\alpha_n(1-\alpha_n)M\|x_n - y_n\| \end{aligned} \tag{3.20}$$

where $M := \sup_{n\geq 0}\|x_0 - y_n\| < \infty$ since $\{y_n\}$ is bounded. Consequently, we have from (3.19) and (3.20) together with (3.17) that

$$\begin{aligned} \|w_n - x^*\|^2 &- \|w_n - y_n\|^2 \leq \alpha_n^2\|x_0 - x^*\|^2 + (1-\alpha_n)^2\|x_n - x^*\|^2 \\ &+2\alpha_n(1-\alpha_n)\langle x_0 - x^*, x_n - x^*\rangle - \alpha_n^2\|y_n - x_0\|^2 \\ &+(1-\alpha_n)^2\Big((1+\epsilon)\|x_n - u_{n-1}\|^2 + \Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 \\ &-\frac{1}{2}\|y_n - y_{n-1}\|^2\Big) + 2\alpha_n(1-\alpha_n)M\|x_n - y_n\| \\ \leq\ & \alpha_n^2\|x_0 - x^*\|^2 + (1-\alpha_n)^2\|x_n - x^*\|^2 + 2\alpha_n(1-\alpha_n)\langle x_0 - x^*, x_n - x^*\rangle \\ &+(1-\alpha_n)^2\Big((1+\epsilon)\|x_n - u_{n-1}\|^2 + \Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\phi_1^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 \\ &-\frac{1}{2}\|y_n - y_{n-1}\|^2\Big) + 2\alpha_n(1-\alpha_n)M\|x_n - y_n\|. \end{aligned} \tag{3.21}$$

Plug (3.21) into (3.5), we have

$$\begin{aligned} \|u_n - x^*\|^2 \leq\ & \alpha_n^2\|x_0 - x^*\|^2 + (1-\alpha_n)^2\|x_n - x^*\|^2 + 2\alpha_n(1-\alpha_n)\langle x_0 - x^*, x_n - x^*\rangle \\ &+(1-\alpha_n)^2\Big((1+\epsilon)\|x_n - u_{n-1}\|^2 + \Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2} - y_{n-1}\|^2 \\ &-\frac{1}{2}\|y_n - y_{n-1}\|^2\Big) + 2\alpha_n(1-\alpha_n)M\|x_n - y_n\| + \frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1} - y_n\|^2. \end{aligned} \tag{3.22}$$

Therefore, it follows from Theorem 3.1 that

$$\begin{aligned}
\|x_{n+1}-x^*\|^2 &\le \alpha_n^2\|x_0-x^*\|^2+(1-\alpha_n)^2\|x_n-x^*\|^2+2\alpha_n(1-\alpha_n)\langle x_0-x^*,x_n-x^*\rangle\\
&\quad+(1-\alpha_n)^2\Big((1+\epsilon)\|x_n-u_{n-1}\|^2+\Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2}-y_{n-1}\|^2\\
&\quad-\frac{1}{2}\|y_n-y_{n-1}\|^2\Big)+2\alpha_n(1-\alpha_n)M\|x_n-y_n\|+\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1}-y_n\|^2-\frac{(2-\gamma)}{\gamma}\|x_{n+1}-u_n\|^2\\
&\le(1-\alpha_n)\Big[\|x_n-x^*\|^2+(1+\epsilon)\|x_n-u_{n-1}\|^2+\Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2}-y_{n-1}\|^2\Big]\\
&\quad+\alpha_n\Big[\alpha_n\|x_0-x^*\|^2+2(1-\alpha_n)\langle x_0-x^*,x_n-x^*\rangle+2(1-\alpha_n)M\|x_n-y_n\|\Big]\\
&\quad-\Big(\frac{(1-\alpha_n)^2}{2}-\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\Big)\|y_n-y_{n-1}\|^2-\frac{(2-\gamma)}{\gamma}\|x_{n+1}-u_n\|^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\|x_{n+1}-x^*\|^2+(1+\epsilon)\|x_{n+1}-u_n\|^2+\Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1}-y_n\|^2\\
&\quad\le(1-\alpha_n)\Big[\|x_n-x^*\|^2+(1+\epsilon)\|x_n-u_{n-1}\|^2+\Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_{n-1}^2\mu^2}{\lambda_n^2}\|y_{n-2}-y_{n-1}\|^2\Big]\\
&\quad+\alpha_n\Big[\alpha_n\|x_0-x^*\|^2+2(1-\alpha_n)\langle x_0-x^*,x_n-x^*\rangle+2(1-\alpha_n)M\|x_n-y_n\|\Big]\\
&\quad-\Big(\frac{(1-\alpha_n)^2}{2}-\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}-\Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\Big)\|y_n-y_{n-1}\|^2-\Big(\frac{(2-\gamma)}{\gamma}-(1+\epsilon)\Big)\|x_{n+1}-u_n\|^2.
\end{aligned}$$

Combine the last inequality with the notation of a_n and the following notation

$$h_n:=\alpha_n\|x_0-x^*\|^2+2(1-\alpha_n)\langle x_0-x^*,x_n-x^*\rangle+2(1-\alpha_n)M\|x_n-y_n\|$$

to deduce

$$\begin{aligned}
a_{n+1} \quad\le\quad &(1-\alpha_n)a_n+\alpha_nh_n-\Big(\frac{(2-\gamma)}{\gamma}-(1+\epsilon)\Big)\|x_{n+1}-u_n\|^2\\
&-\Big(\frac{(1-\alpha_n)^2}{2}-\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}-\Big(1+\frac{1}{\epsilon}\Big)\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\Big)\|y_n-y_{n-1}\|^2. \qquad (3.23)
\end{aligned}$$

Let $\limsup_{i\to\infty} h_{n_i}\le 0$ for each subsequence $\{a_{n_i}\}$ of $\{a_n\}$, then there exists $\liminf_{i\to\infty}(a_{n_{i+1}}-a_{n_i})\ge 0$. By (3.23), we have

$$\begin{aligned}
\limsup_{i\to\infty}\Big[\Big(\frac{(1-\alpha_{n_i})^2}{2}-\Big(2+\frac{1}{\epsilon}\Big)\frac{\lambda_{n_i}^2\mu^2}{\lambda_{n_i+1}^2}\Big)\Big]\|y_{n_{i-1}}-y_{n_i}\|^2 &\le \limsup_{i\to\infty}\big[(a_{n_i}-a_{n_{i+1}})+\alpha_{n_i}(h_{n_i}-a_{n_i})\big]\\
&\le-\liminf_{i\to\infty}(a_{n_{i+1}}-a_{n_i})\le 0,
\end{aligned}$$

which, by the previous property in (3.18), shows

$$\lim_{i\to\infty}\|y_{n_{i-1}}-y_{n_i}\|=0. \qquad (3.24)$$

Similarly,

$$\begin{aligned}
\limsup_{i\to\infty}\Big(\frac{(2-\gamma)}{\gamma}-(1+\epsilon)\Big)\|x_{n_i+1}-u_{n_i}\|^2 &\le \limsup_{i\to\infty}\big[(a_{n_i}-a_{n_{i+1}})+\alpha_{n_i}(h_{n_i}-a_{n_i})\big]\\
&\le-\liminf_{i\to\infty}(a_{n_{i+1}}-a_{n_i})\le 0,
\end{aligned}$$

which, by the region of γ, implies

$$\lim_{i\to\infty} \|x_{n_i+1} - u_{n_i}\| = 0. \tag{3.25}$$

By the fact that F is Lipschitz continuous, we have

$$\lim_{i\to\infty} \|Fy_{n_i-1} - Fy_{n_i}\| = 0. \tag{3.26}$$

So, we get from Algorithm 3.1 that

$$\|u_{n_i} - y_{n_i}\| = \lambda_{n_i}\|Fy_{n_i-1} - Fy_{n_i}\| \to 0, \quad \text{as } i \to \infty,$$

which implies

$$\|x_{n_i+1} - y_{n_i}\| \leq \|x_{n_i+1} - u_{n_i}\| + \|u_{n_i} - y_{n_i}\| \to 0, \quad \text{as } i \to \infty.$$

Thus, as i goes to infinity, we have

$$\begin{cases} \|x_{n_i} - y_{n_i}\| \leq \|x_{n_i} - y_{n_i-1}\| + \|y_{n_i-1} - y_{n_i}\| \to 0, \\ \|x_{n_i+1} - x_{n_i}\| \leq \|x_{n_i} - y_{n_i}\| + \|x_{n_i+1} - y_{n_i}\| \to 0. \end{cases}$$

Since $\{x_{n_i}\}$ is bounded, there exists a subsequence (denoted by $\{x_{n_{i_j}}\}$) of $\{x_{n_i}\}$ such that $x_{n_{i_j}} \rightharpoonup z \in H,\ j \to 0$ and

$$\limsup_{i\to\infty}\langle x_0 - x^*, x_{n_i} - x^*\rangle = \lim_{j\to\infty}\langle x_0 - x^*, x_{n_{i_j}} - x^*\rangle = \langle x_0 - x^*, z - x^*\rangle. \tag{3.27}$$

Therefore, as i goes to infinite, we have $y_{n_i} \rightharpoonup z, w_{n_i} \rightharpoonup z$ and $u_{n_i} \rightharpoonup z$, respectively.

Now, we need to show $z \in \mathcal{S}$. By the definition of y_n, we have $y_{n_i} = P_{\mathcal{C}}(w_{n_i} - \lambda_{n_i}Fy_{n_{i-1}})$ where $w_{n_i} = \alpha_{n_i}x_0 + (1-\alpha_{n_i})x_{n_i}$. From the characterization of $P_{\mathcal{C}}$, we have

$$\langle w_{n_i} - y_{n_i} - \lambda_{n_i}Fy_{n_{i-1}}, x - y_{n_i}\rangle \leq 0,\ \forall\, x \in \mathcal{C},$$

which implies

$$\frac{1}{\lambda_{n_i}}\langle w_{n_i} - y_{n_i}, x - y_{n_i}\rangle + \langle Fy_{n_{i-1}}, y_{n_i} - w_{n_i}\rangle \leq \langle Fy_{n_{i-1}}, x - w_{n_i}\rangle, \quad \forall\, x \in \mathcal{C}. \tag{3.28}$$

Being weakly convergent, $\{x_{n_i}\}$ is bounded. We also have that $\{w_{n_i}\}, \{y_{n_i}\}$ and $\{y_{n_{i-1}}\}$ are bounded. From the Lipschitz continuity of F, we have that $\{Fy_{n_{i-1}}\}$ is bounded. Since $\lim_{i\to\infty}\|y_{n_i} - w_{n_i}\| = 0$, $\{y_{n_i}\}$ is bounded and $\lambda_{n_i} \geq \min\left\{\lambda_1, \frac{\mu}{L}\right\}$, we obtain from (3.28) that

$$\liminf_{i\to\infty}\langle Fy_{n_{i-1}}, x - w_{n_i}\rangle \geq 0,\ \forall\, x \in \mathcal{C}. \tag{3.29}$$

Moreover, we have

$$\langle Fy_{n_i}, x - y_{n_i}\rangle \;=\; \langle Fy_{n_i} - Fy_{n_{i-1}}, x - w_{n_i}\rangle + \langle Fy_{n_i}, w_{n_i} - y_{n_i}\rangle + \langle Fy_{n_{i-1}}, x - w_{n_i}\rangle. \tag{3.30}$$

Combining (3.29), (3.26) and (3.30), we have

$$\liminf_{i\to\infty}\langle Fy_{n_i}, x - y_{n_i}\rangle \geq 0, \quad \forall\, x \in \mathcal{C}.$$

Next, we choose a sequence $\{\epsilon_i\}$ of positive numbers that are decreasing and approaching 0. For each i, let N_i be the smallest positive integers such that

$$\langle Fy_{n_j}, x - y_{n_j}\rangle + \epsilon_i \geq 0, \quad j \geq N_i. \tag{3.31}$$

Since $\{\epsilon_i\}$ is decreasing, we can easily see that $\{N_i\}$ is increasing. Furthermore, for each i, since $\{y_{N_i}\} \subset \mathcal{C}$ we can suppose that $Fy_{N_i} \neq 0$ (otherwise, y_{N_i} is a solution). Let $g_{N_i} = \frac{Fy_{N_i}}{\|Fy_{N_i}\|^2}$. We have $\langle Fy_{N_i}, g_{N_i}\rangle = 1$, for each i. From (3.31), we have for each i that

$$\langle Fy_{N_i}, x + \epsilon_i g_{N_i} - y_{N_i}\rangle \geq 0.$$

From the pseudomonotonicity of F, we obtain

$$\langle F(x + \epsilon_i g_{N_i}), x + \epsilon_i g_{N_i} - y_{N_i}\rangle \geq 0.$$

This implies that

$$\langle Fx, x - y_{N_i}\rangle \geq \langle Fx - F(x + \epsilon_i g_{N_i}), x + \epsilon_i g_{N_i} - y_{N_i}\rangle - \epsilon_i\langle Fx, g_{N_i}\rangle. \tag{3.32}$$

In what follows, we show $\lim\limits_{i\to\infty} \epsilon_i g_{N_i} = 0$. Since $w_{n_i} \rightharpoonup z$ and $\lim\limits_{i\to\infty} \|w_{n_i} - y_{n_i}\| = 0$, we have $y_{N_i} \rightharpoonup z$, as $i \to \infty$. Since $\{y_n\} \subset \mathcal{C}$, we obtain that $z \in \mathcal{C}$. Also,

$$0 < \|Fz\| \leq \liminf_{i\to\infty} \|Fy_{n_i}\|.$$

Since $\{y_{N_i}\} \subset \{y_{n_i}\}$ and $\epsilon_i \to 0,$ *as* $k \to \infty$, we obtain

$$0 \leq \limsup_{i\to\infty} \|\epsilon_i g_{N_i}\| = \limsup_{i\to\infty} \left(\frac{\epsilon_i}{\|Fy_{n_i}\|}\right) \leq \frac{\limsup\limits_{i\to\infty} \epsilon_i}{\liminf\limits_{i\to\infty} \|Fy_{n_i}\|} = 0,$$

which implies that $\lim\limits_{i\to\infty} \epsilon_i g_{N_i} = 0$. Since F is Lipschitz continuous, $\{y_{N_i}\}, \{g_{N_i}\}$ are bounded and $\lim\limits_{i\to\infty} \epsilon_i g_{N_i} = 0$, we obtain from (3.32) that

$$\liminf_{i\to\infty} \langle Fx, x - y_{N_i}\rangle \geq 0.$$

Hence, for all $x \in \mathcal{C}$, we have

$$\langle Fx, x - z\rangle = \lim_{i\to\infty} \langle Fx, x - y_{N_i}\rangle = \liminf_{i\to\infty} \langle Fx, x - y_{N_i}\rangle \geq 0.$$

Recalling Lemma 2.4, we obtain $z \in \mathcal{S}$.

Since $x^* = P_{\mathcal{S}}(x_0)$, it follows from (3.27) that

$$\limsup_{i\to\infty} \langle x_0 - x^*, x_{n_i} - x^*\rangle = \langle x_0 - x^*, z - x^*\rangle \leq 0.$$

Therefore, $\limsup\limits_{i\to\infty} h_{n_i} \leq 0$. By the condition that $\sum\limits_{n\to\infty}^{\infty} \alpha_n = \infty$, we have by Lemma 2.3 that $\lim\limits_{n\to\infty} a_n = 0$. Hence, $\{x_n\}$ converges strongly to $x^* = P_{\mathcal{S}}(x_0)$. $\square$

Assuming that the Lipschitz constant L of the operator F is available, the following result can be easily deduced from Theorem 3.2.

Remark 3.2 *We obtain Theorem 3.2 under the same set of assumptions given in [12, 39, 41]. However, we reduce the two values of F per iteration in the projection and contraction algorithm* (1.9) *utilized in [12, 39, 41] to just one value of F using our Algorithm 3.1.*

Corollary 3.1 *Assume that Assumption 3.1 is fulfilled and $\{(x_n, y_n)\}$ is generated by* (3.1)*. Let $\lambda_n \in [a, b] \subset \left(0, \frac{1}{L}\sqrt{\frac{\epsilon}{2(2\epsilon+1)}}\right)$ and $\gamma \in \left(0, \frac{2}{2+\epsilon}\right)$ for any $\epsilon > 0$. Suppose $\lim\limits_{n\to\infty} \alpha_n = 0$, $\sum\limits_{n=1}^{\infty} \alpha_n = \infty$. Then, the sequences $\{x_n\}$ and $\{y_n\}$ converge strongly to an element $x^* \in \mathcal{S}$, where $x^* = P_{\mathcal{S}}(x_0)$.*

Remark 3.3 *We now clarify the relations between Algorithm 3.1 and three existing algorithms.*

(i) ***Relaxed projection method [45]:*** *Anh proposed a relaxed projection approach built upon a solution mapping with strongly quasi-nonexpansive properties, and established weak convergence for partially pseudomonotone variational inequality problems. PCM-ep differs in that it adopts an explicit projection and contraction structure with a descent direction d_n and a non-negative contraction parameter β_n, and it achieves strong convergence via the Halpern-type parameter α_n rather than via a relaxed projection mapping.*

(ii) ***Halpern relaxed projection method [46]:*** *This method combines Halpern's iterative averaging with relaxed projections to obtain strong convergence under partial pseudomonotonicity. Although PCM-ep also uses a Halpern-type parameter α_n for strong convergence, it additionally incorporates extrapolation from the past and a contraction direction d_n, which allows PCM-ep to use only one operator evaluation per iteration, whereas the method in [46] still requires two evaluations (Fx_n and Fy_n) in its underlying extragradient-like structure.*

(iii) ***Viscosity-projection methods [47]:*** *These methods combine viscosity approximation (which forces strong convergence to a particular solution through a contraction mapping) with standard projection steps. PCM-ep does not use an explicit viscosity term; instead, it achieves strong convergence through the contraction step combined with the Halpern-type parameter α_n, and it reduces the operator evaluation count by reusing Fy_{n-1} from the past iterate.*

In summary, our PCM-ep distinguishes itself from [45, 46, 47] by simultaneously achieving: (a) strong convergence to $P_S(x_0)$, (b) pseudo-monotone and Lipschitz assumptions without cocoercivity, and (c) a single operator evaluation per iteration through extrapolation from the past within a projection–contraction framework.

At the end of this section, we discuss the connections between Algorithm 3.1 and the inertial forward-backward-forward algorithm (IFBF, [48, Algorithm 1]). First of all, Algorithm 3.1 offers several key advantages over existing methods for solving variational inequalities, which makes it theoretically sound and practically efficient for solving large-scale problems:

- **Computational Efficiency:** It requires only *one operator evaluation* per iteration (computing Fy_n), while most strongly convergent projection-contraction methods require two. This significantly reduces computational cost when operator evaluations are expensive.

- **Adaptive Step Size Without Lipschitz Constant:** The step size λ_n is updated adaptively via
$$\lambda_{n+1} = \min\left\{ \frac{\mu\|y_{n-1} - y_n\|}{\|Fy_{n-1} - Fy_n\|},\ \lambda_n + \tau_n \right\},$$
eliminating the need for prior knowledge of the global Lipschitz constant L.

- **Strong Convergence Under Weak Assumptions:** The algorithm achieves *strong convergence* under only *pseudo-monotonicity* and *Lipschitz continuity*, without requiring stronger conditions such as strong monotonicity or cocoercivity.

- **Practical Flexibility:** The contraction step with parameter $\gamma \in \left(0, \frac{2}{2+\epsilon}\right)$ provides additional robustness and can accelerate convergence in practice, as demonstrated in our numerical experiments. Besides, users have the flexibility to choose sequences $\{\alpha_n\}$ and $\{\tau_n\}$, and we just provide a pair of reasonable choices in experiments for fast convergence.

Remark 3.4 *While both Algorithm 3.1 (PCM-ep) and IFBF [48] are strongly convergent algorithms for solving variational inequality problems, they differ fundamentally in their design principles, computational structure, and primary objectives. The key distinction lies in their respective mechanisms for leveraging past information:*

(i) ***Mechanism Difference:*** *IFBF employs inertial extrapolation* $u_m = x_m + \theta(x_m - x_{m-1})$*, a momentum-based acceleration technique. However, PCM-ep uses extrapolation from the past by reusing* Fy_{n-1} *from the previous iteration—a gradient-recycling strategy to reduce computational load.*

(ii) ***Algorithmic Lineage:*** *IFBF is an inertial forward-backward-forward method, while PCM-ep is a projection and contraction method with an added contraction step* (d_n, β_n).

Contributions of PCM-ep over IFBF:

(i) *PCM-ep reduces operator evaluations from two per iteration (as in IFBF) to one per iteration, by reusing* Fy_{n-1}*. This is especially advantageous when operator evaluations are costly.*

(ii) *PCM-ep introduces extrapolation from the past within a projection–contraction framework, a novel combination that maintains strong convergence while lowering per-iteration cost.*

(iii) *By reducing the number of operator calls, PCM-ep is better suited for large-scale or real-time problems where* F *is expensive to compute (e.g., in PDE-constrained optimization, large-scale machine learning).*

(iv) *PCM-ep achieves strong convergence under standard assumptions, yet does so with lower computational overhead than existing strongly convergent PCMs and inertial methods like IFBF.*

(v) *PCM-ep opens a new direction for computation-aware algorithm design in variational inequalities, shifting focus from acceleration alone to efficiency per iteration.*

In summary, while IFBF [48] removes the on-line rule for inertial methods, Algorithm 3.1 addresses a more fundamental issue: reducing the computational cost of strongly convergent methods without sacrificing convergence guarantees. This makes Algorithm 3.1 a distinct and valuable contribution to the field of variational inequality solvers.

4 Numerical Experiments

In this section, we investigate the performance of Algorithm 3.1 by solving large-scale examples in comparison with several well-established methods. All experiments are implemented in MATLAB R2019b (64-bit) and performed on a PC with Windows 10 operating system, equipped with an Intel i7-8565U CPU and 16GB RAM. The following stopping criterion

$$r(x_n) = \|x_n - P_{\mathcal{C}}(x_n - \ Fx_n)\| < 10^{-8}$$

is used with a maximum of 10000 iterations. We also denote $\mathrm{IER}(n) := \|x_{n+1} - x_n\|$.

Consider the so-called Harker-Pang problem [18] with linear mapping $Fx = Wx + w_0$ where $w_0 \in \mathbb{R}^m$ and

$$W = A^\top A + B + \mathrm{diag}(\eta_i).$$

Here, A is a $m \times m$ matrix randomly generated with uniformly distributed entries in the interval $(-5, 5)$, B is a $m \times m$ skew-symmetric matrix generated in the same interval, η_i is an uniformly distributed random variable in the interval $(0, 2)$, and w_0 is randomly generated with uniformly distributed entries in the same interval as A. The feasible set is given by $\mathcal{C} = \{x \in \mathbb{R}^m : 0 \leq x \leq 10l\}$, where $l \in \mathbb{R}^m$ is the vector of ones. It is clear that F is strongly pseudo-monotone and Lipschitz continuous with Lipschitz constant $L = \|W\|$.

4.1 Effects of Algorithmic Parameters

We first test the effect of parameter ϵ on our proposed method (PCM-ep) for solving the Harker-Pang problem with $m = 2000$ and with initial points $x_0 = (1, 1, \cdots, 1)^\top \in \mathbb{R}^m, x_{-1} = y_{-1} = x_0$. Figure 1 depicts the comparison results of the aforementioned qualities $r(x_n)$ and IER(n) obtained by our PCM-ep with $\epsilon \in \{0.01, 0.05, 0.1, 0.5, 1, 2\}$, and with tuned parameters

$$\mu = 0.99\sqrt{\frac{\epsilon}{2(2\epsilon + 1)}}, \gamma = 0.99\frac{2}{2 + \epsilon}, \alpha_n = \frac{1}{m^4(n + 1)}, \tau_n = \frac{1}{(n + 1)^2}, \lambda_0 = 1.6.$$

We can check that the above choice of τ_n satisfies $\sum_{n=0}^{\infty} \tau_n < +\infty$, and α_n satisfies the conditions of Theorem 3.2. Figure 1 shows that the choice of ϵ could affect the performance of our PCM-ep, and choosing $\epsilon = 0.05$ is relatively better than other tested values. Hence, we set $\epsilon = 0.05$ as the default value in the forthcoming comparison experiments.

Next, we investigate the effect of τ_n on PCM-ep. Figure 2 shows the comparison of PCM-ep for solving the problem as above, but with three updating rules of τ_n: (case 1) $\tau_n = \frac{a}{(1+n)^2}$, (case 2) $\tau_n = \frac{a}{n^2+n}$, (case 3) $\tau_n = \frac{a}{n(n+1)(n+2)}$, where $a > 0$. Here, the legends for case 1-1 and case 1-20 in Figure 2 indicate that the algorithm PCM-ep uses case 1 with $a = 1$ and $a = 20$, respectively. It can be observed from Figure 2 that the first and second cases with $a = 20$ perform competitively and much better than the other cases. In the following, we will use case 1 with $a = 20$ as the default setting. Figure 3 also shows the comparison of PCM-ep with different starting points, from which we can see that choosing $x_0 =$ zeros(n,1) as the initial point is relatively better. However, in the forthcoming comparative experiments, we will use $x_0 =$ ones(n,1) as the initial point to maintain consistency with the initial point used by other existing methods.

4.2 Comparative Experiments with Existing Methods

Now, we compare our proposed method with the following well-established methods:

- Double Inertial Tseng's Extragradient Method (DiTEM, [32, Algorithm 2]) with the same parameters $\gamma_0 = 0.25, \alpha = 0.55, \beta = 0.454, \mu = 0.9, \zeta = 0.328$ as that in [32, Example 4.1].
- Tseng's forward-backward-forward method (FBF, [42]) with involved parameter $\lambda_n = 0.99/L$ as suggested by Bot, et al. [6];
- The generalized forward-backward-forward algorithm in (1.5), denoted by FBFp, with parameter $\lambda_n = 1/(2L)$ which is the upper bound suggested in [7];
- The forward-backward-forward algorithm with relaxation parameters (a-FBF, [6, Algorithm 3.1]) with parameters $\rho_n = 1.3$ and $\lambda_n = 0.99/L$;
- Inertial Forward-Backward-forward Algorithm (IFBF, [48]) with involved parameters $\tau = 10^{-4}, \mu = 0.45, \theta = 0.99$ and $\lambda_n = \frac{20}{(1+n)^2}$ and $\alpha_n = \frac{1}{m^4(n+1)}$. These settings perform better than that in [48, Section 4].
- The projection and contraction method (1.7), denoted by PCM, with fixed parameter $\lambda_n = 0.99/L$ as suggested by [8] for fast convergence.
- The standard Alternating Direction Method of Multipliers (ADMM, [40]) to solve the reformulation problem of (1.1): $\langle G(u), z - u\rangle \geq 0$ for any $z \in \mathcal{C} := \{(x, y) \mid x = y, 0 \leq y \leq 10l\}, u = (x; y)$ and $G(u) = (Fx; 0)$. For this method, we use the tuned relaxation parameter $\gamma = 1.6$ and penalty parameter $\beta = n/8$, which performs better than the adaptive tuning strategies presented in [20].
- Projection Algorithm with Extrapolation (PAE, [34]) with tuned parameters $\epsilon = 0.9, \phi_0 = 0.28, \phi_1 = 0.16, \lambda_0 = 0.5, \beta_n = \frac{15}{(1+n)^2}$.

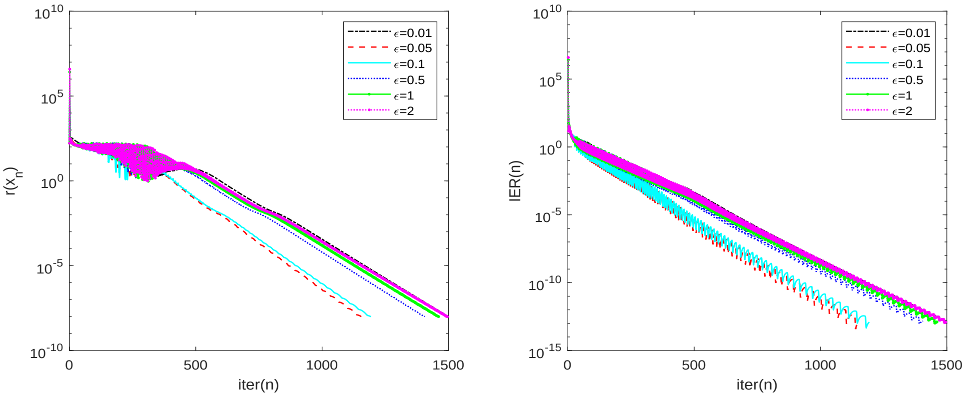

Figure 1: Effect of parameter ϵ on the performance of PCM-ep for Harker-Pang problem with $m = 2000$.

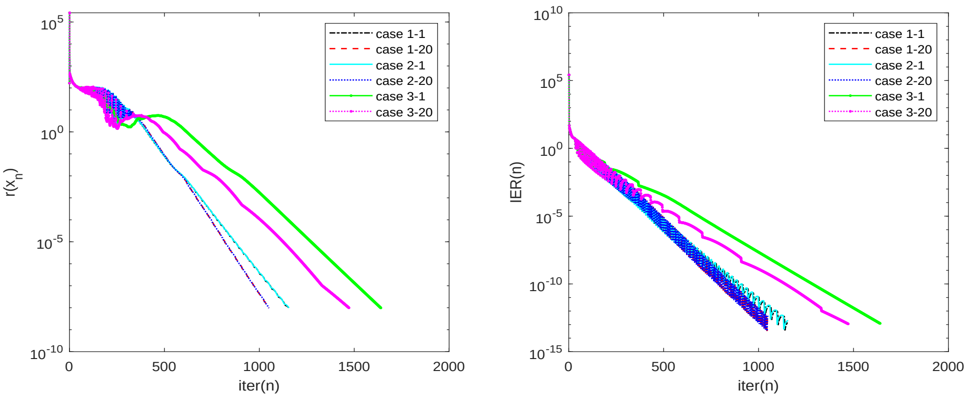

Figure 2: Effect of parameter τ_n on the performance of PCM-ep for Harker-Pang problem with $m = 2000$.

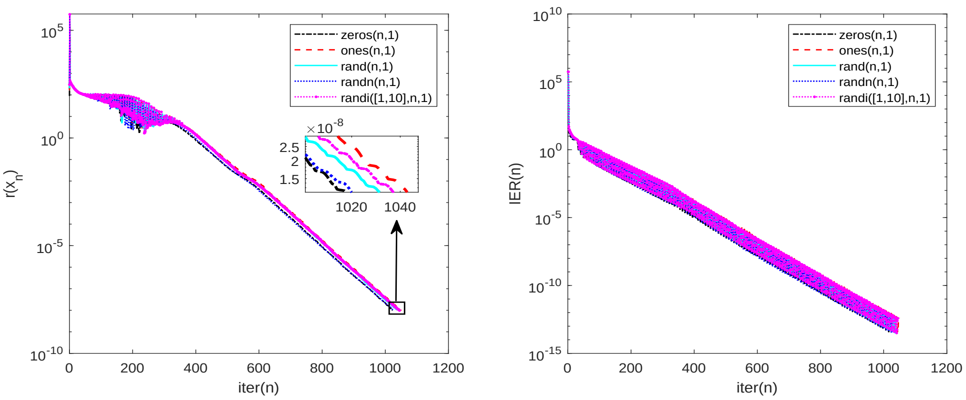

Figure 3: Effect of initial point x_0 on the performance of PCM-ep for Harker-Pang problem with $m = 2000$.

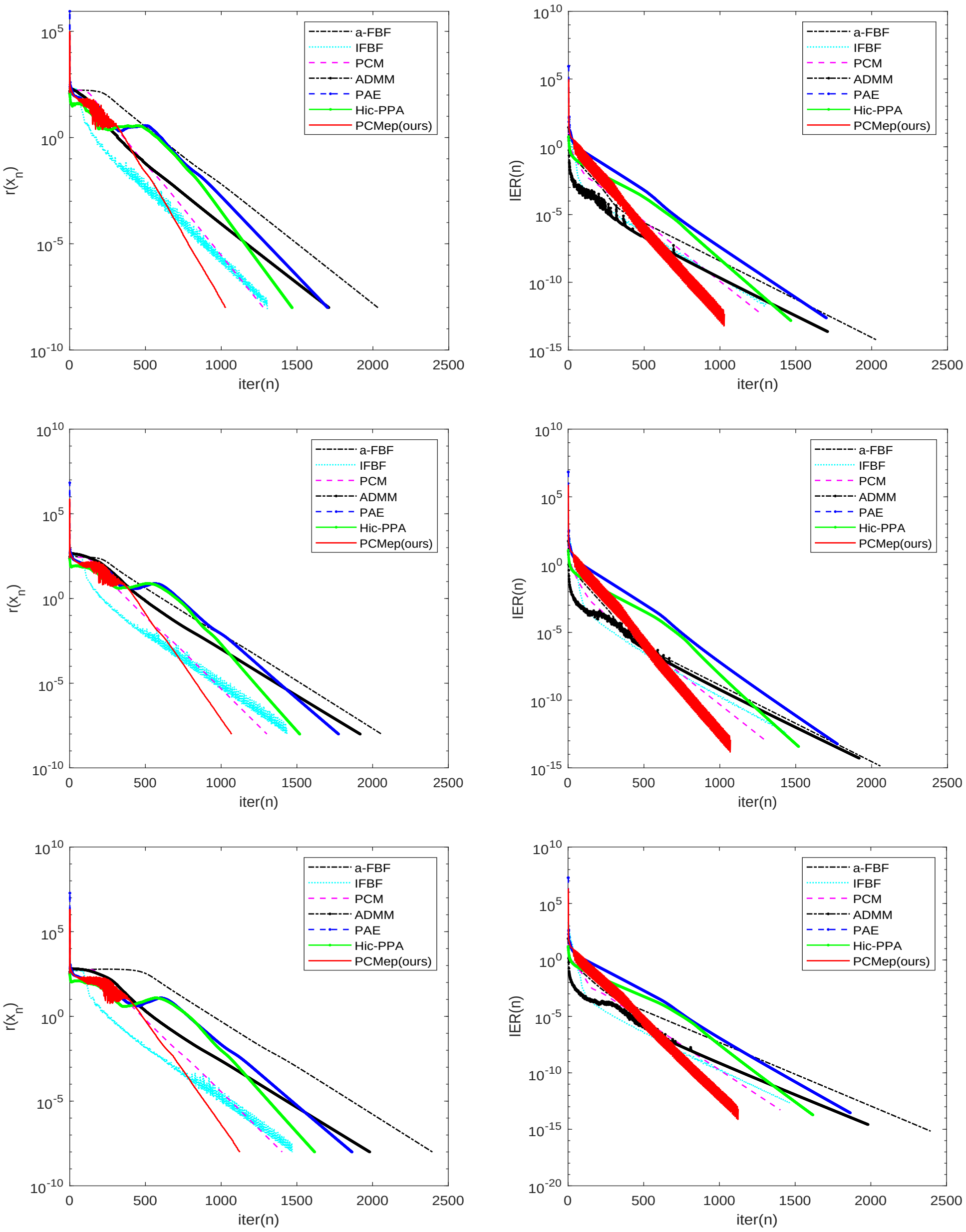

Figure 4: Comparison of different algorithms for Harker-Pang problem: from top to bottom, $m = 1000, 4000, 8000$.

- The hybrid inertial and contraction proximal point algorithm (Hic-PPA, [12]) with $a = 0.5, \gamma = 1.5, \alpha_n = \overline{\alpha}_n, \tau_n = \frac{1}{n^2}, \lambda_n = \frac{1}{2L}, \theta_n = 0.8 - \beta_n$ as the suggested settings in [12], but with $\beta_n = \frac{1}{m^4(n+1)}$ which performs better than the original.

All of the above-mentioned sequences satisfy the convergence conditions of each algorithm. Table 1 reports the comparison results of these algorithms for solving the Harker-Pang problem with $m \in \{1000, 2000, 4000, 6000, 8000\}$, where the bold value denotes the smallest for each test problem. Figure 4 also compares the decreasing trend in the residual function $r(x_n)$ and the iterative residual IER(n) by the last seven methods with the same initial point. Results of the first three methods are not depicted since they require more iterations. We can observe from the reported results that our proposed method, PCM-ep, performs significantly better than other methods to satisfy the desired stopping criterion.

5 Conclusion and Discussion

This paper has studied a strongly convergent projection and contraction-type algorithm with an extrapolation of the operator from past iterates to reduce the two-value computations of the cost operator apparent in the projection and contraction algorithm to a single value computation per iteration. We gave strong convergence results of the proposed algorithm under standard conditions. Large-scale comparison experiments showed that our algorithm outperformed other related algorithms in the literature.

Although the parameter μ is bounded by $\sqrt{\epsilon/(2(2\epsilon+1))}$, the adaptive update (3.2) ensures that the step size λ_n can grow via the term $\lambda_n + \tau_n$. In practice, choosing $\tau_n = O(1/n^2)$ allows λ_n to increase sufficiently, preventing stagnation. Exploring relaxed bounds for μ is an interesting topic to further improve large-step performance while maintaining convergence guarantees. In addition, while this paper focuses on solving variational inequality problems with a single operator F, the consideration of *finite families of contraction mappings* presents several potential benefits that could be explored in future extensions:

– *Hierarchical and Composite Problems:* Many real-world optimization problems involve nested or composite structures that can be modeled using finite families of operators. For instance, hierarchical variational inequalities or systems of interrelated variational inequalities naturally involve multiple operators.

– *Operator Splitting and Decomposition:* Complex operators F can often be decomposed into finitely many simpler contraction mappings:

$$F = \sum_{i=1}^{m} F_i \quad \text{or} \quad F = F_m \circ F_{m-1} \circ \cdots \circ F_1.$$

Algorithms that handle such families directly could leverage this structure for improved efficiency.

– *Parallel and Cyclic Implementation:* Finite families allow for parallel or cyclic algorithmic schemes, where different operators are applied in different iterations. This could lead to (i) reduced computational burden per iteration, (ii) better utilization of distributed computing resources; (iii) improved convergence through strategic operator ordering.

– *Adaptive Operator Selection:* Algorithms could adaptively select which operator(s) to apply based on local problem characteristics, computational cost of evaluating each operator, or progress metrics from previous iterations.

– *Multi-objective and Game-Theoretic Applications:* Finite families naturally arise in multi-objective optimization and Nash equilibrium problems, where each objective or player's optimality condition corresponds to a different operator.

Table 1: Numerical results of different algorithms for the Harker-Pang problem.

Size	DiTEM			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	1.0381	2181	9.8979e-9	1.8181e-13
2000	6.5103	2159	9.9682e-9	9.3702e-14
4000	33.4295	2230	9.9218e-9	4.6277e-14
6000	76.6062	2288	9.9915e-9	3.0769e-14
8000	129.5787	2384	9.8749e-9	2.2764e-14
Size	FBF			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.8812	2646	9.9863e-9	4.2650e-15
2000	8.3471	2885	9.9939e-9	2.1329e-15
4000	39.5787	2679	9.9953e-9	1.0595e-15
6000	101.3622	2967	9.9185e-9	7.0757e-16
8000	187.1944	3118	9.9472e-9	5.3061e-16
Size	FBFp			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.8045	2433	9.9355e-9	1.4886e-13
2000	6.9943	2436	9.9920e-9	7.5415e-14
4000	36.5404	2495	9.9670e-9	3.7613e-14
6000	84.8855	2555	9.9503e-9	2.4900e-14
8000	156.3067	2652	9.9132e-9	1.8606e-14
Size	a-FBF			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.6976	2033	9.9024e-9	5.5141e-15
2000	6.3501	2216	9.9807e-9	2.7775e-15
4000	30.4817	2058	9.9452e-9	1.3745e-15
6000	76.4782	2279	9.9032e-9	9.2137e-16
8000	141.5425	2395	9.9294e-9	6.9060e-16
Size	IFBF			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.4675	1304	8.3816e-9	2.3454e-12
2000	4.1583	1434	8.9787e-9	8.8106e-13
4000	21.2326	1435	9.7078e-9	3.9679e-13
6000	45.9116	1434	9.4176e-9	2.4233e-13
8000	83.8905	1469	9.6006e-9	2.0775e-13
Size	PCM			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.4767	1278	9.8963e-9	4.0810e-13
2000	3.9379	1345	9.8619e-9	2.0345e-13
4000	19.2866	1302	9.9959e-9	1.0218e-13
6000	43.3915	1402	9.8372e-9	6.7803e-14
8000	77.1424	1403	9.9823e-9	5.1429e-14
Size	ADMM			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.6884	1709	9.9657e-9	2.3181e-14
2000	3.7662	1797	9.8917e-9	1.1416e-14
4000	19.0400	1918	9.8841e-9	5.2191e-15
6000	40.4426	1959	9.8932e-9	3.4766e-15
8000	73.1083	1981	9.9733e-9	2.6722e-15

Table 2: Numerical results of different algorithms for the Harker-Pang problem (continued).

Size	PAE			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.4584	1701	9.9534e-9	2.3501e-13
2000	3.4963	1764	9.9509e-9	1.1442e-13
4000	17.9045	1775	9.9132e-9	5.7649e-14
6000	36.8395	1795	9.9124e-9	3.8821e-14
8000	68.2062	1864	9.8598e-9	2.8884e-14
Size	Hic-PPA			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	0.4825	1467	9.7819e-9	1.4852e-13
2000	3.2308	1488	9.9565e-9	7.6196e-14
4000	16.0097	1519	9.7989e-9	3.7493e-14
6000	35.9359	1561	9.9822e-9	2.5336e-14
8000	62.9666	1617	9.9255e-9	1.8882e-14
Size	PCMep(ours)			
m	time(s)	Iter	$r(x_{\text{end}})$	IER(end)
1000	**0.4481**	**1029**	9.7307e-9	4.6253e-13
2000	**2.9785**	**1050**	9.5897e-9	3.8508e-13
4000	**15.0372**	**1070**	9.8932e-9	1.1841e-13
6000	**32.5386**	**1095**	9.8746e-9	1.2347e-13
8000	**59.7496**	**1124**	9.8161e-9	6.3488e-14

Besides, extending our algorithm to non-convex and stochastic settings is another interesting topic, that is, (1.1) with F being a continuous nonconvex operator with stochastic variable:

1. **Non-convex Setting:** We plan to extend the algorithm to variational inequalities involving non-monotone operators. A promising direction involves combining our algorithm with inertial strategies while incorporating adaptive regularization terms to handle non-convexity. Preliminary analysis suggests that under appropriate conditions (e.g., weak Minty variational inequality conditions), convergence to critical points may be established.

2. **Stochastic Setting:** For large-scale or data-driven applications where only stochastic estimates of the operator F are available, we intend to develop a stochastic variant of our algorithm. The main challenge will be to design variance-reduction techniques (such as the popular SVRG or SARAH estimators) compatible with our extrapolation scheme, ensuring convergence with reduced sample complexity.

3. **Applications to Modern Machine Learning:** The algorithm single-operator-evaluation feature makes it particularly suitable for training large neural networks and solving structured non-convex optimization problems. We will investigate its application to adversarial training, generative adversarial networks (GANs), and reinforcement learning, where variational inequality formulations naturally arise.

CRediT authorship contribution statement

Yekini Shehu: Writing - original draft, Investigation, Formal analysis, Conceptualization, Visualization. **Jianchao Bai:** Writing - review & editing, Validation, Software, Methodology, Funding acquisition, 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

Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

Acknowledgments

The authors would like to thank the anonymous referees for providing very constructive comments, which have significantly improved the quality of the paper. This work was partially supported by the National Natural Science Foundation of China (12471298) and the Open Project Program of Center for Applied Mathematics of Fujian Province (S20250605).

References

[1] Aubin, J., Ekeland, I., *Applied Nonlinear Analysis*, Wiley, New York, (1984).

[2] Baiocchi, C., Capelo, A., *Variational and Quasivariational Inequalities: Applications to Free Boundary Problems*, John Wiley & Sons Ltd, (1984).

[3] Bauschke, H., Combettes, P., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, Cham, (2017).

[4] Bai, J., Chen, Y., Yu, X., Zhang, H., *Generalized asymmetric forward-backward-adjoint algorithms for convex-concave saddle-point problem*, Journal of Scientific Computing, 102 (2025), article number 80.

[5] Bai, J., Zhang, M., Zhang, H., *An inexact ADMM for separable nonconvex and nonsmooth optimization*, Computational Optimization and Applications, 90 (2025), pp. 445–479.

[6] Boţ, R., Csetnek, E., Vuong, P., *The forward-backward-forward method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces*, European Journal of Operational Research, 287 (2020), pp. 49–60.

[7] Bohm, A., Sedlmayer, M., Csetnek, E., Boţ, R., *Two steps at a time—taking GAN training in stride with Tseng's method*, SIAM Journal on Mathematics of Data Science, 4 (2022), pp. 750–771.

[8] Cai, X., Gu, G., He, B., *On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators*, Computational Optimization and Applications, 57 (2014), pp. 339–363.

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

[10] Cholamjiak, P., Thong, D., Cho, Y., *A novel inertial projection and contraction method for solving pseudomonotone variational inequality problems*, Acta Applicandae Mathematicae, 169 (2020), pp. 217–245.

[11] Cottle, R., Yao, J., *Pseudomonotone complementary problems in Hilbert space*, Journal of Optimization Theory and Applications, 75 (1992), pp. 281–295.

[12] Dey, S., *A hybrid inertial and contraction proximal point algorithm for monotone variational inclusions*, Numerical Algorithms, 93 (2023), pp. 1–25.

[13] Dong, Q., Cho, Y., Zhong, L., Rassias, Th., *Inertial projection and contraction algorithms for variational inequalities*, Journal of Global Optimization, 70 (2018), pp. 687–704.

[14] Dong, Q., Yang, J., Yuan, H., *The projection and contraction algorithm for solving variational inequality problems in Hilbert space*, Journal of Nonlinear and Convex Analysis, 20 (2019), pp. 111–122.

[15] Fichera, G., *Sul problema elastostatico di Signorini con ambigue condizioni al contorno*, Atti della Accademia Nazionale dei Lincei, Classe di Scienze Fisiche, Matematiche e Naturali, 34 (1963), pp. 138–142.

[16] Gidel, G., Berard, H., Vignoud, G., Vincent, P., Lacoste-Julien, S., *A variational inequality perspective on generative adversarial networks*, Proceedings of the 2019 International Conference on Learning Representations, (2019), pp. 1–38.

[17] Han, D., *Inexact operator splitting methods with selfadaptive strategy for variational inequality problems*, Journal of Optimization Theory and Applications, 132 (2007), pp. 227–243.

[18] Harker, P., Pang, J., *A damped-Newton method for the linear complementarity problem*, Lectures in Applied Mathematics, 26 (1990), pp. 265–284.

[19] He, B., *A class of projection and contraction methods for monotone variational inequalities*, Applied Mathematics and Optimization, 35 (1997), pp. 69–76.

[20] He, B., Yang, H., Wang, S., *Alternating directions method with self-adaptive penalty parameters for monotone variational inequalities*, Journal of Optimization Theory and Applications, 106 (2000), pp. 337–356.

[21] Kinderlehrer, D., Stampacchia, G., *An Introduction to Variational Inequalities and Their Applications*, Academic Press, (1980).

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

[23] Li, M., Yuan, X., *An APPA-based descent method with optimal step-sizes for monotone variational inequalities*, European Journal of Operational Research, 186 (2008), pp. 486–495.

[24] Li, B., Liu, P., Shao, H., Wu, T., Xu, J., *A proximal alternating direction method of multipliers with a proximal-perturbed Lagrangian function for nonconvex and nonsmooth structured optimization*, Optimization Letters, 20 (2026), pp. 109–122.

[25] Liu, P., Jian, J., He, B., Jiang, X., *Convergence of Bregman Peaceman-Rachford splitting method for nonconvex nonseparable optimization*, Journal of the Operations Research Society of China, 11 (2023), pp. 707–733.

[26] Liu, H., Yang, J., *Weak convergence of iterative methods for solving quasimonotone variational inequalities*, Computational Optimization and Applications, 77 (2020), pp. 491–508.

[27] Maingé, P., *Approximation methods for common fixed points of nonexpansive mappings in Hilbert spaces*, Journal of Mathematical Analysis and Applications, 325 (2007), pp. 469–479.

[28] Malitsky, Y., Tam, M., *A forward–backward splitting method for monotone inclusions without cocoercivity*, SIAM Journal on Optimization, 30 (2020), pp. 1451–1472.

[29] Malitsky, Y., *Golden ratio algorithms for variational inequalities*, Mathematical Programming, 184 (2020), pp. 383–410.

[30] Malitsky, Y., *Projected reflected gradient methods for monotone variational inequalities*, SIAM Journal on Optimization, 25 (2015), pp. 502–520.

[31] Mu, X., Zhang, Y., *Grasping force optimization for multi-fingered robotic hands using projection and contraction methods*, Journal of Optimization Theory and Applications, 183 (2019), pp. 592–608.

[32] Pakkaranang, N., *Double inertial extragradient algorithms for solving variational inequality problems with convergence analysis*, Mathematical Methods in the Applied Sciences, 47 (2024), pp. 11642–11669.

[33] Panyanak, B., Khunpanuk, C., Pholasa, N., Pakkaranang, N., *A novel class of forward-backward explicit iterative algorithms using inertial techniques to solve variational inequality problems with quasi-monotone operators*, AIMS Mathematics, 8 (2023), pp. 9692–9715.

[34] Peng, J., Adamu, A., Shehu, Y., Yao, J., *Projection algorithm with extrapolations from the past for variational inequalities*, Journal of Optimization Theory and Applications, 208 (2026), article number 61.

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

[36] Saejung, S., Yotkaew, P., *Approximation of zeros of inverse strongly monotone operators in Banach spaces*, Nonlinear Analysis, 75 (2012), pp. 742–750.

[37] Shehu, Y., Iyiola, O., *Projection methods with alternating inertial steps for variational inequalities: weak and linear convergence*, Applied Numerical Mathematics, 157 (2020), pp. 315–337.

[38] Stampacchia, G., *"Variational Inequalities", in: Theory and Applications of Monotone Operators*, Proceedings of the NATO Advanced Study Institute, Venice, Italy, (1968), pp. 102–192.

[39] Tan, B., Qin, X., Yao, J., *Strong convergence of inertial projection and contraction methods for pseudomonotone variational inequalities with applications to optimal control problems*, Journal of Global Optimization, 82 (2022), pp. 523–557.

[40] Trémolières, R., Lions, J., Glowinski, R., *Numerical Analysis of Variational Inequalities*, North-Holland, Amsterdam, (1981).

[41] Thong, D., Vinh, N., Cho, Y., *New strong convergence theorem of the inertial projection and contraction method for variational inequality problems*, Numerical Algorithms, 84 (2020), pp. 285–305.

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

[43] Tongnoi, B., *A modified Tseng's algorithm with extrapolation from the past for pseudo-monotone variational inequalities*, Taiwanese Journal of Mathematics, 28 (2024), pp. 187–210.

[44] Tongnoi, B., *Tseng's algorithm with extrapolation from the past endowed with variable metrics and error terms*, Numerical Functional Analysis and Optimization, 44 (2023), pp. 87–123.

[45] Anh, P., *Relaxed projection methods for solving variational inequality problems*, Journal of Global Optimization, 90 (2024), pp. 909–930.

[46] Thanh, D., Anh, P., *Halpern relaxed projection methods for solving variational inequality problems*, Optimization Letters, 20 (2026), pp.213–230.

[47] Trang, C., Tu, H., Anh, P., *New viscosity-projection methods for solving variational inequality problems with applications to image restoration problems*, Communications in Nonlinear Science and Numerical Simulation, 152 (2026), article number 109179.

[48] Yao, Y., Adamu, A., Shehu, Y., *Strongly convergent inertial forward-backward-forward algorithm without on-line rule for variational inequalities*, Acta Mathematica Scientia, 44 (2024), pp. 551–566.

[49] Zhang, C., Dong, Q., Chen, J., *Multi-step inertial proximal contraction algorithms for monotone variational inclusion problems*, Carpathian Journal of Mathematics, 36 (2020), pp. 159–177.