Commun Nonlinear Sci Numer Simulat 152 (2026) 109213

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

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

Extragradient algorithms with double inertial for solving variational inequalities and applications to signal processing

Bing Tan [a], Jiawei Chen [a], Songxiao Li [b,*], Xiaoqing Ou [c]

[a] School of Mathematics and Statistics, Southwest University, 400715, Chongqing, China
[b] Department of Mathematics, Shantou University, 515063, Shantou, China
[c] School of Mathematics and Information Science, North Minzu University, 750021, Yinchuan, China

A R T I C L E I N F O

2020 MSC:
47J20
68W10

Keywords:
Variational inequality problem
Double inertial method
Extragradient method
Signal processing
Linear convergence

A B S T R A C T

We propose four double inertial-based extragradient algorithms for solving pseudomonotone (or quasimonotone) variational inequality problems in real Hilbert spaces. The proposed algorithms utilize a nonmonotone adaptive step size rule and require only one projection onto the feasible set per iteration. Under appropriate conditions, we establish the weak convergence, strong convergence, and linear convergence theorems for the proposed algorithms. Moreover, we demonstrate the nonasymptotic $O(1/t)$ convergence rate of the proposed algorithms when the operator is pseudomonotone (or quasimonotone) and provide global error bounds in the case where the operator is strongly pseudomonotone. Numerical experimental results in signal processing illustrate the computational advantages of the proposed algorithms compared to state-of-the-art methods.

1. Introduction

This paper aims to propose four new double inertial extragradient algorithms for addressing pseudomonotone (or quasimonotone) variational inequality problems (shortly, VIP) in real Hilbert spaces. Let $\mathbb{C}$ be a nonempty, closed, and convex subset of a real Hilbert space $\mathbb{H}$, equipped with an inner product $\langle\cdot,\cdot\rangle$ and the induced norm $\|\cdot\|$, and let $\mathcal{G}:\mathbb{C}\to\mathbb{H}$ be a nonlinear mapping. Recall that the classical VIP is stated as follows:

$$\text{Find } a^{\dagger}\in\mathbb{C} \text{ s.t. } \langle\mathcal{G}(a^{\dagger}), a-a^{\dagger}\rangle\geq 0,\quad \forall a\in\mathbb{C}. \tag{VI}$$

The solution set of problem (VI), denoted by $\mathrm{VI}(\mathbb{C},\mathcal{G})$. Along with (VI), its dual VIP (also known as the Minty VIP) of (VI) is formulated as follows:

$$\text{Find } a^{\dagger}\in\mathbb{C} \text{ s.t. } \left\langle\mathcal{G}(a), a-a^{\dagger}\right\rangle\geq 0,\quad \forall a\in\mathbb{C}. \tag{DVI}$$

The solution set of the dual VIP (DVI) is denoted by $\mathrm{DVI}(\mathbb{C},\mathcal{G})$. It is evident that $\mathrm{DVI}(\mathbb{C},\mathcal{G})$ forms a closed convex set, which may be empty. If $\mathcal{G}$ is pseudomonotone and continuous, then $\mathrm{VI}(\mathbb{C},\mathcal{G})$ and $\mathrm{DVI}(\mathbb{C},\mathcal{G})$ are equivalent (see [1, Lemma 2.1]). However, if $\mathcal{G}$ is quasimonotone and continuous, the inclusion $\mathrm{VI}(\mathbb{C},\mathcal{G})\subset\mathrm{DVI}(\mathbb{C},\mathcal{G})$ does not hold (see [2, Example 4.2]).

Variational inequalities represent a central research area in nonlinear analysis and optimization theory, providing a unified framework for addressing diverse problems such as fixed-point problems, split feasibility problems, Nash equilibrium problems, and saddle point problems. This unification underscores their profound theoretical and practical importance. Variational inequality problems

* Corresponding author.
E-mail addresses: bingtan72@gmail.com (B. Tan), j.w.chen713@163.com (J. Chen), jyulsx@163.com (S. Li), ouxiaoqing413@163.com (X. Ou).

https://doi.org/10.1016/j.cnsns.2025.109213
Received 23 December 2024; Received in revised form 13 July 2025; Accepted 10 August 2025
Available online 12 August 2025

also find numerous practical applications, including optimal control problems [3], compressed sensing problems [4], image processing problems [5], and dynamic traffic flow allocation problems [6]. Next, we briefly introduce some of these applications. In traffic flow theory, VIPs can be used to describe the evolution of traffic flow and congestion phenomena. By solving these inequalities, the efficiency and flow distribution of traffic systems can be optimized. In economics, VIPs can be used to describe issues of market competition and resource allocation. By solving these inequalities, equilibrium points and optimal solutions in economic systems can be identified. Variational inequality is widely applied in image restoration, dealing with damaged images, removing noise, or filling in missing image regions. The key lies in constructing an energy function containing a measure of dissimilarity between the original and restored images and constraints on the smoothness of the restored image. By minimizing this energy function, optimal image restoration results can be obtained.

A widely used approach for solving variational inequalities is the extragradient method [7,8]. This method generally involves computing two projections onto the feasible set in each iteration, which can significantly impact computational efficiency, especially when these projections are computationally intensive. To overcome this disadvantage, some methods have been proposed that require only one projection onto the feasible set per iteration; see, e.g., [9–13]. Notably, the convergence analysis of the methods discussed in [9–13] relies on the assumption that the operator is monotone and L-Lipschitz continuous, with the step size constrained to the interval $(0, 1/L)$. Therefore, scholars' efforts are divided into three main areas: the first involves relaxing the monotonicity assumption to pseudomonotonicity or quasimonotonicity (see, e.g., [2,14–18]); the second focuses on replacing the fixed step size with an adaptive step size (see, e.g., [19,20]); and the third involves relaxing the Lipschitz continuity condition to uniform continuity (see, e.g., [21]).

Algorithm convergence speed is a key concern. The inertial method (cf. [22]), inspired by second-order dissipative dynamical systems, serves as an effective numerical technique to enhance the convergence rate of algorithms. The core concept of this approach lies in making the next iteration dependent on the previous two iterations (or multiple earlier iterations). Its advantage lies in its simplicity of implementation and its ability to effectively accelerate the iterative process, thus improving the convergence speed of algorithms. In particular, Beck and Teboulle [23] utilize inertial techniques to elevate the convergence rate of the iterative shrinkage-thresholding algorithm from first order to second order, and demonstrate the advantages and efficiency of the proposed algorithm through applications in image processing. Over the past few decades, researchers have employed a combination of inertial methods with various optimization techniques to address a range of problems; see, e.g., [4,5,24–26]. They demonstrated the computational advantages of these methods over non-inertial-based methods. Recently, double inertial techniques garnered attention and research from scholars as an improvement over single-step inertial methods. In 2018, Dong et al. [27] first introduced a general Mann algorithm with double inertial steps for nonexpansive mapping. Subsequently, Yao et al. [28] combined double inertial techniques with subgradient extragradient algorithms (shortly, SEGA) to solve VIPs and achieved faster convergence speed compared to algorithms with a single inertial step. In recent years, scholars integrated the double inertial method, the SEGA, the Tseng's extragradient algorithm, and the projection and contraction algorithms (shortly, PCA) to solve VIPs and demonstrated its computational efficiency through numerical experiments; see, e.g., [28–36]. To provide a clearer comparison, we summarize the main features of these algorithms in Table 1.

Recently, some scholars focused on solving quasimonotone VIPs due to their prevalence compared to pseudomonotone VIPs. Scholars combined inertial methods (single-step inertial methods and double inertial methods) and extragradient-type algorithms to discover solutions for VIPs with quasimonotone operators; see, e.g., [2,16,30,33]. It is observed that the algorithms mentioned above use the same step size and parameters with small range constraints when calculating the sequences at each iteration, leading to slow convergence. Our questions are:

> *can we design algorithms that use different step sizes to compute the sequences in each iteration and increase the range of the involved parameters? Furthermore, can the algorithms be applied to solve a broader range of pseudomonotone (or quasimonotone) and non-Lipschitz continuous VIPs?*

We answer this question affirmatively. Our contributions in this paper can be summarized in four aspects:

Table 1
Comparison of the proposed algorithms with some existing double inertial methods.

Algorithm	Monotonicity	Continuity	Step size rule	Convergence
[28, Alg. 1]	PM	L	A	W + S + L
[29, Alg. 1]	PM	L	NM	W + L
[30, Alg. 1]	QM	L	A	W + S + L
[31, Alg. 1]	QM	L	A	W + S + L
[32, Alg. 3.1]	QM	L	NM	W + S + L
[33, Alg. 1]	QM	L	AT	W + L
[34, Alg. 3.1]	M	L	AT	W + L
[35, Alg. 1]	QM	L	F / A	W + S + L
[36, Alg. 3.1]	PM	L	A	W
Our Algorithms 3.1–3.4	QM	L / NL	NM	W + S + L

QM = Quasimonotone, PM = Pseudomonotone, M = Monotone.
L = Lipschitz continuous, NL = Non-Lipschitz continuous.
NM = Nonmonotone, A = Adaptive (nonincreasing), AT = Armijo-type, F = Fixed.
W = Weak, S = Strong, L = Linear convergence.

(i) We propose four accelerated double inertial extragradient-type methods for solving pseudomonotone and quasimonotone VIPs, establishing their weak, strong, and linear convergence properties. Additionally, we demonstrate the nonasymptotic $O(1/t)$ convergence rate of the proposed algorithms when the operator is pseudomonotone (or quasimonotone). Furthermore, we derive global error bounds for the proposed methods in the case of strongly pseudomonotone operators, which can serve as a basis for designing stopping criteria to achieve a specified accuracy.
(ii) The stated algorithms employ different step sizes at each iteration, resulting in faster convergence rates. These methods utilize an adaptive step size without requiring a line search, enabling them to operate without prior knowledge of the Lipschitz constant. Moreover, our step size criterion generates a nonmonotone sequence of step sizes, offering an advantage over the nonincreasing step size sequences used in [28,30,31] and the Armijo-type step size sequences in [33].
(iii) Compared to the algorithms for solving pseudomonotone VIPs in [28,29], our algorithms can be extended to tackle quasimonotone VIPs, highlighting their broader applicability. Additionally, the parameters of our algorithms permit a larger range of values than those in [28–34]. Furthermore, we demonstrate that the proposed algorithms can also be extended to solve VIPs involving pseudomonotone and uniformly continuous operators when employing Armijo-type step size rules.
(iv) We present applications of our approaches to signal processing problems, showcasing their computational efficiency in comparison to those in [28–30,33]. Numerical experimental results confirm that the algorithms introduced in this paper are both efficient and robust.

The structure of this paper is as follows. In Section 2, we review key terminology and lemmas that are essential for the subsequent convergence analysis. In Section 3, we propose four efficient extragradient algorithms with double inertial terms and adaptive nonmonotone step size rules and analyze their weak convergence, strong convergence, linear convergence, convergence rate, and global error bound under suitable conditions. In Section 4, we apply our approaches to signal processing problems and compare their performance with several methods. Finally, in Section 5, we conclude the paper with a summary of our findings.

2. Preliminaries

The weak and strong convergence of $\{a_t\}$ converge to t are denoted by $a_t \rightharpoonup t$ and $a_t \to t$, respectively. For each $w, u \in \mathbb{H}$ and $\varphi \in \mathbb{R}$,

$$\|\varphi w + (1-\varphi)u\|^2 = \varphi\|w\|^2 + (1-\varphi)\|u\|^2 - \varphi(1-\varphi)\|w-u\|^2. \tag{1}$$

Let $\mathbf{Proj}_{\mathbb{C}} : \mathbb{H} \to \mathbb{C}$ denote the metric projection. It is known that $\mathbf{Proj}_{\mathbb{C}}$ possesses the following fundamental properties:

$$\langle w - \mathbf{Proj}_{\mathbb{C}}(w), u - \mathbf{Proj}_{\mathbb{C}}(w)\rangle \le 0, \quad \forall w \in \mathbb{H}, u \in \mathbb{C}. \tag{2}$$

$$\|\mathbf{Proj}_{\mathbb{C}}(w) - a\|^2 \le \|w-a\|^2 - \|w - \mathbf{Proj}_{\mathbb{C}}(w)\|^2, \quad \forall w \in \mathbb{H}, a \in \mathbb{C}. \tag{3}$$

Remark 1. Eq. (2) is the well-known projection theorem (see [37, Theorem 3.16, p. 53]). Eq. (3) can be obtained from Eq. (2) (see, e.g., [38], Eq. (2)). Note that projections onto some specific sets can be explicitly computed (see, e.g., [37, Chapter 29] for more details).

(i) Let $\mathbb{H}$ be a real Hilbert space. For a half-space $\mathbb{C}_{u,\varphi} = \{w \in \mathbb{H} : \langle u, w\rangle \le \varphi\}$, the metric projection $\mathbf{Proj}_{\mathbb{C}_{u,\varphi}}(w)$ is given by

$$\mathbf{Proj}_{\mathbb{C}_{u,\varphi}}(w) = w - \max\left\{\frac{\langle u, w\rangle - \varphi}{\|u\|^2}, 0\right\}u.$$

(ii) For a closed ball $\mathbb{C}_{a,r} = \{w \in \mathbb{H} : \|w-a\| \le r\}$ with center $a \in \mathbb{H}$ and radius $r > 0$, we have

$$\mathbf{Proj}_{\mathbb{C}_{a,r}}(w) = a + \frac{r}{\max\{\|w-a\|, r\}}(w-a).$$

(iii) For a box constraint set $\mathbb{C}_{[\alpha,\beta]} = \{w \in \mathbb{H} : \alpha_i \le w_i \le \beta_i \text{ for all } i\}$, the projection is performed componentwise as

$$\mathbf{Proj}_{\mathbb{C}_{[\alpha,\beta]}}(w)_i = \min\left\{\beta_i, \max\left\{w_i, \alpha_i\right\}\right\}, \quad i = 1, \dots, n.$$

Let $a, b \in \mathbb{H}$. Recall that a mapping $\mathcal{G} : \mathbb{H} \to \mathbb{H}$ is said to be:

(i) *L-Lipschitz continuous*: $\|\mathcal{G}(a) - \mathcal{G}(b)\| \le L\|a-b\|$;
(ii) *η-strongly monotone*: $\langle \mathcal{G}(a) - \mathcal{G}(b), a-b\rangle \ge \eta\|a-b\|^2$;
(iii) *monotone*: $\langle \mathcal{G}(a) - \mathcal{G}(b), a-b\rangle \ge 0$;
(iv) *ν-strongly pseudomonotone*: $\langle \mathcal{G}(a), b-a\rangle \ge 0 \implies \langle \mathcal{G}(b), b-a\rangle \ge \nu\|a-b\|^2$;
(v) *pseudomonotone*: $\langle \mathcal{G}(a), b-a\rangle \ge 0 \implies \langle \mathcal{G}(b), b-a\rangle \ge 0$;
(vi) *quasimonotone*: $\langle \mathcal{G}(a), b-a\rangle > 0 \implies \langle \mathcal{G}(b), b-a\rangle \ge 0$;
(vii) *sequentially weakly continuous*: if for each $\{a_t\}$ converging weakly to a, we have $\{\mathcal{G}(a_t)\}$ converging weakly to $\mathcal{G}(a)$;
(viii) *uniformly continuous*: if $\forall\epsilon > 0$, $\exists\delta > 0$ such that $\|a-b\| < \delta \implies \|\mathcal{G}(a) - \mathcal{G}(b)\| < \epsilon$.

Remark 2. It can be check that (ii) $\implies$ (iii) $\implies$ (v), (ii) $\implies$ (iv) $\implies$ (v), and (v) $\implies$ (vi), but the converse not hold in general; see, e.g. [39, Example 1] for an example where the operator satisfies pseudomonotonicity but not monotonicity.

Definition 1. ([40, Chapter 9]) Let $\{a_t\}$ be a sequence in $\mathbb{H}$. The sequence $\{a_t\}$ is said to converge R-linearly to $a^\dagger$ with rate $\rho \in [0,1)$ if there exists a constant $c > 0$ such that for all $t \in \mathbb{N}$,

$$\|a_t - a^\dagger\| \le c\rho^t.$$

Lemma 1. *([41, Lemma 2.2]) Let $\{a_t\}$, $\{\psi_t\}$, $\{\mu_t\}$ be three sequences in $[0,+\infty)$ such that*

$$a_{t+1} \le a_t + \mu_t(a_t - a_{t-1}) + \psi_t, \quad \forall t \ge 1.$$

If $0 \le \mu_t \le \mu < 1$, $\forall t \in \mathbb{N}$ and $\sum_{t=1}^{\infty} \psi_t < \infty$, then

(i) $\sum_{t=1}^{+\infty} [a_t - a_{t-1}]_+ < +\infty$, where $[a]_+ := \max\{a, 0\}$;
(ii) there exists $a^\dagger \in [0,+\infty)$ such that $\lim_{t\to\infty} a_t = a^\dagger$.

Lemma 2. *([42]) Let $\mathbb{C} \subset \mathbb{H}$, and let $\{a_t\}$ be a sequence in $\mathbb{H}$ satisfying: (i) For every $a \in \mathbb{C}$, $\lim_{t\to\infty} \|a_t - a\|$ exists; and (ii) Every sequential weak cluster point of $\{a_t\}$ belongs to $\mathbb{C}$. Then, $\{a_t\}$ converges weakly to a point in $\mathbb{C}$.*

3. Main results

We present four algorithms that employ different step sizes at each iteration to solve pseudomonotone and quasimonotone VIPs. To begin, we assume that the proposed algorithms satisfy the following conditions.

(C1) Let $\mathbb{C} \subset \mathbb{H}$ and $\mathrm{VI}(\mathbb{C}, \mathcal{G}) \neq \emptyset$.
(C2) The mapping $\mathcal{G}\colon \mathbb{H} \to \mathbb{H}$ is pseudomonotone, L-Lipschitz continuous on $\mathbb{H}$, and sequentially weakly continuous on $\mathbb{C}$.
(C3) Assume that $\alpha_1 > 0$, $\theta \in (0,1)$, $\{\delta_t\} \subset [1,\infty)$ such that $\sum_{t=1}^{\infty}(\delta_t - 1) < +\infty$, and $\{\rho_t\} \subset [0,\infty)$ such that $\sum_{t=1}^{\infty} \rho_t < +\infty$.

3.1. The first type of double inertial SEGA

In this subsection, we present a modified version of the enhanced SEGA with double inertial, which integrates adaptive step size criteria, the relaxation technique, the SEGA, and the double inertial approach. The algorithm is outlined in Algorithm 3.1.

Algorithm 3.1 The first type of double inertial SEGA.

Initialization: Take $\theta \in (0,1)$, $\alpha_1 > 0$, $\psi \in [0,1)$, $\mu \in [0,1]$, $\zeta \in (0,1)$, and $\upsilon \in (0, 2/(1+\theta))$. Choose $\{\delta_t\}$ and $\{\rho_t\}$ satisfy Condition (C3). Let $a_0, a_1 \in \mathbb{H}$. Set $t = 1$.
Step 1. Compute

$$b_t = a_t + \psi(a_t - a_{t-1}), c_t = a_t + \mu(a_t - a_{t-1}). \tag{4}$$

Step 2. Compute $d_t = \mathbf{Proj}_{\mathbb{C}}(c_t - \alpha_t \mathcal{G}(c_t))$. If $c_t = d_t = a_t$, then $d_t \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$.
Step 3. Compute $f_t = \mathbf{Proj}_{H_t}(c_t - \upsilon\alpha_t \mathcal{G}(d_t))$, where

$$H_t := \{w \in \mathbb{H} : \langle c_t - \alpha_t\mathcal{G}(c_t) - d_t, w - d_t\rangle \le 0\}.$$

Step 4. Compute $a_{t+1} = (1-\zeta)b_t + \zeta f_t$, and update α_{t+1} by

$$\alpha_{t+1} = \begin{cases} \min\left\{ \theta \dfrac{\|c_t - d_t\|^2 + \|f_t - d_t\|^2}{2l_t}, \delta_t\alpha_t + \rho_t \right\}, & \text{if } l_t > 0, \\ \delta_t\alpha_t + \rho_t, & \text{otherwise,} \end{cases} \tag{5}$$

where $l_t := \langle \mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle$.
Set $t \leftarrow t+1$ and go to *Step 1*.

Lemma 3. *Suppose that Condition* (C3) *holds. Then $\{\alpha_t\}$ formed by Eq. (5) is well defined and $\lim_{t\to\infty} \alpha_t$ exists.*

Proof. According to the fact that $\mathcal{G}$ is L-Lipschitz continuous, one has $\|\mathcal{G}(c_t) - \mathcal{G}(d_t)\| \le L\|c_t - d_t\|$. Using Eq. (5), we have

$$\theta \frac{\|c_t - d_t\|^2 + \|f_t - d_t\|^2}{2\langle \mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle} \ge \frac{\theta}{L}.$$

From Condition (C3), it should be noted that $\delta_t \ge 1$ and $\rho_t \ge 0$. Thus

$$\alpha_{t+1} \ge \min\left\{\frac{\theta}{L}, \delta_t\alpha_t + \rho_t\right\} \ge \min\left\{\frac{\theta}{L}, \alpha_t\right\}.$$

By induction, one obtains that $\{\alpha_t\}$ is bounded. It follows from the definition of Eq. (5) that $\alpha_{t+1} \le \delta_t\alpha_t + \rho_t$ for all $t \ge 1$. This combining with Condition (C3) and [43, Lemma 1] gives that $\lim_{t\to\infty} \alpha_t$ exists, and thus $\{\alpha_t\}$ is bounded. This completes the proof. □

Lemma 4. *Assume that Condition* (C2) *holds. Let* $\{f_t\}$ *be a sequence created by Algorithm 3.1. Then*

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - v_t^\dagger\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right), \quad \forall a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G}),$$

where $v_t^\dagger = 2 - v - v\theta\alpha_t\alpha_{t+1}^{-1}$ *if* $v \in [1, 2/(1+\theta))$ *and* $v_t^\dagger = v - v\theta\alpha_t\alpha_{t+1}^{-1}$ *if* $v \in (0,1)$.

Proof. Since $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ and $d_t \in \mathbb{C}$, one sees that $\langle \mathcal{G}(a^\dagger), d_t - a^\dagger\rangle \ge 0$. According to the pseudomonotonicity of $\mathcal{G}$, one obtains

$$\langle \mathcal{G}(d_t), d_t - a^\dagger\rangle \ge 0. \tag{6}$$

From the definition of f_t, Eqs. (3) and (6), one has

$$\begin{aligned}
\left\|f_t - a^\dagger\right\|^2 &= \left\|\mathbf{Proj}_{H_t}(c_t - v\alpha_t\mathcal{G}(d_t)) - a^\dagger\right\|^2 \\
&\le \left\|c_t - v\alpha_t\mathcal{G}(d_t) - a^\dagger\right\|^2 - \|c_t - v\alpha_t\mathcal{G}(d_t) - f_t\|^2 \\
&= \left\|c_t - a^\dagger\right\|^2 - \|c_t - f_t\|^2 - 2v\alpha_t\langle \mathcal{G}(d_t), f_t - d_t\rangle \\
&\quad - 2v\alpha_t\left\langle \mathcal{G}(d_t), d_t - a^\dagger\right\rangle \\
&\le \left\|c_t - a^\dagger\right\|^2 - \|c_t - f_t\|^2 - 2v\alpha_t\langle \mathcal{G}(d_t), f_t - d_t\rangle.
\end{aligned} \tag{7}$$

According to $f_t \in H_t$ and definition of H_t, one obtains

$$\langle c_t - \alpha_t\mathcal{G}(c_t) - d_t, f_t - d_t\rangle \le 0. \tag{8}$$

According to the definition of α_{t+1} in Eq. (5), one gives

$$\langle \mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle \le \frac{\theta}{2\alpha_{t+1}}\|c_t - d_t\|^2 + \frac{\theta}{2\alpha_{t+1}}\|f_t - d_t\|^2. \tag{9}$$

Note that

$$\begin{aligned}
&\langle c_t - d_t, f_t - d_t\rangle \\
&= \langle c_t - d_t - \alpha_t\mathcal{G}(c_t) + \alpha_t\mathcal{G}(c_t) - \alpha_t\mathcal{G}(d_t) + \alpha_t\mathcal{G}(d_t), f_t - d_t\rangle \\
&= \langle c_t - \alpha_t\mathcal{G}(c_t) - d_t, f_t - d_t\rangle + \alpha_t\langle \mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle \\
&\quad + \alpha_t\langle \mathcal{G}(d_t), f_t - d_t\rangle.
\end{aligned} \tag{10}$$

From Eqs. (8)–(10), we have

$$\begin{aligned}
-\|c_t - f_t\|^2 &= -\|c_t - d_t\|^2 - \|d_t - f_t\|^2 + 2\langle c_t - d_t, f_t - d_t\rangle \\
&\le -\left(1 - \theta\alpha_t\alpha_{t+1}^{-1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right) \\
&\quad + 2\alpha_t\langle \mathcal{G}(d_t), f_t - d_t\rangle.
\end{aligned}$$

This gives

$$\begin{aligned}
-2v\alpha_t\langle \mathcal{G}(d_t), f_t - d_t\rangle &\le -v\left(1 - \theta\alpha_t\alpha_{t+1}^{-1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right) \\
&\quad + v\|c_t - f_t\|^2.
\end{aligned} \tag{11}$$

By Eqs. (7) and (11), one sees that

$$\begin{aligned}
\left\|f_t - a^\dagger\right\|^2 &\le \left\|c_t - a^\dagger\right\|^2 - v\left(1 - \theta\alpha_t\alpha_{t+1}^{-1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right) \\
&\quad - (1 - v)\|c_t - f_t\|^2.
\end{aligned} \tag{12}$$

Note that

$$\|c_t - f_t\|^2 \le 2\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right). \tag{13}$$

Next, we consider two cases of v.

Case 1: Consider $v \in (0,1)$. Combining Eqs. (12) and (13), we have

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - v\left(1 - \theta\alpha_t\alpha_{t+1}^{-1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right), \quad \forall v \in (0,1).$$

Case 2: Consider $v \in [1, 2/(1+\theta))$. It follows from Eq. (13) that

$$-(1 - v)\|c_t - f_t\|^2 \le -2(1 - v)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right), \quad \forall v \in [1, 2/(1+\theta)).$$

This together with Eq. (12) yields

$$\begin{aligned}
&\left\|f_t - a^\dagger\right\|^2 \\
&\le \left\|c_t - a^\dagger\right\|^2 - \left(2 - v - v\theta\alpha_t\alpha_{t+1}^{-1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right), \quad \forall v \in [1, 2/(1+\theta)).
\end{aligned}$$

This completes the proof of the lemma. □

Remark 3. The step size criterion Eq. (5) in Algorithm 3.1 can be replaced by the following step size criterion Eq. (14):

$$\alpha_{t+1}=\begin{cases}\min\left\{\dfrac{\theta\|c_t-d_t\|}{\|\mathcal{G}(c_t)-\mathcal{G}(d_t)\|},\ \delta_t\alpha_t+\rho_t\right\}, & \text{if } \mathcal{G}(c_t)\neq\mathcal{G}(d_t),\\ \delta_t\alpha_t+\rho_t, & \text{otherwise.}\end{cases} \tag{14}$$

Indeed, it is easy to derive from Eq. (14) that

$$\langle\mathcal{G}(c_t)-\mathcal{G}(d_t), f_t-d_t\rangle\leq\frac{\theta}{2\alpha_{t+1}}\|c_t-d_t\|^2+\frac{\theta}{2\alpha_{t+1}}\|f_t-d_t\|^2.$$

In this way, we obtain Eq. (9). The rest of the proof is the same as Lemma 4. In Eq. (14), if we set $\delta_t=1$ and $\rho_t\neq0$, then we obtain the nonmonotone step size criterion used in [29, Algorithm 1]; if we choose $\delta_t=1$ and $\rho_t=0$, then we obtain a nonincreasing step size criterion applied in many works (see, e.g., [20,28,30,31]).

Lemma 5. *Suppose that Conditions* (C1) *and* (C2) *hold. Let* $\{c_t\}$ *and* $\{d_t\}$ *be given by Algorithm 3.1. If there exists a subsequence* $\left\{c_{t_k}\right\}$ *of* $\{c_t\}$ *such that* $\left\{c_{t_k}\right\}$ *converges weakly to* $z\in\mathbb{H}$ *and* $\lim_{k\to\infty}\|c_{t_k}-d_{t_k}\|=0$, *then* $z\in\mathrm{VI}(\mathbb{C},\mathcal{G})$.

Proof. From the fact that $\left\{c_{t_k}\right\}$ is a weakly convergent sequence, one obtains that $\left\{c_{t_k}\right\}$ is bounded. This together with the Lipschitz continuity of $\mathcal{G}$ gives that $\left\{\mathcal{G}(c_{t_k})\right\}$ is bounded. By using $c_{t_k}\rightharpoonup z$ and $\lim_{k\to\infty}\left\|c_{t_k}-d_{t_k}\right\|=0$, one sees that $\left\{d_{t_k}\right\}$ is also bounded and $d_{t_k}\rightharpoonup z$. Moreover, we have $\lim_{k\to\infty}\left\|\mathcal{G}(c_{t_k})-\mathcal{G}(d_{t_k})\right\|=0$ by the fact that $\mathcal{G}$ is Lipschitz continuous on $\mathbb{H}$. From Lemma 3, one has $\alpha_{t_k}\geq\min\left\{\alpha_1,\theta/L\right\}$. From $d_{t_k}=\mathbf{Proj}_{\mathbb{C}}\left(c_{t_k}-\alpha_{t_k}\mathcal{G}(c_{t_k})\right)$ and Eq. (2), we have

$$\left\langle c_{t_k}-\alpha_{t_k}\mathcal{G}(c_{t_k})-d_{t_k}, a-d_{t_k}\right\rangle\leq0,\quad\forall a\in\mathbb{C}.$$

Thus we obtain

$$\alpha_{t_k}^{-1}\left\langle c_{t_k}-d_{t_k}, a-d_{t_k}\right\rangle+\left\langle\mathcal{G}(c_{t_k}), d_{t_k}-c_{t_k}\right\rangle\leq\left\langle\mathcal{G}(c_{t_k}), a-c_{t_k}\right\rangle,\quad\forall a\in\mathbb{C}.$$

Then

$$\liminf_{k\to\infty}\left\langle\mathcal{G}(c_{t_k}), a-c_{t_k}\right\rangle\geq0,\quad\forall a\in\mathbb{C}. \tag{15}$$

Note that

$$\begin{aligned}\left\langle\mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle=&\left\langle\mathcal{G}(d_{t_k})-\mathcal{G}(c_{t_k}), a-c_{t_k}\right\rangle+\left\langle\mathcal{G}(c_{t_k}), a-c_{t_k}\right\rangle\\&+\left\langle\mathcal{G}(d_{t_k}), c_{t_k}-d_{t_k}\right\rangle.\end{aligned} \tag{16}$$

From Eqs. (15) and (16), one obtains

$$\liminf_{k\to\infty}\left\langle\mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle\geq0,\quad\forall a\in\mathbb{C}.$$

Next, we define a decreasing positive sequence $\{\epsilon_k\}$ such that $\lim_{k\to\infty}\epsilon_k=0$. For each $k\geq1$, let N_k denote the smallest positive integer such that

$$\left\langle\mathcal{G}(d_{t_j}), a-d_{t_j}\right\rangle+\epsilon_k\geq0,\quad\forall j\geq N_k. \tag{17}$$

Since $\{\epsilon_k\}$ is decreasing, one obtains that $\{N_k\}$ is increasing. In addition, for each $k\geq1$, by using $\left\{d_{N_k}\right\}\subset\mathbb{C}$ we can suppose that $\mathcal{G}(d_{N_k})\neq0$ (otherwise, $d_{N_k}\in\mathrm{VI}(\mathbb{C},\mathcal{G})$). Set

$$v_{N_k}:=\frac{\mathcal{G}(d_{N_k})}{\left\|\mathcal{G}(d_{N_k})\right\|^2}.$$

Then we have

$$\left\langle\mathcal{G}(d_{N_k}), v_{N_k}\right\rangle=1,\quad\forall k\geq1. \tag{18}$$

According to $d_{t_k}\rightharpoonup z$ and $\{d_t\}\subset\mathbb{C}$, one has $z\in\mathbb{C}$. Assume that $\mathcal{G}(z)\neq0$. Since $\mathcal{G}$ is sequentially weakly continuous on $\mathbb{C}$, one obtains $\left\{\mathcal{G}(d_{t_k})\right\}$ converges weakly to $\mathcal{G}(z)$. By the fact that norm mapping is sequentially weakly lower semicontinuous, one arrives at

$$0<\|\mathcal{G}(z)\|\leq\liminf_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|.$$

Since $\left\{d_{N_k}\right\} \subset \left\{d_{t_k}\right\}$ and $\lim_{t\to\infty} \epsilon_k = 0$, one has

$$0 \le \limsup_{k\to\infty} \left\|\epsilon_k v_{N_k}\right\| = \limsup_{k\to\infty} \frac{\epsilon_k}{\left\|\mathcal{G}(d_{t_k})\right\|} \le \frac{\limsup_{k\to\infty} \epsilon_k}{\liminf_{k\to\infty} \left\|\mathcal{G}(d_{t_k})\right\|} = 0,$$

which implies that

$$\lim_{k\to\infty} \epsilon_k v_{N_k} = 0. \tag{19}$$

Now, we can deduce from Eqs. (17) and (18) that

$$\left\langle \mathcal{G}(d_{N_k}), a + \epsilon_k v_{N_k} - d_{N_k} \right\rangle \ge 0, \quad \forall k \ge 1.$$

Since $\mathcal{G}$ is pseudomonotone on $\mathbb{H}$, one has $\left\langle \mathcal{G}\left(a + \epsilon_k v_{N_k}\right), a + \epsilon_k v_{N_k} - d_{N_k} \right\rangle \ge 0$. Thus

$$\left\langle \mathcal{G}(a), a - d_{N_k} \right\rangle \ge \left\langle \mathcal{G}(a) - \mathcal{G}\left(a + \epsilon_k v_{N_k}\right), a + \epsilon_k v_{N_k} - d_{N_k} \right\rangle - \left\langle \mathcal{G}(a), \epsilon_k v_{N_k} \right\rangle. \tag{20}$$

Combining the boundedness of $\left\{d_{N_k}\right\}$, Eqs. (19), (20), and the Lipschitz continuity of $\mathcal{G}$, one has

$$\liminf_{k\to\infty} \left\langle \mathcal{G}(a), a - d_{N_k} \right\rangle \ge 0, \quad \forall a \in \mathbb{C}.$$

Hence we have

$$\langle \mathcal{G}(a), a - z \rangle = \lim_{k\to\infty} \left\langle \mathcal{G}(a), a - d_{N_k} \right\rangle = \liminf_{k\to\infty} \left\langle \mathcal{G}(a), a - d_{N_k} \right\rangle \ge 0,$$

which together with [1, Lemma 2.1] shows $z \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$. This finishes the proof. □

Remark 4. From the proof of Lemma 5, we observe that the weak continuity requirement of the sequence in Condition (C2) can be substituted with the following condition:

$$\text{whenever } \left\{a_t\right\} \subset \mathbb{H} \text{ and } a_t \rightharpoonup v^\dagger, \text{ one has } \left\|\mathcal{G}(v^\dagger)\right\| \le \liminf_{t\to\infty} \left\|\mathcal{G}(a_t)\right\|. \tag{21}$$

To the best of our knowledge, we know that Assumption (21) has been first employed in [39], and then applied in many paper (see, e.g., [28–31]). It can be verified that Condition (21) is a strictly weaker requirement than the assumption of sequential weak continuity. For example, if we define $\mathcal{G}(a) = a\|a\|$ for all $a \in \mathbb{C}$, the operator $\mathcal{G}$ satisfies Eq. (21) but does not exhibit sequential weak continuity (see [39, Remark 3.2] for further details). Additionally, if the operator $\mathcal{G}$ is either monotone or strongly pseudomonotone, the weak continuity assumption of the sequence in Condition (C2) or Eq. (21) can be omitted (see [14, Remark 3.4]).

Theorem 1. *Suppose that Conditions* (C1)–(C3) *hold. Let* $v \in \left(0, \frac{2}{1+\theta}\right)$*, and let parameters* μ*,* ψ*, and* ζ *fulfill the following condition*

$$\begin{aligned} &\psi \in [0,1), \quad \mu \in [0,1], \quad \zeta \in (0,\zeta_1) \text{ if } \psi \ne \mu \text{ and } \zeta \in (0,\zeta_2) \text{ if } \psi = \mu, \\ &\text{where } \zeta_1 = \frac{b - \sqrt{b^2 - 4ac}}{2a}, \quad \zeta_2 = \frac{c}{b}, \\ &a = \psi(1+\psi) - \mu(1+\mu),\ b = 1 + 2\psi^2 - \psi,\ c = (1-\psi)^2. \end{aligned} \tag{22}$$

Then the sequence $\left\{a_t\right\}$ *generated by Algorithm 3.1 converges weakly to* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$*. Moreover, there exist positive constants* M_1 *and* β *such that*

$$\min_{t_0 \le k \le t} \left\|c_k - d_k\right\| \le \left(\frac{\Gamma_{t_0} + \frac{1}{1-\eta}\left[\Psi_{t_0}\right]_+ + \frac{1}{1-\eta} M_1}{\zeta\beta(t - t_0 + 1)} \right)^{1/2}, \tag{23}$$

where $\Gamma_t := \left\|a_t - a^\dagger\right\|^2$*,* $\Psi_t := \Gamma_t - \Gamma_{t-1}$*.*

Proof. Since $\lim_{t\to\infty} \alpha_t$ exists, and by Lemma 4 we have

$$\lim_{t\to\infty} v_t^\dagger = \begin{cases} 2 - v - v\theta, & \text{if } v \in \left[1, 2/(1+\theta)\right), \\ v - v\theta, & \text{if } v \in (0,1). \end{cases}$$

Therefore, we have $\lim_{t\to\infty} v_t^\dagger \in (0,1)$ for all $v \in \left(0, 2/(1+\theta)\right)$. From Lemma 4, there exists a natural number $t_0 \ge 1$ such that

$$\left\|f_t - a^\dagger\right\| \le \left\|c_t - a^\dagger\right\|, \quad \forall t \ge t_0. \tag{24}$$

Using the definition of a_{t+1}, one gives

$$\|f_t - b_t\| = \zeta^{-1}\|a_{t+1} - b_t\|. \tag{25}$$

From Eqs. (1), (24), and (25), one has

$$\begin{aligned}\left\|a_{t+1} - a^{\dagger}\right\|^2 &= \left\|(1-\zeta)\left(b_t - a^{\dagger}\right) + \zeta\left(f_t - a^{\dagger}\right)\right\|^2\\ &= (1-\zeta)\left\|b_t - a^{\dagger}\right\|^2 + \zeta\left\|f_t - a^{\dagger}\right\|^2 - \zeta(1-\zeta)\|b_t - f_t\|^2\\ &\le (1-\zeta)\left\|b_t - a^{\dagger}\right\|^2 + \zeta\left\|c_t - a^{\dagger}\right\|^2 - \zeta^{-1}(1-\zeta)\|a_{t+1} - b_t\|^2.\end{aligned} \tag{26}$$

By c_t and Eq. (1), one has

$$\begin{aligned}\left\|c_t - a^{\dagger}\right\|^2 &= \left\|(1+\mu)\left(a_t - a^{\dagger}\right) - \mu\left(a_{t-1} - a^{\dagger}\right)\right\|^2\\ &= (1+\mu)\left\|a_t - a^{\dagger}\right\|^2 - \mu\left\|a_{t-1} - a^{\dagger}\right\|^2 + \mu(1+\mu)\|a_t - a_{t-1}\|^2.\end{aligned} \tag{27}$$

According to the definition of b_t and Eq. (1), one sees that

$$\left\|b_t - a^{\dagger}\right\|^2 = (1+\psi)\left\|a_t - a^{\dagger}\right\|^2 - \psi\left\|a_{t-1} - a^{\dagger}\right\|^2 + \psi(1+\psi)\|a_t - a_{t-1}\|^2. \tag{28}$$

Using the Cauchy-Schwarz inequality, one obtains

$$\begin{aligned}&\|a_{t+1} - b_t\|^2\\ &= \|a_{t+1} - a_t\|^2 + \psi^2\|a_t - a_{t-1}\|^2 - 2\psi\left\langle a_{t+1} - a_t, a_t - a_{t-1}\right\rangle\\ &\ge \|a_{t+1} - a_t\|^2 + \psi^2\|a_t - a_{t-1}\|^2 - 2\psi\|a_{t+1} - a_t\|\|a_t - a_{t-1}\|\\ &\ge (1-\psi)\|a_{t+1} - a_t\|^2 + \left(\psi^2 - \psi\right)\|a_t - a_{t-1}\|^2.\end{aligned} \tag{29}$$

Substituting Eqs. (27)–(29) into Eq. (26), we have

$$\begin{aligned}\left\|a_{t+1} - a^{\dagger}\right\|^2 \le{}& \left\|a_t - a^{\dagger}\right\|^2 + \eta\left(\left\|a_t - a^{\dagger}\right\|^2 - \left\|a_{t-1} - a^{\dagger}\right\|^2\right)\\ &- \rho\|a_{t+1} - a_t\|^2 + \sigma\|a_t - a_{t-1}\|^2, \quad \forall t \ge t_0,\end{aligned} \tag{30}$$

where

$$\begin{aligned}&\eta := \zeta\mu + \psi(1-\zeta), \quad \rho := \zeta^{-1}(1-\zeta)(1-\psi),\\ &\sigma := (1-\zeta)\psi(1+\psi) + \zeta\mu(1+\mu) - \zeta^{-1}(1-\zeta)\left(\psi^2 - \psi\right).\end{aligned} \tag{31}$$

According to the condition on ζ in Eq. (22) and the graph of the function $f(\zeta)$ (see Eq. (34) below), it can be verified that $\zeta_1 < 1$ and $\zeta_2 < 1$ for any $\psi \in [0,1)$ and $\mu \in [0,1]$. Thus we have $\eta < \max\{\mu, \psi\} \le 1$, $\rho > 0$, and $\sigma > 0$.

Define

$$\Theta_t := \left\|a_t - a^{\dagger}\right\|^2 - \eta\left\|a_{t-1} - a^{\dagger}\right\|^2 + \sigma\|a_t - a_{t-1}\|^2, \quad \forall t \ge t_0.$$

By using Eq. (30), one obtains

$$\Theta_{t+1} - \Theta_t \le -(\rho - \sigma)\|a_{t+1} - a_t\|^2. \tag{32}$$

It follows from the definitions of ρ and σ in Eq. (31) that

$$\begin{aligned}\rho - \sigma &= \zeta^{-1}(1-\zeta)(1-\psi)^2 - (1-\zeta)\psi(1+\psi) - \zeta\mu(1+\mu)\\ &= \zeta^{-1}\left(\left(\psi(1+\psi) - \mu(1+\mu)\right)\zeta^2 - \left((1-\psi)^2 + \psi(1+\psi)\right)\zeta + (1-\psi)^2\right).\end{aligned} \tag{33}$$

Let

$$\begin{aligned}&f(\zeta) := a\zeta^2 - b\zeta + c,\\ &\text{where } a = \psi(1+\psi) - \mu(1+\mu),\ b = (1-\psi)^2 + \psi(1+\psi),\ c = (1-\psi)^2.\end{aligned} \tag{34}$$

It follows from $\psi \in [0,1)$ and $\mu \in [0,1]$ that

$$\Delta = b^2 - 4ac = (1-3\psi)^2 + 4\mu(1+\mu)(1-\psi)^2 \ge 0.$$

In addition, noting that $f(0) > 0$ and $f(1) \le 0$. By using the condition of ζ in Eq. (22), one has $\rho - \sigma > 0$. Therefore, from Eqs. (32) and (33), we have

$$\Theta_{t+1} - \Theta_t \le -\omega\|a_{t+1} - a_t\|^2, \quad \forall t \ge t_0, \tag{35}$$

where $\omega := \rho - \sigma > 0$. Therefore, the sequence $\{\Theta_t\}$ is nonincreasing. From the definition of Θ_t and note that $\sigma > 0$, one sees that

$$\Theta_t \ge \left\|a_t - a^{\dagger}\right\|^2 - \eta\left\|a_{t-1} - a^{\dagger}\right\|^2.$$

Thus

$$\begin{aligned}\left\|a_t - a^\dagger\right\|^2 &\le \eta\left\|a_{t-1} - a^\dagger\right\|^2 + \Theta_t \le \eta\left\|a_{t-1} - a^\dagger\right\|^2 + \Theta_{t_0}\\ &\le \eta^{t-t_0}\left\|a_{t_0} - a^\dagger\right\|^2 + \Theta_{t_0}\left(\eta^{t-t_0-1} + \cdots + 1\right)\\ &\le \eta^{t-t_0}\left\|a_{t_0} - a^\dagger\right\|^2 + \frac{\Theta_{t_0}}{1-\eta}.\end{aligned} \tag{36}$$

Consequently, the sequence $\left\{\left\|a_t - a^\dagger\right\|\right\}$ is bounded and thus $\{a_t\}$ is also bounded. From the definition of Θ_{t+1} and note that $\sigma > 0$, one has $\Theta_{t+1} \ge -\eta\left\|a_t - a^\dagger\right\|^2$. This together with Eq. (36) implies

$$-\Theta_{t+1} \le \eta\left\|a_t - a^\dagger\right\|^2 \le \eta^{t-t_0+1}\left\|a_{t_0} - a^\dagger\right\|^2 + \frac{\eta\Theta_{t_0}}{1-\eta},$$

which combining with Eq. (35) gives

$$\begin{aligned}\omega\sum_{k=t_0}^{t}\|a_{k+1} - a_k\|^2 &\le \Theta_{t_0} - \Theta_{t+1}\\ &\le \eta^{t-t_0+1}\left\|a_{t_0} - a^\dagger\right\|^2 + \frac{\Theta_{t_0}}{1-\eta} \le \left\|a_{t_0} - a^\dagger\right\|^2 + \frac{\Theta_{t_0}}{1-\eta}.\end{aligned}$$

Therefore,

$$\sum_{t=1}^{\infty}\|a_{t+1} - a_t\|^2 < +\infty. \tag{37}$$

From Eqs. (30), (37), and Lemma 1, one sees that $\lim_{t\to\infty}\left\|a_t - a^\dagger\right\|$ exists. Owing to Eq. (37), one has

$$\lim_{t\to\infty}\|a_{t+1} - a_t\| = 0. \tag{38}$$

By using the definitions of b_t and c_t, and Eq. (38), one obtains

$$\lim_{t\to\infty}\|a_t - b_t\| = 0, \quad \lim_{t\to\infty}\|a_t - c_t\| = 0. \tag{39}$$

By Eqs. (38) and (39), one has

$$\lim_{t\to\infty}\|a_{t+1} - b_t\| = 0, \quad \lim_{t\to\infty}\|a_{t+1} - c_t\| = 0, \quad \lim_{t\to\infty}\|c_t - b_t\| = 0. \tag{40}$$

According to the definition of a_{t+1}, one arrives at

$$\|f_t - b_t\| = \zeta^{-1}\|a_{t+1} - b_t\| \to 0, \quad \text{as } t \to \infty. \tag{41}$$

By using Eqs. (40) and (41), one gives $\lim_{t\to\infty}\|c_t - f_t\| = 0$. This together with Lemma 4 yields

$$\begin{aligned}v_t^\dagger\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right) &\le \left\|c_t - a^\dagger\right\|^2 - \left\|f_t - a^\dagger\right\|^2\\ &= \left(\left\|c_t - a^\dagger\right\| + \left\|f_t - a^\dagger\right\|\right)\left(\left\|c_t - a^\dagger\right\| - \left\|f_t - a^\dagger\right\|\right)\\ &\le M_0\|c_t - f_t\| \to 0, \quad \text{as } t \to \infty\end{aligned}$$

for some $M_0 > 0$. Thanks to $\lim_{t\to\infty} v_t^\dagger > 0$, we obtain

$$\lim_{t\to\infty}\|c_t - d_t\| = 0, \quad \lim_{t\to\infty}\|f_t - d_t\| = 0. \tag{42}$$

From Eq. (36), one gives that $\{a_t\}$ is bounded. Let $v^\dagger$ be a weak cluster point of $\{a_t\}$. Now we can select a subsequence $\left\{a_{t_k}\right\}$ of $\{a_t\}$ such that $a_{t_k} \rightharpoonup v^\dagger$. From Eqs. (39) and (42), we obtain $c_{t_k} \rightharpoonup v^\dagger$ and $\lim_{k\to\infty}\|c_{t_k} - d_{t_k}\| = 0$, respectively. Those together with Lemma 5 mean that $v^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$. That is to say, every sequential weak cluster point of $\{a_t\}$ is in $\mathrm{VI}(\mathbb{C}, \mathcal{G})$. This combining with $\lim_{t\to\infty}\|a_t - a^\dagger\|$ exists for any $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ and Lemma 2 concludes that $\{a_t\}$ converges weakly to a point in $\mathrm{VI}(\mathbb{C}, \mathcal{G})$.

Next we prove inequality Eq. (23). Let $\beta < \lim_{t\to\infty} v_t^\dagger$. By using Lemma 4, one has

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - v_t^\dagger\|c_t - d_t\|^2.$$

Then we have

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - \beta\|c_t - d_t\|^2, \quad \forall t \ge t_0.$$

Based on a similar proof process in Eqs. (25)–(30), one arrives at

$$\begin{aligned}\left\|a_{t+1} - a^\dagger\right\|^2 \le&\left\|a_t - a^\dagger\right\|^2 + \eta\left(\left\|a_t - a^\dagger\right\|^2 - \left\|a_{t-1} - a^\dagger\right\|^2\right)\\ &+ \sigma\|a_t - a_{t-1}\|^2 - \zeta\beta\|c_t - d_t\|^2,\end{aligned} \tag{43}$$

where η and σ are defined in Eq. (31). Let $\Gamma_t := \left\|a_t - a^\dagger\right\|^2$, $\Psi_t := \Gamma_t - \Gamma_{t-1}$, and $\Xi_t := \sigma\|a_t - a_{t-1}\|^2$. Thanks to Eq. (37), one sees that there exists a positive constant M satisfies $\sum_{t=1}^{\infty}\|a_{t+1} - a_t\|^2 \le M$. Then we obtain $\sum_{t=1}^{\infty}\Xi_t \le M_1$, where $M_1 = \sigma M$. By using Eq. (43) and noticing that $\zeta\beta > 0$, one obtains

$$\Psi_{t+1} \le \eta\Psi_t + \Xi_t \le \eta\left[\Psi_t\right]_+ + \Xi_t.$$

Thus we deduce

$$\left[\Psi_{t+1}\right]_+ \le \eta\left[\Psi_t\right]_+ + \Xi_t \le \eta^{t-t_0+1}\left[\Psi_{t_0}\right]_+ + \sum_{k=t_0}^{t}\eta^{t-k}\Xi_k.$$

Consequently,

$$\begin{aligned}
\sum_{t=t_0}^{\infty}\left[\Psi_{t+1}\right]_+ &\le \sum_{t=t_0}^{\infty}\eta^{t-t_0+1}\left[\Psi_{t_0}\right]_+ + \sum_{t=t_0}^{\infty}\sum_{k=t_0}^{t}\eta^{t-k}\Xi_k \\
&\le \frac{\eta}{1-\eta}\left[\Psi_{t_0}\right]_+ + \frac{1}{1-\eta}\sum_{t=t_0}^{\infty}\Xi_t \le \frac{\eta}{1-\eta}\left[\Psi_{t_0}\right]_+ + \frac{1}{1-\eta}M_1.
\end{aligned} \tag{44}$$

It follows from Eq. (43) that

$$\zeta\beta\|c_t - d_t\|^2 \le \Gamma_t - \Gamma_{t+1} + \eta\Psi_t + \Xi_t \le \Gamma_t - \Gamma_{t+1} + \eta\left[\Psi_t\right]_+ + \Xi_t.$$

This together with Eq. (44) yields (noting that $\eta < 1$ by means of Eq. (31)

$$\begin{aligned}
\zeta\beta\sum_{k=t_0}^{t}\|d_k - c_k\|^2 &\le \Gamma_{t_0} - \Gamma_{t+1} + \eta\sum_{k=t_0}^{t}\left[\Psi_k\right]_+ + \sum_{k=t_0}^{t}\Xi_k \\
&\le \Gamma_{t_0} + \left[\Psi_{t_0}\right]_+ + \eta\sum_{k=t_0}^{t}\left[\Psi_{k+1}\right]_+ + M_1 \\
&\le \Gamma_{t_0} + \left[\Psi_{t_0}\right]_+ + \frac{\eta^2}{1-\eta}\left[\Psi_{t_0}\right]_+ + \frac{\eta}{1-\eta}M_1 + M_1 \\
&\le \Gamma_{t_0} + \frac{1}{1-\eta}\left[\Psi_{t_0}\right]_+ + \frac{1}{1-\eta}M_1.
\end{aligned}$$

Hence we conclude

$$\min_{t_0\le k\le t}\|c_k - d_k\|^2 \le \frac{\Gamma_{t_0} + \frac{1}{1-\eta}\left[\Psi_{t_0}\right]_+ + \frac{1}{1-\eta}M_1}{\zeta\beta(t - t_0 + 1)}.$$

This is the desired inequality Eq. (23) by a simple transformation. □

Remark 5. In Theorem 1, we establish the nonasymptotic $O(1/t)$ convergence rate of Algorithm 3.1 as given in Eq. (23). This result relies on the observation that if $c_t = d_t$, then d_t serves as a solution to Problem (VI). Additionally, it is important to note that our Condition Eq. (22) is weaker than the conditions used in [28,32,34]. Specifically, the algorithm proposed by Yao et al. [28] uses the following condition:

$$\mu \in [0,1], \quad \psi \in \left(0, \frac{\epsilon - \sqrt{2\epsilon}}{\epsilon}\right), \quad \zeta \in \left(0, \frac{1}{1+\epsilon}\right), \text{where } \epsilon \in (2,\infty).$$

Together this condition with Eq. (33), one obtains

$$\begin{aligned}
\rho - \sigma &= \zeta^{-1}(1-\zeta)(1-\psi)^2 - (1-\zeta)\psi(1+\psi) - \zeta\mu(1+\mu) \\
&> \zeta^{-1}(1-\zeta)(1-\psi)^2 - 2(1-\zeta) - 2\zeta \\
&> \epsilon(1-\psi)^2 - 2 > 0.
\end{aligned}$$

Thong et al. [32, Alg. 1] apply the following condition:

$$\mu \in [0,1], \quad \psi \in [0,1), \quad \zeta \in \left(0, \frac{(1-\psi)^2}{(1-\psi)^2 + \max\{\psi(1+\psi), \mu(1+\mu)\}}\right).$$

This condition combining with Eq. (33) yields

$$\rho - \sigma \ge \zeta^{-1}(1-\psi)^2 - (1-\psi)^2 - \max\{\psi(1+\psi), \mu(1+\mu)\} > 0.$$

Wang et al. [34, Alg. 3.1] take the following condition:

$$\mu \in [0,1], \quad \psi \in \left(0, \frac{3 + 2\epsilon - \sqrt{8\epsilon + 17}}{2\epsilon}\right), \quad \zeta \in \left(0, \frac{1}{1+\epsilon}\right], \text{where } \epsilon \in (1,\infty).$$

Alongside this condition and Eq. (33), one has

$$\begin{aligned}\rho-\sigma &\geq \zeta^{-1}(1-\zeta)(1-\psi)^2-(1-\zeta)\psi(1+\psi)-2\zeta\\ &\geq \epsilon\left(1-2\psi+\psi^2\right)-\frac{\epsilon}{1+\epsilon}\left(\psi+\psi^2\right)-\frac{2}{1+\epsilon}\\ &=\frac{1}{1+\epsilon}\left[\epsilon^2\psi^2-\left(3\epsilon+2\epsilon^2\right)\psi+\left(\epsilon^2+\epsilon-2\right)\right]>0.\end{aligned}$$

In particular, we did not use any bounding techniques when solving for the range of ζ based on $\rho-\sigma>0$ in Eq. (33). Therefore, our convergence conditions are significantly weaker than those of the algorithms in [28,32,34]. Indeed, Yao et al. [28] require ζ to be less than $1/3$, while Wang et al. [34] extend the range of ζ but require $\zeta \in (0,1/2)$. Additionally, the range of ζ in Thong et al. [32] is restricted, especially when $\psi \ll \mu$. To illustrate this point more clearly, Table 2 provides the values of the parameter ζ for our Algorithm 3.1 and the algorithms in [28,32,34] when ψ and μ are fixed.

Table 2
Compare the values of the parameter ζ for different algorithms when ψ and μ are fixed.

ψ	μ					
	0	0.2	0.4	0.6	0.8	1.0
0	1.000 (1.000)	0.833 (0.806)	0.714 (0.641)	0.625 (0.510)	0.556 (0.410)	0.500 (0.333)
0.1	1.000 (0.880)	0.792 (0.771)	0.664 (0.591)	0.575 (0.458)	0.508 (0.360)	0.455 (0.288)
0.2	1.000 (0.727)	0.727 (0.727)	0.597 (0.533)	0.512 (0.400)	0.451 (0.308)	0.403 (0.242)
0.4	0.643 (0.391)	0.467 (0.391)	0.391 (0.391)	0.341 (0.273)	0.303 (0.200)	0.274 (0.153)
0.6	0.167 (0.143)	0.159 (0.143)	0.151 (0.143)	0.143 (0.143)	0.135 (0.100)	0.128 (0.074)
0.8	0.028 (0.027)	0.028 (0.027)	0.027 (0.027)	0.027 (0.027)	0.027 (0.027)	0.027 (0.020)

Note: the first value in each cell represents our algorithms, and the second is Thong et al. [32].

ϵ	ψ	ζ	ψ	ζ
	Wang et al. [34]		Yao et al. [28]	
1.01	0.006	0.498	–	–
2.01	0.316	0.332	0.002	0.332
4.01	0.501	0.200	0.294	0.200
6.01	0.578	0.143	0.423	0.143
8.01	0.625	0.111	0.500	0.111
10.01	0.658	0.091	0.553	0.091

From Table 2, it is easy to see that our parameter ζ allows for larger values compared to those in [28,32,34]. This is important, as demonstrated by the subsequent numerical experiments (see Section 4 for more information).

For the subsequent convergence analysis, we require the following conditions:

(C1′) Let $\mathbb{C} \subset \mathbb{H}$ and $\mathrm{DVI}(\mathbb{C},\mathcal{G}) \neq \emptyset$.
(C2′) The mapping $\mathcal{G}: \mathbb{H} \to \mathbb{H}$ is quasimonotone, L-Lipschitz continuous on $\mathbb{H}$, and sequentially weakly continuous on $\mathbb{C}$.
(C2″) The mapping $\mathcal{G}: \mathbb{H} \to \mathbb{H}$ is ν-strongly pseudomonotone and L-Lipschitz continuous on $\mathbb{H}$.

Next, we prove the weak convergence of Algorithm 3.1 under the condition that the operator $\mathcal{G}$ is quasimonotone. Similar to Lemma 5, we first establish the following result using [16, Lemma 3.3].

Lemma 6. *Assume that Conditions* (C1′) *and* (C2′) *hold. Let* $\{c_t\}$ *be created by Algorithm 3.1. If there exists a subsequence* $\{c_{t_k}\}$ *convergent weakly to* $z \in \mathbb{H}$ *and* $\lim_{k\to\infty}\|c_{t_k}-d_{t_k}\|=0$, *then* $z \in \mathrm{DVI}(\mathbb{C},\mathcal{G})$ *or* $\mathcal{G}(z)=0$.

Proof. This proof is similar to the one of Lemma 5 and we divide it into two parts.

Case 1: Consider $\limsup_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|=0$. It follows that

$$\lim_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|=\liminf_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|=0.$$

Combining $d_{t_k} \rightharpoonup z$ and the sequential weak continuity of $\mathcal{G}$ on $\mathbb{C}$, we obtain $\mathcal{G}(d_{t_k}) \rightharpoonup \mathcal{G}(z)$. Note that

$$0 \leq \|\mathcal{G}(z)\| \leq \liminf_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|=0.$$

Thus $\mathcal{G}(z)=0$.

Case 2: Consider $\limsup_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|>0$. For convenience, we set $\lim_{k\to\infty}\left\|\mathcal{G}(d_{t_k})\right\|=M>0$. Then there exists a constant $K \in \mathbb{N}$ such that

$$\left\|\mathcal{G}(d_{t_k})\right\|>\frac{M}{2}, \quad \forall k \geq K. \tag{45}$$

By Lemma 5, one has

$$\liminf_{k\to\infty}\left\langle \mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle \geq 0, \quad \forall a\in\mathbb{C}. \tag{46}$$

We complete the rest proof by considering the following two cases.

Case 2(i): If $\limsup_{k\to\infty}\left\langle \mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle > 0$, then there exists a subsequence $\{d_{t_{k_j}}\}$ of $\{d_{t_k}\}$ such that $\lim_{j\to\infty}\langle \mathcal{G}(d_{t_{k_j}}), a-d_{t_{k_j}}\rangle > 0$. Therefore, there exists $j_0\in\mathbb{N}$ such that $\langle \mathcal{G}(d_{t_{k_j}}), a-d_{t_{k_j}}\rangle > 0$ for all $j\geq j_0$. This together with the quasimonotonicity of $\mathcal{G}$ yields that $\langle \mathcal{G}(a), a-d_{t_{k_j}}\rangle > 0$ for all $j\geq j_0$. Letting $j\to\infty$, one obtain $\langle \mathcal{G}(a), a-z\rangle > 0$ for all $a\in\mathbb{C}$ and thus $z\in \mathrm{DVI}(\mathbb{C},\mathcal{G})$.

Case 2(ii): Consider $\limsup_{k\to\infty}\left\langle \mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle = 0$. By Eq. (46), one has

$$\lim_{k\to\infty}\left\langle \mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle = 0.$$

Let $\epsilon_k := \left|\left\langle \mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle\right| + \frac{1}{k+1}$. Therefore,

$$\left\langle \mathcal{G}(d_{t_k}), a-d_{t_k}\right\rangle + \epsilon_k > 0, \quad \forall k\geq 1. \tag{47}$$

Since $\left\{d_{t_k}\right\}\subset\mathbb{C}$, we can assume that $\mathcal{G}(d_{t_k})\neq 0$ for each $k\geq 1$. Set

$$p_{t_k} := \frac{\mathcal{G}(d_{t_k})}{\left\|\mathcal{G}(d_{t_k})\right\|^2}.$$

We have $\left\langle \mathcal{G}(d_{t_k}), p_{t_k}\right\rangle = 1$ for each $k\geq 1$. From Eq. (47), one has

$$\left\langle \mathcal{G}(d_{t_k}), a+\epsilon_k p_{t_k} - d_{t_k}\right\rangle > 0.$$

Since $\mathcal{G}$ is quasimonotone, we obtain $\left\langle \mathcal{G}\left(a+\epsilon_k p_{t_k}\right), a+\epsilon_k p_{t_k} - d_{t_k}\right\rangle \geq 0$. This together with the Lipschitz continuity of $\mathcal{G}$, the definition of p_{t_k}, and Eq. (45) yields

$$\begin{aligned}\left\langle \mathcal{G}(a), a+\epsilon_k p_{t_k} - d_{t_k}\right\rangle &\geq \left\langle \mathcal{G}(a)-\mathcal{G}\left(a+\epsilon_k p_{t_k}\right), a+\epsilon_k p_{t_k} - d_{t_k}\right\rangle \\ &\geq -L\epsilon_k\left\|p_{t_k}\right\|\left\|a+\epsilon_k p_{t_k} - d_{t_k}\right\| \\ &\geq -2L\epsilon_k M^{-1}\left\|a+\epsilon_k p_{t_k} - d_{t_k}\right\|, \quad \forall k\geq K.\end{aligned} \tag{48}$$

Note that $\lim_{k\to\infty}\epsilon_k = 0$, which combining with the boundedness of $\left\{\left\|a+\epsilon_k p_{t_k} - d_{t_k}\right\|\right\}$, $d_{t_k}\rightharpoonup z$, and Eq. (48) implies that $\langle \mathcal{G}(a), a-z\rangle \geq 0$ for all $a\in\mathbb{C}$. That is $z\in \mathrm{DVI}(\mathbb{C},\mathcal{G})$. □

Theorem 2. *Suppose that Conditions* (C1′), (C2′), *and* (C3) *hold. Let $\upsilon\in\left(0, 2/(1+\theta)\right)$ and let Eq. (22) holds. Assume that $\mathcal{G}(a)\neq 0$ for all $a\in\mathbb{C}$ (otherwise $a\in\mathrm{VI}(\mathbb{C},\mathcal{G})$). Then the sequence $\left\{a_t\right\}$ generated by Algorithm 3.1 converges weakly to $a^\dagger\in\mathrm{DVI}(\mathbb{C},\mathcal{G})$. Moreover, we have the error estimate Eq. (23).*

Proof. Let $a^\dagger\in\mathrm{DVI}(\mathbb{C},\mathcal{G})$. It follows from $d_t\in\mathbb{C}$ that $\langle \mathcal{G}(d_t), d_t-a^\dagger\rangle \geq 0$. The detailed steps are omitted here for brevity, as they follow closely from the arguments presented in Lemma 4 and Theorem 1 (noting that we need to replace Lemma 5 with Lemma 6). □

Next, we verify two strong convergence theorems of Algorithm 3.1 under the strong pseudomonotonicity of the operator $\mathcal{G}$ and different parameter conditions.

Theorem 3. *Suppose that Conditions* (C1), (C2″), *and* (C3) *hold. Let $\upsilon\in\left(0, 2/(1+\theta)\right)$ and let Eq. (22) holds. Then $\left\{a_t\right\}$ generated by Algorithm 3.1 converges strongly to the unique solution $a^\dagger\in\mathrm{VI}(\mathbb{C},\mathcal{G})$. Moreover, we have the error estimate Eq. (23) and the global bound*

$$\frac{1-\alpha_t L}{1+\alpha_t L}\|c_t-d_t\| \leq \|c_t-a^\dagger\| \leq \left(1+\frac{1+\alpha_t L}{\nu\alpha_t}\right)\|c_t-d_t\|. \tag{49}$$

Proof. Under Condition (C2″), Problem (VI) admits a unique solution, denoted by $a^\dagger$ (see [44, Theorem 2.1]). Since $a^\dagger\in\mathrm{VI}(\mathbb{C},\mathcal{G})$ and $d_t\in\mathbb{C}$, it follows that

$$\langle \mathcal{G}(a^\dagger), d_t-a^\dagger\rangle \geq 0.$$

By the ν-strong pseudomonotonicity of $\mathcal{G}$, we further have

$$\langle \mathcal{G}(d_t), d_t-a^\dagger\rangle \geq \nu\|d_t-a^\dagger\|^2.$$

Following a similar argument as in the proof of Lemma 4, we obtain

$$\left\|f_t-a^\dagger\right\|^2 \leq \left\|c_t-a^\dagger\right\|^2 - \upsilon_t^\dagger\left(\|c_t-d_t\|^2+\|f_t-d_t\|^2\right) - 2\upsilon\alpha_t\nu\|d_t-a^\dagger\|^2, \tag{50}$$

where $\upsilon_t^\dagger$ is defined in Lemma 4. According to $\lim_{t\to\infty}\upsilon_t^\dagger>0$ for all $\upsilon\in\big(0,2/(1+\theta)\big)$ and Eq. (50), one gives

$$\left\|f_t-a^\dagger\right\|^2\le\left\|c_t-a^\dagger\right\|^2-2\upsilon\alpha_t\nu\|d_t-a^\dagger\|^2. \tag{51}$$

Note that $\lim_{t\to\infty}\alpha_t=\alpha$ and thus $\alpha_t\ge\alpha$ for all $t\ge1$. Hence, there exists a constant, still denoted by t_0 making $\alpha_t\ge\alpha$ for any $t\ge t_0$. Combining Eq. (51) and utilizing proof techniques similar to those in Eqs. (25)–(30), we have

$$\begin{aligned}\left\|a_{t+1}-a^\dagger\right\|^2\le{}&\left\|a_t-a^\dagger\right\|^2+\eta\left(\left\|a_t-a^\dagger\right\|^2-\left\|a_{t-1}-a^\dagger\right\|^2\right)\\&-2\zeta\upsilon\alpha\nu\|d_t-a^\dagger\|^2+\sigma\left\|a_t-a_{t-1}\right\|^2,\quad\forall t\ge t_0.\end{aligned}$$

By induction, one has

$$\begin{aligned}2\zeta\upsilon\alpha\nu\sum_{k=t_0}^{t}\left\|d_k-a^\dagger\right\|^2\le{}&\left\|a_{t_0}-a^\dagger\right\|^2-\left\|a_{t+1}-a^\dagger\right\|^2+\sigma\sum_{k=t_0}^{t}\left\|a_k-a_{k-1}\right\|^2\\&+\eta\left(\left\|a_t-a^\dagger\right\|^2-\left\|a_{t_0-1}-a^\dagger\right\|^2\right).\end{aligned}$$

Thus we have $\sum_{k=t_0}^{\infty}\left\|d_k-a^\dagger\right\|^2<\infty$ by means of the sequence $\{a_t\}$ is bounded (cf. Eq. (36)) and $\sum_{k=t_0}^{\infty}\left\|a_k-a_{k-1}\right\|^2<\infty$ (cf. Eq. (37)). Therefore, one has $\lim_{t\to\infty}\left\|d_t-a^\dagger\right\|=0$. This together with Eqs. (39) and (42) yields

$$\left\|a_t-a^\dagger\right\|\le\|a_t-c_t\|+\|c_t-d_t\|+\left\|d_t-a^\dagger\right\|\to0,\quad\text{as } t\to\infty.$$

That is, $a_t\to a^\dagger$. The required error estimate Eq. (23) can be obtained using the same method as in Theorem 1.

Next we show the error bound of $\|c_t-a^\dagger\|$. From the definition of d_t and Eq. (2), one has

$$\left\langle c_t-\alpha_t\mathcal{G}(c_t)-d_t,a^\dagger-d_t\right\rangle\le0.$$

This together with the Lipschitz continuity of $\mathcal{G}$ and $\langle\mathcal{G}(d_t),d_t-a^\dagger\rangle\ge\nu\|d_t-a^\dagger\|^2$ implies

$$\begin{aligned}\left\langle c_t-d_t,a^\dagger-d_t\right\rangle&\le\alpha_t\left\langle\mathcal{G}(c_t),a^\dagger-d_t\right\rangle\\&=\alpha_t\left(\left\langle\mathcal{G}(c_t)-\mathcal{G}(d_t),a^\dagger-d_t\right\rangle+\left\langle\mathcal{G}(d_t),a^\dagger-d_t\right\rangle\right)\\&\le\alpha_t\left(L\|c_t-d_t\|\|a^\dagger-d_t\|-\nu\|d_t-a^\dagger\|^2\right).\end{aligned}\tag{52}$$

Thus we have

$$\begin{aligned}\nu\alpha_t\|d_t-a^\dagger\|^2&\le\alpha_tL\|c_t-d_t\|\|a^\dagger-d_t\|-\left\langle c_t-d_t,a^\dagger-d_t\right\rangle\\&\le\alpha_tL\|c_t-d_t\|\|a^\dagger-d_t\|+\|c_t-d_t\|\|a^\dagger-d_t\|\\&=(1+\alpha_tL)\|c_t-d_t\|\|a^\dagger-d_t\|.\end{aligned}$$

This combined with the Cauchy-Schwarz inequality further gives

$$\|c_t-a^\dagger\|\le\|c_t-d_t\|+\|d_t-a^\dagger\|\le\left(1+\frac{1+\alpha_tL}{\nu\alpha_t}\right)\|c_t-d_t\|.$$

Now we obtained an upper bound for $\|c_t-a^\dagger\|$. Let us proceed to demonstrate a lower bound for $\|c_t-a^\dagger\|$. From Eq. (52) and the Cauchy-Schwarz inequality, one sees that

$$\begin{aligned}\alpha_tL\|c_t-d_t\|\|a^\dagger-d_t\|&\ge\left\langle c_t-d_t,a^\dagger-d_t\right\rangle\\&=\|c_t-d_t\|^2-\left\langle d_t-c_t,a^\dagger-c_t\right\rangle\\&\ge\|c_t-d_t\|^2-\|c_t-d_t\|\|a^\dagger-c_t\|.\end{aligned}$$

Note that

$$\alpha_tL\|c_t-d_t\|\left(\|a^\dagger-c_t\|+\|c_t-d_t\|\right)\ge\alpha_tL\|c_t-d_t\|\|a^\dagger-d_t\|.$$

Hence

$$\frac{1-\alpha_tL}{1+\alpha_tL}\|c_t-d_t\|\le\|c_t-a^\dagger\|.$$

This completes the proof. □

Remark 6. Eq. (49) is highly valuable for establishing the convergence rate of Algorithm 3.1 in solving (VI). It also plays a crucial role in developing practical stopping criteria for our method, ensuring that the final iterate meets a specified level of accuracy. However, it is important to note that utilizing Eq. (49) necessitates knowledge of the Lipschitz constant L of the operator $\mathcal{G}$. However, if we use the step size criterion in Eq. (14), then Eq. (49) becomes the following

$$\frac{1-\theta\alpha_t/\alpha_{t+1}}{1+\theta\alpha_t/\alpha_{t+1}}\|c_t-d_t\|\le\|c_t-a^\dagger\|\le\left(\frac{1+\theta\alpha_t/\alpha_{t+1}}{\nu\alpha_t}+1\right)\|c_t-d_t\|. \tag{53}$$

Using the above equation to determine the error bound for $\|c_t-a^\dagger\|$ does not require the Lipschitz constant L.

Theorem 4. *Suppose that Conditions* (C1), (C2″), *and* (C3) *hold. Let* $\upsilon \in \big(0, 2/(1+\theta)\big)$. *Let* $\lim_{t\to\infty} \upsilon_t^\dagger = \hat{\upsilon}$ *and define* $\epsilon = \min\{\hat{\upsilon}/2, \theta L^{-1}\upsilon\nu\}$. *Assume that*

$$\zeta \in (0, 1/2], \quad 0 \le \mu < \psi \le \min\left\{\frac{1-2\zeta}{1-\zeta}, \frac{\zeta\epsilon}{1-\zeta\epsilon}\right\}. \tag{54}$$

Then $\{a_t\}$ *generated by Algorithm 3.1 converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ *with an* R*-linear rate. Moreover, we have the error estimate Eq. (23) and the global bound Eq. (49).*

Proof. Suppose that $\alpha_1 \ge \theta/L$, and by Lemma 3 we obtain $\lim_{t\to\infty} \alpha_t \ge \min\{\alpha_1, \theta/L\} = \theta/L$. Combining this with Eq. (54), we have

$$\lim_{t\to\infty} \upsilon_t^\dagger = \hat{\upsilon} \ge 2\epsilon, \quad \lim_{t\to\infty} \upsilon\alpha_t\nu \ge \theta L^{-1}\upsilon\nu \ge \epsilon.$$

Consequently, there exists a constant, which remains denoted by t_0 such that $\upsilon_t^\dagger \ge 2\epsilon$ and $\upsilon\alpha_t\nu \ge \epsilon$ for all $t \ge t_0$. Because of Eq. (50), one has

$$\begin{aligned}\left\|f_t - a^\dagger\right\|^2 &\le \left\|c_t - a^\dagger\right\|^2 - 2\epsilon\|c_t - d_t\|^2 - 2\epsilon\|d_t - a^\dagger\|^2 \\ &\le (1-\epsilon)\left\|c_t - a^\dagger\right\|^2, \quad \forall t \ge t_0.\end{aligned}$$

This combining with similar derivation processes as in Eqs. (25)–(30) gives

$$\begin{aligned}\left\|a_{t+1} - a^\dagger\right\|^2 &\le \eta_1\left\|a_t - a^\dagger\right\|^2 - \big(\psi(1-\zeta) + \zeta(1-\epsilon)\mu\big)\left\|a_{t-1} - a^\dagger\right\|^2 \\ &\quad - \rho\|a_{t+1} - a_t\|^2 + \sigma_1\|a_t - a_{t-1}\|^2 \\ &\le \eta_1\left\|a_t - a^\dagger\right\|^2 - \rho\|a_{t+1} - a_t\|^2 + \sigma_1\|a_t - a_{t-1}\|^2,\end{aligned} \tag{55}$$

where

$$\begin{aligned}\eta_1 &:= (1-\zeta)(1+\psi) + \zeta(1-\epsilon)(1+\mu), \quad \rho := \zeta^{-1}(1-\zeta)(1-\psi), \\ \sigma_1 &:= (1-\zeta)\psi(1+\psi) + \zeta(1-\epsilon)\mu(1+\mu) + \zeta^{-1}(1-\zeta)\big(\psi - \psi^2\big).\end{aligned}$$

It follows from Eq. (54) that $\psi \le (1-2\zeta)/(1-\zeta)$, and hence $\rho \ge 1$. Since $\mu < \psi$ and $\psi \le \zeta\epsilon/(1-\zeta\epsilon)$, one has

$$0 < \eta_1 < (1-\zeta\epsilon)(1+\psi) \le 1.$$

Note that $\hat{\upsilon} \in (0,1)$ and thus $\epsilon \in (0, 1/2)$. Since $\psi \le \zeta\epsilon/(1-\zeta\epsilon) < \zeta/(1-\zeta)$, we have $\psi \le \zeta(1+\psi)$. Therefore,

$$(1-\zeta)\big(\psi - \psi^2\big) < \zeta(1-\zeta)(1-\psi)(1+\psi) + \zeta^2(1-\epsilon)(1-\mu)(1+\mu).$$

This together with the definitions of σ_1 and η_1 implies that $\sigma_1 < \eta_1$. By using $\rho \ge 1$ and Eq. (55), we have

$$\begin{aligned}\left\|a_{t+1} - a^\dagger\right\|^2 + \|a_{t+1} - a_t\|^2 &\le \left\|a_{t+1} - a^\dagger\right\|^2 + \rho\|a_{t+1} - a_t\|^2 \\ &\le \eta_1\left(\left\|a_t - a^\dagger\right\|^2 + \|a_t - a_{t-1}\|^2\right).\end{aligned}$$

By induction, one deduces that

$$\begin{aligned}\left\|a_{t+1} - a^\dagger\right\|^2 &\le \eta_1\left(\left\|a_t - a^\dagger\right\|^2 + \|a_t - a_{t-1}\|^2\right) \\ &\le \eta_1^{t-t_0+1}\left(\|a_{t_0} - a^\dagger\|^2 + \|a_{t_0} - a_{t_0-1}\|^2\right) \le M\eta_1^t\end{aligned}$$

for some $M > 0$. This implies that $\{a_t\}$ converges R-linearly to $a^\dagger$. The error estimate Eq. (23) needed can be derived in the same way as Theorem 1. Additionally, the error bound for $\|c_t - a^\dagger\|$ is proven in the same manner as in Theorem 3. □

Remark 7. Both Theorems 3 and 4 establish the strong convergence of Algorithm 3.1. However, Theorem 4 provides additional insight by demonstrating the R-linear convergence of Algorithm 3.1 under more stringent parameter conditions. Specifically, when Algorithm 3.1 adopts the step size criterion Eq. (14) instead of Eq. (5), we are able to derive the error bound in Eq. (53). This error bound offers a more refined understanding of the convergence behavior of the algorithm.

Nevertheless, there are limitations when applying Eqs. (53) and (49) to determine the error bound for $\|c_t - a^\dagger\|$ in the case of Theorem 4. The key challenge lies in the requirement to know the Lipschitz constant L, which is necessary to determine the value of the parameter ϵ in condition Eq. (54). The necessity of the Lipschitz constant introduces an additional layer of complexity in practical applications, as it may not always be readily available. This constraint limits the general applicability of the error bounds, particularly in situations where an explicit value for L is difficult to obtain or estimate. Thus, while the theoretical convergence rates are promising, the practical use of these bounds requires careful consideration of the availability of the required parameters.

3.2. The second type of double inertial SEGA

In this subsection, we introduce a novel double inertial SEGA to solve VIPs. The difference between this algorithm and Algorithm 3.1 lies in the computation of d_t and f_t. However, it shares the same advantage of enhancing the computational speed of the original SEGA [12]. Now, this method is presented in Algorithm 3.2.

Algorithm 3.2 The second type of double inertial SEGA.

Initialization: Take $\theta \in (0,1)$, $\alpha_1 > 0$, $\psi \in [0,1)$, $\mu \in [0,1]$, $\zeta \in (0,1)$, and $\upsilon \in (1/(2-\theta), 1/\theta)$. Choose $\{\delta_t\}$ and $\{\rho_t\}$ satisfy Conditions (C3). Let $a_0, a_1 \in \mathbb{H}$. Set $t = 1$.
Step 1. Compute b_t and c_t according to (3.1).
Step 2. Compute $d_t = \mathbf{Proj}_{\mathbb{C}}\big(c_t - \upsilon\alpha_t\mathcal{G}(c_t)\big)$. If $c_t = d_t = a_t$, then $d_t \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$.
Step 3. Compute $f_t = \mathbf{Proj}_{Q_t}\big(c_t - \alpha_t\mathcal{G}(d_t)\big)$, where

$$Q_t := \big\{w \in \mathbb{H} : \langle c_t - \upsilon\alpha_t\mathcal{G}(c_t) - d_t, w - d_t\rangle \le 0\big\}.$$

Step 4. Compute $a_{t+1} = (1-\zeta)b_t + \zeta f_t$, and update α_{t+1} by (3.2).
Set $t \leftarrow t+1$ and go to *Step 1*.

Lemma 7. *Assume that Condition* (C2) *holds. Let* $\{f_t\}$ *be generated by Algorithm 3.2. Then*

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - \upsilon_t^\dagger\Big(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\Big), \quad \forall a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G}),$$

where $\upsilon_t^\dagger = 2 - \upsilon^{-1} - \theta\alpha_t\alpha_{t+1}^{-1}$ *if* $\upsilon \in ((2-\theta)^{-1}, 1]$ *and* $\upsilon_t^\dagger = \upsilon^{-1} - \theta\alpha_t\alpha_{t+1}^{-1}$ *if* $\upsilon \in (1, \theta^{-1})$.

Proof. By using $f_t \in Q_t$ and the definition of Q_t, one sees that

$$\langle c_t - \upsilon\alpha_t\mathcal{G}(c_t) - d_t, f_t - d_t\rangle \le 0. \tag{56}$$

Note that

$$\begin{aligned}
&\langle c_t - d_t, f_t - d_t\rangle \\
&\quad = \langle c_t - d_t - \upsilon\alpha_t\mathcal{G}(c_t) + \upsilon\alpha_t\mathcal{G}(c_t) - \upsilon\alpha_t\mathcal{G}(d_t) + \upsilon\alpha_t\mathcal{G}(d_t), f_t - d_t\rangle \\
&\quad = \langle c_t - \upsilon\alpha_t\mathcal{G}(c_t) - d_t, f_t - d_t\rangle + \upsilon\alpha_t\langle\mathcal{G}(d_t), f_t - d_t\rangle \\
&\quad + \upsilon\alpha_t\langle\mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle.
\end{aligned} \tag{57}$$

Combining Eqs. (9), (56), and (57), we have

$$\begin{aligned}
-\|c_t - f_t\|^2 &= -\|c_t - d_t\|^2 - \|d_t - f_t\|^2 + 2\langle c_t - d_t, f_t - d_t\rangle \\
&\le -\big(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\big)\Big(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\Big) + 2\upsilon\alpha_t\langle\mathcal{G}(d_t), f_t - d_t\rangle.
\end{aligned}$$

This is equivalent to

$$\begin{aligned}
&-2\alpha_t\langle\mathcal{G}(d_t), f_t - d_t\rangle \le -\big(\upsilon^{-1} - \theta\alpha_t/\alpha_{t+1}\big)\Big(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\Big) \\
&\quad + \upsilon^{-1}\|c_t - f_t\|^2, \quad \forall \upsilon > 0.
\end{aligned} \tag{58}$$

Following a process similar to that in Eq. (7), we obtain

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - \|c_t - f_t\|^2 - 2\alpha_t\langle\mathcal{G}(d_t), f_t - d_t\rangle. \tag{59}$$

From Eqs. (58) and (59), one has

$$\begin{aligned}
\left\|f_t - a^\dagger\right\|^2 \le{}& \left\|c_t - a^\dagger\right\|^2 - \big(\upsilon^{-1} - \theta\alpha_t/\alpha_{t+1}\big)\Big(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\Big) \\
&- \big(1 - \upsilon^{-1}\big)\|c_t - f_t\|^2.
\end{aligned} \tag{60}$$

We consider two cases for $\upsilon \in \big((2-\theta)^{-1}, \theta^{-1}\big)$.

Case 1: Consider $\upsilon \in \big((2-\theta)^{-1}, 1\big]$. Note that

$$\|c_t - f_t\|^2 \le 2\Big(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\Big),$$

which implies that $\big($for any $\upsilon \in ((2-\theta)^{-1}, 1]\big)$

$$-\big(1 - \upsilon^{-1}\big)\|c_t - f_t\|^2 \le -2\big(1 - \upsilon^{-1}\big)\Big(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\Big).$$

This together with Eq. (60) yields

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - \left(2 - \upsilon^{-1} - \theta\alpha_t/\alpha_{t+1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right),\ \forall \upsilon \in \left((2-\theta)^{-1}, 1\right].$$

Case 2: Consider $\upsilon \in \left(1, \theta^{-1}\right)$. Then it follows from Eq. (60) that

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - \left(\upsilon^{-1} - \theta\alpha_t/\alpha_{t+1}\right)\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right), \quad \forall \upsilon \in \left(1, \theta^{-1}\right).$$

This completes the proof. □

Lemma 8. *Let $\{c_t\}$ and $\{d_t\}$ be formulated by Algorithm 3.2. If $\left\{c_{t_k}\right\} \rightharpoonup z \in \mathbb{H}$ and $\lim_{k\to\infty} \|c_{t_k} - d_{t_k}\| = 0$, and*

(i) if the conditions (C1) and (C2) hold, then $z \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$.
(ii) if the conditions (C1′) and (C2′) hold, then $z \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$ or $\mathcal{G}(z) = 0$.

Proof. The conclusions can be obtained by applying a similar statement in Lemmas 5 and 6, respectively. □

Now we show the convergence theorems of our Algorithm 3.2 under different conditions.

Theorem 5. *Suppose that Condition (C3) holds. Let $\{a_t\}$ be generated by Algorithm 3.2. Let $\upsilon \in \left((2-\theta)^{-1}, \theta^{-1}\right)$. Then the following statements hold.*

(i) Let μ, ψ, and ζ fulfill Eq. (22).
(1) If Conditions (C1) and (C2) hold, then $\{a_t\}$ converges weakly to $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$. Additionally, we have the error estimate Eq. (23).
(2) If Conditions (C1′) and (C2′) hold, and assume that $\mathcal{G}(a) \ne 0$ for all $a \in \mathbb{C}$, then $\{a_t\}$ converges weakly to $a^\dagger \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$. Furthermore, we have the error estimate Eq. (23).
(3) If Conditions (C1) and (C2″) hold, then $\{a_t\}$ converges strongly to the unique solution $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$. In addition, we have the error estimate Eq. (23) and the global bound

$$\frac{1 - \upsilon\alpha_t L}{1 + \upsilon\alpha_t L}\|c_t - d_t\| \le \|c_t - a^\dagger\| \le \left(1 + \frac{1 + \upsilon\alpha_t L}{\nu\upsilon\alpha_t}\right)\|c_t - d_t\|. \tag{61}$$

In addition, if Algorithm 3.2 use the step size rule Eq. (14) instead of Eq. (5), then we have the global bound

$$\frac{1 - \upsilon\theta\alpha_t/\alpha_{t+1}}{1 + \upsilon\theta\alpha_t/\alpha_{t+1}}\|c_t - d_t\| \le \|c_t - a^\dagger\| \le \left(1 + \frac{1 + \upsilon\theta\alpha_t/\alpha_{t+1}}{\nu\upsilon\alpha_t}\right)\|c_t - d_t\|, \tag{62}$$

which is not related to the Lipschitz constant.

(ii) Let $\lim_{t\to\infty} \upsilon_t^\dagger = \hat{\upsilon}$ and define $\epsilon = \min\{\hat{\upsilon}/2, \theta L^{-1}\nu\}$. Let μ, ψ, and ζ fulfill the condition Eq. (54). If Conditions (C1) and (C2″) hold, then $\{a_t\}$ converges strongly to the unique solution $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ with an R-linear rate. Moreover, we have the error estimate Eq. (23) and the global bound Eq. (61) (or Eq. (62) when Algorithm 3.2 use the step size criterion Eq. (14)).

Proof. Since $\lim_{t\to\infty} \alpha_t$ exists, one has

$$\lim_{t\to\infty} \upsilon_t^\dagger = \begin{cases} 2 - \upsilon^{-1} - \theta, & \text{if } \upsilon \in (1/(2-\theta), 1], \\ \upsilon^{-1} - \theta, & \text{if } \upsilon \in (1, \theta^{-1}). \end{cases}$$

Thus we obtain $\lim_{t\to\infty} \upsilon_t^\dagger \in (0, 1)$ for all $\upsilon \in (1/(2-\theta), 1/\theta)$. The remaining proofs closely follow the proof processes of Theorems 1–4 (notice that we need to replace Lemmas 5 and 6 with Lemma 8). To avoid redundant descriptions, we omit the proof. □

Remark 8. The difference between Algorithms 3.1 and 3.2 lies in their step sizes for calculating d_t and f_t. Specifically, when $\upsilon = 1$, Algorithms 3.1 and 3.2 are equivalent. Moreover, if Algorithm 3.2 employs the step size rule Eq. (14) instead of Eq. (5), then the conclusions in Remarks 6 and 7 are also true. Specifically, for Algorithms 3.1 and 3.2, it appears that by setting $\upsilon = 1$ and using the step size criterion Eq. (5) with $\delta_t = 1$ and $\rho_t = 0$, one can recover the double inertial SEGA proposed in Yao et al. [28].

3.3. The first type of double inertial PCA

In this subsection, inspired by the double inertial method, the PCA [9], the SEGA [12], and the relaxation method, we propose a modified double inertial PCA to solve VIPs in real Hilbert spaces. The key distinction between this algorithm and Algorithm 3.2 lies in the step size used for computing f_t. Furthermore, unlike Algorithm 3.1, both the step sizes used for calculating d_t and f_t are different. Our numerical results in Section 4 demonstrate that this modification significantly improves the convergence speed of the algorithm. Now, we proceed to present Algorithm 3.3.

The following lemmas are useful for the convergence analysis of Algorithm 3.3.

Lemma 9. *Suppose that Condition (C3) holds. Then the step size sequence $\{\alpha_t\}$ formed by Eq. (64) is well defined and $\lim_{t\to\infty} \alpha_t$ exists.*

Algorithm 3.3 The first modified double inertial PCA.

Initialization: Take $\alpha_1 > 0$, $\theta \in (0,1)$, $\kappa \in (0, 2/\theta)$, $\psi \in [0,1)$, $\mu \in [0,1]$, $\zeta \in (0,1)$, and $\upsilon \in (\kappa/2, 1/\theta)$. Choose $\{\delta_t\}$ and $\{\rho_t\}$ satisfy Conditions (C3). Let $a_0, a_1 \in \mathbb{H}$. Set $t = 1$.

Step 1. Compute b_t and c_t according to (3.1).

Step 2. Compute $d_t = \mathbf{Proj}_{\mathbb{C}}(c_t - \upsilon\alpha_t \mathcal{G}(c_t))$. If $c_t = d_t = a_t$, then $d_t \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$.

Step 3. Compute $f_t = \mathbf{Proj}_{H_t}(c_t - \kappa w_t \alpha_t \mathcal{G}(d_t))$, where

$$H_t := \left\{ w \in \mathbb{H} : \langle c_t - \upsilon\alpha_t \mathcal{G}(c_t) - d_t, w - d_t \rangle \le 0 \right\},$$

and

$$w_t := \frac{\langle c_t - d_t, \eta_t \rangle}{\|\eta_t\|^2}, \quad \eta_t := c_t - d_t - \upsilon\alpha_t(\mathcal{G}(c_t) - \mathcal{G}(d_t)). \tag{63}$$

Step 4. Compute $a_{t+1} = (1-\zeta)b_t + \zeta f_t$, and update α_{t+1} by

$$\alpha_{t+1} = \begin{cases} \min\left\{ \dfrac{\theta\|c_t - d_t\|}{\|\mathcal{G}(c_t) - \mathcal{G}(d_t)\|}, \delta_t\alpha_t + \rho_t \right\}, & \text{if } \mathcal{G}(c_t) \ne \mathcal{G}(d_t), \\ \delta_t\alpha_t + \rho_t, & \text{otherwise.} \end{cases} \tag{64}$$

Set $t \leftarrow t+1$ and go to *Step 1*.

Proof. Since mapping $\mathcal{G}$ is L-Lipschitz continuous, one has

$$\frac{\theta\|c_t - d_t\|}{\|\mathcal{G}(c_t) - \mathcal{G}(d_t)\|} \ge \frac{\theta\|c_t - d_t\|}{L\|c_t - d_t\|} = \frac{\theta}{L}.$$

The remaining proof follows the proof process of Lemma 3. □

Lemma 10. *Suppose that Conditions* (C1) *and* (C2) *hold. Let* $\{c_t\}$, $\{d_t\}$, *and* $\{f_t\}$ *be three sequences formed by Algorithm 3.3. Then, for every* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$, *there exists* $t_0 > 0$ *such that*

$$\|f_t - a^\dagger\|^2 \le \|c_t - a^\dagger\|^2 - \frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\frac{\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}\|c_t - d_t\|^2, \quad \forall t \ge t_0. \tag{65}$$

Proof. By using Eq. (7), one has

$$\|f_t - a^\dagger\|^2 \le \|c_t - a^\dagger\|^2 - \|f_t - c_t\|^2 - 2\kappa\alpha_t w_t \langle \mathcal{G}(d_t), f_t - d_t \rangle. \tag{66}$$

From the definition of η_t and Eq. (64), one obtains

$$\|\eta_t\| \ge \|c_t - d_t\| - \upsilon\alpha_t\|\mathcal{G}(c_t) - \mathcal{G}(d_t)\| \ge \left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)\|c_t - d_t\|.$$

We can also show that

$$\|\eta_t\| \le \left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)\|c_t - d_t\|. \tag{67}$$

Therefore, we conclude that

$$\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)\|c_t - d_t\| \le \|\eta_t\| \le \left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)\|c_t - d_t\|.$$

It follows from Lemma 9 that $\lim_{t\to\infty}\alpha_t$ exists. Consequently, there exists a constant t_0 such that $1 - \upsilon\theta\alpha_t/\alpha_{t+1} > 0$ for all $t \ge t_0$ (noting that $\upsilon < 1/\theta$). It should be noted that $w_t > 0$ for all $t \ge t_0$. Indeed, by using the definitions of w_t and η_t, and Eq. (64), we have

$$\begin{aligned} w_t = \frac{\langle c_t - d_t, \eta_t \rangle}{\|\eta_t\|^2} &= \frac{\|c_t - d_t\|^2 - \langle c_t - d_t, \upsilon\alpha_t(\mathcal{G}(c_t) - \mathcal{G}(d_t)) \rangle}{\|\eta_t\|^2} \\ &\ge \frac{\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)\|c_t - d_t\|^2}{\|\eta_t\|^2} > 0. \end{aligned} \tag{68}$$

Combining Eqs. (67) and (68), we deduce

$$w_t \ge \frac{1 - \upsilon\theta\alpha_t/\alpha_{t+1}}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2} > 0, \quad \forall t \ge t_0. \tag{69}$$

According to the definition of H_t and $f_t \in H_t$, one obtains

$$\langle c_t - \upsilon\alpha_t\mathcal{G}(c_t) - d_t, f_t - d_t \rangle \le 0.$$

This shows that

$$\langle c_t - d_t - \upsilon\alpha_t(\mathcal{G}(c_t) - \mathcal{G}(d_t)), f_t - d_t \rangle \le \upsilon\alpha_t\langle \mathcal{G}(d_t), f_t - d_t \rangle. \tag{70}$$

From the definitions of η_t and w_t, and Eq. (70), we have

$$\begin{aligned}
&-2\kappa\alpha_t w_t\langle \mathcal{G}(d_t), f_t - d_t\rangle \\
&\le -2\kappa\upsilon^{-1} w_t\langle \eta_t, f_t - d_t\rangle \\
&= -2\kappa\upsilon^{-1} w_t\langle \eta_t, c_t - d_t\rangle + 2\kappa\upsilon^{-1} w_t\langle \eta_t, c_t - f_t\rangle \\
&= -2\kappa\upsilon^{-1} w_t^2\|\eta_t\|^2 + 2\kappa\upsilon^{-1} w_t\langle \eta_t, c_t - f_t\rangle \\
&= -2\kappa\upsilon^{-1} w_t^2\|\eta_t\|^2 + \|c_t - f_t\|^2 + \kappa^2/\upsilon^2 w_t^2\|\eta_t\|^2 - \|c_t - f_t - \kappa\upsilon^{-1} w_t\eta_t\|^2.
\end{aligned} \tag{71}$$

It follows from Eq. (68) that $w_t\|\eta_t\|^2 \ge \left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)\|c_t - d_t\|^2$. This together with Eq. (67) yields

$$w_t^2\|\eta_t\|^2 = \frac{w_t^2\|\eta_t\|^4}{\|\eta_t\|^2} \ge \frac{\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}\|c_t - d_t\|^2. \tag{72}$$

Combining Eqs. (66), (71), and (72), we conclude that (65) holds, as desired. □

We proceed to prove the convergence theorems for Algorithm 3.3.

Theorem 6. *Suppose that Condition* (C3) *holds. Let the sequence* $\{a_t\}$ *be generated by Algorithm 3.3. Let* $\upsilon \in (\kappa/2, 1/\theta)$ *and* $\kappa \in (0, 2/\theta)$. *Then the following statements hold.*

(i) Let μ, ψ, and ζ fulfill the condition Eq. (22).
- *(1) If Conditions* (C1) *and* (C2) *hold, then* $\{a_t\}$ *converges weakly to* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$. *Also, we obtain the error estimate Eq. (23).*
- *(2) If Conditions (C1′) and (C2′) hold, and assume that* $\mathcal{G}(a) \neq 0$ *for all* $a \in \mathbb{C}$, *then* $\{a_t\}$ *converges weakly to* $a^\dagger \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$. *Moreover, we have the error estimate Eq. (23).*
- *(3) If Conditions* (C1) *and* (C2″) *hold, then* $\{a_t\}$ *converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$. *Likewise, we have the error estimate Eq. (23) and the global bound*

$$\frac{1 - \theta\upsilon\alpha_t/\alpha_{t+1}}{1 + \theta\upsilon\alpha_t/\alpha_{t+1}}\|c_t - d_t\| \le \|c_t - a^\dagger\| \le \left(\frac{1 + \theta\upsilon\alpha_t/\alpha_{t+1}}{\nu\upsilon\alpha_t} + 1\right)\|c_t - d_t\|. \tag{73}$$

(ii) Define

$$\epsilon := \frac{1}{2}\min\left\{\frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\left(\frac{1 - \upsilon\theta}{1 + \upsilon\theta}\right)^2, 2\frac{\theta}{L}\kappa\nu\frac{1 - \upsilon\theta}{(1 + \upsilon\theta)^2}\right\}.$$

Let μ, ψ, and ζ fulfill the condition Eq. (54). If Conditions (C1) *and* (C2″) *hold, then* $\{a_t\}$ *converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ *with an R-linear rate. Furthermore, we obtain the error estimate Eq. (23) and the global bound Eq. (73).*

Proof. We divide this proof into four parts.

The proof of (i)(1): From the fact that $\lim_{t\to\infty}\alpha_t$ exists, $\theta \in (0, 1)$, $\kappa \in (0, 2/\theta)$, and $\upsilon \in (\kappa/2, 1/\theta)$, we have

$$\lim_{t\to\infty}\frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\frac{\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2} = \frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\left(\frac{1 - \upsilon\theta}{1 + \upsilon\theta}\right)^2 > 0.$$

Let $f(\kappa) := \kappa(2\upsilon - \kappa)/\upsilon^2$. It can be checked that the maximum value of $f(\kappa)$ is 1 when $\kappa = \upsilon$. Thus we have

$$0 < \frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\left(\frac{1 - \upsilon\theta}{1 + \upsilon\theta}\right)^2 < 1. \tag{74}$$

Thus there exists a constant $t_0 \in \mathbb{N}$ such that

$$\frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\frac{\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2} \ge \frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\left(\frac{1 - \upsilon\theta}{1 + \upsilon\theta}\right)^2 > 0, \quad \forall t \ge t_0.$$

Combining Lemmas 8, 10, and the proof of Theorem 1, we can easily conclude that $\{a_t\}$ converges weakly to $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ when Conditions (C1) and (C2) hold. The required error estimate Eq. (23) follows from the same technique as Theorem 1.

The proof of (i)(2): If Conditions (C1′) and (C2′) hold, then $\{a_t\}$ converges weakly to $a^\dagger \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$ by means of the proof Theorem 2.

The proof of (i)(3): On the other hand, if Conditions (C1) and (C2″) hold, then it follows from the strong pseudomonotonicity of $\mathcal{G}$, Eqs. (65) and (69) that

$$\begin{aligned}
\|f_t - a^\dagger\|^2 \le{}& \|c_t - a^\dagger\|^2 - \frac{\kappa}{\upsilon^2}(2\upsilon - \kappa)\frac{\left(1 - \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}\|c_t - d_t\|^2 \\
&- 2\kappa\alpha_t\nu\frac{1 - \upsilon\theta\alpha_t/\alpha_{t+1}}{\left(1 + \upsilon\theta\alpha_t/\alpha_{t+1}\right)^2}\|d_t - a^\dagger\|^2.
\end{aligned}$$

Since $\lim_{t\to\infty}\alpha_t$ exists (assume that $\lim_{t\to\infty}\alpha_t=\alpha$), one has

$$\begin{aligned}\|f_t-a^\dagger\|^2\le{}&\|c_t-a^\dagger\|^2-\frac{\kappa}{\upsilon^2}(2\upsilon-\kappa)\Big(\frac{1-\upsilon\theta}{1+\upsilon\theta}\Big)^2\|c_t-d_t\|^2\\&-2\kappa\alpha\nu\frac{1-\upsilon\theta}{(1+\upsilon\theta)^2}\|d_t-a^\dagger\|^2,\quad\forall t\ge t_1.\end{aligned}\tag{75}$$

By using Eqs. (74) and (75), one obtains

$$\|f_t-a^\dagger\|^2\le\|c_t-a^\dagger\|^2-2\kappa\alpha\nu\frac{1-\upsilon\theta}{(1+\upsilon\theta)^2}\|d_t-a^\dagger\|^2,\quad\forall t\ge t_1.$$

By employing a proof process similar to that of Theorem 3, we can obtain that $\{a_t\}$ converges strongly to the unique solution $a^\dagger\in\mathrm{VI}(\mathbb{C},\mathcal{G})$ when Conditions (C1) and (C2′′) hold. Moreover, the error bound for $\|c_t-a^\dagger\|$ defined in Eq. (73) can be obtained through a proof similar to that of Theorem 3.

The proof of (ii): Finally, we prove that $\{a_t\}$ converges strongly to the unique solution $a^\dagger\in\mathrm{VI}(\mathbb{C},\mathcal{G})$ with an R-linear rate when Conditions (C1), (C2′′), and Eq. (54) hold. Indeed, it follows from the definition of ϵ and Eq. (75) that $\|f_t-a^\dagger\|^2\le(1-\epsilon)\|c_t-a^\dagger\|^2$. The rest of the proof follows the same procedure as the subsequent proof of Theorem 4. □

Remark 9. The proposed Algorithms 3.2 and 3.3 use the same search direction when calculating f_t. Indeed, if we replace κw_t in the proposed Algorithm 3.3 by 1, then it reduces to the update form of the suggested Algorithm 3.2. The numerical results by Cai et al. [38] demonstrate that the proposed projection and contraction method achieves twice the convergence speed of the extragradient method [7]. In our numerical experiments in Section 4, we also observe the superior performance of the proposed double inertial projection and contraction Algorithm 3.3 compared to the suggested double inertial subgradient extragradient Algorithm 3.2. On the other hand, by setting $\upsilon=1$ and using the step size criterion in Eq. (64) with $\delta_t=1$ and $\rho_t=0$, Algorithm 3.3 can recover the double inertial projection and contraction method in Li et al. [33].

3.4. *The second type of double inertial PCA*

In this subsection, we present the last accelerated algorithm of this paper, as shown in Algorithm 3.4, which is a variant of the proposed Algorithm 3.3. Note that the same step size length is used in Algorithms 3.3 and 3.4 even if the search directions are different when computing f_t; they are a pair of geminate directions (i.e., $\alpha_t\mathcal{G}(d_t)$ and η_t) from the PCA [9]. Moreover, the suggested Algorithm 3.4 only computes the projection onto the feasible set once per iteration and does not involve projection onto a half-space.

Algorithm 3.4 The second type of double inertial PCA.

Initialization: Take $\alpha_1>0$, $\theta\in(0,1)$, $\kappa\in(0,2)$, $\psi\in[0,1)$, $\mu\in[0,1]$, $\zeta\in(0,1)$, and $\upsilon\in(0,1/\theta)$. Choose $\{\delta_t\}$ and $\{\rho_t\}$ satisfy Conditions (C3). Let $a_0,a_1\in\mathbb{H}$. Set $t=1$.
Step 1. Compute b_t and c_t according to (3.1).
Step 2. Compute $d_t=\mathbf{Proj}_{\mathbb{C}}(c_t-\upsilon\alpha_t\mathcal{G}(c_t))$. If $c_t=d_t=a_t$, then $d_t\in\mathrm{VI}(\mathbb{C},\mathcal{G})$.
Step 3. Compute $f_t=c_t-\kappa w_t\eta_t$, where w_t and η_t are defined in (3.60).
Step 4. Compute $a_{t+1}=(1-\zeta)b_t+\zeta f_t$, and update α_{t+1} by (3.61).
Set $t\leftarrow t+1$ and go to *Step 1*.

The following lemma is crucial for the convergence analysis of Algorithm 3.4.

Lemma 11. *Suppose that Conditions* (C1) *and* (C2) *hold. Let* $\{c_t\}$, $\{d_t\}$, *and* $\{f_t\}$ *be three sequences generated by Algorithm 3.4. Then there exists* $t\ge t_0$ *such that*

$$\|f_t-a^\dagger\|^2\le\|c_t-a^\dagger\|^2-(2-\kappa)\kappa\frac{\big(1-\upsilon\theta\alpha_t/\alpha_{t+1}\big)^2}{\big(1+\upsilon\theta\alpha_t/\alpha_{t+1}\big)^2}\|c_t-d_t\|^2,\quad\forall a^\dagger\in\mathrm{VI}(\mathbb{C},\mathcal{G}).$$

Proof. By using $a^\dagger\in\mathrm{VI}(\mathbb{C},\mathcal{G})$, $d_t\in\mathbb{C}$, and the pseudomonotonicity of $\mathcal{G}$, we deduce

$$\langle\mathcal{G}(d_t),d_t-a^\dagger\rangle\ge0.\tag{76}$$

In view of $d_t=\mathbf{Proj}_{\mathbb{C}}(c_t-\upsilon\alpha_t\mathcal{G}(c_t))$ and Eq. (2), we have

$$\langle c_t-d_t-\upsilon\alpha_t\mathcal{G}(c_t),d_t-a^\dagger\rangle\ge0.\tag{77}$$

From Eqs. (76) and (77), we obtain

$$\langle d_t-a^\dagger,\eta_t\rangle\ge0.\tag{78}$$

According to the definition of w_t, one has $\langle c_t-d_t,\eta_t\rangle=w_t\|\eta_t\|^2$. By using the definition of η_t and Eq. (78), one sees that

$$\langle c_t-a^\dagger,\eta_t\rangle=\langle c_t-d_t,\eta_t\rangle+\langle d_t-a^\dagger,\eta_t\rangle\ge w_t\|\eta_t\|^2.\tag{79}$$

By the definition of f_t and Eq. (79), one has

$$\begin{aligned}\|f_t - a^\dagger\|^2 &= \|c_t - \kappa w_t\eta_t - a^\dagger\|^2\\ &= \|c_t - a^\dagger\|^2 - 2\kappa w_t\langle c_t - a^\dagger, \eta_t\rangle + \kappa^2 w_t^2\|\eta_t\|^2\\ &\le \|c_t - a^\dagger\|^2 - 2\kappa w_t^2\|\eta_t\|^2 + \kappa^2 w_t^2\|\eta_t\|^2\\ &= \|c_t - a^\dagger\|^2 - (2-\kappa)\kappa\|w_t\eta_t\|^2, \quad \forall t \ge t_0.\end{aligned}$$

This combines with Eq. (72) to generate the desired inequality. The proof is completed. □

Let us demonstrate the weak, strong, and linear convergence theorems of Algorithm 3.4.

Theorem 7. *Suppose that Condition* (C3) *holds. Let the sequence* $\{a_t\}$ *be generated by Algorithm 3.4. Let* $\upsilon \in (0, 1/\theta)$ *and* $\kappa \in (0, 2)$*. Then the following statements hold.*

(i) Let μ*,* ψ*, and* ζ *fulfill the condition Eq. (22).*
  (1) If Conditions (C1) *and* (C2) *hold, then* $\{a_t\}$ *converges weakly to* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$*. In addition, there is the error estimate Eq. (23).*
  (2) If Conditions (C1′) *and* (C2′) *hold, and assume that* $\mathcal{G}(a) \ne 0$ *for all* $a \in \mathbb{C}$*, then* $\{a_t\}$ *converges weakly to* $a^\dagger \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$*. Additionally, we have the error estimate Eq. (23).*
  (3) If Conditions (C1) *and* (C2″) *hold, then* $\{a_t\}$ *converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$*. Similarly, we have the error estimate Eq. (23) and the global bound Eq. (73).*

(ii) Define

$$\epsilon := \frac{1}{2}\min\left\{\kappa(2-\kappa)\left(\frac{1-\upsilon\theta}{1+\upsilon\theta}\right)^2, 2\frac{\theta}{L}\upsilon\kappa\nu\frac{1-\upsilon\theta}{(1+\upsilon\theta)^2}\right\}.$$

Let μ*,* ψ*, and* ζ *fulfill the condition Eq. (54). If Conditions* (C1) *and* (C2″) *hold, then* $\{a_t\}$ *converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ *with an R-linear rate. Moreover, we have the error estimate Eq. (23) and the global bound Eq. (73).*

Proof. We divide this proof into four parts.

The proof of (i)(1): Combining $\lim_{t\to\infty}\alpha_t$ exists, $\theta \in (0,1)$, $\kappa \in (0,2)$, and $\upsilon \in (0, 1/\theta)$, we have

$$\lim_{t\to\infty}\kappa(2-\kappa)\frac{(1-\upsilon\theta\alpha_t/\alpha_{t+1})^2}{(1+\upsilon\theta\alpha_t/\alpha_{t+1})^2} = \kappa(2-\kappa)\left(\frac{1-\upsilon\theta}{1+\upsilon\theta}\right)^2 > 0.$$

Let $f(\kappa) := \kappa(2-\kappa)$. Then the maximum value of $f(\kappa)$ is 1 when $\kappa = 1$. Hence we obtain

$$0 < \kappa(2-\kappa)\left(\frac{1-\upsilon\theta}{1+\upsilon\theta}\right)^2 < 1, \quad \forall t \ge t_0.$$

If Conditions (C1) and (C2) hold, then we deduce that $\{a_t\}$ converges weakly to $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ by means of Lemmas 8, 11 and the proof of Theorem 1. The error estimate Eq. (23) can be obtained in the same manner as Theorem 1.

The proof of (i)(2): In addition, we also have $\{a_t\}$ converges weakly to $a^\dagger \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$ using the proof of Theorem 2 in the case that Conditions (C1′) and (C2′) hold.

The proof of (i)(3): Combining Condition (C2″), the proof of Lemma 11, Eq. (69), and $\alpha := \lim_{t\to\infty}\alpha_t$, we conclude that

$$\begin{aligned}\|f_t - a^\dagger\|^2 \le{}& \|c_t - a^\dagger\|^2 - \kappa(2-\kappa)\left(\frac{1-\upsilon\theta}{1+\upsilon\theta}\right)^2\|c_t - d_t\|^2\\ &- 2\upsilon\kappa\alpha\nu\frac{1-\upsilon\theta}{(1+\upsilon\theta)^2}\|d_t - a^\dagger\|^2, \quad \forall t \ge t_2.\end{aligned} \tag{80}$$

Then

$$\|f_t - a^\dagger\|^2 \le \|c_t - a^\dagger\|^2 - 2\upsilon\kappa\alpha\nu\frac{1-\upsilon\theta}{(1+\upsilon\theta)^2}\|d_t - a^\dagger\|^2, \quad \forall t \ge t_2.$$

By employing a proof process similar to that of Theorem 3, we obtain that $\{a_t\}$ converges strongly to the unique solution $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ when Conditions (C1) and (C2″) hold. The error bound analysis in Eq. (73) is similar to the one in Theorem 3 and is therefore omitted.

The proof of (ii): It follows from the definition of ϵ and Eq. (80) that $\|f_t - a^\dagger\|^2 \le (1-\epsilon)\|c_t - a^\dagger\|^2$. We can conclude that $\{a_t\}$ converges strongly to the unique solution $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ with an R-linear rate follows the same procedure as the subsequent proof of Theorem 4. □

3.5. Convergence theorems of the proposed algorithms without the assumption of Lipschitz continuity

In this subsection, we present the weak convergence theorems of the proposed Algorithms 3.1–3.4 without assuming Lipschitz continuity. Before proceeding, we first introduce the following condition, which weakens the Lipschitz continuity in Conditions (C2) to uniform continuity.

Armijo type step size criterion-I

Let $\gamma > 0$, $\ell \in (0,1)$, and $\theta \in (0,1)$. The step size $\alpha_t = \gamma\ell^m$ is chosen to be the largest $\{\gamma, \gamma\ell, \gamma\ell^2, \dots\}$ such that

$$\alpha_t\langle \mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle \le \frac{\theta}{2}\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right). \tag{81}$$

Armijo type step size criterion-II

Let $\gamma > 0$, $\ell \in (0,1)$, and $\theta \in (0,1)$. The step size $\alpha_t = \gamma\ell^m$ is chosen to be the largest $\{\gamma, \gamma\ell, \gamma\ell^2, \dots\}$ such that

$$\alpha_t\|\mathcal{G}(c_t) - \mathcal{G}(d_t)\| \le \theta\|c_t - d_t\|. \tag{82}$$

(C2-1) The mapping $\mathcal{G}\colon \mathbb{H} \to \mathbb{H}$ is pseudomonotone, uniformly continuous on bounded subsets of $\mathbb{H}$, and sequentially weakly continuous on $\mathbb{C}$.

We also need the following two Armijo-type adaptive step size rules, which correspond to the step size criteria Eqs. (5) and (14), respectively.

Remark 10. Notice that Eq. (81) can be deduced from Eq. (82). Indeed, it follows from the Cauchy-Schwarz inequality and Eq. (82) that

$$\begin{aligned}\alpha_t\langle \mathcal{G}(c_t) - \mathcal{G}(d_t), f_t - d_t\rangle &\le \alpha_t\|\mathcal{G}(c_t) - \mathcal{G}(d_t)\|\|f_t - d_t\| \\ &\le \theta\|c_t - d_t\|\|f_t - d_t\| \le \frac{\theta}{2}\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right).\end{aligned}$$

Note that step size rule Eq. (81) needs to be embedded in Step 3 of the proposed Algorithms 3.1–3.4, while step size rule Eq. (82) should be embedded in Step 2. The advantage of utilizing step size rule Eq. (81) is that it uses the information from the sequence f_t when determining the step size α_t, which helps improve the convergence speed of the algorithms.

On the other hand, if $\mathcal{G}$ is Lipschitz continuous, then it follows from Eq. (81) (or (Eq. 82)) that $\ell^{-1}\alpha_t L\|c_t - d_t\| > \theta\|c_t - d_t\|$, i.e., $\alpha_t > \ell\theta L^{-1}$ for all $t \ge 1$.

The following three lemmas are crucial for the convergence analysis of Algorithm 3.1 under the step size criteria Eq. (81) (or Eq. (82)).

Lemma 12. *Suppose that $\mathcal{G}\colon \mathbb{H} \to \mathbb{H}$ is uniformly continuous on bounded subsets of $\mathbb{H}$. The Armijo-like criteria generated in Eq. (81) (or Eq. (82)) is well defined.*

Proof. If $c_t \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$, then $c_t = \mathbf{Proj}_{\mathbb{C}}\left(c_t - \gamma\mathcal{G}(c_t)\right)$ by using Eq. (2). This implies that $d_t = c_t$ and thus $m = 0$. If $c_t \notin \mathrm{VI}(\mathbb{C}, \mathcal{G})$, then we suppose that the opposite of Eq. (81) holds. That is

$$\alpha_m\langle \mathcal{G}(c_t) - \mathcal{G}(d_m), f_m - d_m\rangle > \frac{\theta}{2}\left(\|c_t - d_m\|^2 + \|f_m - d_m\|^2\right).$$

where $d_m := \mathbf{Proj}_{\mathbb{C}}\left(c_t - \alpha_m\mathcal{G}(c_t)\right)$, $f_m := \mathbf{Proj}_{\mathbb{C}}\left(c_t - \upsilon\alpha_m\mathcal{G}(d_m)\right)$ and $\alpha_m := \gamma\ell^m$. By using the Cauchy-Schwarz inequality, one has

$$\begin{aligned}\alpha_m\|\mathcal{G}(c_t) - \mathcal{G}(d_m)\|\|f_m - d_m\| &> \frac{\theta}{2}\left(\|c_t - d_m\|^2 + \|f_m - d_m\|^2\right) \\ &\ge \theta\|c_t - d_m\|\|f_m - d_m\|.\end{aligned}$$

This is equivalent to $\alpha_m\|\mathcal{G}(c_t) - \mathcal{G}(d_m)\| > \theta\|c_t - d_m\|$ (It can be obtained directly if Eq. (82) is violated). Therefore, we have

$$\|\mathcal{G}(c_t) - \mathcal{G}(d_m)\| > \theta\|c_t - d_m\|\alpha_m^{-1}. \tag{83}$$

We study two cases of c_t.

Case 1: Consider $c_t \notin \mathbb{C}$. It follows from $\lim_{m\to\infty}\alpha_m = 0$ and the uniform continuity of G that $\lim_{m\to\infty}\|c_t - d_m\| > 0$ and $\lim_{m\to\infty}\alpha_m\|\mathcal{G}(c_t) - \mathcal{G}(d_m)\| = 0$. Combining these inequalities with Eq. (83) creates a contradiction.

Case 2: Consider $c_t \in \mathbb{C}$. From the definition of d_m and Eq. (2, one arrives at

$$\langle d_m - c_t + \alpha_m\mathcal{G}(c_t), a - d_m\rangle \ge 0, \quad \forall a \in \mathbb{C}.$$

This gives

$$\left\langle (d_m - c_t)/\alpha_m, a - d_m\right\rangle + \langle \mathcal{G}(c_t), a - d_m\rangle \ge 0, \quad \forall a \in \mathbb{C}. \tag{84}$$

Since $\mathcal{G}$ is uniformly continuous and $\lim_{m\to\infty}\alpha_m = 0$, one obtains $\lim_{m\to\infty}\|c_t - d_m\| = 0$ and $\lim_{m\to\infty}\|\mathcal{G}(c_t) - \mathcal{G}(d_m)\| = 0$. Thanks to Eq. (83), one sees that $\lim_{m\to\infty}\|c_t - d_m\|/\alpha_m = 0$. This together with Eqs. (83) and (84) implies that $\langle \mathcal{G}(c_t), a - c_t\rangle \ge 0$ for all $a \in \mathbb{C}$, i.e. $c_t \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$, which contradicts the hypothesis $c_t \notin \mathrm{VI}(\mathbb{C}, \mathcal{G})$. The proof is completed. ☐

Lemma 13. *Assume that Condition* (C2-1) *holds. Let* $\{f_t\}$ *be a sequence created by Algorithm 3.1 with step size criterion Eq. (81) (or Eq. (82)). Then*

$$\left\|f_t - a^\dagger\right\|^2 \le \left\|c_t - a^\dagger\right\|^2 - v_t^\dagger\left(\|c_t - d_t\|^2 + \|f_t - d_t\|^2\right), \quad \forall a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G}),$$

where $v_t^\dagger = 2 - \upsilon - \upsilon\theta$ *if* $\upsilon \in [1, 2/(1+\theta))$ *and* $v_t^\dagger = \upsilon - \upsilon\theta$ *if* $\upsilon \in (0,1)$.

Proof. The proof of this lemma is similar to that Lemma 4. □

Lemma 14. *Suppose that Conditions* (C1) *and* (C2-1) *hold. Let* $\{c_t\}$ *and* $\{d_t\}$ *be two sequences generated by Algorithm 3.1 with step size criterion Eq. (81) (or Eq. (82)). If there exists a subsequence* $\{c_{t_k}\}$ *of* $\{c_t\}$ *such that* $\{c_{t_k}\}$ *converges weakly to* $z \in \mathbb{H}$ *and* $\lim_{k\to\infty}\|c_{t_k} - d_{t_k}\| = 0$*, then* $z \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$.

Proof. Following the proof process of Lemma 5, one has

$$\alpha_{t_k}^{-1}\left\langle c_{t_k} - d_{t_k}, a - d_{t_k}\right\rangle + \left\langle \mathcal{G}(c_{t_k}), d_{t_k} - c_{t_k}\right\rangle \le \left\langle \mathcal{G}(c_{t_k}), a - c_{t_k}\right\rangle, \quad \forall a \in \mathbb{C}. \tag{85}$$

Next, we need to demonstrate $\liminf_{k\to\infty}\left\langle \mathcal{G}(c_{t_k}), a - c_{t_k}\right\rangle \ge 0$ for all $a \in \mathbb{C}$ by considering two possible cases of α_{t_k}.

Case 1: Consider $\liminf_{k\to\infty}\alpha_{t_k} > 0$. From the fact that $\{c_{t_k}\}$ is bounded and mapping $\mathcal{G}$ is uniformly continuous on bounded subset of $\mathbb{H}$, one sees that $\{\mathcal{G}(c_{t_k})\}$ is bounded. By using $\liminf_{k\to\infty}\left\|c_{t_k} - d_{t_k}\right\| = 0$ and Eq. (85), we have $\liminf_{k\to\infty}\left\langle \mathcal{G}(c_{t_k}), a - c_{t_k}\right\rangle \ge 0$.

Case 2: Consider $\liminf_{k\to\infty}\alpha_{t_k} = 0$. Set $y_{t_k} = \mathbf{Proj}_{\mathbb{C}}\left(c_{t_k} - \alpha_{t_k}\ell^{-1}\mathcal{G}(c_{t_k})\right)$, by [21, Lemma 2] we have $\ell\left\|c_{t_k} - d_{t_k}\right\| \le \left\|c_{t_k} - y_{t_k}\right\|$ and thus $\lim_{k\to\infty}\left\|c_{t_k} - y_{t_k}\right\| = 0$. Therefore, one obtains $y_{t_k} \rightharpoonup z \in \mathbb{C}$ and hence $\{y_{t_k}\}$ is bounded. This together with the uniform continuity of mapping $\mathcal{G}$ yields $\lim_{k\to\infty}\left\|\mathcal{G}(c_{t_k}) - \mathcal{G}(y_{t_k})\right\| = 0$. From Eq. (81) (or Eq. (82)) and the Cauchy-Schwarz inequality, one gives

$$\alpha_{t_k}\ell^{-1}\left\|\mathcal{G}(c_{t_k}) - \mathcal{G}(y_{t_k})\right\| > \theta\left\|c_{t_k} - y_{t_k}\right\|.$$

This combining with $\lim_{k\to\infty}\left\|\mathcal{G}(c_{t_k}) - \mathcal{G}(y_{t_k})\right\| = 0$ yields

$$\lim_{k\to\infty}\alpha_{t_k}^{-1}\ell\left\|c_{t_k} - y_{t_k}\right\| = 0.$$

From the definition of y_{t_k} and Eq. (2), one has

$$\alpha_{t_k}^{-1}\ell\left\langle c_{t_k} - y_{t_k}, a - y_{t_k}\right\rangle + \left\langle \mathcal{G}(c_{t_k}), y_{t_k} - c_{t_k}\right\rangle \le \left\langle \mathcal{G}(c_{t_k}), a - c_{t_k}\right\rangle, \quad \forall a \in \mathbb{C}.$$

The desired conclusion is obtained by taking the limits of the left and right sides of the above equation.

The rest of the proof of this lemma basically follows from the subsequent proof of Lemma 5 and is therefore omitted. □

Now we can show the convergence theorem of the proposed Algorithm 3.1 with step size criterion Eq. (81) (or Eq. (82)).

Theorem 8. *Suppose that Conditions* (C1) *and* (C3) *hold. Let the sequence* $\{a_t\}$ *be generated by Algorithm 3.1 with step size criterion Eq. (81) (or Eq. (82)). Let* $\upsilon \in (0, 2/1+\theta)$*. Then the following statements hold.*

(i) Let μ, ψ, *and* ζ *fulfill Eq. (22).*

(1) If Conditions (C2-1) *and Eq. (81) (or Eq. (82)) hold, then* $\{a_t\}$ *converges weakly to* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$*. Additionally, there exist positive constants* M_1 *and* β *such that*

$$\min_{1\le k\le t}\|c_k - d_k\| \le \left(\frac{\Gamma_1 + \frac{1}{1-\eta}[\Psi_1]_+ + \frac{1}{1-\eta}M_1}{\zeta\beta t}\right)^{1/2}, \tag{86}$$

where $\Gamma_t := \left\|a_t - a^\dagger\right\|^2$ *and* $\Psi_t := \Gamma_t - \Gamma_{t-1}$.

(2) If Conditions (C1′) and (C2′) *hold, and assume that* $\mathcal{G}(a) \ne 0$ *for all* $a \in \mathbb{C}$*, then* $\{a_t\}$ *converges weakly to* $a^\dagger \in \mathrm{DVI}(\mathbb{C}, \mathcal{G})$*. Additionally, we have the error estimate Eq. (86).*

(3) If Condition (C2″) and Eq. (82) hold, then $\{a_t\}$ *converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$*. In addition, we have the global error estimate Eq. (86) and the global bound*

$$\frac{1-\theta}{1+\theta}\|c_t - d_t\| \le \|c_t - a^\dagger\| \le \left(1 + \frac{1+\theta}{\nu\alpha_t}\right)\|c_t - d_t\|. \tag{87}$$

(ii) Let

$$v_t^\dagger = \begin{cases} 2 - \upsilon - \upsilon\theta, & \text{if } \upsilon \in [1, 2/(1+\theta)), \\ \upsilon - \upsilon\theta, & \text{if } \upsilon \in (0,1), \end{cases} \qquad \epsilon = \min\left\{\frac{v_t^\dagger}{2}, \frac{\theta\ell}{L}\nu\right\}.$$

Let μ, ψ, *and* ζ *fulfill the Eq. (54). If Condition* (C2″) *and Eq. (82) hold, then* $\{a_t\}$ *converges strongly to the unique solution* $a^\dagger \in \mathrm{VI}(\mathbb{C}, \mathcal{G})$ *with an R-linear rate. Moreover, we have the error estimate Eq. (86) and the global bound Eq. (87).*

Proof. Combining Theorems 1–4 with Lemmas 6 and 12–14, the desired conclusions can be obtained. □

Remark 11. Through the proofs of Lemmas 6 and 12–14, as well as Theorems 5–7, Algorithms 3.2–3.4 also yield similar conclusions as those in Theorem 8.

Remark 12. To provide a more balanced perspective, we highlight some potential limitations of the proposed algorithms as follows.

(i) *Dependence on projection computations:* All proposed algorithms involve projections onto the feasible set at each iteration. Although this is standard in variational inequality methods, the computational cost may become significant when the projection lacks a closed-form expression or is costly to evaluate numerically.
(ii) *Parameter sensitivity:* While our methods permit a broader range of parameter choices compared to existing algorithms, their practical performance may still be sensitive to the selection of inertial and step size parameters. Improper tuning may lead to slow convergence or instability.
(iii) *Limited generalization to non-Euclidean settings:* The current convergence analysis is established in Euclidean spaces. Extending these algorithms to non-Euclidean geometries such as Riemannian manifolds or Banach spaces requires nontrivial modifications, and the theoretical guarantees do not directly carry over.

4. Numerical experiments

In this section, we provide some numerical experiments to demonstrate the computational efficiency and advantages of the proposed algorithms compared to those in [28–30,33]. All our code was implemented in MATLAB R2023b and executed on a MacBook with 8GB of memory.

Before starting our numerical experiments, it is important to introduce the *performance profiles* introduced by Dolan and Moré [45]. It is a tool used to evaluate and compare the performance of optimization methods. The main idea behind performance profiles is to provide a comprehensive and statistically robust method for benchmarking various optimization algorithms on a set of test problems. A performance profile is essentially a cumulative distribution function that quantifies the performance of each solver relative to the best solver for each problem in the test set. The steps to construct a performance profile are as follows.

(i) *Define performance measure*: Select a performance metric, such as runtime, number of iterations, or accuracy of the solution.
(ii) *Compute performance ratios*: For each solver and each test problem, compute the performance ratio, which is the solver's performance measure divided by the best performance measure among all solvers for that problem.
(iii) *Construct cumulative distribution*: For each solver, the performance profile is the cumulative distribution function of these performance ratios. This function represents the proportion of problems for which the solver's performance ratio is within a given factor of the best possible performance ratio.

More precisely, the performance ratio $c_{p,s}$ for solver s on problem p is defined as:

$$c_{p,s} := \frac{a_{p,s}}{\min\{a_{p,k} : k \in S\}},$$

where $a_{p,s}$ is the performance measure (e.g., runtime) of solver s on problem p, and $\min\{a_{p,k} \mid k \in S\}$ is the best (i.e., minimum) performance measure obtained by any solver k on problem p. It is clear to see that $c_{p,s} \geq 1$ for all $k \in S$ and $c_{p,s} = 1$ indicates that the solver achieves maximum efficiency. The scaled performance profile $\rho_s(\omega)$ for solver s at a performance ratio ω can be defined as:

$$\rho_s(\omega) := \frac{1}{n}\,\mathrm{size}\big\{p \in P : \log_2(c_{p,s}) \leq \omega\big\},$$

where n is the number of problems in the test set P, and $\log_2(c_{p,s})$ is the scaled performance ratio for solver s on problem p.

Through the definition of $\rho_s(\omega)$, we can draw the following conclusions that facilitate the analysis of the efficiency and robustness of each method: (1) when $\omega = 0$ (i.e. $c_{p,s} = 1$), it means the solver's performance is exactly the best performance for that problem; (2) as ω increases, it allows for a higher performance ratio, indicating worse performance compared to the best; (3) the value $\rho_s(\omega)$ ranges from 0 to 1, representing the proportion of problems within the test set P that solver s solves within a factor of 2^{ω} of the best performance; (4) if $\rho_s(\omega)$ is high for low values of ω, it indicates that the solver performs well on a large number of problems, achieving near-best performance; and (5) solvers that maintain high values of $\rho_s(\omega)$ over a wide range of ω are considered robust and efficient, as they consistently perform well across various problems.

Performance profiles offer several advantages: (1) they provide a clear and concise comparison of multiple solvers over a range of problems; (2) they are less sensitive to outliers and variations in individual problem difficulty; and (3) they allow for an easy visual interpretation of a solver's overall efficiency and robustness. By analyzing performance profiles, users can identify which optimization solvers perform best on average, which are most robust, and which might be suitable for specific types of problems.

Example 1. Signal processing models often involve manipulating and analyzing signals to extract useful information, enhance certain features, or remove noise. Assume we have a signal processing problem where we want to find a signal $\mathbf{t} \in \mathbb{R}^n$ that satisfies certain conditions. These conditions might include fidelity to observed data $\mathbf{y} \in \mathbb{R}^n$, and regularization to promote smoothness or sparsity. A common approach is to frame the signal processing problem as an optimization problem:

$$\min_{\mathbf{t}\in\mathbb{R}^n} \{F(\mathbf{t}) = f(\mathbf{t}) + g(\mathbf{t})\},$$

where $f(\mathbf{t})$ is a data fidelity term, often a least squares term like $\|\mathbf{St}-\mathbf{y}\|^2$ ($\mathbf{S}$ is a system matrix); $g(\mathbf{t})$ is a regularization term, such as $\lambda\|\mathbf{t}\|_1$ for promoting sparsity or $\lambda\|\nabla\mathbf{t}\|^2$ for promoting smoothness. One powerful approach for solving signal processing problems is to convert them into variational inequality models. It provides a unified framework for dealing with various constraints and optimization problems encountered in signal processing.

Here we consider a simple denoising problem where the signal $\mathbf{t}$ is to be recovered from noisy observations $\mathbf{y}$ using an ℓ_2 fidelity term and ℓ_1 regularization (Lasso problem). The optimization problem is

$$\min_{\mathbf{t}\in\mathbb{R}^n}\left\{\frac{1}{2}\|\mathbf{St}-\mathbf{y}\|^2+\gamma\|\mathbf{t}\|_1\right\}.$$

It is known that the above unconstrained optimization problem is equivalent to the following constrained optimization problem

$$\min_{\mathbf{t}\in\mathbb{R}^n}\left\{\frac{1}{2}\|\mathbf{St}-\mathbf{y}\|^2\right\},\quad \text{subject to } \|\mathbf{t}\|_1\le\zeta.$$

Define $\mathbb{C}:=\{\mathbf{t}:\ \|\mathbf{t}\|_1\le\zeta\}$. The corresponding variational inequality formulation is

$$\text{Find } \mathbf{t}^\dagger\in\mathbb{C} \text{ such that } \langle\mathbf{S}^\top(\mathbf{St}^\dagger-\mathbf{y}),\mathbf{z}-\mathbf{t}^\dagger\rangle\ge 0 \text{ for all } \mathbf{z}\in\mathbb{C}, \tag{88}$$

where $\mathbf{S}^\top$ denotes the transpose of $\mathbf{S}$. It is easy to check that the operator $\mathcal{G}(\mathbf{t})=\mathbf{S}^\top(\mathbf{St}-\mathbf{y})$ in the variational inequality model Eq. (88) is monotone and Lipschitz continuous with constant $L=\|\mathbf{S}^\top\mathbf{S}\|$. We can now use algorithms for solving variational inequalities to solve signal denoising problems.

In the following numerical experiments, we generate a matrix $\mathbf{S}\in\mathbb{R}^{m\times n}$ from a normal distribution. The original signal $\mathbf{t}\in\mathbb{R}^n$, containing k nonzero elements with values of 1 or -1, is randomly generated. The observed signal is obtained through $\mathbf{y}=\mathbf{St}+\epsilon$, where ϵ is random noise with a mean of 0 and a variance of 0.001. We uniformly set $n=1024$ and $m=512$, and choose $\zeta=k$ in the model Eq. (88). We use the mean squared error $\mathrm{MSE}=\frac{1}{n}\|\mathbf{t}^\dagger-\mathbf{t}\|^2$ to measure the iterative error between the recovered signal $\mathbf{t}^\dagger$ and the original signal $\mathbf{t}$. The iterations start from the initial values $\mathbf{t}_0=\mathbf{t}_1=\mathbf{0}$ and stop when MSE $<10^{-6}$ or after a maximum of 2000 iterations. Our numerical experiments are divided into four parts. The first three parts test the impact of different parameters on the convergence speed of the proposed Algorithms 3.1–3.4 under $k=60$, while the last part compares the stated methods with some known ones in the literature [28–30,33] under different $k\in\{40,60,80,100\}$.

1. *Testing the influence of different parameters ψ, μ, and ζ on the computational speed of the proposed algorithms.* Set $\delta_t=1+(t+1)^{-2}$, $\rho_t=(t+1)^{-1.1}$, $\zeta=0.25$ (or $\zeta=0.9\zeta_1/\zeta=0.9\zeta_2$ based on the relationship between ψ and μ as determined by Eq. (22)), $\alpha_1=0.006$, $\theta=0.6$, and $\upsilon=0.9$ for our Algorithms 3.1–3.4. Choose $\kappa=1.5$ for the proposed Algorithms 3.3 and 3.4. We consider different choices for parameter $\psi\in\{0,0.02,0.04,0.06,0.08,0.1\}$ and parameter $\mu\in\{0,0.1,0.2,0.4,0.6,0.8,1.0\}$. Table 3 shows the number of iterations required for termination of the proposed Algorithms 3.1–3.4 under different parameters ψ, μ, and considering different relaxation parameter ζ. In Table 3, the first result under different parameters indicates when $\zeta=0.25$, and the second value denotes when $\zeta=0.9\zeta_1$ or $\zeta=0.9\zeta_2$.
2. *Testing the influence of different parameters υ and ζ on the computational speed of the proposed algorithms.* Set $\delta_t=1+(t+1)^{-2}$, $\rho_t=(t+1)^{-1.1}$, $\psi=0.1$, $\mu=1.0$, $\alpha_1=0.006$, and $\theta=0.6$ for our Algorithms 3.1–3.4. Choose $\kappa=1.5$ for the proposed Algorithms 3.3 and 3.4. We consider different choices for parameter $\upsilon\in\{0.8,0.9,1.0,1.1,1.2\}$ and parameter $\zeta\in\{0.1,0.15,0.2,0.25,0.35,0.41\}$. The number of iterations required for termination under different parameters υ and ζ for the proposed algorithms is shown in Table 4.
3. *Testing the influence of different step sizes α_t on the computational speed of the proposed algorithms.* Set $\psi=0.1$, $\mu=1.0$, $\zeta=0.41$, $\alpha_1=0.006$, $\theta=0.6$, and $\upsilon=0.9$ (or $\upsilon=1.0$) for our Algorithms 3.1–3.4. Choose $\kappa=1.5$ for the proposed Algorithms 3.3 and 3.4. We consider four different choices for the step size α_t (cf. Eqs. (5) and (64)) as follows:

   Case 1: $\delta_t=1,\ \rho_t=0$, Case 2: $\delta_t=1,\ \rho_t=(t+1)^{-1.1}$,

   Case 3: $\delta_t=1+(t+1)^{-2},\ \rho_t=0$, Case 4: $\delta_t=1+(t+1)^{-2},\ \rho_t=(t+1)^{-1.1}$.

   Table 5 presents the numerical results of the proposed algorithms under different choices of step size selection and parameter υ. In Table 5, “Iter.” denotes the number of iterations required for termination, and “Time” indicates the execution time in seconds.
4. *Comparing the computational speed of the proposed algorithms with some double inertial methods under different sparsity levels.* For simplicity of presentation, we abbreviate the compared algorithms as follows: the Algorithm 1 of Yao et al. [28] is referred to as YIS Alg. 1, the Algorithm 1 of Thong et al. [29] is labeled as TDAT Alg. 1, the Algorithm 1 of Wang et al. [30] is termed as WWIS Alg. 1, and the Algorithm 1 of Li et al. [33] is denoted as LWW Alg. 1. The parameter settings for these algorithms are as follows. Set $\delta_t=1+(t+1)^{-2}$, $\rho_t=(t+1)^{-1.1}$, $\psi=0.1$, $\mu=1.0$, $\zeta=0.41$, $\upsilon=0.9$, $\alpha_1=0.006$, and $\theta=0.6$ for our Algorithms 3.1–3.4. Choose $\kappa=1.5$ for the proposed Algorithms 3.3 and 3.4. Select $\psi=0.0019$, $\mu=1.0$, $\zeta=0.33$, $\alpha_1=0.006$, and $\theta=0.6$ for all four algorithms compared. Take $\rho_t=(t+1)^{-1.1}$ for TDAT Alg. 1. Pick $\kappa=1.5$ for WWIS Alg. 1 and LWW Alg. 1. Set $\sigma=2$ and $\ell=0.5$ for LWW Alg. 1. It is worth noting that we did not choose the parameter values from the original papers [28–30,33] for these comparison algorithms, as their performance with the original parameter values was worse in this example. We selected appropriate parameters for the comparison algorithms by following the principle of consistency in meeting the conditions. Next, we test the numerical performance of all algorithms with $m=512$, $n=1204$, and different parameters $k\in\{40,60,80,100\}$. To clearly illustrate the recovery effect, Fig. 1 shows the original and degraded signals, and as an example, Fig. 2 displays the recovery results of the proposed Algorithm 3.3 under different conditions. Table 6 presents the number of iterations to termination and execution time for all algorithms under various conditions. Fig. 3 visually demonstrates the trend of MSE versus iterations for all algorithms

under different conditions. Finally, we designed a set of test problems with $m = 512$, $n \in \{1024, 1074, 1124, 1174, \dots, 4074\}$, and $k = 40$. Figs. 4 and 5 illustrate the performance profiles of termination iteration count and execution time for all algorithms on this benchmark test problem, respectively.

Remark 13. From the numerical results of Example 1, specifically Tables 3–6 and Figs. 1–5, we can draw the following preliminary conclusions:

Table 3
Numerical results of the proposed algorithms with different parameters ψ, μ, and ζ.

Algorithms	ψ	μ						
		0	0.1	0.2	0.4	0.6	0.8	1.0
Our Algorithm 3.1	0	1345 (373)	1310 (**377**)	1276 (**381**)	1207 (**388**)	1139 (**396**)	1071 (**403**)	1003 (**410**)
	0.02	1324 (373)	1289 (378)	1255 (382)	1187 (391)	1118 (399)	1050 (407)	983 (415)
	0.04	1303 (372)	1269 (378)	1234 (384)	1166 (394)	1097 (403)	1029 (411)	962 (420)
	0.06	1282 (371)	1248 (379)	1213 (386)	1145 (397)	1077 (407)	1009 (416)	941 (425)
	0.08	1261 (371)	1227 (381)	1192 (388)	1124 (401)	1056 (412)	988 (421)	921 (430)
	0.1	**1240** (**370**)	**1206** (382)	**1172** (391)	**1103** (405)	**1035** (417)	**967** (427)	**900** (436)
Our Algorithm 3.2	0	1087 (296)	1059 (**299**)	1032 (**302**)	976 (**309**)	921 (**315**)	866 (**322**)	813 (**328**)
	0.02	1070 (295)	1042 (300)	1015 (303)	959 (311)	904 (318)	849 (325)	796 (331)
	0.04	1053 (295)	1026 (300)	998 (305)	942 (313)	887 (321)	832 (328)	779 (335)
	0.06	1036 (294)	1009 (301)	981 (306)	925 (316)	870 (324)	815 (331)	762 (338)
	0.08	1019 (293)	992 (302)	964 (308)	908 (318)	853 (327)	798 (334)	745 (342)
	0.1	**1002** (**293**)	**975** (303)	**947** (310)	**892** (321)	**836** (330)	**781** (338)	**728** (347)
Our Algorithm 3.3	0	708 (194)	690 (**195**)	673 (**198**)	638 (**203**)	603 (**206**)	569 (**210**)	534 (**216**)
	0.02	697 (197)	680 (198)	662 (203)	627 (204)	593 (208)	558 (212)	523 (218)
	0.04	687 (191)	669 (196)	652 (199)	617 (203)	582 (210)	547 (214)	512 (221)
	0.06	676 (196)	658 (199)	641 (202)	606 (208)	571 (214)	536 (218)	501 (223)
	0.08	665 (**190**)	648 (197)	630 (201)	595 (208)	560 (216)	526 (221)	491 (225)
	0.1	**654** (194)	**637** (198)	**619** (204)	**585** (211)	**550** (217)	**515** (221)	**480** (227)
Our Algorithm 3.4	0	773 (208)	754 (**209**)	735 (212)	697 (**216**)	659 (**221**)	622 (**226**)	584 (**231**)
	0.02	761 (207)	742 (209)	723 (**211**)	686 (216)	648 (223)	610 (226)	572 (232)
	0.04	749 (207)	730 (210)	712 (212)	674 (219)	636 (224)	598 (230)	560 (235)
	0.06	738 (207)	719 (210)	700 (213)	662 (220)	624 (226)	587 (232)	549 (237)
	0.08	726 (207)	707 (211)	688 (215)	650 (221)	613 (227)	575 (233)	537 (239)
	0.1	**714** (**207**)	**695** (212)	**677** (216)	**639** (223)	**601** (229)	**563** (235)	**525** (241)

Table 4
Numerical results of the proposed algorithms with different parameters υ and ζ.

Algorithms	υ	ζ					
		0.1	0.15	0.2	0.25	0.35	0.41
Our Algorithm 3.1	0.8	2000	1947	1366	1016	623	486
	0.9	2000	1728	1210	900	553	436
	1.0	2000	1551	1084	802	500	397
	1.1	2000	1404	978	728	458	361
	1.2	2000	**1281**	**889**	**669**	**423**	**335**
Our Algorithm 3.2	0.8	**1987**	**1246**	**871**	**645**	**394**	**304**
	0.9	2000	1399	979	728	444	347
	1.0	2000	1551	1084	802	500	397
	1.1	2000	1696	1172	882	551	436
	1.2	2000	1857	1298	967	604	478
Our Algorithm 3.3	0.8	**1284**	**809**	**571**	**426**	**260**	**200**
	0.9	1439	906	640	480	290	227
	1.0	1592	1003	708	531	326	249
	1.1	1742	1098	775	581	360	278
	1.2	1886	1189	839	629	389	297
Our Algorithm 3.4	0.8	1581	996	704	529	**317**	242
	0.9	1568	989	699	525	318	**241**
	1.0	1555	981	694	522	321	244
	1.1	1540	971	688	517	322	247
	1.2	**1518**	**958**	**678**	**511**	319	247

Table 5
Numerical results of the proposed algorithms with different parameters υ and α_t.

Cases	Our Algorithm 3.1 Iter.	Our Algorithm 3.1 Time (s)	Our Algorithm 3.2 Iter.	Our Algorithm 3.2 Time (s)	Our Algorithm 3.3 Iter.	Our Algorithm 3.3 Time (s)	Our Algorithm 3.4 Iter.	Our Algorithm 3.4 Time (s)
				$\upsilon = 0.9$				
Case 1	553	0.523	2000	1.953	351	0.481	370	0.425
Case 2	436	0.455	347	0.306	227	0.293	241	0.258
Case 3	508	0.550	2000	1.858	287	0.510	297	0.318
Case 4	**436**	0.374	**347**	0.300	**227**	0.290	**241**	0.265
				$\upsilon = 1.0$				
Case 1	504	0.527	504	0.482	330	0.449	335	0.398
Case 2	397	0.338	397	0.545	249	0.320	244	0.282
Case 3	455	0.378	455	0.398	280	0.505	257	0.298
Case 4	**397**	0.297	**397**	0.332	**249**	0.308	**244**	0.281

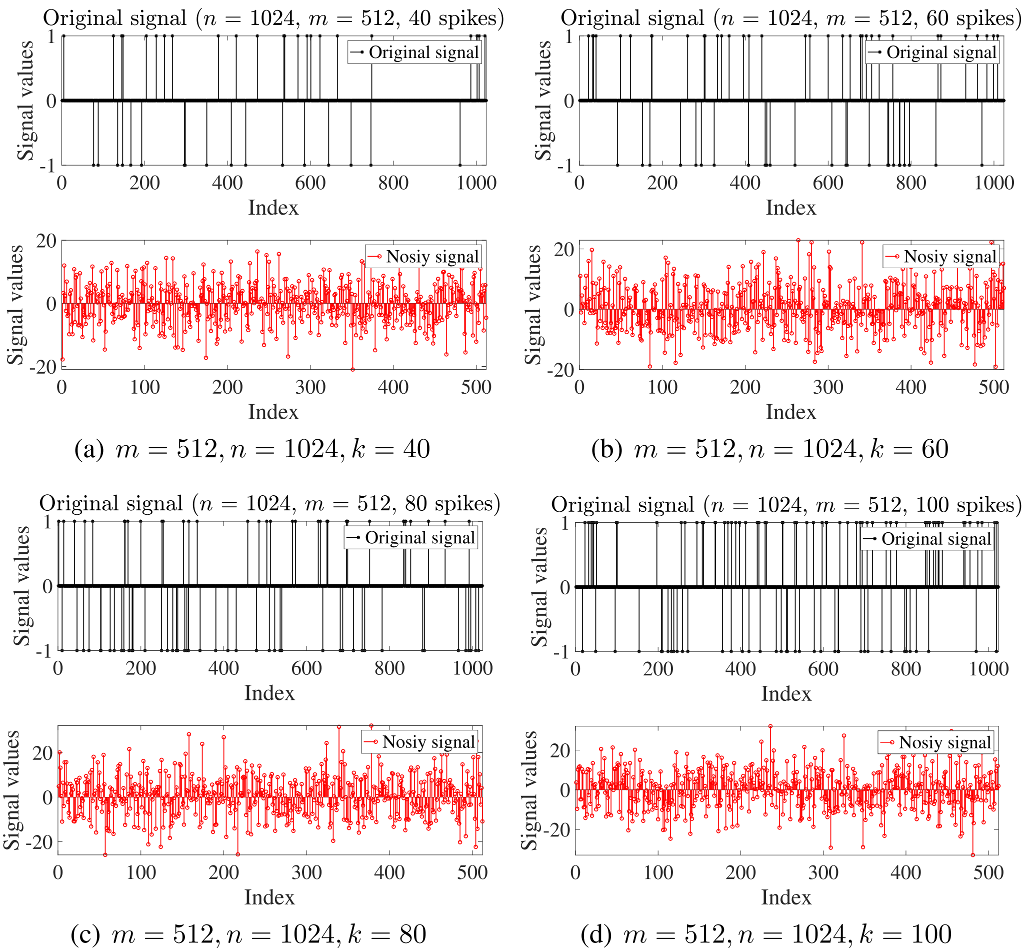
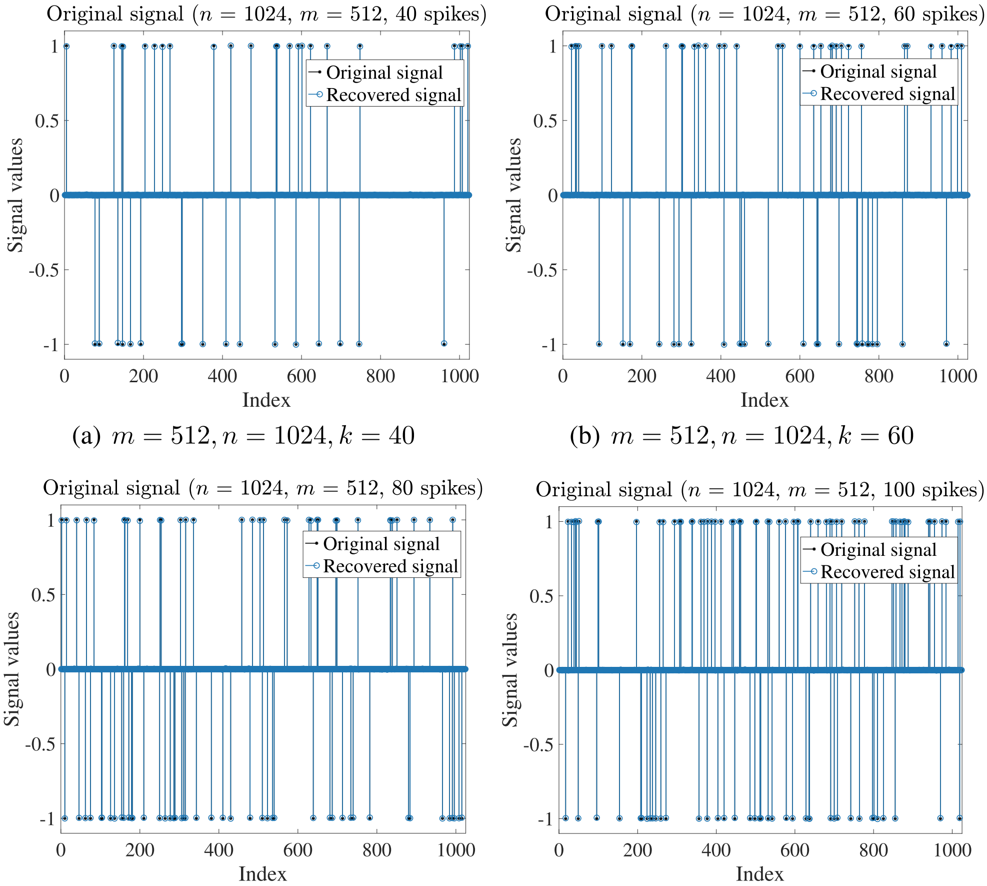


Fig. 1. The original and noisy signals at different sparsities.

(i) As shown in Table 3, when ψ is fixed and $\zeta = 0.25$ is selected, the number of iterations required by Algorithms 3.1–3.4 decreases as μ increases. However, this phenomenon reverses when $\zeta = 0.9\zeta_1$ (or $\zeta = 0.9\zeta_2$) is chosen while ψ is fixed. Additionally, when μ is fixed and $\zeta = 0.25$ is selected, the number of iterations required by Algorithms 3.1–3.4 decreases as ψ increases. Similarly, this phenomenon reverses when $\zeta = 0.9\zeta_1$ (or $\zeta = 0.9\zeta_2$) is chosen while μ is fixed. These observations indicate that, when ζ is fixed, larger parameters ψ and μ play a positive role in enhancing computational speed of the proposed algorithms. However, when different values of ζ are selected based on ψ and μ, smaller parameters ψ and μ can improve computational efficiency (also noting that when $\mu = 0$, larger values of ψ result in fewer iterations). It is important to note that choosing larger values of ζ



(a) $m = 512, n = 1024, k = 40$ (b) $m = 512, n = 1024, k = 60$

(c) $m = 512, n = 1024, k = 80$ (d) $m = 512, n = 1024, k = 100$

Fig. 2. The original signal and the signal recovered by our Algorithm 3.3 under different sparsities.

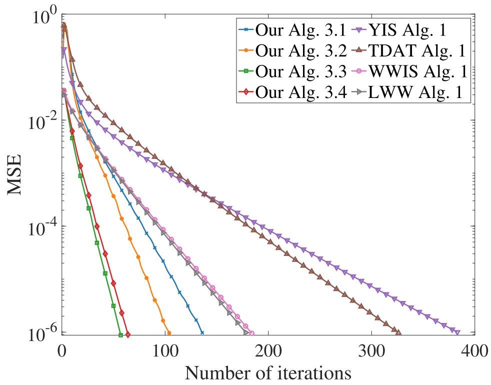
Table 6
Numerical results of the proposed algorithms with other methods at different sparsities.

Algorithms	Iter. $k = 40$	Time (s)	Iter. $k = 60$	Time (s)	Iter. $k = 80$	Time (s)	Iter. $k = 100$	Time (s)
Our Algorithm 3.1	136	0.172	277	0.288	337	0.326	678	0.699
Our Algorithm 3.2	104	0.107	213	0.354	247	0.377	515	0.557
Our Algorithm 3.3	57	0.080	122	0.235	148	0.207	315	0.385
Our Algorithm 3.4	64	0.082	134	0.200	165	0.193	340	0.348
YIS Alg. 1	383	0.331	816	0.744	987	0.803	2000	1.787
TDAT Alg. 1	326	0.341	635	0.684	761	0.815	1483	1.522
WWIS Alg. 1	184	0.238	393	0.598	472	0.485	953	0.921
LWW Alg. 1	178	1.543	370	3.538	447	4.258	884	8.358

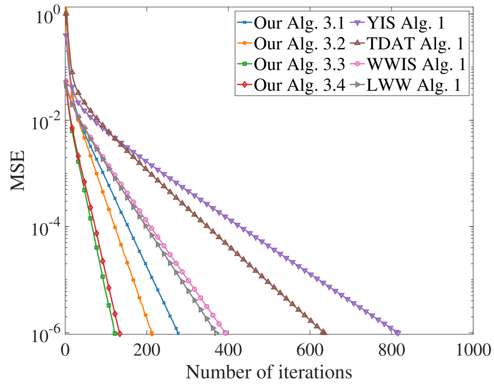
significantly increases the computational speed when ψ and μ are fixed, highlighting the crucial role of the relaxation parameter ζ in computational speed. Therefore, the broader range of ζ values allowed by our algorithms is beneficial.

(ii) According to Table 4, we obtain: (1) when υ is fixed, the number of iterations required by Algorithms 3.1–3.4 decreases as ζ increases; (2) when ζ is fixed, Algorithm 3.1 performs better with larger values of υ, while Algorithms 3.2–3.4 achieve faster computation speeds with smaller values of υ. These results also underscore the significant role of parameters υ and ζ in enhancing the computational efficiency of the algorithms.

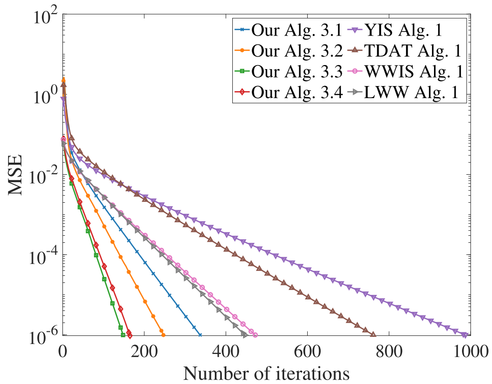
(iii) Table 5 demonstrates that the use of a nonmonotone step size rule significantly improves the computational speed compared to using a nonincreasing step size rule. Specifically, Algorithms 3.1–3.4 achieve fewer iterations when employing a nonmonotone step size rule (see Case 2–Case 4) compared to a nonincreasing step size rule (see Case 1) when υ is fixed. Additionally, Table 5 illustrates the combined impact of using different step size rules and different values of υ on the computational speed of Algorithms 3.1–3.4. It is also noteworthy that the number of iterations required by Algorithms 3.1–3.4 under step size Case 2 and step size Case 4 are identical, due to the differences in the step size sequences generated by Case 2 and Case 4 are minimal in this example. Moreover,



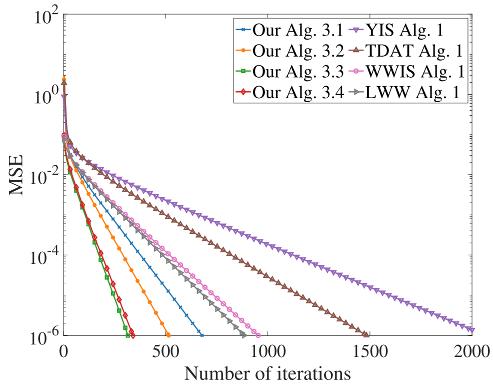
(a) $m = 512, n = 1024, k = 40$



(b) $m = 512, n = 1024, k = 60$



(c) $m = 512, n = 1024, k = 80$



(d) $m = 512, n = 1024, k = 100$

Fig. 3. Numerical performance of MSE for all algorithms under different sparsities.

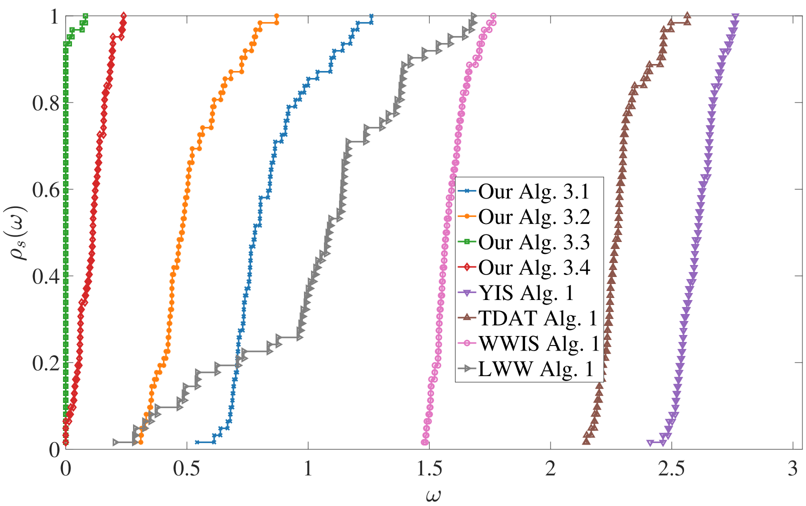


Fig. 4. Performance profiles of all algorithms based on number of iterations.

it is observed that the number of iterations required by Algorithms 3.1 and 3.2 is the same when $\upsilon = 1$. This can be easily explained because Algorithms 3.1 and 3.2 are equivalent when $\upsilon = 1$.

(iv) Figs. 1 and 2 clearly show that the proposed Algorithm 3.3 can effectively recover the original signals under different conditions.

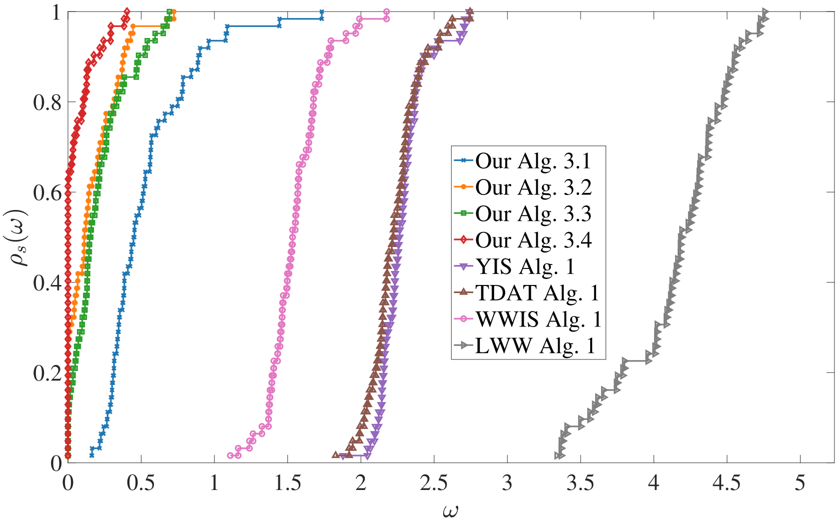


Fig. 5. Performance profiles of all algorithms based on execution time.

(v) The information from Table 6 and Fig. 3 indicates that the proposed Algorithms 3.1–3.4 significantly outperform the ones in [28–30,33] in terms of the required number of iterations and computation time under various conditions. This demonstrates that our algorithms are both efficient and robust.

(vi) From Fig. 4, we have: (1) our Algorithm 3.3 shows the best performance, reaching $\rho_s(\omega) = 1$ almost immediately as ω increases from 0. This indicates that it consistently achieves the best or near-best number of iterations across all problems; (2) our Algorithm 3.4 also performs very well, reaching $\rho_s(\omega) = 1$ quickly but slightly after our Algorithm 3.3; (3) our Algorithms 3.1 and 3.2 show good performance but are not as consistently efficient as the previous two. They reach $\rho_s(\omega) = 1$ at a slightly higher ω value, indicating some problems where they do not perform as well; and (4) the compared methods show significantly lower performance with respect to our proposed algorithms. They reach $\rho_s(\omega) = 1$ at much higher ω values, demonstrating that they are less efficient in terms of number of iterations across the test problems. Similarly, as shown in Fig. 5, we observe the following: (1) the proposed Algorithm 3.4 achieves the best performance on most of the test problems in terms of execution time, and the proposed Algorithms 3.1–3.3 can achieve near-optimal performance with smaller values of ω (e.g., $\omega = 1$); and (2) the compared algorithms are significantly less efficient in terms of execution time, as they reach $\rho_s(\omega) = 1$ at much higher ω values. This indicates that they require more execution time to solve the problems compared to our proposed algorithms.

(vii) From Tables 3–6 and Figs. 4 and 5, it is evident that the projection and contraction algorithms (i.e., Algorithm 3.3, Algorithm 3.4, WWIS Alg. 1 [30], and LWW Alg. 1 [33]) converge significantly faster than the subgradient extragradient algorithms (i.e., Algorithms 3.1and 3.2, YIS Alg. 1 [28], and TDAT Alg. 1 [29]). In some cases, the convergence speed is twice as fast. This further validates the efficiency of the projection and contraction algorithms.

In conclusion, our proposed algorithms outperform the other ones in [28–30,33] significantly in both execution time and number of iterations, demonstrating superior overall performance.

5. Conclusions

In this paper, we explore four modified extragradient algorithms with double inertial steps and relaxation steps designed to solve VIPs in real Hilbert spaces. These algorithms are inspired by the subgradient extragradient algorithm, projection and contraction algorithms, and the double inertial method. The key advantage of our approaches lies in the incorporation of double inertial steps and adaptive step sizes, which accelerate the convergence of the algorithms. In our convergence analysis, we prove weak convergence theorems under the assumption of pseudomonotonicity (or quasimonotonicity) of the operator. Additionally, we establish strong and linear convergence properties when the operator is strongly pseudomonotone. The proposed algorithms achieve a nonasymptotic $O(1/t)$ convergence rate under pseudomonotonicity (or quasimonotonicity), and we derive global error bounds when the operator is strongly pseudomonotone. Finally, numerical experiments demonstrate the computational efficiency of the proposed algorithms in comparison to several state-of-the-art methods.

As future work, several directions are worth exploring. First, since the proposed algorithms rely on projection computations at each iteration, it is of interest to develop projection-free or projection-reduced variants, particularly for problems where the projection onto the feasible set is expensive or lacks a closed-form expression. Second, designing more robust parameter selection strategies, including adaptive or self-tuning schemes, would improve the practical performance and stability of the algorithms, especially in ill-conditioned settings. Lastly, extending the current framework to non-Euclidean spaces, such as Riemannian manifolds or general

Banach spaces, remains a challenging yet promising direction that would broaden the applicability of these methods to more complex geometric settings.

Author contributions

Bing Tan, Jiawei Chen, Songxiao Li, and Xiaoqing Ou wrote and edited the original manuscript. Bing Tan conducted the numerical experiments. Songxiao Li provided research suggestions. All authors read and approved the final manuscript for publication.

Ethical approval

Not applicable.

Conflict of interest

The authors declare no competing interests.

CRediT authorship contribution statement

Bing Tan: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization; **Jiawei Chen:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Formal analysis, Conceptualization; **Songxiao Li:** Writing – review & editing, Writing – original draft, Investigation, Funding acquisition, Formal analysis, Conceptualization; **Xiaoqing Ou:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Conceptualization.

Data availability

Data sharing is not applicable for this article as no datasets were generated or analyzed during the current study.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bing Tan reports financial support was provided by Natural Science Foundation Project of Chongqing. Songxiao Li reports financial support was provided by National Natural Science Foundation of China. Jiawei Chen reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

B. Tan thanks the supports of the Natural Science Foundation of Chongqing (No. CSTB2024NSCQ-MSX0354), the Fundamental Research Funds for the Central Universities (No. SWU-KQ24052 and No. SWU-KR25013), and the National Natural Science Foundation of China (No. 12471473). J.W. Chen was supported by the National Natural Science Foundation of China (Nos. 12071379 and 12471299), the National Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0925, cstc2022ycjh-bgzxm0097). S. Li acknowledges the National Natural Science Foundation of China (No. 12371131), GuangDong Basic and Applied Basic Research Foundation (Nos. 2024A1515012404 and 2023A1515010614), and STU Scientific Research Initiation Grant (No. NTF23004).

References

[1] Cottle RW, Yao JC. Pseudo-monotone complementarity problems in Hilbert space. J Optim Theory Appl 1992;75:281–95.
[2] Ye M, He Y. A double projection method for solving variational inequalities without monotonicity. Comput Optim Appl 2015;60:141–50.
[3] 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.
[4] Tan B, Qin X. Two relaxed inertial forward-backward-forward algorithms for solving monotone inclusions and an application to compressed sensing. Canad J Math 2025. https://doi.org/10.4153/S0008414X24000889
[5] Ma X, Liu H, Li X. Two optimization approaches for solving split variational inclusion problems with applications. J Sci Comput 2022;91:58.
[6] Han K, Eve G, Friesz TL. Computing dynamic user equilibria on large-scale networks with software implementation. Netw Spat Econ 2019;19:869–902.
[7] Korpelevich GM. The extragradient method for finding saddle points and other problems. Ekonomika i Mat Metody 1976;12:747–56.
[8] Antipin AS. On a method for convex programs using a symmetrical modification of the lagrange function. Ekonomika i Mat Metody 1976;12:1164–73.
[9] He BS. A class of projection and contraction methods for monotone variational inequalities. Appl Math Optim 1997;35:69–76.
[10] Tseng P. A modified forward-backward splitting method for maximal monotone mappings. SIAM J Control Optim 2000;38:431–46.
[11] Censor Y, Gibali A, Reich S. Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space. Optim Methods Softw 2011;26:827–45.
[12] 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.
[13] Malitsky Y. Projected reflected gradient methods for monotone variational inequalities. SIAM J Optim 2015;25:502–20.
[14] Vuong PT. On the weak convergence of the extragradient method for solving pseudo-monotone variational inequalities. J Optim Theory Appl 2018;176:399–409.
[15] Ceng LC, Petruşel A, Qin X, Yao JC. Pseudomonotone variational inequalities and fixed points.; Fixed Point Theory 2021;22:543–58.
[16] Liu H, Yang J. Weak convergence of iterative methods for solving quasimonotone variational inequalities. Comput Optim Appl 2020;77:491–508.
[17] Ogwo GN, Izuchukwu C, Shehu Y, Mewomo OT. Convergence of relaxed inertial subgradient extragradient methods for quasimonotone variational inequality problems. J Sci Comput 2022;90:10.

[18] Hu S, Wang Y, Dong QL. Convergence analysis of a new Bregman extragradient method for solving fixed point problems and variational inequality problems in reflexive Banach spaces. J Sci Comput 2023;96:19.
[19] Yang J, Liu H, Liu Z. Modified subgradient extragradient algorithms for solving monotone variational inequalities. Optimization 2018;67:2247–58.
[20] Yang J, Liu H. Strong convergence result for solving monotone variational inequalities in Hilbert space. Numer Algorithms 2019;80:741–52.
[21] Denisov SV, Semenov VV, Chabak LM. Convergence of the modified extragradient method for variational inequalities with non-Lipschitz operators. Cybern Syst Anal 2015;51:757–65.
[22] Polyak BT. Some methods of speeding up the convergence of iterates methods. USSR Comput Math Math Phys 1964;4:1–17.
[23] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imaging Sci 2009;2:183–202.
[24] Izuchukwu C, Shehu Y, Yao JC. New inertial forward-backward type for variational inequalities with quasi-monotonicity. J Global Optim 2022;84:441–64.
[25] Shehu Y, Iyiola OS, Reich S. A modified inertial subgradient extragradient method for solving variational inequalities. Optim Eng 2022;23:421–49.
[26] Dong QL, Cho YJ, Zhong LL, Rassias TM. Inertial projection and contraction algorithms for variational inequalities. J Glob Optim 2018a;70:687–704.
[27] Dong QL, Cho YJ, Rassias TM. General inertial Mann algorithms and their convergence analysis for nonexpansive mappings; Vol. 134 Cham. Springer; 2018.
[28] Yao Y, Iyiola OS, Shehu Y. Subgradient extragradient method with double inertial steps for variational inequalities. J Sci Comput 2022;90:71.
[29] Thong DV, Dung VT, Anh PK, Thang HV. A single projection algorithm with double inertial extrapolation steps for solving pseudomonotone variational inequalities in Hilbert space. J Comput Appl Math 2023;426:115099.
[30] Wang K, Wang Y, Iyiola OS, Shehu Y. Double inertial projection method for variational inequalities with quasi-monotonicity. Optimization 2024;73:707–39.
[31] Wang K, Wang Y, Shehu Y, Jiang BN. Double inertial forward-backward-forward method with adaptive step-size for variational inequalities with quasi-monotonicity. Commun Nonlinear Sci Numer Simul 2024;132:107924.
[32] Thong DV, Li XH, Dung VT, Thang HV, Long LV. A modified subgradient extragradient method with non-monotonic step sizes for solving quasimonotone variational inequalities. Comput Appl Math 2024;43:198.
[33] Li H, Wang X, Wang F. Projection and contraction method with double inertial steps for quasi-monotone variational inequalities. Optimization 2025;74:1643–74.
[34] Wang Z, Lei Z, Long X, Chen Z. A modified Tseng splitting method with double inertial steps for solving monotone inclusion problems. J Sci Comput 2023;96:92.
[35] Yao Y, Adamu A, Shehu Y, Yao JC. Forward-reflected-backward algorithms with double inertial extrapolations for variational inequalities with applications to image processing. Numer Algorithms 2025. https://doi.org/10.1007/s11075-025-02095-9
[36] Huang Y, You L, Cai G, Dong QL. Subgradient extragradient algorithm with double inertial steps for solving variational inequality problems and fixed point problems in Hilbert spaces. Rend Circ Mat Palermo, II Ser 2025;74:129.
[37] Bauschke HH, Combettes PL. Convex analysis and monotone operator theory in Hilbert spaces. CMS books in mathematics. Cham; Springer; 2017.
[38] 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. Comput Optim Appl 2014;57:339–63.
[39] 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.
[40] Ortega JM, Rheinboldt WC. Iterative solution of nonlinear equations in several variables. New York: Academic Press; 1970.
[41] Maingé P-E. Convergence theorems for inertial KM-type algorithms. J Comput Appl Math 2008;219:223–36.
[42] Opial Z. Weak convergence of the sequence of successive approximations for nonexpansive mappings. Bull Amer Math Soc 1967;73:591–7.
[43] Osilike MO, Aniagbosor SC. Weak and strong convergence theorems for fixed points of nonasymptotically nonexpansive mappings. Math Comput Model 2000;32:1181–91.
[44] Kim DS, Vuong PT, Khanh PD. Qualitative properties of strongly pseudomonotone variational inequalities. Optim Lett 2016;10:1669–79.
[45] Dolan ED, Moré JJ. Benchmarking optimization software with performance profiles. Math Program 2002;91:201–13.