AN EXTRA GRADIENT ANDERSON-ACCELERATED ALGORITHM FOR PSEUDOMONOTONE VARIATIONAL INEQUALITIES

XIN QU, WEI BIAN, AND XIAOJUN CHEN

ABSTRACT. This paper proposes an extra gradient Anderson-accelerated algorithm for solving pseudomonotone variational inequalities, which uses the extra gradient scheme with line search to guarantee the global convergence and Anderson acceleration to have fast convergent rate. We prove that the sequence generated by the proposed algorithm from any initial point converges to a solution of the pseudomonotone variational inequality problem without assuming the Lipschitz continuity and contractive condition, which are used for convergence analysis of the extra gradient method and Anderson-accelerated method, respectively in existing literatures. Numerical experiments, particular emphasis on Harker-Pang problems, fractional programming problems, nonlinear complementarity problems, PDE problems with free boundary and linear complementarity problems, are conducted to validate the effectiveness and good performance of the proposed algorithm comparing with the extra gradient method and Anderson-accelerated method.

1. INTRODUCTION

In this paper, we consider the following variational inequality (VI) problem: find an $x^* \in \Omega$ such that

$$\langle H(x^*), x - x^* \rangle \geq 0, \quad \forall x \in \Omega, \tag{1.1}$$

where Ω is a closed convex set of $\mathbb{R}^n$ and $H : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous function, and pseudomonotone on Ω, but not necessarily smooth or even Lipschitz continuous. Throughout this paper, we denote problem (1.1) by $\mathrm{VI}(\Omega, H)$ and the solution set of (1.1) by $\mathrm{SOL}(\Omega, H)$, and assume $\mathrm{SOL}(\Omega, H) \neq \emptyset$.

Variational inequalities (VIs) provide a unified framework for representing various important concepts in applied mathematics such as nonlinear equation systems, complementarity problems, optimality conditions for optimization problems and network equilibrium problems. Thus, VIs have a wide range of applications in physics, economics, engineering sciences and so on [6, 16, 25, 39, 45]. One of the most interesting topics in VIs is to develop efficient and fast iterative algorithms to find solutions.

As a class of effective numerical methods for solving VIs, projection methods have received a lot of attention from many researchers. The earliest projection method for solving $\mathrm{VI}(\Omega, H)$ is the gradient projection (PG) method [22]

$$x_{k+1} = P_\Omega(x_k - tH(x_k)), \tag{1.2}$$

where P_Ω denotes the projection onto the set Ω and $t > 0$. To guarantee the convergence of the PG method, it is usually assumed that either H is strongly monotone and L-Lipschitz

2020 *Mathematics Subject Classification.* Primary 47J20, 65K10, 65Y20; Secondary 90C30, 65H10.

Key words and phrases. Pseudomonotone variational inequality, extra gradient algorithm, Anderson acceleration, sequence convergence, PDE problem with free boundary.

This paper is supported in part by National Natural Science Foundation of China grants (12271127, 62176073) and Hong Kong Research Grant Council grants (PolyU15300123, PolyU15300021).

continuous on Ω[1], or H is cocoercive. In order to weaken the strong monotonicity of H, many projection methods have been proposed to solve monotone VIs [1, 11, 18, 12, 34, 35, 47, 48, 49]. These methods include using some acceleration techniques to solve monotone VIs. For example, an adaptive golden ratio algorithm was proposed in [1] for solving VIs where the operator is monotone and locally Lipschitz continuous. Specifically, when the feasible set is $\mathbb{R}^n$, the monotonicity assumption of the operator in [1] can be replaced by the existence of a weak Minty solution to VIs. Under these assumptions, convergence rates were provided for both the gap function and the residual function. In [18], the authors introduced an inertial projection and contraction algorithm and proved that, under the conditions that the operator is monotone and Lipschitz continuous, the sequence generated by the algorithm weakly converges to a solution of VIs in a Hilbert space.

Very recently, developing algorithms for different nonmonotone VIs has attracted great attention due to applications in machine learning [10, 32, 42, 43, 46, 52]. In this paper, we focus on solving $\mathrm{VI}(\Omega, H)$ with a pseudomonotone H, which is a widely used class of nonmonotone VIs. A most commonly used algorithm in the literatures for solving pseudomonotone VIs is the Korpelevich's extra gradient (EG) method [31]. The EG method was originally used to solve $\mathrm{VI}(\Omega, H)$ with a monotone and L-Lipschitz continuous H, and was later extended by Pang and Facchinei [19] to solve the pseudomonotone VIs. After that, the EG method has been intensively studied and extended in various ways [10, 11, 12, 17, 23, 27, 32, 33, 35, 42, 43, 46, 49, 50]. It is worth pointing out that besides the L-Lipschitz continuity of H on Ω, some other conditions on H are often required to guarantee the convergence of the EG method and its variants, such as Minty condition, quasimonotonicity, pseudomonotonicity and weak monotonicity.

To the best of our knowledge, current research on the Minty condition cannot simultaneously provide an analysis of both sequence convergence and convergence rates. One of the main objectives of this paper is to design an algorithm that ensures both the sequence convergence and a fast convergence rate under the pseudomonotonicity of H on Ω. In subsection 1.1, we will review and summarize the development of EG methods for nonmonotone VIs in recent years in detail. A summary of some main comparisons is provided in Table 1.

It is known that $\mathrm{VI}(\Omega, H)$ is equivalent to the following fixed point problem [19]:

$$x = G(x) := P_\Omega(x - tH(x)), \tag{1.3}$$

where $t > 0$. Thus, the study of PG method in (1.2), which is a fixed point method, can help to improve the performance of the algorithms for solving VIs. Anderson acceleration is efficient to improve the convergence rate of fixed point methods, but the existing convergence analysis of Anderson acceleration requires the fixed point mapping G to be contractive and piecewise smooth. However, due to the nonmonotonicity of H, the mapping G is not contractive, and may be even expansive, which means that we cannot directly use Anderson acceleration to solve the corresponding fixed point problem of $\mathrm{VI}(\Omega, H)$ in (1.3). Moreover, G is nonsmooth, because of the projection operator P_Ω. In this paper, we introduce Anderson acceleration technique into the EG method to ensure the global sequence convergence of the algorithm and improve the convergence rate of the EG method. We will review the recent development of Anderson acceleration for fixed point problems in subsection 1.2.

[1] H is called L-Lipschitz continuous on Ω if $\|H(x) - H(y)\| \le L\|x - y\|$ for any $x, y \in \Omega$.

1.1. **EG methods for solving nonmonotone VIs.** The structure of the EG method proceeds as follows:

$$\begin{cases} x_{k+0.5} := P_{\Omega}(x_k - tH(x_k)) \\ x_{k+1} := P_{\Omega}(x_k - tH(x_{k+0.5})) \end{cases}$$

with $t > 0$. It is important to note that the convergence results for the EG method require H to be L-Lipschitz continuous on Ω and the stepsize t to satisfy $tL < 1$ [31]. Iusem [2] proposed a modified EG method with an updated stepsize to guarantee the efficiency of the proposed algorithm for $\mathrm{VI}(\Omega, H)$, in which H is monotone and continuous. Recently, similar extensions have been developed not only for monotone operators but also for pseudomonotone operators [10, 42, 43]. However, the convergence rates are not mentioned in these works. Most results on convergence rates of EG methods for VIs are established based on the L-Lipschitz continuity of H, resulting in a sublinear rate of convergence for the best-iterate of the residual term. In particular, when H is monotone and L-Lipschitz continuous, we known that the EG method converges to a solution of $\mathrm{VI}(\Omega, H)$ in terms of $\min_{0\le k\le N} \|x_k - x_{k+0.5}\|$ with a rate of $O(1/\sqrt{N})$ [40], which has been extended to the EG method for solving $\mathrm{VI}(\Omega, H)$ with a pseudomonotone and L-Lipschitz continuous H in [19, Lemma 12.1.10].

Furthermore, research on nonmonotone VIs under Minty condition has been conducted. We say $\mathrm{VI}(\Omega, H)$ satisfies the Minty condition if there exists an $x^* \in \Omega$ such that

$$\langle H(x), x - x^* \rangle \ge 0, \ \forall x \in \Omega. \tag{1.4}$$

Under the Minty condition, Ye and He in [52] introduced a double projection algorithm (DPA) with global convergence on the sequence, which requires computing the projection onto the intersection of a finite number of halfspaces and the closed convex set Ω. Subsequently, Lei and He in [32] proposed a new extra gradient method (NEG) that does not involve adding halfspaces during the projection computation for solving this class of VIs under the same assumptions as in [52]. Then a new extra gradient type projection algorithm (NEGTP) was presented in [46] to solve a class of continuous quasimonotone VIs satisfying $H(x) \ne \mathbf{0}, \forall x \in \Omega$. All the algorithms in [32, 46, 52] have the global sequence convergence, but do not have the estimation on the convergence rate. Ye [51] proved the global convergence of the sequence for the proposed algorithm under the Minty condition. However, in each iteration, the algorithm in [51] requires selecting the half-space that has the largest distance from x_k to some generated half-spaces. As k increases, more information needs to be computed and stored. Approximation-based Regularized Extra-gradient method (ARE), a p^{th}-order ($p \ge 1$) algorithm, was proposed in [27] for solving monotone VIs with a convergence rate of $O(1/N^{\frac{p+1}{2}})$ on the gap function. In [28], it was stated that ARE also can solve the nonmonotone VIs satisfying the Minty condition with the convergence rate of $O(1/\sqrt{N})$ for the residual function and $O(1/N^{\frac{p}{2}})$ for the gap function. However, the algorithms in [27, 28] need the Lipschitz continuity of H and do not have the sequence convergence on the iterates.

1.2. **Anderson acceleration for fixed point problems.** Anderson acceleration was first proposed by Anderson in 1965 in the context of integral equations [4]. This technique aims to improve the convergence rate of fixed point iteration by utilizing the history of search directions. It is not necessary to compute the Jacobian of G, which allows it to perform effectively in various fields, including electronic structure computations [4, 13], machine learning [26], radiation diffusion and nuclear physics [3]. Anderson acceleration is formally described in the following algorithm, commonly referred to as Anderson(m).

Methods	Assumptions	Sequence convergence	Convergence rate (residual function)
EG [19]	pseudomonotone Lipschitz continuous	$\surd$	$O(\frac{1}{\sqrt{N}})$
ARE [28] (the order $p=1$)	Minty condition Lipschitz continuous	$\times$	$O(\frac{1}{\sqrt{N}})$
DPA [52]	Minty condition continuous	$\surd$	$\times$
NEG [32]	Minty condition continuous	$\surd$	$\times$
NEGTP [46]	Minty condition quasimonotone continuous $H(x)\neq 0, \forall x\in\Omega$	$\surd$	$\times$
EG-Anderson(1) **[This paper]**	pseudomonotone continuous	$\surd$	$O(\frac{1}{\sqrt{N}})$ (locally Lipschitz continuous)

TABLE 1. Summary of results on algorithms with global convergence to non-monotone VIs

Algorithm: Anderson(m)

Choose $x_0\in\mathbb{R}^n$ and a positive integer m. Set $x_1=G(x_0)$ and $F_0=G(x_0)-x_0$.
for $k=1,2,\ldots,$ **do**
set $F_k=G(x_k)-x_k$;
choose $m_k=\min\{m,k\}$;
solve

$$\min\left\|\sum_{j=0}^{m_k}\theta_j F_{k-m_k+j}\right\| \quad \text{s.t.} \quad \sum_{j=0}^{m_k}\theta_j=1 \tag{1.5}$$

to find a solution $\{\theta_j^k : j=0,\ldots,m_k\}$, and set

$$x_{k+1}=\sum_{j=0}^{m_k}\theta_j^k G(x_{k-m_k+j}).$$

end for

Even after a long period of use and attention, the first mathematical convergence result for Anderson acceleration had not been given until 2015 by Toth and Kelley [44]. They showed that when G is Lipschitz continuously differentiable and contractive, Anderson(m) has local r-linear convergence with r-factor $\hat{c}\in(c,1)$, and Anderson(1) has q-linear convergence with q-factor c, where c is the contraction coefficient of the fixed point mapping. In 2019, Chen and Kelley [13] weakened the condition of G, proving that this conclusion can be obtained as long as G is a continuously differentiable operator. Additionally, Bian, Chen and Kelley [8] demonstrated the q-linear convergence of Anderson(1) for general non-smooth fixed point problems in a Hilbert space, and r-linear convergence of Anderson(m) for a special nonsmooth operator. Then, Bian and Chen [7] proved that Anderson(1) is

q-linear convergent for the composite max fixed point problem with a smaller q-factor than the existing q-factors. Zhang et al. [53] introduced a variant of Anderson acceleration that guaranteed global convergence for nonsmooth fixed point problems, but did not provide a convergence rate. Ouyang et al. [37] established a globalization strategy for Anderson acceleration incorporating a nonmonotone trust-region framework. They demonstrated that the algorithm has global convergence for a class of nonexpansive mappings and showed a local r-linear convergence for contractive mappings. Moreover, the local properties of Anderson acceleration with restarting were investigated in [36] in terms of function values when applied to a basic gradient scheme.

In this paper, we will utilize the Anderson method to accelerate the convergence of the EG method while ensuring the global convergence of the sequence. The contributions of this paper include the following two aspects.

(1) We propose a new algorithm to solve pseudomonotone VI(Ω, H) by combining Anderson(1) with the EG method. We prove that the sequence generated by the proposed algorithm converges to a solution of VI(Ω, H) without assuming the Lipschitz continuity and contractive condition of G.
(2) Under the condition that H is locally Lipschitz continuous, the convergence rate of the proposed algorithm on the residual function is not worse than the EG method. This condition is weaker than the requirement of the EG method that H is Lipschitz continuous. Moreover, in numerical experiments, the proposed algorithm has been found to outperform Anderson(1) and the EG methods.

This paper is organized as follows. In Section 2, we briefly review some related concepts and recall some preliminary results used in this paper. In Section 3, we use the idea of Anderson(1) to develop an extra gradient Anderson-accelerated algorithm to solve the continuous VI(Ω, H). Furthermore, the sequence convergence of the algorithm is analyzed and the convergence rate is provided. Finally, we use five numerical experiments to illustrate the good performance of the proposed algorithm in Section 4.

2. Preliminaries

Let $\|\cdot\|$ denote Euclidean norm in $\mathbb{R}^n$. For a matrix $A \in \mathbb{R}^{m\times n}$, $\|A\|$ represents its 2-norm. We begin by introducing two operators, which play a crucial role in the proposed algorithm. Additionally, we present some definitions and lemmas that will be used for the convergence analysis of the proposed algorithm.

Define the following operators

$$G_t(x) := P_\Omega(x - tH(x)) \quad \text{and} \quad \tilde{G}_t(x) := P_\Omega(x - tH(G_t(x))),$$

where $t > 0$. Let

$$F_t(x) := G_t(x) - x \quad \text{and} \quad \tilde{F}_t(x) := \tilde{G}_t(x) - x.$$

Definition 2.1. [30] The mapping $H : \mathbb{R}^n \to \mathbb{R}^n$ is said to be pseudomonotone on Ω, if for any $x, y \in \Omega$ it holds

$$\langle H(x), y - x\rangle \geq 0 \Rightarrow \langle H(y), y - x\rangle \geq 0.$$

Lemma 2.2. [21] *For any $x \in \mathbb{R}^n$, the following statements hold.*

(i) $\|P_\Omega(x) - P_\Omega(y)\|^2 \leq \langle P_\Omega(x) - P_\Omega(y), x - y\rangle,\ \forall y \in \mathbb{R}^n$;

(ii) $\langle x - P_\Omega(x), y - P_\Omega(x)\rangle \leq 0,\ \forall y \in \Omega$.

Lemma 2.3. [17] *For any $x \in \mathbb{R}^n$ and $t_1 \geq t_2 > 0$, the following inequalities hold:*

$$\frac{\|x - P_\Omega(x - t_1 H(x))\|}{t_1} \leq \frac{\|x - P_\Omega(x - t_2 H(x))\|}{t_2},$$
$$\|x - P_\Omega(x - t_2 H(x))\| \leq \|x - P_\Omega(x - t_1 H(x))\|.$$

Lemma 2.4. [38] [Opial's Lemma] *Let S be a nonempty subset of $\mathbb{R}^n$, and $\{x_k\}$ a sequence of elements in $\mathbb{R}^n$. Assume that*

(i) *every sequential cluster point of $\{x_k\}$, as $k \to \infty$, belongs to S;*

(ii) *for every $z \in S$, $\lim_{k\to\infty} \|x_k - z\|$ exists.*

Then the sequence $\{x_k\}$ converges as $k \to \infty$ to a point in S.

Lemma 2.5. [19, Proposition 1.5.8, Exercise 1.8.29] *$x^* \in \mathrm{SOL}(\Omega, H)$ if and only if it is a fixed point of G_t, and if and only if it is a fixed point of $\tilde{G}_t$, where t can be any positive number.*

Lemma 2.6. [23] *Suppose that the mapping $H : \mathbb{R}^n \to \mathbb{R}^n$ is continuous. Then, for all bounded sequences $\{x_k\}$, $\{y_k\} \subseteq \mathbb{R}^n$ satisfying $\lim_{k\to\infty} \|x_k - y_k\| = 0$, it holds that $\lim_{k\to\infty} \|H(x_k) - H(y_k)\| = 0$.*

Lemma 2.7. [5] *Let $\{a_k\}$ and $\{\varepsilon_k\}$ be real sequences. Assume that $\{a_k\}$ is bounded from below, $\sum_{k=1}^{\infty} \varepsilon_k < \infty$ and*

$$a_{k+1} - a_k \leq \varepsilon_k$$

for every k. Then $\lim_{k\to\infty} a_k$ exists.

3. Proposed algorithm and its convergence analysis

In this section, based on Anderson acceleration and the EG methods, we propose the EG-Anderson(1) algorithm for solving $\mathrm{VI}(\Omega, H)$. In addition, we give the convergence analysis of this algorithm.

3.1. Proposed algorithm.

The proposed algorithm is presented in Algorithm 1, where the line search framework in [10] is used in Step 2.

From the notations and definitions for $y_{k+0.5}$ and y_{k+1} in the EG-Anderson(1) algorithm, we find that

$$y_{k+0.5} = G_{t_k}(x_k) \quad \text{and} \quad y_{k+1} = \tilde{G}_{t_k}(x_k);$$
$$y_{k+0.5} - x_k = F_{t_k}(x_k) \quad \text{and} \quad y_{k+1} - x_k = \tilde{F}_{t_k}(x_k).$$

3.2. Convergence analysis.

In this subsection, we will analyze the convergence properties of the EG-Anderson(1) algorithm, including the global convergence of the sequence and the convergence rate evaluated by the residual function. In order to categorize the iteration counts, we divide them into two subsets:

$$K_{AA} = \{k_0, k_1, \cdots\} \quad \text{and} \quad K_{EG} = \{l_0, l_1, \cdots\},$$

where K_{AA} consists of iterations setting by (3.3) and K_{EG} includes the remaining iterations setting by (3.4).

If the EG-Anderson(1) algorithm is terminated in finite times, then the final output point is a solution of $\mathrm{VI}(\Omega, H)$. Therefore, in the following analysis we assume that the EG-Anderson(1) algorithm loops infinitely.

Algorithm 1: EG-Anderson(1)

Initialization: Choose $x_0 \in \Omega$, $\omega \geq 0, \gamma > 0$, $\tau > \frac{1}{2}$, $\rho, \mu \in (0,1)$ and $\sigma_0 = 1$. Give a sufficiently large $M > 0$.

for $k = 0, 1, 2, \ldots,$ **do**

Step 1: Compute $F_\gamma(x_k) = P_\Omega(x_k - \gamma H(x_k)) - x_k$.

If $F_\gamma(x_k) = \mathbf{0}$, then stop.

Otherwise, let $t_k = \gamma$ and go to **Step 2**.

Step 2: Compute $y_{k+0.5} = P_\Omega(x_k - t_k H(x_k))$ and $y_{k+1} = P_\Omega(x_k - t_k H(y_{k+0.5}))$.

If

$$t_k \langle H(y_{k+0.5}) - H(x_k), y_{k+0.5} - y_{k+1} \rangle \leq \frac{\mu}{2} \left(\|x_k - y_{k+0.5}\|^2 + \|y_{k+0.5} - y_{k+1}\|^2 \right), \tag{3.1}$$

go to **Step 3**.

Otherwise, set $t_k = \rho t_k$ and repeat **Step 2**.

Step 3: Compute $F_{t_k}(x_k) = y_{k+0.5} - x_k$ and $\tilde{F}_{t_k}(x_k) = y_{k+1} - x_k$.

If $\|\tilde{F}_{t_k}(x_k)\| < \min\{\|F_{t_k}(x_k)\|, \omega \sigma_k^{-\tau}\}$, set

$$\alpha_k = \frac{\langle \tilde{F}_{t_k}(x_k), \tilde{F}_{t_k}(x_k) - F_{t_k}(x_k) \rangle}{\|\tilde{F}_{t_k}(x_k) - F_{t_k}(x_k)\|^2}, \tag{3.2}$$

Otherwise, set $\alpha_k = M + 1$.

Step 4: **If** $|\alpha_k| \leq M$, set

$$x_{k+1} = \alpha_k x_k + (1 - \alpha_k) y_{k+1}, \quad \sigma_{k+1} = \sigma_k + 1. \tag{3.3}$$

Otherwise, set

$$x_{k+1} = y_{k+1}, \quad \sigma_{k+1} = \sigma_k. \tag{3.4}$$

end for

Remark 3.1. For $k_i \in K_{AA}$, note that $\|\tilde{F}_{t_{k_i}}(x_{k_i}) - F_{t_{k_i}}(x_{k_i})\| \neq 0$ due to $\|\tilde{F}_{t_{k_i}}(x_{k_i})\| < \|F_{t_{k_i}}(x_{k_i})\|$, thus α_{k_i} is well-defined. Moreover, α_{k_i} is the optimal solution of

$$\min \left\| \alpha F_{t_{k_i}}(x_{k_i}) + (1 - \alpha) \tilde{F}_{t_{k_i}}(x_{k_i}) \right\|.$$

We start the convergence analysis of the EG-Anderson(1) algorithm by proving that (3.1) terminates after a finite number of loops.

Lemma 3.2. *The EG-Anderson(1) algorithm is well-defined.*

Proof We will show that the EG-Anderson(1) algorithm is well-defined by proving that for every k there exists t_k satisfying (3.1) when $x_k \notin \mathrm{SOL}(\Omega, H)$.

From the updated form of t_k in the EG-Anderson(1) algorithm, it can be reformulated as $t_k = \gamma \rho^{m_k}$, where m_k is the smallest nonnegative integer m satisfying

$$\gamma \rho^m \left\langle H\left(y_{k+0.5}^{(m)}\right) - H\left(x_k\right), y_{k+0.5}^{(m)} - y_{k+1}^{(m)} \right\rangle \leq \frac{\mu}{2} \left(\left\| x_k - y_{k+0.5}^{(m)} \right\|^2 + \left\| y_{k+0.5}^{(m)} - y_{k+1}^{(m)} \right\|^2 \right), \tag{3.5}$$

where $y_{k+0.5}^{(m)} := P_\Omega\left(x_k - \gamma \rho^m H\left(x_k\right)\right)$ and $y_{k+1}^{(m)} := P_\Omega\left(x_k - \gamma \rho^m H\left(y_{k+0.5}^{(m)}\right)\right)$.

If there exists a nonnegative integer $\bar{m}$ such that $y_{k+0.5}^{(\bar{m})} = y_{k+1}^{(\bar{m})}$, then there must exist an integer m_k between 0 and $\bar{m}$ such that (3.5) holds. We consider the situation $y_{k+0.5}^{(m)} \neq y_{k+1}^{(m)}$ for any nonnegative integer m and assume the contrary that for all m we have

(3.6)
$$\gamma\rho^m \left\langle H\left(y_{k+0.5}^{(m)}\right) - H(x_k), y_{k+0.5}^{(m)} - y_{k+1}^{(m)} \right\rangle > \frac{\mu}{2}\left(\left\|x_k - y_{k+0.5}^{(m)}\right\|^2 + \left\|y_{k+0.5}^{(m)} - y_{k+1}^{(m)}\right\|^2\right).$$

On one hand, by Cauchy–Schwarz inequality, we obtain

(3.7)
$$\gamma\rho^m \left\langle H\left(y_{k+0.5}^{(m)}\right) - H\left(x_k\right), y_{k+0.5}^{(m)} - y_{k+1}^{(m)} \right\rangle \leq \gamma\rho^m \left\|H\left(y_{k+0.5}^{(m)}\right) - H(x_k)\right\| \left\|y_{k+0.5}^{(m)} - y_{k+1}^{(m)}\right\|.$$

On the other hand, we also find

$$\left\|x_k - y_{k+0.5}^{(m)}\right\|^2 + \left\|y_{k+0.5}^{(m)} - y_{k+1}^{(m)}\right\|^2 \geq 2\left\|x_k - y_{k+0.5}^{(m)}\right\| \left\|y_{k+0.5}^{(m)} - y_{k+1}^{(m)}\right\|. \tag{3.8}$$

Combining (3.6) with (3.7) and (3.8), we deduce that

$$\frac{\left\|x_k - y_{k+0.5}^{(m)}\right\|}{\gamma\rho^m} \leq \frac{1}{\mu}\left\|H\left(y_{k+0.5}^{(m)}\right) - H(x_k)\right\|. \tag{3.9}$$

Since $x_k \notin \mathrm{SOL}(\Omega, H)$, we discuss the following two cases.

(i) If $x_k \in \Omega$, from the definition of $y_{k+0.5}^{(m)}$ and the continuity of P_Ω, we have

$$\lim_{m\to\infty}\left\|x_k - y_{k+0.5}^{(m)}\right\| = 0.$$

In view of the continuity of H, we get $\lim_{m\to\infty}\left\|H(x_k) - H\left(y_{k+0.5}^{(m)}\right)\right\| = 0$. This together with (3.9) yields

$$\lim_{m\to\infty}\frac{\left\|x_k - y_{k+0.5}^{(m)}\right\|}{\gamma\rho^m} = 0. \tag{3.10}$$

By the definition of $y_{k+0.5}^{(m)}$ and using Lemma 2.2-(ii), we get

$$\left\langle y_{k+0.5}^{(m)} - x_k + \gamma\rho^m H(x_k), x - y_{k+0.5}^{(m)} \right\rangle \geq 0, \forall x \in \Omega,$$

which implies

$$\left\langle \frac{y_{k+0.5}^{(m)} - x_k}{\gamma\rho^m} + H(x_k), x - y_{k+0.5}^{(m)} \right\rangle \geq 0, \forall x \in \Omega. \tag{3.11}$$

Taking the limit $m \to \infty$ in (3.11) and using (3.10) and $\lim_{m\to\infty} y_{k+0.5}^{(m)} = x_k$, we obtain $\langle H(x_k), x - x_k\rangle \geq 0, \forall x \in \Omega$. It can be deduced that $x_k \in \mathrm{SOL}(\Omega, H)$ and this leads to a contraction.

(ii) If $x_k \notin \Omega$, we can conclude that

$$\lim_{m\to\infty}\left\|x_k - y_{k+0.5}^{(m)}\right\| = \|x_k - P_\Omega(x_k)\| > 0, \tag{3.12}$$

and

$$\lim_{m\to\infty}\gamma\rho^m\left\|H\left(y_{k+0.5}^{(m)}\right) - H(x_k)\right\| = 0. \tag{3.13}$$

Rearranging the terms in (3.9), we find

$$\left\|x_k - y_{k+0.5}^{(m)}\right\| \leq \frac{1}{\mu}\gamma\rho^m\left(\left\|H\left(y_{k+0.5}^{(m)}\right) - H(x_k)\right\|\right).$$

Taking the limit $m \to \infty$ in the above inequality, we can find a contradiction with (3.12) and (3.13). Hence, the proof is fully established. □

The next two lemmas are instrumental in establishing the key findings of this section.

Lemma 3.3. *Let $\{x_k\}$ be the sequence generated by the EG-Anderson(1) algorithm. Then we have*

$$\left(1-\sqrt{\frac{\mu}{2-\mu}}\right)\|F_{t_k}(x_k)\| \le \|\tilde{F}_{t_k}(x_k)\| \le \left(1+\sqrt{\frac{\mu}{2-\mu}}\right)\|F_{t_k}(x_k)\|.$$

Proof For any k, by the condition of t_k in (3.1) and Lemma 2.2-(i), we obtain

$$\begin{aligned}\|y_{k+0.5}-y_{k+1}\|^2 &= \|P_\Omega(x_k - t_k H(x_k)) - P_\Omega(x_k - t_k H(y_{k+0.5}))\|^2\\ &\le \langle y_{k+0.5}-y_{k+1}, t_k\left(H(y_{k+0.5})-H(x_k)\right)\rangle\\ &\le \frac{\mu}{2}\|x_k-y_{k+0.5}\|^2 + \frac{\mu}{2}\|y_{k+0.5}-y_{k+1}\|^2,\end{aligned}$$

which implies

$$\|y_{k+1}-y_{k+0.5}\|^2 \le \frac{\mu}{2-\mu}\|x_k-y_{k+0.5}\|^2. \tag{3.14}$$

Since $\mu \in (0,1)$, then $\frac{\mu}{2-\mu} \in (0,1)$. From the triangle inequality and (3.14), we have

$$\begin{aligned}\|\tilde{F}_{t_k}(x_k)\| = \|y_{k+1}-x_k\| &\ge \|y_{k+0.5}-x_k\| - \|y_{k+1}-y_{k+0.5}\|\\ &\ge \|y_{k+0.5}-x_k\| - \sqrt{\frac{\mu}{2-\mu}}\|y_{k+0.5}-x_k\|\\ &= \left(1-\sqrt{\frac{\mu}{2-\mu}}\right)\|F_{t_k}(x_k)\|\end{aligned}$$

and

$$\begin{aligned}\|\tilde{F}_{t_k}(x_k)\| = \|y_{k+1}-x_k\| &\le \|y_{k+0.5}-x_k\| + \|y_{k+1}-y_{k+0.5}\|\\ &\le \left(1+\sqrt{\frac{\mu}{2-\mu}}\right)\|F_{t_k}(x_k)\|.\end{aligned}$$

The proof is completed. □

Lemma 3.4. *Let $\{x_k\}$ be the sequence generated by the EG-Anderson(1) algorithm and $x^* \in \mathrm{SOL}(\Omega, H)$. For every k, it holds that*

$$\|y_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - (1-\mu)\|y_{k+0.5}-x_k\|^2 - (1-\mu)\|y_{k+1}-y_{k+0.5}\|^2.$$

Proof In view of the pseudomonotonicity of H and $x^* \in \mathrm{SOL}(\Omega, H)$, we deduce that $\langle t_k H(y_{k+0.5}), x^* - y_{k+0.5}\rangle \le 0$, which gives

$$\langle t_k H(y_{k+0.5}), x^*-y_{k+1}\rangle \le \langle t_k H(y_{k+0.5}), y_{k+0.5}-y_{k+1}\rangle. \tag{3.15}$$

By the definition of $y_{k+0.5}$, Lemma 2.2-(ii) and (3.1), we obtain

$$\begin{aligned}&\langle x_k - t_k H(y_{k+0.5}) - y_{k+0.5}, y_{k+1}-y_{k+0.5}\rangle\\ &= \langle x_k - t_k H(x_k) - y_{k+0.5}, y_{k+1}-y_{k+0.5}\rangle + t_k\langle H(x_k)-H(y_{k+0.5}), y_{k+1}-y_{k+0.5}\rangle\\ &\overset{(3.1)}{\le} \langle x_k - t_k H(x_k) - P_\Omega(x_k - t_k H(x_k)), y_{k+1}-P_\Omega(x_k-t_kH(x_k))\rangle\\ &\quad + \frac{\mu}{2}\|x_k-y_{k+0.5}\|^2 + \frac{\mu}{2}\|y_{k+0.5}-y_{k+1}\|^2\\ &\le \frac{\mu}{2}\|x_k-y_{k+0.5}\|^2 + \frac{\mu}{2}\|y_{k+0.5}-y_{k+1}\|^2.\end{aligned} \tag{3.16}$$

Based on the definition of y_{k+1}, Lemma 2.2-(ii), (3.15) and (3.16), and using the same idea as in [10, Lemma 3.3], we conclude that

$$\begin{aligned}\|y_{k+1}-x^*\|^2 &\le \|x_k-x^*\|^2-\|x_k-y_{k+1}\|^2+2\langle t_kH(y_{k+0.5}),x^*-y_{k+1}\rangle\\ &\overset{(3.15)}{\le} \|x_k-x^*\|^2-\|x_k-y_{k+1}\|^2+2\langle t_kH(y_{k+0.5}),y_{k+0.5}-y_{k+1}\rangle\\ &= \|x_k-x^*\|^2-\|x_k-y_{k+0.5}\|^2-\|y_{k+0.5}-y_{k+1}\|^2\\ &\quad +2\langle x_k-t_kH(y_{k+0.5})-y_{k+0.5},y_{k+1}-y_{k+0.5}\rangle\\ &\overset{(3.16)}{\le} \|x_k-x^*\|^2-(1-\mu)\|x_k-y_{k+0.5}\|^2-(1-\mu)\|y_{k+0.5}-y_{k+1}\|^2.\end{aligned}$$

□

Now, utilizing the Opial's Lemma, we can state and prove our main convergence result in what follows.

Theorem 3.5. *Let $\{x_k\}$ be the sequence generated by the EG-Anderson(1) algorithm. Then the sequence $\{x_k\}$ converges to a solution of* VI(Ω,H).

Proof We will prove this theorem from three steps.

Step 1: $\{x_k\}$ is bounded.

Let x^* be a solution of problem VI(Ω,H). If $k_i\in K_{AA}$, let $\beta_{k_i}:=1-\alpha_{k_i}$ and by the definition of x_{k_i+1} in (3.3), we know that

$$\begin{aligned}\|x_{k_i+1}-x^*\|^2 &= \|\alpha_{k_i}x_{k_i}+\beta_{k_i}y_{k_i+1}-x^*\|^2\\ &= \alpha_{k_i}^2\|x_{k_i}-x^*\|^2+\beta_{k_i}^2\|y_{k_i+1}-x^*\|^2+2\alpha_{k_i}\beta_{k_i}\langle x_{k_i}-x^*,y_{k_i+1}-x^*\rangle\\ &= \alpha_{k_i}^2\|x_{k_i}-x^*\|^2+\beta_{k_i}^2\|y_{k_i+1}-x^*\|^2\\ &\quad +2\alpha_{k_i}\beta_{k_i}\left(\frac{1}{2}\|x_{k_i}-x^*\|^2+\frac{1}{2}\|y_{k_i+1}-x^*\|^2-\frac{1}{2}\|x_{k_i}-y_{k_i+1}\|^2\right)\\ &= \alpha_{k_i}\|x_{k_i}-x^*\|^2+\beta_{k_i}\|y_{k_i+1}-x^*\|^2-\alpha_{k_i}\beta_{k_i}\|x_{k_i}-y_{k_i+1}\|^2.\end{aligned}\tag{3.17}$$

From Lemma 3.4, we can obtain

$$\|y_{k_i+1}-x^*\|^2\le\|x_{k_i}-x^*\|^2-(1-\mu)\|y_{k_i+0.5}-x_{k_i}\|^2-(1-\mu)\|y_{k_i+1}-y_{k_i+0.5}\|^2.\tag{3.18}$$

Since $\|\tilde{F}_{t_{k_i}}(x_{k_i})\|<\|F_{t_{k_i}}(x_{k_i})\|$ when $k_i\in K_{AA}$, then

$$\langle F_{t_{k_i}}(x_{k_i}),\tilde{F}_{t_{k_i}}(x_{k_i})\rangle<\|F_{t_{k_i}}(x_{k_i})\|^2,$$

which implies

$$\langle\tilde{F}_{t_{k_i}}(x_{k_i}),\tilde{F}_{t_{k_i}}(x_{k_i})-F_{t_{k_i}}(x_{k_i})\rangle<\|\tilde{F}_{t_{k_i}}(x_{k_i})-F_{t_{k_i}}(x_{k_i})\|^2.$$

By (3.2), we have $\alpha_{k_i}<1$. Thus $\beta_{k_i}=1-\alpha_{k_i}>0$.

Introducing (3.18) into (3.17), we deduce that

$$\begin{aligned}\|x_{k_i+1}-x^*\|^2\le\|x_{k_i}-x^*\|^2&-(1-\mu)\beta_{k_i}\|y_{k_i+0.5}-x_{k_i}\|^2\\ &-(1-\mu)\beta_{k_i}\|y_{k_i+0.5}-y_{k_i+1}\|^2-\alpha_{k_i}\beta_{k_i}\|x_{k_i}-y_{k_i+1}\|^2.\end{aligned}\tag{3.19}$$

In view of $|\alpha_{k_i}|\le M$ and $\beta_{k_i}=1-\alpha_{k_i}$, we get $|\beta_{k_i}|\le M+1$. This together with (3.19), $\mu\in(0,1)$, $\beta_{k_i}>0$ and $\|x_{k_i}-y_{k_i+1}\|=\|\tilde{F}_{t_{k_i}}(x_{k_i})\|\le\omega\sigma_{k_i}^{-\tau}=\omega(\sigma_0+i)^{-\tau}=\omega(1+i)^{-\tau}$, we deduce that

$$\begin{aligned}\|x_{k_i+1}-x^*\|^2&\le\|x_{k_i}-x^*\|^2+|\alpha_{k_i}||\beta_{k_i}|\|\tilde{F}_{t_{k_i}}(x_{k_i})\|^2\\ &\le\|x_{k_i}-x^*\|^2+\varepsilon_{k_i},\end{aligned}\tag{3.20}$$

where $\varepsilon_{k_i} := M(M+1)\omega^2(1+i)^{-2\tau}$.

If $l_j \in K_{EG}$, by Lemma 3.4 and $\mu \in (0,1)$, we conclude that

$$\begin{aligned}\|x_{l_j+1}-x^*\|^2 &= \|y_{l_j+1}-x^*\|^2\\ &\le \|x_{l_j}-x^*\|^2-(1-\mu)\|y_{l_j+0.5}-x_{l_j}\|^2-(1-\mu)\|y_{l_j+1}-y_{l_j+0.5}\|^2\\ &\le \|x_{l_j}-x^*\|^2.\end{aligned} \tag{3.21}$$

By defining $\varepsilon_{l_j}=0$ and combining (3.20) and (3.21), for every k, we find

$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2+\varepsilon_k, \tag{3.22}$$

where $\sum_{k=0}^{\infty}\varepsilon_k=\sum_{i=0}^{\infty}\varepsilon_{k_i}<\infty$ because of $\tau>\frac{1}{2}$. From (3.22), we further have

$$\|x_k-x^*\|^2 \le \|x_0-x^*\|^2+\sum_{k=0}^{\infty}\varepsilon_k<\infty.$$

Hence $\{x_k\}$ is bounded.

Step 2: any cluster point of $\{x_k\}$ belongs to $\mathrm{SOL}(\Omega,H)$, i.e. a solution of $\mathrm{VI}(\Omega,H)$.

By the boundedness of $\{x_k\}$, it has at least one cluster point, denoted by $\bar{x}$ with convergence subsequence $\{x_{p_j}\}$ satisfying $\lim_{j\to\infty}x_{p_j}=\bar{x}$. Next, we will prove $\bar{x}\in\mathrm{SOL}(\Omega,H)$ from the following two steps.

Step 2.1: $\lim_{k\to\infty}\|F_{t_k}(x_k)\|=0$.

We know that $K_{AA}\cup K_{EG}$ is infinite. The following proof is divided into two cases for discussion.

(i) We first consider the case that both K_{AA} and K_{EG} are infinite. Rearranging the terms in (3.21) and using $\mu\in(0,1)$, we infer that

$$(1-\mu)\|F_{l_j}(x_{l_j})\|^2=(1-\mu)\|y_{l_j+0.5}-x_{l_j}\|^2\le\|x_{l_j}-x^*\|^2-\|x_{l_j+1}-x^*\|^2. \tag{3.23}$$

Combining (3.20) and (3.23), we deduce that

$$\sum_{j=0}^{\infty}\|F_{l_j}(x_{l_j})\|^2\le\frac{1}{1-\mu}\|x_0-x^*\|^2+\frac{1}{1-\mu}\sum_{i=0}^{\infty}\varepsilon_{k_i}<\infty. \tag{3.24}$$

Moreover, we know $\|\tilde{F}_{t_{k_i}}(x_{k_i})\|\le\omega(1+i)^{-\tau}$ when $k_i\in K_{AA}$. Together with Lemma 3.3, we have

$$\|F_{t_{k_i}}(x_{k_i})\|^2\le\left(1-\sqrt{\frac{\mu}{2-\mu}}\right)^{-2}\|\tilde{F}_{t_{k_i}}(x_{k_i})\|^2\le\left(1-\sqrt{\frac{\mu}{2-\mu}}\right)^{-2}\omega^2(1+i)^{-2\tau}.$$

Thus

$$\sum_{i=0}^{\infty}\|F_{t_{k_i}}(x_{k_i})\|^2\le\left(1-\sqrt{\frac{\mu}{2-\mu}}\right)^{-2}\omega^2\sum_{i=0}^{\infty}(1+i)^{-2\tau}<\infty. \tag{3.25}$$

Since $\tau>\frac{1}{2}$, there exists a $C_0>0$ such that $\sum_{i=0}^{\infty}(1+i)^{-2\tau}\le C_0$. Combining (3.24) and (3.25), we deduce that

$$\sum_{k=0}^{\infty}\|F_{t_k}(x_k)\|^2=\sum_{i=0}^{\infty}\|F_{t_{k_i}}(x_{k_i})\|^2+\sum_{j=0}^{\infty}\|F_{t_{l_j}}(x_{l_j})\|^2\le C, \tag{3.26}$$

where $C:=\left(\left(1-\sqrt{\frac{\mu}{2-\mu}}\right)^{-2}+\frac{1}{1-\mu}M(M+1)\right)\omega^2C_0+\frac{1}{1-\mu}\|x_0-x^*\|^2$. Hence we conclude that $\lim_{k\to\infty}\|F_{t_k}(x_k)\|=0$.

(ii) We then consider the case that either K_{AA} or K_{EG} is finite. In this situation, the proof becomes simpler, as we only need to use either (3.24) or (3.25) to obtain $\sum_{k=0}^{\infty}\|F_{t_k}(x_k)\|^2<\infty$. Therefore, we have $\lim_{k\to\infty}\|F_{t_k}(x_k)\|=0$.

Step 2.2: $\lim_{k\to\infty}\|F_\gamma(x_k)\|=0$.

Let $\tilde{y}_{k+0.5}:=P_\Omega(x_k-t_k\rho^{-1}H(x_k))$. Applying Lemma 2.3 and $t_k\rho^{-1}>t_k$, we get

$$\|x_k-\tilde{y}_{k+0.5}\|\le\rho^{-1}\|x_k-y_{k+0.5}\|=\rho^{-1}\|F_{t_k}(x_k)\|, \tag{3.27}$$

which implies $\lim_{k\to\infty}\|x_k-\tilde{y}_{k+0.5}\|=0$.

Below, we will estimate $\frac{1}{t_k}\|x_k-\tilde{y}_{k+0.5}\|$ by dividing it into two cases based on the value of t_k.

(i) If $t_k=\gamma$, we have

$$\frac{1}{t_k}\|x_k-\tilde{y}_{k+0.5}\|=\frac{1}{\gamma}\|x_k-\tilde{y}_{k+0.5}\|. \tag{3.28}$$

(ii) If $t_k<\gamma$, let $\tilde{y}_{k+1}:=P_\Omega(x_k-t_k\rho^{-1}H(\tilde{y}_{k+0.5}))$. From the condition of t_k in (3.1), we know that $t_k\rho^{-1}$ satisfies

$$t_k\rho^{-1}\langle H(\tilde{y}_{k+0.5})-H(x_k),\tilde{y}_{k+0.5}-\tilde{y}_{k+1}\rangle>\frac{\mu}{2}\left(\|x_k-\tilde{y}_{k+0.5}\|^2+\|\tilde{y}_{k+0.5}-\tilde{y}_{k+1}\|^2\right), \tag{3.29}$$

which implies $\tilde{y}_{k+0.5}\neq\tilde{y}_{k+1}$. Then for (3.29), based on a similar estimation as for (3.6), we can conclude that

$$\frac{1}{t_k}\|x_k-\tilde{y}_{k+0.5}\|<\mu^{-1}\rho^{-1}\|H(x_k)-H(\tilde{y}_{k+0.5})\|. \tag{3.30}$$

Combining (3.28) and (3.30), we can obtain

$$\frac{1}{t_k}\|x_k-\tilde{y}_{k+0.5}\|\le\max\left\{\frac{1}{\gamma}\|x_k-\tilde{y}_{k+0.5}\|,\mu^{-1}\rho^{-1}\|H(x_k)-H(\tilde{y}_{k+0.5})\|\right\}. \tag{3.31}$$

Since $\{x_k\}$ is bounded and $\lim_{k\to\infty}\|x_k-\tilde{y}_{k+0.5}\|=0$, we find $\{\tilde{y}_{k+0.5}\}$ is bounded. Combining the continuity of H and Lemma 2.6, we conclude that

$$\lim_{k\to\infty}\|H(x_k)-H(\tilde{y}_{k+0.5})\|=0.$$

This together with $\lim_{k\to\infty}\|x_k-\tilde{y}_{k+0.5}\|=0$ and (3.31) yields

$$\lim_{k\to\infty}\frac{1}{t_k}\|x_k-\tilde{y}_{k+0.5}\|=0.$$

Again by Lemma 2.3 and $t_k\rho^{-1}>t_k$, we get

$$\|x_k-\tilde{y}_{k+0.5}\|\ge\|x_k-y_{k+0.5}\|. \tag{3.32}$$

Then, we obtain

$$\lim_{k\to\infty}\frac{1}{t_k}\|x_k-y_{k+0.5}\|\le\lim_{k\to\infty}\frac{1}{t_k}\|x_k-\tilde{y}_{k+0.5}\|=0,$$

which means $\lim_{k\to\infty}\frac{1}{t_k}\|F_{t_k}(x_k)\|=\lim_{k\to\infty}\frac{1}{t_k}\|x_k-y_{k+0.5}\|=0$. Hence, we conclude that $\lim_{k\to\infty}\frac{1}{t_k}\|F_{t_k}(x_k)\|=0$.

Recalling Lemma 2.3 and $t_k\le\gamma$, we have $\frac{1}{t_k}\|F_{t_k}(x_k)\|\ge\frac{1}{\gamma}\|F_\gamma(x_k)\|$ and $\|F_{t_k}(x_k)\|\le\|F_\gamma(x_k)\|$. Then we conclude that

$$\|F_{t_k}(x_k)\|\le\|F_\gamma(x_k)\|\le\frac{\gamma}{t_k}\|F_{t_k}(x_k)\|, \tag{3.33}$$

which gives

$$\lim_{k\to\infty}\|F_\gamma(x_k)\| = 0.$$

Together with $\lim_{j\to\infty} x_{p_j} = \bar{x}$ and Lemma 2.5, we deduce that $\bar{x} \in \mathrm{SOL}(\Omega, H)$.

Step 3: $\{x_k\}$ is convergent to a solution of $\mathrm{VI}(\Omega, H)$.

Since $\|x_k - x^*\|^2$ is nonnegative, $\sum_{k=0}^{\infty}\varepsilon_k < \infty$ and (3.22) holds, by applying Lemma 2.7 with $a_k = \|x_k - x^*\|^2$, $\lim_{k\to\infty}\|x_k - x^*\|^2$ exists for any $x^* \in \mathrm{SOL}(\Omega, H)$. Together this with Step 2, the proof is completed by Lemma 2.4. □

If H is also locally Lipschitz continuous at any solution of $\mathrm{VI}(\Omega, H)$, we derive the following conclusion about the convergence rate on the residual function.

Theorem 3.6. *(Best-iterate convergence rate) Suppose that H is locally Lipschitz continuous at any solution of $\mathrm{VI}(\Omega, H)$. Let $\{x_k\}$ be the sequence generated by the EG-Anderson(1) algorithm. Then there exists a positive integer N_0 such that*

$$\min_{N_0+1\le k\le N}\|F_\gamma(x_k)\|^2 = O\left(\frac{1}{N}\right).$$

Proof By Theorem 3.5, the sequence $\{x_k\}$ converges to a solution x^* of $\mathrm{VI}(\Omega, H)$ and $\lim_{k\to\infty}\|x_k - \tilde{y}_{k+0.5}\| = 0$. Thus, we know that the sequence $\{\tilde{y}_{k+0.5}\}$ also converges to x^*. From the locally Lipschitz continuity of H, there exist an $r \in (0,1)$, an $L^* > 0$ and a positive integer N_0 such that for $k > N_0$, we have $\|x_k - x^*\| \le r$ and

$$\|H(x_k) - H(\tilde{y}_{k+0.5})\| \le L^*\|x_k - \tilde{y}_{k+0.5}\|.$$

First, we will discuss the relationship between $\|F_\gamma(x_k)\|$ and $\|F_{t_k}(x_k)\|$ in two cases.

(i) For $t_k = \gamma$, we have

$$\|F_\gamma(x_k)\| = \|F_{t_k}(x_k)\|. \tag{3.34}$$

(ii) For $t_k < \gamma$, combining (3.27), (3.30), (3.32) and (3.33) yields that

$$\begin{aligned}
\|F_\gamma(x_k)\| &\overset{(3.33)}{\le} \frac{\gamma}{t_k}\|F_{t_k}(x_k)\| \\
&\overset{(3.32)}{\le} \frac{\gamma}{t_k}\|x_k - \tilde{y}_{k+0.5}\| \\
&\overset{(3.30)}{<} \gamma\rho^{-1}\mu^{-1}\|H(x_k) - H(\tilde{y}_{k+0.5})\| \\
&\le \gamma\rho^{-1}\mu^{-1}L^*\|x_k - \tilde{y}_{k+0.5}\| \\
&\overset{(3.27)}{\le} \gamma\rho^{-2}\mu^{-1}L^*\|F_{t_k}(x_k)\|, \quad \forall k > N_0.
\end{aligned} \tag{3.35}$$

Combining (3.34) and (3.35), we know that for any $k > N_0$, we have

$$\|F_\gamma(x_k)\| \le \max\{\gamma\rho^{-2}\mu^{-1}L^*, 1\}\|F_{t_k}(x_k)\|. \tag{3.36}$$

Next, we will prove that $\sum_{k=N_0+1}^{\infty}\|F_{t_k}(x_k)\|^2$ is finite. We only consider the case that both K_{AA} and K_{EG} are infinite, as the analysis is similar or simpler when either of them is finite. For the aforementioned N_0, there exist i_0 and j_0 such that $\{k_i : i \ge i_0\} \cup \{l_j : j \ge j_0\} = \{k : k \ge N_0 + 1\}$. Then we know

$$\sum_{k=N_0+1}^{\infty}\|F_{t_k}(x_k)\|^2 = \sum_{i=i_0}^{\infty}\|F_{t_{k_i}}(x_{k_i})\|^2 + \sum_{j=j_0}^{\infty}\|F_{t_{l_j}}(x_{l_j})\|^2. \tag{3.37}$$

From (3.20) and (3.23), we get

$$0 \leq \sum_{i=i_0}^{\infty} \left(\|x_{k_i} - x^*\|^2 - \|x_{k_i+1} - x^*\|^2 \right) + \sum_{i=i_0}^{\infty} \varepsilon_{k_i} \tag{3.38}$$

and

$$(1-\mu) \sum_{j=j_0}^{\infty} \|F_{t_{l_j}}(x_{l_j})\|^2 \leq \sum_{j=j_0}^{\infty} \left(\|x_{l_j} - x^*\|^2 - \|x_{l_j+1} - x^*\|^2 \right), \tag{3.39}$$

respectively.

Adding (3.38) and (3.39), and using $\|x_{N_0+1} - x^*\| \leq r < 1$, we obtain

$$\begin{aligned} \sum_{j=j_0}^{\infty} \|F_{t_{l_j}}(x_{l_j})\|^2 &\leq \frac{1}{1-\mu} \|x_{N_0+1} - x^*\|^2 + \frac{1}{1-\mu} \sum_{i=0}^{\infty} \varepsilon_{k_i} \\ &\leq \frac{1}{1-\mu} + \frac{1}{1-\mu} M(M+1)\omega^2 C_0. \end{aligned} \tag{3.40}$$

By (3.25), (3.37) and (3.40), we conclude that

$$\sum_{k=N_0+1}^{\infty} \|F_{t_k}(x_k)\|^2 \leq \tilde{C},$$

where $\tilde{C} := \left(\left(1 - \sqrt{\frac{\mu}{2-\mu}} \right)^{-2} + \frac{1}{1-\mu} M(M+1) \right) \omega^2 C_0 + \frac{1}{1-\mu}$.

This together with (3.36) yields

$$\sum_{k=N_0+1}^{\infty} \|F_\gamma(x_k)\|^2 \leq \max\{\gamma\rho^{-2}\mu^{-1}L^*, 1\}^2 \sum_{k=N_0+1}^{\infty} \|F_{t_k}(x_k)\|^2 \leq C^*$$

with $C^* := \max\{\gamma\rho^{-2}\mu^{-1}L^*, 1\}^2 \tilde{C}$. Hence we conclude that

$$\min_{N_0+1 \leq k \leq N} \|F_\gamma(x_k)\|^2 \leq \frac{1}{N-N_0} \sum_{k=N_0+1}^{N} \|F_\gamma(x_k)\|^2 \leq \frac{1}{N-N_0} C^*.$$

□

Remark 3.7. If the pseudomonotone operator H is L-Lipschitz continuous, we will no longer need the line search step (3.1) in the EG-Anderson(1) algorithm, in which case we can take t_k to be the constant $t > 0$ satisfying $tL < 1$. At this situation, let $\{x_k\}$ be the sequence generated by the EG-Anderson(1) algorithm, then the following statements hold.

(i) (**Sequence convergence**) The sequence $\{x_k\}$ converges to a solution of $\mathrm{VI}(\Omega, H)$;

(ii) (**Best-iterate convergence rate**) $\min_{0 \leq k \leq N} \|F_t(x_k)\|^2 = O\left(\frac{1}{N}\right)$.

It can be seen that the EG-Anderson(1) algorithm can guarantee the sequence convergence as well as the EG method under the condition $tL < 1$, and we will show that it is faster than the EG method by numerical experiments.

4. Numerical experiments

In this section, we perform some numerical examples to compare the EG-Anderson(1) with Anderson(1) [7] and the EG method [31]. All the codes were written in Matlab (R2023b) and run on a MacBook Air (16.00GB of RAM).

In the following numerical experiments, the stopping rule for Examples 4.1-4.4 is set by

$$\|r(x_k)\| := \|F_1(x_k)\| = \|x_k - P_\Omega(x_k - H(x_k))\| < 10^{-8}$$

or the maximum iteration exceeds 10^4 times. The parameters in the EG-Anderson(1) are set as follows

$$\omega = 30, M = 5000, \tau = 0.6, \rho = 0.8, \mu = 0.5.$$

In the figures and tables of this section, 'Sec.' represents the CPU time in seconds and 'Iter.' represents the number of iterations. Specifically, the numbers in parentheses represent the number of Anderson step (3.3) executed. Moreover, '\' indicates that the number of iterations exceeds 10^4, and the corresponding CPU time is not counted, represented by –. Furthermore, the best performing algorithm in terms of the average number of iterations and CPU time is highlighted in bold for each combination of dimension n and parameter γ.

Example 4.1. [24] Consider the Harker-Pang problem with linear mapping $H(x) := Wx + w_0$, where $w_0 \in \mathbb{R}^n$ and

$$W := A^{\mathrm{T}}A + S + D.$$

Here, A is an $n \times n$ matrix, S is an $n \times n$ skew-symmetric matrix and D is an $n \times n$ diagonal matrix with nonnegative diagonal entries. Therefore, it follows that W is positive semidefinite. Let the feasible set be $\Omega := \{x \in \mathbb{R}^n : \mathbf{0} \le x \le 20l\}$, where $l = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^n$. It is clear that H is monotone and Lipschitz continuous.

We can easily obtain that the Lipschitz constant of H is $L = \|W\|$. Applying the EG-Anderson(1) to this example, instead of using line search, we can do experiment with a constant stepsize t that satisfies $tL < 1$.

In the following experiments, we let $t = \frac{0.7}{L}$, and every entry of the skew-symmetric matrix S is uniformly generated from $(-5, 5)$, and every diagonal entry of D is uniformly generated from $(0, 2)$, and A, w_0 are randomly generated.

Figure 1 compares the decreasing on the residual function by the EG-Anderson(1), Anderson(1) and the EG algorithms at the same random initial point for Example 4.1 with $n = 500$ and $n = 2000$, respectively. For different dimension n, Table 2 illustrates the average number of iterations and CPU time of the corresponding experiments at ten random initial points, where we see that the superiorities of the EG-Anderson(1) over Anderson(1) and the EG algorithms gradually emerges as the dimension increases.

Example 4.2. [41] Consider the quadratic fractional programming problem

$$\begin{array}{ll} \min & \varphi(x) := \frac{x^T Qx + a^T x + a_0}{b^T x + b_0} \\ \text{s.t.} & x \in \Omega := \{x \in \mathbb{R}^n : 2l \le x \le 10l\} \end{array}$$

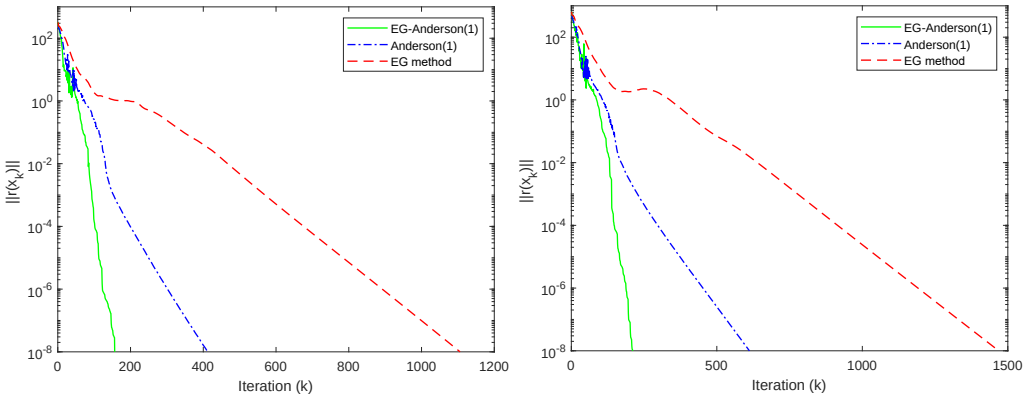
with

$$Q := Q_0^{\mathrm{T}} Q_0 + I, a := l + c, b := l + d, a_0 := 1 + c_0, b_0 := 1 + d_0,$$

where $I \in \mathbb{R}^{n \times n}$ is the identity matrix, l represents the vector that was defined in Example 4.1 and $Q_0 \in \mathbb{R}^{n \times n}$, $c, d \in \mathbb{R}^n$, $c_0, d_0 \in \mathbb{R}$ are randomly generated from a uniform distribution.

It is easily verified that $\Omega \subseteq \{x \in \mathbb{R}^n : b^T x + b_0 > 0\}$ and Q is positive definite, and consequently φ is pseudoconvex on Ω. Thus, $H(x) := \nabla\varphi(x)$ in $\mathrm{VI}(\Omega, H)$ can be written in the following explicit form:

$$H(x) = \frac{(b^T x + b_0)(2Qx + a) - b(x^T Qx + a^T x + a_0)}{(b^T x + b_0)^2}.$$



(a) $n = 500$ (b) $n = 2000$

FIGURE 1. Comparisons of the convergence behaviours of the EG-Anderson(1), Anderson(1) and the EG algorithms for Example 4.1

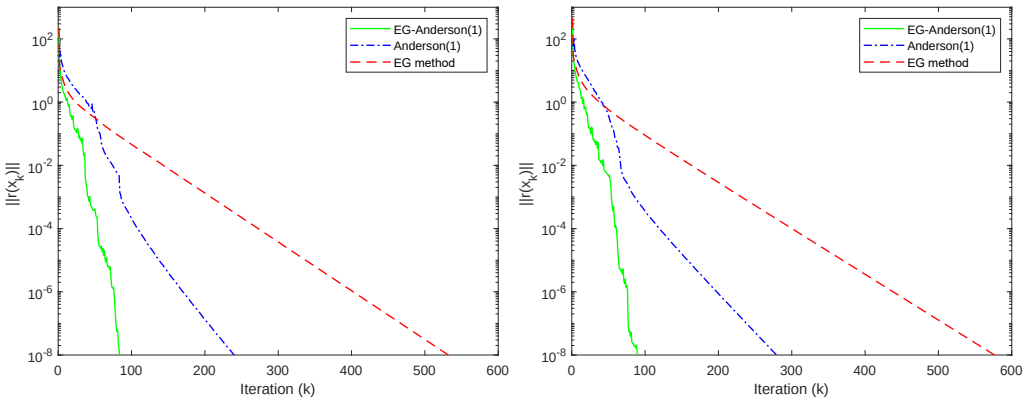
Sec.(avr) Iter.(avr)	EG-Anderson(1)	Anderson(1)	EG
$n = 100$	0.0038 **97.8** (96.6)	**0.0028** 226.4	0.0040 501.7
$n = 500$	**0.0233** **169** (166.4)	0.0243 402.6	0.1171 1103.8
$n = 1000$	**0.0972** **197.5** (195.6)	0.1172 545.8	0.5764 1417.8
$n = 2000$	**0.3462** **214.5** (211.6)	0.4789 595.4	2.2768 1465.2
$n = 5000$	**2.2158** **244** (240.6)	2.7644 600.9	14.1645 1538
$n = 10000$	**9.6154** **260.3** (256.7)	10.5268 563.4	63.4244 1685

TABLE 2. Comparisons of the three algorithms for Example 4.1

Let $\gamma = 0.6$. Starting from the same random initial point, Figure 2 shows the comparisons of the results obtained by the EG-Anderson(1), Anderson(1) and the EG algorithms for Example 4.2 with $n = 500$ and $n = 2000$. Table 3 presents the average number of iterations and CPU time for the experiments with more cases on n, conducted with ten different random initial points. Thus, the results indicate that the EG-Anderson(1) outperforms both Anderson(1) and the EG algorithms across various dimensions in terms of iterations and CPU time.

Example 4.3. [9] Consider the following nonlinear complementarity problem (NCP)

$$H(x) \geq 0,\ x \geq 0,\ x^{\mathrm{T}} H(x) = 0, \tag{4.1}$$



(a) $n = 500$ (b) $n = 2000$

FIGURE 2. Comparisons of the convergence behaviours of the EG-Anderson(1), Anderson(1) and the EG algorithms for Example 4.2

Sec.(avr) Iter.(avr)	EG-Anderson(1)	Anderson(1)	EG
$n = 100$	0.0051 **99** (99)	**0.0044** 299.7	0.0105 609
$n = 500$	**0.0439** **93.2** (91.4)	0.0446 238.5	0.1819 531
$n = 1000$	**0.1631** **102.6** (101.3)	0.1937 285.2	0.7851 588.8
$n = 2000$	**0.6839** **101.2** (99.9)	0.8337 285.3	3.4400 576
$n = 5000$	**5.5632** **95.3** (93.7)	6.9468 276	27.9809 538
$n = 10000$	**18.7886** **102.3** (100.9)	20.7397 262	89.8027 556

TABLE 3. Comparisons of the three algorithms for Example 4.2

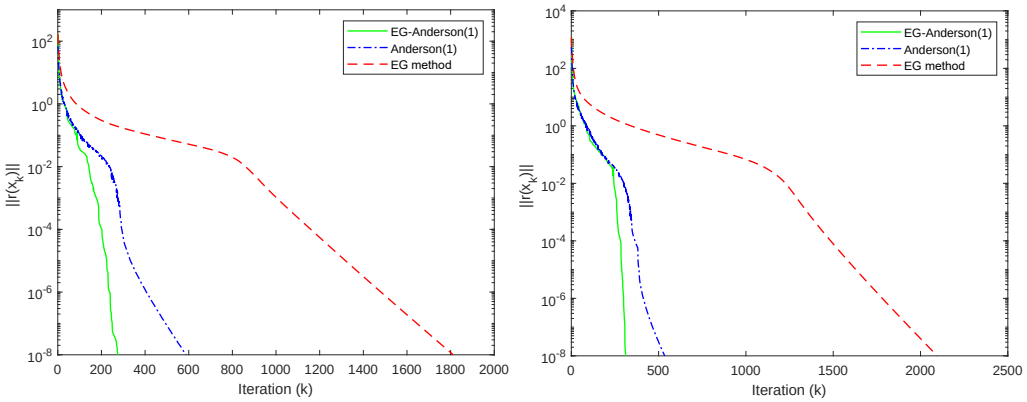
where $H : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$H(x) = (e^{-x^T Ux} + \kappa)(Px + \iota).$$

Here U is an $n \times n$ positive definite matrix, P is an $n \times n$ positive semidefinite matrix, $\iota \in \mathbb{R}^n$ and $\kappa > 0$. It is also shown that H is pseudomonotone [9]. Furthermore, the NCP in (4.1) can be equivalently formulated as VI(Ω, H) with $\Omega := \{x \in \mathbb{R}^n : x \geq \mathbf{0}\}$.

In numerical tests, we take $\kappa = 0.01, P = P_0^T P_0, U = U_0^T U_0$, where matrices $P_0, U_0 \in \mathbb{R}^{n \times n}$ and vector $\iota \in \mathbb{R}^n$ are randomly generated as follows with an input integer n:

$$P_0 = \text{randn}(n,n); U_0 = \text{randn}(n,n); \hat{x} = \max(\mathbf{0}, \text{randn}(n,1));$$



(a) $n = 500$ (b) $n = 2000$

FIGURE 3. Comparisons of the convergence behaviours of the EG-Anderson(1), Anderson(1) and the EG algorithms for Example 4.3

$$\iota = (-P * \hat{x}).*(\hat{x} > 0) + (-P * \hat{x} + \text{rand}(n,1)).*(\hat{x} == 0).$$

From the above generation, we know that $\hat{x}$ is a solution of (4.1).

In Figure 3, we present a comparative analysis of the outcomes achieved by applying the EG-Anderson(1), Anderson(1) and the EG algorithms to Example 4.3. All three methods in Figure 3 are initialized with the same random initial point. Table 4 summarizes the average number of iterations and CPU time for each method, where the corresponding experiments are repeated ten times with different random initial points.

In all dimensions, the average number of iterations and CPU time of the EG-Anderson(1) are significantly less than those of Anderson(1) and the EG algorithms. The advantages of the EG-Anderson(1) become more pronounced as the dimension increases. For instance, in the case of $n = 10000$, the EG-Anderson(1) achieves an average number of iterations of 552.2 and an average CPU time of 31.5962 seconds, significantly outperforming Anderson(1) and the EG algorithms.

Example 4.4. Consider the following partial differential equation (PDE) problem with free boundary

$$\begin{aligned} -\triangle u + \frac{9}{(1-p)^2} u^p + \delta e^{-u} &= c(\xi,\varsigma) && \text{in } \Lambda_+ \\ u &= 0 && \text{in } \Lambda_0 \\ u = |\nabla u| &= 0 && \text{on } \Gamma \\ u &= v(\xi,\varsigma) && \text{on } \partial\Lambda, \end{aligned} \tag{4.2}$$

where $p \in (0,1)$, $\delta \geq 0$, $\Lambda = (0,1)\times(0,1)$, $\Lambda_+ = \{(\xi,\varsigma) \in \Lambda : u(\xi,\varsigma) > 0\}$, $\Lambda_0 = \{(\xi,\varsigma) \in \Lambda : u(\xi,\varsigma) = 0\}$ and $\Gamma = \partial\Lambda_0 = \partial\Lambda_+ \cap \Lambda$ are unknown. Let $r^2 = \xi^2 + \varsigma^2$. We choose

$$c(\xi,\varsigma) = D(r,p)R(r,p) + \delta e^{-R(r,p)}$$

and

$$v(\xi,\varsigma) = R(r,p),$$

Sec.(avr) Iter.(avr)	EG-Anderson(1)	Anderson(1)	EG
$n = 100$ $\gamma = 0.27$	0.0041 **128.2** (125.4)	**0.0027** 256.7	0.0090 710.3
$n = 500$ $\gamma = 0.03$	**0.0405** **253.2** (249)	0.0424 566.2	0.2523 1754.4
$n = 1000$ $\gamma = 0.02$	**0.1802** **232.9** (228.8)	0.1823 504	1.0553 1513.1
$n = 2000$ $\gamma = 0.008$	**0.9466** **321.2** (316.6)	0.9478 622.3	6.1649 2122.6
$n = 5000$ $\gamma = 0.002$	**7.3880** **485.2** (478.6)	8.3522 1097	63.3128 4146.3
$n = 10000$ $\gamma = 0.0009$	**31.5962** **552.2** (544.5)	33.8899 1173.7	287.9682 4836.6

TABLE 4. Comparisons of the three algorithms for Example 4.3

where

$$D(r,p) := -\frac{3(3-p)\left[(3r-1)(1-p)+6r\right]}{r(1-p)^2(3r-1)^2} + \frac{27}{2(1-p)^2}\left(\frac{2}{3}\right)^p\left(\frac{3r-1}{2}\right)^{p-3}$$

and $R(r,p) := \left(\frac{3r-1}{2}\right)^{\frac{2}{1-p}} \max\left(0, r-\frac{1}{3}\right)$.

Then problem (4.2) has a solution as follows

$$u(\xi,\varsigma) = R(r,p) = \left(\frac{3r-1}{2}\right)^{\frac{2}{1-p}} \max\left(0, r-\frac{1}{3}\right).$$

Dividing the interval (0,1) into N subintervals of equal width h provides mesh points (ξ_i, ς_j) where

$$\xi_i = ih,\ i = 0,1,\ldots,N$$
$$\varsigma_j = jh,\ j = 0,1,\ldots,N.$$

Using the five point finite difference method for the problem (4.2) at grid (ξ_i, ς_j) gives

$$-u_{i,j+1} - u_{i,j-1} + 4u_{i,j} - u_{i+1,j} - u_{i-1,j} + \frac{9h^2}{(1-p)^2}u_{i,j}^p + h^2\delta e^{-u_{i,j}} = h^2 c_{i,j},\ (\xi_i,\varsigma_j) \in \Lambda_+$$

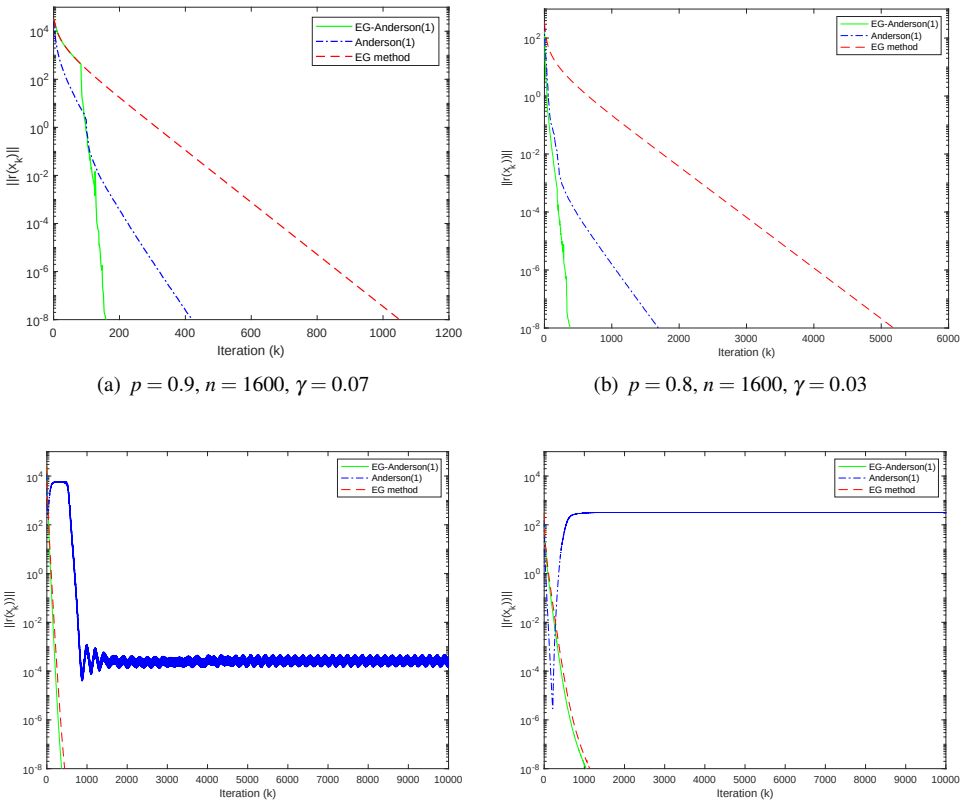
and

$$u_{i,j} = v_{i,j},\ (\xi_i,\varsigma_j) \in \partial\Lambda.$$

Let $x := (u_{1,1}, u_{2,1}, \ldots, u_{N-1,1}, u_{1,2}, \ldots, u_{N-1,N-1})^{\mathrm{T}} \in \mathbb{R}^{(N-1)^2}$ and $\tilde{c}, \tilde{v} \in \mathbb{R}^{(N-1)^2}$ be the corresponding vectors transformed by $c_{i,j}$ and $v_{i,j}$. Then, we obtain an NCP with

$$H(x) := Bx + E\max(0,x)^p + Vf(x) + q,$$

where B is a block tri-diagonal positive definite matrix of dimension $(N-1)^2 \times (N-1)^2$, E and V are both $(N-1)^2 \times (N-1)^2$ dimensional diagonal matrices with the diagonal elements being $\frac{9h^2}{(1-p)^2}$ and δh^2 respectively, $f(x) := (e^{-u_{1,1}}, e^{-u_{2,1}}, \ldots, e^{-u_{N-1,N-1}})^{\mathrm{T}} \in \mathbb{R}^{(N-1)^2}$



(a) $p = 0.9, n = 1600, \gamma = 0.07$ (b) $p = 0.8, n = 1600, \gamma = 0.03$

(c) $p = 0.9, n = 900, \gamma = 0.39$ (d) $p = 0.8, n = 900, \gamma = 0.39$

FIGURE 4. Comparisons of the convergence behaviours of the EG-Anderson(1), Anderson(1) and the EG algorithms for Example 4.4

and $q = -h^2\tilde{c} - \tilde{v} \in \mathbb{R}^{(N-1)^2}$. Note that the dimension of the corresponding complementarity problem is $n = (N-1)^2$. Furthermore, the NCP can be equivalently formulated as $\mathrm{VI}(\Omega, H)$ with $\Omega := \{x \in \mathbb{R}^n : x \geq \mathbf{0}\}$.

In the experiments, let $\delta = 1$. We set $p = 0.9$ and $p = 0.8$, respectively. The effectiveness of the EG-Anderson(1), Anderson(1) and the EG algorithms for Example 4.4 with $n = 900$ and $n = 1600$ is compared in Figure 4. All three methods are started from the same random initial point for each case.

As can be seen from Table 5, the average number of iterations for the EG-Anderson(1) is consistently smaller than that of Anderson(1) and the EG algorithms in all dimensions. Additionally, the EG-Anderson(1) has a shorter CPU usage time, highlighting its higher computational efficiency. In high-dimensional problems, such as $n = 8100$ and $n = 10000$, the EG-Anderson(1) outperforms the other two algorithms in terms of the average number of iterations and CPU time, demonstrating higher efficiency and performance.

Similar to Table 5, Table 6 presents the average number of iterations and CPU time required for each algorithm in Example 4.4, but with a different value of p (0.8 instead of 0.9).

Sec.(avr) Iter.(avr)	EG-Anderson(1)	Anderson(1)	EG
$n = 100$	**0.0036**	0.0056	0.0135
$\gamma = 0.01$	**64.7** (47.7)	280	530
$n = 900$	0.0165	**0.0158**	0.0604
$\gamma = 0.06$	**121.1** (73.1)	300.3	656
$n = 1600$	**0.0236**	0.0246	0.1022
$\gamma = 0.07$	**147.2** (79.2)	398.6	964
$n = 2500$	**0.0349**	0.0412	0.1740
$\gamma = 0.08$	**186.9** (97.9)	526.7	1288
$n = 4900$	**0.0857**	0.1050	0.5587
$\gamma = 0.09$	**296.8** (156.8)	771.8	2178
$n = 8100$	**0.1693**	0.1972	1.0951
$\gamma = 0.1$	**421.3** (228.3)	1064.9	3181
$n = 10000$	**0.2659**	0.3225	1.8438
$\gamma = 0.1$	**502.9** (276.9)	1270.7	3899

TABLE 5. Comparisons of the three algorithms for Example 4.4 with $p = 0.9$

Sec.(avr) Iter.(avr)	EG-Anderson(1)	Anderson(1)	EG
$n = 100$	0.0115	**0.0088**	0.0426
$\gamma = 0.005$	**371.2** (371.2)	774	2448.3
$n = 900$	**0.0383**	0.0683	0.3788
$\gamma = 0.02$	**376.4** (376.4)	1478.4	4512.3
$n = 1600$	**0.0843**	0.1161	0.8202
$\gamma = 0.03$	**375.1** (375.1)	1427	5183.8
$n = 2500$	**0.0640**	0.1019	0.6231
$\gamma = 0.05$	**397.9** (397.9)	1417.5	4761.1
$n = 4900$	**0.1355**	0.2959	1.8412
$\gamma = 0.06$	**474.4** (474.4)	2275.9	7554.5
$n = 8100$	**0.2472**	0.4831	–
$\gamma = 0.07$	**639.4** (638.4)	2692.5	\
$n = 10000$	**0.3341**	0.6177	–
$\gamma = 0.08$	**671.7** (669.7)	2660.8	\

TABLE 6. Comparisons of the three algorithms for Example 4.4 with $p = 0.8$

392 **Example 4.5.** Consider the following linear complementarity problem (LCP)

$$\tilde{M}x + \tilde{q} \geq 0,\ x \geq 0,\ x^{\mathrm{T}}(\tilde{M}x + \tilde{q}) = 0, \tag{4.3}$$

where $\tilde{M}$ is an $n \times n$ P-matrix and $\tilde{q} \in \mathbb{R}^n$ is a vector. Additionally, the LCP in (4.3) can be equivalently expressed as a nonmonotone VI(Ω, H), where $H(x) = \tilde{M}x + \tilde{q}$ and $\Omega = \{x \in \mathbb{R}^n : x \geq \mathbf{0}\}$.

Note that an $n \times n$ matrix $\tilde{M} = (m_{ij})$ is called a P-matrix if all principal minors of $\tilde{M}$ are positive. We refer to Example 4.4 from [14] to generate the matrix $\tilde{M}$ and vector $\tilde{q}$. First, we randomly generate a dense matrix $\tilde{A} \in \mathbb{R}^{n\times n}$ and a vector $\tilde{q} \in \mathbb{R}^n$, with elements uniformly distributed in the range $(-5, 5)$. By applying the QR decomposition to $\tilde{A}$, we obtain an upper triangular matrix $\tilde{N}$. Next, we replace the diagonal elements of $\tilde{N}$ with their absolute values, resulting in a triangular matrix $\tilde{M}$ with positive diagonal entries. This ensures that $\tilde{M}$ is a P-matrix.

Below, we compare the performance of the EG-Anderson(1) algorithm, iPCA from [18], IPA$_L$ from [51] and the semi-smooth Newton algorithm[2] from [20] for solving (4.3). The stopping criterion for the experiments is the same as that for the semi-smooth Newton algorithm, specifically,

$$0.5 \left\| s_k - x_k - (\tilde{M}x_k + \tilde{q}) \right\|^2 < 10^{-8}$$

or the maximum iteration exceeds 10^4 times, where $s_k = ((s_k)_1, (s_k)_2, \dots)^{\mathrm{T}}$ with $(s_k)_i = \sqrt{(x_k)_i^2 + (\tilde{M}x_k + \tilde{q})_i^2}$, $i = 1, \dots, n$. Note that for an n-dimensional vector z, $(z)_i$ represents the i-th component of z.

We can find that the Lipschitz constant of H is $\|\tilde{M}\|$. When we run the EG-Anderson(1) to this example, rather than employing a line search, we do experiment with a constant step size t that meets the condition $t\|\tilde{M}\| < 1$. In the following experiments, we let $t = \frac{0.7}{\|\tilde{M}\|}$. In both the iPCA and the IPA$_L$, the step size is taken to be the same as that in the EG-Anderson(1) algorithm. In addition, we set the parameter $\gamma = 1.5$ in the iPCA.

Table 7 presents the average number of iterations and CPU time required by the four algorithms to meet the stopping criterion for (4.3) across different dimensions, starting from ten random initial points. It can be observed that the EG-Anderson(1) algorithm consistently outperforms the other three algorithms in terms of CPU time across various dimensions.

5. Conclusions

This paper proposes an algorithm, called EG-Anderson(1) algorithm, for solving the pseudomonotone variational inequalities VI(Ω, H). This algorithm is based on the EG method and Anderson acceleration. Firstly, the global sequence convergence of the EG-Anderson(1) algorithm is proven without relying on the Lipschitz continuity and contractive condition that are required for the convergence analysis of the EG method and Anderson acceleration in prior research. Moreover, when H is locally Lipschitz continuous, the convergence rate of the residual function is analyzed and shown to be no worse than that of the EG method. Finally, the effectiveness of the EG-Anderson(1) algorithm has been validated through numerical experiments. The results demonstrate that it outperforms both Anderson(1) and the EG algorithms in terms of the number of iterations and CPU time, especially in the context of solving Harker-Pang problems, fractional programming problems, nonlinear complementarity problem, PDE problems with free boundary and linear complementarity problems.

[2]The solver, developed by Y. Tassa, https://www.mathworks.com/matlabcentral/fileexchange/20952-lcp-mcp-solver-newton-based

Sec.(avr) Iter.(avr)	EG-Anderson(1)	semi-smooth Newton	iPCA	IPA$_L$
$n = 100$	**0.0038** 64 (63.2)	0.0237 **29.4**	0.0044 149	0.0215 281.5
$n = 500$	**0.0188** 125.2 (124.8)	0.3223 **45.7**	0.0464 286.8	0.4225 638.7
$n = 1000$	**0.1372** 293.6 (288.2)	1.3163 **50.8**	0.7139 1221.2	19.1573 2718.1
$n = 2000$	**0.5563** 306.4 (293.8)	8.0195 **68.9**	2.3217 913.1	32.1018 2048.6
$n = 5000$	**4.3059** **408.3** (337.7)	17.1193 578.4	959.8955 976.8	854.2569 4994.4

TABLE 7. Comparisons of the four algorithms for Example 4.5

REFERENCES

[1] A. Alacaoglu, A. Böhm, Y. Malitsky, *Beyond the golden ratio for variational inequality algorithms*, J. Mach. Learn. Res., 24 (2023): 1-33.

[2] N. I. Alfredo, *An iterative algorithm for the variational inequality problem*, Comput. Appl. Math., 13 (1994): 103-114.

[3] H. An, X. Jia, H. F. Walker, *Anderson acceleration and application to the three-temperature energy equations*, J. Comput. Phys., 347 (2017): 1-19.

[4] D. G. Anderson, *Iterative procedures for nonlinear integral equations*, J. ACM, 12 (1965): 547-560.

[5] H. Attouch, M. O. Czarnecki, J. Peypouquet, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM J. Optim., 21 (2011): 1251-1274.

[6] V. Barbu, M. Röckner, *Stochastic variational inequalities and applications to the total variation flow perturbed by linear multiplicative noise*, Arch. Ration. Mech. Anal., 209 (2013): 797-834.

[7] W. Bian, X. Chen. *Anderson acceleration for nonsmooth fixed point problems*, SIAM J. Numer. Anal., 60 (2022): 2565-2591.

[8] W. Bian, X. Chen, C. T. Kelley, *Anderson acceleration for a class of nonsmooth fixed-point problems*, SIAM J. Sci. Comput., 43 (2021): S1-S20.

[9] R. I. Bot, E. R. Csetnek, P. T. Vuong, *The Forward-Backward-Forward method from discrete and continuous perspective for pseudomonotone variational inequalities in Hilbert spaces*, Eur. J. Oper. Res., 287 (2020): 49-56.

[10] G. Cai, Q. L. Dong, Y. Peng, *Strong convergence theorems for solving variational inequality problems with pseudomonotone and non-Lipschitz operators*, J. Optim. Theory Appl., 188 (2021): 447-472.

[11] Y. Censor, A. Gibali, S. Reich, *Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space*, Optim. Method Softw., 26 (2011): 827-845.

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

[13] X. Chen, C. T. Kelley, *Convergence of the EDIIS algorithm for nonlinear equations*, SIAM J. Sci. Comput., 41 (2019): A365-A379.

[14] X. Chen, Y. Ye, *On homotopy-smoothing methods for box-constrained variational inequalities*, SIAM J. Control Optim., 37 (1999): 589-616.

[15] R. W. Cottle, J. C. Yao, *Pseudo-monotone complementarity problems in Hilbert space*, J. Optim. Theory Appl., 75 (1992): 281-295.

[16] S. Dafermos, A. Nagurney, *A network formulation of market equilibrium problems and variational inequalities*, Oper. Res. Lett., 3 (1984): 247-250.

[17] S. V. Denisov, V. V. Semenov, L. M. Chabak, *Convergence of the modified extragradient method for variational inequalities with non-Lipschitz operators*, Cybern. Syst. Anal., 51 (2015): 757-765.

[18] Q. L. Dong, Y. J. Cho, L. L. Zhong, T. M. Rassias, *Inertial projection and contraction algorithms for variational inequalities*, J. Glob. Optim., 70 (2018): 687-704.
[19] F. Facchinei, J. S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, New York: Springer-Verlag, 2003.
[20] A. Fischer, *A Newton-type method for positive-semidefinite linear complementarity problems*, J. Optim. Theory Appl., 86 (1995): 585-608.
[21] K. Goebel, S. Reich, *Uniform Convexity, hperbolic geometry, and nonexpansive mappings*, Marcel Dekker, New York, 1984.
[22] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Amer. Math. Soc., 70 (1964): 709-710.
[23] T. N. Hai, *Two modified extragradient algorithms for solving variational inequalities*, J. Glob. Optim., 78 (2020): 91-106.
[24] P. T. Harker, J. S. Pang, *A damped-Newton method for the linear complementarity problem*, Lectures in Appl. Math., 26 (1990): 265-284.
[25] C. Heinemann, K. Sturm, *Shape optimization for a class of semilinear variational inequalities with applications to damage models*, SIAM J. Math. Anal., 48 (2016): 3579-3617.
[26] N. J. Higham, N. Strabić, *Anderson acceleration of the alternating projections method for computing the nearest correlation matrix*, Numer. Algorithms, 72 (2016): 1021-1042.
[27] K. Huang, S. Zhang, *An approximation-based regularized extra-gradient method for monotone variational inequalities*, arXiv preprint arXiv:2210.04440, 2022.
[28] K. Huang, S. Zhang, *Beyond monotone variational inequalities: Solution methods and iteration complexities*, arXiv preprint arXiv:2304.04153, 2023.
[29] S. Karamardian, *Complementarity problems over cones with monotone and pseudomonotone maps*, J. Optim. Theory Appl., 18 (1976): 445-454.
[30] S. Karamardian, S. Schaible, *Seven kinds of monotone maps*, J. Optim. Theory Appl., 66 (1990): 37-46.
[31] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Ekon. Mat. Metody, 12 (1976): 747-756.
[32] M. Lei, Y. He, *An extragradient method for solving variational inequalities without monotonicity*, J. Optim. Theory Appl., 188 (2021): 432-446.
[33] P. E. Maingé, *A hybrid extragradient-viscosity method for monotone operators and fixed point problems*, SIAM J. Control Optim., 47 (2008): 1499-1515.
[34] Y. Malitsky, *Projected reflected gradient methods for monotone variational inequalities*, SIAM J. Optim., 25 (2015): 502-520.
[35] Y. V. Malitsky, V.V. Semenov, *An extragradient algorithm for monotone variational inequalities*, Cybern. Syst. Anal., 50 (2014): 271-277.
[36] W. Ouyang, Y. Liu, A. Milzarek, *Descent Properties of an Anderson Accelerated Gradient Method with Restarting*, SIAM J. Optim., 34 (2024): 336-365.
[37] W. Ouyang, J. Tao, A. Milzarek, B. Deng, *Nonmonotone globalization for Anderson acceleration via adaptive regularization*, J. Sci. Comput., 96 (2023): 5.
[38] J. Peypouquet, S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. Convex Anal., 17 (2010): 1113-1163.
[39] R. T. Rockafellar, J. Sun, *Solving Lagrangian variational inequalities with applications to stochastic programming*, Math. Program., 181 (2020): 435-451.
[40] M. V. Solodov, B. F. Svaiter, *A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator*, Set-Valued Anal., 7 (1999): 323-345.
[41] D. V. Thong, V. T. Dung, P. K. Anh, H. Van Thang, *A single projection algorithm with double inertial extrapolation steps for solving pseudomonotone variational inequalities in Hilbert space*, J. Comput. Appl. Math., 426 (2023): 115099.
[42] D. V. Thong, A. Gibali, *Extragradient methods for solving non-Lipschitzian pseudomonotone variational inequalities*, J. Fixed Point Theory Appl., 21 (2019): 1-19.
[43] D. V. Thong, Y. Shehu, O. S. Iyiola, *Weak and strong convergence theorems for solving pseudomonotone variational inequalities with non-Lipschitz mappings*, Numer. Algorithms, 84 (2020): 795-823.
[44] A. Toth, C. T. Kelley, *Convergence analysis for Anderson acceleration*, SIAM J. Numer. Anal., 53 (2015): 805-819.
[45] P. T. Vuong, Y. Shehu, *Convergence of an extragradient-type method for variational inequality with applications to optimal control problems*, Numer. Algorithms, 81 (2019): 269-291.
[46] W. Y. Wang, B. B. Ma, *Extragradient type projection algorithm for solving quasimonotone variational inequalities*, Optimization, 73 (2024): 2639-2655.

[47] J. Yang, H. Liu, *A modified projected gradient method for monotone variational inequalities*, J. Optim. Theory Appl., 179 (2018): 197-211.
[48] J. Yang, H. Liu, *Strong convergence result for solving monotone variational inequalities in Hilbert space*, Numer. Algorithms, 80 (2019): 741-752.
[49] J. Yang, H. Liu, Z. Liu, *Modified subgradient extragradient algorithms for solving monotone variational inequalities*, Optimization, 67 (2018): 2247-2258.
[50] Y. Yao, O. S. Iyiola, Y. Shehu, *Subgradient extragradient method with double inertial steps for variational inequalities*, J. Sci. Comput., 90 (2022): 1-29.
[51] M. Ye, *An infeasible projection type algorithm for nonmonotone variational inequalities*, Numer. Algorithms, 89 (2022): 1723-1742.
[52] M. Ye, Y. He, *A double projection method for solving variational inequalities without monotonicity*, Comput. Optim. Appl., 60 (2015): 141-150.
[53] J. Zhang, B. O'Donoghue, S. Boyd, *Globally convergent type-I Anderson acceleration for nonsmooth fixed-point iterations*, SIAM J. Optim., 30 (2020): 3170-3197.
[54] X. Zhang, W. Zhao, G. Zhou, W. Liu, *An accelerated monotonic convergent algorithm for a class of non-Lipschitzian NCP(F) involving an M-matrix*, J. Comput. Appl. Math., 397 (2021): 113624.

SCHOOL OF MATHEMATICS, HARBIN INSTITUTE OF TECHNOLOGY, HARBIN, CHINA
Current address: Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
Email address: hitxin.qu@connect.polyu.hk

SCHOOL OF MATHEMATICS, HARBIN INSTITUTE OF TECHNOLOGY, HARBIN, CHINA
Email address: bianweilvse520@163.com

DEPARTMENT OF APPLIED MATHEMATICS, THE HONG KONG POLYTECHNIC UNIVERSITY, HUNG HOM, KOWLOON, HONG KONG, CHINA
Email address: xiaojun.chen@polyu.edu.hk