TAIWANESE JOURNAL OF MATHEMATICS
Vol. 28, No. 1, pp. 187–210, February 2024
DOI: 10.11650/tjm/230906

A Modified Tseng's Algorithm with Extrapolation from the Past for Pseudo-monotone Variational Inequalities

Buris Tongnoi

Abstract. We present Tseng's forward-backward-forward method with extrapolation from the past for pseudo-monotone variational inequalities in Hilbert spaces. In addition, we propose a variable stepsize scheme of the extrapolated Tseng's algorithm governed by the operator which is pseudo-monotone, Lipschitz continuous and sequentially weak-to-weak continuous. We also investigate the algorithm's adaptive stepsize scenario, which arises when it is impossible to calculate the Lipschitz constant of a pseudo-monotone operator correctly. Finally, we prove a weak convergence theorem and conduct a numerical experiment to support it.

1. Introduction

Variational inequalities (VIs) are beneficial mathematical models for solving various problems, like saddle problems, equilibrium problems, obstacle problems, and others; see [12,19] and reference therein.

In this paper, we are concerned with the *variational inequality* (VI) in the type of Hartman-Stampacchia (Stampacchia type) [14]:

$$\text{find } x^* \in C \text{ such that } \langle F(x^*), x - x^* \rangle \geq 0, \forall\, x \in C, \tag{1.1}$$

where C is a nonempty closed convex subset of Hilbert space $\mathscr{H}$ endowed with an inner product $\langle \cdot\,, \cdot \rangle$ and the corresponding norm $\| \cdot \|$ and F is monotone and L-Lipschitz continuous operator for $L \geq 0$. We denote the inequality of (1.1) by $\mathrm{VI}(F, C)$ and assume that its solution set is represented by $\Omega \neq \emptyset$.

The projected-gradient algorithm is the simplest method for solving variational inequalities, and it is defined as follows: for a starting point $x_0 \in \mathscr{H}$,

$$x_{n+1} = P_C(x_n - \lambda F(x_n)), \quad \forall\, n \geq 0,$$

where P_C denotes the projection operator onto the closed convex set $C \subseteq \mathscr{H}$ and $\lambda > 0$. However, this method does not necessarily provide the convergence if F is only monotone

Received February 28, 2023; Accepted September 25, 2023.
Communicated by Jein-Shan Chen.
2020 *Mathematics Subject Classification.* 47J20, 65K10, 65K15, 47H05.
Key words and phrases. variational inequalities, pseudo-monotonicity, Tseng's FBF algorithm, extrapolation from the past.

(see [12, Example 12.1.3] for example). In order to converge, the method requires additional assumptions like F being cocoercive (or inverse strongly monotone, see [1, 29]) or strongly (pseudo) monotone (see [12, 17]).

The extragradient method proposed by Korpelevich in [20], which is used to solve variational inequalities governed by Lipschitz continuous and pseudo-monotone operators, reads as for a starting point $x_0 \in \mathscr{H}$,

$$y_n = P_C(x_n - \lambda F(x_n)), \quad x_{n+1} = P_C(x_n - \lambda F(y_n)),$$

where $\lambda \in (0, 1/L)$. We can see that the algorithm needs to compute the projection onto C two times, which sometimes affects the method's efficiency if the projection is not easy to calculate. The extragradient method has gained overwhelming attention from several authors to modify this method in multiple ways; see, e.g., [8, 15, 18]. The work of Popov [24] gives us a sophisticated idea that we can reuse and store the computation of the projection term in each iterative scheme for extragradient, namely

$$y_n = P_C(x_n - \lambda F(y_{n-1})), \quad x_{n+1} = P_C(x_n - \lambda F(y_n)),$$

where $\lambda \in (0, 1/(3L))$. This modern concept called the *extrapolated technique or extrapolation from the past* which is given in [13] and note that the latter approach is working within a smaller stepsize.

As an alternative extragradient method applied to solving the monotone inclusions, Tseng [28] proposed the forward-backward-forward (FBF) algorithm. Furthermore, the combination of Tseng's algorithm and the extrapolated technique is also suggested in [3, 27] as known as Tseng's algorithm with extrapolation from the past (FBF-EP). We can demonstrate the general iterative scheme as following statement:

$$\text{Tseng-G} \qquad y_n = P_C(x_n - \lambda F(s_n)), \quad x_{n+1} = y_n + \lambda(F(s_n) - F(y_n)),$$

for any choice of s_n, we obtain that

1. If $s_n = x_n$, it becomes an approach based on Tseng's forward-backward-forward (FBF) algorithm, see also [3, 6, 28].

2. If $s_n = y_{n-1}$, it turns into the method, which relies on the *Tseng's algorithm with extrapolation from the past* (FBF-EP):

$$y_n = P_C(x_n - \lambda F(y_{n-1})), \quad x_{n+1} = y_n + \lambda(F(y_{n-1}) - F(y_n)).$$

In addition, if we substitute x_{n+1} into the first step of FBF-EP at y_{n+1}, we get that

$$y_{n+1} = P_C(y_n - 2\lambda F(y_n) + \lambda F(y_{n-1})),$$

which is established from *forward-reflected-backward* algorithm in [21]. Notably, in case C is a whole space $\mathscr{H}$, both methods turn into *optimistic mirror descent* (see [25, 26]) in the unconstrained case shown as

$$y_{n+1} = y_n - 2\lambda F(y_n) + \lambda F(y_{n-1}),$$

known as *Optimistic Gradient Descent Ascent* (OGDA) for Generative Adversarial Network (GANs), see [6, 10, 11] for more details.

Both FBF and FBF-EP require only one projection per each iterative computation and provide the weak convergence of $(x_n)_{n\geq 0}$ to a solution of $\mathrm{VI}(F, C)$ for monotone Lipschitz operator F, see [7,27,28]. Because of its simplicity and generality, this algorithm has attracted a lot of attention from many researchers, for example in [4, 7, 27]. The improvement of Tseng's method (FBF) goes further in the relaxation version for the pseudo-monotone and sequentially weak-to-weak continuous operator F based on FBF, proposed by Boţ et al. in [5]. The algorithm reads as follows: for the starting point $x_0 \in \mathscr{H}$,

$$y_n = P_C(x_n - \lambda F(x_n)), \quad x_{n+1} = \rho_n(y_n + \lambda(F(x_n) - F(y_n))) + (1 - \rho_n)x_n, \tag{1.2}$$

where $\lambda > 0$ is the stepsize and $(\rho_n)_{n\geq 0} \subseteq [0, 1]$ is the sequence of relaxation parameters. This work has shown that the convergence result holds when F is a pseudo-monotone (not necessary to be monotone), Lipschitz continuous and sequentially weak-to-weak-continuous operator. Sometimes, the Lipschitz constant L is unknown or is not simple to calculate. Therefore, Boţ et al. [5] introduced an adaptive stepsize scenario for the method (1.2).

The work of Boţ et al. [5] motivates us to investigate the convergence of Tseng's algorithm with extrapolation from the past (FBF-EP) in case the operator F is pseudo-monotone, Lipschitz continuous and sequentially weak-to-weak-continuous. In addition, we propose an adaptive stepsize approach for FBF-EP, which does not depend on the knowledge of the Lipschitz constant. We also prove that the convergence statement holds in finite dimensional spaces under weaker assumptions on F. In the final section of this work, we additionally perform a numerical experiment on pseudo-monotone variational inequalities.

Considering that the relaxation version of FBF has already been proposed for the pseudo-monotone and sequentially weak-to-weak continuous assumptions, it would be of great interest to investigate the relaxation version of FBF-EP with the same operator F assumptions in this work. We have attempted to solve the relaxation version of FBF-EP for both pseudo-monotone and sequentially weak-to-weak continuous properties, as well as monotone Lipschitzian properties of F. However, a conclusive weak convergence result cannot be determined at this time. Indeed, the convergence results for Tseng's forward-backward-forward algorithm (FBF) are typically established through the Fejér monotone sequence, which satisfies $\|x_{n+1} - x\| \leq \|x_n - x\|$, $\forall\, x \in C \subseteq \mathscr{H}$, $\forall\, n \in \mathbb{N}$, see [1, Definition 5.1]. In contrast, Tseng's algorithm with extrapolation from the past (FBF-EP), which yields a similar inequality, can be expressed as shown in equation (3.9). Due to the appearance of a negative term on the right-hand side of the inequality, which is challenging to eliminate, drawing any definitive conclusions has become a significant challenge. Consequently, we present an open question to the reader below.

Open Question. The weak convergence of Tseng's algorithm with extrapolation in the relaxation version for monotone Lipschitz (and for both pseudo-monotone and sequentially weak-to-weak continuous) operators has remained unresolved due to the presence of a perturbed term in the inequality of Fejer-monotone sequence, which poses a significant obstacle to obtaining weak convergence results.

2. Preliminaries

Before we present the main results, let us give the relevant background knowledge. Throughout this work, the symbol $\rightharpoonup$ denotes weak convergence, $\to$ stands for strong convergence, and $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_{++}$ represent the set of all natural numbers, the set of all real numbers and the set of all positive real numbers, respectively.

We describe various properties of the operator F as follows:

Definition 2.1. Let C be a nonempty subset of the real Hilbert space $\mathscr{H}$. The mapping $F\colon \mathscr{H} \to \mathscr{H}$ is said to be

(a) *pseudo-monotone* on C if it holds that for every $x, y \in C$,

$$\langle F(x), y - x\rangle \geq 0 \quad \Longrightarrow \quad \langle F(y), y - x\rangle \geq 0;$$

(b) *monotone* on C if it holds that for every $x, y \in C$,

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

Note that every monotone operator is pseudo-monotone but the pseudo-monotone is not necessarily monotone (see, [16] for example).

The operator $F\colon \mathscr{H} \to \mathscr{H}$ is called *Lipschitz continuous* with Lipschitz constant $L > 0$, if for every $x, y \in \mathscr{H}$ it holds that

$$\|F(x) - F(y)\| \leq L\|x - y\|,$$

and we say that the operator F is *sequentially weak-to-weak continuous*, if for every sequence $(x_n)_{n\geq 0}$ that converges weakly to x the sequence $(F(x_n))_{n\geq 0}$ also converges weakly to $F(x)$.

The characterization of projection mapping is a useful tool that we will employ to present our main results, as demonstrated in the theorem below.

Theorem 2.2. [1, Theorem 3.14] *Let C be a nonempty closed convex subset of $\mathscr{H}$. Then for every x and p in $\mathscr{H}$,*

$$p = P_C x \quad \iff \quad [p \in C \text{ and } \forall\, y \in C, \langle y - p, x - p\rangle \leq 0].$$

Since the set C is a nonempty, closed, and convex subset of the Hilbert space $\mathscr{H}$, there exists a related property of C that is presented in the following theorem.

Theorem 2.3. [1, Theorem 3.32] *Let C be a convex subset of $\mathscr{H}$. Then the following statements are equivalent:*

(i) *C is weakly sequentially closed.*
(ii) *C is sequentially closed.*
(iii) *C is closed.*
(iv) *C is weakly closed.*

Next, we present the fertile lemma confirming the sequence's convergence and summability associated with a specific form of inequality. This lemma is essential and applied multiple times in this work. Additionally, we provide the equivalent property of the convex function in the subsequent theorem.

Lemma 2.4. [7, Lemma 2.1] *Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $[0, +\infty)$ (bounded from below), let $(\beta_n)_{n\in\mathbb{N}}$ be a sequence in $[0, +\infty)$, and let $(\epsilon_n)_{n\in\mathbb{N}}$ be a summable sequence (i.e., $\sum_{n\in\mathbb{N}} \epsilon_n < +\infty$) in $[0, +\infty)$ such that $\forall\, n \in \mathbb{N}$, $\alpha_{n+1} \leq \alpha_n - \beta_n + \epsilon_n$. Then $(\alpha_n)_{n\in\mathbb{N}}$ has a limit, and $(\beta_n)_{n\in\mathbb{N}}$ is summable.*

Theorem 2.5. [1, Theorems 9.1 or 10.23] *Let $f \colon \mathscr{H} \to (-\infty, +\infty]$ be (quasi) convex. Then the following statements are equivalent:*

(i) *f is weakly sequentially lower semicontinuous.*

(ii) *f is sequentially lower semicontinuous.*

(iii) *f is lower semicontinuous.*

(iv) *f is weakly lower semicontinuous.*

Now, we present the proposition of a continuous pseudo-monotone operator on a nonempty, convex, and closed set. Additionally, this proposition establishes an identical relationship between the Stampacchia and Minty types of variational inequality (VI).

Proposition 2.6. [9, Lemma 2.1] *Let C be a nonempty, convex and closed subset of the real Hilbert space $\mathscr{H}$ and let $F\colon \mathscr{H} \to \mathscr{H}$ be an operator which is pseudo-monotone on C and continuous. Then for every $x \in C$ we have*

$$\langle F(x), y - x\rangle \geq 0 \quad \Longleftrightarrow \quad \langle F(y), y - x\rangle \geq 0, \quad \forall\, y \in C.$$

Another version of variational inequality is called *Minty type*:

$$\text{Find } x^* \in C \text{ such that } \langle F(x), x - x^*\rangle \geq 0, \forall\, x \in C.$$

Notably, Proposition 2.6 implies that the solution sets of two variational inequalities in both Stampacchia type (1.1) and Minty type are the same when they are performed over a nonempty closed convex set and governed by pseudo-monotone and continuous operators.

To demonstrate the weak convergence of our proposed algorithm in this work, we utilize a valid theorem known as Opial's Lemma, presented below.

Lemma 2.7. [23, Opial's Lemma] *Let S be a nonempty subset of $\mathscr{H}$, and $(x_k)_{k\geq 0}$ a sequence of elements of $\mathscr{H}$. Assume that*

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

(ii) *every weak sequential limit point of $(x_k)_{k\geq 0}$ belongs to S as $k \to +\infty$.*

Then x_k converges weakly as $k \to +\infty$ to a point in S.

In the next section we will demonstrate our approach and its proof using the appropriate tools.

3. Main results

In this section, we introduce Tseng's algorithm with extrapolation from the past for pseudo-monotone variational inequalities, and we show a weak convergence result using Opial's lemma.

Theorem 3.1. *Let $\Omega \neq \emptyset$ be the solution set of the variational inequality problem of F on C, namely $\mathrm{VI}(F, C)$, let $F\colon \mathscr{H} \to \mathscr{H}$ be pseudo-monotone on $\mathscr{H}$, Lipschitz continuous with constant L and sequentially weak-to-weak continuous. Let C be a nonempty, convex and closed subset of the real Hilbert space $\mathscr{H}$. Let $x_0, y_{-1} \in \mathscr{H}$. Assume that the sequence $(x_n)_{n\geq 0}$ is generated by the following algorithm*

$$y_n = P_C(x_n - \lambda_n F(y_{n-1})), \quad x_{n+1} = y_n + \lambda_n(F(y_{n-1}) - F(y_n)) \tag{3.1}$$

with $0 < \liminf_{n\to\infty} \lambda_n \leq \limsup_{n\to\infty} \lambda_n < 1/(2L)$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Ω.

The idea of the proof. We first try to show the inequality (3.11) by using $\Omega \neq \emptyset$, the pseudo-monotonicity of F, Theorem 2.2 (property of the projection operator) and the Lipschitz continuity of F. Then we rearrange formula (3.11) and apply Lemma 2.4 to conclude that $\lim_{n\to+\infty} \|x_n - x^*\|^2$ exists which is the first condition of Opial's lemma (see Lemma 2.7). For the rest of the proof, we need to verify the second condition of Opial's lemma, namely, if $\widehat{x}$ is a weak sequential cluster point of $(x_n)_{n\geq 0}$, then $\widehat{x} \in C$.

Let x^* be an arbitrary element in Ω and let $n \geq 0$ be fixed. Then we have

$$\langle F(x^*), y - x^* \rangle \geq 0, \quad \forall\, y \in C.$$

Substituting $y := y_n \in C$ into the inequality yields

$$\langle F(x^*), y_n - x^* \rangle \geq 0.$$

From the pseudo-monotonicity of F on C, it follows that

$$\langle F(y_n), y_n - x^* \rangle \geq 0. \tag{3.2}$$

Since $y_n = P_C(x_n - \lambda_n F(y_{n-1}))$ according to Theorem 2.2, we get

$$\langle y - y_n, y_n - x_n + \lambda_n F(y_{n-1}) \rangle \geq 0, \quad \forall\, y \in C, \tag{3.3}$$

which yields for $y = x^* \in C$:

$$\langle x^* - y_n, y_n - x_n + \lambda_n F(y_{n-1}) \rangle \geq 0. \tag{3.4}$$

Multiplying both sides of (3.2) by $\lambda_n > 0$ we have

$$\langle \lambda_n F(y_n), y_n - x^* \rangle \geq 0 \quad (\text{or } \langle x^* - y_n, -\lambda_n F(y_n) \rangle \geq 0),$$

adding the above inequality to (3.4) yields

$$\langle x^* - y_n, y_n - x_n + \lambda_n F(y_{n-1}) - \lambda_n F(y_n) \rangle \geq 0,$$

or, equivalently,

$$\langle x^* - y_n, x_{n+1} - x_n \rangle \geq 0.$$

Then, using (3.1) we obtain that

$$\begin{aligned}
&\langle x_{n+1} - x^*, x_{n+1} - x_n \rangle \\
&\leq \langle x_{n+1} - y_n, x_{n+1} - x_n \rangle \\
&= \|x_{n+1} - x_n\|^2 + \langle x_n - y_n, x_{n+1} - x_n \rangle \\
&= \|x_{n+1} - x_n\|^2 + \langle x_n - y_n, y_n + \lambda_n F(y_{n-1}) - \lambda_n F(y_n) - x_n \rangle \\
&= \|x_{n+1} - x_n\|^2 - \|x_n - y_n\|^2 + \lambda_n \langle x_n - y_n, F(y_{n-1}) - F(y_n) \rangle.
\end{aligned} \tag{3.5}$$

On the other hand, we have

$$\|x_{n+1}-x^*\|^2-\|x_n-x^*\|^2+\|x_{n+1}-x_n\|^2=2\langle x_{n+1}-x^*,x_{n+1}-x_n\rangle. \tag{3.6}$$

Combining (3.5) and (3.6), we obtain that

$$\begin{aligned}
&\|x_{n+1}-x^*\|^2\\
&=\|x_n-x^*\|^2-\|x_{n+1}-x_n\|^2+2\langle x_{n+1}-x^*,x_{n+1}-x_n\rangle\\
&\le\|x_n-x^*\|^2-\|x_{n+1}-x_n\|^2\\
&\quad+2\big[\|x_{n+1}-x_n\|^2-\|x_n-y_n\|^2+\lambda_n\langle x_n-y_n,F(y_{n-1})-F(y_n)\rangle\big]\\
&=\|x_n-x^*\|^2+\|x_{n+1}-x_n\|^2-2\|x_n-y_n\|^2+2\lambda_n\langle x_n-y_n,F(y_{n-1})-F(y_n)\rangle.
\end{aligned} \tag{3.7}$$

Using the Lipschitz continuity of F and (3.1), we obtain that

$$\begin{aligned}
&\|x_{n+1}-x_n\|^2\\
&=\|y_n+\lambda_nF(y_{n-1})-\lambda_nF(y_n)-x_n\|^2\\
&=\|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n,F(y_{n-1})-F(y_n)\rangle+\lambda_n^2\|F(y_{n-1})-F(y_n)\|^2\\
&\le\|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n,F(y_{n-1})-F(y_n)\rangle+\lambda_n^2L^2\|y_{n-1}-y_n\|^2.
\end{aligned} \tag{3.8}$$

From (3.7) and (3.8), we derive

$$\begin{aligned}
&\|x_{n+1}-x^*\|^2\\
&\le\|x_n-x^*\|^2+\big[\|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n,F(y_{n-1})-F(y_n)\rangle+\lambda_n^2L^2\|y_{n-1}-y_n\|^2\big]\\
&\quad-2\|x_n-y_n\|^2+2\lambda_n\langle x_n-y_n,F(y_{n-1})-F(y_n)\rangle\\
&=\|x_n-x^*\|^2-\|y_n-x_n\|^2+\lambda_n^2L^2\|y_{n-1}-y_n\|^2.
\end{aligned} \tag{3.9}$$

By parallelogram identity, we know that $\|y_n-y_{n-1}\|^2+\|(x_n-y_n)+(x_n-y_{n-1})\|^2=2\|x_n-y_n\|^2+2\|x_n-y_{n-1}\|^2$, hence together with the Lipschitz property of F we obtain that

$$\begin{aligned}
\|x_n-y_n\|^2&\ge-\|x_n-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\\
&=-\|y_{n-1}+\lambda_{n-1}(F(y_{n-2})-F(y_{n-1}))-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\\
&\ge-(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2.
\end{aligned} \tag{3.10}$$

It follows from (3.9) and (3.10) that

$$\begin{aligned}
&\|x_{n+1}-x^*\|^2\\
&\le\|x_n-x^*\|^2-\left[-(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\right]+\lambda_n^2L^2\|y_{n-1}-y_n\|^2.
\end{aligned}$$

Thus, we have

$$(3.11)\quad \|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-(\lambda_n L)^2\right)\|y_n-y_{n-1}\|^2 \le \|x_n-x^*\|^2+(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2.$$

Since $\limsup_{n\to\infty}\lambda_n < 1/(2L)$, then we have $\limsup_{n\to\infty}[2(\lambda_n L)^2] < 1/2$. This means that

$$\liminf_{n\to\infty}\left\{\frac{1}{2}-2(\lambda_n L)^2\right\} > 0,$$

and so there exists $n_0 \in \mathbb{N}$ such that

$$(3.12)\qquad \frac{1}{2}-2(\lambda_n L)^2 > \eta > 0, \quad \forall\, n > n_0, \text{ for some } \eta \in \mathbb{R}_{++}.$$

Now, from (3.11) we can derive the following inequality, for all $n > n_0$,

$$(3.13)\qquad \begin{aligned} &\|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-2(\lambda_n L)^2\right)\|y_n-y_{n-1}\|^2+(\lambda_n L)^2\|y_n-y_{n-1}\|^2 \\ &\le \|x_n-x^*\|^2+(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2. \end{aligned}$$

By (3.12), (3.13) and Lemma 2.4, we obtain that $\|x_n-x^*\|^2+(\lambda_{n-1}L)^2\|y_{n-2}-y_{n-1}\|^2$ converges (hence it is bounded) and $\sum_{n\in\mathbb{N}}\big(1/2-2(\lambda_n L)^2\big)\|y_n-y_{n-1}\|^2 < +\infty$. Moreover, from (3.12) we have

$$(3.14)\qquad \sum_{n\in\mathbb{N}}\|y_n-y_{n-1}\|^2 < +\infty,$$

which implies $y_n - y_{n-1} \to 0$ as $n \to +\infty$. Recall from (3.9) that $\|x_{n+1}-x^*\|^2 \le \|x_n-x^*\|^2-\|y_n-x_n\|^2+(\lambda_n L)^2\|y_{n-1}-y_n\|^2$. Since $\limsup_{n\to+\infty}\lambda_n < 1/(2L)$ (and $0 < \liminf_{n\in\mathbb{N}}\lambda_n$) and (3.14) by using Lemma 2.4 again, we obtain

$$(3.15)\qquad \sum_{n\in\mathbb{N}}\|y_n-x_n\|^2 < +\infty,$$

and $\lim_{n\to+\infty}\|x_n-x^*\|^2$ exists. Furthermore, we also obtain that $\lim_{n\to+\infty}\|y_n-x_n\|^2 = 0$.

In this part of the proof we will show the second part of Opial's Lemma. We proceed similarly as in [5, Theorem 2.1]. Let $\widehat{x} \in \mathscr{H}$ be a weak sequential cluster point of x_n as $n \to +\infty$. Since $\lim_{n\to+\infty}\|x_n-y_n\| = 0$, we also have $y_n \rightharpoonup \widehat{x}$ as $n \to +\infty$. Furthermore, since F is Lipschitz continuous, $\|F(y_{n-1})-F(y_n)\| \to 0$ as $n \to +\infty$. We want to show that $\widehat{x} \in \Omega$. We assume that $F(\widehat{x}) \ne 0$, otherwise the conclusion follows automatically. For every $n \ge 0$. Since $(y_n)_{n\ge 0} \subseteq C$, and C is weakly closed (by Theorem 2.3), then we have $\widehat{x} \in C$. From (3.3), for all $y \in C$,

$$\langle y-y_n, x_n-\lambda_n F(y_{n-1})-y_n\rangle \le 0,$$

or, equivalently,

$$\frac{1}{\lambda_n}\langle x_n - y_n, y - y_n\rangle \leq \langle F(y_{n-1}) - F(y_n), y - y_n\rangle + \langle F(y_n), y - y_n\rangle. \tag{3.16}$$

Considering the inequality (3.16) and taking into account that $\lim_{n\to+\infty}\|x_n - y_n\| = 0$, $\|F(y_{n-1}) - F(y_n)\| \to 0$ (as $n \to +\infty$), $(y_n)_{n\geq 0}$ is bounded and $\liminf_{n\to+\infty}\lambda_n > 0$, it follows

$$\forall\, y \in C, \quad 0 \leq \liminf_{n\to+\infty}\langle F(y_n), y - y_n\rangle.$$

On the other hand, we have that $(y_n)_{n\geq 0}$ converges weakly to $\widehat{x}$ as $n \to +\infty$. Since F is sequentially weak-to-weak continuous, $(F(y_n))_{n\geq 0}$ converges weakly to $F(\widehat{x})$ as $n \to +\infty$. Because the norm mapping is convex (or quasi convex) by Theorem 2.5, it is weakly sequentially lower semicontinuous. So we have $0 < \|F(\widehat{x})\| \leq \liminf_{n\to+\infty}\|F(y_n)\|$. Then there exists $n_{-1} \geq 0$ such that $F(y_n) \neq 0$ for all $n \geq n_{-1}$. Let $(\epsilon_k)_{k\geq 0}$ be a positive strictly decreasing sequence which converges to 0 as $k \to +\infty$ and $y \in C$. Since $\sup_{N\geq 0}\inf_{n\geq N}\langle F(y_n), y - y_n\rangle = \liminf_{n\to+\infty}\langle F(y_n), y - y_n\rangle > -\epsilon_0$, there exits $N_0 \geq 0$ such that $\inf_{n\geq N_0}\langle F(y_n), y - y_n\rangle > -\epsilon_0$. Taking $n_0 > \max\{N_0, n_{-1}\}$, we have $\langle F(y_{n_0}), y - y_{n_0}\rangle + \epsilon_0 > 0$ and $F(y_{n_0}) \neq 0$. We can continue this construction inductively and assume to this end that $n_0 < n_1 < \cdots < n_k$ are given. Then there exists $N_{k+1} \geq 0$ such that $\inf_{n\geq N_{k+1}}\langle F(y_n), y - y_n\rangle > -\epsilon_{k+1}$ $(> -\epsilon_0)$. Taking $n_{k+1} > \max\{N_{k+1}, n_k\}$, we have

$$\langle F(y_{n_{k+1}}), y - y_{n_{k+1}}\rangle + \epsilon_{k+1} \geq 0 \quad \text{and} \quad F(y_{n_{k+1}}) \neq 0.$$

In this way, we obtain a strictly increasing sequence $(n_k)_{k\geq 0}$ with the property that

$$\langle F(y_{n_k}), y - y_{n_k}\rangle + \epsilon_k \geq 0 \quad \text{and} \quad F(y_{n_k}) \neq 0, \quad \forall\, k \geq 0. \tag{3.17}$$

Setting for every $k \geq 0$,

$$z_k := \frac{F(y_{n_k})}{\|F(y_{n_k})\|^2},$$

it holds that $\langle F(y_{n_k}), z_k\rangle = 1$. According to (3.17) we have that

$$\begin{aligned}0 \leq \langle F(y_{n_k}), y - y_{n_k}\rangle + \epsilon_k &= \langle F(y_{n_k}), y - y_{n_k}\rangle + \langle F(y_{n_k}), \epsilon_k z_k\rangle \\ &= \langle F(y_{n_k}), y + \epsilon_k z_k - y_{n_k}\rangle, \quad \forall\, k \geq 0,\ \forall\, y \in C.\end{aligned}$$

Since F is pseudo-monotone on $\mathscr{H}$, it yields

$$\langle F(y + \epsilon_k z_k), y + \epsilon_k z_k - y_{n_k}\rangle \geq 0, \quad \forall\, k \geq 0. \tag{3.18}$$

Using that $(F(y_{n_k}))_{n\geq 0}$ is bounded (since $(F(y_n))_{n\geq 0}$ converges weakly to $F(\widehat{x})$), we have

$$\lim_{k\to+\infty}\|\epsilon_k z_k\| = \lim_{k\to+\infty}\frac{\epsilon_k}{\|F(y_{n_k})\|} = 0.$$

Taking the limit in (3.18) as $k \to +\infty$, we obtain

$$\langle F(y), y - \widehat{x} \rangle \geq 0, \quad \forall\, k \geq 0.$$

As y was arbitrarily chosen in C, it follows from Proposition 2.6 that $\widehat{x} \in \Omega$. Hence, by Opial's lemma (see Lemma 2.7), we can conclude that the sequence x_n converges weakly to a point in Ω. □

Remark 3.2. (i) Following the proof of Theorem 3.1, we can demonstrate the first part of Opial's lemma ($\lim_{n\to+\infty} \|x_n - x^*\|$ exists) by using the assumption that F is pseudo-monotone on C (not necessary on $\mathscr{H}$) and $\Omega \neq \emptyset$. Moreover, we also obtain from this part of the proof that $\sum_{n\in\mathbb{N}} \|y_n - y_{n-1}\|^2 < +\infty$ and $\sum_{n\in\mathbb{N}} \|y_n - x_n\|^2 < +\infty$.

(ii) For the second part, the sequentially weak-to-weak continuity and pseudo-monotonicity on $\mathscr{H}$ of F are essential to prove that every weak cluster point of $(x_n)_{n\geq 0}$ belongs to the solution set of $\mathrm{VI}(F, C)$.

(iii) Notice from the proof of Theorem 3.1 that if we suppose D is an open set of $\mathscr{H}$ containing C, it has seen that $y \in C \subset D$ and there exists $\delta > 0$ such that $\mathscr{B}(y, \delta) \subset D$. Since $\epsilon_k z_k \to 0$, then $y + \epsilon_k z_k \to y \in C \subset D$ (as $k \to 0$). Hence, there is $k' > 0$ such that $y + \epsilon_k z_k \in \mathscr{B}(y, \delta) \subset D$, $\forall\, k \geq k'$. Therefore, we can relax the assumption of F as a pseudo-monotone on D. Note that $\mathscr{B}(y, \delta)$ is an open ball with center y and radius δ.

Remark 3.3 (Adaptive stepsize strategy). On the other hand, when (an upper bound of) the Lipschitz constant of F is not available, we can use in our algorithm the following stepsize strategy, see also [5]:

$$\lambda_{n+1} := \begin{cases} \min\left\{\frac{\mu\|y_{n-1} - y_n\|}{\|F(y_{n-1}) - F(y_n)\|}, \lambda_n\right\} & \text{if } F(y_{n-1}) - F(y_n) \neq 0, \\ \lambda_n & \text{otherwise,} \end{cases}$$

where $\mu \in (0, 1/2)$ and $\lambda_0 > 0$. The sequence $(\lambda_n)_{n\geq 0}$ is nonincreasing. If $F(y_{n-1}) - F(y_n) \neq 0$, for $n \geq 0$, then it holds

$$\frac{\mu\|y_{n-1} - y_n\|}{\|F(y_{n-1}) - F(y_n)\|} \geq \frac{\mu\|y_{n-1} - y_n\|}{L\|y_{n-1} - y_n\|} = \frac{\mu}{L},$$

which shows that $(\lambda_n)_{n\geq 0}$ is bounded from below by $\min\{\lambda_0, \mu/L\}$ (this means $\lim_{n\to+\infty} \lambda_n$ exits). Notice that, if $\lambda_0 \leq \mu/L$, then $(\lambda_n)_{n\geq 0}$ is a constant sequence, which leads to a fixed stepsize strategy. Consequently, the $\lim_{n\to+\infty} \lambda_n$ exists and it is a positive real number. We can adapt the proof of Theorem 3.1 to the new adaptive stepsize strategy. On the other hand, from (3.8), we can write

$$\begin{aligned}\|x_{n+1}-x_n\|^2 &= \|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n, F(y_{n-1})-F(y_n)\rangle+\lambda_n^2\|F(y_{n-1})-F(y_n)\|^2\\ &\leq \|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n, F(y_{n-1})-F(y_n)\rangle+\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1}-y_n\|^2.\end{aligned}$$

Then it follows from (3.7) and the above inequality that

$$\begin{aligned}&\|x_{n+1}-x^*\|^2\\ &\leq \|x_n-x^*\|^2+\left[\|y_n-x_n\|^2+2\lambda_n\langle y_n-x_n, F(y_{n-1})-F(y_n)\rangle+\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1}-y_n\|^2\right]\\ &\quad -2\|x_n-y_n\|^2+2\lambda_n\langle x_n-y_n, F(y_{n-1})-F(y_n)\rangle\\ &= \|x_n-x^*\|^2-\|x_n-y_n\|^2+\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1}-y_n\|^2.\end{aligned} \tag{3.19}$$

By parallelogram identity, we know that $\|y_n-y_{n-1}\|^2+\|(x_n-y_n)+(x_n-y_{n-1})\|^2 = 2\|x_n-y_n\|^2+2\|x_n-y_{n-1}\|^2$, then

$$\begin{aligned}\|x_n-y_n\|^2 &\geq -\|x_n-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\\ &= -\|y_{n-1}+\lambda_{n-1}(F(y_{n-2})-F(y_{n-1}))-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\\ &= -\|\lambda_{n-1}(F(y_{n-2})-F(y_{n-1}))\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\\ &\geq -\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2\|y_{n-2}-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2.\end{aligned} \tag{3.20}$$

It follows from (3.19) and (3.20) that

$$\begin{aligned}\|x_{n+1}-x^*\|^2 \leq \|x_n-x^*\|^2 &- \left[-\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2\|y_{n-2}-y_{n-1}\|^2+\frac{1}{2}\|y_n-y_{n-1}\|^2\right]\\ &+\frac{\lambda_n^2\mu^2}{\lambda_{n+1}^2}\|y_{n-1}-y_n\|^2.\end{aligned}$$

Thus

$$\begin{aligned}&\|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right)\|y_n-y_{n-1}\|^2\\ &\leq \|x_n-x^*\|^2+\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2\|y_{n-2}-y_{n-1}\|^2.\end{aligned}$$

Hence

$$\|x_{n+1}-x^*\|^2+\left(\frac{1}{2}-2\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right)\|y_n-y_{n-1}\|^2+\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\|y_n-y_{n-1}\|^2$$
$$\leq \|x_n-x^*\|^2+\left(\frac{\lambda_{n-1}\mu}{\lambda_n}\right)^2\|y_{n-2}-y_{n-1}\|^2.$$

Since $\mu\in(0,1/2)$, then

$$0<\frac{1}{2}-2\mu^2=\lim_{n\to+\infty}\left(\frac{1}{2}-2\left(\frac{\lambda_n\mu}{\lambda_{n+1}}\right)^2\right).$$

This means that there exists $n_{00}\in\mathbb{N}$ such that $\frac{1}{2}-2\big(\frac{\lambda_n\mu}{\lambda_{n+1}}\big)^2>\eta>0$, $\forall\, n>n_{00}$ for some $\eta\in\mathbb{R}$.

The proof of convergence is similar to the proof in Theorem 3.1. By using Lemma 2.4, we have that the sequence $(x_n)_{n\geq 0}$ is bounded and $\sum_{n=2}^{\infty}\big(\frac{1}{2}-2\big(\frac{\lambda_n\mu}{\lambda_{n+1}}\big)^2\big)\|y_n-y_{n-1}\|^2<+\infty$ and moreover $\sum_{n\in\mathbb{N}}\|y_n-y_{n-1}\|^2<+\infty$. Now from (3.19), the fact that $\lim_{n\to+\infty}\lambda_n$ exists and $\sum_{n\in\mathbb{N}}\|y_n-y_{n-1}\|^2<+\infty$, by using Lemma 2.4 again, we obtain that $\sum_{n\in\mathbb{N}}\|y_n-x_n\|^2<+\infty$. Furthermore, we also get that $\lim_{n\to+\infty}\|x^*-x_n\|^2=0$. The rest of the proof is similar to one of Theorem 3.1.

Next, we will show that the convergence result in Theorem 3.1 holds in finite dimensional spaces under a weaker assumption: F is pseudo-monotone only on C ($\subset\mathscr{H}$) instead of $\mathscr{H}$.

Theorem 3.4. *Let $\mathscr{H}$ be a finite-dimensional real Hilbert space. Assume that the solution set Ω is nonempty, F is pseudo-monotone on C and Lipschitz continuous with constant $L>0$, and $0<\liminf_{n\to+\infty}\lambda_n\leq\limsup_{n\to+\infty}\lambda_n<1/(2L)$. Then the sequence $(x_n)_{n\geq 0}$ generated by* (3.1) *converges to a solution of* $\mathrm{VI}(F,C)$.

Proof. Let $x^*\in\Omega$ be fixed. The first part of Opial's lemma follows directly from Remark 3.2(i). Thus we have $\lim_{n\to+\infty}\|x_n-x^*\|$ exists. In addition, we have also that $\sum_{n\in\mathbb{N}}\|y_n-x_n\|^2<+\infty$, hence $\lim_{n\to+\infty}\|y_n-x_n\|=0$ (see also (3.15)). Let us prove the second part of Opial's lemma. Let $\widehat{x}$ be a cluster point of $(x_n)_{n\geq 0}$. Then there exists a subsequence $(x_{n_k})_{k\geq 0}$ of $(x_n)_{n\geq 0}$, which converges to $\widehat{x}$ as $k\to+\infty$. Since $\lim_{n\to+\infty}\|y_n-x_n\|=0$, then $(y_{n_k})_{k\geq 0}$ also converges to $\widehat{x}$ as $k\to+\infty$. Let $y\in C$ be fixed. It follows from (3.3) that

$$\langle y-y_{n_k},y_{n_k}-x_{n_k}+\lambda_{n_k}F(y_{n_k-1})\rangle\geq 0,\quad\forall\, k\geq 0. \tag{3.21}$$

Because the sequence $(\lambda_{n_k})_{k\geq 0}$ is bounded, it has a subsequence which converges to $\widetilde{\lambda}>0$ (since $0<\liminf_{n\to+\infty}\lambda_n$). Taking the limit along this subsequence in (3.21) and using that F is continuous, we obtain

$$\langle y - \widehat{x}, F(\widehat{x}) \rangle \geq 0.$$

Since $y \in C$ was chosen arbitrarily, it follows that $\widehat{x}$ is a solution of VI(F, C). Now we can apply the Opial's Lemma (see Lemma 2.7) to complete the proof. □

4. Numerical experiments for pseudo-monotone variational inequalities

In this part, we consider a numerical experiment which is carried out in order to compare the classical Tseng's algorithm (FBF) and our algorithm (FBF-EP) for solving pseudo-monotone variational inequalities. We implemented the numerical codes in MATLAB and performed all computations on a Windows desktop with an Intel(R) Core(TM) i5-8250U processor at 1.60 GHz up to 1.8 GHz and RAM of 8 GB. In this experiment, we considered variational inequalities governed by a pseudo-monotone operator, which is not monotone.

Example 4.1. We follow the construction of the experiment in [5] and give our example in a higher dimension (10 and 20 dimensions). It will be shown as below.

We consider the VI(F, C) with

$$C = \left\{ \mathbf{x} \in \mathbb{R}^m : \sum_{i=1}^{m} x_i \leq 5, 0 \leq x_i \leq 5, \forall\, i = 1, \ldots, m \right\}$$

and

$$F \colon \mathbb{R}^m \to \mathbb{R}^m, \quad F(\mathbf{x}) = (e^{-\|\mathbf{x}\|^2} + \alpha)(\mathbf{M}\mathbf{x} + \mathbf{p}),$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^m$, $\alpha = 0.1$, $\mathbf{p}$ is a given vector in $\mathbb{R}^m$, e is an exponential function and choose the matrix $\mathbf{M}$ as: For $m = 10$, we pick

$$\mathbf{M}_{10} := \begin{bmatrix} 25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 \\ -5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 \\ 10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 \\ 0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 \\ 10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 \\ 5 & 5 & 1 & -2 & 6 & 5 & 6 & -1 & 1 & 1 \\ 15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 \\ 5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 \\ 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 \\ 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 \end{bmatrix},$$

and for $m = 20$, we choose

$$\mathbf{M}_{20} := \begin{bmatrix}
25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 & 25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 \\
-5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 & -5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 \\
10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 & 10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 \\
0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 & 0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 \\
10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 & 10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 \\
5 & 5 & 1 & -2 & 6 & 5 & 6 & -1 & 1 & 1 & 5 & 5 & 1 & 2 & 6 & 5 & 6 & -1 & 1 & 1 \\
15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 & 15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 \\
5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 & 5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 \\
10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 & 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 \\
0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 & 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 \\
25 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 & 41 & -5 & 10 & 0 & 10 & 5 & 15 & 5 & 10 & 0 \\
-5 & 37 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 & -5 & 46 & -8 & 18 & -2 & 5 & -3 & 11 & -8 & 0 \\
10 & -8 & 14 & -3 & 7 & 1 & 3 & 6 & 14 & 9 & 10 & -8 & 18 & -3 & 7 & 1 & 3 & 6 & 14 & 9 \\
0 & 18 & -3 & 34 & 0 & -2 & 10 & 21 & 2 & 0 & 0 & 18 & -3 & 35 & 0 & -2 & 10 & 21 & 2 & 0 \\
10 & -2 & 7 & 0 & 21 & 6 & 17 & 0 & 7 & 11 & 10 & -2 & 7 & 0 & 25 & 6 & 17 & 0 & 7 & 11 \\
5 & 5 & 1 & -2 & 6 & 5 & 6 & -1 & 1 & 1 & 5 & 5 & 1 & -2 & 6 & 14 & 6 & -1 & 1 & 1 \\
15 & -3 & 3 & 10 & 17 & 6 & 31 & 2 & 7 & 3 & 15 & -3 & 3 & 10 & 17 & 6 & 47 & 2 & 7 & 3 \\
5 & 11 & 6 & 21 & 0 & -1 & 2 & 29 & 15 & 6 & 5 & 11 & 6 & 21 & 0 & -1 & 2 & 54 & 15 & 6 \\
10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 56 & 10 & 10 & -8 & 14 & 2 & 7 & 1 & 7 & 15 & 72 & 10 \\
0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 41 & 0 & 0 & 9 & 0 & 11 & 1 & 3 & 6 & 10 & 50
\end{bmatrix},$$

which are positive definite matrices (the matrices were constructed from the upper triangular matrices in both 10 and 20 dimensions and followed Theorem 8.3.3 in [22]). In general, the operator F is not monotone (see Bianchi et al. [2]). Moreover, we can show that this operator is pseudo-monotone (see Boţ et al. [5]), i.e., for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ such that $\langle F(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0$ and since $g(\mathbf{x}) := e^{-\|\mathbf{x}\|^2} + \alpha \geq 0$ then we have $\langle \mathbf{M}\mathbf{x} + \mathbf{p}, \mathbf{y} - \mathbf{x} \rangle \geq 0$ and thus

$$\begin{aligned}
\langle \mathbf{F}(\mathbf{y}), \mathbf{y} - \mathbf{x} \rangle &= g(\mathbf{y}) \langle \mathbf{M}\mathbf{y} + \mathbf{p}, \mathbf{y} - \mathbf{x} \rangle \geq g(\mathbf{y}) (\langle \mathbf{M}\mathbf{y} + \mathbf{p}, \mathbf{y} - \mathbf{x} \rangle - \langle \mathbf{M}\mathbf{x} + \mathbf{p}, \mathbf{y} - \mathbf{x} \rangle) \\
&= g(\mathbf{y}) (\langle \mathbf{M}(\mathbf{y} - \mathbf{x}) + \mathbf{p}, \mathbf{y} - \mathbf{x} \rangle) \geq 0.
\end{aligned}$$

We computed the unique solution $\mathbf{x}^*$ of the variational inequality $\mathrm{VI}(F, C)$ by running 10000 iterations of Tseng's algorithm for all $n \geq 0$ and stepsize $\lambda_n = 0.49/L$.

In the first trial, we give $\mathbf{p} = \overline{\mathbf{1}}_m$, a vector in $\mathbb{R}^m$ which all elements are equal to one. We compared the performances of the Tseng's algorithm (FBF) and the Tseng's algorithm with extrapolation (FBF-EP) by considering the random initial points

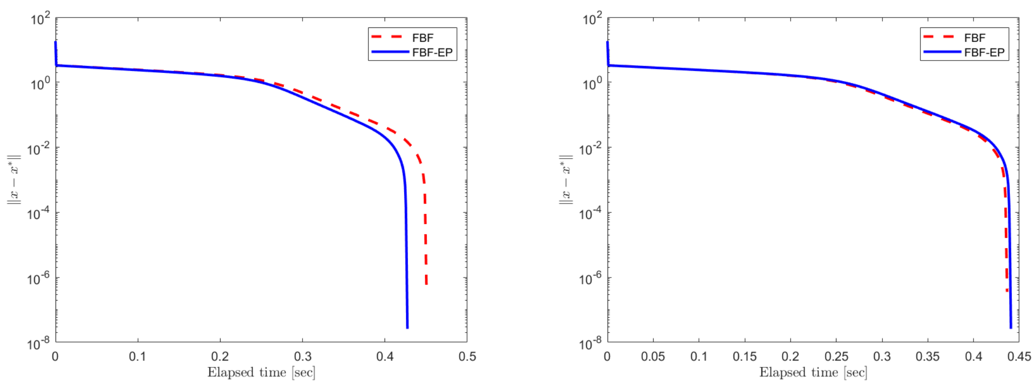
For 10 dimensions:
$\mathbf{x}_0^{10} = (-4, 1, 8, -9, 0, -1, 8, 3, 10, 2)^T$,
$\mathbf{y}_{-1}^{10} = \widehat{\mathbf{y}}_{-1}^{10} := (8, -10, -8, 5, -2, -2, 2, -10, -2, -9)^T$,

For 20 dimensions:

$\mathbf{x}_0^{20} = (-5, 1, 3, -9, 0, -1, 8, -5, 3, -2, -1, 0, 2, -8, 4, 0, -3, -10, 1, 2)^T$,

$\mathbf{y}_{-1}^{20} = \widehat{\mathbf{y}}_{-1}^{20} := (11, -2, 10, 7, -8, 4, -6, 7, -1, 10, -17, 9, 13, -1, 0, 3, 12, -8, 9, 15)^T$,

($\widehat{\mathbf{y}}_{-1}$ denoted to be a fixed vector of $\mathbf{y}_{-1}$) and $\|\mathbf{x}_n - \mathbf{x}^*\| \leq 10^{-6}$ as stopping criterion. The projection on C was computed by using the quadprog function in MATLAB. Figure 4.1, Table 4.1 and Figure 4.2, Table 4.2 show that at the first implementation of this trial for 10 dimensions and 20 dimensions, respectively. We can see that our algorithm (Tseng's algorithm with extrapolation from the past or FBF-EP) spends less time than the classical Tseng's algorithm (FBF), whereas at the second implementation it is not the case. Therefore, one may see that our algorithm is quite sensitive with respect to computer errors or engine computational power.

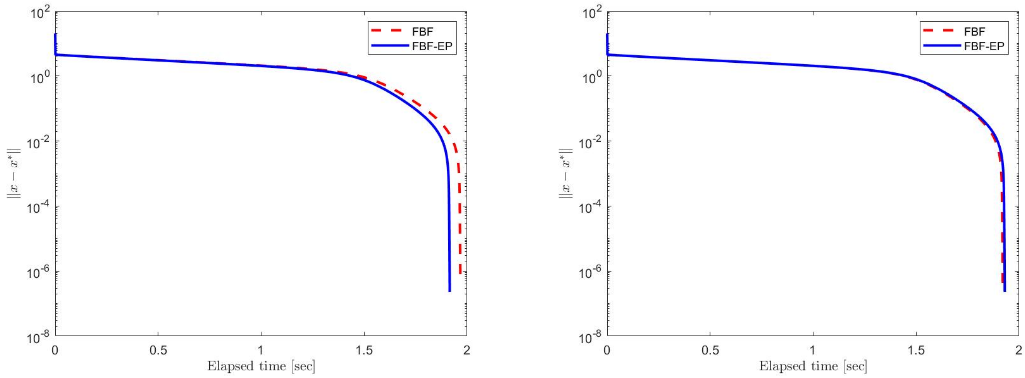


(a) First execution. (b) Second execution.

Figure 4.1: [10 dimensions] A figure of two graphs for comparison between the classical Tseng's algorithm (FBF) and Tseng's algorithm with extrapolation from the past (FBF-EP) for two different executions with $\mathbf{x}_0^{10} = (-4, 1, 8, -9, 0, -1, 8, 3, 10, 2)^T$ and $\mathbf{y}_{-1}^{10} = \widehat{\mathbf{y}}_{-1}^{10}$ (when the stepsize is $\lambda_n = 0.49/L$).

Table 4.1: [10 dimensions] The table of the performances for 2 attempts of FBF and FBF-EP with $\mathbf{x}_0^{10} = (-4, 1, 8, -9, 0, -1, 8, 3, 10, 2)^T$ and $\mathbf{y}_{-1}^{10} = \widehat{\mathbf{y}}_{-1}^{10}$ (when the stepsize is $\lambda_n = 0.49/L$).

Attempt	FBF			FBF-EP		
	$\|\mathbf{x}_n - \mathbf{x}^*\|$	No. iter	CPU-time	$\|\mathbf{x}_n - \mathbf{x}^*\|$	No. iter	CPU-time
1	3.567×10^{-7}	464	0.44943	2.6077×10^{-7}	456	0.42628
2	3.567×10^{-7}	464	0.43565	2.6077×10^{-7}	456	0.43983



(a) First execution. (b) Second execution.

Figure 4.2: [20 dimensions] A figure of two graphs for comparison between the classical Tseng's algorithm (FBF) and Tseng's algorithm with extrapolation from the past (FBF-EP) for two different executions with $\mathbf{x}_0^{20} = (-5, 1, 3, -9, 0, -1, 8, -5, 3, -2, -1, 0, 2, -8, 4, 0, -3, -10, 1, 2)^T$ and $\mathbf{y}_{-1}^{20} = \widehat{\mathbf{y}}_{-1}^{20}$ (when the stepsize is $\lambda_n = 0.49/L$).

Table 4.2: [20 dimensions] The table of the performances for 2 attempts of FBF and FBF-EP with $\mathbf{x}_0^{20} = (-5, 1, 3, -9, 0, -1, 8, -5, 3, -2, -1, 0, 2, -8, 4, 0, -3, -10, 1, 2)^T$ and $\mathbf{y}_{-1}^{20} = \widehat{\mathbf{y}}_{-1}^{20}$ (when the stepsize is $\lambda_n = 0.49/L$).

Attempt	FBF			FBF-EP		
	$\|\mathbf{x}_n - \mathbf{x}^*\|$	No. iter	CPU-time	$\|\mathbf{x}_n - \mathbf{x}^*\|$	No. iter	CPU-time
1	4.0665×10^{-7}	1976	1.9663	2.295×10^{-7}	1973	1.9147
2	4.0665×10^{-7}	1976	1.9194	2.295×10^{-7}	1973	1.9293

In order to confirm the effectiveness of our method, we designed the experiment as follows. We randomise an initial vector $\mathbf{x}_0$ in which each coordinate is an integer shuffled from the interval $[-10, 10]$, denoted by $\text{rand}_{[-10,10]}$ (in MATLAB). We also select $\mathbf{y}_{-1}$ to be $\mathbf{x}_0$, $\widehat{\mathbf{y}}_{-1}$ and $\text{rand}_{[-10,10]}$, and put vector $\mathbf{p}$ as following vectors:

For 10 dimensions:
$\mathbf{p}_1^{10} = \overline{\mathbf{1}}_{10}, \quad \mathbf{p}_2^{10} = (1, 2, 1, 0, -1, 2, 0, 1, -1, 2)^T,$
$\mathbf{p}_3^{10} = (5, 1, 0, -2, -1, 0, -5, 4, -5, -1)^T,$
For 20 dimensions:
$\mathbf{p}_1^{20} = \overline{\mathbf{1}}_{20}, \quad \mathbf{p}_2^{20} = (1, -2, 1, 2, -1, 2, 0, 1, -1, 2, 1, -2, 1, 0, -1, 2, 0, 1, -1, 2)^T,$
$\mathbf{p}_3^{20} = (5, 1, 0, -2, -1, 0, -5, 4, -5, -1, 5, 1, 0, -2, -1, 0, -5, 4, -5, -1)^T.$

Table 4.3: The table shows the result of satisfaction after 100 executions in 10 dimensions by choosing $\mathbf{x}_0 = \text{rand}_{[-10,10]}$ (for 10 dimensions), $\mathbf{y}_{-1} = \mathbf{x}_0^{10}$, $\widehat{\mathbf{y}}_{-1}^{10}$, $\text{rand}_{[-10,10]}$ (for 10 dimensions) and $\mathbf{p} = \mathbf{p}_1^{10}$, $\mathbf{p}_2^{10}$ and $\mathbf{p}_2^{10}$ (when the stepsize is $\lambda_n = 0.49/L$).

$\mathbf{x}_0$	$\mathbf{y}_{-1}$	vector$\mathbf{p}$	satisfaction[1]
$\text{rand}_{[-10,10]}$	$\mathbf{x}_0^{10}$		69 out of 100
$\text{rand}_{[-10,10]}$	$\widehat{\mathbf{y}}_{-1}^{10}$	$\mathbf{p}_1^{10}$	65 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		66 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x}_0^{10}$		66 out of 100
$\text{rand}_{[-10,10]}$	$\widehat{\mathbf{y}}_{-1}^{10}$	$\mathbf{p}_2^{10}$	70 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		57 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x}_0^{10}$		61 out of 100
$\text{rand}_{[-10,10]}$	$\widehat{\mathbf{y}}_{-1}^{10}$	$\mathbf{p}_3^{10}$	60 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		60 out of 100

[1] The number of executions in which the CPU time for Tseng-EP algorithm is less than the Tseng algorithm, within 100 executions.

Table 4.4: The table shows the result of satisfaction after 100 executions in 20 dimensions by choosing $\mathbf{x}_0 = \text{rand}_{[-10,10]}$ (for 20 dimensions), $\mathbf{y}_{-1} = \mathbf{x}_0^{20}$, $\widehat{\mathbf{y}}_{-1}^{20}$ and $\text{rand}_{[-10,10]}$ (for 20 dimensions), and $\mathbf{p} = \mathbf{p}_1^{20}$, $\mathbf{p}_2^{20}$, $\mathbf{p}_3^{20}$ (when the stepsize is $\lambda_n = 0.49/L$).

$\mathbf{x}_0$	$\mathbf{y}_{-1}$	vector$\mathbf{p}$	satisfaction[1]
$\text{rand}_{[-10,10]}$	$\mathbf{x}_0^{20}$		62 out of 100
$\text{rand}_{[-10,10]}$	$\widehat{\mathbf{y}}_{-1}^{20}$	$\mathbf{p}_1^{20}$	61 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		56 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x}_0^{20}$		51 out of 100
$\text{rand}_{[-10,10]}$	$\widehat{\mathbf{y}}_{-1}^{20}$	$\mathbf{p}_2^{20}$	53 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		53 out of 100
$\text{rand}_{[-10,10]}$	$\mathbf{x}_0^{20}$		58 out of 100
$\text{rand}_{[-10,10]}$	$\widehat{\mathbf{y}}_{-1}^{20}$	$\mathbf{p}_3^{20}$	55 out of 100
$\text{rand}_{[-10,10]}$	$\text{rand}_{[-10,10]}$		51 out of 100

[1] The number of executions in which the CPU time for Tseng-EP algorithm is less than the Tseng algorithm, within 100 executions.

Then, we run 100 executions for FBF and FBF-EP algorithms. Finally, we collect the achievements when FBF-EP takes less time than FBF and call this value "satisfaction". As a result, we could deduce from Table 4.3 (and Table 4.4) that the FBF-EP algorithm overcomes the FBF algorithm by more than half of 100 executions. In 10 dimensions, the highest satisfaction value is 70 (62 in 20 dimensions), and the lowest is 57 (51 in 20 dimensions) for this experiment.

As the referee suggested, it would be better to show another example in Hilbert space, which is not Euclidean space. Then, we consider our proposed algorithm for solving the variational inequality problem in L^2-space as below.

Example 4.2. Let $\mathscr{H} = L^2([0,1])$ with norm $\|x\| := \left(\int_0^1 |x(t)|^2\,dt\right)^{1/2}$ and inner product $\langle x, y\rangle := \int_0^1 x(t)y(t)\,dt$, $\forall\, x, y \in \mathscr{H}$. Define an operator $F\colon \mathscr{H} \to \mathscr{H}$ by

$$(Fx)(t) = \max\{x(t), 0\}, \quad x \in \mathscr{H},\ t \in [0,1].$$

It is easy to show that F is monotone (pseudo-monotone) and Lipschitz continuous with Lipschitz constant $L = 1$. We provide the feasible set as the ball $C := \{x \in \mathscr{H} : \|x\| \le 2\}$. To implement, we consider either the FBF or FBF-EP methods, including their adaptive stepsize strategy. We represent FBF and FBF-EP with adaptive stepsize strategy by aFBF and aFBF-EP, respectively, and the stopping criterion is $\|x_n - x_{n-1}\| \le \epsilon$ with $\epsilon = 10^5$. Note that the aFBF algorithm is obtained from [5]. The parameter values of the algorithm in Example 4.2 is determined as follows.

Case I.

FBF : $\mathbf{x}_0 = t^3,\ \lambda_n = 0.49$;

FBF-EP (our proposed algorithm) : $\mathbf{x}_0 = \mathbf{y}_{-1} = t^3,\ \lambda_n = 0.49$;

aFBF [5] : $\mathbf{x}_0 = t^3,\ \mu = 0.49,\ \lambda_0 = 0.7$;

aFBF-EP (our proposed algorithm) : $\mathbf{x}_0 = \mathbf{y}_{-1} = t^3,\ \mu = 0.49,\ \lambda_0 = 0.7$.

Case II.

FBF : $\mathbf{x}_0 = \dfrac{\cos(-3t) + \sin(-10)}{200},\ \lambda_n = 0.49$;

FBF-EP (our proposed algorithm) : $\mathbf{x}_0 = \mathbf{y}_{-1} = \dfrac{\cos(-3t) + \sin(-10)}{200},\ \lambda_n = 0.49$;

aFBF [5] : $\mathbf{x}_0 = \dfrac{\cos(-3t) + \sin(-10)}{200},\ \mu = 0.49,\ \lambda_0 = 0.7$;

aFBF-EP (our proposed algorithm) : $\mathbf{x}_0 = \mathbf{y}_{-1} = \dfrac{\cos(-3t) + \sin(-10)}{200},\ \mu = 0.49,$

$$\lambda_0 = 0.7.$$

Case III.

FBF : $\mathbf{x}_0 = (10t^3 - 3t^2)/20,\ \lambda_n = 0.49;$

FBF-EP (our proposed algorithm) : $\mathbf{x}_0 = \mathbf{y}_{-1} = (10t^3 - 3t^2)/20,\ \lambda_n = 0.49;$

aFBF [5] : $\mathbf{x}_0 = (10t^3 - 3t^2)/20,\ \mu = 0.49,\ \lambda_0 = 0.7;$

aFBF-EP (our proposed algorithm) : $\mathbf{x}_0 = \mathbf{y}_{-1} = (10t^3 - 3t^2)/20,\ \mu = 0.49,\ \lambda_0 = 0.7.$

Case IV. In this case, we use the same values of starting points $\mathbf{x}_0$, $\mathbf{y}_{-1}$, μ and λ_n in the methods without adaptive stepsize strategy as in Case III, and we replace instead $\lambda_0 = 0.9$ in the adaptive stepsize scheme.

Table 4.5: The result of computation in Example 4.2.

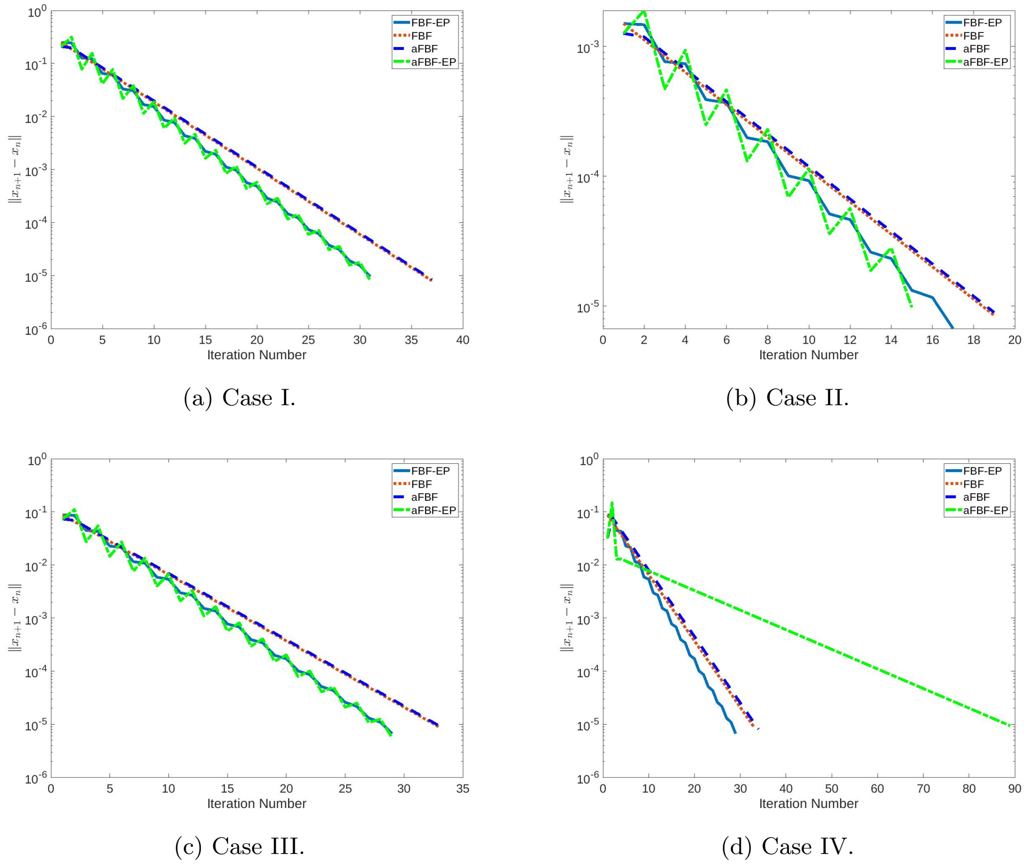
Algorithms	Case I			Case II		
	$\Vert\mathbf{x}_{n+1} - \mathbf{x}_n\Vert$	No. iter	CPU-time	$\Vert\mathbf{x}_{n+1} - \mathbf{x}_n\Vert$	No. iter	CPU-time
FBF	7.980980×10^{-6}	38	15.9776	8.472908×10^{-6}	20	79.437
FBF-EP*	9.608509×10^{-6}	32	7.9064	6.706328×10^{-6}	18	53.6874
aFBF	8.405511×10^{-6}	38	18.2491	8.923607×10^{-6}	20	85.596
aFBF-EP**	8.08286×10^{-6}	32	9.7161	9.771265×10^{-6}	16	47.0231

Algorithms	Case III			Case IV		
	$\Vert\mathbf{x}_{n+1} - \mathbf{x}_n\Vert$	No. iter	CPU-time	$\Vert\mathbf{x}_{n+1} - \mathbf{x}_n\Vert$	No. iter	CPU-time
FBF	8.823636×10^{-6}	34	14.9177	8.823636×10^{-7}	34	14.7945
FBF-EP*	6.640949×10^{-6}	30	7.59	6.640949×10^{-6}	30	8.0234
aFBF	9.292991×10^{-6}	34	16.8252	8.029509×10^{-6}	35	17.3107
aFBF-EP**	5.520665×10^{-6}	30	9.4287	9.317460×10^{-6}	90	28.3885

* Our proposed algorithm.

** Our proposed algorithm with adaptive stepsize strategy.

Table 4.5 and Figure 4.3 depict the results of the experiment by showing $\|x_{n+1} - x_n\|$, the number of iteration (No. iter), CPU-time (second) in the table, and plotting between $\|x_{n+1} - x_n\|$ and their iteration numbers in different cases. One can notice that our proposed algorithm, FBF-EP, is always faster than FBF method, which sometimes consumes twice the CPU-time more significantly than the CPU-time of FBF-EP. Further, the number of iterations of FBF-EP is less than the iteration numbers of FBF. For this numerical experiment, we know that the Lipschitz constant L equals to 1 ($L = 1$). Therefore, we have to choose $\lambda_n < 1/(2L)$ ($= 1/2$, because $L = 1$) for the FBF-EP method (see Theorem 3.1). Nevertheless, the constant L might not figure out in certain cases. Then, we can use the adaptive stepsize strategy described in Remark 3.3 with $\mu = 0.49 < 1/2$



(a) Case I.

(b) Case II.

(c) Case III.

(d) Case IV.

Figure 4.3: The graphs plot the value between $\|x_{n+1} - x_n\|$ and the iteration number of the experiments in Cases I, II, III and IV.

and $\lambda_0 > 0$. We have seen that the starting stepsize λ_0 have a significant effect in both aFBF-EP and aFBF. If the stepsize is a suitable one, it can reduce the CPU-time and turn the method to the fastest one, for example, the FBF-EP in Case II; however, since the time of the process to finding λ_n in the next iteration of the adaptive stepsize strategy has been counted in computation then the method with adaptive stepsize strategy can give the time performance, which is slower than the normal method without the adaptive stepsize strategy. In addition, the adaptive stepsize strategy can provide a small stepsize, which then impacts the number of iterations of the iterative method shown in Case IV.

Acknowledgments

This work was supported by the Development and Promotion of Science and Technology Talents Project (DPST) scholarship, Royal Government of Thailand scholarship. The author are very thankful to Dr. habil. Ernö Robert Csetnek for his careful guidance. The author wants to thank the guidance in programming from Dr. Phan Tu Vuong and the helpful comments from the referees.

References

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

[2] M. Bianchi, N. Hadjisavvas and S. Schaible, *On pseudomonotone maps T for which $-T$ is also pseudomonotone*, J. Convex Anal. **10** (2003), no. 1, 149–168.

[3] A. Böhm, M. Sedlmayer, E. R. Csetnek and R. I. Boţ, *Two steps at a time—Taking GAN training in stride with Tseng's method*, SIAM J. Math. Data Sci. **4** (2022), no. 2, 750–771.

[4] R. I. Boţ and E. R. Csetnek, *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algorithms **71** (2016), no. 3, 519–540.

[5] R. I. Boţ, E. R. Csetnek and P. T. Vuong, *The forward-backward-forward method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces*, European J. Oper. Res. **287** (2020), no. 1, 49–60.

[6] R. I. Boţ, M. Sedlmayer and P. T. Vuong, *A relaxed inertial forward-backward-forward algorithm for solving monotone inclusions with application to GANs*, J. Mach. Learn. Res. **24** (2023), no. 8, 1–37.

[7] L. M. Briceño-Arias and P. L. Combettes, *A monotone skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim. **21** (2011), no. 4, 1230–1250.

[8] Y. Censor, A. Gibali and S. Reich, *The subgradient extragradient method for solving variational inequalities in Hilbert space*, J. Optim. Theory Appl. **148** (2011), no. 2, 318–335.

[9] R. W. Cottle and J. C. Yao, *Pseudo-monotone complementarity problems in Hilbert space*, J. Optim. Theory Appl. **75** (1992), no. 2, 281–295.

[10] C. Daskalakis, A. Ilyas, V. Syrgkanis and H. Zeng, *Training GANs with optimism*, arXiv:1711.00141.

[11] C. Daskalakis and I. Panageas, *The limit points of (optimistic) gradient descent in min-max optimization*, Adv. Neural Inf. Process. Syst. **31** (2018), 1–11.

[12] F. Facchinei and J.-S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems II*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2003.

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

[14] P. Hartman and G. Stampacchia, *On some non-linear elliptic differential-functional equations*, Acta Math. **115** (1966), 271–310.

[15] A. N. Iusem and B. F. Svaiter, *A variant of Korpelevich's method for variational inequalities with a new search strategy*, Optimization **42** (1997), no. 4, 309–321.

[16] S. Karamardian and S. Schaible, *Seven kinds of monotone maps*, J. Optim. Theory Appl. **66** (1990), no. 1, 37–46.

[17] P. D. Khanh and P. T. Vuong, *Modified projection method for strongly pseudomonotone variational inequalities*, J. Global Optim. **58** (2014), no. 2, 341–350.

[18] E. N. Khobotov, *Modification of the extra-gradient method for solving variational inequalities and certain optimization problems*, USSR Comput. Math. Math. Phys. **27** (1987), no. 5, 120–127.

[19] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Classics Appl. Math. **31**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000.

[20] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Matekon **12** (1976), 747–756.

[21] Y. Malitsky and M. K. Tam, *A forward-backward splitting method for monotone inclusions without cocoercivity*, SIAM J. Optim. **30** (2020), no. 2, 1451–1472.

[22] W. K. Nicholson, *Linear Algebra with Applications*, Lyryx: Create Space, 2019.

[23] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc. **73** (1967), 591–597.

[24] L. D. Popov, *A modification of the Arrow–Hurwicz method for search of saddle points*, Mathematical Notes of the Academy of Sciences of the USSR **28** (1980), no. 5, 845–848.

[25] A. Rakhlin and K. Sridharan, *Online learning with predictable sequences*, PMLR 2013: Proceedings of the 26th Annual Conference on Learning Theory **30** (2013), 993–1019.

[26] ———, *Optimization, learning, and games with predictable sequences*, NIPS 2013: Proceedings of the 26th International Conference on Neural Information Processing Systems **2** (2013), 3066–3074.

[27] B. Tongnoi, *Tseng's algorithm with extrapolation from the past endowed with variable metrics and error terms*, Numer. Funct. Anal. Optim. **44** (2023), no. 2, 87–123.

[28] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim. **38** (2000), no. 2, 431–446.

[29] D. L. Zhu and P. Marcotte, *Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities*, SIAM J. Optim. **6** (1996), no. 3, 714–726.

Buris Tongnoi
Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, Vienna 1090, Austria
E-mail address: `buris.tongnoi@univie.ac.at`