THE FORWARD-BACKWARD-FORWARD METHOD FROM CONTINUOUS AND DISCRETE PERSPECTIVE FOR PSEUDO-MONOTONE VARIATIONAL INEQUALITIES IN HILBERT SPACES

R. I. BOŢ*, E. R. CSETNEK†, AND P. T. VUONG‡

Abstract. Tseng's forward-backward-forward algorithm is a valuable alternative for Korpelevich's extragradient method when solving variational inequalities over a convex and closed set governed by monotone and Lipschitz continuous operators, as it requires in every step only one projection operation. However, it is well-known that Korpelevich's method converges and can therefore be used also for solving variational inequalities governed by pseudo-monotone and Lipschitz continuous operators. In this paper, we first associate to a pseudo-monotone variational inequality a forward-backward-forward dynamical system and carry out an asymptotic analysis for the generated trajectories. The explicit time discretization of this system results into Tseng's forward-backward-forward algorithm with relaxation parameters, which we prove to converge also when it is applied to pseudo-monotone variational inequalities. In addition, we show that linear convergence is guaranteed under strong pseudo-monotonicity. Numerical experiments are carried out for pseudo-monotone variational inequalities over polyhedral sets and fractional programming problems.

Key words. convex programming, variational inequalities, pseudo-monotonicity, dynamical system, Tseng's FBF algorithm

AMS subject classifications. 47J20, 90C25, 90C30, 90C52

1. Introduction and preliminaries. In this paper, the object of our investigation is the following variational inequality of *Stampacchia type*:

Find $x^* \in C$ such that

$$(1) \qquad \langle F(x^*), x - x^* \rangle \geq 0 \quad \forall x \in C,$$

where C is a nonempty, convex and closed subset of the real Hilbert space H, endowed with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$, and $F : H \to H$ is a Lipschitz continuous operator. We abbreviate the problem (1) as $\mathrm{VI}(F, C)$ and denote its solution set by Ω.

Variational inequalities (VIs) are powerful mathematical models which unify important concepts in applied mathematics, like systems of nonlinear equations, optimality conditions for optimization problems, complementarity problems, obstacle problems, and network equilibrium problems (see, for instance, [14, 19]). In the last decades, various solution methods for solving problems of type $\mathrm{VI}(F, C)$ have been proposed (see [14, 19]). These methods typically require certain monotonicity properties for the operator F (see [17]).

The most popular algorithm for solving variational inequalities is the so-called projected-gradient method, which generates, for a starting point $x_0 \in H$, a sequence

that approaches the solution set Ω by

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

where P_C is the projection operator onto the convex and closed set C and λ is a positive stepsize. It is known that the sequence $(x_n)_{n \geq 0}$ converges, if F is cococercive (inverse strongly monotone) ([3, 34]) or F is strongly (pseudo-) monotone ([14, 20]). The projected-gradient method with variable step sizes was proved to convergence also for variational inequalities governed by (not necessarily single-valued) maximally monotone and paramonotone operators ([4]). If F is "only" monotone, then $(x_n)_{n \geq 0}$ does not necessarily convergence (see [14] for an example). Very recently, Malitsky [23] introduced a modification of the projected-gradient method, called projected-reflected-gradient method, which, for a starting point $x_0 \in H$, reads

$$x_{n+1} = P_C(x_n - \lambda F(2x_n - x_{n-1})) \qquad \forall n \geq 0.$$

The sequence $(x_n)_{n \geq 0}$ is shown to converge to an element in Ω, if F is monotone. Further extensions of this method can be found in [24, 25].

The mostly used algorithm in the literature to solve variational inequalities governed by Lipschitz continuous and pseudo-monotone operators is Korpelevich's extragradient method (see [21]) or variants of it. All these methods share the feature to perform two projections per iteration. Korpelevich's extragradient method generates, for a starting point $x_0 \in H$, a sequence $(x_n)_{n \geq 0}$ approaching the solution set Ω as follows

$$\begin{cases} y_n = P_C(x_n - \lambda F(x_n)) \\ x_{n+1} = P_C(x_n - \lambda F(y_n)) \end{cases} \qquad \forall n \geq 0.$$

This algorithm was originally introduced for solving monotone VIs in finite dimensional spaces, however, it was shown in [14, Theorem 12.2.11] that it converges even when F is a pseudo-monotone operator. In the last years, the extragradient method has attracted a lot of attention from the research community (see, for instance, [8, 10, 14, 16, 29, 30, 33]). In infinite dimensional spaces, Ceng, Teboulle and Yao proved in [8] that, if F is additionally *sequentially weak-to-strong continuous* (which is however not satisfied by the identity operator), then the sequence $(x_n)_{n \geq 0}$ converges weakly to an element in Ω. It was recently proved in [33] that this statement remains true even if the operator F is *sequentially weak-to-weak continuous*.

A challenging task when designing efficient algorithms for solving variational inequalities is to keep the number of projection operations performed at each iteration as low as possible. Projection operations may be very expensive, in particular when for these no closed formulas are available. Censor, Gibali and Reich proposed in [9, 10], for a starting point $x_0 \in H$, the following numerical scheme, called subgradient-extragradient method

$$\begin{cases} y_n = P_C(x_n - \lambda F(x_n)) \\ x_{n+1} = P_{T_n}(x_n - \lambda F(y_n)) \end{cases} \qquad \forall n \geq 0,$$

where

$$T_n = \{w \in H : \langle x_n - \lambda F(x_n) - y_n, w - y_n \rangle \leq 0\}.$$

The projection onto the half-space T_n can be explicitly given (see, for instance, [3]), thus, the subgradient-extragradient method requires the computation of only one projection per iteration and outperforms from this point of view the extragradient

method. The subgradient-extragradient method converges for monotone VIs (see [10]), but also for pseudo-monotone VIs (see [9, 31]).

In this paper, we first attach to VI(F, C) a dynamical system of forward-backward-forward-type (see (4)) and carry out a convergence analysis for the generated trajectories to an element in Ω, in the case when F is a pseudo-monotone operator. If F is assumed to be strongly pseudo-monotone, we prove that the trajectory converges exponentially to the unique solution of VI(F, C). Dynamical systems of forward-backward-forward type were first studied in [2] in the context of approaching the set of the zeros of the sum of a maximally monotone operator and a monotone and Lipschitz continuous operator by continuous trajectories.

The explicit time discretization of (4) leads to Tseng's forward-backward-forward algorithm with relaxation parameters ([32]). When applied to the solving of monotone operators, this algorithm, which requires the computation of only one projection per iteration, is known to generate a sequence, which weakly converges to a solution of VI(F, C). In this paper we show that this convergence result remains true even if F is a pseudo-monotone and *sequentially weak-to-weak-continuous* operator, for both an underrelaxed and an overrelaxed variant of Tseng's algorithm, and provide examples of operators that fulfill the assumptions of the convergence theorems. We also prove that the convergence statement remains true in finite dimensional spaces under less restrictive assumptions on F. In addition, we propose an adaptive stepsize strategy, which does not require the knowledge of the Lipschitz constant of the governing operator. This shows that Tseng's algorithm is a method to be considered when solving constrained pseudo-convex differentiable optimization problems. We also show that linear convergence is guaranteed when the pseudo-monotonicity for F is replaced by strong pseudo-monotonicity. In the last section we carry out numerical experiments which show that, when applied to pseudo-monotone variational inequalities over polyhedral sets and to fractional programming problems, Tseng's method outperforms Korpelevich's extragradient method and even the subgradient-extragradient method.

We want to notice that a single projection method of Halpern-type for pseudo-monotone variational inequalities in Hilbert spaces, which consequently generates a strongly convergent sequence to a solution, has been recently provided in [28].

We close this section by recalling some notions and results which will be useful within this paper.

DEFINITION 1.1. *Let C be a nonempty subset of the real Hilbert space H. The mapping $F : H \to H$ is said to be*
(a) *pseudo-monotone on C, if for every $x, y \in C$ it holds*

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

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

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

(c) *γ-strongly pseudo-monotone on C with $\gamma > 0$, if for every $x, y \in C$ it holds*

$$\langle F(x), y - x \rangle \geq 0 \ \Rightarrow \ \langle F(y), y - x \rangle \geq \gamma \|x - y\|^2;$$

(d) *γ-strongly monotone on C with $\gamma > 0$, if for every $x, y \in C$ it holds*

$$\langle F(y) - F(x), y - x \rangle \geq \gamma \|x - y\|^2.$$

3

For a survey on pseudo-monotone operators and their applications in consumer theory of mathematical economics we refer to [15].

We recall that the operator $F : H \to H$ is called *Lipschitz continuous with Lipschitz constant $L > 0$*, if for every $x, y \in H$ it holds

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

The operator F is called *sequential weak-to-weak continuous*, if for every sequence $(x_n)_{n \geq 0}$ that converges weakly to x the sequence $(F(x_n))_{n \geq 0}$ converges weakly to $F(x)$.

For a nonempty, convex and closed set $C \subseteq H$ and an arbitrary element $x \in H$, there exists a unique element in C, denoted by $P_C(x)$, such that

$$\|x - P_C(x)\| \leq \|x - y\| \quad \forall y \in C.$$

The operator $P_C : H \to C$ is the *projection operator* onto C. For all $x \in H$ and $y \in C$ it holds

(2)
$$\langle x - P_C(x), y - P_C(x) \rangle \leq 0.$$

One can also easily see that, for $\lambda > 0$, x^* is a solution of $\mathrm{VI}(F, C)$ if and only if $x^* = P_C(x^* - \lambda F(x^*))$. We recall the following characterization of the solution set of a pseudo-monotone variational inequality ([12, Lemma 2.1]).

PROPOSITION 1.1. *Let C be a nonempty, convex and closed subset of the real Hilbert space H and $F : H \to H$ an operator which is pseudo-monotone on C and continuous. Then for every $x \in C$ we have*

(3)
$$\langle F(x), y - x \rangle \geq 0 \quad \forall y \in C \iff \langle F(y), y - x \rangle \geq 0 \quad \forall y \in C.$$

The variational inequality:
Find $x^* \in C$ such that

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

is called of *Minty type*. Proposition 1.1 shows that the two variationaly inequalities have the same set of solutions when they are formulated over a nonempty, convex and closed set and governed by pseudo-monotone and continuous operators. Existence results for solutions of variational inequalities have been obtained for instance in [19, 22, 27].

2. A dynamical system of forward-backward-forward type. In this section we will approach the solution set of $\mathrm{VI}(F, C)$ from a continuous perspective by means of trajectories generated by the following dynamical system of forward-backward-forward type

(4)
$$\begin{cases} y(t) = P_C(x(t) - \lambda F(x(t))) \\ \dot{x}(t) + x(t) = y(t) + \lambda\left[F(x(t)) - F(y(t))\right] \\ x(0) = x_0, \end{cases}$$

where $\lambda > 0$ and $x_0 \in H$. The formulation of (4) has its roots in [2], where the continuous counterpart of Tseng's algorithm has been considered in the more general context of a monotone inclusion problem. The existence and uniqueness of the

trajectory $x \in \mathcal{C}^1([0, +\infty), H)$ generated by (4) has been established in [2], as a consequence of the global Cauchy-Lipschitz Theorem and by making use of the Lipschitz continuity of F. Here we study the convergence of $x(t)$ and $y(t)$ to an element in Ω as $t \to +\infty$, in the case when F is pseudo-monotone.

REMARK 2.1. The explicit time discretization of the dynamical system (4) with step size $\rho_n > 0$ and initial point $x_0 \in H$ yields for every $n \geq 0$ the following equation

$$\frac{x_{n+1} - x_n}{\rho_n} + x_n = P_C(x_n - \lambda F(x_n)) + \lambda F(x_n) - \lambda F[P_C(x_n - \lambda F(x_n))].$$

Denoting $y_n := P_C(x_n - \lambda F(x_n))$, we can rewrite this scheme as

(5) $\qquad \begin{cases} y_n = P_C(x_n - \lambda F(x_n)) \\ x_{n+1} = \rho_n \left(y_n + \lambda(F(x_n) - F(y_n)) \right) + (1 - \rho_n)x_n \end{cases} \qquad \forall n \geq 0,$

which is precisely Tseng's forward-backward-forward algorithm with relaxation parameters $(\rho_n)_{n \geq 0}$. In the case $\rho_n = 1$ for every $n \geq 0$, this iterative scheme reduces to the classical forward-backward-forward algorithm as it was introduced in [32]. In Section 3 we prove the convergence of the algorithm in (5).

In the following we will investigate the asymptotic behaviour of the trajectory generated by the dynamical system (4). To this end we will use the following two results. The first one (see [1, Lemma 5.2]) is the continuous counterpart of a result which states the convergence of quasi-Fejér monotone sequences. The second one (see [1, Lemma 5.3]) is the continuous version of the Opial Lemma.

LEMMA 2.1. If $1 \leq p < \infty, 1 \leq r < \infty, A : [0, +\infty) \to [0, +\infty)$ is locally absolutely continuous, $A \in L^p([0, +\infty))$, $B : [0, +\infty) \to \mathbb{R}, B \in L^r([0, +\infty))$ and for almost every $t \in [0, +\infty)$

$$\frac{d}{dt} A(t) \leq B(t),$$

then $\lim_{t \to +\infty} A(t) = 0$.

LEMMA 2.2. Let $\Omega \subseteq H$ be a nonempty set and $x : [0, +\infty) \to H$ a given map. Assume that

(i) for every $x^* \in \Omega$ the limit $\lim_{t \to +\infty} \|x(t) - x^*\|$ exists;

(ii) every weak sequential cluster point of the map x belongs to Ω.

Then there exists an element $x^\infty \in \Omega$ such that $x(t)$ converges weakly to x^∞ as $t \to +\infty$.

We start our asymptotic analysis with two preliminary results.

PROPOSITION 2.1. Assume that the solution set Ω is nonempty, F is pseudo-monotone on C and Lipschitz continuous with constant $L > 0$. Then for every solution $x^* \in \Omega$ it holds

$$\langle \dot{x}(t), x(t) - x^* \rangle \leq -(1 - \lambda L) \|x(t) - y(t)\|^2 \leq 0 \quad \forall t \in [0, +\infty).$$

Proof. Since $x^* \in \Omega$ and $y(t) \in C$ it holds

$$\langle F(x^*), y(t) - x^* \rangle \geq 0 \quad \forall t \in [0, +\infty).$$

By the pseudo-monotonicity of F it holds

(6) $\qquad \langle F(y(t)), y(t) - x^* \rangle \geq 0 \quad \forall t \in [0, +\infty).$

5

On the other hand, since $y(t) = P_C(x(t) - \lambda F(x(t)))$, we obtain from (2) that

(7) $\qquad\qquad \langle x(t) - \lambda F(x(t)) - y(t), y(t) - x^* \rangle \geq 0 \quad \forall t \in [0, +\infty).$

Combining (6) and (7) we obtain for every $t \in [0, +\infty)$

$$\langle x(t) - y(t) - \lambda \left[F(x(t)) - F(y(t)) \right], y(t) - x^* \rangle \geq 0$$

or, equivalently, by taking into account the formulation of the dynamical system (4)

$$\langle x(t) - y(t) - \lambda \left[F(x(t)) - F(y(t)) \right], y(t) - x(t) \rangle - \langle \dot{x}(t), x(t) - x^* \rangle \geq 0.$$

This implies that

$$\begin{aligned}
\langle \dot{x}(t), x(t) - x^* \rangle &\leq \langle x(t) - y(t) - \lambda \left[F(x(t)) - F(y(t)) \right], y(t) - x(t) \rangle \\
&= -\|x(t) - y(t)\|^2 + \lambda \langle F(x(t)) - F(y(t)), x(t) - y(t) \rangle \\
&\leq -(1 - \lambda L) \|x(t) - y(t)\|^2 \quad \forall t \in [0, +\infty). \qquad \square
\end{aligned}$$

PROPOSITION 2.2. *Assume that the solution set Ω is nonempty, F is pseudo-monotone on C and Lipschitz continuous with constant $L > 0$, and $0 < \lambda < \frac{1}{L}$. Then, for every solution $x^* \in \Omega$, the function $t \to \|x(t) - x^*\|^2$ is nonincreasing and it holds*

$$\int_0^{+\infty} \|x(t) - y(t)\|^2 dt < +\infty \text{ and } \lim_{t \to +\infty} \|x(t) - y(t)\| = 0.$$

Proof. Using Proposition 2.1, for every $t \in [0, +\infty)$ it holds

$$\frac{1}{2} \frac{d}{dt} \|x(t) - x^*\|^2 = \langle x(t) - x^*, \dot{x}(t) \rangle \leq -(1 - \lambda L) \|x(t) - y(t)\|^2 \leq 0,$$

which shows that $t \to \|x(t) - x^*\|^2$ is nonincreasing. Let be $T > 0$. Integrating the previous inequality from 0 to T it yields

$$(1 - \lambda L) \int_0^T \|x(t) - y(t)\|^2 dt \leq \frac{1}{2} \left(\|x(0) - x^*\|^2 - \|x(T) - x^*\|^2 \right) \leq \frac{1}{2} \|x(0) - x^*\|^2.$$

Letting $T \to +\infty$, it follows that $\int_0^{+\infty} \|x(t) - y(t)\|^2 dt < +\infty$.

Since P_C is nonexpansive and F is Lipschitz continuous with constant L, we get that $P_C \circ (I - \lambda F)$ is Lipschitz continuous with constant $1 + \lambda L$. Using that

$$y(t) = P_C \circ (I - \lambda F)(x(t)) \quad \forall t \in [0, +\infty),$$

if follows that the trajectory y is locally absolutely continuous and that for almost every $t \in [0, +\infty)$ it holds

$$\|\dot{y}(t)\| \leq (1 + \lambda L) \|\dot{x}(t)\|.$$

On the other hand,

$$\|\dot{x}(t)\| = \|x(t) - y(t) - \lambda \left[F(x(t)) - F(y(t)) \right] \| \leq (1 + \lambda L) \|x(t) - y(t)\| \quad \forall t \in [0, +\infty).$$

Thus, for almost every $t \in [0, +\infty)$,

$$\begin{aligned}
\frac{d}{dt} \|x(t) - y(t)\|^2 &= 2 \langle x(t) - y(t), \dot{x}(t) - \dot{y}(t) \rangle \\
&\leq 2 \left(\|\dot{x}(t)\| + \|\dot{y}(t)\| \right) \|x(t) - y(t)\| \\
&\leq 2 \left(1 + \lambda L + (1 + \lambda L)^2 \right) \|x(t) - y(t)\|^2.
\end{aligned}$$

From here, according to Lemma 2.1, we obtain

$$\lim_{t \to +\infty} \|x(t) - y(t)\| = 0. \qquad \qquad \square$$

We come now to the main theorem of this section.

THEOREM 2.1. *Assume that the solution set Ω is nonempty, F is pseudo-monotone on H, Lipschitz continuous with constant $L > 0$ and sequentially weak-to-weak continuous, and $0 < \lambda < \frac{1}{L}$. Then the trajectories $x(t)$ and $y(t)$ generated by (4) converge weakly to a solution of* VI(F, C) *as $t \to +\infty$.*

Proof. Let $\hat{x} \in H$ be a weak sequential cluster point of $x(t)$ as $t \to +\infty$ and $(t_n)_{n \geq 0}$ be a sequence in $[0, +\infty)$ with $t_n \to +\infty$ and $x(t_n) \rightharpoonup \hat{x}$ as $n \to +\infty$. Since $\lim_{t \to +\infty} \|x(t) - y(t)\| = 0$, we also have $y(t_n) \rightharpoonup \hat{x}$ as $n \to +\infty$. Furthermore, since F is Lipschitz continuous, $\|F(x(t_n)) - F(y(t_n))\| \to 0$ as $n \to +\infty$. We will prove that $\hat{x} \in \Omega$. For convenience, we denote $x_n := x(t_n)$ and $y_n := y(t_n)$ for every $n \geq 0$. Since $(y_n)_{n \geq 0} \subseteq C$ and C is weakly closed, we have $\hat{x} \in C$. We assume that $F(\hat{x}) \neq 0$, otherwise the conclusion follows automatically.

Let $y \in C$ be fixed. For every $n \geq 0$ we have

$$y_n = P_C(x_n - \lambda F(x_n)),$$

thus

$$\langle x_n - \lambda F(x_n) - y_n, y - y_n \rangle \leq 0$$

or, equivalently,

(8) $$\frac{1}{\lambda} \langle x_n - y_n, y - y_n \rangle \leq \langle F(x_n) - F(y_n), y - y_n \rangle + \langle F(y_n), y - y_n \rangle.$$

Letting in the last inequality $n \to +\infty$ and taking into account that $\lim_{n \to +\infty} \|x_n - y_n\| = 0$, $\lim_{n \to +\infty} \|F(x_n) - F(y_n)\| = 0$ and $(y_n)_{n \geq 0}$ is bounded, it follows

$$\liminf_{n \to +\infty} \langle F(y_n), y - y_n \rangle \geq 0.$$

On the other hand, we have that $(y_n)_{n \geq 0}$ converges weakly to $\hat{x}$ as $n \to +\infty$. Since F is sequentially weak-to-weak continuous, $(F(y_n))_{n \geq 0}$ converges weakly to $F(\hat{x})$ as $n \to +\infty$. Since the norm mapping is sequentially weakly lower semicontinuous, we have

$$0 < \|F(\hat{x})\| \leq \liminf_{n \to +\infty} \|F(y_n)\|.$$

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 $k \to +\infty$.

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 exists $N_0 \geq 0$ 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

$$\langle F(y_{n_0}), y - y_{n_0} \rangle + \epsilon_0 \geq 0 \text{ and } F(y_{n_0}) \neq 0.$$

We can continue this construction inductively and assume to this end that $n_0 < n_1 < \ldots < n_k$ are given. Then there exists $N_{k+1} \geq 0$ such that $\inf_{n \geq N_{k+1}} \langle F(y_n), y - y_n \rangle > -\epsilon_{k+1}$. Taking $n_{k+1} > \max\{N_{k+1}, n_k\}$, we have

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

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

$$(9) \qquad \langle F(y_{n_k}), y - y_{n_k} \rangle + \epsilon_k \geq 0 \text{ and } F(y_{n_k}) \neq 0 \quad \forall k \geq 0.$$

Setting for every $k \geq 0$

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

it holds $\langle F(y_{n_k}), z_k \rangle = 1$. According to (9) we have that

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

Since F is pseudo-monotone on H, it yields

$$(10) \qquad \langle F(y + \epsilon_k z_k), y + \epsilon_k z_k - y_{n_k} \rangle \geq 0 \quad \forall k \geq 0.$$

Using that $(F(y_{n_k}))_{k \geq 0}$ is bounded we have

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

Taking in (10) the limit as $k \to +\infty$ we obtain

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

As y was arbitrarily chosen in C, it follows from Proposition 1.1 that $\hat{x} \in \Omega$.

On the other hand, by Proposition 2.2, for every $x^* \in \Omega$, $\|x(t) - x^*\|$ converges as $t \to +\infty$. Thus, according to the Lemma 2.2, $x(t)$ converges weakly to an element of Ω as $t \to +\infty$. Since, due to Proposition 2.2, we have that

$$\lim_{t \to +\infty} \|x(t) - y(t)\| = 0,$$

it follows that $y(t)$ converges weakly to the same element of Ω as $t \to +\infty$. □

The following example introduces a class of operators which are pseudo-monotone, Lipschitz continuous and sequentially weak-to-weak continuous on H, but are not necessarily monotone.

EXAMPLE 2.1. Let $F : H \to H$ be defined as

$$F(x) := g(x)(Mx + p),$$

where $M : H \to H$ is a linear bounded operator satisfying

$$\langle Mx, x \rangle \geq 0 \quad \forall x \in H,$$

$p \in H$, and $g : H \to (0, +\infty)$ is a function taking positive values.

Such operators have been considered in [5] in the case when H is a finite dimensional and M is a skew operator, i.e. $\langle Mx, x \rangle = 0$ for every $x \in H$, under the name *pseudo-affine operators*. In general F is not monotone, see [5]. This fact is reflected by Figure 1 in the case when $H = \mathbb{R}$.

We show that F is pseudo-monotone on H. Indeed, let $x, y \in H$ be such that $\langle F(x), y - x \rangle \geq 0$. Since $g(x) > 0$, we have

$$\langle Mx + p, y - x \rangle \geq 0.$$

Hence

$$\langle F(y), y - x \rangle = g(y)\langle My + p, y - x \rangle \geq g(y)(\langle My + p, y - x \rangle - \langle Mx + p, y - x \rangle)$$
$$= g(y)\langle M(y - x), y - x \rangle \geq 0,$$

which leads to the desired conclusion.

Since every linear bounded operator $M : H \to H$ is sequentially weak-to-weak continuous, the operator F is sequentially weak-to-weak continuous if g is weakly continuous. This is for instance the case when g has the expression $g(x) := \eta(\langle a, x \rangle)$ for a fixed vector $a \in H$ and a continuous function $\eta : \mathbb{R} \to (0, +\infty)$.

In addition, for some choices of H, a and η the operator F is Lipschitz continuous. Indeed, for $H = \ell_2$, $a = e_1 = (1, 0, 0, ...) \in \ell_2$, $\eta(t) = e^{-t^2}$,

$$M : \ell_2 \to \ell_2, M(x_1, x_2, ...) = (x_1, 0, 0, ...), \text{ and } p = 0 \in \ell_2,$$

the operator $F : \ell_2 \to \ell_2, F(x_1, x_2, ...) = (x_1 e^{-x_1^2}, 0, 0, ...)$, is Lipschitz continuous. This follows by the Mean Value Theorem, since it is easy to see by direct computation that there exists $L > 0$ such that $\|\nabla F(x)\| \leq L$ for every $x \in \ell_2$. We illustrate in Figure 1 this choice of F.
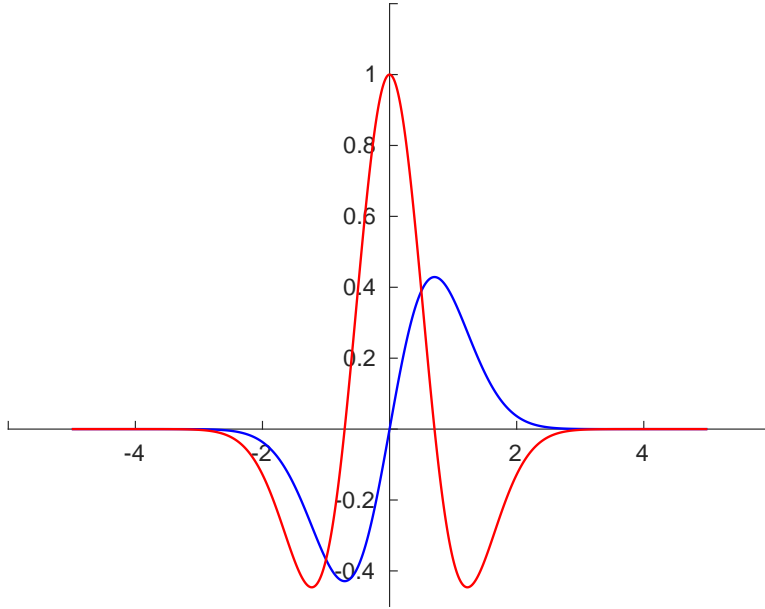

FIG. 1. *The graph of $F : \mathbb{R} \to \mathbb{R}, F(x) = xe^{-x^2}$, is in blue and the graph of $\nabla F : \mathbb{R} \to \mathbb{R}, \nabla F(x) = (1 - 2x^2)e^{-x^2}$, is in red.*

For the important particular case of strongly pseudo-monotone operators we will show exponential convergence of the trajectories to the unique solution of VI(F, C). To this end we need the following lemma.

LEMMA 2.3. *Assume that F is γ-strongly pseudo-monotone on C with $\gamma > 0$ and Lipschitz continuous with constant $L > 0$. Then for every $t \in [0, +\infty)$ we have*

$$\|x(t) - x^*\| \leq \frac{1 + \lambda L + \lambda \gamma}{\lambda \gamma} \|x(t) - y(t)\|.$$

Proof. Let $x^* \in C$ be the unique solution to $\mathrm{VI}(F, C)$ (see, for instance, [18]) and $t \in [0, +\infty)$ fixed. Since $y(t) \in C$ we have

$$\langle F(x^*), y(t) - x^* \rangle \geq 0,$$

which implies, according to the strong pseudo-monotonicity of F on C, that

$$\langle F(y(t)), y(t) - x^* \rangle \geq \gamma \|y(t) - x^*\|^2.$$

Using the Lipschitz continuity of F we get

$$\begin{aligned}
\langle F(x(t)), x^* - y(t) \rangle &= \langle F(x(t)) - F(y(t)), x^* - y(t) \rangle - \langle F(y(t)), y(t) - x^* \rangle \\
&\leq \|F(x(t)) - F(y(t))\| \|y(t) - x^*\| - \gamma \|y(t) - x^*\|^2 \\
&\leq L \|x(t) - y(t)\| \|y(t) - x^*\| - \gamma \|y(t) - x^*\|^2,
\end{aligned}$$

which, in combination with (7), gives

$$\begin{aligned}
\langle x^* - y(t), x(t) - y(t) \rangle &\leq \lambda \langle F(x(t)), x^* - y(t) \rangle \\
&\leq \lambda L \|x(t) - y(t)\| \|y(t) - x^*\| - \lambda \gamma \|y(t) - x^*\|^2
\end{aligned}$$

and, further,

$$\begin{aligned}
\lambda \gamma \|y(t) - x^*\|^2 &\leq \lambda L \|x(t) - y(t)\| \|y(t) - x^*\| - \langle x^* - y(t), x(t) - y(t) \rangle \\
&\leq \lambda L \|x(t) - y(t)\| \|y(t) - x^*\| + \|x^* - y(t)\| \|x(t) - y(t)\| \\
&= (1 + \lambda L) \|x(t) - y(t)\| \|y(t) - x^*\|.
\end{aligned}$$

This implies

$$\|y(t) - x^*\| \leq \frac{1 + \lambda L}{\lambda \gamma} \|x(t) - y(t)\|$$

and, further,

$$(11) \qquad \|x(t) - x^*\| \leq \|x(t) - y(t)\| + \|y(t) - x^*\| \leq \frac{1 + \lambda L + \lambda \gamma}{\lambda \gamma} \|x(t) - y(t)\|. \qquad \square$$

THEOREM 2.2. *Assume that F is γ-strongly pseudo-monotone on C with $\gamma > 0$ and Lipschitz continuous with constant $L > 0$, and $0 < \lambda < \frac{1}{L}$. Then for every $t \in [0, +\infty)$ we have*

$$(12) \qquad \|x(t) - x^*\|^2 \leq \|x(0) - x^*\|^2 \exp(-\alpha t),$$

where $\alpha =: 2(1 - \lambda L) \left(\frac{\lambda \gamma}{1 + \lambda L + \lambda \gamma} \right)^2$ and x^* is the unique solution of $\mathrm{VI}(F, C)$.

Proof. From Lemma 2.3 we have that for every $t \in [0, +\infty)$

$$\|x(t) - x^*\| \leq \frac{1 + \lambda L + \lambda \gamma}{\lambda \gamma} \|x(t) - y(t)\|,$$

10

which, in combination with Proposition 2.1, leads to

$$\frac{1}{2}\frac{d}{dt}\|x(t) - x^*\|^2 = \langle x(t) - x^*, \dot{x}(t)\rangle$$
$$\leq -(1 - \lambda L)\|x(t) - y(t)\|^2$$
$$\leq -(1 - \lambda L)\left(\frac{\lambda\gamma}{1 + \lambda L + \lambda\gamma}\right)^2 \|x(t) - x^*\|^2.$$

Relation (12) is a direct consequence of Gronwall's Lemma.   □

EXAMPLE 2.2. If $M : H \to H$ is such that

$$\langle Mx, x\rangle \geq \gamma\|x\|^2 \quad \forall x \in H,$$

for some $\gamma > 0$, then one can show that the operator $F : H \to H$ in Example 2.1 is $\alpha\gamma$-strongly pseudo-monotone on H. On the other hand, F is in general not monotone, as one can see in Figure 2 for a particular operator.
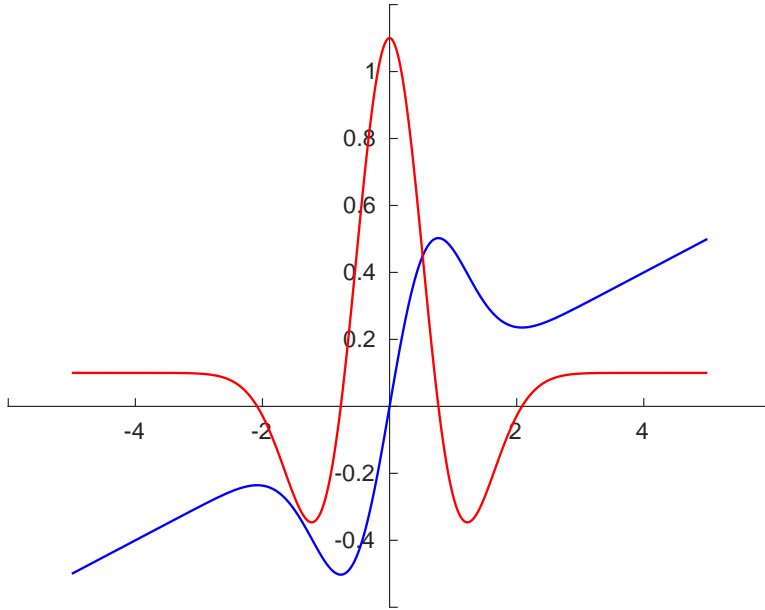

FIG. 2. *The graph of* $F : \mathbb{R} \to \mathbb{R}, F(x) = xe^{-x^2} + 0.1x$, *is in blue and the graph of* $\nabla F : \mathbb{R} \to \mathbb{R}, \nabla F(x) = (1 - 2x^2)e^{-x^2} + 0.1$, *is in red.*

EXAMPLE 2.3. Let $C = \{x \in [-5, 5]^3 : x_1 + x_2 + x_3 = 0\} \subseteq \mathbb{R}^3$ and $F : \mathbb{R}^3 \to \mathbb{R}^3$ be defined as

$$F(x) = \left(e^{-\|x\|^2} + q\right)Mx,$$

11

where $q = 0.2$ and

$$M = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1.5 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$

As mentioned in Example 2.2, F is γ-strongly pseudo-monotone on $\mathbb{R}^3$ with constant $\gamma := q \cdot \lambda_{min} \approx 0.0764$, where λ_{min} is the smallest eigenvalue of M, and Lipschitz continuous with constant $L \approx 5.0679$. Since for $x = (-1, 0, 0)^T, y = (-2, 0, 0)^T \in \mathbb{R}^3$

$$\langle F(x) - F(y), x - y \rangle = -0.1312 < 0,$$
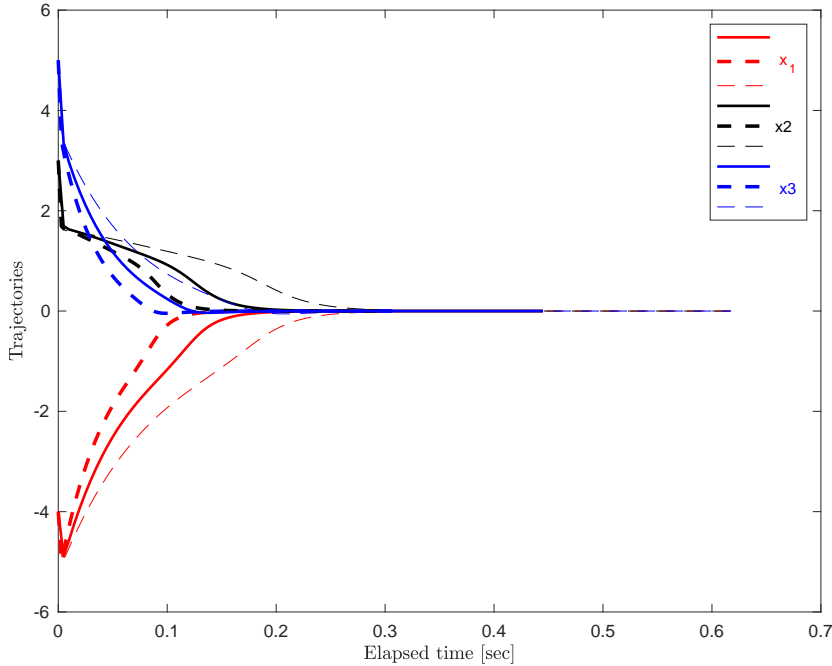
F is not monotone.

FIG. 3. *Trajectories generated by the dynamical system* (4) *for $x_0 = (-4, 3, 5)^T$ and $\lambda = 0.99/L$ (continuous lines), $\lambda = 0.8/L$ (dashed thick lines), and $\lambda = 0.5/L$ (dashed thin lines).*

Figure 3 displays the trajectories generated by the dynamical system (4) attached to VI(F, C), with starting point $x_0 = (-4, 3, 5)^T$ and different values of λ. These are represented for $\lambda = 0.99/L$ by continuous lines, for $\lambda = 0.8/L$ by dashed thick lines, and for $\lambda = 0.5/L$ by dashed thin lines. They all converge exponentially to the unique solution $x^* = (0, 0, 0)^T$ of VI(F, C). One can clearly see that the choice of λ influences the speed of convergence, namely, the smaller the values of λ, the worse the convergence of the trajectories become.

3. The forward-backward-forward algorithm with relaxation parameters. In this section we analyze the convergence of Tseng's forward-backward-forward algorithm with relaxation parameters derived in Remark 2.1 by the time discretization

12

of the dynamical system (4) in the context of solving pseudo-monotone variational inequalities.

ALGORITHM 3.1. ***Initialization:*** *Choose the starting point $x_0 \in H$, the step size $\lambda > 0$, and the sequence of relaxation parameters $(\rho_n)_{n \geq 0}$. Set $n = 0$.*

Step 1: *Compute*

$$y_n = P_C(x_n - \lambda F(x_n)).$$

If $y_n = x_n$ or $F(y_n) = 0$, then STOP: y_n is a solution.

Step 2: *Set*

$$x_{n+1} = \rho_n \left(y_n + \lambda(F(x_n) - F(y_n)) \right) + (1 - \rho_n)x_n,$$

update n to $n + 1$ and go to **Step 1**.

REMARK 3.1. *If $\rho_n = 1$ for every $n \geq 0$, then Algorithm 3.1 reduces to the classical forward-backward-forward method proposed by Tseng in [32].*

For the convergence analysis we assume that Algorithm 3.1 does not terminate after a finite number of iterations. In other words, we assume that for every $n \geq 0$ it holds $x_n \neq y_n$ and $F(y_n) \neq 0$.

PROPOSITION 3.1. *Assume that the solution set Ω is nonempty and F is pseudo-monotone on C and Lipschitz continuous with constant L. Let $t_n := y_n + \lambda(F(x_n) - F(y_n))$ for every $n \geq 0$. Then for every solution $x^* \in \Omega$ and every $n \geq 0$ it holds*

$$(13) \quad \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - \rho_n \left(1 - \lambda^2 L^2 \right) \|y_n - x_n\|^2 - \rho_n(1 - \rho_n)\|t_n - x_n\|^2.$$

Proof. Let x^* be an arbitrary element in Ω and $n \geq 0$ 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 this inequality it yields

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

From the pseudo-monotonicity of F on C it follows

$$(14) \quad \langle F(y_n), y_n - x^* \rangle \geq 0.$$

Since $y_n = P_C(x_n - \lambda F(x_n))$, according to (2), we get

$$\langle y - y_n, y_n - x_n + \lambda F(x_n) \rangle \geq 0 \quad \forall y \in C,$$

which yields

$$(15) \quad \langle x^* - y_n, y_n - x_n + \lambda F(x_n) \rangle \geq 0.$$

Multiplying both sides of (14) by $\lambda > 0$ and adding the resulting inequality to (15), it yields

$$\langle x^* - y_n, y_n - x_n + \lambda F(x_n) - \lambda F(y_n) \rangle \geq 0$$

or, equivalently,

$$\langle x^* - y_n, t_n - x_n \rangle \geq 0.$$

13

This implies that

$$\langle t_n - x^*, t_n - x_n \rangle \leq \langle t_n - y_n, t_n - x_n \rangle$$
$$= \|t_n - x_n\|^2 + \langle x_n - y_n, t_n - x_n \rangle$$
$$= \|t_n - x_n\|^2 + \langle x_n - y_n, y_n + \lambda(F(x_n) - F(y_n)) - x_n \rangle$$
(16)
$$= \|t_n - x_n\|^2 - \|y_n - x_n\|^2 + \lambda \langle x_n - y_n, F(x_n) - F(y_n) \rangle.$$

On the other hand, we have

(17) $$\|t_n - x^*\|^2 - \|x_n - x^*\|^2 + \|t_n - x_n\|^2 = 2 \langle t_n - x^*, t_n - x_n \rangle.$$

Combining (16) and (17) we obtain

$$\|t_n - x^*\|^2 \leq \|x_n - x^*\|^2 + \|t_n - x_n\|^2 - 2\|y_n - x_n\|^2$$
(18)
$$+ 2\lambda \langle x_n - y_n, F(x_n) - F(y_n) \rangle.$$

Using the Lipschitz continuity of F we obtain

$$\|t_n - x_n\|^2 = \|y_n + \lambda(F(x_n) - F(y_n)) - x_n\|^2$$
$$= \|y_n - x_n\|^2 + 2\lambda \langle y_n - x_n, F(x_n) - F(y_n) \rangle + \lambda^2 \|F(x_n) - F(y_n)\|^2$$
(19)
$$\leq \|y_n - x_n\|^2 + 2\lambda \langle y_n - x_n, F(x_n) - F(y_n) \rangle + \lambda^2 L^2 \|x_n - y_n\|^2.$$

Finally, from (18) and (19) it yields

$$\|t_n - x^*\|^2 \leq \|x_n - x^*\|^2 - \left(1 - \lambda^2 L^2\right) \|y_n - x_n\|^2.$$

Moreover,

$$\|x_{n+1} - x^*\|^2 = \|\rho_n(t_n - x^*) + (1 - \rho_n)(x_n - x^*)\|^2$$
$$= \rho_n \|t_n - x^*\|^2 + (1 - \rho_n)\|x_n - x^*\|^2 - \rho_n(1 - \rho_n)\|t_n - x_n\|^2.$$

By plugging this equality in the inequality above, we obtain the desired result. $\quad\square$

REMARK 3.2. One can notice that the pseudo-monotonicity of F was used in the proof of Proposition 3.1 in order to obtain relation (14). This means that the pseudo-monotonicity of F can be actually replaced by the following weaker assumption (see [13, 29])

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

which basically requires that every solution of the variational inequality of Stampacchia type is a solution of the variational inequality of Minty type.

REMARK 3.3. In contrast to the extragradient method, the sequence $(x_n)_{n \geq 0}$ generated by Algorithm 3.1 may not be feasible. This is why we need to ask in the convergence analysis that F is Lipschitz continuous on the whole space H. However, if the feasible set C is bounded, then we can weaken this assumption by asking that F is Lipschitz continuous on the bounded set

$$D := \{x + y : x \in C, \|y\| \leq d\},$$

where d denotes the diameter of C. Notice that $C \subseteq D$. In this case, if we start Algorithm 3.1 with an element $x_0 \in C$ and choose $0 < \lambda < \frac{1}{L}$, then from (13) and $\rho_0 \in [0, 1]$ we have

$$\|x_1 - x^*\|^2 \leq \|x_0 - x^*\|^2 - \rho_0 \left(1 - \lambda^2 L^2\right) \|y_0 - x_0\|^2,$$

which implies that $\|x_1 - x^*\| \le \|x_0 - x^*\| \le d$. Since $x_1 = x^* + x_1 - x^*$, we have $x_1 \in D$. By induction, we obtain $\|x_n - x^*\| \le d$ and therefore $x_n \in D$ for every $n \ge 0$.

The following theorem states the convergence of the underrelaxed Tseng's method for pseudo-monotone variational inequalities.

THEOREM 3.1. *Assume that the solution set Ω is nonempty, F is pseudo-monotone on H, Lipschitz continuous with constant $L > 0$ and sequentially weak-to-weak continuous, and $0 < \lambda < \frac{1}{L}$. Assume also that $(\rho_n)_{n\ge0} \subseteq [0,1]$ and $\liminf_{n\to+\infty}\rho_n > 0$. Then the sequence $(x_n)_{n\ge0}$ generated by Algorithm 3.1 converges weakly to a solution of $\mathrm{VI}(F,C)$.*

Proof. Let $x^* \in \Omega$ be fixed. Since $(\rho_n)_{n\ge0} \subseteq [0,1]$ and $0 < \lambda < \frac{1}{L}$, (13) yields that the sequence $\left(\|x_n - x^*\|^2\right)_{n\ge0}$ is monotonically decreasing and therefore convergent.

To obtain the convergence of the sequence $(x_n)_{n\ge0}$ to an element in Ω we only need to prove that every weak sequential cluster point of the sequence belongs to Ω. The conclusion will follow from the Opial Lemma (see [3, Theorem 5.5]).

Relation (13) also implies that

$$\lim_{n\to+\infty} \rho_n(1 - \lambda^2 L^2)\|y_n - x_n\| = 0,$$

which, since $\liminf_{n\to+\infty}\rho_n > 0$, further leads to

$$\lim_{n\to+\infty} \|y_n - x_n\| = 0.$$

Since F is Lipschitz continuous on H, we have

$$\|F(x_n) - F(y_n)\| \le L\|x_n - y_n\| \quad \forall n \ge 0,$$

hence,

$$\lim_{n\to+\infty} \|F(x_n) - F(y_n)\| = 0.$$

Further, we consider $\hat{x}$, a weak sequential cluster point of $(x_n)_{n\ge0}$, and a subsequence $(x_{n_k})_{k\ge0}$ of $(x_n)_{n\ge0}$ which converges weakly to $\hat{x}$ as $k \to +\infty$. Since $\lim_{k\to+\infty}\|x_{n_k} - y_{n_k}\| = 0$, $(y_{n_k})_{k\ge0}$ also converges weakly to $\hat{x}$ as $k \to +\infty$.

We are now in the same situation as in the proof of Theorem 2.1, the role of the sequences $(x_n)_{n\ge0}$ and $(y_n)_{n\ge0}$ being played by $(x_{n_k})_{k\ge0}$ and $(y_{n_k})_{k\ge0}$, respectively. Thus, arguing as in the proof of this theorem, we obtain that $\hat{x} \in \Omega$. $\qquad\square$

REMARK 3.4. The conclusion of Theorem 3.1 remains valid even if we replace in every iteration of Algorithm 3.1 the fixed stepsize $\lambda > 0$ by a variable stepsize λ_n, where the sequence $(\lambda_n)_{n\ge0}$ fulfills

$$0 < \inf_{n\ge0}\lambda_n \le \sup_{n\ge0}\lambda_n < \frac{1}{L}.$$

On the other hand, when (an upper bound of) the Lipschitz constant of F is not available, we can use in Algorithm 3.1 the following adaptive stepsize strategy

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

where $\mu \in (0,1)$ and $\lambda_0 > 0$. The sequence $(\lambda_n)_{n\ge0}$ is nonincreasing. If, for $n \ge 0$, $F(x_n) - F(y_n) \ne 0$, then it holds

$$\frac{\mu\|x_n - y_n\|}{\|F(x_n) - F(y_n)\|} \ge \frac{\mu\|x_n - y_n\|}{L\|x_n - y_n\|} = \frac{\mu}{L},$$

which shows that $(\lambda_n)_{n\geq 0}$ is bounded from below by $\min\left\{\lambda_0, \frac{\mu}{L}\right\} > 0$. Notice that, if $\lambda_0 \leq \frac{\mu}{L}$, then $(\lambda_n)_{n\geq 0}$ is a constant sequence, which leads to a fixed stepsize strategy. Consequently, the limit $\lim_{n\to+\infty} \lambda_n$ exists and it is a positive real number.

This means that we can adapt the proof of Proposition 3.1 to the new adaptive stepsize strategy and, by taking into consideration (19), we get instead of (13)

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - \rho_n \left(1 - \frac{\lambda_n^2 \mu^2}{\lambda_{n+1}^2}\right) \|y_n - x_n\|^2 \ \forall n \geq 0.$$

Due to $\liminf_{n\to+\infty} \rho_n > 0$ and $\lim_{n\to+\infty} \left(1 - \frac{\lambda_n^2 \mu^2}{\lambda_{n+1}^2}\right) = 1 - \mu^2 > 0$, there exists $N > 0$ such that

$$\|x_{n+1} - x^*\| \leq \|x_n - x^*\| \quad \forall n \geq N,$$

which implies that $\lim_{n\to+\infty} \|x_n - x^*\|$ exists and $\lim_{n\to+\infty} \|x_n - y_n\| = 0$. From here, one can carry out the same convergence analysis as for the fixed stepsize strategy.

REMARK 3.5. If the operator F is monotone on C, then it is not necessary to impose that F is sequentially weak-to-weak continuous. Indeed, for $y \in C$ fixed, we obtain for the subsequences $(x_{n_k})_{k\geq 0}$ and $(y_{n_k})_{k\geq 0}$ arising in the proof of Theorem 3.1 (see also (8) and the proof of Theorem 2.1)

$$\frac{1}{\lambda} \langle x_{n_k} - y_{n_k}, y - y_{n_k} \rangle \leq \langle F(x_{n_k}) - F(y_{n_k}), y - y_{n_k} \rangle + \langle F(y_{n_k}), y - y_{n_k} \rangle$$

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

Letting $k \to +\infty$ we get

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

and this leads to the desired conclusion.

In the following we will show that the convergence result in Theorem 3.1 follows in finite dimensional spaces under weaker assumptions.

THEOREM 3.2. *Let 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 $0 < \lambda < \frac{1}{L}$. Assume also that $(\rho_n)_{n\geq 0} \subseteq [0,1]$ and $\liminf_{n\to+\infty} \rho_n > 0$. Then the sequence $(x_n)_{n\geq 0}$ generated by Algorithm 3.1 converges to a solution of* VI(F, C).

Proof. Let $x^* \in \Omega$ be fixed. Since $0 < \lambda < \frac{1}{L}$, from (13) it follows that the sequence $\left(\|x_n - x^*\|^2\right)_{n\geq 0}$ is monotonically decreasing and therefore convergent. In addition we have

$$\lim_{n\to+\infty} \|y_n - x_n\| = 0.$$

As $(x_n)_{n\geq 0}$ is bounded, there exists a subsequence $(x_{n_k})_{k\geq 0}$ of it, which converges to an element $\hat{x}$ as $k \to +\infty$. Since $\lim_{n\to+\infty} \|x_{n_k} - y_{n_k}\| = 0$, $(y_{n_k})_{k\geq 0}$ also converges to $\hat{x}$ as $k \to +\infty$.

Let now $y \in C$ be fixed. Then we have that

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

Taking the limit as $k \to +\infty$ and using that F is continuous, we obtain

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

Since $y \in C$ has been arbitrarily chosen, it follows that $\hat{x}$ is a solution of VI(F, C).

Replacing in (13) x^* with $\hat{x}$, it yields that the sequence $\left(\|x_n - \hat{x}\|\right)_{n\geq 0}$ is convergent. Since $\lim_{k\to+\infty} \|x_{n_k} - \hat{x}\| = 0$, it follows that $\lim_{n\to+\infty} x_n = \hat{x}$. □

In the next theorem we show that one can consider even an overrelaxation of the forward-backward-forward algorithm without altering its convergence behaviour.

THEOREM 3.3. *Assume that the solution set* Ω *is nonempty,* F *is pseudo-monotone on* H, *Lipschitz continuous with constant* $L > 0$ *and sequentially weak-to-weak continuous, and* $0 < \lambda < \frac{1}{L}$. *Assume also that* $(\rho_n)_{n \geq 0} \subseteq [1, 2)$ *and* $\limsup_{n \to +\infty} \rho_n < 2 - \frac{2\lambda L}{1 + \lambda L}$. *Then the sequence* $(x_n)_{n \geq 0}$ *generated by Algorithm* 3.1 *converges weakly to a solution of* $\mathrm{VI}(F, C)$.

Proof. In view of Proposition 3.1 we have that

(20)
$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - \rho_n \left(1 - \lambda^2 L^2\right) \|y_n - x_n\|^2 + \rho_n(\rho_n - 1)\|t_n - x_n\|^2 \ \forall n \geq 0.$$

By the Lipschitz continuity of F we have for all $n \geq 0$ that

$$\|t_n - x_n\|^2 = \|y_n - x_n + \lambda\left(F(x_n) - F(y_n)\right)\|^2 \leq (1 + \lambda L)^2 \|y_n - x_n\|^2.$$

Therefore, from (20) we obtain

$$\begin{aligned}
\|x_{n+1} - x^*\|^2 &\leq \|x_n - x^*\|^2 - \rho_n \left(1 - \lambda^2 L^2\right) \|y_n - x_n\|^2 \\
&\quad + \rho_n(\rho_n - 1)(1 + \lambda L)^2 \|y_n - x_n\|^2 \\
&= \|x_n - x^*\|^2 - \rho_n \left(1 - \lambda^2 L^2 - (\rho_n - 1)(1 + \lambda L)^2\right) \|y_n - x_n\|^2.
\end{aligned}$$

Since $(\rho_n)_{n \geq 0} \subseteq [1, 2)$ and $\limsup_{n \to +\infty} \rho_n < 2 - \frac{2\lambda L}{1 + \lambda L}$, it is easy to check that

$$\liminf_{n \to +\infty} \left(1 - \lambda^2 L^2 - (\rho_n - 1)(1 + \lambda L)^2\right) > 0.$$

Hence, there exists $N \geq 0$ such that the sequence $\left(\|x_n - x^*\|^2\right)_{n \geq N}$ is monotonically decreasing and therefore convergent. In addition, we have

$$\lim_{n \to +\infty} \|y_n - x_n\| = 0.$$

The rest of the proof can be done in analogy to the proof of of Theorem 3.1, relying on the Opial Lemma and on arguments from the proof of Theorem 2.1. □

EXAMPLE 3.1. A differentiable function $f : E \to \mathbb{R}$, where $E \subseteq \mathbb{R}^n$ is an open set, is called *pseudo-convex on* E, if for every $x, y \in E$ it holds

$$\langle \nabla f(x), y - x \rangle \geq 0 \quad \Rightarrow \quad f(y) \geq f(x).$$

It is well-known that f is pseudo-convex on E if and only if ∇f is pseudo-monotone on E ([17]). Algorithm 3.1 can be used to solve optimization problems of the form

$$\min_{x \in C} f(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function with Lipschitz continuous gradient which is also pseudo-convex on an open set $E \subseteq \mathbb{R}^n$, and $C \subseteq E$ is a nonempty, convex and closed set. The class of pseudo-convex functions has been investigated in [26], while characterizations of quadratic pseudo-convex functions have been provided in [11].

A important subclass of the one of pseudo-convex functions are ratios of convex and concave functions. Indeed, if $E \subseteq \mathbb{R}^n$ is a convex set, $g : E \to [0, +\infty)$ is a convex

function, $h : E \to (0, +\infty)$ is a concave function, and both g and h are differentiable on E, then the function

$$f : E \to [0, +\infty), \quad f(x) := \frac{g(x)}{h(x)},$$

is pseudo-convex on E ([6]).

In the following we show that when F is strongly pseudo-monotone on C, then Algorithm 3.1 generates a sequence which converges linearly to the unique solution of VI(F, C). We extend in this way a result proved by Tseng in [32] for strongly monotone operators.

THEOREM 3.4. *Assume that F is γ-strongly pseudo-monotone on C with $\gamma > 0$ and Lipschitz continuous with constant $L > 0$, and $0 < \lambda < \frac{1}{L}$. Assume also that $(\rho_n)_{n \geq 0} \subseteq [0, 1]$. Let x^* be the unique solution of the problem VI(F, C). Then*

$$\|x_{n+1} - x^*\| \leq \delta_n \|x_n - x^*\| \quad \forall n \geq 0,$$

where $\delta_n := \left(1 - \rho_n \left(1 - \lambda^2 L^2 \right) \left(\frac{\lambda\gamma}{1+\lambda L+\lambda\gamma} \right)^2 \right)^{1/2} \in (0, 1)$.

In addition, if $\liminf_{n \to +\infty} \rho_n > 0$, then the sequence $(x_n)_{n \geq 0}$ converges linearly to x^*.

Proof. Let $n \geq 0$ be fixed. As in the proof of Lemma 2.3 (see (11)), one can show that

(21) $$\|x_n - x^*\| \leq \|x_n - y_n\| + \|y_n - x^*\| \leq \frac{1 + \lambda L + \lambda\gamma}{\lambda\gamma} \|x_n - y_n\|.$$

From (21) and (13) we obtain

$$\|x_{n+1} - x^*\|^2 \leq \left(1 - \rho_n \left(1 - \lambda^2 L^2 \right) \left(\frac{\lambda\gamma}{1 + \lambda L + \lambda\gamma} \right)^2 \right) \|x_n - x^*\|^2,$$

therefore,

$$\|x_{n+1} - x^*\| \leq \delta_n \|x_n - x^*\|,$$

where $\delta_n := \left(1 - \rho_n \left(1 - \lambda^2 L^2 \right) \left(\frac{\lambda\gamma}{1+\lambda L+\lambda\gamma} \right)^2 \right)^{1/2} \in (0, 1)$. Now, if $\liminf_{n \to +\infty} \rho_n > 0$, then we have $\limsup_{n \to +\infty} \delta_n < 1$, which means that $(x_n)_{n \geq 0}$ converges linearly to x^*. □

One can prove in a similar way linear convergence for the sequence generated by the overrelaxed variant of the forward-backward-forward algorithm.

THEOREM 3.5. *Assume that $0 < \lambda < \frac{1}{L}$ and F is γ-strongly pseudo-monotone on C with $\gamma > 0$ and Lipschitz continuous with constant $L > 0$. Assume also that $(\rho_n)_{n \geq 0} \subseteq [1, 2)$ and $\limsup_{n \to +\infty} \rho_n < 2 - \frac{2\lambda L}{1+\lambda L}$. Then the sequence $(x_n)_{n \geq 0}$ converges linearly to the unique solution x^* of VI(F, C).*

4. Numerical experiments. In this section we present two numerical experiments which we carried out in order to compare Algorithm 3.1 with other algorithms in the literature designed for solving pseudo-monotone variational inequalities. We implemented the numerical codes in Matlab and performed all computations on a Linux desktop with an Intel(R) Core(TM) i5-4670S processor at 3.10GHz. In our experiments we considered only variational inequalities governed by pseudo-monotone operators which are not monotone.

Let be VI(F, C) with

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

and

$$F : \mathbb{R}^5 \to \mathbb{R}^5, F(x) = \left(e^{-\|x\|^2} + \alpha \right) (Mx + p),$$

where $\| \cdot \|$ denotes the Euclidean norm on $\mathbb{R}^5$, $\alpha = 0.1$, $p = (-1, 2, 1, 0, -1)^T$ and

(22)
$$M := \begin{bmatrix} 5 & -1 & 2 & 0 & 2 \\ -1 & 6 & -1 & 3 & 0 \\ 2 & -1 & 3 & 0 & 1 \\ 0 & 3 & 0 & 5 & 0 \\ 2 & 0 & 1 & 0 & 4 \end{bmatrix}$$

is a positive definite matrix. We computed the unique solution x^* of the variational inequality VI(F, C) by running 10000 iterations of Algorithm 3.1 with $\rho_n = 1$ for all $n \geq 0$ and stepsize $\lambda = \frac{0.5}{L}$.

In a first experiment, we considered different variants of Algorithm 3.1 with constant relaxation parameters $\rho_n = \rho$ for all $n \geq 0$, chosen such that $\rho < 2 - \frac{2\lambda L}{1+\lambda L} = \frac{4}{3}$. The aim was to see to which extend the relaxation parameter does influence the convergence behaviour of the method. We considered $x_0 = (1, 3, 2, 1, 4)^T$ as starting point and $\|x_n - x^*\| \leq 10^{-6}$ as stopping criterion. The projection on the set C was computed by using the `quadprog` function in Matlab.

In Table 1 we present the performances of the algorithm for different values of the relaxation parameter. It can be seen that the larger the values of the relaxation parameter ρ, the better the algorithm performs. This shows how important is it to investigate overrelaxed algorithms from both theoretical and numerical perspective.

TABLE 1
Comparison of the performances of Algorithm 3.1 for different values of ρ.

ρ	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3
Iterations	236	195	166	144	127	112	90	93	88
Time (sec)	2.06	1.10	0.89	0.74	0.60	0.55	0.47	0.56	0.44

In a second experiment we compared for the same problem the performances of the forward-backward-forward method without relaxation, the extragradient method and the subgradient-extragradient method, by considering for all three methods as stepsize $\lambda = 0.99/L$. It can be seen in Figure 4 that the forward-backward-forward method outperforms the extragradient method, being at least two times faster. This is not surprising, since the extragradient method requires two projections on the set C at each iteration, while the forward-backward-forward method requires only one. It can be also notice that the latter also slightly outperforms the subgradient-extragradient method.

In a third experiment we considered the quadratic fractional programming problem

(23)
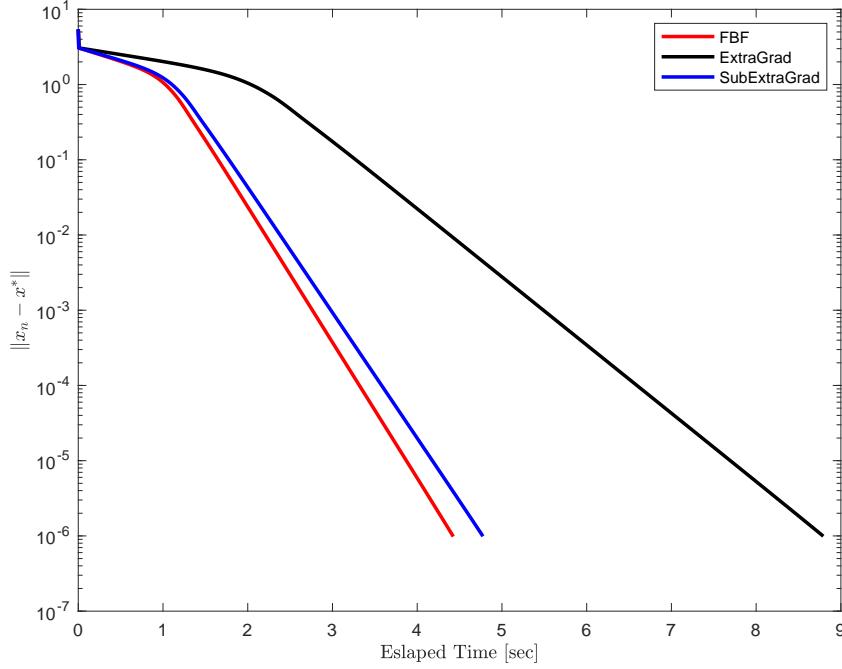$$\min_{x \in C} f(x) := \frac{x^T M x + a^T x + c}{b^T x + d},$$

19

Fig. 4. *Comparison of the convergence behaviour of the forward-backward-forward method without relaxation (FBF), the extragradient method (ExtraGrad), and the subgradient-extragradient method (SubExtraGrad) with stepsize $\lambda = 0.99/L$.*

with

$$C = \{x \in \mathbb{R}^5 : 1 \le x_i \le 3, \ i = 1, 2, 3, 4, 5\},$$

M taken as in (22), $a = (1, 2, -1, -2, 1)^T$, $b = (1, 0, -1, 0, 1)^T$, $c = -2$ and $d = 20$. According to the discussion in Example 3.1, f is pseudo-convex on the open set $E := \{x \in \mathbb{R}^5 : b^T x + d = x_1 - x_3 + x_5 + 20 > 0\}$, which implies that

$$F : \mathbb{R}^5 \to \mathbb{R}^5, \ F(x) = \nabla f(x) := \frac{\left(b^T x + d\right)(2Mx + a) - b\left(x^T M x + a^T x + c\right)}{\left(b^T x + d\right)^2},$$

is pseudo-monotone on E. One can also notice that $C \subseteq E$.

In order to show the Lipschitz continuity of F, since C is bounded, according to Remark 3.3 it is enough to prove that this property holds on the set

$$
\begin{aligned}
D &= \{x + y \in \mathbb{R}^5 : x \in C, \|y\| \le 2\sqrt{5}\} \\
&= \{x \in \mathbb{R}^5 : 1 - 2\sqrt{5} \le x_i \le 3 + 2\sqrt{5}, i = 1, 2, 3, 4, 5\}.
\end{aligned}
$$

Notice that $C \subseteq D \subseteq E$.

One can easily see that $\|\nabla F(x)\| \le 148.68 =: L > 0$ for all $x \in D$, which means according to the Mean Value Theorem that F is Lipschitz continuous on D with constant L. For this numerical experiment we assumed that the constant L is
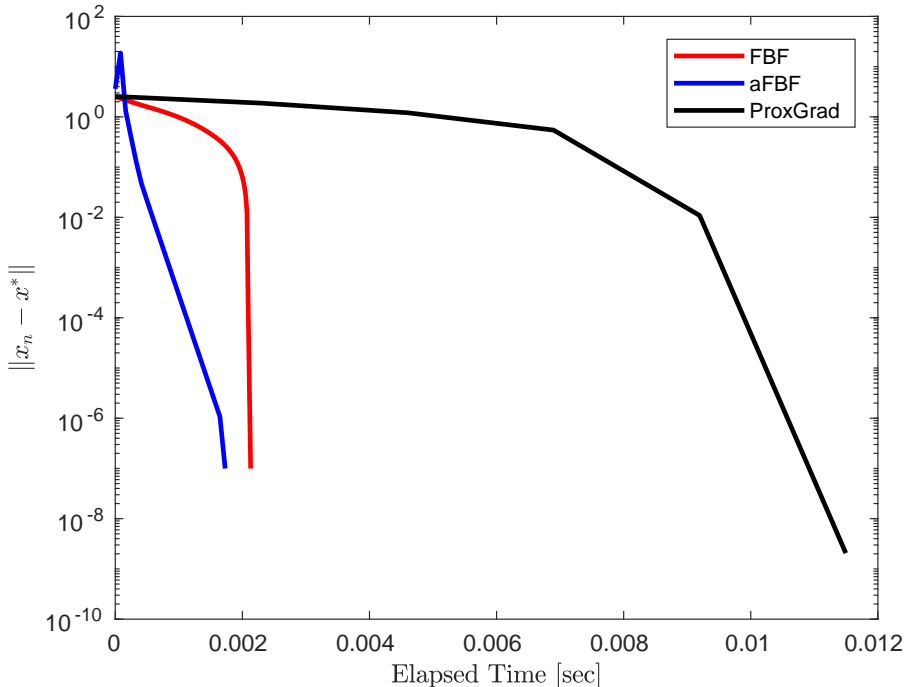
20

FIG. 5. *Comparison of the convergence behaviour of the forward-backward-forward method (FBF) with fixed stepsize, the one with adaptive stepsize (aFBF) and the proximal-gradient method (ProxGrad) for the fractional programming problem* (23).

not known in advance and used the adaptive stepsize strategy described in Remark 3.4 with $\mu = 0.9$ and $\lambda_0 = 1$. We compared the forward-backward-forward method (FBF) with fixed stepsize $\lambda = 0.9/L$, with the one with adaptive stepsize (aFBF) and the proximal-gradient (ProxGrad) method for fractional programming proposed in [7, Algorithm 6]. We considered as starting point $x_0 = (3, 1.5, 2, 1.5, 2)^T$ and as stopping criterion $\|x_n - x^*\| \leq 10^{-6}$. The optimal solution of (23) $x^* = (1, 1, 1, 1, 1)^T$ was obtained by running 10000 iterations of Algorithm 3.1 with $\rho_n = 1$ for all $n \geq 0$. We solved the quadratic subproblem in [7, Algorithm 6] by using the *quadprog* function in Matlab. The numerical performances of the three methods are displayed in Figure 5. One can notice that the adaptive method aFBF is faster than FBF. Moreover, both FBF and aFBF outperform the proximal-gradient method from [7, Algorithm 6]. A possible reason is that, while for the first two methods the projection on the set C is computed explicitly, in every iteration of the proximal-gradient method a subproblem is solved by an external solver.

5. Conclusions and further research. The object of our investigation was a variational inequality of Stampacchia type over a nonempty, convex and closed set governed by a pseudo-monotone and Lipschitz continuous operator. We associated to it a forward-backward-forward dynamical system and carried out a Lyapunov-type analysis in order to prove the asymptotic convergence of the generated trajectories to a solution of the variational inequality. The explicit time discretization of the dy-

21

namical system leads to Tseng's forward-backward-forward algorithm with relaxation parameters. We proved convergence of the generated sequence of iterates to a solution of the variational inequality as well as linear convergence rate under strong pseudo-monotonicity. Numerical experiments show that, when applied to pseudo-monotone variational inequalities over polyhedral sets, the overrelaxed variant algorithm has a better convergence behaviour when compared to other variants and also that Tseng's method outperforms Korpelevich's extragradient method and also the subgradient-extragradient method.

A topic of current interest is the formulation of numerical algorithms for minimax problems, due to its relevance for the training of generative adversarial networks (GANs). We want to investigate the convergence property of Tseng's forward-backward-forward method when solving the variational inequality to which the optimality conditions for the minimax problem give rise in both a deterministic and a stochastic setting and possibly to derive ergodic convergence rates for the gap function associated to the minimax problem.

REFERENCES

[1] ABBAS, B., ATTOUCH, H. AND SVAITER, B. F.: *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, J. Optim. Theory Appl. 161(2) (2014), pp. 331–360.

[2] BANERT, S. AND BOŢ, R. I.: *A forward-backward-forward differential equation and its asymptotic properties*, J. Conv. Anal. 25(2) (2018), pp. 371–388.

[3] BAUSCHKE, H. H. AND COMBETTES, P. L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York (2011).

[4] BELLO CRUZ, J. Y.; IUSEM, A.N.: *Convergence of direct methods for paramonotone variational inequalities*, Comput. Optim. Appl. 46(2) (2010), pp. 247–263.

[5] BIANCHI, M., HADJISAVVAS, N. AND SCHAIBLE, S.: *On pseudomonotone maps T for which −T is also pseudomonotone*, J. Conv. Anal. 10 (2003), pp. 149–168.

[6] BORWEIN, J. M. AND LEWIS, A. S.: *Convex Analysis and Nonlinear Optimization: Theory and Examples* Springer Science and Business Media, New York (2006).

[7] BOŢ, R. I. AND CSETNEK, E. R.: *Proximal-gradient algorithms for fractional programming*, Optimization 66(8) (2017), pp. 1383–1396.

[8] CENG, L.C., TEBOULLE, M. AND YAO, J.-C.: *Weak convergence of an iterative method for pseudomonotone variational inequalities and fixed-point problems*, J. Optim. Theory Appl. 146 (2010), pp. 19–31.

[9] CENSOR, Y., GIBALI, A. AND REICH, S.: *Extensions of Korpelevich's extragradient method for the variational inequality problem in Euclidean space*, Optimization 61(9) (2012), pp. 1119–1132.

[10] CENSOR, Y., GIBALI, A. AND REICH, S.:*The subgradient extragradient method for solving variational inequalities in Hilbert space*, J. Optim. Theory Appl. 148 (2011), pp 318–335.

[11] COTTLE, R.W. AND FERLAND, J.A.: *On pseudo-convex functions of nonnegative variables*, Math. Programm. 1 (1971), pp. 95–101.

[12] COTTLE, R. W. AND YAO, J. C.: *Pseudo-monotone complementarity problems in Hilbert space*, J. Optim. Theory Appl. 75 (1992), pp. 281–295.

[13] DANG, C. D. AND LAN, G.: *On the convergence properties of non-Euclidean extragradient methods for variational inequalities with generalized monotone operators*, Comput. Optim. Appl. 60 (2015), pp. 277–310.

[14] FACCHINEI, F. AND PANG, J.-S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York (2003).

[15] HADJISAVVAS, N., SCHAIBLE, S. AND WONG, N.-C.: *Pseudomonotone operators: a survey of the theory and its applications*, J. Optim. Theory Appl. 152 (2012), pp. 1–20.

[16] HARKER, P. T. AND PANG, J.-S.: *A damped-Newton method for the linear complementarity problem*, in: E.L. Allower, K. Georg, Computational Solution of Nonlinear Systems of Equations, AMS Lectures on Applied Mathematics, Vol. 26 (1990), pp. 265–284.

[17] KARAMARDIAN, S. AND SCHAIBLE, S.: *Seven kinds of monotone maps*, J. Optim. Theory Appl. 66 (1990) pp. 37-46.

[18] KIM, D. S., VUONG, P. T. AND KHANH, P. D. : *Qualitative properties of strongly pseudomonotone variational inequalities*, Opt. Lett. 10 (2016), pp.1669–1679.

[19] KINDERLEHRER, D. AND STAMPACCHIA, G.: *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York (1980).

[20] KHANH, P. D. AND VUONG, P. T.: *Modified projection method for strongly pseudomonotone variational inequalities*, J. Global Optim. 58 (2014), pp.341–350.

[21] KORPELEVICH, G. M.: *The extragradient method for finding saddle points and other problems*, Ekonomika i Mat. Metody 12 (1976), pp.747–756.

[22] LÁSZLÓ, S.C.: *Some existence results of solutions for general variational inequalities*, J. Optim. Theory Appl. 150 (2011), pp. 425–443.

[23] MALITSKY, Y.: *Projected reflected gradient methods for monotone variational inequalities*, SIAM J. Optim. 25 (2015), pp. 502–520.

[24] MALITSKY, Y.: *Proximal extrapolated gradient methods for variational inequalities*, Optim. Meth. Softw. 33(1) (2018), pp. 140-164.

[25] MALITSKY, Y.: *Golden ratio algorithms for variational inequalities*, arXiv:1803.08832 (2018).

[26] MANGASARIAN O.L.: *Pseudo-convex functions*, SIAM J. Control Optim. 3 (1965), pp. 281-290.

[27] MAUGERI, A. AND RACITI, F.: *On existence theorems for monotone and nonmonotone variational inequalities*, J. Conv. Analy. 16 (2009), pp. 899–911.

[28] SHEHU, Y., DONG, Q.-L. AND JIANG, D.: *Single projection method for pseudo-monotone variational inequality in Hilbert spaces*, Optimization 68(1) (2019), pp. 385–409.

[29] SOLODOV, M. V. AND SVAITER, B. F.: *A new projection method for variational inequality problems*, SIAM J. Control Optim. 37 (1999), pp. 765–776.

[30] SOLODOV, M. V. AND TSENG, P.: *Modified projection-type methods for monotone variational inequalities*, SIAM J. Control Optim. 34 (1996), pp. 1814–1830.

[31] THONG, D.V., SHEHU, Y. AND IYIOLA, O.S.: *Weak and strong convergence theorems for solving pseudo-monotone variational inequalities with non-Lipschitz mappings*, Numerical Algorithms, DOI: 10.1007/s11075-019-00780-0

[32] TSENG, P.: *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim. 38 (2000), pp. 431–446.

[33] VUONG, P. T.: *On the weak convergence of the extragradient method for solving variational inequalities*, J. Optim. Theory Appl. 176(2) (2018), pp. 399–409.

[34] ZHU, D. L. AND MARCOTTE, P.: *Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities*, SIAM J. Control Optim. 6 (1996), pp. 714–726.