

A SECOND ORDER DYNAMICAL SYSTEM AND ITS DISCRETIZATION FOR STRONGLY PSEUDO-MONOTONE VARIATIONAL INEQUALITIES

PHAN TU VUONG*

Abstract. We consider a second order dynamical system for solving variational inequalities in Hilbert spaces. Under standard conditions, we prove the existence and uniqueness of strong global solution of the proposed dynamical system. The exponential convergence of trajectories is established under strong pseudo-monotonicity and Lipschitz continuity assumptions. A discrete version of the proposed dynamical system leads to a relaxed inertial projection algorithm whose linear convergence is proved under suitable conditions on parameters. We discuss the possibility of extension to general monotone inclusion problems. Finally some numerical experiments are reported demonstrating the theoretical results.

Key words. Variational inequality, Dynamical system, Strong pseudo-monotonicity, Exponential convergence, Linear convergence, Monotone inclusion

AMS subject classifications. 47J20, 49J40, 90C30, 90C52

1. Introduction. Let H be a real Hilbert space endowed with an inner product and its induced norm denoted $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. Let C be a nonempty closed convex subset of H and let $A : H \rightarrow H$ be a continuous operator. The variational inequality $VI(A, C)$ consists in finding a point $x^* \in C$ such that

$$(1.1) \quad \langle Ax^*, y - x^* \rangle \geq 0 \quad \text{for every } y \in C.$$

Variational inequality (VI) is a general mathematical framework arising naturally in many theoretical and applied fields, such as economics, engineering mechanics, transportation, and many more; see for example, [2, 9, 17]. The VIs theoretical, algorithmic foundations and applications have been extensively studied in the literature. For the current state-of-the-art results, see, for instance [17, 28] and the references quoted therein. For solving VIs, numerous projection methods such as basic projection, extragradient projection, and hyperplane projection methods, have been designed to solve different classes of VIs [2, 17, 28, 34]. In principle, each method is confined to a certain class of VIs so that the convergence of the algorithm can be guaranteed.

In recent years, dynamical systems have been widely investigated for solving optimization problems, fixed point problems, variational inequalities and monotone inclusions [4, 11, 13, 14, 15, 16, 20, 30, 32, 37, 38]. For solving strongly pseudo-monotone VIs, some first order projected dynamical systems were proposed and investigated [16, 20, 24, 37]. In this paper, we continue this research direction by considering second order dynamical systems. The work is motivated by Antipin [4], who studied the following second order dynamical system

$$(1.2) \quad \begin{cases} \ddot{x}(t) + \alpha \dot{x}(t) + x(t) - P_C(x(t) - \lambda \nabla g(x(t))) = 0, \\ x(0) = x_0, \quad \dot{x}(0) = v_0, \end{cases}$$

for minimizing a smooth and convex function $g : H \rightarrow \mathbb{R}$ over the closed convex set $C \subset H$, where $x_0, v_0 \in H$, $\alpha, \lambda > 0$ and P_C is the projection operator onto C .

*School of Mathematical Sciences, University of Southampton, SO17 1BJ, Southampton, UK.
Email: t.v.phan@soton.ac.uk

Antipin proved in [4, Theorem 5] that if g is smooth and strongly convex then (1.2) has a unique equilibrium point, and this point is exponentially stable. An iterative analogue of the dynamical system (1.2) leads to a projection method with momentum

$$(1.3) \quad x_{k+1} = P_C(x_k - \lambda \nabla g(x_k) + \theta(x_k - x_{k-1})),$$

where $\theta > 0$, which was studied by Antipin in [3], see also [33]. A generalization of (1.3) for solving a composite optimization problem was investigated in [31], where linear convergence was obtained under a strong convexity condition.

It is well known that the (generalized) convexity of a smooth function g is characterized by the (generalized) monotonicity of its gradient ∇g (see e.g., [25]). Replacing ∇g in (1.2) by the general operator A , we consider the following dynamical system for solving $VI(A, C)$

$$(1.4) \quad \begin{cases} \ddot{x}(t) + \alpha(t)\dot{x}(t) + \beta(t)[x(t) - P_C(x(t) - \lambda Ax(t))] = 0, \\ x(0) = x_0, \quad \dot{x}(0) = v_0, \end{cases}$$

where $\alpha, \beta : [0, +\infty) \rightarrow [0, +\infty)$ be Lebesgue measurable functions. Under Lipschitz continuity of A , we prove that dynamical system (1.4) has a unique global solution. In addition, when A is strongly pseudo-monotone, we show that the trajectories generated by this dynamical system converge globally exponentially to the unique solution of the VI problem.

The second order dynamical systems (1.2) and (1.4) have a close connection with the heavy-ball method [3, 4, 8, 33] and numerical methods with inertial effect for solving convex optimization problems [1, 7] and monotone inclusions [5, 6, 12, 13, 29]. Motivated by the recent works of Attouch and Cabot [5, 6], we propose a discrete version of dynamical system (1.4), which leads to a relaxed inertial projection algorithm

$$(1.5) \quad \begin{cases} y_k = x_k + \theta_k(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho_k)y_k + \rho_k P_C(y_k - \lambda Ay_k). \end{cases}$$

We prove the linear convergence of the iterations to the unique solution when A is strongly pseudo-monotone and Lipschitz continuous. As far as we know, this result is new even in strongly monotone settings.

We also extend the obtained results to general monotone inclusion problems: Find $x^* \in H$ such that

$$0 \in Ax^* + Bx^*,$$

where A is monotone and $B : H \rightrightarrows H$ is a set-valued maximal monotone operator. In this case, we rediscover the relaxed inertial forward-backward (RIFB) [6], where the weak convergence was obtained provided that A is co-coercive. We investigate the linear convergence of (RIFB) when one of the involved operators A or B is strongly monotone, which is also new, to the best of our knowledge. Some other convergence results without inertial effect θ_k can be found in the literature. For example, in [19], the author obtained the linear convergence when A is co-coercive and B is strongly monotone.

It is worth emphasizing that the convergence analysis of algorithm (1.5) is carried out independently with the results of dynamical system (1.4). In this paper, we consider both continuous and discrete systems to provide a comprehensive study.

The remaining part of the paper is organized as follows. Section 2 describes the global exponential convergence of the trajectories of dynamical system (1.4). A

discrete version of the dynamical system and its linear convergence are investigated in Section 3. Section 4 discusses the possibility of extension to general monotone inclusions. Finally, some numerical experiments are reported in Section 5 to illustrate the theoretical results.

2. Global exponential convergence. In this section we will approach the solution set of $VI(A, C)$ from a continuous perspective by means of the trajectories generated by the dynamical system (1.4). We will recall some basic definitions and properties before proving the existence and uniqueness of the trajectories of (1.4). For each $x \in H$, there exists a unique point in C , (see, e.g., [28]), denoted by $P_C(x)$, such that

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

Some well-known properties of the metric projection $P_C : H \rightarrow C$ are given in the following lemma [18, 28].

LEMMA 2.1. *Assume that the set C is a closed convex subset of H . Then we have the following:*

(a) $P_C(\cdot)$ is a nonexpansive operator, i.e., for all $x, y \in H$, it holds that

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|.$$

(b) For any $x \in H$ and $y \in C$, it holds that

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

REMARK 2.1. For any $\lambda > 0$, x is a solution of $VI(A, C)$ if and only if $x = P_C(x - \lambda Ax)$; see, e.g. [17, 28].

One often considers $VI(A, C)$ with some additional properties imposed on the operator A such as Lipschitz continuity, (strong) monotonicity and (strong) pseudo-monotonicity of A . Let us recall some well-known definitions (see, e.g. [25]).

DEFINITION 2.2. *The operator $A : H \rightarrow H$ is*

(a) *strongly monotone with modulus $\gamma > 0$ on C if*

$$\langle Ax - Ay, x - y \rangle \geq \gamma \|x - y\|^2 \quad \forall x, y \in C;$$

(b) *strongly pseudo-monotone with modulus $\gamma > 0$ on C if*

$$\langle Ax, y - x \rangle \geq 0 \Rightarrow \langle Ay, y - x \rangle \geq \gamma \|x - y\|^2$$

for all $x, y \in C$;

(c) *Lipschitz continuous with modulus $L > 0$ on C if*

$$\|Ax - Ay\| \leq L \|x - y\| \quad \forall x, y \in C.$$

REMARK 2.2. If A is strongly pseudo-monotone and continuous then $VI(A, C)$ has a unique solution [27].

DEFINITION 2.3. A function $x : [0, b] \rightarrow H$ (where $b > 0$) is said to be absolutely continuous if one of the following equivalent properties holds:

(i) There exists an integrable function $y : [0, b] \rightarrow H$ such that

$$x(t) = x(0) + \int_0^t y(s) ds \quad \forall t \in [0, b].$$

- (ii) x is continuous and its distributional derivative \dot{x} is Lebesgue integrable on $[0, b]$.

Before proving the existence and uniqueness of the trajectory of (1.4), we need to recall the definition of its strong global solution.

DEFINITION 2.4. *We say that $x : [0, +\infty) \rightarrow H$ is a strong global solution of dynamical system (1.4) if the following properties are satisfied:*

- (i) $x, \dot{x} : [0, +\infty) \rightarrow H$ are locally absolutely continuous, in other word, absolutely continuous on each interval $[0, b]$ for $0 < b < +\infty$.
- (ii) $\ddot{x}(t) + \alpha(t)\dot{x}(t) + \beta(t)[x(t) - P_C(x(t) - \lambda Ax(t))] = 0$ for almost every $t \in [0, +\infty)$.
- (iii) $x(0) = x_0$ and $\dot{x}(0) = v_0$.

The existence and uniqueness of the trajectory of (1.4) are stated in the following result, where the proof is adapted from [12, Theorem 4].

THEOREM 2.5. *Let $\alpha, \beta : [0, +\infty) \rightarrow [0, +\infty)$ be Lebesgue measurable functions such that $\alpha, \beta \in L^1_{loc}([0, +\infty))$ (that is $\alpha, \beta \in L^1_{loc}([0, b])$ for every $0 < b < +\infty$). Let A be an L -Lipschitz continuous operator. Then for each $x_0, v_0 \in H$, there exists a unique strong global solution of the dynamical system (1.4).*

Proof. For all $x \in H$, if we define $T : H \rightarrow H$ by

$$Tx := x - P_C(x - \lambda Ax),$$

then dynamical system (1.4) can be rewritten equivalently as

$$(2.1) \quad \begin{cases} \ddot{x}(t) + \alpha(t)\dot{x}(t) + \beta(t)Tx(t) = 0, \\ x(0) = x_0, \quad \dot{x}(0) = v_0. \end{cases}$$

From Remark 2.1, we know that the solution set of $VI(A, C)$ coincides with the zeros set of T . By Lemma 2.1(a), the Cauchy-Schwarz inequality and the Lipschitz continuity of A , we obtain for all $x, \bar{x} \in H$ and $\lambda > 0$

$$\begin{aligned} \|Tx - T\bar{x}\| &= \|x - P_C(x - \lambda Ax) - \bar{x} + P_C(\bar{x} - \lambda A\bar{x})\| \\ &\leq \|x - \bar{x}\| + \|P_C(x - \lambda Ax) - P_C(\bar{x} - \lambda A\bar{x})\| \\ &\leq \|x - \bar{x}\| + \|x - \lambda Ax - \bar{x} + \lambda A\bar{x}\| \\ &\leq 2\|x - \bar{x}\| + \lambda\|Ax - A\bar{x}\| \\ &\leq (2 + \lambda L)\|x - \bar{x}\|. \end{aligned}$$

i.e., T is Lipschitz continuous with modulus $l = 2 + \lambda L > 0$. The dynamical system (2.1) can be equivalently written as the following first order dynamical system in the product space $H \times H$

$$(2.2) \quad \begin{cases} \dot{Y}(t) = F(t, Y(t)), \\ Y(0) = (u_0, v_0), \end{cases}$$

where

$$Y : [0, +\infty) \rightarrow H \times H, \quad Y(t) = (x(t), \dot{x}(t))$$

and

$$F : [0, +\infty) \times H \times H \rightarrow H \times H, \quad F(t, u, v) = (v, -\alpha(t)v - \beta(t)Tu).$$

We endow $H \times H$ with scalar product $\langle (u, v), (\bar{u}, \bar{v}) \rangle_{H \times H} = \langle u, \bar{u} \rangle + \langle v, \bar{v} \rangle$ and corresponding norm $\|(u, v)\|_{H \times H} = \sqrt{\|u\|^2 + \|v\|^2}$.

For arbitrary $u, \bar{u}, v, \bar{v} \in H$, by using the Lipschitz continuity of T , we obtain for all $t \geq 0$ that

$$\begin{aligned} \|F(t, u, v) - F(t, \bar{u}, \bar{v})\|_{H \times H} &= \sqrt{\|v - \bar{v}\|^2 + \|\alpha(t)(\bar{v} - v) + \beta(t)(T\bar{u} - Tu)\|^2} \\ &\leq \sqrt{(1 + 2\alpha^2(t))\|v - \bar{v}\|^2 + 2l^2\beta^2(t)\|\bar{u} - u\|^2} \\ &\leq \sqrt{(1 + 2\alpha^2(t)) + 2l^2\beta^2(t)} \|(u, \bar{u}) - (v, \bar{v})\|_{H \times H} \\ &\leq (1 + \sqrt{2}\alpha(t)) + l\sqrt{2}\beta(t) \|(u, \bar{u}) - (v, \bar{v})\|_{H \times H} \end{aligned}$$

As $\alpha, \beta \in L^1_{loc}([0, +\infty))$, the Lipschitz constant of $F(t, \cdot, \cdot)$ is locally integrable.

We show that

$$(2.3) \quad \forall u, v \in H, \forall b > 0, \quad F(\cdot, u, v) \in L^1([0, b], H \times H).$$

Indeed, given arbitrary $u, v \in H$ and $b > 0$, it holds that

$$\begin{aligned} \int_0^b \|F(t, u, v)\|_{H \times H} dt &= \int_0^b \sqrt{\|v\|^2 + \|\alpha(t)v + \beta(t)Tu\|^2} dt \\ &\leq \int_0^b \sqrt{(1 + 2\alpha^2(t))\|v\|^2 + 2\beta^2(t)\|Tu\|^2} dt \\ &\leq \int_0^b \left((1 + \sqrt{2}\alpha(t))\|v\| + \sqrt{2}\beta(t)\|Tu\| \right) dt \end{aligned}$$

and from here, by using the assumptions made on α, β , (2.3) follows.

Therefore, the existence and uniqueness of a strong global solution for (2.2) follow from the Cauchy-Lipschitz-Picard Theorem for first order dynamical systems (see, for example, [23, Proposition 6.2.1]). The conclusion is a consequence of the equivalence of (1.4), (2.1) and (2.2). \square

We investigate the exponential convergence of the trajectories $x(t)$ generated by dynamical system (1.4). The following result will play an important role in our convergence analysis.

PROPOSITION 2.6. *Let C be a nonempty closed convex subset of a real Hilbert space H . Let A be γ -strongly pseudo-monotone and L -Lipschitz continuous on C . Let x^* be the unique solution of $VI(A, C)$. For all $\lambda > 0$ and $x \in H$, denote $z := P_C(x - \lambda Ax)$. Then*

$$(2.4) \quad \langle x - z, x - x^* \rangle \geq \left(1 - \frac{\lambda L^2}{4\gamma}\right) \|x - z\|^2$$

and

$$(2.5) \quad \|x - x^*\| \leq \frac{1 + \lambda\gamma + \lambda L}{\lambda\gamma} \|x - z\|.$$

Proof. Since $z = P_C(x - \lambda Ax)$, from Lemma 2.1(b) we have

$$\langle x - \lambda Ax - z, y - z \rangle \leq 0 \quad \forall y \in C.$$

Substituting $y = x^* \in C$ into the last inequality and combining the result with the Lipschitz continuity of A we obtain

$$\begin{aligned}
\langle x - z, z - x^* \rangle &\geq \lambda \langle Ax, z - x^* \rangle \\
&= \lambda \langle Ax - Az, z - x^* \rangle + \lambda \langle Az, z - x^* \rangle \\
(2.6) \quad &\geq -\lambda L \|x - z\| \|z - x^*\| + \lambda \langle Az, z - x^* \rangle.
\end{aligned}$$

Since x^* is the unique solution of $VI(A, C)$ and $z \in C$, it holds that $\langle Ax^*, z - x^* \rangle \geq 0$. Then by the γ -strong pseudo-monotonicity of A we have $\langle Az, z - x^* \rangle \geq \gamma \|z - x^*\|^2$. It follows from (2.6) that

$$(2.7) \quad \langle x - z, z - x^* \rangle \geq \lambda \gamma \|z - x^*\|^2 - \lambda L \|x - z\| \|z - x^*\|.$$

Hence

$$\begin{aligned}
\langle x - z, x - x^* \rangle &= \langle x - z, x - z + z - x^* \rangle \\
&= \|x - z\|^2 + \langle x - z, z - x^* \rangle \\
&\geq \|x - z\|^2 + \lambda \gamma \|z - x^*\|^2 - \lambda L \|x - z\| \|z - x^*\| \\
&= \left(1 - \frac{\lambda L^2}{4\gamma}\right) \|x - z\|^2 \\
&\quad + \frac{\lambda L^2}{4\gamma} \|x - z\|^2 + \lambda \gamma \|z - x^*\|^2 - \lambda L \|x - z\| \|z - x^*\| \\
&\geq \left(1 - \frac{\lambda L^2}{4\gamma}\right) \|x - z\|^2,
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality in the last estimation. From (2.7) and the Cauchy-Schwarz inequality we have

$$\lambda \gamma \|z - x^*\|^2 \leq \|x - z\| \|z - x^*\| + \lambda L \|x - z\| \|z - x^*\|,$$

which implies

$$\|z - x^*\| \leq \frac{1 + \lambda L}{\lambda \gamma} \|x - z\|.$$

Hence

$$\|x - x^*\| \leq \|x - z\| + \|z - x^*\| \leq \frac{1 + \lambda \gamma + \lambda L}{\lambda \gamma} \|x - z\|. \quad \square$$

REMARK 2.3. *It is clear from (2.5) that the quantity $\|x - z\|$ provides an upper error bound for the distance from any arbitrary vector $x \in H$ to the unique solution x^* of $VI(F, C)$. As a direct corollary of (2.4), using the Cauchy-Schwarz inequality, we can see that if $\lambda \in (0, \frac{4\gamma}{L^2})$, then $\|x - z\|$ also provides a lower error bound for the distance from any vector $x \in H$ to the unique solution x^* of $VI(F, C)$. These error bounds are particularly useful for stopping criteria when designing algorithms for solving VIs.*

Since any strongly monotone operator is strongly pseudo-monotone, we have immediately the following corollary.

COROLLARY 2.7. *Let C be a nonempty closed convex subset of a real Hilbert space H . Let A be γ -strongly (pseudo)-monotone and L -Lipschitz continuous on C . Let*

x^* be the unique solution of $VI(A, C)$. For all $\lambda \in (0, \frac{4\gamma}{L^2})$ and $x \in H$, denote $z := P_C(x - \lambda Ax)$. Then the following error bounds hold

$$(2.8) \quad \left(1 - \frac{\lambda L^2}{4\gamma}\right) \|x - z\| \leq \|x - x^*\| \leq \frac{1 + \lambda\gamma + \lambda L}{\lambda\gamma} \|x - z\|.$$

We are now in the position to establish the main result of this section.

THEOREM 2.8. *Let C be a nonempty closed convex subset of a real Hilbert space H . Let A be γ -strongly pseudo-monotone and L -Lipschitz continuous on C . Let x^* be the unique solution of $VI(A, C)$, $0 < \lambda < \frac{4\gamma}{L^2}$ and $\kappa = 1 - \frac{\lambda L^2}{4\gamma} > 0$. Let $\alpha, \beta : [0, +\infty) \rightarrow [0, +\infty)$ be locally absolutely continuous functions fulfilling for every $t \in [0, +\infty)$*

$$(i) \quad 1 < \alpha \leq \alpha(t) \leq \frac{\kappa \lambda^2 \gamma^2}{(1 + \lambda\gamma + \lambda L)^2} \beta(t) + 1;$$

$$(ii) \quad \dot{\alpha}(t) \leq 0 \text{ and } \frac{d}{dt} \left(\frac{\alpha(t)}{\beta(t)} \right) \leq 0;$$

$$(iii) \quad \alpha^2(t) - \alpha(t) - \frac{2\beta(t)}{\kappa} \geq 0.$$

Then the trajectories $x(t)$ generated by dynamical system (1.4) converge exponentially to x^ as $t \rightarrow \infty$, i.e., there exist positive numbers μ, η such that*

$$\|x(t) - x^*\| \leq \mu \|x(0) - x^*\| e^{-\eta t} \quad \forall t \geq 0.$$

Proof. Consider for every $t \in [0, +\infty)$ the function $h(t) = \frac{1}{2} \|x(t) - x^*\|^2$. Then

$$\dot{h}(t) = \langle x(t) - x^*, \dot{x}(t) \rangle, \quad \ddot{h}(t) = \|\dot{x}(t)\|^2 + \langle x(t) - x^*, \ddot{x}(t) \rangle.$$

Setting $z(t) := P_C(x(t) - \lambda Ax(t))$ and taking into account of (1.4) we obtain for every $t \in [0, +\infty)$ that

$$\ddot{h}(t) + \alpha(t)\dot{h}(t) + \beta(t) \langle x(t) - x^*, x(t) - z(t) \rangle = \|\dot{x}(t)\|^2,$$

which, together with Proposition 2.6 implies

$$\ddot{h}(t) + \alpha(t)\dot{h}(t) + \kappa\beta(t)\|x(t) - z(t)\|^2 \leq \|\dot{x}(t)\|^2,$$

where $\kappa = 1 - \frac{\lambda L^2}{4\gamma} > 0$. Again it follows from (1.4) that

$$(2.9) \quad \ddot{h}(t) + \alpha(t)\dot{h}(t) + \frac{\kappa}{2}\beta(t)\|x(t) - z(t)\|^2 + \frac{\kappa}{2\beta(t)} \|\ddot{x}(t) + \alpha(t)\dot{x}(t)\|^2 \leq \|\dot{x}(t)\|^2.$$

Applying (2.5), we obtain from the last inequality

$$(2.10) \quad \begin{aligned} & \ddot{h}(t) + \alpha(t)\dot{h}(t) + \kappa_1\beta(t)h(t) + \frac{\kappa}{2\beta(t)} \|\ddot{x}(t)\|^2 \\ & + \left(\frac{\kappa\alpha^2(t)}{2\beta(t)} - 1 \right) \|\dot{x}(t)\|^2 + \frac{\kappa\alpha(t)}{\beta(t)} \langle \ddot{x}(t), \dot{x}(t) \rangle \leq 0, \end{aligned}$$

where $\kappa_1 = \frac{\kappa\lambda^2\gamma^2}{(1 + \lambda\gamma + \lambda L)^2}$. Since $\frac{d}{dt} \|\dot{x}(t)\|^2 = 2 \langle \ddot{x}(t), \dot{x}(t) \rangle$, setting for every $t \in [0, +\infty)$

$$a(t) := \kappa_1\beta(t), \quad b(t) := \frac{\kappa\alpha(t)}{2\beta(t)}, \quad c(t) := \frac{\kappa\alpha^2(t)}{2\beta(t)} - 1, \quad u(t) := \|\dot{x}(t)\|^2$$

and eliminating a nonnegative term $\frac{\kappa}{2\beta(t)} \|\ddot{x}(t)\|^2$ in (2.10) we obtain

$$(2.11) \quad \ddot{h}(t) + \alpha(t)\dot{h}(t) + a(t)h(t) + b(t)\dot{u}(t) + c(t)u(t) \leq 0.$$

Multiplying both sides of (2.11) with $e^t > 0$, and using the identities

$$\begin{aligned} e^t \ddot{h}(t) &= \frac{d}{dt} (e^t \dot{h}(t)) - e^t \dot{h}(t) \\ e^t \dot{h}(t) &= \frac{d}{dt} (e^t h(t)) - e^t h(t) \\ e^t \dot{u}(t) &= \frac{d}{dt} (e^t u(t)) - e^t u(t) \end{aligned}$$

we obtain

$$\begin{aligned} & \frac{d}{dt} (e^t \dot{h}(t)) + (\alpha(t) - 1) \frac{d}{dt} (e^t h(t)) + (a(t) + 1 - \alpha(t)) e^t h(t) \\ (2.12) \quad & + b(t) \frac{d}{dt} (e^t u(t)) + (c(t) - b(t)) e^t u(t) \leq 0. \end{aligned}$$

From assumptions (i) and (iii) we have

$$a(t) + 1 - \alpha(t) \geq 0, \quad c(t) - b(t) \geq 0 \quad \forall t \in [0, +\infty).$$

Hence from (2.12) we can write

$$(2.13) \quad \frac{d}{dt} (e^t \dot{h}(t)) + (\alpha(t) - 1) \frac{d}{dt} (e^t h(t)) + b(t) \frac{d}{dt} (e^t u(t)) \leq 0.$$

Since

$$\begin{aligned} (\alpha(t) - 1) \frac{d}{dt} (e^t h(t)) &= \frac{d}{dt} [(\alpha(t) - 1) e^t h(t)] - \dot{\alpha}(t) e^t h(t) \\ b(t) \frac{d}{dt} (e^t u(t)) &= \frac{d}{dt} (b(t) e^t u(t)) - \dot{b}(t) e^t u(t), \end{aligned}$$

we have from (2.13)

$$(2.14) \quad \frac{d}{dt} (e^t \dot{h}(t)) + \frac{d}{dt} [(\alpha(t) - 1) e^t h(t)] - \dot{\alpha}(t) e^t h(t) + \frac{d}{dt} (b(t) e^t u(t)) - \dot{b}(t) e^t u(t) \leq 0.$$

By assumption (ii), $\dot{\alpha}(t) \leq 0$ and $\dot{b}(t) \leq 0$ for all $t \in [0, +\infty)$. Therefore, we have from the last inequality

$$\frac{d}{dt} [e^t \dot{h}(t) + (\alpha(t) - 1) e^t h(t) + b(t) e^t u(t)] \leq 0.$$

This implies that the function

$$t \rightarrow e^t \dot{h}(t) + (\alpha(t) - 1) e^t h(t) + b(t) e^t u(t)$$

is monotonically decreasing, hence there exists $M > 0$ such that

$$e^t \dot{h}(t) + (\alpha(t) - 1) e^t h(t) + b(t) e^t u(t) \leq M.$$

Since $b(t), u(t) \geq 0$, we get

$$\dot{h}(t) + (\alpha(t) - 1) h(t) \leq M e^{-t},$$

hence

$$\dot{h}(t) + (\alpha - 1) h(t) \leq M e^{-t}$$

for every $t \in [0, \infty)$. This implies that

$$\frac{d}{dt} \left[e^{(\alpha-1)t} h(t) \right] \leq M e^{(\alpha-2)t}$$

for every $t \in [0, \infty)$. By integration, we have

(i) if $1 < \alpha < 2$ then

$$0 \leq h(t) \leq \left(h(0) + \frac{M}{\alpha-2} \right) e^{-(\alpha-1)t};$$

(ii) if $2 < \alpha$ then

$$0 \leq h(t) \leq h(0) e^{-(\alpha-1)t} + \frac{M}{2-\alpha} e^{-t} \leq \left(h(0) + \frac{M}{2-\alpha} \right) e^{-t};$$

(ii) if $\alpha = 2$ then

$$0 \leq h(t) \leq (h(0) + Mt) e^{-t}.$$

This implies that $x(t)$ converges exponentially to x^* . \square

REMARK 2.4. We notice that it is easy to find functions α, β satisfying assumptions (i)-(iii) in Theorem 2.8. For example, if we choose $\alpha(t) = \alpha + \frac{1}{t+1}$ and $\beta(t) = \beta - \frac{1}{t+1}$ for all $t \in [0, +\infty)$ and $\alpha > 1$, then (ii) is fulfilled. Assumption (iii) is equivalent to

$$\alpha^2 - \alpha - \frac{2}{\kappa} \beta + \frac{1}{(t+1)^2} + \frac{1}{t+1} + \frac{2}{\kappa(t+1)} \geq 0,$$

which can be guaranteed if

$$\alpha^2 - \alpha - \frac{2}{\kappa} \beta \geq 0,$$

or

$$\alpha \geq 1 + \sqrt{1 + \frac{8\beta}{\kappa}}.$$

Assumption (i) reads as

$$\alpha + \frac{1}{t+1} \leq \kappa_1 \left(\beta - \frac{1}{t+1} \right) + 1,$$

i.e.,

$$\alpha \leq \kappa_1 \beta - \frac{\kappa_1 + 1}{t+1} + 1,$$

which is guaranteed whenever

$$\alpha \leq \kappa_1 \beta - (\kappa_1 + 1) + 1 = (\beta - 1) \kappa_1.$$

Therefore, to fulfill assumptions (i)-(iii), it is sufficient to choose β (large enough) and α satisfying

$$(2.15) \quad 1 + \sqrt{1 + \frac{8\beta}{\kappa}} \leq \alpha \leq (\beta - 1) \kappa_1.$$

Also any constant functions $\alpha(t) = \alpha$ and $\beta(t) = \beta$ for all $t \in [0, +\infty)$ with α and β satisfying

$$\sqrt{1 + \frac{8\beta}{\kappa}} \leq \alpha \leq \beta \kappa_1,$$

fulfill assumptions (i)-(iii).

3. Linear convergence of a discrete system. A finite-difference scheme for (1.4) with respect to the time variable t , with stepsize $h_k > 0$, relaxation variable $\beta_k > 0$, damping variable $\alpha_k > 0$, and initial points x_0 and x_1 yields the following iterative scheme:

$$(3.1) \quad \frac{1}{h_k^2} (x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha_k}{h_k} (x_k - x_{k-1}) + \beta_k (y_k - P_C(y_k - \lambda A y_k)) = 0,$$

where y_k is an extrapolated point from x_k and x_{k-1} that will be chosen later. We note that since $T := I - P_C(I - \lambda A(\cdot))$ is Lipschitz continuous, there is some flexibility in this choice. We can write (3.1) as

$$x_{k+1} = x_k + (1 - \alpha_k h_k)(x_k - x_{k-1}) - h_k^2 \beta_k (y_k - P_C(y_k - \lambda A y_k)).$$

Setting $\theta_k = 1 - \alpha_k h_k$, $\rho_k = h_k^2 \beta_k$ and choosing $y_k := x_k + \theta_k(x_k - x_{k-1})$ we can write the above scheme as

$$(3.2) \quad \begin{cases} y_k = x_k + \theta_k(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho_k)y_k + \rho_k P_C(y_k - \lambda A y_k), \end{cases}$$

which is a relaxed inertial projection algorithm (RIPA). In this section, we will investigate the convergence properties of (3.2). For the sake of simplicity, we only consider the case where all parameters are constants, i.e., $\theta_k = \theta$ and $\rho_k = \rho$ for all k .

$$(3.3) \quad (\mathbf{RIPA}) \begin{cases} y_k = x_k + \theta(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho)y_k + \rho P_C(y_k - \lambda A y_k). \end{cases}$$

We also make the following assumptions on the parameters:

(A1): $\eta > 0$ and $0 < \lambda < \frac{2\gamma}{\eta L^2}$;

(A2):

$$(3.4) \quad 0 < \rho < \min \left\{ \frac{1}{1-q}, 1 + q \left(1 - \frac{1}{\eta} \right) \right\}$$

where $q := \frac{1}{1 + \lambda(2\gamma - \lambda\eta L^2)}$.

(A3):

$$(3.5) \quad 0 \leq \theta \leq \min \left\{ \frac{1 - \rho + \rho q}{3}, \frac{\left(1 - \frac{1}{\eta}\right) q + 1 - \rho}{\left(1 - \frac{1}{\eta}\right) q + 1 + \rho} \right\}.$$

REMARK 3.1. Note that (A2) allows overrelaxation, i.e. $\rho > 1$, which can accelerate the convergence speed in certain examples. If $\rho = 1$ and (A1) is fulfilled, then (A2) holds for any $\eta > 1$ and (A3) becomes

$$0 \leq \theta \leq \min \left\{ \frac{q}{3}, \frac{\left(1 - \frac{1}{\eta}\right) q}{\left(1 - \frac{1}{\eta}\right) q + 2} \right\}.$$

In this case, (3.2) reduces to inertial projection algorithm (IPA) for solving $VI(A, C)$

$$(\mathbf{IPA}) \begin{cases} y_k = x_k + \theta(x_k - x_{k-1}) \\ x_{k+1} = P_C(y_k - \lambda A y_k). \end{cases}$$

If the inertial term $\theta = 0$, then we recover the projection algorithm (PA) for solving $VI(A, C)$ considered in [26]

$$(3.6) \quad x_{k+1} = P_C(x_k - \lambda A x_k).$$

REMARK 3.2. From (A3), we see that the upper bound for θ is

$$\theta(\rho) = \min \left\{ \frac{1 - \rho + \rho q}{3}, \frac{\left(1 - \frac{1}{\eta}\right) q + 1 - \rho}{\left(1 - \frac{1}{\eta}\right) q + 1 + \rho} \right\}.$$

Hence, there is a trade-off between the inertial parameter θ and the relaxed parameter ρ , i.e. θ is decreasing when ρ is increasing and vice versa. If $\rho \searrow 0$, then $\theta(\rho) \nearrow 1/3$. If $\rho \nearrow \min \left\{ \frac{1}{1-q}, 1 + q \left(1 - \frac{1}{\eta}\right) \right\}$, then $\theta(\rho) \searrow 0$.

The linear convergence of scheme (RIPA) is established as follows.

THEOREM 3.1. Let C be a nonempty closed convex subset of a real Hilbert space H . Let A be γ -strongly pseudo-monotone and L -Lipschitz continuous on C . Let the parameters λ, ρ, θ be such that assumptions (A1), (A2) and (A3) are fulfilled. Then the sequence $\{x_k\}$ generated by (3.3) converges linearly to the unique solution x^* of $VI(A, C)$.

Proof. Setting $z_k := P_C(y_k - \lambda A y_k) \in C$, similarly to (2.7) we can deduce

$$(3.7) \quad \langle y_k - z_k, z_k - x^* \rangle \geq \lambda \gamma \|z_k - x^*\|^2 - \lambda L \|y_k - z_k\| \|z_k - x^*\|.$$

Hence

$$\begin{aligned} -2\lambda \gamma \|z_k - x^*\|^2 + 2\lambda L \|y_k - z_k\| \|z_k - x^*\| &\geq 2 \langle y_k - z_k, x^* - z_k \rangle \\ &= \|z_k - x^*\|^2 - \|y_k - x^*\|^2 + \|y_k - z_k\|^2, \end{aligned}$$

which implies

$$\begin{aligned} (1 + 2\lambda \gamma) \|z_k - x^*\|^2 &\leq \|y_k - x^*\|^2 - \|y_k - z_k\|^2 + 2\lambda L \|y_k - z_k\| \|z_k - x^*\| \\ &\leq \|y_k - x^*\|^2 - \|y_k - z_k\|^2 + \frac{1}{\eta} \|y_k - z_k\|^2 + \eta \lambda^2 L^2 \|z_k - x^*\|^2 \\ &= \|y_k - x^*\|^2 - \left(1 - \frac{1}{\eta}\right) \|y_k - z_k\|^2 + \eta \lambda^2 L^2 \|z_k - x^*\|^2, \end{aligned}$$

or equivalently,

$$(3.8) \quad [1 + \lambda(2\gamma - \lambda \eta L^2)] \|z_k - x^*\|^2 \leq \|y_k - x^*\|^2 - \left(1 - \frac{1}{\eta}\right) \|y_k - z_k\|^2.$$

Therefore,

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &= \|(1 - \rho)y_k + \rho z_k - x^*\|^2 \\ &= \|(1 - \rho)(y_k - x^*) + \rho(z_k - x^*)\|^2 \\ &= (1 - \rho)\|y_k - x^*\|^2 + \rho\|z_k - x^*\|^2 - \rho(1 - \rho)\|y_k - z_k\|^2 \\ &\leq (1 - \rho)\|y_k - x^*\|^2 + \frac{\rho}{1 + \lambda(2\gamma - \lambda \eta L^2)} \|y_k - x^*\|^2 \\ &\quad - \left(1 - \frac{1}{\eta}\right) \frac{\rho}{(1 + \lambda(2\gamma - \lambda \eta L^2))} \|y_k - z_k\|^2 - \rho(1 - \rho)\|y_k - z_k\|^2 \\ (3.9) \quad &= \mu_1 \|y_k - x^*\|^2 - \mu_2 \rho \|y_k - z_k\|^2, \end{aligned}$$

with

$$\mu_1 := 1 - \rho + \frac{\rho}{1 + \lambda(2\gamma - \lambda\eta L^2)} = 1 - \rho + \rho q \in (0, 1)$$

and

$$\mu_2 := \left(1 - \frac{1}{\eta}\right) \frac{1}{1 + \lambda(2\gamma - \lambda\eta L^2)} + 1 - \rho = \left(1 - \frac{1}{\eta}\right) q + 1 - \rho > 0$$

by assumptions (A1) and (A2).

On the one hand

$$\begin{aligned} \|y_k - x^*\|^2 &= \|(1 + \theta)(x_k - x^*) - \theta(x_{k-1} - x^*)\|^2 \\ (3.10) \quad &= (1 + \theta)\|x_k - x^*\|^2 - \theta\|x_{k-1} - x^*\|^2 + \theta(1 + \theta)\|x_k - x_{k-1}\|^2. \end{aligned}$$

On the other hand

$$\begin{aligned} \rho^2 \|y_k - z_k\|^2 &= \|x_{k+1} - y_k\|^2 \\ &= \|x_{k+1} - x_k - \theta(x_k - x_{k-1})\|^2 \\ &= \|x_{k+1} - x_k\|^2 + \theta^2 \|x_k - x_{k-1}\|^2 - 2\theta \langle x_{k+1} - x_k, x_k - x_{k-1} \rangle \\ &\geq \|x_{k+1} - x_k\|^2 + \theta^2 \|x_k - x_{k-1}\|^2 - 2\theta \|x_{k+1} - x_k\| \|x_k - x_{k-1}\| \\ &\geq \|x_{k+1} - x_k\|^2 + \theta^2 \|x_k - x_{k-1}\|^2 - \theta \|x_{k+1} - x_k\|^2 - \theta \|x_k - x_{k-1}\|^2 \\ (3.11) \quad &= (1 - \theta)\|x_{k+1} - x_k\|^2 - \theta(1 - \theta)\|x_k - x_{k-1}\|^2. \end{aligned}$$

Combining (3.9), (3.10) and (3.11), we obtain

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \mu_1(1 + \theta)\|x_k - x^*\|^2 - \mu_1\theta\|x_{k-1} - x^*\|^2 + \mu_1\theta(1 + \theta)\|x_k - x_{k-1}\|^2 \\ &\quad + \frac{\mu_2\theta(1 - \theta)}{\rho}\|x_k - x_{k-1}\|^2 - \frac{\mu_2(1 - \theta)}{\rho}\|x_{k+1} - x_k\|^2. \end{aligned}$$

Since $\mu_1 \in (0, 1)$, the last inequality implies

$$\begin{aligned} &\|x_{k+1} - x^*\|^2 - \theta\|x_k - x^*\|^2 + \frac{\mu_2(1 - \theta)}{\rho}\|x_{k+1} - x_k\|^2 \\ &\leq \mu_1 \left[\|x_k - x^*\|^2 - \theta\|x_{k-1} - x^*\|^2 + \frac{\mu_2(1 - \theta)}{\rho}\|x_k - x_{k-1}\|^2 \right] \\ (3.12) \quad &- \left(\frac{\mu_1\mu_2(1 - \theta)}{\rho} - \frac{\mu_2\theta(1 - \theta)}{\rho} - \mu_1\theta(1 + \theta) \right) \|x_k - x_{k-1}\|^2. \end{aligned}$$

Setting

$$a_k := \|x_k - x^*\|^2 - \theta\|x_{k-1} - x^*\|^2 + \frac{\mu_2(1 - \theta)}{\rho}\|x_k - x_{k-1}\|^2,$$

we can prove that $a_k \geq 0$ for all $k \geq 0$. Indeed, from assumption (A3) we have $\theta \leq \frac{\mu_2}{\mu_2 + 2\rho}$, hence $\theta \leq \frac{\mu_2(1 - \theta)}{2\rho}$. Therefore, using the following inequality with $\kappa > 0$

$$a^2 + \kappa b^2 \geq \min\{1, \kappa\} (a^2 + b^2) \geq \min\{1/2, \kappa/2\} (a - b)^2,$$

we obtain

$$a_k \geq \min\left\{\frac{1}{2}, \frac{\mu_2(1 - \theta)}{2\rho}\right\} \|x_{k-1} - x^*\|^2 - \theta\|x_{k-1} - x^*\|^2 \geq 0.$$

Moreover from (A3) we also have

$$0 \leq \theta \leq \min \left\{ \frac{\mu_1}{3}, \frac{\mu_2}{\mu_2 + 2\rho} \right\} < \frac{1}{3}.$$

Hence

$$\begin{aligned} \frac{\mu_2\theta(1-\theta)}{\rho} &\leq \frac{\mu_1\mu_2\theta(1-\theta)}{3\rho} \\ \mu_1\theta(1+\theta) &< \mu_1 \frac{\mu_2(1-\theta)}{2\rho} \frac{4}{3} = \frac{2\mu_1\mu_2(1-\theta)}{3\rho}, \end{aligned}$$

which implies

$$\frac{\mu_1\mu_2(1-\theta)}{\rho} - \frac{\mu_2\theta(1-\theta)}{\rho} - \mu_1\theta(1+\theta) \geq \sigma > 0,$$

where $\sigma := \frac{2\mu_1\mu_2(1-\theta)}{3\rho} - \mu_1\theta(1+\theta) > 0$. Hence it follows from (3.12) that

$$a_{k+1} \leq \mu_1 a_k - \sigma \|x_k - x_{k-1}\|^2 \leq \mu_1 a_k,$$

for all $k \geq 0$, from which we deduce

$$a_k \leq a_0 \mu_1^k \quad \text{and} \quad \sigma \|x_k - x_{k-1}\|^2 \leq \mu_1 a_k \leq a_0 \mu_1^{k+1}$$

i.e., $\{a_k\}$ and $\{\|x_k - x_{k-1}\|\}$ converge linearly to 0, and this immediately implies that the sequence $\{x_k\}$ converges linearly to the unique solution x^* . \square

REMARK 3.3. In the case $\rho = 1$, we obtain the convergence rate $q \in (0, 1)$ for (IPA), which is a generalization of the basic projection algorithm (3.6) considered in [26].

Finally, we emphasize that the result established in Theorem 3.1 seems to be new even in the strongly monotone case.

COROLLARY 3.2. Let C be a nonempty closed convex subset of a real Hilbert space H . Let A be γ -strongly monotone and L -Lipschitz continuous on C . Let the parameters λ, ρ, θ be such that assumptions (A1), (A2) and (A3) are fulfilled. Then the sequence $\{x_k\}$ generated by (3.3) converges linearly to the unique solution x^* of $VI(A, C)$.

4. Generalize to monotone inclusions. The $VI(A, C)$ (1.1) can be equivalently rewritten as

$$(4.1) \quad \text{find } x^* \in H \text{ such that } 0 \in Ax^* + N_C x^*,$$

where N_C is the normal cone of C at x^* , which is a maximal monotone operator (see e.g. [10] for the definition and properties). Replacing N_C in (4.1) by a general maximal monotone set-valued operator $B : H \rightrightarrows H$, we will discuss in this section the following monotone inclusion problem:

$$(4.2) \quad \text{find } x^* \in H \text{ such that } 0 \in Ax^* + Bx^*.$$

In this general form, the projection P_C is replaced by the resolvent $J_{\lambda B} := (I + \lambda B)^{-1}$ of the operator λB , where I is the identity operator and $\lambda > 0$. The dynamical system (1.4) has the form

$$(4.3) \quad \begin{cases} \ddot{x}(t) + \alpha(t)\dot{x}(t) + \beta(t)[x(t) - J_{\lambda B}(x(t) - \lambda Ax(t))] = 0, \\ x(0) = x_0, \quad \dot{x}(0) = v_0, \end{cases}$$

whose the asymptotic analysis was discussed in [12, 13]. The corresponding discretization of (4.3) leads to the following relaxed inertial forward-backward algorithm

$$(4.4) \quad (\mathbf{RIFB}) \begin{cases} y_k = x_k + \theta(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho)y_k + \rho J_{\lambda B}(I - \lambda A)y_k. \end{cases}$$

The weak convergence of (RIFB) was studied by Attouch and Cabot in [6] when A is co-coercive, i.e., there exists $L > 0$ such that

$$\langle Ax - Ay, x - y \rangle \geq \frac{1}{L} \|Ax - Ay\|^2 \quad \forall x, y \in H.$$

In particular, co-coercivity holds when $A = \nabla f$ is L -Lipschitz continuous, where ∇f is the gradient of a convex and differentiable function f . It is not clear whether the linear convergence of (RIFB) is still guaranteed when A is strongly pseudo-monotone. We conject that the answer is negative. The reason could be that while the sum of a monotone operator and a (strongly) monotone operator is (strongly) monotone [10], this is not the case for a (strongly) pseudo-monotone operator. For example, the operator

$$A(x_1, x_2) := (x_1^2 + x_2^2)(-x_2, x_1)^T$$

is pseudo-monotone but $A + \epsilon I$ is not (pseudo)-monotone for any $\epsilon > 0$ (see [35, Counterexample 2.1]). Hence, the theory of maximal monotone operators cannot be applied in this setting. Nevertheless, the linear convergence of (RIFB) holds when either A or B is strongly monotone. Indeed, we have the following new convergence result.

THEOREM 4.1. *Let A be monotone and L -Lipschitz continuous, B be maximal monotone and either A or B be γ -strongly monotone. Let the parameters λ, ρ, θ be such that assumptions (A1), (A2) and (A3) are fulfilled. Then the sequence $\{x_k\}$ generated by (4.4) converges linearly to the unique solution x^* of the monotone inclusion (4.2).*

Proof. Having a careful look at the proof of Theorem 3.1, we see that it is sufficient to prove that

$$(4.5) \quad \langle y_k - z_k, z_k - x^* \rangle \geq \lambda \gamma \|z_k - x^*\|^2 - \lambda L \|y_k - z_k\| \|z_k - x^*\|,$$

where $z_k = J_{\lambda B}(I - \lambda A)y_k$, then the rest of the proof follows. Since $z_k = J_{\lambda B}(I - \lambda A)y_k$ we have

$$y_k - \lambda A y_k \in (I + \lambda B)z_k,$$

or equivalently

$$\frac{1}{\lambda}(y_k - z_k) - A y_k \in B z_k.$$

We consider two cases:

(i) A is γ -strongly monotone. Since $-Ax^* \in Bx^*$ and B is monotone, we have

$$(4.6) \quad \left\langle \frac{1}{\lambda}(y_k - z_k) - A y_k + A x^*, z_k - x^* \right\rangle \geq 0.$$

It follows from the last inequality, and the strong monotonicity and Lipschitz continuity of A that

$$\begin{aligned}
(4.7) \quad \frac{1}{\lambda} \langle y_k - z_k, z_k - x^* \rangle &\geq \langle Ay_k - Ax^*, z_k - x^* \rangle \\
&= \langle Ay_k - Az_k, z_k - x^* \rangle + \langle Az_k - Ax^*, z_k - x^* \rangle \\
&\geq -\|Ay_k - Az_k\| \|z_k - x^*\| + \gamma \|z_k - x^*\|^2 \\
&\geq -L \|y_k - z_k\| \|z_k - x^*\| + \gamma \|z_k - x^*\|^2,
\end{aligned}$$

which implies (4.5).

(ii) B is γ -strongly monotone. Then (4.6) can be rewritten as

$$\left\langle \frac{1}{\lambda} (y_k - z_k) - Ay_k + Ax^*, z_k - x^* \right\rangle \geq \gamma \|z_k - x^*\|^2.$$

Hence (4.7) is obtained without requiring the strong monotonicity of A . \square

REMARK 4.1. In [19], the author provided the linear convergence analysis for a relaxed version of the forward-backward method without inertial effect, i.e., $\theta = 0$ in (4.4). The result is obtained when A is co-coercive and B is strongly monotone. Since co-coercivity implies monotonicity and Lipschitz continuity, the result we obtained in case (ii) above is stronger.

From the proofs of Proposition 2.6 and Theorem 4.1, it is clear that the error bounds stated in Corollary 2.7 are still valid for monotone inclusion (4.2).

COROLLARY 4.2. Let A be monotone and L -Lipschitz continuous, B be maximal monotone and either A or B be γ -strongly monotone. Let x^* be the unique solution of the inclusion (4.2). For all $\lambda \in (0, \frac{4\gamma}{L^2})$ and $x \in H$, denote $z := J_{\lambda B}(x - \lambda Ax)$. Then the following error bounds hold

$$\left(1 - \frac{\lambda L^2}{4\gamma}\right) \|x - z\| \leq \|x - x^*\| \leq \frac{1 + \lambda\gamma + \lambda L}{\lambda\gamma} \|x - z\|.$$

We consider now the optimization problem of the form

$$(4.8) \quad \min_{x \in H} f(x) + g(x),$$

where $f : H \rightarrow \mathbb{R}$ is a strongly convex and differentiable function with L -Lipschitz continuous gradient for $L > 0$ and $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function. The set of minimizers of (4.8) coincides with the solution set of the monotone inclusion problem

$$(4.9) \quad \text{find } x^* \in H \text{ such that } 0 \in \nabla f(x^*) + \partial g(x^*).$$

Here, ∇f is the gradient of f and $\partial g : H \rightrightarrows H$, defined by

$$\partial g(x) = \{u \in H : g(y) \geq g(x) + \langle u, y - x \rangle \ \forall y \in H\},$$

if $g(x) \in \mathbb{R}$ and $\partial g(x) = \emptyset$, otherwise, denotes the convex subdifferential of g , which is a maximally monotone operator, provided that g is proper, convex and lower semicontinuous. We notice that, for $\lambda > 0$, the resolvent of $\lambda \partial g$ is given by $J_{\lambda \partial g} = \text{prox}_{\lambda g}$, where $\text{prox}_{\lambda g} : H \rightarrow H$,

$$\text{prox}_{\lambda g}(x) = \arg \min \left\{ g(y) + \frac{1}{2\lambda} \|y - x\|^2, \quad y \in H \right\}$$

denotes the proximal operator of λg . Let us also notice that f is γ -strongly convex if and only if ∇f is γ -strongly monotone. Then the (RIFB) method for solving (4.8) is the relaxed inertial proximal method read as

$$(4.10) \quad (\mathbf{RIPM}) \quad \begin{cases} y_k = x_k + \theta(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho)y_k + \rho \operatorname{prox}_{\lambda g}(y_k - \lambda \nabla f(y_k)). \end{cases}$$

The following result is a direct consequence of Theorem 4.1.

THEOREM 4.3. *Let f be a γ -strongly convex and differentiable function with L -Lipschitz gradient and let g be a proper, convex and lower semicontinuous function. Let the parameters λ, ρ, θ be such that assumptions (A1), (A2) and (A3) are fulfilled. Then the sequence $\{x_k\}$ generated by (4.10) converges linearly to the unique solution x^* of the optimization problem (4.8).*

REMARK 4.2. *The linear convergence of another forward-backward type method with inertial effect for solving (4.8) was studied in [31]. The algorithm was named iPiasco and reads as*

$$x_{k+1} = \operatorname{prox}_{\lambda g}(x_k - \lambda \nabla f(x_k) + \beta(x_k - x_{k-1})),$$

which is a generalization of the heavy-ball method studied by Polyak [33].

5. Examples and numerical illustrations. In this section, we consider some numerical results to illustrate the global exponential convergence of dynamical system (1.4) and the linear convergence of (RIPA). Codes are implemented in MATLAB 2019b running on a Macbook Pro laptop with an Intel core CPU i7 at 2.6 GHz and 16 GB memory.

In the first experiment, we focus on a class of strongly pseudo-monotone (but not monotone) operators. The stopping condition is $\|x(t) - x^*\| \leq \epsilon$ for all test problems, where $\epsilon = 10^{-5}$ and x^* is the unique solution of $VI(A, C)$. Let $A : H \rightarrow H$ be defined as

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

where $p \in H$, $M : H \rightarrow H$ is a linear bounded operator satisfying

$$(5.1) \quad \langle Mx, x \rangle \geq \gamma \|x\|^2 \quad \forall x \in H,$$

and $g : H \rightarrow (0, +\infty)$ is a function taking positive values, i.e., $g(x) > \kappa > 0$ for all $x \in H$. We show that A is strongly pseudo-monotone on H . Indeed, let $x, y \in H$ be such that $\langle Ax, y - x \rangle \geq 0$. Since $g(x) > \kappa > 0$, we have

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

Hence

$$\begin{aligned} \langle Ay, y - x \rangle &= g(y) \langle My + p, y - x \rangle \\ &\geq g(y) (\langle My + p, y - x \rangle - \langle Mx + p, y - x \rangle) \\ &= g(y) \langle M(y - x), y - x \rangle \\ &\geq \kappa \gamma \|x - y\|^2, \end{aligned}$$

which leads to the desired conclusion. In general, A is not monotone, as seen in examples below.

EXAMPLE 5.1. Let $A : H \rightarrow H$ be defined as

$$(5.2) \quad Ax = \left(e^{-\|x\|^2} + q \right) Mx,$$

where $q > 0$ and $M : H \rightarrow H$ satisfying (5.1), then A is strongly pseudo-monotone. A direct computation shows that there exists $L > 0$ such that $\|\nabla Ax\| \leq L$ for every $x \in H$, hence by the mean value theorem, A is Lipschitz continuous. In our illustration, we choose

$$C = \{x \in [-5, 5]^3 : x_1 + x_2 + x_3 = 0\} \subseteq \mathbb{R}^3,$$

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

and $q = 2$, then A is γ -strongly pseudo-monotone on \mathbb{R}^3 with constant $\gamma := q \cdot \lambda_{\min} \approx 0.764$, where λ_{\min} is the smallest eigenvalue of M , and Lipschitz continuous with constant $L \approx 5.0679$. Since for $x = (-1, 0, 0)^T, y = (-2, 0, 0)^T \in \mathbb{R}^3$

$$\langle Ax - Ay, x - y \rangle = -0.1312 < 0,$$

A is not monotone. It is clear that $x^* = 0$ is the unique solution of $VI(A, C)$. According to Theorem 2.8, the trajectory of dynamical system (1.4) globally exponentially converges to the unique solution x^* . This is numerically confirmed in Figure 5.1 with $\alpha(t) = \alpha, \beta(t) = \beta$ satisfying (2.15), $\eta = 1.5, \rho = 1, x_0 = x_1 = (-3, 2, 3)$ and $\lambda = \frac{\gamma}{\eta L^2}$.

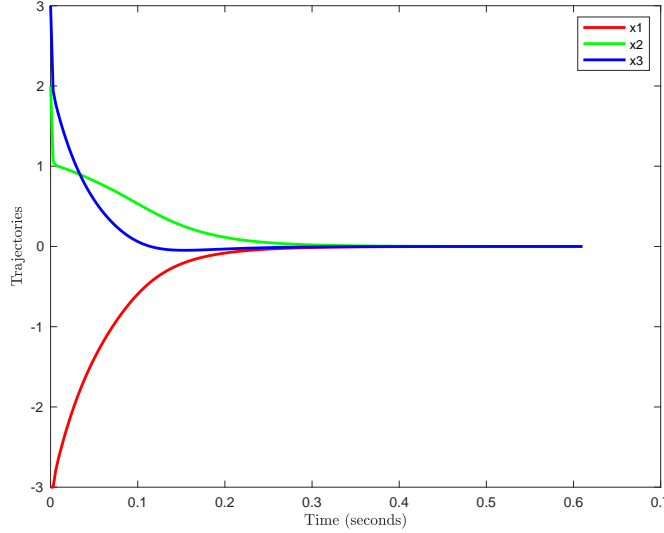


FIG. 5.1. Global exponential convergence for Example 5.1

We compare linear convergence of (RIPA) with some of its other special cases. We choose $\eta = 1.5$, $\lambda = \frac{1.99\gamma}{\eta L^2}$, and $x_0 = x_1 = (-3, 2, 3)$. For overrelaxed (oRPA), we choose maximum ρ satisfying (3.4), i.e., $\rho = 1.26$ and this implies $\theta = 0$. We choose $\rho = 1.1$ satisfying (3.4) for overrelaxed inertial (oRIPA), $\rho = 1$ for nonrelaxed inertial (IPA) and $\rho = 0.5$ for inner relaxed inertial (iRIPA). For these algorithms we choose

maximum inertial θ satisfying (3.5). Finally, for classical projection algorithm (PA), $\rho = 1$ and $\theta = 0$. It can be seen from Figure 5.2 that (oRPA) is the best choice among different algorithms.

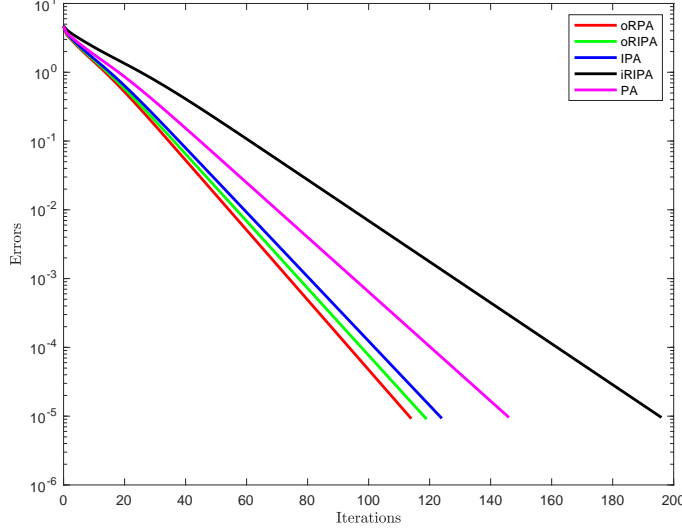


FIG. 5.2. Comparison of different versions of (RIPA) for Example 5.1

EXAMPLE 5.2. Let $H = \ell_2$, the real Hilbert space whose elements are the square-summable sequences of real scalars. Let $\delta, \xi \in \mathbb{R}$ be such that $\xi > \delta > \frac{\xi}{2} > 0$. Put

$$C := \{x \in H : \|x\| \leq \delta\}, \quad Ax := (\xi - \|x\|)x,$$

where δ and ξ are parameters. It is easy to verify that $x^* = 0$ is the unique solution of $VI(A, C)$. It was proved in [26] that A is Lipschitz continuous with $L := \xi + 2\delta$ and strongly pseudo-monotone with modulus $\gamma := \xi - \delta > 0$ on C . Note that A is neither strongly monotone nor monotone on C . Indeed, it suffices to choose $x = (\frac{\xi}{2}, 0, \dots, 0, \dots)$, $y = (\delta, 0, \dots, 0, \dots) \in C$ and note that

$$\langle Ax - Ay, x - y \rangle = \left(\frac{\xi}{2} - \delta\right)^3 < 0.$$

According to Theorem 2.8, the trajectory of dynamical system (1.4) globally exponentially converges to the unique solution $x^* = 0$. In the following, we will confirm this behavior numerically. Figure 5.3 shows an example confirming the global exponential convergence. We choose $H = \mathbb{R}^5$, $\xi = 8$, $\delta = 5$, then $\gamma = 3$, $L = 18$, and $\alpha(t) = \alpha$, $\beta(t) = \beta$ satisfying (2.15), $x(0) = (5, 4, 3, -1, -3)^T$, $\dot{x}(0) = 0$. It is clear from Figure 5.3 that the trajectories $x(t)$ converge exponentially to the unique solution $x^* = 0$.

Figure 5.4 compares linear convergence of (RIPA) with some of its others special cases. We choose $H = \mathbb{R}^{5000}$, $\xi = 8$, $\delta = 5$, then $\gamma = 3$, $L = 18$. $x_0 = x_1 = (1, 1, \dots, 1)$ and other parameters are chosen as in Example 5.1. Again from Figure 5.4, (oRPA) is the best choice among different algorithms.

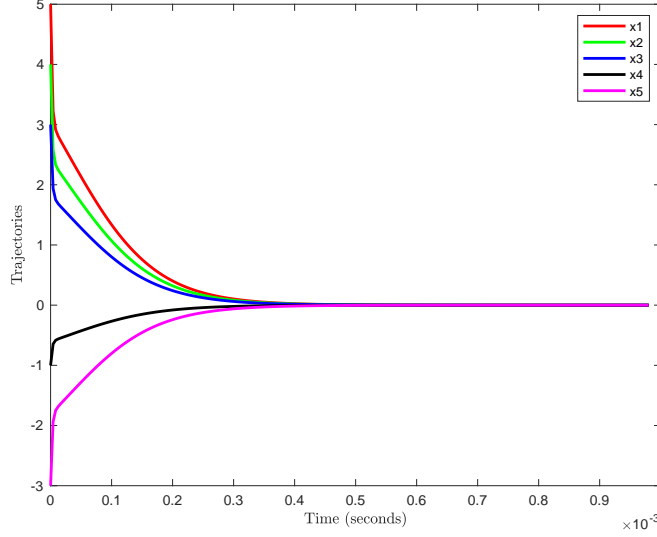


FIG. 5.3. *Global exponential convergence for Example 5.2*

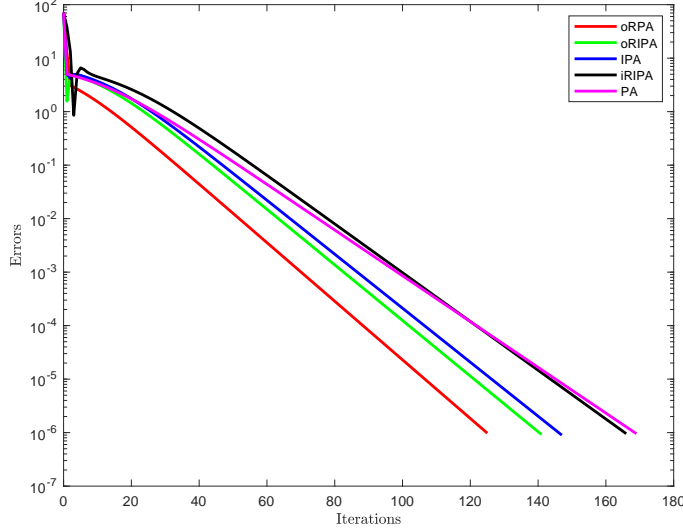


FIG. 5.4. *Comparison of different versions of (RIPA) for Example 5.2 with $H = \mathbb{R}^{5000}$*

Sparse binary tomography. To illustrate the monotone inclusions, we consider the sparse binary tomography problem. The goal is to reconstruct a sparse vectorized binary image $x \in \{0, 1\}^n$ from a limited number of tomographic projections

$$Tx = b.$$

Each pixel of the image x , denoted by $x_i \in [0, 1]$, and each entry in $b \in \mathbb{R}^m$, called *tomographic measurement* or *single projection*, corresponds to the integrated gray values of x along the single ray. Each matrix entry $T_{ij} \geq 0$ corresponds to the length of

the intersection of the i -th ray with the j -th pixel. If ray i and pixel j do not intersect then $T_{ij} = 0$. Stacking all equations for all rays together leads to the linear equation $Tx = b$, and the measurements are described such that $T = (T_{\theta_1}^t \ T_{\theta_2}^t \ \dots \ T_{\theta_{na}}^t)^t$ and each block matrix T_{θ_i} corresponds to a different projecting angle (t denotes the transpose). Mathematically, the problem is formed as

$$\min_x \frac{1}{2} \|Tx - b\|_2^2 + \lambda \|x\|_0,$$

with $\|\cdot\|_0$ being the l_0 norm, which counts the number of nonzero entries of x ; that is

$$\|x\|_0 = |\{x_i \mid x_i \neq 0\}|$$

where $|\cdot|$ denotes here the cardinality, i.e., the number of elements of a set. Since l_0 minimization is NP-hard a popular approach is to replace l_0 by the convex norm l_1 , and hence solve the well-known LASSO problem [36]

$$\min_x \frac{1}{2} \|Tx - b\|_2^2 + \lambda \|x\|_1,$$

or equivalently the monotone inclusion

$$0 \in Ax + Bx,$$

where $Ax := \nabla(\frac{1}{2}\|Tx - b\|^2) = T^*(Tx - b)$ and $Bx := \partial\|x\|_1$. Recall that the subdifferential of $\partial\|\cdot\|_1$, denoted by $\partial\|\cdot\|_1$, is defined as

$$(\partial\|x\|_1)_i = \begin{cases} \text{sign}(x_i) & \text{if } x_i \neq 0; \\ \text{any element of } [-1, 1] & \text{if } x_i = 0. \end{cases}$$

The resolvent of B at $v \in H$ is computed by

$$[J_{\lambda B}(v)]_i = \left[\text{prox}_{\lambda\|\cdot\|_1}(v) \right]_i = \text{sign}(v_i) \max\{|v_i| - \lambda, 0\}.$$

For the experiments we use the MATLAB routine `parallel_tomo.m` from the AIR Tools package [21, 22] that implements such a tomographic matrix for a given vector of angles. For the experiment we choose a peppers image of size $N = 512$ (this means 512×512 pixels) and the number of parallel beams $nA = 100$. We use the default value of the number of parallel rays, that is $p = \text{round}(\text{sqrt}(2) * N)$. The matrix size depends on the parameters N, nA, p as $T \in \mathbb{R}^{(nA * p) \times (N^2)}$, i.e. a problem in 262144 dimensions. We also include unknown additive random noise $\varepsilon \in \mathbb{R}^m$ so we get

$$Tx = b + \varepsilon.$$

We choose $\lambda = 10^{-5}$ and $\rho = 1.26, \theta = 0$ for oRPA, $\rho = 1, \theta = 0.2$ for IPA and $\rho = 0.5, \theta = 0.2$ for iRIPA. The results after 1000 iterations are displayed in Figures 5.5 and 5.6. It can be seen again that (oRPA) outperforms the other chosen algorithms.

6. Concluding remarks. We have proposed a second order dynamical system for solving strong pseudo-monotone VIs and investigated its global exponential convergence. A discretization of the proposed dynamical system leads to a new relaxed inertial projection algorithm for which we establish the linear convergence of the iterations. Extension to general monotone inclusions are also discussed. Relaxing the strong-pseudo-monotonicity assumption is an interesting topic for future research.

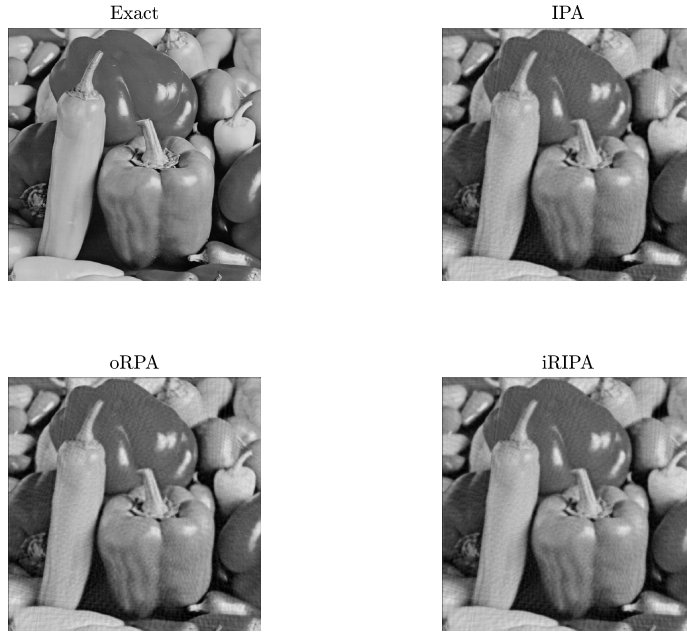


FIG. 5.5. *Reconstruction results of different versions of (RIPA) for the sparse binary tomography problem*

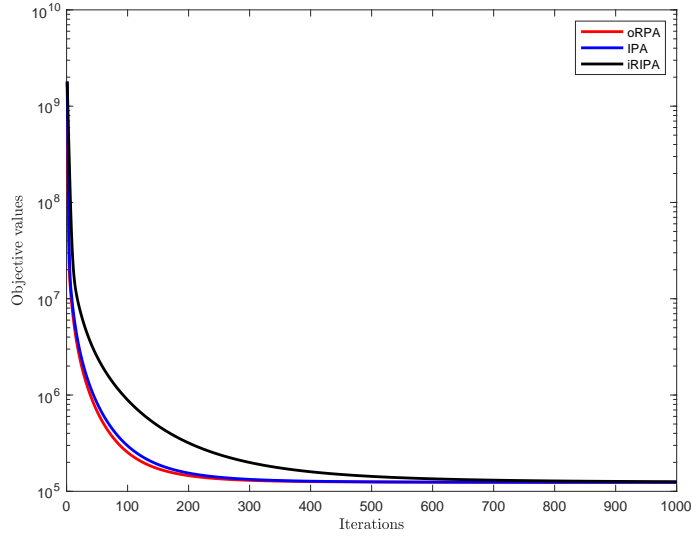


FIG. 5.6. *Comparison of different versions of (RIPA) for the sparse binary tomography problem*

Acknowledgement. The author is grateful to both anonymous referees for their constructive comments, which helped improve the presentation of the paper. This work was supported by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) project 101.01-2019.320.

REFERENCES

- [1] F. Alvarez, H. Attouch: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Analysis* 9(2001), 3–11.
- [2] A.S. Antipin: On a method for convex programs using a symmetrical modification of the Lagrange function. *Ekonomika i Mat. Metody.*, 12(1976), 1164–1173.
- [3] A.S. Antipin: Continuous and Iterative Processes with Projection Operators and Projection Like Operators. AN SSSR, Scientific Counsel on the Complex Problem Cybernetics, Moscow, 5-43 (1989).
- [4] A.S. Antipin: Minimization of convex functions on convex sets by means of differential equations. (Russian) *Differentsial'nye Uravneniya* 30(9), 1475–1486, 1994; translation in *Differential Equations* 30(1994), 1365–1375
- [5] H. Attouch, A. Cabot: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. *Mathematical Programming*, DOI 10.1007/s10107-019-01412-0 (2019)
- [6] H. Attouch, A. Cabot: Convergence of a relaxed inertial forward–backward algorithm for structured monotone inclusions. *Applied Mathematics and Optimization*, 80 (2019), 547–598
- [7] H. Attouch, J. Peypouquet, P. Redont: A dynamical approach to an inertial forward- backward algorithm for convex minimization. *SIAM J Optim.* 24 (2014), 232–256.
- [8] H. Attouch, X. Goudou, P. Redont: The heavy ball with friction method. I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics* 2(2000), 1–34.
- [9] M. Avriel, W.E. Diewert, S. Schaible, I. Zang: *Generalized Concavity*. Society for Industrial and Applied Mathematics (2010)
- [10] H. Bauschke, P. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
- [11] J. Bolte: Continuous gradient projection method in Hilbert spaces. *J. Optim. Theory. Appl.* 119(2003), 235–259.
- [12] R. I. Boţ, E.R. Csetnek: Second order forward-backward dynamical systems for monotone inclusion problems, *SIAM J. Control Optim.* 54(2016), 1423–1443
- [13] R. I. Boţ, E. R. Csetnek: A dynamical system associated with the fixed points set of a nonexpansive operator. *Journal of Dynamics and Differential Equations* 29 (2017), 155–168
- [14] R. I. Boţ, E. R. Csetnek, P.T. Vuong: The Forward-Backward-Forward Method from discrete and continuous perspective for pseudo-monotone variational inequalities in Hilbert Spaces. *Eur. J. Oper. Res.*, DOI 10.1016/j.ejor.2020.04.035 (2020)
- [15] R. I. Boţ, M. Sedlmayer, P.T. Vuong: A Relaxed Inertial Forward-Backward-Forward Algorithm for Solving Monotone Inclusions with Application to GANs, arXiv: 2003.07886 (2020)
- [16] E. Cavazzuti, P. Pappalardo, M. Passacantando: Nash Equilibria, Variational Inequalities, and Dynamical Systems. *J. Optim. Theory Appl.* 114 (2002), 491–506
- [17] F. Facchinei, and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Vols. I and II, Springer-Verlag, New York, 2003.
- [18] K. Goebel, S. Reich: *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*. Marcel Dekker, New York (1984)
- [19] K. Guo: On the linear convergence rate of a relaxed forward–backward splitting method, *Optimization*, DOI 10.1080/02331934.2020.1783260 (2020)
- [20] N.T.T.Ha, J.J. Strodiot, P.T. Vuong: On the global exponential stability of a projected dynamical system for strongly pseudomonotone variational inequalities. *Opt. Lett.* 12 (2018), 1625–1638
- [21] P.C. Hansen, M. Saxild-Hansen: AIR Tools — A MATLAB package of Iterative Reconstruction Methods, *J. Comput. Appl. Math.* 236 (2012), 2167–2178
- [22] P.C. Hansen, J.S. Jørgensen: AIR Tools II: Algebraic Iterative Reconstruction Methods, Improved Implementation, *Num. Alg.*, 79 (2018), 107–137
- [23] A. Haraux, *Systemes Dynamiques Dissipatifs et Applications*, Recherches en Mathematiques Appliquees 17, Masson, Paris, 1991
- [24] X. Hu, J. Wang: Global stability of a recurrent neural network for solving pseudomonotone variational inequalities. *Proc. IEEE Int. Symp. Circuits Syst.*, Island of Kos, Greece, May

- 21-24, 755-758 (2006)
- [25] S. Karamardian, S. Schaible: Seven kinds of monotone maps, *J. Optim. Theory Appl.* 66 (1990), 37–46
 - [26] P.D. Khanh, P.T. Vuong: Modified projection method for strongly pseudomonotone variational inequalities. *J. Glob. Optim.* 58 (2014), 341-350
 - [27] D.S. Kim, P.T. Vuong, P.D. Khanh: Qualitative properties of strongly pseudomonotone variational inequalities. *Opt. Lett.* 10 (2016), 1669–1679
 - [28] D. Kinderlehrer, G. Stampacchia: An introduction to variational inequalities and their applications. New York: Academic, (1980)
 - [29] D.A. Lorenz, T. Pock: An Inertial Forward-Backward Algorithm for Monotone Inclusions, *J. Math. Imaging Vis.* 51 (2015), 311–325
 - [30] A. Nagurney, D. Zhang, D.: Projected dynamical systems and variational inequalities with applications. Kluwer Academic. (1996)
 - [31] P. Ochs, T. Brox, T. Pock: iPiasco: Inertial Proximal Algorithm for strongly convex Optimization. *Journal of Mathematical Imaging and Vision*, 53(2015), 171–181
 - [32] M. Pappalardo, M. Passacantando: Stability for equilibrium problems: From variational inequalities to dynamical systems. *J. Optim. Theory Appl.* 113 (2002), 567–582
 - [33] B.T. Polyak: Some methods of speeding up the convergence of iteration methods. *U.S.S.R. Comput. Math. Math. Phys.* 4, (1964), 1–17
 - [34] M.V. Solodov, B.F. Svaiter: A new projection method for variational inequality problems. *SIAM J. Control Optim.*, 37 (1999), 765–776
 - [35] N. N. Tam, J. C. Yao, N. D. Yen: Solution methods for pseudomonotone variational inequalities. *J. Optim. Theory Appl.* 138 (2008), 253–273
 - [36] R. Tibshirani: Regression shrinkage and selection via the LASSO. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 58 (1996), 267–288
 - [37] P.T. Vuong: The global exponential stability of a dynamical system for solving variational inequalities. *Networks and Spatial Economics* (2019), DOI 10.1007/s11067-019-09457-6
 - [38] P.T. Vuong, J. J. Strodiot: A dynamical system for strongly pseudo-monotone equilibrium problems. *J. Optim. Theory Appl.* 185 (2020) 767–784