



# Second Order Splitting Dynamics with Vanishing Damping for Additively Structured Monotone Inclusions

Radu Ioan Boţ<sup>1</sup> · David Alexander Hulett<sup>1</sup>

Received: 4 January 2022 / Revised: 4 January 2022 / Accepted: 22 March 2022 /  
Published online: 19 April 2022  
© The Author(s) 2022

## Abstract

In the framework of a real Hilbert space, we address the problem of finding the zeros of the sum of a maximally monotone operator  $A$  and a cocoercive operator  $B$ . We study the asymptotic behaviour of the trajectories generated by a second order equation with vanishing damping, attached to this problem, and governed by a time-dependent forward–backward-type operator. This is a splitting system, as it only requires forward evaluations of  $B$  and backward evaluations of  $A$ . A proper tuning of the system parameters ensures the weak convergence of the trajectories to the set of zeros of  $A + B$ , as well as fast convergence of the velocities towards zero. A particular case of our system allows to derive fast convergence rates for the problem of minimizing the sum of a proper, convex and lower semicontinuous function and a smooth and convex function with Lipschitz continuous gradient. We illustrate the theoretical outcomes by numerical experiments.

**Keywords** Asymptotic stabilization · Damped inertial dynamics · Lyapunov analysis · Vanishing viscosity · Splitting system · Monotone inclusions

**Mathematics Subject Classification** 37N40 · 46N10 · 65K05 · 65K10 · 90B50 · 90C25

---

Research partially supported by the Doctoral Programme *Vienna Graduate School on Computational Optimization (VGSCO)* which is funded by FWF (Austrian Science Fund), project W1260-N35.

---

✉ Radu Ioan Boţ  
radu.bot@univie.ac.at

David Alexander Hulett  
david.alexander.hulett@univie.ac.at

<sup>1</sup> Faculty of Mathematics, University of Vienna, Vienna, Austria

## 1 Introduction

### 1.1 Problem Formulation and a Continuous Time Splitting Scheme with Vanishing Damping

Let  $\mathcal{H}$  be a real Hilbert,  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  a maximally monotone operator and  $B : \mathcal{H} \rightarrow \mathcal{H}$  a  $\beta$ -cocoercive operator for some  $\beta > 0$  such that  $\text{zer}(A + B) \neq \emptyset$ . Devising fast convergent continuous and discrete time dynamics for solving monotone inclusions of the type

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in (A + B)(x) \quad (1)$$

is of great importance in many fields, including, but not limited to, optimization, equilibrium theory, economics and game theory, partial differential equations, and statistics. One of our main motivations comes from the fact that solving the convex optimization problem

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

where  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  is proper, convex and lower semicontinuous and  $g : \mathcal{H} \rightarrow \mathbb{R}$  is convex and Fréchet differentiable with a Lipschitz continuous gradient, is equivalent to solving the monotone inclusion

$$0 \in (\partial f + \nabla g)(x).$$

We want to exploit the additive structure of (1) and approach  $A$  and  $B$  separately, in the spirit of the splitting paradigm.

For  $t \geq t_0 > 0$ ,  $\alpha > 1$ ,  $\xi \geq 0$ , and functions  $\lambda, \gamma : [t_0, +\infty) \rightarrow (0, +\infty)$ , we will study the asymptotic behaviour of the trajectories of the second order differential equation

$$\text{(Split-DIN-AVD)} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \left( \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right) + T_{\lambda(t), \gamma(t)}(x(t)) = 0, \quad (2)$$

where, for  $\lambda, \gamma > 0$ , the operator  $T_{\lambda, \gamma} : \mathcal{H} \rightarrow \mathcal{H}$  is given by

$$T_{\lambda, \gamma} = \frac{1}{\lambda} \left[ \text{Id} - J_{\gamma A} \circ (\text{Id} - \gamma B) \right].$$

The sets of zeros of  $A + B$  and of  $T_{\lambda, \gamma}$ , for  $\lambda, \gamma > 0$ , coincide. The nomenclature (Split-DIN-AVD) comes from the splitting feature of the continuous time scheme, as well as the link with the (DIN-AVD) system developed by Attouch and László in [9] (Dynamic Inertial Newton—Asymptotic Vanishing Damping), which we will emphasize later. We will discuss the existence and uniqueness of the trajectories generated (Split-DIN-AVD), and also show their weak convergence to the set of zeros of  $A + B$  as well as the fast convergence of the velocities to zero, and convergence rates for  $T_{\lambda(t), \gamma(t)}(x(t))$  and  $\frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t))$  as  $t \rightarrow +\infty$ .

For the particular case  $B = 0$ , we are left with the monotone inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in A(x),$$

and the attached system

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \left( \frac{d}{dt} A_{\lambda(t), \gamma(t)}(x(t)) \right) + A_{\lambda(t), \gamma(t)}(x(t)) = 0,$$

where, for  $\lambda, \gamma > 0$ , the operator  $A_{\lambda,\gamma} : \mathcal{H} \rightarrow \mathcal{H}$  can be seen as a *generalized Moreau envelope* of the operator  $A$ , i.e.,

$$A_{\lambda,\gamma} = \frac{1}{\lambda} [\text{Id} - J_{\gamma A}].$$

In particular, we will be able to set  $\gamma(t) = \lambda(t)$  for every  $t \geq t_0$ . Since for  $\lambda > 0$ ,  $A_{\lambda,\lambda} = A_\lambda$ , this allows us to recover the (DIN-AVD) system

$$\text{(DIN-AVD)} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \left( \frac{d}{dt} A_{\lambda(t)}(x(t)) \right) + A_{\lambda(t)}(x(t)) = 0,$$

addressed by Attouch and László in [9].

If  $A = 0$ , and after properly redefining some parameters, we obtain the following system

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \left( \frac{d}{dt} \frac{1}{\eta(t)} Bx(t) \right) + \frac{1}{\eta(t)} Bx(t) = 0,$$

with  $\eta : [t_0, +\infty) \rightarrow (0, +\infty)$ , which addresses the monotone equation

$$\text{find } x \in \mathcal{H} \text{ such that } B(x) = 0.$$

This dynamical system approaches the cocoercive operator  $B$  directly through a forward evaluation, which is more natural, instead of having to resort to its Moreau envelope, as in (DIN-AVD).

### 1.2 Notation and Preliminaries

In this subsection, we will explain the notions which were mentioned in the previous subsection, and we will introduce some definitions and preliminary results that will be required later. Throughout the paper, we will be working in a real Hilbert space  $\mathcal{H}$  with inner product  $\langle \cdot, \cdot \rangle$  and corresponding norm  $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ .

Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a *set-valued* operator, that is,  $Ax$  is a subset of  $\mathcal{H}$  for every  $x \in \mathcal{H}$ . The operator  $A$  is totally characterized by its *graph*  $\text{gra } A = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ . The *inverse* of  $A$  is the operator  $A^{-1} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  well-defined through the equivalence  $x \in A^{-1}u$  if and only if  $u \in Ax$ . The *set of zeros* of  $A$  is the set  $\text{zer } A = \{x \in \mathcal{H} : 0 \in Ax\}$ . For a subset  $C \subseteq \mathcal{H}$ , we say that  $A(C) = \cup_{x \in C} Ax$ . The *range* of  $A$  is the set  $\text{ran } A = A(\mathcal{H})$ .

A set-valued operator  $A$  is said to be *monotone* if  $\langle v - u, y - x \rangle \geq 0$  whenever  $(x, u), (y, v) \in \text{gra } A$ , and *maximally monotone* if it is monotone and the following implication holds:

$$\tilde{A} \text{ is monotone, } \text{gra } A \subseteq \text{gra } \tilde{A} \implies A = \tilde{A}.$$

Let  $\lambda > 0$ . The *resolvent* of index  $\lambda$  of  $A$  is the operator  $J_{\lambda A} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  given by

$$J_{\lambda A} = (\text{Id} + \lambda A)^{-1},$$

and the *Moreau envelope* (or *Yosida approximation* or *Yosida regularization*) of index  $\lambda$  of  $A$  is the operator  $A_\lambda : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  given by

$$A_\lambda = \frac{1}{\lambda} (\text{Id} - J_{\lambda A}),$$

where  $\text{Id} : \mathcal{H} \rightarrow \mathcal{H}$ , defined by  $\text{Id}(x) = x$  for every  $x \in \mathcal{H}$ , is the *identity* operator of  $\mathcal{H}$ . For  $\lambda_1, \lambda_2 > 0$ , it holds  $(A_{\lambda_1})_{\lambda_2} = A_{\lambda_1 + \lambda_2}$ .

A single-valued operator  $B : \mathcal{H} \rightarrow \mathcal{H}$  is said to be  $\beta$ -cocoercive for some  $\beta > 0$  if for every  $x, y \in \mathcal{H}$  we have

$$\beta \|Bx - By\|^2 \leq \langle Bx - By, x - y \rangle.$$

In this case,  $B$  is  $\frac{1}{\beta}$ -Lipschitz continuous, namely, for every  $x, y \in \mathcal{H}$  we have

$$\|Bx - By\| \leq \frac{1}{\beta} \|x - y\|.$$

We say  $B$  is *nonexpansive* if it is 1-Lipschitz continuous, and *firmly nonexpansive* if it is 1-cocoercive. For  $\alpha \in (0, 1)$ , we say  $B$  is  $\alpha$ -averaged if there exists a nonexpansive operator  $R : \mathcal{H} \rightarrow \mathcal{H}$  such that

$$B = (1 - \alpha) \text{Id} + \alpha R.$$

Let  $\lambda > 0$  and  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ . According to *Minty’s Theorem*,  $A$  is maximally monotone if and only if  $\text{ran}(\text{Id} + \lambda A) = \mathcal{H}$ . In this case  $J_{\lambda A}$  is single-valued and firmly nonexpansive,  $A_{\lambda}$  is single-valued,  $\lambda$ -cocoercive, and for every  $x \in \mathcal{H}$  and every  $\lambda_1, \lambda_2 > 0$  we have

$$\|J_{\lambda_1 A}(x) - J_{\lambda_2 A}(x)\| \leq |\lambda_1 - \lambda_2| \|A_{\lambda_1}(x)\|.$$

Let  $B : \mathcal{H} \rightarrow \mathcal{H}$  be a single-valued operator. If  $B$  is  $\alpha$ -averaged for some  $\alpha \in (0, 1)$ , then  $\text{Id} - B$  is  $\frac{1}{2\alpha}$ -cocoercive. If  $B$  is monotone and continuous, then it is maximally monotone.

The following concepts and results show the strong interplay between the theory of monotone operators and the convex analysis.

Let  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper, convex and lower semicontinuous function. We denote the infimum of  $f$  over  $\mathcal{H}$  by  $\min_{\mathcal{H}} f$  and the set of global minimizers of  $f$  by  $\text{argmin}_{\mathcal{H}} f$ . The *subdifferential* of  $f$  is the operator  $\partial f : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  defined, for every  $x \in \mathcal{H}$ , by

$$\partial f(x) = \{x^* \in \mathcal{H} : \langle x^*, y - x \rangle + f(x) \leq f(y) \quad \forall y \in \mathcal{H}\}.$$

The subdifferential operator of  $f$  is maximally monotone and  $\bar{x} \in \text{zer } \partial f \Leftrightarrow \bar{x}$  is a global minimizer of  $f$ .

Let  $\lambda > 0$ . The *proximal operator* of  $f$  of index  $\lambda$  is the operator  $\text{prox}_{\lambda f} : \mathcal{H} \rightarrow \mathcal{H}$  defined, for every  $x \in \mathcal{H}$ , by

$$\text{prox}_{\lambda f}(x) = J_{\lambda \partial f}(x) = \text{argmin}_{y \in \mathcal{H}} \left[ f(y) + \frac{1}{2\lambda} \|x - y\|^2 \right],$$

which also means that  $\text{prox}_{\lambda f}$  is firmly nonexpansive. The *Moreau envelope* of  $f$  of index  $\lambda$  is the function  $f_{\lambda} : \mathcal{H} \rightarrow \mathbb{R}$  given, for every  $x \in \mathcal{H}$ , by

$$f_{\lambda}(x) = f(\text{prox}_{\lambda f}(x)) + \frac{1}{2\lambda} \|x - \text{prox}_{\lambda f}(x)\|^2.$$

The function  $f_{\lambda}$  is Fréchet differentiable and

$$\nabla f_{\lambda}(x) = \frac{1}{\lambda} (x - \text{prox}_{\lambda f}(x)) = (\partial f)_{\lambda}(x) \quad \forall x \in \mathcal{H}.$$

Finally, if  $f : \mathcal{H} \rightarrow \mathbb{R}$  has full domain and is Fréchet differentiable with  $\frac{1}{\beta}$ -Lipschitz continuous gradient, for  $\beta > 0$ , then, according to *Baillon–Haddad’s Theorem*,  $\nabla f$  is  $\beta$ -cocoercive.

### 1.3 A Brief History of Inertial Systems Attached to Optimization Problems and Monotone Inclusions

In the last years there have been many advances in the study of continuous time inertial systems with vanishing damping attached to monotone inclusion problems. We briefly visit them in the following paragraphs.

#### 1.3.1 The Heavy Ball Method with Friction

Consider a convex and continuously differentiable function  $f : \mathcal{H} \rightarrow \mathbb{R}$  with at least one minimizer. The *heavy ball with friction* system

$$(HBF) \quad \ddot{x}(t) + \mu\dot{x}(t) + \nabla f(x(t)) = 0 \quad (3)$$

was introduced by Álvarez in [2] as a suitable continuous time scheme to approach the minimization of the function  $f$ . This system can be seen as the equation of the horizontal position  $x(t)$  of an object that moves, under the force of gravity, along the graph of the function  $f$ , subject to a kinetic friction represented by the term  $\mu\dot{x}(t)$  (a nice derivation can be seen in the work done by Attouch-Goudou-Redont in [8]). It is known that, if  $x$  is a solution of (HBF), then  $x$  converges weakly to a minimizer of  $f$  and  $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t}\right)$  as  $t \rightarrow +\infty$ .

In recent times, the question was raised whether the damping coefficient  $\mu$  could be chosen to be time-dependent. An important contribution was made by Su–Boyd–Candés (in [20]) who studied the case of an Asymptotic Vanishing Damping coefficient  $\mu(t) = \frac{\alpha}{t}$ , namely,

$$(AVD) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \quad (4)$$

and proved when  $\alpha \geq 3$  the rate of convergence for the functional values  $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$  as  $t \rightarrow +\infty$ . This second order system can be seen as a continuous counterpart to Nesterov's accelerated gradient method from [19]. Weak convergence of the trajectories generated by (AVD) when  $\alpha > 3$  has been shown by Attouch-Chbani-Peypouquet-Redont [6] and May [18], with the improved rate of convergence for the functional values  $f(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right)$  as  $t \rightarrow +\infty$ . For  $\alpha = 3$ , the convergence of the trajectories remains an open question, except for the one dimensional case (see [7]). In the subcritical case  $\alpha \leq 3$ , it has been shown by Apidopoulos-Aujol-Dossal [5] and Attouch-Chbani-Riahi [7] that the objective values converge at a rate  $\mathcal{O}(t^{-\frac{2\alpha}{3}})$  as  $t \rightarrow +\infty$ .

#### 1.3.2 Heavy Ball Dynamics and Cocoercive Operators

If  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, convex and lower semicontinuous function which is not necessarily differentiable, then we cannot make direct use of (3). However, since for  $\lambda > 0$  we have  $\operatorname{argmin} f = \operatorname{argmin} f_\lambda$ , we can replace  $f$  by its Moreau envelope  $f_\lambda$ , and the system now becomes

$$\ddot{x}(t) + \mu\dot{x}(t) + \nabla f_\lambda(x(t)) = 0.$$

In line with this idea, and in analogy with (3), Álvarez and Attouch [3] and Attouch and Maingé [11] studied the dynamics

$$\ddot{x}(t) + \mu\dot{x}(t) + B(x(t)) = 0, \quad (5)$$

where  $B : \mathcal{H} \rightarrow \mathcal{H}$  is a  $\beta$ -cocoercive operator. They were able to prove that the solutions of this system weakly converge to elements of  $\text{zer } B$  provided that the cocoercivity parameter  $\beta$  and the damping coefficient  $\mu$  satisfy  $\beta\mu^2 > 1$ . For a maximally monotone operator  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ , we know that its Moreau envelope is  $\lambda$ -cocoercive and thus, under the condition  $\lambda\mu^2 > 1$ , the trajectories of

$$\ddot{x}(t) + \mu\dot{x}(t) + A_{\lambda}(x(t)) = 0$$

converge weakly to elements of  $\text{zer } A_{\lambda} = \text{zer } A$ .

Also related to (5), Boţ-Csetnek [16] considered the system

$$\ddot{x}(t) + \mu(t)\dot{x}(t) + \nu(t)Bx(t) = 0, \tag{6}$$

where  $B : \mathcal{H} \rightarrow \mathcal{H}$  is again  $\beta$ -cocoercive. Under the assumption that  $\mu$  and  $\nu$  are locally absolutely continuous,  $\dot{\mu}(t) \leq 0 \leq \dot{\nu}(t)$  for almost every  $t \in [0, +\infty)$  and  $\inf_{t \geq 0} \frac{\mu^2(t)}{\nu(t)} > \frac{1}{\beta}$ , the authors were able to prove that the solutions to this system converge weakly to zeros of  $B$ .

In [12], Attouch and Peypouquet addressed the system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \tag{7}$$

where  $\alpha > 1$  and the time-dependent regularizing parameter  $\lambda(t)$  satisfies  $\lambda(t) \frac{\alpha^2}{t^2} > 1$  for every  $t \geq t_0 > 0$ . As well as ensuring the weak convergence of the trajectories towards elements of  $\text{zer } A$ , choosing the regularizing parameter in such a fashion allowed the authors to obtain fast convergence of the velocities and accelerations towards zero.

### 1.3.3 Inertial Dynamics with Hessian Damping

Let us return briefly to the (AVD) system (4). In addition to the viscous vanishing damping term  $\frac{\alpha}{t}\dot{x}(t)$ , the following system with Hessian-driven damping was considered by Attouch-Peypouquet-Redont in [13]

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \xi \nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0,$$

where  $\xi \geq 0$ . While preserving the fast convergence properties of the Nesterov accelerated method, the Hessian-driven damping term reduces the oscillatory aspect of the trajectories. In [9], Attouch and László studied a version of (7) with an added Hessian-driven damping term:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \xi \left( \frac{d}{dt} A_{\lambda(t)}(x(t)) \right) + A_{\lambda(t)}(x(t)) = 0.$$

While preserving the convergence results of (7), the main benefit of the introduction of this damping term is the fast convergence rates that can be obtained for  $A_{\lambda(t)}(x(t))$  and  $\frac{d}{dt} A_{\lambda(t)}(x(t))$  as  $t \rightarrow +\infty$ . The regularizing parameter  $\lambda(t)$  is again chosen to be time-dependent; in the general case, the authors take  $\lambda(t) = \lambda t^2$ , and in [12] it is shown that taking  $\lambda(t)$  this way is critical. However, in the case where  $A = \partial f$  for a proper, convex and lower semicontinuous function  $f$ , it is also allowed to take  $\lambda(t) = \lambda t^r$  with  $r \geq 0$ .

### 1.4 Layout of the Paper

In Sect. 2, we give the proof for the existence and uniqueness of strong global solutions to (Split-DIN-AVD) by means of a Cauchy–Lipschitz–Picard argument. In Sect. 3 we state the main theorem of this work, and we show the weak convergence of the solutions of (2) to elements of  $\text{zer}(A + B)$ , as well as the fast convergence of the velocities and accelerations to zero. We also provide convergence rates for  $T_{\lambda(t),\gamma(t)}(x(t))$  and  $\frac{d}{dt}T_{\lambda(t),\gamma(t)}(x(t))$  as  $t \rightarrow +\infty$ . We explore the particular cases  $A = 0$  and  $B = 0$ , and show improvements with respect to previous works. In Sect. 4, we address the convex minimization case, namely, when  $A = \partial f$  and  $B = \nabla g$ , where  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, convex and lower semicontinuous function and  $g : \mathcal{H} \rightarrow \mathbb{R}$  is a convex and Fréchet differentiable function with Lipschitz continuous gradient, and derive, in addition, a fast convergence rate for the function values. In Sect. 5, we illustrate the theoretical results by numerical experiments. In Sect. 5, we provide an algorithm that arises from a time discretization of (Split-DIN-AVD) and discuss its convergence properties.

## 2 Existence and Uniqueness of Trajectories

In this section, we show the existence and uniqueness of strong global solutions to (Split-DIN-AVD). For the sake of clarity, first we state the definition of a strong global solution.

**Definition 2.1** We say that  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  is a *strong global solution* of (Split-DIN-AVD) with Cauchy data  $(x_0, u_0) \in \mathcal{H} \times \mathcal{H}$  if

- (i)  $x, \dot{x} : [t_0, +\infty) \rightarrow \mathcal{H}$  are locally absolutely continuous;
- (ii)  $\ddot{x}(t) + \frac{\alpha}{t}\dot{x} + \xi \left( \frac{d}{dt}T_{\lambda(t),\gamma(t)}(x(t)) \right) + T_{\lambda(t),\gamma(t)}(x(t)) = 0$  for almost every  $t \in [t_0, +\infty)$ ;
- (iii)  $x(t_0) = x_0, \dot{x}(t_0) = u_0$ .

A classic solution is just a strong global solution which is  $C^2$ . Sometimes we will mention the terms *strong global solution* or *classic global solution* without explicit mention of the Cauchy data.

The following lemma will be used to prove the existence of strong global solutions of our system, and we will need it in the proof of the main theorem as well.

**Lemma 2.2** Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a maximally monotone operator and  $B : \mathcal{H} \rightarrow \mathcal{H}$  a  $\beta$ -cocoercive operator for some  $\beta > 0$ . Then, the following statements hold:

- (i) For  $\lambda > 0$  and  $\gamma \in (0, 2\beta)$ ,  $T_{\lambda,\gamma}$  is a  $\lambda \frac{4\beta-\gamma}{4\beta}$ -cocoercive operator. In particular, this also implies that  $T_{\lambda,\gamma}$  is  $\frac{\lambda}{2}$ -cocoercive.
- (ii) Choose  $\lambda_1, \lambda_2 > 0, \gamma_1, \gamma_2 \in (0, 2\beta)$  and  $x, y \in \mathcal{H}$ . Then, for  $\bar{x} \in \text{zer}(A + B)$  it holds

$$\begin{aligned} \|\lambda_1 T_{\lambda_1,\gamma_1}(x) - \lambda_2 T_{\lambda_2,\gamma_2}(y)\| &\leq 4\|x - y\| + \frac{4\beta|\gamma_1 - \gamma_2|}{\gamma_1} \|B(x)\| \\ &\quad + \frac{2|\gamma_1 - \gamma_2|}{\gamma_1} \|x - \bar{x}\|, \\ \|T_{\lambda_1,\gamma_1}(x) - T_{\lambda_2,\gamma_2}(y)\| &\leq \frac{1}{\lambda_1} \left[ 4\|x - y\| + 4\beta \frac{|\gamma_1 - \gamma_2|}{\gamma_1} \|B(x)\| \right. \\ &\quad \left. + 2 \frac{|\gamma_1 - \gamma_2|}{\gamma_1} \|x - \bar{x}\| \right] + 2 \frac{|\lambda_2 - \lambda_1|}{\lambda_1 \lambda_2} \|y - \bar{x}\|. \end{aligned}$$

(iii) If  $x$  is a classic global solution to (2) and  $\bar{x} \in \text{zer}(A + B)$ , then, for every  $t \geq t_0$ , we have

$$\left\| \frac{d}{dt} (\lambda(t)T_{\lambda(t),\gamma(t)}(x(t))) \right\| \leq 4\|\dot{x}(t)\| + 4\beta \frac{|\dot{\gamma}(t)|}{\gamma(t)} \|B(x(t))\| + 2 \frac{|\dot{\gamma}(t)|}{\gamma(t)} \|x(t) - \bar{x}\|.$$

**Proof** (i) From [14, Proposition 26.1(iv)(d)] we know that the operator  $J_{\gamma A} \circ (\text{Id} - \gamma B)$  is  $\alpha = \frac{2\beta}{4\beta - \gamma}$ -averaged. From [14, Proposition 4.39], we obtain that  $\text{Id} - J_{\gamma A} \circ (\text{Id} - \gamma B)$  is  $\frac{1}{2\alpha}$ -cocoercive, namely, it is  $\frac{4\beta - \gamma}{4\beta}$ -cocoercive. Since  $\gamma \in (0, 2\beta)$ , we have  $\frac{4\beta - \gamma}{4\beta} > \frac{2\beta}{4\beta} = \frac{1}{2}$ , which implies that  $\text{Id} - J_{\gamma A} \circ (\text{Id} - \gamma B)$  is  $\frac{1}{2}$ -cocoercive and thus

$$T_{\lambda,\gamma} \text{ is } \lambda \frac{4\beta - \gamma}{4\beta}\text{-cocoercive and } T_{\lambda,\gamma} \text{ is } \frac{\lambda}{2}\text{-cocoercive.}$$

(ii) We have

$$\begin{aligned} \|\lambda_1 T_{\lambda_1, \gamma_1}(x) - \lambda_2 T_{\lambda_2, \gamma_2}(y)\| &\leq \|x - y\| + \|J_{\gamma_1 A}(x - \gamma_1 B(x)) - J_{\gamma_2 A}(y - \gamma_2 B(y))\| \\ &\leq \|x - y\| + \|J_{\gamma_1 A}(x - \gamma_1 B(x)) - J_{\gamma_2 A}(x - \gamma_1 B(x))\| \\ &\quad + \|J_{\gamma_2 A}(x - \gamma_1 B(x)) - J_{\gamma_2 A}(y - \gamma_2 B(y))\| \\ &\leq 2\|x - y\| + |\gamma_1 - \gamma_2| \|A_{\gamma_1}(x - \gamma_1 B(x))\| + \|\gamma_1 B(x) - \gamma_2 B(y)\| \\ &\leq 2\|x - y\| + |\gamma_1 - \gamma_2| \|A_{\gamma_1}(x - \gamma_1 B(x))\| \\ &\quad + \|\gamma_1 B(x) - \gamma_2 B(x)\| + \|\gamma_2 B(x) - \gamma_2 B(y)\| \\ &= 2\|x - y\| + |\gamma_1 - \gamma_2| \|A_{\gamma_1}(x - \gamma_1 B(x))\| \\ &\quad + |\gamma_1 - \gamma_2| \|B(x)\| + \gamma_2 \|B(x) - B(y)\|. \end{aligned}$$

Now, notice that

$$\begin{aligned} A_{\gamma_1}(x - \gamma_1 B(x)) &= \frac{1}{\gamma_1} (\text{Id} - J_{\gamma_1 A})(x - \gamma_1 B(x)) = -B(x) + \frac{1}{\gamma_1} (x - J_{\gamma_1 A}(x - \gamma_1 B(x))) \\ &= -B(x) + T_{\gamma_1, \gamma_1}(x), \end{aligned}$$

so using (i) and the fact that  $T_{\gamma_1, \gamma_2}(\bar{x}) = 0$ , we obtain

$$\begin{aligned} \|A_{\gamma_1}(x - \gamma_1 B(x))\| &= \|-B(x) + T_{\gamma_1, \gamma_2}(x)\| \leq \|B(x)\| + \|T_{\gamma_1, \gamma_2}(x) - T_{\gamma_1, \gamma_2}(\bar{x})\| \\ &\leq \|B(x)\| + \frac{2}{\gamma_1} \|x - \bar{x}\|. \end{aligned} \tag{8}$$

Altogether, plugging (8) into our initial inequality yields

$$\begin{aligned} \|\lambda_1 T_{\lambda_1, \gamma_2}(x) - \lambda_2 T_{\lambda_2, \gamma_2}(y)\| &\leq 2\|x - y\| + 2|\gamma_1 - \gamma_2| \|B(x)\| + \frac{2|\gamma_1 - \gamma_2|}{\gamma_1} \|x - \bar{x}\| \\ &\quad + \gamma_2 \|B(x) - B(y)\| \\ &\leq 2\|x - y\| + \frac{4\beta|\gamma_1 - \gamma_2|}{\gamma_1} \|B(x)\| + \frac{2|\gamma_1 - \gamma_2|}{\gamma_1} \|x - \bar{x}\| + 2\beta \left(\frac{1}{\beta}\right) \|x - y\|. \end{aligned}$$

To show the second inequality, we use the previous one. We have

$$\begin{aligned} \|T_{\lambda_1, \gamma_1}(x) - T_{\lambda_2, \gamma_2}(y)\| &= \frac{1}{\lambda_1} \|\lambda_1 T_{\lambda_1, \gamma_1}(x) - \lambda_2 T_{\lambda_2, \gamma_2}(y) + (\lambda_2 - \lambda_1) T_{\lambda_2, \gamma_2}(y)\| \\ &\leq \frac{1}{\lambda_1} \left[ 4\|x - y\| + 4\beta \frac{|\gamma_1 - \gamma_2|}{\gamma_1} \|B(x)\| + 2 \frac{|\gamma_1 - \gamma_2|}{\gamma_1} \|x - \bar{x}\| \right] + \frac{|\lambda_2 - \lambda_1|}{\lambda_1} \|T_{\lambda_2, \gamma_2}(y)\| \end{aligned}$$



$$\leq \frac{1}{\lambda_1} \left[ 4\|x - y\| + 4\beta \frac{|\gamma_1 - \gamma_2|}{\gamma_1} \|Bx\| + 2 \frac{|\gamma_1 - \gamma_2|}{\gamma_1} \|x - \bar{x}\| \right] + 2 \frac{|\lambda_2 - \lambda_1|}{\lambda_1 \lambda_2} \|y - \bar{x}\|,$$

where the last line is a consequence of  $T_{\lambda_2, \gamma_2}$  being  $\frac{\lambda_2}{2}$ -cocoercive, and hence  $\frac{2}{\lambda_2}$ -Lipschitz continuous (see (i)).

(iii) For  $t, s \geq t_0$  set

$$x = x(t), \quad y = x(s), \quad \lambda_1 = \lambda(t), \quad \gamma_1 = \gamma(t), \quad \lambda_2 = \lambda(s), \quad \gamma_2 = \gamma(s)$$

and use (ii) to obtain, for every  $t \geq t_0$ ,

$$\frac{\|\lambda(t)T_{\lambda(t), \gamma(t)}(x(t)) - \lambda(s)T_{\lambda(s), \gamma(s)}(x(s))\|}{|t - s|} \leq 4 \frac{\|x(t) - x(s)\|}{|t - s|} + \frac{4\beta}{\gamma(t)} \frac{|\gamma(t) - \gamma(s)|}{|t - s|} \|B(x(t))\| + \frac{2}{\gamma(t)} \frac{|\gamma(t) - \gamma(s)|}{|t - s|} \|x(t) - \bar{x}\|.$$

Hence, by taking the limit as  $s \rightarrow t$  we get, for any  $t \geq t_0$ ,

$$\left\| \frac{d}{dt} \lambda(t)T_{\lambda(t), \gamma(t)}(x(t)) \right\| \leq 4\|\dot{x}(t)\| + 4\beta \frac{|\dot{\gamma}(t)|}{\gamma(t)} \|B(x(t))\| + 2 \frac{|\dot{\gamma}(t)|}{\gamma(t)} \|x(t) - \bar{x}\|. \quad \square$$

The next theorem concerns the existence and uniqueness of strong global solutions to (Split-DIN-AVD).

**Theorem 2.3** *Assume that  $\lambda, \gamma : [t_0, +\infty) \rightarrow (0, +\infty)$  are Lebesgue measurable functions and that  $\inf_{t \geq t_0} \lambda(t) > 0$ . Then, for any  $(x_0, u_0) \in \mathcal{H} \times \mathcal{H}$  there exists a unique strong global solution  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  of the system (2) that satisfies  $x(t_0) = x_0$  and  $\dot{x}(t_0) = u_0$ .*

**Proof** We will rely on [17, Proposition 6.2.1] and distinguish between the cases  $\xi > 0$  and  $\xi = 0$ . For each case, we will check that the conditions of the aforementioned proposition are fulfilled. We will be working in the real Hilbert space  $\mathcal{H} \times \mathcal{H}$  endowed with the norm  $\|(x, y)\| = \|x\| + \|y\|$ . Let  $\bar{x} \in \text{zer}(A + B)$  be fixed.

*The Case  $\xi > 0$ .* First, it can be easily checked (see also [4, 9, 13]) that for all  $t \geq t_0$  the following dynamical systems are equivalent

$$\begin{aligned} * \quad & \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \left( \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right) + T_{\lambda(t), \gamma(t)}(x(t)) = 0, \\ * \quad & \begin{cases} \dot{x}(t) + \xi T_{\lambda(t), \gamma(t)}(x(t)) - \left( \frac{1}{\xi} - \frac{\alpha}{t} \right) x(t) + \frac{1}{\xi} y(t) = 0, \\ \dot{y}(t) - \left( \frac{1}{\xi} - \frac{\alpha}{t} + \frac{\alpha \xi}{t^2} \right) x(t) + \frac{1}{\xi} y(t) = 0. \end{cases} \end{aligned}$$

In other words, (2) with Cauchy data  $(x_0, u_0) = (x(t_0), \dot{x}(t_0))$  is equivalent to the first order system

$$\begin{cases} \dot{z}(t) = F(t, z(t)), \\ z(t_0) = (x_0, y_0), \end{cases}$$

where  $z(t) = (x(t), y(t))$ ,  $F$  is given, for every  $t \geq t_0$ , by

$$F(t, (x, y)) = \left[ -\xi T_{\lambda(t), \gamma(t)}(x) + \left( \frac{1}{\xi} - \frac{\alpha}{t} \right) x - \frac{1}{\xi} y, \left( \frac{1}{\xi} - \frac{\alpha}{t} + \frac{\alpha \xi}{t^2} \right) x - \frac{1}{\xi} y \right]$$

and the Cauchy data is  $x_0 = x(t_0), y_0 = -\xi \left( u_0 + \xi T_{\lambda(t_0), \gamma(t_0)}(x_0) - \left( \frac{1}{\xi} - \frac{\alpha}{t_0} \right) x_0 \right)$ .

(i) Let  $t \in [t_0, +\infty)$  be fixed. We need to verify the Lipschitz continuity of  $F$  on the  $z$  variable. Set  $z = (x, y)$ ,  $w = (u, v)$ . We have

$$\begin{aligned} \|F(t, z) - F(t, w)\| &= \left\| -\xi \left( T_{\lambda(t), \gamma(t)}(x) - T_{\lambda(t), \gamma(t)}(u) + \left( \frac{1}{\xi} - \frac{\alpha}{t} \right) \right. \right. \\ &\quad \left. \left. (x - u) - \frac{1}{\xi}(y - v) \right) \right\| \\ &\quad + \left\| \left( \frac{1}{\xi} - \frac{\alpha}{t} + \frac{\alpha\xi}{t^2} \right) (x - u) - \frac{1}{\xi}(y - v) \right\|. \end{aligned}$$

Set  $\underline{\lambda} := \inf_{t \geq t_0} \lambda(t) > 0$ . According to Lemma 2.2(i), the term involving the operator  $T_{\lambda(t), \gamma(t)}$  satisfies

$$\|T_{\lambda(t), \gamma(t)}(x) - T_{\lambda(t), \gamma(t)}(u)\| \leq \frac{2}{\underline{\lambda}(t)} \|x - u\| \leq \frac{2}{\underline{\lambda}} \|x - u\|.$$

It follows that, if we take

$$K(t) := \max \left\{ \frac{2\xi}{\underline{\lambda}} + \left| \frac{1}{\xi} - \frac{\alpha}{t} \right| + \left| \frac{1}{\xi} - \frac{\alpha}{t} + \frac{\alpha\xi}{t^2} \right|, \frac{2}{\xi} \right\} \quad \forall t \geq t_0,$$

then we have  $K \in L^1_{\text{loc}}([t_0, +\infty), \mathbb{R})$  and

$$\|F(t, z) - F(t, w)\| \leq K(t) \|z - w\| \quad \forall t \geq t_0.$$

(ii) Now, we claim that  $F$  fulfills a boundedness condition. For  $t \in [t_0, +\infty)$  and  $z = (x, y) \in \mathcal{H} \times \mathcal{H}$  we have

$$\|F(t, z)\| = \left\| -\xi T_{\lambda(t), \gamma(t)}(x) + \left( \frac{1}{\xi} - \frac{\alpha}{t} \right) x - \frac{1}{\xi} y \right\| + \left\| \left( \frac{1}{\xi} - \frac{\alpha}{t} + \frac{\alpha\xi}{t^2} \right) x - \frac{1}{\xi} y \right\|.$$

By Lemma 2.2(i), we have, for every  $t \geq t_0$ ,

$$\|T_{\lambda(t), \gamma(t)}(x)\| = \|T_{\lambda(t), \gamma(t)}(x) - T_{\lambda(t), \gamma(t)}(\bar{x})\| \leq \frac{2}{\lambda(t)} \|x - \bar{x}\|.$$

Hence, if we take

$$P(t) = \max \left\{ \frac{2\xi}{\lambda(t)} + \left| \frac{1}{\xi} - \frac{\alpha}{t} \right| + \left| \frac{1}{\xi} - \frac{\alpha}{t} + \frac{\alpha\xi}{t^2} \right|, \frac{2\xi}{\lambda(t)}, \frac{2}{\xi} \right\} \quad \forall t \geq t_0,$$

then we have  $P \in L^1_{\text{loc}}([t_0, +\infty), \mathbb{R})$  and

$$\|F(t, z)\| \leq P(t)(1 + \|z\|).$$

We have checked that the conditions of [17, Proposition 6.2.1] hold. Therefore, there exists a unique locally absolutely continuous solution  $t \mapsto x(t)$  of (2) that satisfies  $x(t_0) = x_0$  and  $\dot{x}(t_0) = u_0$ .

The Case  $\xi = 0$ . Now, (2) is easily seen to be equivalent to

$$\begin{cases} \dot{z}(t) = F(t, z(t)), \\ z(t_0) = (x_0, u_0), \end{cases}$$

where  $z(t) = (x(t), y(t))$  and  $F$  is given, for every  $t \geq t_0$ , by

$$F(t, (x, y)) = \left[ y, -\frac{\alpha}{t} y - T_{\lambda(t), \gamma(t)}(x) \right].$$

Showing that  $F$  fulfills the required properties is straightforward. □

### 3 The Convergence Properties of the Trajectories

In this section, we will study the asymptotic behaviour of the trajectories of the system

$$\text{(Split-DIN-AVD)} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \xi \frac{d}{dt} (T_{\lambda(t),\gamma(t)}(x(t))) + T_{\lambda(t),\gamma(t)}(x(t)) = 0,$$

where

$$T_{\lambda,\gamma}(x) = \frac{1}{\lambda} [\text{Id} - J_{\gamma A} \circ (\text{Id} - \gamma B)].$$

We will show weak convergence of the trajectories generated by (2) to elements of  $\text{zer}(A + B)$ , as well as the fast convergence of the velocities and accelerations to zero. Additionally, we will provide convergence rates for  $T_{\lambda(t),\gamma(t)}(x(t))$  and  $\frac{d}{dt}T_{\lambda(t),\gamma(t)}(x(t))$  as  $t \rightarrow +\infty$ . To avoid repetition of the statement “for almost every  $t$ ”, in the following theorem we will assume we are working with a classic global solution of our system.

**Theorem 3.1** *Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a maximally monotone operator and  $B : \mathcal{H} \rightarrow \mathcal{H}$  a  $\beta$ -cocoercive operator for some  $\beta > 0$  such that  $\text{zer}(A + B) \neq \emptyset$ . Assume that  $\alpha > 1, \xi \geq 0, \lambda(t) = \lambda t^2$  for  $\lambda > \frac{2}{(\alpha-1)^2}$  and all  $t \geq t_0$ , and that  $\gamma : [t_0, +\infty) \rightarrow (0, 2\beta)$  is a differentiable function that satisfies  $\frac{\dot{\gamma}(t)}{\gamma(t)} = \mathcal{O}(\frac{1}{t})$  as  $t \rightarrow +\infty$ . Then, for a solution  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  to (Split-DIN-AVD), the following statements hold:*

- (i)  $x$  is bounded.
- (ii) We have the estimates

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} t^3 \|\ddot{x}(t)\|^2 dt < +\infty,$$

$$\int_{t_0}^{+\infty} \frac{\gamma^2(t)}{t} \|A_{\gamma(t)}[x(t) - \gamma(t)Bx(t)] + Bx(t)\|^2 dt < +\infty.$$

- (iii) We have the convergence rates

$$\|\dot{x}(t)\| = o\left(\frac{1}{t}\right), \quad \|\ddot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^2}\right),$$

$$\|A_{\gamma(t)}[x(t) - \gamma(t)Bx(t)] + Bx(t)\| = o\left(\frac{1}{\gamma(t)}\right),$$

$$\left\| \frac{d}{dt} (A_{\gamma(t)}[x(t) - \gamma(t)Bx(t)] + Bx(t)) \right\| = \mathcal{O}\left(\frac{1}{t\gamma(t)}\right) + o\left(\frac{t^2 \left| \frac{d}{dt} \frac{\gamma(t)}{\lambda(t)} \right|}{\gamma^2(t)}\right)$$

as  $t \rightarrow +\infty$ .

- (iv) If  $0 < \inf_{t \geq t_0} \gamma(t) \leq \sup_{t \geq t_0} \gamma(t) < 2\beta$ , then  $x(t)$  converges weakly to an element of  $\text{zer}(A + B)$  as  $t \rightarrow +\infty$ .

**Proof** *Integral Estimates and Rates.* To develop the analysis, we will fix  $\bar{x} \in \text{zer}(A + B)$  and make use of the Lyapunov function  $\mathcal{E} : [t_0, +\infty) \rightarrow \mathbb{R} \cup \{+\infty\}$  given by

$$\mathcal{E}(t) := \frac{1}{2} \left\| \frac{\alpha - 1}{2} (x(t) - \bar{x}) + t(\dot{x}(t) + \xi T_{\lambda(t),\gamma(t)}(x(t))) \right\|^2 + \frac{(\alpha - 1)^2}{8} \|x(t) - \bar{x}\|^2. \quad (9)$$

Differentiation of  $\mathcal{E}$  with respect to time yields, for every  $t \geq t_0$ ,

$$\begin{aligned} \dot{\mathcal{E}}(t) = & \left\langle \frac{\alpha - 1}{2}(x(t) - \bar{x}) + t (\dot{x}(t) + \xi T_{\lambda(t),\lambda(t)}(x(t))), \right. \\ & \left. \frac{\alpha + 1}{2}\dot{x}(t) + \xi T_{\lambda(t),\gamma(t)}(x(t)) + t \left( \ddot{x}(t) + \xi \frac{d}{dt} (T_{\lambda(t),\gamma(t)}(x(t))) \right) \right\rangle \\ & + \frac{(\alpha - 1)^2}{4} \langle x(t) - \bar{x}, \dot{x}(t) \rangle. \end{aligned}$$

After reduction and employing (2), we get, for every  $t \geq t_0$ ,

$$\begin{aligned} \dot{\mathcal{E}}(t) = & \frac{(\alpha - 1)(\xi - t)}{2} \langle x(t) - \bar{x}, T_{\lambda(t),\gamma(t)}(x(t)) \rangle + \frac{(1 - \alpha)t}{2} \|\dot{x}(t)\|^2 \\ & + \left( -t^2 + \frac{\xi(3 - \alpha)t}{2} \right) \langle T_{\lambda(t),\gamma(t)}(x(t)), \dot{x}(t) \rangle + \xi(\xi - t)t \|T_{\lambda(t),\gamma(t)}(x(t))\|^2. \end{aligned}$$

Now, by Lemma 2.2(i), we know that  $T_{\lambda(t),\gamma(t)}$  is  $\frac{\lambda(t)}{2}$ -cocoercive for every  $t \geq t_0$ . Using this on the first summand of the right hand side of the previous inequality yields, for  $t \geq t_1 = \max\{\xi, t_0\}$ ,

$$\begin{aligned} \dot{\mathcal{E}}(t) \leq & \frac{(1 - \alpha)t}{2} \|\dot{x}(t)\|^2 + \left( -t^2 + \frac{\xi(3 - \alpha)t}{2} \right) \langle T_{\lambda(t),\gamma(t)}(x(t)), \dot{x}(t) \rangle \\ & + \left( \frac{(\alpha - 1)(\xi - t)\lambda(t)}{4} + \xi(\xi - t)t \right) \|T_{\lambda(t),\gamma(t)}(x(t))\|^2. \end{aligned} \tag{10}$$

Now, since  $\lambda > \frac{2}{(\alpha - 1)^2}$ , we can choose  $\epsilon > 0$  such that

$$0 < \epsilon < \alpha - 1 - \sqrt{\frac{2}{\lambda}} < \alpha - 1. \tag{11}$$

From (10) we get, for every  $t \geq t_1$ ,

$$\begin{aligned} \dot{\mathcal{E}}(t) + & \frac{\epsilon}{2}t\|\dot{x}(t)\|^2 + \frac{\epsilon}{4}t\lambda(t) \|T_{\lambda(t),\gamma(t)}(x(t))\|^2 \\ \leq & \left( \frac{1 - \alpha}{2} + \frac{\epsilon}{2} \right) t\|\dot{x}(t)\|^2 + \left( -t^2 + \frac{\xi(3 - \alpha)t}{2} \right) \langle T_{\lambda(t),\gamma(t)}(x(t)), \dot{x}(t) \rangle \\ & + \left( \left( \frac{(\alpha - 1)(\xi - t)}{2} + \frac{\epsilon}{2}t \right) \frac{\lambda(t)}{2} + \xi(\xi - t)t \right) \|T_{\lambda(t),\gamma(t)}(x(t))\|^2. \end{aligned} \tag{12}$$

By (11) and the definition of  $\lambda(t)$ , we know that  $\frac{1 - \alpha}{2} + \frac{\epsilon}{2} < 0$ , and

$$\left( \left( \frac{(\alpha - 1)(\xi - t)}{2} + \frac{\epsilon}{2}t \right) \frac{\lambda(t)}{2} + \xi(\xi - t)t \right) = \underbrace{\left( \frac{1 - \alpha}{2} + \frac{\epsilon}{2} \right)}_{<0} \frac{\lambda}{2} t^3 + \mathcal{O}(t^2),$$

so we can find  $t_2 \geq t_1$  such that for every  $t \geq t_2$  the previous expression becomes nonpositive. According to Lemma A.2, the right hand side of (12) is nonpositive whenever

$$\begin{aligned} R(t) := & \left( -t^2 + \frac{\xi(3 - \alpha)t}{2} \right)^2 - 4 \left( \frac{1 - \alpha}{2} + \frac{\epsilon}{2} \right) \\ & t \left( \left( \frac{(\alpha - 1)(\xi - t)}{2} + \frac{\epsilon}{2}t \right) \frac{\lambda(t)}{2} + \xi(\xi - t)t \right) \leq 0. \end{aligned}$$

This quantity can be rewritten as

$$R(t) = \left(1 + 4 \left(\frac{1 - \alpha}{2} + \frac{\epsilon}{2}\right) \left(\frac{\alpha - 1}{2} - \frac{\epsilon}{2}\right) \frac{\lambda}{2}\right) t^4 + \mathcal{O}(t^3) \quad \text{as } t \rightarrow +\infty.$$

Since  $\epsilon < \alpha - 1 - \sqrt{\frac{2}{\lambda}}$ , we have  $\frac{\lambda}{2} > \frac{1}{(\alpha - 1 - \epsilon)^2}$ . Hence,

$$1 + 4 \left(\frac{1 - \alpha}{2} + \frac{\epsilon}{2}\right) \left(\frac{\alpha - 1}{2} - \frac{\epsilon}{2}\right) \frac{\lambda}{2} = 1 - (\alpha - 1 - \epsilon)^2 \frac{\lambda}{2} < 0.$$

This means we can find  $t_3 \geq t_2$  such that for every  $t \geq t_3$  we have  $R(t) \leq 0$ , that is, for every  $t \geq t_3$  we have

$$\dot{\mathcal{E}}(t) + \frac{\epsilon}{2} t \|\dot{x}(t)\|^2 + \frac{\epsilon}{4} t \lambda(t) \|T_{\lambda(t), \gamma(t)}(x(t))\|^2 \leq 0. \tag{13}$$

Now, integrating (13) from  $t_3$  to  $t$  we obtain

$$\mathcal{E}(t) + \frac{\epsilon}{2} \int_{t_3}^t s \|\dot{x}(s)\|^2 ds + \frac{\epsilon}{4} \lambda \int_{t_3}^t s^3 \|T_{\lambda(s), \gamma(s)}(x(s))\|^2 ds \leq \mathcal{E}(t_3). \tag{14}$$

From (13) and the form of  $\mathcal{E}$  we immediately obtain

$$t \mapsto \|x(t) - \bar{x}\| \text{ is bounded,} \tag{15}$$

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \tag{16}$$

$$\int_{t_0}^{+\infty} t^3 \|T_{\lambda(t), \gamma(t)}(x(t))\|^2 dt < +\infty, \tag{17}$$

$$\sup_{t \geq t_0} \left\| \left(\frac{\alpha - 1}{2}\right) (x(t) - \bar{x}) + t (\dot{x}(t) + \xi T_{\lambda(t), \gamma(t)}(x(t))) \right\| < +\infty. \tag{18}$$

From Lemma 2.2(i), we know that for every  $t \geq t_0$  the operator  $T_{\lambda(t), \gamma(t)}$  is  $\frac{2}{\lambda(t)}$ -Lipschitz continuous, which gives, for every  $t \geq t_0$ ,

$$\|T_{\lambda(t), \gamma(t)}(x(t))\| = \|T_{\lambda(t), \gamma(t)}(x(t)) - T_{\lambda(t), \gamma(t)}(\bar{x})\| \leq \frac{2}{\lambda(t)} \|x(t) - \bar{x}\|.$$

Thus, from (15) and recalling that  $\lambda(t) = \lambda t^2$  we arrive at

$$\|T_{\lambda(t), \gamma(t)}(x(t))\| = \mathcal{O}\left(\frac{1}{t^2}\right) \quad \text{as } t \rightarrow +\infty. \tag{19}$$

By combining (15), (18) and (19) we obtain  $\sup_{t \geq t_0} t \|\dot{x}(t)\| < +\infty$  and therefore

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right) \quad \text{as } t \rightarrow +\infty. \tag{20}$$

From Lemma 2.2, (15), (20) and the fact that  $B$  is  $\frac{1}{\beta}$ -Lipschitz continuous we deduce that, as  $t \rightarrow +\infty$ ,

$$\left\| \frac{d}{dt} \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\| \leq 4 \|\dot{x}(t)\| + 4\beta \frac{|\dot{\gamma}(t)|}{\gamma(t)} \|B(x(t))\| + 2 \frac{|\dot{\gamma}(t)|}{\gamma(t)} \|x(t) - \bar{x}\| = \mathcal{O}\left(\frac{1}{t}\right). \tag{21}$$

On the other hand, for every  $t \geq t_0$  we have

$$\left\| \frac{d}{dt} \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\| = \left\| \dot{\lambda}(t) T_{\lambda(t), \gamma(t)}(x(t)) + \lambda(t) \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\|, \tag{22}$$

so by combining (19), (21), (22) and the fact that  $\dot{\lambda}(t) = 2\lambda t$  we arrive at

$$\left\| \lambda(t) \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\| \leq \underbrace{\left\| \frac{d}{dt} \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|}_{\mathcal{O}\left(\frac{1}{t}\right)} + \underbrace{\dot{\lambda}(t) \left\| T_{\lambda(t), \gamma(t)}(x(t)) \right\|}_{\mathcal{O}\left(\frac{1}{t^2}\right)} = \mathcal{O}\left(\frac{1}{t}\right)$$

as  $t \rightarrow +\infty$ ,

which yields

$$\left\| \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\| = \frac{1}{\lambda(t)} \mathcal{O}\left(\frac{1}{t}\right) = \mathcal{O}\left(\frac{1}{t^3}\right) \text{ as } t \rightarrow +\infty. \tag{23}$$

Let us now improve (19) and show that

$$\left\| T_{\lambda(t), \gamma(t)}(x(t)) \right\| = o\left(\frac{1}{t^2}\right) \text{ as } t \rightarrow +\infty. \tag{24}$$

According to (19) and (21) there exists a constant  $K > 0$  such that for every  $t \geq t_0$  it holds

$$\begin{aligned} \left| \frac{d}{dt} \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^4 \right| &= \left| 4 \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2 \right. \\ &\quad \times \left. \left\langle \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)), \frac{d}{dt} \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\rangle \right| \\ &\leq 4 \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2 \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\| \\ &\quad \times \left\| \frac{d}{dt} \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\| \\ &\leq \frac{4K}{t} \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2. \end{aligned}$$

By (17), the right hand side belongs to  $L^1([t_0, +\infty), \mathbb{R})$ , so we get

$$\frac{d}{dt} \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^4 \in L^1([t_0, +\infty), \mathbb{R}),$$

hence the limit

$$\lim_{t \rightarrow +\infty} \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^4$$

exists. Obviously, this implies the existence of  $L := \lim_{t \rightarrow +\infty} \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2$ . By using (17) again we come to

$$\int_{t_0}^{+\infty} \frac{1}{t} \left\| \lambda(t) T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2 dt = \lambda^2 \int_{t_0}^{+\infty} t^3 \left\| T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2 dt < +\infty,$$

and so we must have  $L = 0$ , which gives

$$\left\| T_{\lambda(t), \gamma(t)}(x(t)) \right\| = o\left(\frac{1}{t^2}\right) \text{ as } t \rightarrow +\infty. \tag{25}$$

By combining (2), (19), (20) and (23) we obtain, as  $t \rightarrow +\infty$ ,

$$\left\| \ddot{x}(t) \right\| = \left\| -\frac{\alpha}{t} \dot{x}(t) - \xi \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) - T_{\lambda(t), \gamma(t)}(x(t)) \right\|$$

$$\leq \frac{\alpha}{t} \underbrace{\|\dot{x}(t)\|}_{\mathcal{O}\left(\frac{1}{t}\right)} + \xi \underbrace{\left\| \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\|}_{\mathcal{O}\left(\frac{1}{t^3}\right)} + \underbrace{\|T_{\lambda(t), \gamma(t)}(x(t))\|}_{\mathcal{O}\left(\frac{1}{t}\right)} = \mathcal{O}\left(\frac{1}{t^2}\right).$$

Moreover, by using the well-known inequality  $\|a + b + c\|^2 \leq 3\|a\|^2 + 3\|b\|^2 + 3\|c\|^2$  for every  $a, b, c \in \mathcal{H}$ , for every  $t \geq t_0$  it holds

$$\begin{aligned} t^3 \|\ddot{x}(t)\|^2 &\leq t^3 \left\| -\frac{\alpha}{t} \dot{x}(t) - \xi \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) - T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2 \\ &\leq 3\alpha t \|\dot{x}(t)\|^2 + 3\xi^2 t^3 \left\| \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\|^2 + 3t^3 \|T_{\lambda(t), \gamma(t)}(x(t))\|^2. \end{aligned}$$

From (16), (23) and (17) it follows

$$\int_{t_0}^{+\infty} t^3 \|\ddot{x}(t)\|^2 dt < +\infty. \tag{26}$$

To see that  $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$  as  $t \rightarrow +\infty$ , we write, for every  $t \geq t_0$ ,

$$\frac{d}{dt} (t^2 \|\dot{x}(t)\|^2) = 2t \|\dot{x}(t)\|^2 + 2t^2 \langle \dot{x}(t), \ddot{x}(t) \rangle \leq 3t \|\dot{x}(t)\|^2 + t^3 \|\ddot{x}(t)\|^2.$$

From (16) and (26) we deduce that the left hand side belongs to  $L^1([t_0, +\infty), \mathbb{R})$ , from which we infer that the limit  $\lim_{t \rightarrow +\infty} t^2 \|\dot{x}(t)\|^2$  exists. Using (16) again, we get

$$\int_{t_0}^{+\infty} \frac{1}{t} (t^2 \|\dot{x}(t)\|^2) dt = \int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty,$$

from which we finally deduce  $\lim_{t \rightarrow +\infty} t^2 \|\dot{x}(t)\|^2 = 0$ , therefore

$$\|\dot{x}(t)\| = o\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty. \tag{27}$$

Notice that we can write for every  $t \geq t_0$

$$T_{\lambda(t), \gamma(t)} = \frac{1}{\lambda(t)} \left[ \text{Id} - J_{\gamma(t)A} (\text{Id} - \gamma(t)B) \right] = \frac{\gamma(t)}{\lambda(t)} \left( A_{\gamma(t)} [x(t) - \gamma(t)Bx(t)] + Bx(t) \right).$$

Hence, multiplying both sides of (25) by  $\frac{\lambda(t)}{\gamma(t)}$  and remembering the definition of  $\lambda(t)$  we obtain

$$\left\| A_{\gamma(t)} [x(t) - \gamma(t)Bx(t)] + Bx(t) \right\| = o\left(\frac{1}{\gamma(t)}\right) \text{ as } t \rightarrow +\infty. \tag{28}$$

For every  $t \geq t_0$ , we have

$$\begin{aligned} \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) &= \frac{d}{dt} \left( \frac{\gamma(t)}{\lambda(t)} \right) \left( A_{\gamma(t)} [x(t) - \gamma(t)Bx(t)] + Bx(t) \right) \\ &\quad + \frac{\gamma(t)}{\lambda(t)} \frac{d}{dt} \left( A_{\gamma(t)} [x(t) - \gamma(t)Bx(t)] + Bx(t) \right). \end{aligned}$$

Therefore, by using (23) and (28), and recalling that  $\lambda(t) = \lambda t^2$ , we obtain

$$\left\| \frac{d}{dt} \left( A_{\gamma(t)} \left[ x(t) - \gamma(t) Bx(t) \right] + Bx(t) \right) \right\| = \mathcal{O} \left( \frac{1}{t\gamma(t)} \right) + o \left( \frac{t^2 \left| \frac{d}{dt} \frac{\gamma(t)}{\lambda(t)} \right|}{\gamma^2(t)} \right) \text{ as } t \rightarrow +\infty.$$

The fact that  $\|\ddot{x}(t)\| = \mathcal{O} \left( \frac{1}{t^2} \right)$  as  $t \rightarrow +\infty$  comes from (2), (27), (23) and (24).

*Weak Convergence of the Trajectories.* Let  $\bar{x} \in \text{zer}(A + B)$ . We will work with the energy function  $h : [t_0, +\infty) \rightarrow \mathbb{R}$  given by

$$h(t) := \frac{1}{2} \|x(t) - \bar{x}\|^2.$$

For every  $t \geq t_0$ , we have

$$\dot{h}(t) = \langle x(t) - \bar{x}, \dot{x}(t) \rangle, \quad \ddot{h}(t) = \langle x(t) - \bar{x}, \ddot{x}(t) \rangle + \|\dot{x}(t)\|^2. \tag{29}$$

Combining (2) and (29) gives us, for every  $t \geq t_0$ ,

$$\ddot{h}(t) + \frac{\alpha}{t} \dot{h}(t) + \langle T_{\lambda(t), \gamma(t)}(x(t)), x(t) - \bar{x} \rangle = \|\dot{x}(t)\|^2 + \left\langle -\xi \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)), x(t) - \bar{x} \right\rangle.$$

By using the  $\frac{\lambda(t)}{2}$ -cocoercitivity of  $T_{\lambda(t), \gamma(t)}$  on the left hand side, Cauchy–Schwarz on the right hand side and multiplying both sides by  $t$ , the previous inequality entails, for every  $t \geq t_0$ ,

$$t\ddot{h}(t) + \alpha\dot{h}(t) + t \frac{\lambda(t)}{2} \|T_{\lambda(t), \gamma(t)}(x(t))\| \leq t\|\dot{x}(t)\|^2 + \xi t \left\| \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\| \|x(t) - \bar{x}\| \quad \forall t \geq t_0.$$

Now, puttin together results in

$$k(t) := t\|\dot{x}(t)\|^2 + \xi t \left\| \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) \right\| \|x(t) - \bar{x}\| \in L^1([t_0, +\infty), \mathbb{R}).$$

Now apply Lemma A.1 with  $\theta(t) := t \frac{\lambda(t)}{2} \|T_{\lambda(t), \gamma(t)}(x(t))\|$  for every  $t \geq t_0$  to deduce that the limit

$$\lim_{t \rightarrow +\infty} h(t)$$

exists, which fulfills the first condition of Opial’s Lemma A.3.

Let us now move on to the second condition. Suppose  $\hat{x}$  is a weak sequential cluster point of  $t \mapsto x(t)$ , that is, there exists a sequence  $(t_n)_{n \in \mathbb{N}} \subseteq [t_0, +\infty)$  such that  $t_n \rightarrow +\infty$  and  $x_n := x(t_n)$  converges weakly to  $\hat{x}$  as  $n \rightarrow +\infty$ . Define

$$U_\gamma := \text{Id} - J_{\gamma A} \circ (\text{Id} - \gamma B).$$

According to (25), we have  $U_{\gamma(t)}(x(t)) = \lambda(t)T_{\lambda(t), \gamma(t)}(x(t)) \rightarrow 0$  as  $t \rightarrow +\infty$ . Now, since  $\gamma(t) \in [\delta, 2\beta - \delta]$  for all  $t \geq t_0$  for some  $\delta > 0$ , we can extract a subsequence  $(\gamma(t_{n_k}))_{k \in \mathbb{N}}$  such that  $\gamma(t_{n_k}) \rightarrow \bar{\gamma} \in (0, 2\beta)$  as  $k \rightarrow +\infty$ . We may assume without loss of generality then that  $\gamma_n := \gamma(t_n) \rightarrow \bar{\gamma}$  as  $n \rightarrow +\infty$ . We now have for every  $n \in \mathbb{N}$

$$\begin{aligned} \|U_{\gamma_n}(x_n) - U_{\bar{\gamma}}(x_n)\| &= \|J_{\gamma_n A}(x_n - \gamma_n B(x_n)) - J_{\bar{\gamma} A}(x_n - \bar{\gamma} B(x_n))\| \\ &= \|J_{\gamma_n A}(x_n - \gamma_n B(x_n)) - J_{\gamma_n A}(x_n - \bar{\gamma} B(x_n))\| \end{aligned}$$



$$\begin{aligned}
 &+ \|J_{\gamma_n A}(x_n - \bar{\gamma} B(x_n)) - J_{\bar{\gamma} A}(x_n - \bar{\gamma} B(x_n))\| \\
 &\leq |\bar{\gamma} - \gamma_n| \|B(x_n)\| + |\bar{\gamma} - \gamma_n| \|A_{\bar{\gamma}}(x_n - \bar{\gamma} B(x_n))\|.
 \end{aligned}$$

Now, since every weakly convergent sequence is bounded and the operators  $B$  and  $A_{\bar{\gamma}}$  are Lipschitz-continuous we deduce that the right-hand side of the previous inequality approaches zero as  $n \rightarrow +\infty$ , therefore getting

$$U_{\bar{\gamma}}(x_n) = U_{\gamma_n}(x_n) + (U_{\bar{\gamma}}(x_n) - U_{\gamma_n}(x_n)) \rightarrow 0$$

as  $n \rightarrow +\infty$ . Now, from the proof of part (i) of Lemma 2.2, we know that  $U_{\bar{\gamma}}$  is  $\frac{4\beta - \bar{\gamma}}{4\beta}$ -cocoercive, thus monotone and Lipschitz continuous and therefore maximally monotone. Summarizing, we have

1.  $U_{\bar{\gamma}}$  is maximally monotone and thus its graph is closed in the weak  $\times$  strong topology of  $\mathcal{H} \times \mathcal{H}$  (see [14, Proposition 20.38(ii)]),
2.  $x_n$  converges weakly to  $\hat{x}$  and  $U_{\bar{\gamma}}(x_n) \rightarrow 0$  as  $n \rightarrow +\infty$ ,

which allows us to conclude that  $U_{\bar{\gamma}}(\hat{x}) = 0$ , and gives finally  $\hat{x} \in \text{zer}(A + B)$ . Now we just invoke Opial’s Lemma to achieve that  $x(t)$  converges weakly to  $\bar{x}$  as  $t \rightarrow +\infty$  for some  $\bar{x} \in \text{zer}(A + B)$ . □

In the following subsections, we explore the particular cases  $B = 0$  and  $A = 0$ , and we will show improvements with respect to previous results from the literature addressing continuous time approaches to monotone inclusions.

### 3.1 The Case $B = 0$

If we let  $B = 0$  in the (Split-DIN-AVD) system (2), then, attached to the monotone inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in A(x),$$

we obtain the dynamics

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \frac{d}{dt} (A_{\lambda(t), \gamma(t)}(x(t)) + A_{\lambda(t), \gamma(t)}(x(t))) = 0, \tag{30}$$

where

$$A_{\lambda, \gamma}(x) = \frac{1}{\lambda} (\text{Id} - J_{\gamma A}).$$

We can state the following theorem.

**Theorem 3.2** *Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a maximally monotone operator such that  $\text{zer } A \neq \emptyset$ . Assume that  $\alpha > 1$ ,  $\xi \geq 0$ ,  $\lambda(t) = \lambda t^2$  for  $\lambda > \frac{1}{(\alpha-1)^2}$  and all  $t \geq t_0$ , and that  $\gamma : [t_0, +\infty) \rightarrow (0, +\infty)$  is a differentiable function that satisfies  $\frac{|\dot{\gamma}(t)|}{\gamma(t)} = \mathcal{O}(\frac{1}{t})$  as  $t \rightarrow +\infty$ . Then, for a solution  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  to (30), the following statements hold:*

- (i)  $x$  is bounded.
- (ii) We have the estimates

$$\begin{aligned}
 &\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} t^3 \|\ddot{x}(t)\|^2 dt < +\infty, \\
 &\int_{t_0}^{+\infty} \frac{\gamma^2(t)}{t} \|A_{\gamma(t)}(x(t))\|^2 dt < +\infty.
 \end{aligned}$$

(iii) We have the convergence rates

$$\|\dot{x}(t)\| = o\left(\frac{1}{t}\right), \quad \|\ddot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^2}\right),$$

$$\|A_{\gamma(t)}(x(t))\| = o\left(\frac{1}{\gamma(t)}\right), \quad \left\|\frac{d}{dt}A_{\gamma(t)}(x(t))\right\| = \mathcal{O}\left(\frac{1}{t\gamma(t)}\right) + o\left(\frac{t^2 \left|\frac{d}{dt}\frac{\gamma(t)}{\lambda(t)}\right|}{\gamma^2(t)}\right)$$

as  $t \rightarrow +\infty$ .

(iv) If  $0 < \inf_{t \geq t_0} \gamma(t)$ , then  $x(t)$  converges weakly to an element of  $\text{zer } A$  as  $t \rightarrow +\infty$ .

**Proof** The proof proceeds in the exact same way as the proof of Theorem 3.1. However, a few comments are in order: first of all, now we have  $T_{\lambda,\gamma} = \frac{1}{\lambda}(\text{Id} - J_{\gamma A}) = A_{\lambda,\gamma}$ . Since  $J_{\lambda A}$  is firmly nonexpansive, by [14, Proposition 4.4] so is  $\text{Id} - J_{\lambda A}$ . In other words,  $\text{Id} - J_{\gamma A}$  is 1-cocoercive, therefore  $A_{\lambda,\gamma} = \frac{1}{\lambda}(\text{Id} - J_{\gamma A})$  is  $\lambda$ -cocoercive, so now the condition on  $\lambda$  becomes  $\lambda > \frac{1}{(\alpha-1)^2}$ .

The proof also changes when we verify the second part of the Opial’s Lemma, to get weak convergence of the trajectories  $t \mapsto x(t)$ . This is in order to allow for  $\gamma(t)$  not to be necessarily bounded. We do need, however, the assumption  $0 < \inf_{t \geq t_0} \gamma(t)$ . Indeed, from  $\|A_{\lambda(t),\gamma(t)}(x(t))\| = o\left(\frac{1}{t^2}\right)$  as  $t \rightarrow +\infty$ , we obtain

$$y(t) := x(t) - J_{\gamma(t)A}x(t) = \lambda(t)A_{\lambda(t),\gamma(t)}(x(t)) \rightarrow 0$$

as  $t \rightarrow +\infty$ . Using the definition of the resolvent, we come to

$$J_{\gamma(t)A}x(t) = x(t) - y(t) \Leftrightarrow y(t) \in \gamma(t)A(x(t) - y(t)) \Leftrightarrow \frac{1}{\gamma(t)}y(t) \in A(x(t) - y(t)).$$

for all  $t \geq t_0$ . If  $(t_n)_{n \in \mathbb{N}} \subseteq [t_0, +\infty)$  is such that  $t_n \rightarrow +\infty$  and  $x(t_n)$  converges weakly to  $\widehat{x}$  as  $n \rightarrow +\infty$ , then the previous inclusion, together with the assumption on  $\gamma$  gives

$$x(t_n) - y(t_n) \text{ converges weakly to } \widehat{x} \quad \text{and} \quad \frac{1}{\gamma(t)}y(t) \rightarrow 0 \quad \text{as } n \rightarrow +\infty,$$

and by the closedness of the graph of  $A$  in the weak  $\times$  strong topology of  $\mathcal{H} \times \mathcal{H}$ , we deduce that  $\widehat{x} \in \text{zer } A$ . □

**Remark 3.3** The hypotheses required for  $\gamma$  are fulfilled at least by two families of functions. First, take  $r \geq 0$  and set  $\gamma(t) = e^{t^{-r}}$ . Then, we have

$$\frac{\dot{\gamma}(t)}{\gamma(t)} = \frac{-rt^{-(r+1)}e^{t^{-r}}}{e^{t^{-r}}} = -\frac{r}{t^{r+1}} = \mathcal{O}\left(\frac{1}{t}\right) \quad \text{as } t \rightarrow +\infty,$$

and

$$\gamma(t) = e^{t^{-r}} \geq e^0 = 1 \quad \forall t \geq 0.$$

If  $\gamma$  is a polynomial of degree  $n$  for some  $n \in \mathbb{N}$ , the conditions are also fulfilled. Assume  $\gamma(t) = a_n t^n + a_{n-1} t^{n-1} + \dots + a_0$  for all  $t \geq t_0$ , for some  $a_i \in \mathbb{R}$  for  $i \in \{0, \dots, n\}$  and  $a_n > 0$ . Then, we have

$$t \cdot \frac{\dot{\gamma}(t)}{\gamma(t)} = t \cdot \frac{na_n t^{n-1} + (n-1)a_{n-1} t^{n-2} + \dots + a_1}{a_n t^n + a_{n-1} t^{n-1} + \dots + a_0}$$

$$\rightarrow \frac{na_n}{a_n} = n \quad \text{as } t \rightarrow +\infty,$$

so  $\frac{\dot{\gamma}(t)}{\gamma(t)} = \mathcal{O}\left(\frac{1}{t}\right)$  as  $t \rightarrow +\infty$ . Since we also have  $\gamma(t) \rightarrow +\infty$  as  $t \rightarrow +\infty$ , the condition  $\inf_{t \geq t_0} \gamma(t) > 0$  is fulfilled for large enough  $t_0$ .

In particular, we can choose  $\gamma(t) = \lambda(t) = \lambda t^2$ , which fulfills  $\gamma(t) \geq \lambda t_0^2 > 0$  for any  $t \geq t_0$  and any  $t_0$ . Since  $A_{\lambda, \lambda} = A_\lambda$  for  $\lambda > 0$ , this choice of  $\gamma$  allows us to recover the (DIN-AVD) system studied by Attouch and László in [9]. Notice the way the convergence rates for  $A_{\gamma(t)}(x(t))$  and  $\frac{d}{dt}A_{\gamma(t)}(x(t))$  exhibited in part (iii) of Theorem 3.2 depend on  $\gamma(t)$ . If we set  $\gamma(t) = t^n$  for every  $t \geq t_0$  for any natural number  $n > 2$ , (Split-DIN-AVD) performs from this point of view better than (DIN-AVD) without increasing the complexity of the governing operator.

### 3.2 The Case $A = 0$

Let us return to (Split-DIN-AVD) dynamics (2). Set  $A = 0$ , and for every  $t \geq t_0$  take  $\gamma(t) = \gamma \in (0, 2\beta)$  and  $\eta(t) = \eta t^2$  with  $\eta = \lambda/\gamma$ . Then, associated to the problem

$$\text{find } x \in \mathcal{H} \text{ such that } B(x) = 0,$$

we obtain the system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \xi \frac{d}{dt} \left( \frac{1}{\eta(t)} Bx(t) \right) + \frac{1}{\eta(t)} Bx(t) = 0. \tag{31}$$

The conditions  $\lambda > \frac{2}{(\alpha-1)^2}$  and  $\gamma \in (0, 2\beta)$  imply

$$\eta = \frac{\lambda}{\gamma} > \frac{2}{\gamma(\alpha-1)^2} > \frac{2}{2\beta(\alpha-1)^2} = \frac{1}{\beta(\alpha-1)^2}.$$

With the previous observation, we are able to state the following theorem.

**Theorem 3.4** *Let  $B : \mathcal{H} \rightarrow \mathcal{H}$  be a  $\beta$ -cocoercive operator for some  $\beta > 0$  such that  $\text{zer } B \neq \emptyset$ . Assume that  $\alpha > 1$ ,  $\xi \geq 0$  and  $\eta(t) = \eta t^2$  for  $\eta > \frac{1}{\beta(\alpha-1)^2}$  and all  $t \geq t_0$ . Take  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  a solution to (31). Then, the following hold:*

- (i)  $x$  is bounded, and  $x(t)$  converges weakly to an element of  $\text{zer } B$  as  $t \rightarrow +\infty$ .
- (ii) We have the estimates

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} t^3 \|\ddot{x}(t)\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} \frac{1}{t} \|Bx(t)\|^2 dt < \infty.$$

- (iii) We have the convergence rates

$$\|\dot{x}(t)\| = o\left(\frac{1}{t}\right), \quad \|\ddot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^2}\right)$$

as well as the limit

$$\|Bx(t)\| \rightarrow 0$$

as  $t \rightarrow +\infty$ .

**Proof** Since  $\eta > \frac{1}{\beta(\alpha-1)^2}$ , we can find  $\epsilon \in (0, \beta)$  such that  $\eta > \frac{1}{(\beta-\epsilon)(\alpha-1)^2}$ , equivalently,  $2(\beta-\epsilon)\eta > \frac{2}{(\alpha-1)^2}$ . Since (31) is equivalent to (Split-DIN-AVD) with  $A = 0$  and parameters  $\lambda = 2(\beta-\epsilon)\eta > \frac{2}{(\alpha-1)^2}$  and  $\gamma(t) \equiv 2(\beta-\epsilon) \in (0, 2\beta)$ , the conclusion follows from Theorem 3.1. □

**Remark 3.5** (a) As we mentioned in the introduction, the dynamical system (31) provides a way of finding the zeros of a cocoercive operator directly through forward evaluations, instead of having to resort to its Moreau envelope when following the approach in [9].  
 (b) The dynamics (31) bear some resemblance to the system (6) (see also [16]) with  $\mu(t) = \frac{\alpha}{t}$  and  $\nu(t) = \frac{1}{\eta(t)}$ , with an additional Hessian-driven damping term. In our case, since  $\eta > \frac{1}{\beta(\alpha-1)^2}$ , the parameters satisfy

$$\dot{\mu}(t) = -\frac{\alpha}{t^2} \leq 0, \quad \frac{\mu^2(t)}{\nu(t)} = \frac{\alpha^2 \eta t^2}{t^2} = \alpha^2 \eta > \frac{1}{\beta} \quad \forall t \geq t_0.$$

However, we have

$$\dot{\nu}(t) = -\frac{2}{\lambda t^3} \leq 0 \quad \forall t \geq t_0,$$

so one of the hypotheses which is needed in (6) is not fulfilled, which shows that one cannot address the dynamical system (31) as a particular case of it; indeed, for (6) a vanishing damping is not allowed. With our system, we obtain convergence rates for  $\dot{x}(t)$  and  $\ddot{x}(t)$  as  $t \rightarrow +\infty$ , which are not obtained in [16].

### 4 Structured Convex Minimization

We can specialize the previous results to the case of convex minimization, and show additionally the convergence of functional values along the generated trajectories to the optimal objective value at a rate that will depend on the choice of  $\gamma$ . Let  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper, convex and lower semicontinuous function, and let  $g : \mathcal{H} \rightarrow \mathbb{R}$  be a convex and Fréchet differentiable function with  $L_{\nabla g}$ -Lipschitz continuous gradient. Assume that  $\text{argmin}_{\mathcal{H}}(f + g) \neq \emptyset$ , and consider the minimization problem

$$\min_{x \in \mathcal{H}} f(x) + g(x). \tag{32}$$

Fermat’s rule tells us that  $\bar{x}$  is a global minimum of  $f + g$  if and only if

$$0 \in \partial(f + g)(\bar{x}) = \partial f(\bar{x}) + \nabla g(\bar{x}).$$

Therefore, solving (32) is equivalent solving the monotone inclusion  $0 \in (A + B)(x)$  addressed in the first section, with  $A = \partial f$  and  $B = \nabla g$ . Moreover, recall that if  $\nabla g$  is  $L_{\nabla g}$ -Lipschitz then it is  $\frac{1}{L_{\nabla g}}$ -cocoercive (Baillon–Haddad’s Theorem, see [14, Corollary 18.17]). Therefore, associated to the problem (32) we have the dynamics

$$\begin{aligned} \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \frac{d}{dt} \left( \frac{\gamma(t)}{\lambda(t)} (\nabla f_{\gamma(t)}(u(t)) + \nabla g(x(t))) \right) \\ + \frac{\gamma(t)}{\lambda(t)} (\nabla f_{\gamma(t)}(u(t)) + \nabla g(x(t))) = 0, \end{aligned} \tag{33}$$

where we have denoted  $u(t) = x(t) - \gamma(t)\nabla g(x(t))$  for all  $t \geq t_0$  for convenience.

**Theorem 4.1** *Let  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper, convex and lower semicontinuous function, and let  $g : \mathcal{H} \rightarrow \mathbb{R}$  be a convex and Fréchet differentiable function with a  $L_{\nabla g}$ -Lipschitz continuous gradient such that  $\text{argmin}_{\mathcal{H}}(f + g) \neq \emptyset$ . Assume that  $\alpha > 1$ ,  $\xi \geq 0$ ,  $\lambda(t) = \lambda t^2$  for  $\lambda > \frac{2}{(\alpha-1)^2}$  and all  $t \geq t_0$ , and that  $\gamma : [t_0, +\infty) \rightarrow \left(0, \frac{2}{L_{\nabla g}}\right)$  is*

a differentiable function that satisfies  $\frac{\dot{\gamma}(t)}{\gamma(t)} = \mathcal{O}(1/t)$  as  $t \rightarrow +\infty$ . Then, for a solution  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  to (33), the following statements hold:

- (i)  $x$  is bounded.
- (ii) We have the estimates

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} t^3 \|\ddot{x}(t)\|^2 dt < +\infty,$$

$$\int_{t_0}^{+\infty} \frac{\gamma^2(t)}{t} \left\| \nabla f_{\gamma(t)} [x(t) - \gamma(t) \nabla g(x(t))] + \nabla g(x(t)) \right\|^2 dt < +\infty.$$

- (iii) We have the convergence rates

$$\| \dot{x}(t) \| = o\left(\frac{1}{t}\right), \quad \| \ddot{x}(t) \| = \mathcal{O}\left(\frac{1}{t^2}\right),$$

$$\left\| \nabla f_{\gamma(t)} [x(t) - \gamma(t) \nabla g(x(t))] + \nabla g(x(t)) \right\| = o\left(\frac{1}{\gamma(t)}\right),$$

$$\left\| \frac{d}{dt} \left( \nabla f_{\gamma(t)} [x(t) - \gamma(t) \nabla g(x(t))] + \nabla g(x(t)) \right) \right\| = \mathcal{O}\left(\frac{1}{t\gamma(t)}\right) + o\left(\frac{t^2 \left| \frac{d}{dt} \frac{\gamma(t)}{\lambda(t)} \right|}{\gamma^2(t)}\right)$$

as  $t \rightarrow +\infty$ .

- (iv) If  $0 < \inf_{t \geq t_0} \gamma(t) \leq \sup_{t \geq t_0} \gamma(t) < \frac{2}{L_{\nabla g}}$ , then  $x(t)$  converges to a minimizer of  $f + g$  as  $t \rightarrow +\infty$ .
- (v) Additionally, if  $0 < \gamma(t) \leq \frac{1}{L_{\nabla g}}$  for every  $t \geq t_0$  and we set  $u(t) := x(t) - \gamma(t) \nabla g(x(t))$ , then

$$f(\text{prox}_{\gamma(t)f}(u(t))) + g(\text{prox}_{\gamma(t)f}(u(t))) - \min_{\mathcal{H}}(f + g) = o\left(\frac{1}{\gamma(t)}\right)$$

as  $t \rightarrow +\infty$ . Moreover,  $\|\text{prox}_{\gamma(t)f}(u(t)) - x(t)\| \rightarrow 0$  as  $t \rightarrow +\infty$ .

**Proof** Parts (i)–(iv) are a direct consequence of Theorem 3.1. For checking (v), first notice that for all  $t \geq t_0$  we have

$$\begin{aligned} T_{\lambda(t), \gamma(t)}(x(t)) &= \frac{1}{\lambda(t)} \left[ \text{Id} - J_{\gamma(t)\partial f} \circ (\text{Id} - \gamma(t) \nabla g) \right](x(t)) \\ &= \frac{1}{\lambda(t)} \left[ x(t) - \text{prox}_{\gamma(t)f}(u(t)) \right]. \end{aligned} \tag{34}$$

Now, let  $\bar{x} \in \text{argmin}_{\mathcal{H}}(f + g)$ . According to [15, Lemma 2.3], for every  $t \geq t_0$ , we have the inequality

$$\begin{aligned} & f(\text{prox}_{\gamma(t)f}(u(t))) + g(\text{prox}_{\gamma(t)f}(u(t))) - \min_{\mathcal{H}}(f + g) \\ & \leq f(\text{prox}_{\gamma(t)f}(u(t))) + g(\text{prox}_{\gamma(t)f}(u(t))) - f(\bar{x}) - g(\bar{x}) \\ & \leq -\frac{1}{2\gamma(t)} \left\| \text{prox}_{\gamma(t)f}(u(t)) - x(t) \right\|^2 + \frac{1}{\gamma(t)} \langle x(t) - x^*, x(t) - \text{prox}_{\gamma(t)f}(u(t)) \rangle. \end{aligned}$$

After summing the norm squared term and using the Cauchy–Schwarz inequality, for every  $t \geq t_0$  we obtain

$$\frac{1}{2\gamma(t)} \left\| \text{prox}_{\gamma(t)f}(u(t)) - x(t) \right\|^2$$

$$\begin{aligned}
 &\leq f(\text{prox}_{\gamma(t)f}(u(t))) + g(\text{prox}_{\gamma(t)f}(u(t))) + \\
 &\frac{1}{2\gamma(t)} \|\text{prox}_{\gamma(t)f}(u(t)) - x(t)\|^2 - \min_{\mathcal{H}}(f + g) \\
 &\leq \left\langle \frac{1}{\gamma(t)}(x(t) - \text{prox}_{\gamma(t)f}(u(t))), x(t) - \bar{x} \right\rangle \\
 &\leq \left\| \frac{1}{\gamma(t)}(x(t) - \text{prox}_{\gamma(t)f}(u(t))) \right\| \|x(t) - \bar{x}\| \\
 &= \frac{\lambda(t)}{\gamma(t)} \|T_{\lambda(t),\gamma(t)}(x(t))\| \|x(t) - \bar{x}\| \\
 &= o\left(\frac{1}{\gamma(t)}\right) \text{ as } t \rightarrow +\infty,
 \end{aligned}$$

which follows as a consequence of  $x$  being bounded and  $\|T_{\lambda(t),\gamma(t)}(x(t))\| = o\left(\frac{1}{\sqrt{t}}\right)$  as  $t \rightarrow +\infty$ . □

**Remark 4.2** It is also worth mentioning the system we obtain in the case where  $g \equiv 0$ , since we also get some improved rates for the objective functional values when we compare (Split-DIN-AVD) to (DIN-AVD) [9]. In this case, we have the system

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \frac{d}{dt} \left( \frac{\gamma(t)}{\lambda(t)} \nabla f_{\gamma(t)}(x(t)) \right) + \frac{\gamma(t)}{\lambda(t)} \nabla f_{\gamma(t)}(x(t)) = 0 \tag{35}$$

attached to the convex optimization problem

$$\min_{x \in \mathcal{H}} f(x).$$

If we assume  $\lambda > \frac{1}{(\alpha-1)^2}$ , allow  $\gamma : [t_0, +\infty) \rightarrow (0, +\infty)$  to be unbounded from above and otherwise keep the hypotheses of Theorem 4.1, for a solution  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  to (35), the following statements hold:

- (i)  $x$  is bounded,
- (ii) We have the estimates

$$\begin{aligned}
 \int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_0}^{+\infty} t^3 \|\ddot{x}(t)\|^2 dt < +\infty, \\
 \int_{t_0}^{+\infty} \frac{\gamma^2(t)}{t} \|\nabla f_{\gamma(t)}(x(t))\|^2 dt < +\infty,
 \end{aligned}$$

- (iii) We have the convergence rates

$$\begin{aligned}
 \|\dot{x}(t)\| &= o\left(\frac{1}{t}\right), \quad \|\ddot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^2}\right), \\
 \|\nabla f_{\gamma(t)}(x(t))\| &= o\left(\frac{1}{\gamma(t)}\right), \quad \left\| \frac{d}{dt} \nabla f_{\gamma(t)}(x(t)) \right\| = \mathcal{O}\left(\frac{1}{t\gamma(t)}\right) + o\left(\frac{t^2 \left| \frac{d}{dt} \frac{\gamma(t)}{\lambda(t)} \right|}{\gamma^2(t)}\right)
 \end{aligned}$$

as  $t \rightarrow +\infty$ .

- (iv) If  $0 < \inf_{t \geq t_0} \gamma(t)$ , then  $x(t)$  converges weakly to a minimizer of  $f$  as  $t \rightarrow +\infty$ .

(v) We also obtain the rate

$$f_{\gamma(t)}(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{\gamma(t)}\right) \text{ as } t \rightarrow +\infty,$$

which entails

$$f(\text{prox}_{\gamma(t)f}(x(t))) - \min_{\mathcal{H}} f = o\left(\frac{1}{\gamma(t)}\right) \text{ and } \|\text{prox}_{\gamma(t)f}(x(t)) - x(t)\| \rightarrow 0$$

as  $t \rightarrow +\infty$ .

Parts (i)–(iv) are a direct consequence of Theorem 3.2 for the case  $A = \partial f$ . For showing part (v), first notice that for  $\lambda > 0$  and  $u \in \mathcal{H}$  we have, according to the definition of  $f_\lambda$  and  $\text{prox}_{\lambda f}$ ,

$$f_\lambda(u) = f(\text{prox}_{\lambda f}(u)) + \frac{1}{2\lambda} \|\text{prox}_{\lambda f}(u) - u\|^2 \leq f(u).$$

Let  $\bar{x} \in \mathcal{H}$  be a minimizer of  $f$ . We apply the gradient inequality to  $f_{\gamma(t)}$ , from which we obtain, for every  $t \geq t_0$

$$\begin{aligned} f_{\gamma(t)}(x(t)) - \min_{\mathcal{H}} f &= f_{\gamma(t)}(x(t)) - f(\bar{x}) \leq f_{\gamma(t)}(x(t)) - f_{\lambda(t)}(\bar{x}) \\ &\leq \langle \nabla f_{\gamma(t)}(x(t)), x(t) - \bar{x} \rangle \leq \|\nabla f_{\gamma(t)}(x(t))\| \|x(t) - \bar{x}\|, \end{aligned}$$

where the last inequality follows from the Cauchy–Schwarz inequality. Since  $\|\nabla f_{\gamma(t)}(x(t))\| = o\left(\frac{1}{\gamma(t)}\right)$  as  $t \rightarrow +\infty$  and  $x$  is bounded, the previous inequality entails the first statement of (v). Again recalling the definition of the Moreau envelope of  $f$ , this finally gives

$$\begin{aligned} f(\text{prox}_{\gamma(t)f}(x(t))) + \frac{1}{2\gamma(t)} \|\text{prox}_{\gamma(t)f}(x(t)) - x(t)\|^2 - \min_{\mathcal{H}} f &= f_{\gamma(t)}(x(t)) \\ - \min_{\mathcal{H}} f &= o\left(\frac{1}{\gamma(t)}\right) \end{aligned}$$

as  $t \rightarrow +\infty$ , which implies the last two statements and concludes the proof.

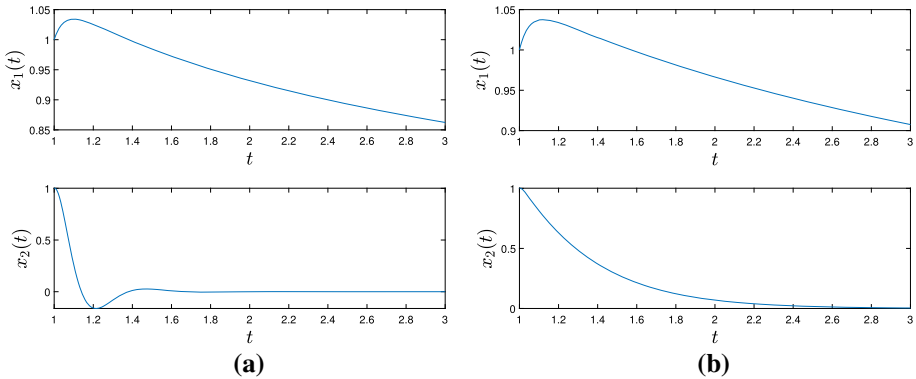
As pointed out in Remark 3.3, we can choose  $\gamma(t) = \lambda t^2$  for every  $t \geq t_0$  and recover the (DIN-AVD) system for nonsmooth convex minimization problems studied in [9]. Moreover, we can also set  $\gamma(t) = t^n$  for a natural number  $n > 3$  and all  $t \geq t_0$ . Now, not only are the convergence rates for  $\nabla f_{\gamma(t)}(x(t))$  and  $\frac{d}{dt} \nabla f_{\gamma(t)}(x(t))$  as  $t \rightarrow +\infty$  improved with respect to the system in [9], but (Split-DIN-AVD) also provides a better rate for the convergence of  $f_{\gamma(t)}(x(t))$  to  $\min_{\mathcal{H}} f$  as  $t \rightarrow +\infty$ .

## 5 Numerical Experiments

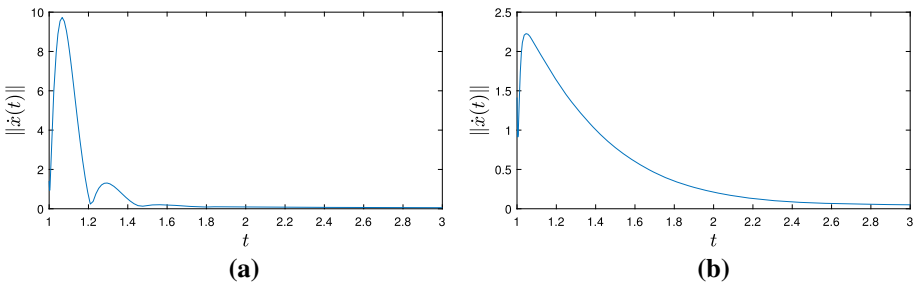
In the following paragraphs we describe some numerical experiments that portray some aspects of the theory.

### 5.1 Minimizing a Smooth and Convex Function

As an example of a continuous time scheme minimizing a convex and Fréchet differentiable function  $g : \mathcal{H} \rightarrow \mathbb{R}$  with  $L_{\nabla g}$ -Lipschitz continuous gradient via (Split-DIN-AVD), we



**Fig. 1** Trajectories of (Split-DIN-AVD) for  $B = \nabla g$



**Fig. 2** Fast convergence of the velocities

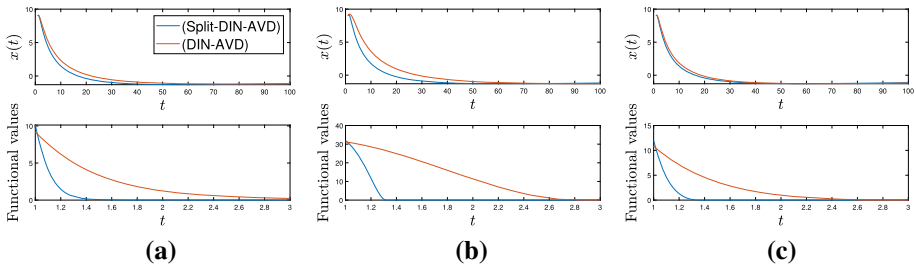
consider the system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \xi \frac{d}{dt} \left( \frac{1}{\eta(t)} \nabla g(x(t)) \right) + \frac{1}{\eta(t)} \nabla g(x(t)) = 0, \tag{36}$$

where for  $(x_1, x_2) \in \mathbb{R}^2$  we set  $g(x_1, x_2) = \frac{1}{2}(x_1^2 + 100x_2^2)$  and therefore  $\nabla g(x_1, x_2) = (x_1, 100x_2)$ . A trajectory generated by (36) is a pair  $x(t) = (x_1(t), x_2(t))$ . Figure 1 plots both components of the solution to (36) with initial Cauchy data  $x_0 = (1, 1), u_0 = (1, 1)$ . Notice that the Lipschitz constant of  $\nabla g$  is  $L_{\nabla g} = 100$ , which means that the cocoercivity modulus of  $\nabla g$  is  $\beta = \frac{1}{L_{\nabla g}} = \frac{1}{100}$ . To fulfill  $\eta > \frac{1}{\beta(\alpha-1)^2} = \frac{100}{(\alpha-1)^2}$ , we choose  $\alpha = 20, \eta = 0.278$ . Figure 1a corresponds to the case with no Hessian damping, that is,  $\xi = 0$ . Figure 1b corresponds to a Hessian damping parameter  $\xi = 0.2$ .

Figure 2 depicts the fast convergence of the velocities to zero for the cases  $\xi = 0$  (Fig. 2a) and  $\xi = 0.2$  (Fig. 2b). In both figures, notice the effect of the damping parameter  $\xi > 0$ , which attenuates the oscillations of the second component of the trajectories, as well as the oscillations present in the velocities.





**Fig. 3** Trajectories and objective function values in the case  $A = \partial f$

### 5.2 Minimizing a Nonsmooth and Convex Function

As an example of a continuous time scheme minimizing a proper, convex and lower semi-continuous function  $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$  via (Split-DIN-AVD), we consider the system

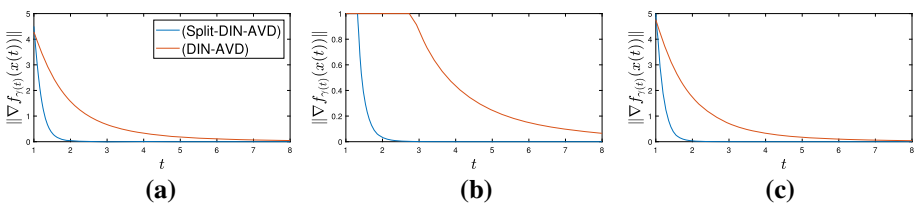
$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \xi \frac{d}{dt} \left( \frac{\gamma(t)}{\lambda(t)} \nabla f_{\gamma(t)}(x(t)) \right) + \frac{\gamma(t)}{\lambda(t)} \nabla f_{\gamma(t)}(x(t)) = 0. \tag{37}$$

We will consider three options for  $f$  and plot for each of them the trajectories, the objective function values and the gradients of the Moreau envelopes as follows:

- $f(x) = \frac{1}{2}x^2$  (Figs. 3a and 4a),
- $f(x) = |x|$  (Figs. 3b and 4b),
- $f(x) = |x| + \frac{1}{2}x^2$  (Figs. 3c and 4c).

In order to fulfill  $\alpha > 1$  and  $\lambda > \frac{1}{(\alpha-1)^2}$ , we choose the parameters  $\alpha = 2$ ,  $\lambda = 1.1$ , and we take  $\xi = 0$  and  $\gamma(t) = t^8$ . We compare the results given by (DIN-AVD) (that is, when  $\gamma(t) = \lambda t^2$ ) and the ones given by our system (Split-DIN-AVD). The choice of  $\xi$  does not seem to change the plots in a significant way for the examples we have chosen.

Figure 3 depicts the trajectories  $x(t)$  of (37) and the function values  $f(\text{prox}_{\gamma(t)}(x(t)))$  for our choices of  $f$  as  $t \rightarrow +\infty$ . Figure 4 portrays the fast convergence to zero of  $\|\nabla f_{\gamma(t)}(x(t))\|$  as  $t \rightarrow +\infty$ . Notice the big improvement over (DIN-AVD) for nonsmooth convex minimization in [9] when choosing  $\gamma(t) = t^8$ , a result which we already knew theoretically. Polynomials of high degree seem to be the ones which give the biggest improvements in terms of rates.



**Fig. 4** Gradients of the Moreau envelopes of  $f$

### 5.3 An Example with Operator Splitting

Now we consider the monotone inclusion problem (1) for  $A(x_1, x_2) = (-x_2, x_1)$  and  $B(x_1, x_2) = (x_1, x_2)$  for every  $(x_1, x_2) \in \mathbb{R}^2$ . For every  $(x_1, x_2) \in \mathbb{R}^2$ , an easy calculation gives

$$J_{\gamma A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{1+\gamma^2} & \frac{\gamma}{1+\gamma^2} \\ \frac{-\gamma}{1+\gamma^2} & \frac{1}{1+\gamma^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and so

$$\begin{aligned} & (\text{Id} - J_{\gamma A}(\text{Id} - \gamma \text{Id})) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - (1 - \gamma) \begin{bmatrix} \frac{1}{1+\gamma^2} & \frac{\gamma}{1+\gamma^2} \\ \frac{-\gamma}{1+\gamma^2} & \frac{1}{1+\gamma^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{\gamma^2+\gamma}{1+\gamma^2} & \frac{\gamma-1}{1+\gamma^2} \\ \frac{1-\gamma}{1+\gamma^2} & \frac{\gamma^2+\gamma}{1+\gamma^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \end{aligned}$$

and

$$T_{\lambda, \gamma} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{\gamma^2+\gamma}{\lambda(1+\gamma^2)} & \frac{\gamma-1}{\lambda(1+\gamma^2)} \\ \frac{1-\gamma}{\lambda(1+\gamma^2)} & \frac{\gamma^2+\gamma}{\lambda(1+\gamma^2)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

(Split-DIN-AVD) now reads

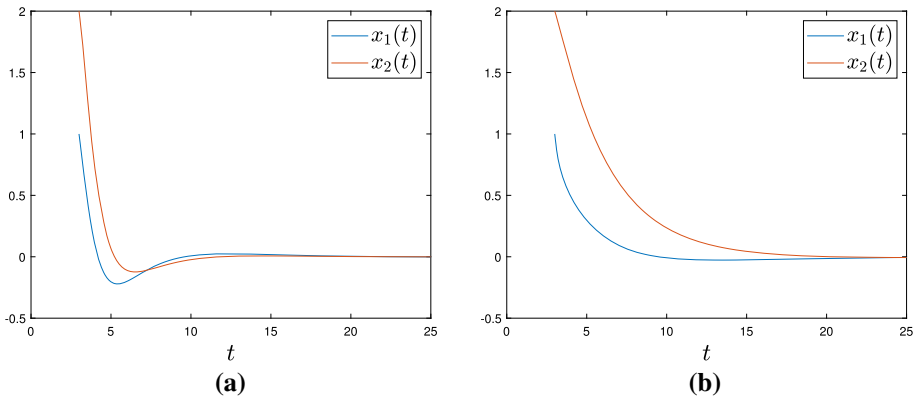
$$\begin{aligned} & \begin{bmatrix} \ddot{x}_1(t) \\ \ddot{x}_2(t) \end{bmatrix} + \frac{\alpha}{t} \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} + \xi \frac{d}{dt} \left( \begin{bmatrix} \frac{\gamma^2(t)+\gamma(t)}{\lambda(t)(1+\gamma^2(t))} & \frac{\gamma(t)-1}{\lambda(t)(1+\gamma^2(t))} \\ \frac{1-\gamma(t)}{\lambda(t)(1+\gamma^2(t))} & \frac{\gamma^2(t)+\gamma(t)}{\lambda(t)(1+\gamma^2(t))} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \right) \\ &+ \begin{bmatrix} \frac{\gamma^2(t)+\gamma(t)}{\lambda(t)(1+\gamma(t)^2)} & \frac{\gamma(t)-1}{\lambda(t)(1+\gamma^2(t))} \\ \frac{1-\gamma(t)}{\lambda(1+\gamma^2(t))} & \frac{\gamma^2(t)+\gamma(t)}{\lambda(t)(1+\gamma^2(t))} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

We choose the parameters  $\alpha = 7, \lambda = 0.056, \gamma(t) \equiv 1.5$ , and the Cauchy data  $x_0 = (1, 2)$  and  $u_0 = (-1, -1)$ . Figure 5a corresponds to the case  $\xi = 0$ , and Fig. 5b depicts the trajectory when the Hessian damping parameter is  $\xi = 0.8$ . Again, notice how, not only for optimization problems, but also for monotone inclusions which cannot be reduced to the former, the presence of  $\xi$  seems to attenuate the oscillations present in the trajectories.

### 6 A Numerical Algorithm

In the following we will derive via time discretization of (Split-DIN-AVD) a numerical algorithm for solving the monotone inclusion problem (1). We perform a discretization of (Split-DIN-AVD) with stepsize 1 and set, for an integer  $k \geq 1, x(k) := x_k, \lambda(k) := \lambda_k, \gamma(k) := \gamma_k$ . We make the approximations

$$\begin{aligned} \ddot{x}(t) &\approx x_{k+1} - 2x_k + x_{k-1}, & \frac{\alpha}{t} \dot{x}(t) &\approx \frac{\alpha}{k}(x_k - x_{k-1}), \\ \frac{d}{dt} T_{\lambda(t), \gamma(t)}(x(t)) &\approx T_{\lambda_k, \gamma_k}(x_k) - T_{\lambda_{k-1}, \gamma_{k-1}}(x_{k-1}), & T_{\lambda(t), \gamma(t)}(x(t)) &\approx T_{\lambda_{k+1}, \gamma_{k+1}}(x_{k+1}), \end{aligned}$$



**Fig. 5** Trajectories of (Split-DIN-AVD) for finding the zeros of  $A + B$

so we get, for every  $k \geq 1$ ,

$$x_{k+1} - 2x_k - x_{k-1} + \frac{\alpha}{k}(x_k - x_{k-1}) + \xi (T_{\lambda_k, \gamma_k}(x_k) - T_{\lambda_{k-1}, \gamma_{k-1}}(x_{k-1})) + T_{\lambda_{k+1}, \gamma_{k+1}}(x_{k+1}) = 0. \tag{38}$$

After rearranging the terms of (38), for every  $k \geq 1$  we obtain

$$x_{k+1} + T_{\lambda_{k+1}, \gamma_{k+1}}(x_{k+1}) = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) - \xi (T_{\lambda_k, \gamma_k}(x_k) - T_{\lambda_{k-1}, \gamma_{k-1}}(x_{k-1})). \tag{39}$$

In other words, after setting  $\alpha_k = 1 - \frac{\alpha}{k}$  and denoting the right hand side of (39) by  $y_k$  for every  $k \geq 1$ , we obtain the following iterative scheme

$$(\forall k \geq 1) \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) - \xi (T_{\lambda_k, \gamma_k}(x_k) - T_{\lambda_{k-1}, \gamma_{k-1}}(x_{k-1})), \\ x_{k+1} = (\text{Id} + T_{\lambda_{k+1}, \gamma_{k+1}})^{-1}(y_k). \end{cases} \tag{40}$$

Observe that the second step in (40) is always well-defined. Indeed, for  $\lambda, \gamma > 0$ ,  $T_{\lambda, \gamma}$  is  $\frac{\lambda}{2}$ -cocoercive, hence monotone (see Lemma 2.2(i)). This also implies that  $T_{\lambda, \gamma}$  is  $\frac{2}{\lambda}$ -Lipschitz continuous, and a monotone and continuous operator is maximally monotone, according to [14, Corollary 20.28]. Hence, by Minty’s Theorem (see [14, Theorem 21.1]), we know that  $\text{Id} + T_{\lambda, \gamma} : \mathcal{H} \rightarrow \mathcal{H}$  is surjective.

We are in conditions of stating the main theorem concerning our previous algorithm.

**Theorem 6.1** *Let  $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a maximally monotone operator and  $B : \mathcal{H} \rightarrow \mathcal{H}$  a  $\beta$ -cocoercive operator for some  $\beta \geq 0$  such that  $\text{zer}(A + B) \neq \emptyset$ . Choose  $x_0, x_1 \in \mathcal{H}$  any initial points. Let  $\alpha > 1$ ,  $\xi \geq 0$ , and  $(\lambda_k)_{k \geq 0}, (\gamma_k)_{k \geq 0}$  sequences of positive numbers that fulfill*

$$\lambda_k = \lambda k^2 \quad \forall k \geq 1, \quad \text{with } \lambda > \frac{4\xi + 2}{(\alpha - 1)^2},$$

$$0 < \inf_{k \geq 0} \gamma_k \leq \sup_{k \geq 0} \gamma_k < 2\beta \quad \text{and} \quad \frac{\gamma_k - \gamma_{k-1}}{\gamma_k} = \mathcal{O}\left(\frac{1}{k}\right) \quad \text{as } k \rightarrow +\infty.$$

*Now, consider the sequences  $(y_k)_{k \geq 1}$  and  $(x_k)_{k \geq 0}$  generated by algorithm (40). The following properties are satisfied:*

(i) We have the estimates

$$\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right) \quad \text{and} \quad \|A_{\gamma_k}(x_k - \gamma_k Bx_k) + Bx_k\| = o\left(\frac{1}{\gamma_k}\right) \quad \text{as } k \rightarrow +\infty.$$

(ii) The sequence  $(x_k)_{k \geq 0}$  converges weakly to an element of  $\text{zer}(A + B)$ .

(iii) The sequence  $(y_k)_{k \geq 1}$  converges weakly to an element of  $\text{zer}(A + B)$ . Precisely, we have  $\|x_k - y_k\| = \mathcal{O}\left(\frac{1}{k}\right)$  as  $k \rightarrow +\infty$ .

The proof can be done by transposing the techniques used in the continuous time case to the discrete time case. Algorithm (40) can be seen as a splitting version of the (PRINAM) algorithm studied by Attouch and László in [10].

**Remark 6.2** The second step in (40) can be quite complicated to compute. However, if  $B = 0$ , we can resort to the fact that  $(A_{\lambda_1})_{\lambda_2} = A_{\lambda_1 + \lambda_2}$  for  $\lambda_1, \lambda_2 > 0$ . We now have, for  $\lambda, \gamma > 0$ ,

$$T_{\lambda, \gamma} = \frac{1}{\lambda} [\text{Id} - J_{\gamma A}] = \frac{\gamma}{\lambda} A_{\gamma},$$

which gives

$$(\text{Id} + T_{\gamma, \lambda})^{-1} = J_{\frac{\gamma}{\lambda} A} = -\frac{\gamma}{\lambda} (A_{\lambda})_{\frac{\gamma}{\lambda}} + \text{Id} = \text{Id} - \frac{\gamma}{\lambda} A_{\lambda + \frac{\gamma}{\lambda}}.$$

It is now possible to write (40) in terms of the resolvents of  $A$ . We have, for every  $k \geq 1$ ,

$$\begin{aligned} T_{\lambda_k, \gamma_k}(x_k) - T_{\lambda_{k-1}, \gamma_{k-1}}(x_{k-1}) &= \frac{1}{\lambda_k} [x_k - J_{\gamma_k A}(x_k)] - \frac{1}{\lambda_{k-1}} [x_{k-1} - J_{\gamma_{k-1} A}(x_{k-1})] \\ &= \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_{k-1}}\right)x_k + \frac{1}{\lambda_{k-1}}(x_k - x_{k-1}) \\ &\quad - \left(\frac{1}{\lambda_k} J_{\gamma_k A}(x_k) - \frac{1}{\lambda_{k-1}} J_{\gamma_{k-1} A}(x_{k-1})\right), \\ y_k - \frac{\gamma_{k+1}}{\lambda_{k+1}} A_{\lambda_{k+1} + \frac{\gamma_{k+1}}{\lambda_{k+1}}}(y_k) &= y_k - \frac{\gamma_{k+1}}{\lambda_{k+1}} \frac{1}{\frac{\lambda_{k+1}^2 + \gamma_{k+1}}{\lambda_{k+1}}} \left[y_k - J_{\left(\lambda_{k+1} + \frac{\gamma_{k+1}}{\lambda_{k+1}}\right)A}(y_k)\right] \\ &= \frac{\lambda_{k+1}^2}{\lambda_{k+1}^2 + \gamma_{k+1}} y_k + \frac{\gamma_k}{\lambda_{k+1}^2 + \gamma_{k+1}} J_{\left(\lambda_{k+1} + \frac{\gamma_{k+1}}{\lambda_{k+1}}\right)A}(y_k). \end{aligned}$$

So now (40) becomes

$$(\forall k \geq 1) \begin{cases} y_k = \left(1 - \xi \left(\frac{1}{\lambda_k} - \frac{1}{\lambda_{k-1}}\right)\right)x_k + \left(\alpha_k - \frac{\xi}{\lambda_{k-1}}\right)(x_k - x_{k-1}) \\ \quad + \xi \left(\frac{1}{\lambda_k} J_{\gamma_k A}(x_k) - \frac{1}{\lambda_{k-1}} J_{\gamma_{k-1} A}(x_{k-1})\right), \\ x_{k+1} = \frac{\lambda_{k+1}^2}{\lambda_{k+1}^2 + \gamma_{k+1}} y_k + \frac{\gamma_k}{\lambda_{k+1}^2 + \gamma_{k+1}} J_{\left(\lambda_{k+1} + \frac{\gamma_{k+1}}{\lambda_{k+1}}\right)A}(y_k). \end{cases} \tag{41}$$

Now, if we assume  $0 < \inf_{k \geq 0} \gamma_k$  and  $\lambda > \frac{2\xi + 1}{(\alpha - 1)^2}$  and otherwise keep the hypotheses of Theorem 6.1, then for the sequences  $(x_k)_{k \geq 0}$  and  $(y_k)_{k \geq 1}$  generated by (41), the following statements hold:

(i) We have the estimates

$$\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right) \quad \text{and} \quad \|A_{\gamma_k}(x_k)\| = o\left(\frac{1}{\gamma_k}\right) \quad \text{as } k \rightarrow +\infty.$$

- (ii) The sequence  $(x_k)_{k \geq 0}$  converges weakly to an element of  $\text{zer } A$ .
- (iii) The sequence  $(y_k)_{k \geq 1}$  converges weakly to an element of  $\text{zer } A$  as well. Precisely, we have  $\|x_k - y_k\| = \mathcal{O}\left(\frac{1}{k}\right)$  as  $k \rightarrow +\infty$ .

Notice that the condition required for  $(\gamma_k)_{k \geq 0}$  is fulfilled in particular for  $\gamma_k = k^n$  for every  $k \geq 1$  and a natural number  $n \geq 1$ . Thus, by choosing large  $n$ , we obtain a fast convergence rate for  $A_{\gamma_k}(x_k)$  as  $k \rightarrow +\infty$ .

**Funding** Open access funding provided by University of Vienna.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## A Appendix

The following are three auxiliary lemmas that are used in the proof of Theorem 3.1. The proof for Lemma A.1 can be found in [12], while the proof of Lemma A.2 is straightforward. For the proof of Opial’s Lemma, we refer the reader to [1, Lemma 1.10].

**Lemma A.1** *Let  $t_0 > 0$ , and let  $u : [t_0, +\infty) \rightarrow \mathbb{R}$  be a continuously differentiable function which is bounded from below. Given  $\alpha > 1$ , a nonnegative function  $\theta : [t_0, +\infty) \rightarrow \mathbb{R}$  and a nonnegative function  $k \in L^1([t_0, +\infty), \mathbb{R})$ , let us assume that*

$$t\ddot{u}(t) + \alpha\dot{u}(t) + \theta(t) \leq k(t)$$

*for almost every  $t \geq t_0$ . Then, the positive part  $[\dot{u}]_+$  of  $\dot{u}$  belongs to  $L^1([t_0, +\infty), \mathbb{R})$  and  $\lim_{t \rightarrow +\infty} u(t)$  exists. Moreover, we have  $\int_{t_0}^{+\infty} \theta(t) dt < +\infty$ .*

**Lemma A.2** *Let  $A, B, C \in \mathbb{R}$  and  $\mathcal{H}$  a real Hilbert space. Then the inequality*

$$A\|X\|^2 + 2C\langle X, Y \rangle + B\|Y\|^2 \leq 0$$

*holds for every  $X, Y \in \mathcal{H}$  if and only if  $A, B \leq 0$  and  $C^2 - AB \leq 0$ .*

**Lemma A.3** (Opial’s Lemma) *Let  $S \subseteq \mathcal{H}$  be a nonempty set and  $x : [t_0, +\infty) \rightarrow \mathcal{H}$  a given map, where  $t_0 > 0$ . Assume that*

- (i) *for every  $x^* \in S$ ,  $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$  exists;*
- (ii) *every weak sequential cluster point of the map  $x$  belongs to  $S$ .*

*Then, there exists  $x_\infty \in S$  such that  $x(t)$  converges weakly to  $x_\infty$  as  $t \rightarrow +\infty$ .*

## References

1. Abbas, B., Attouch, H.: Dynamical systems and forward–backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator. *Optimization* **64**(10), 2223–2252 (2015)

2. Álvarez, F.: On the minimizing property of a second-order dissipative system in Hilbert spaces. *SIAM J. Control Optim.* **38**(4), 1102–1119 (2020)
3. Álvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. *Set-Valued Anal.* **9**(1–2), 3–11 (2001)
4. Álvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *J. Math. Pures Appl.* **81**(8), 747–779 (2002)
5. Apidopoulos, V., Aujol, J.-F., Dossal, C.: Convergence rate of inertial forward–backward algorithm beyond Nesterov’s rule. *Math. Program.* **180**, 137–156 (2020)
6. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.* **168**, 123–175 (2018)
7. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case  $\alpha \leq 3$ . In: *ESAIM: COCV*, vol. 25, Article number 2 (2019)
8. Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system. *Commun. Contemp. Math.* **2**(1), 1–34 (2000)
9. Attouch, H., László, S.C.: Continuous Newton-like inertial dynamics for monotone inclusions. *Set-Valued Var. Anal.* **29**, 555–581 (2021)
10. Attouch, H., László, S.C.: Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. *SIAM J. Optim.* **30**(4), 3252–3283 (2020)
11. Attouch, H., Maingé, P.E.: Asymptotic behavior of second order dissipative evolution equations combining potential with non-potential effects. *ESAIM Control Optim. Calculus Var.* **17**(3), 836–857 (2011)
12. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators. *Math. Program.* **174**(1–2), 391–432 (2019)
13. Attouch, H., Peypouquet, J., Redont, P.: Fast convex minimization via inertial dynamics with Hessian driven damping. *J. Differ. Equ.* **210**(10), 5734–5783 (2016)
14. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory*, CMS Books in Mathematics, 2nd edn. Springer, Berlin (2017)
15. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**(1), 183–202
16. Boj, R.I., Csetnek, E.R.: Second order forward–backward dynamical systems for monotone inclusion problems. *SIAM J. Control Optim.* **54**, 1423–1443 (2016)
17. Haraux, A.: *Systèmes Dynamiques Dissipatifs et Applications*. Masson (1991)
18. May, R.: Asymptotic for a second order evolution equation with convex potential and vanishing damping term. *Turk. J. Math.* **41**(3), 681–685 (2017)
19. Nesterov, Y.: A method of solving a convex programming problem with convergence rate  $\mathcal{O}(1/k^2)$ . *Dokl. Akad. Nauk SSSR* **269**(3), 543–547 (1983)
20. Su, W.J., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *Neural Inf. Process. Syst.* **27**, 2510–2518 (2014)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.