Set-Valued and Variational Analysis (2021) 29:555–581
https://doi.org/10.1007/s11228-020-00564-y

Continuous Newton-like Inertial Dynamics for Monotone Inclusions

Hedy Attouch[1] · **Szilárd Csaba László**[2]

Received: 14 May 2020 / Accepted: 1 October 2020 / Published online: 15 October 2020

Abstract

In a Hilbert framework $\mathcal{H}$, we study the convergence properties of a Newton-like inertial dynamical system governed by a general maximally monotone operator $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$. When A is equal to the subdifferential of a convex lower semicontinuous proper function, the dynamic corresponds to the introduction of the Hessian-driven damping in the continuous version of the accelerated gradient method of Nesterov. As a result, the oscillations are significantly attenuated. According to the technique introduced by Attouch-Peypouquet (Math. Prog. 2019), the maximally monotone operator is replaced by its Yosida approximation with an appropriate adjustment of the regularization parameter. The introduction into the dynamic of the Newton-like correction term (corresponding to the Hessian driven term in the case of convex minimization) provides a well-posed evolution system for which we will obtain the weak convergence of the generated trajectories towards the zeroes of A. We also obtain the fast convergence of the velocities towards zero. The results tolerate the presence of errors, perturbations. Then, we specialize our results to the case where the operator A is the subdifferential of a convex lower semicontinuous function, and obtain fast optimization results.

Keywords Damped inertial dynamics · Hessian damping · Maximally monotone operators · Newton method · Vanishing viscosity · Yosida regularization

Mathematics Subject Classification (2010) 37N40 · 46N10 · 49M30 · 65K05 · 65K10 · 90B50 · 90C25

✉ Hedy Attouch
hedy.attouch@umontpellier.fr

Szilárd Csaba László
szilard.laszlo@math.utcluj.ro

1 IMAG, University of Montpellier, CNRS, Montpellier, France

2 Department of Mathematics, Technical University of Cluj-Napoca, Memorandumului 28, Cluj-Napoca, Romania

1 Introduction

Let $\mathcal{H}$ be a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. Given a general maximally monotone operator $A : \mathcal{H} \to 2^{\mathcal{H}}$, we will study the asymptotic behavior, as $t \to +\infty$, of the second-order in time evolution equation

$$\text{(DIN-AVD)}_{\alpha,\beta,\lambda,e} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right) + A_{\lambda(t)}(x(t)) = e(t).$$

The operators $J_{\lambda A} : \mathcal{H} \to \mathcal{H}$ and $A_\lambda : \mathcal{H} \to \mathcal{H}$ which are defined by

$$J_{\lambda A} = (I + \lambda A)^{-1} \text{ and } A_\lambda = \frac{1}{\lambda}(I - J_{\lambda A}),$$

are respectively the resolvent of A and the Yosida regularization of A of index $\lambda > 0$. The coefficients α, β are positive damping parameters. The tuning of the time dependent parameter $\lambda(t)$ which enters the Yosida regularization of A will play a crucial role in the asymptotic analysis. The second member e takes account of perturbations, errors. Without ambiguity, we refer briefly to the dynamic as (DIN-AVD). The terminology reflects the link of this dynamic with the Dynamic Inertial Newton method and to the Asymptotic Vanishing Damping, as explained in the next paragraph. According to the Lipschitz continuity property of the Yosida approximation, (DIN-AVD) is relevant to the Cauchy-Lipschitz theorem, which provides existence and uniqueness of the corresponding Cauchy problem. On the basis of an appropriate adjustment of the parameters, we will obtain the weak convergence of the generated trajectories towards the zeroes of A, *i.e.* solutions of the monotone inclusion

$$0 \in Ax. \tag{1}$$

This dynamic is the support of a recent study by the authors concerning rapidly converging algorithms to solve (1) and which make use only of the resolvents of A, see [12]. It is a difficult problem of fundamental importance in optimization, equilibrium theory, economics and game theory, partial differential equations, statistics, among other subjects. An in-depth study of (DIN-AVD) is a key to going further in the analysis of the associated algorithms. Our study is based on several recent advances in the study of damped inertial dynamics for solving optimization problems and monotone inclusions. We describe them briefly in the following paragraphs.

1.1 Inertial Dynamics and Algorithms with Vanishing Damping Coefficient

The use of inertial dynamics to accelerate the gradient method in optimization was first considered by B. Polyak in [34]. He considered the inertial system with a fixed viscous damping coefficient $\gamma > 0$

$$\text{(HBF)} \qquad \ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) = 0,$$

which, because of its natural mechanical interpretation, is called the Heavy Ball with Friction method. This system was further developed by Attouch-Goudou-Redont [11] as a tool to explore the local minima of f. For a strongly convex function f, and γ judiciously chosen, (HBF) provides convergence at exponential rate of $f(x(t))$ to $\min_{\mathcal{H}} f$. For a general convex function f, the asymptotic convergence rate of (HBF) is $\mathcal{O}(\frac{1}{t})$ (in the worst case). This is however not better than the steepest descent. A decisive step to obtain a faster asymptotic convergence was taken by Su-Boyd-Candès [36] who considered the case of an Asymptotic Vanishing Damping coefficient $\gamma(t) = \frac{\alpha}{t}$, that is

$$\text{(AVD)}_{\alpha} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0.$$

For a general convex differentiable function f, and $\alpha = 3$, it provides a continuous version of the accelerated gradient method of Nesterov [33].[1] For $\alpha \geq 3$, each trajectory $x(\cdot)$ of (AVD)_{α} satisfies the asymptotic convergence rate of the values

$$f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(1/t^2\right) \text{ as } t \to +\infty.$$

As a specific feature, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes (tends to zero) as time t goes to infinity, hence the terminology. The case $\alpha = 3$, which corresponds to Nesterov's historical algorithm, is critical. In the case $\alpha = 3$, the question of the convergence of the trajectories remains an open problem (except in one dimension where convergence holds [10]). For $\alpha > 3$, it has been shown by Attouch-Chbani-Peypouquet-Redont [8] that each trajectory converges weakly to a minimizer. For $\alpha > 3$, it is shown in [17] and [32] that the asymptotic convergence rate of the values is actually $o(1/t^2)$. The subcritical case $\alpha \leq 3$ has been examined by Apidopoulos-Aujol-Dossal[4] and Attouch-Chbani-Riahi [10], with the convergence rate of the objective values $\mathcal{O}\left(t^{-\frac{2\alpha}{3}}\right)$. These rates are optimal, that is, they can be reached, or approached arbitrarily close. The corresponding inertial algorithms are in line with the Nesterov accelerated gradient method

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} = y_k - s\nabla f(y_k). \end{cases}$$

They enjoy similar properties to the continuous case. They were first obtained by Chambolle-Dossal [27], see [6, 8], and [30] for further results, and the extension to proximal-gradient algorithms for structured optimization.

1.2 Hessian Damping

The following inertial system combines asymptotic vanishing damping with Hessian-driven damping

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

It was considered by Attouch-Peypouquet-Redont in [18]. At first glance, the presence of the Hessian may seem to entail numerical difficulties. However, this is not the case as the Hessian intervenes in the above ODE in the form $\nabla^2 f(x(t))\dot{x}(t)$, which is nothing but the derivative with respect to time of $\nabla f(x(t))$. So, the temporal discretization of this dynamic provides first-order algorithms of the form

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) - \beta_k\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) \\ x_{k+1} = y_k - s\nabla f(y_k). \end{cases}$$

As a specific feature, and by comparison with the accelerated gradient method of Nesterov, these algorithms contain a correction term which is equal to the difference of the

[1] When f is not convex, the convergence of the trajectories generated by (AVD)_{α} is a largely open question. Recent progress has been made in [24] where the convergence of the trajectories of a system, which can be considered as a perturbation of (AVD)_{α}, has been obtained in a non-convex setting.

gradients at two consecutive steps. While preserving the convergence properties of the Nesterov accelerated method, they provide fast convergence to zero of the gradients, and reduce the oscillatory aspects. Several recent studies have been devoted to this subject, see Attouch-Chbani-Fadili-Riahi [7], Boţ-Csetnek-László [25], Kim [29], Lin-Jordan [31], Shi-Du-Jordan-Su [35].

1.3 Inertial Dynamics and Cocoercive Operators

Let's come to the transposition of these techniques to the case of maximally monotone operators. Álvarez-Attouch [2] and Attouch-Maingé [13] studied the equation

$$\ddot{x}(t) + \gamma \dot{x}(t) + A(x(t)) = 0, \tag{2}$$

when A is a cocoercive [2] (and hence maximally monotone) operator, (see also [23]). Cocoercivity plays an important role in the study of (2), not only to ensure the existence of solutions, but also to analyze their long-term behavior. Assuming that the cocoercivity parameter λ and the damping coefficient γ satisfy the inequality $\lambda\gamma^2 > 1$, they showed that each trajectory of (2) converges weakly to a zero of A. Since for $\lambda > 0$, the operator A_λ is λ-cocoercive and $A_\lambda^{-1}(0) = A^{-1}(0)$, we immediately deduce that, under the condition $\lambda\gamma^2 > 1$, given a general maximally monotone operator A, each trajectory of

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

converges weakly to a zero of A. In the quest for faster convergence, the system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0,$$

involves a time-dependent regularizing parameter $\lambda(\cdot)$ satisfying the inequality

$$\lambda(t) \times \frac{\alpha^2}{t^2} > 1, \quad t > t_0 > 0,$$

see Attouch-Peypouquet [16]. Time discretization of this dynamic gives the Relaxed Inertial Proximal Algorithm

$$\text{(RIPA)} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = (1-\rho_k)y_k + \rho_k J_{\mu_k A}(y_k), \end{cases}$$

whose convergence properties have been analyzed by Attouch-Cabot [5], Attouch-Chbani-Riahi [9], Attouch-Peypouquet [16].

1.4 Link with Newton-like Methods for Solving Monotone Inclusions

Let us specify the link between our study and Newton's method for solving (1). To overcome the ill-posed character of the continuous Newton method for a general maximally monotone operator A, the following first order evolution system was studied by Attouch-Svaiter [20],

$$\begin{cases} v(t) \in A(x(t)) \\ \gamma(t)\dot{x}(t) + \beta\dot{v}(t) + v(t) = 0. \end{cases}$$

This system can be considered as a continuous version of the Levenberg-Marquardt method, which acts as a regularization of the Newton method. Remarkably, under a fairly general

[2] $A : \mathcal{H} \to \mathcal{H}$ is λ-cocoercive ($\lambda > 0$) if for all $x, y \in \mathcal{H}$ $\langle Ay - Ax, y - x\rangle \geq \lambda\|Ay - Ax\|^2$.

assumption on the regularization parameter $\gamma(t)$, this system is well posed and generates trajectories that converge weakly to equilibria. Parallel results have been obtained for the associated proximal algorithms obtained by implicit temporal discretization, see [1, 15, 19]. Formally, this system writes as

$$\gamma(t)\dot{x}(t) + \beta\frac{d}{dt}(A(x(t))) + A(x(t)) = 0.$$

Thus, (DIN-AVD)$_{\alpha,\beta}$ can be considered as an inertial and regularized version of this dynamical system.

1.5 Organization of the Paper

In Section 2, we show the existence and the uniqueness of a strong global solution to the Cauchy problem associated with (DIN-AVD). In Section 3, based on an appropriate tuning of the parameters, we study the convergence properties as $t \to +\infty$ of the trajectories generated by (DIN-AVD). Our study takes into account the presence of perturbations, errors. Section 4 is devoted to numerical experiments. In Section 5, we specialize our study in the case where the operator A is the subdifferential of a convex lower semicontinuous function. In this case, we get fast minimization properties. Finally, we present some lines of research for the future.

2 Existence and Uniqueness Results for (DIN-AVD)

We rely on the first-order equivalent formulation of (DIN-AVD) which is valid when $\beta > 0$. It was first considered by Alvarez-Attouch-Bolte-Redont [3] and Attouch-Peypouquet-Redont [18] in the framework of convex minimization, that is $A = \partial f$ with f convex. Specifically, in our context, we have the following equivalence, which follows from a simple differential and algebraic calculation.

Proposition 1 *The following are equivalent:* $(i) \Longleftrightarrow (ii)$

$$(i) \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right) + A_{\lambda(t)}(x(t)) = e(t).$$

$$(ii) \quad \begin{cases} \dot{x}(t) + \beta A_{\lambda(t)}(x(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x(t) + \frac{1}{\beta}y(t) = 0; \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) + \frac{1}{\beta}y(t) = -\beta e(t). \end{cases}$$

Note that (ii) is different from the Hamiltonian formulation of (i). On the one hand, the formulation (ii) is written as an evolution system in the product space $\mathcal{H} \times \mathcal{H}$, which is governed by the sum of a time dependent maximally monotone operator and a time dependent continuous linear operator. From this, we will deduce in the following theorem the existence and uniqueness of a strong global solution for the associated Cauchy problem. This first order equivalent formulation offers many applications. It was used by Attouch-Maingé-Redont in [14] for the modeling of damped shocks in mechanics, and by Castera-Bolte-Févotte-Pauwels in deep learning [26].

On the other hand, when one works with (ii), one loses the mechanical interpretation of the dynamic, and the intuition of the energy notions which are attached to it. So, in the following sections, to develop a Lyapunov analysis for (DIN-AVD), we will work with the initial formulation (i). The analysis of (DIN-AVD) is essentially based on the use of time-dependent Yosida parameters $\lambda(t)$. We have gathered in the following lemma some technical results concerning the properties of the mapping $(\lambda, x) \mapsto A_\lambda(x)$, which will be useful later in the proofs.

Lemma 1 *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $\gamma, \nu > 0$, and $x, y \in \mathcal{H}$. Then, the following inequalities are satisfied*

a) $\|\gamma A_\gamma(x) - \nu A_\nu(y)\| \le 2\|x - y\| + |\gamma - \nu|\|A_\gamma(x)\|$.

b) $\|A_\gamma(x) - A_\nu(y)\| \le \frac{2}{\gamma}\|x - y\| + \frac{|\gamma-\nu|}{\gamma}(\|A_\gamma(x)\| + \|A_\nu(y)\|)$.

c) *Consider $x : [t_0, +\infty[\to \mathcal{H}$ a differentiable function. Assume further $\lambda : [t_0, +\infty[\to]0, +\infty[$ is a derivable function. Then, for every $t \in [t_0, +\infty[$ and every $z \in A^{-1}(0)$,*

$$(c_1) \quad \left\| \frac{d}{dt}\lambda(t)A_{\lambda(t)}(x(t)) \right\| \le 2\|\dot{x}(t)\| + \lambda'(t)\|A_{\lambda(t)}(x(t))\|.$$

$$(c_2) \quad \left\| \frac{d}{dt}A_{\lambda(t)}(x(t)) \right\| \le \frac{2}{\lambda(t)}\|\dot{x}(t)\| + 2\frac{|\lambda'(t)|}{\lambda(t)}\|A_{\lambda(t)}(x(t))\|.$$

$$(c_3) \quad \left\| \frac{d}{dt}\lambda(t)A_{\lambda(t)}(x(t)) \right\| \le 2\|\dot{x}(t)\| + \frac{|\lambda'(t)|}{\lambda(t)}\|x(t) - z\|.$$

Proof a) It follows from the proof of [16, Lemma A.4]. However, for the convenience of the reader, we are providing full proof. Note that according to [21, Proposition 23.28 (iii)]

$$\|J_{\gamma A}(x) - J_{\nu A}(x)\| \le |\gamma - \nu|\|A_\gamma(x)\|.$$

Hence,

$$\|\gamma A_\gamma(x) - \nu A_\nu(x)\| = \|(x - y) - (J_{\gamma A}(x) - J_{\nu A}(x)\| \le \|x - y\| + |\gamma - \nu|\|A_\gamma(x)\|.$$

Moreover, by the $\frac{1}{\nu}$−Lipschitz continuity of the Yosida approximation A_ν, we have

$$\|\nu A_\nu(x) - \nu A_\nu(y)\| \le \|x - y\|.$$

Finally,

$$\begin{aligned}\|\gamma A_\gamma(x) - \nu A_\nu(y)\| &= \|(\gamma A_\gamma(x) - \nu A_\nu(x)) + (\nu A_\nu(x) - \nu A_\nu(y))\| \\ &\le \|x - y\| + |\gamma - \nu|\|A_\gamma(x)\| + \|x - y\|.\end{aligned}$$

b) By using a) we have

$$\begin{aligned}\|A_\gamma(x) - A_\nu(y)\| &= \frac{1}{\gamma}\|(\gamma A_\gamma(x) - \nu A_\nu(y)) + (\nu - \gamma)A_\nu(y)\| \\ &\le \frac{1}{\gamma}\|\gamma A_\gamma(x) - \nu A_\nu(y)\| + \frac{|\nu - \gamma|}{\gamma}\|A_\nu(y)\| \\ &\le \frac{2}{\gamma}\|x - y\| + \frac{|\gamma - \nu|}{\gamma}(\|A_\gamma(x)\| + \|A_\nu(y)\|).\end{aligned}$$

(c_1) Let $h > 0$. From a) we have

$$\begin{aligned}&\|\lambda(t + h)A_{\lambda(t+h)}(x(t + h)) - \lambda(t)A_{\lambda(t)}(x(t))\| \le 2\|x(t + h) - x(t)\| \\ &+ |\lambda(t + h) - \lambda(t)|\|A_{\lambda(t+h)}(x(t + h))\|.\end{aligned}$$

Dividing by h, and letting $h \to 0$, we get the claim.

(c_2) Let $h > 0$. From $b)$ we have

$$\|A_{\lambda(t+h)}(x(t+h)) - A_{\lambda(t)}(x(t))\| \leq$$
$$\frac{2}{\lambda(t+h)}\|x(t+h) - x(t)\| + \frac{|\lambda(t+h)-\lambda(t)|}{\lambda(t+h)}(\|A_{\lambda(t+h)}(x(t+h))\| + \|A_{\lambda(t)}(x(t))\|).$$

Dividing by h and letting $h \to 0$, we get the claim.

(c_3) Let $z \in A^{-1}(0)$. According to $A_{\lambda(t)}(z) = 0$, and the $\frac{1}{\lambda(t)}$-Lipschitz continuity of $A_{\lambda(t)}$ we have

$$\|A_{\lambda(t)}(x(t))\| \leq \frac{1}{\lambda(t)}\|x(t) - z\|.$$

Combining this inequality with (c_1) gives the claim. □

Theorem 1 *Suppose that* $\beta \geq 0$. *Suppose that* $\lambda : [t_0, +\infty[\to]0, +\infty[$ *is a measurable function, and that there exists* $\underline{\lambda} > 0$ *such that* $\lambda(t) \geq \underline{\lambda}$ *for all* $t \geq t_0$. *Suppose that* $e \in \mathrm{L}^1(t_0, T; \mathcal{H})$ *for all* $T > t_0$.
Then, for any $(x_0, x_1) \in \mathcal{H}\times\mathcal{H}$, *there exists a unique strong global solution* $x : [t_0, +\infty[\to \mathcal{H}$ *of the continuous dynamic* (DIN-AVD) *which satisfies the Cauchy data* $x(t_0) = x_0$, $\dot{x}(t_0) = x_1$.

Proof a) **Case** $\beta > 0$. According to Proposition 1, it is equivalent to solve the first-order system (ii) with the Cauchy data $x(t_0) = x_0$ and $y(t_0) = y_0 := -\beta\Big(x_1 + \beta A_{\lambda(t_0)}(x_0) - \left(\frac{1}{\beta} - \frac{\alpha}{t_0}\right)x_0\Big)$. This system writes

$$\dot{Z}(t) + F(t, Z(t)) = 0, \quad Z(t_0) = (x_0, y_0)$$

where $Z(t) = (x(t), y(t)) \in \mathcal{H} \times \mathcal{H}$ and

$$F(t, (x, y)) = \left(\beta A_{\lambda(t)}(x) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x + \frac{1}{\beta}y, -\left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x + \frac{1}{\beta}y + \beta e(t)\right).$$

According to the Lipschitz continuity property of the Yosida approximation of a maximally monotone operator, the existence and uniqueness of a strong global solution $(x, y) : [t_0, +\infty[\to \mathcal{H} \times \mathcal{H}$ is relevant of the Cauchy-Lipschitz theorem. Precisely, we can apply the non-autonomous version of this theorem given in [28, Proposition 6.2.1]. Thus, we obtain a strong solution, that is, $t \mapsto \dot{x}(t)$ is locally absolutely continuous. If, moreover, we suppose that the functions $\lambda(\cdot)$ and $e(\cdot)$ are continuous, then the solution is a classical solution of class $\mathcal{C}^2$.

b) **Case** $\beta = 0$. We then get the system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = e(t).$$

In this case, we use the equivalent Hamiltonian formulation

$$\begin{cases} \dot{x}(t) - y(t) = 0; \\ \dot{y}(t) + \frac{\alpha}{t}y(t) + A_{\lambda(t)}(x(t)) - e(t) = 0, \end{cases}$$

which, thanks to the Lipschitz continuity of A_λ, is relevant of the classical Cauchy-Lipschitz theorem. □

3 Asymptotic Convergence Properties of (DIN-AVD)

Given $\alpha > 1$, $\beta \geq 0$ and $e \in L^1_{loc}(t_0, +\infty; \mathcal{H})$, we consider the evolution system

$$\text{(DIN-AVD)} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right) + A_{\lambda(t)}(x(t)) = e(t), \quad t > t_0 > 0.$$

The existence of strong global solutions to this system is guaranteed by Theorem 1. The convergence properties as $t \to +\infty$ of the trajectories generated by this system are summarized in the following theorem.

Theorem 2 *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $S = A^{-1}(0) \neq \emptyset$. Consider the evolution equation* (DIN-AVD) *where the parameters satisfy the following conditions*

$$\alpha > 1,\ \beta \geq 0,\ \textit{and}\ \lambda(t) = \lambda t^2\ \textit{with}\ \lambda > \frac{1}{(\alpha-1)^2}.$$

Assume further that

$$\int_{t_0}^{+\infty} t^3\|e(t)\|^2 dt < +\infty \quad \textit{and} \quad \int_{t_0}^{+\infty} t\|e(t)\| dt < +\infty.$$

Then, for any trajectory $x : [t_0, +\infty[\to \mathcal{H}$ of (DIN-AVD) *the following properties are satisfied:*

(i) *(convergence) $x(t)$ is bounded, and $x(t)$ converges weakly, as $t \to +\infty$, to an element of S.*

(ii) *(integral estimates)* $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty$, $\int_{t_0}^{+\infty} t^3\|\ddot{x}(t)\|^2 dt < +\infty$ *and*
$$\int_{t_0}^{+\infty} t^3\|A_{\lambda(t)}(x(t))\|^2 dt < +\infty.$$

(iii) *(pointwise estimates)* $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$, $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ *as $t \to +\infty$ and, for all $0 < \eta < 1$, one has*
$$\|A_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t^2}\right), \quad \left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| = o\left(\frac{1}{t^{2+\eta}}\right) \ \textit{as}\ t \to +\infty.$$

Proof **Lyapunov analysis.** Take $z \in S$. For $0 < b < \alpha - 1$ consider the energy functional

$$\mathcal{E}_b(t) := \tfrac{1}{2}\|b(x(t)-z) + t(\dot{x}(t) + \beta A_{\lambda(t)}(x(t)))\|^2 + \tfrac{b(\alpha-1-b)}{2}\|x(t)-z\|^2. \tag{3}$$

Using the classical derivation chain rule and (DIN-AVD), we get

$$\begin{aligned}&\dot{\mathcal{E}}_b(t) = \\ &\left\langle (b+1-\alpha)\dot{x}(t) + (\beta - t)A_{\lambda(t)}(x(t)) + te(t), b(x(t)-z) + t(\dot{x}(t) + \beta A_{\lambda(t)}(x(t)))\right\rangle \\ &+ b(\alpha-1-b)\langle \dot{x}(t), x(t)-z\rangle.\end{aligned} \tag{4}$$

After reduction, we obtain

$$\begin{aligned}\dot{\mathcal{E}}_b(t) \;=\; & b(\beta - t)\left\langle A_{\lambda(t)}(x(t)), x(t)-z\right\rangle + (-t^2 + \beta(b+2-\alpha)t)\left\langle A_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle \\ &+ (b+1-\alpha)t\|\dot{x}(t)\|^2 + \beta(\beta-t)t\|A_{\lambda(t)}(x(t))\|^2 \\ &+ bt\langle e(t), x(t)-z\rangle + t^2\langle e(t), \dot{x}(t)\rangle + \beta t^2\langle e(t), A_{\lambda(t)}(x(t))\rangle.\end{aligned} \tag{5}$$

We have for all $p_1 > 0$

$$bt\langle e(t), x(t) - z\rangle \le bt\|e(t)\|\|x(t) - z\|, \tag{6}$$

$$t^2\langle e(t), \dot{x}(t)\rangle \le p_1 t^3\|e(t)\|^2 + \frac{t}{4p_1}\|\dot{x}(t)\|^2, \tag{7}$$

$$\beta t^2\langle e(t), A_{\lambda(t)}(x(t))\rangle \le \frac{\beta}{2}\left(t^3\|e(t)\|^2 + t\|A_{\lambda(t)}(x(t))\|^2\right). \tag{8}$$

Combining (5), (6), (7) and (8), we obtain

$$\begin{aligned}\dot{\mathcal{E}}_b(t) \le\ & b(\beta - t)\left\langle A_{\lambda(t)}(x(t)), x(t) - z\right\rangle + \left(b + 1 - \alpha + \frac{1}{4p_1}\right)t\|\dot{x}(t)\|^2\\ & + (-t^2 + \beta(b + 2 - \alpha)t)\left\langle A_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle + \left(\beta(\beta - t)t + \frac{\beta}{2}t\right)\|A_{\lambda(t)}(x(t))\|^2\\ & + bt\|e(t)\|\|x(t) - z\| + \left(p_1 + \frac{\beta}{2}\right)t^3\|e(t)\|^2. \end{aligned}\tag{9}$$

Now, using the fact that $A_{\lambda(t)}$ is $\lambda(t)$ cocoercive, and that $z \in S$, we get, for all $t \ge t_1 = \max(t_0, \beta)$

$$b(\beta - t)\left\langle A_{\lambda(t)}(x(t)), x(t) - z\right\rangle \le b(\beta - t)\lambda(t)\|A_{\lambda(t)}(x(t))\|^2. \tag{10}$$

Let us choose $b = \dfrac{\alpha - 1}{2} > 0$. According to the assumption $\lambda > \frac{1}{(\alpha-1)^2}$, we can find ϵ such that

$$0 < \epsilon < \alpha - 1 - \frac{1}{\sqrt{\lambda}} < 2(\alpha - 1 - b), \tag{11}$$

where the last inequality comes from the choice of b. Further, take $p_1 = \dfrac{1}{\epsilon}$. Then, (9) and (10) lead to

$$\begin{aligned}&\dot{\mathcal{E}}_b(t) + \frac{\epsilon}{4}t\|\dot{x}(t)\|^2 + \frac{\epsilon}{2}t\lambda(t))\|A_{\lambda(t)}(x(t))\|^2 \le \left(b + 1 - \alpha + \frac{\epsilon}{2}\right)t\|\dot{x}(t)\|^2\\ &+(-t^2 + \beta(b + 2 - \alpha)t)\left\langle A_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle\\ &+\left(\left(b(\beta - t) + \frac{\epsilon}{2}t\right)\lambda(t) + \beta(\beta - t)t + \frac{\beta}{2}t\right)\|A_{\lambda(t)}(x(t))\|^2\\ &+bt\|e(t)\|\|x(t) - z\| + \left(\frac{1}{\epsilon} + \frac{\beta}{2}\right)t^3\|e(t)\|^2, \end{aligned}\tag{12}$$

for all $t \ge t_1$. By (11) we have $b + 1 - \alpha + \frac{\epsilon}{2} < 0$. Moreover, still by (11) we have $-b + \frac{\epsilon}{2} = -\frac{\alpha-1}{2} + \frac{\epsilon}{2} < 0$. Since $\lambda(t) = \lambda t^2$ with $\lambda > 0$, we deduce that there exists $t_2 \ge t_1$ such that, for all $t \ge t_2$

$$\left(b(\beta - t) + \frac{\epsilon}{2}t\right)\lambda(t) + \beta(\beta - t)t + \frac{\beta}{2}t < 0.$$

According to Lemma 2 (see Appendix), we deduce that the sum

$$\begin{aligned}\mathcal{S}(t) =\ & \left(b + 1 - \alpha + \frac{\epsilon}{2}\right)t\|\dot{x}(t)\|^2 + (-t^2 + \beta(b + 2 - \alpha)t)\left\langle A_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle\\ & + \left(\left(b(\beta - t) + \frac{\epsilon}{2}t\right)\lambda(t) + \beta(\beta - t)t + \frac{\beta}{2}t\right)\|A_{\lambda(t)}(x(t))\|^2\end{aligned}$$

in the right hand side of (12) is nonpositive whenever

$$R(t) := (-t^2+\beta(b+2-\alpha)t)^2 \\ -4\left(b+1-\alpha+\frac{\epsilon}{2}\right)t\left(\left(b(\beta-t)+\frac{\epsilon}{2}t\right)\lambda(t)+\beta(\beta-t)t+\frac{\beta}{2}t\right)\leq 0.$$

Taking into account the fact that $\lambda(t)=\lambda t^2$ we obtain

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

Since $\epsilon<\alpha-1-\frac{1}{\sqrt{\lambda}}$, we get that $\lambda>\frac{1}{(\alpha-1-\epsilon)^2}$. From $b=\frac{\alpha-1}{2}$ we deduce that

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

Consequently, there exists $t_3\geq t_2$ such that $R(t)<0$, for all $t\geq t_3$. Hence, (12) leads to, for all $t\geq t_3$

$$\dot{\mathcal{E}}_b(t)+\frac{\epsilon}{4}t\|\dot{x}(t)\|^2+\frac{\epsilon}{2}t\lambda(t))\|A_{\lambda(t)}(x(t))\|^2\leq bt\|e(t)\|\|x(t)-z\|+\left(\frac{1}{\epsilon}+\frac{\beta}{2}\right)t^3\|e(t)\|^2. \tag{13}$$

Estimates. By integrating (13) on an interval $[t_3,t]$, and by denoting

$$C_0=\left(\frac{1}{\epsilon}+\frac{\beta}{2}\right)\int_{t_3}^{+\infty}t^3\|e(t)\|^2dt+\mathcal{E}_b(t_3)<+\infty$$

we obtain that for all $t\geq t_3$

$$\mathcal{E}_b(t)+\frac{\epsilon}{4}\int_{t_3}^t s\|\dot{x}(s)\|^2ds+\frac{\epsilon}{2}\int_{t_3}^t s\lambda(s)\|A_{\lambda(s)}(x(s))\|^2ds \\ \leq C_0+b\int_{t_3}^t s\|e(s)\|\|x(s)-z\|ds. \tag{14}$$

Taking into account the form of the energy functional $\mathcal{E}_b(t)$ and the fact that $b=\frac{\alpha-1}{2}$, the latter relation leads to

$$\frac{(\alpha-1)^2}{8}\|x(t)-z\|^2\leq C_0+\frac{\alpha-1}{2}\int_{t_3}^t s\|e(s)\|\|x(s)-z\|ds.$$

More precisely, we have

$$\frac{1}{2}\|x(t)-z\|^2\leq C_1+\frac{2}{\alpha-1}\int_{t_3}^t s\|e(s)\|\|x(s)-z\|ds, \tag{15}$$

where $C_1=\frac{4C_0}{(\alpha-1)^2}$. Now, applying the Gronwall lemma (see [22, Lemma A.5]) to (15) we obtain

$$\|x(t)-z\|\leq\sqrt{2C_1}+\frac{2}{\alpha-1}\int_{t_3}^t s\|e(s)\|ds.$$

Therefore, $\|x(t)-z\|$ and consequently $x(t)$ are bounded.
Further, the boundedness of $\|x(t)-z\|$ and the assumption on e leads to

$$\int_{t_3}^{+\infty}s\|e(s)\|\|x(s)-z\|ds<+\infty.$$

Therefore, (14) becomes

$$\mathcal{E}_b(t) + \frac{\epsilon}{4}\int_{t_3}^{t} s\|\dot{x}(s)\|^2 ds + \frac{\epsilon}{2}\int_{t_3}^{t} s\lambda(s)\|A_{\lambda(s)}(x(s))\|^2 ds \le C, \tag{16}$$

where $C = C_0 + b\int_{t_3}^{+\infty} s\|e(s)\|\|x(s) - z\|ds < +\infty$.
From this we immediately deduce that

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

$$\int_{t_0}^{+\infty} t^3\|A_{\lambda(t)}(x(t))\|^2 dt < +\infty, \tag{18}$$

$$\sup_{t\ge t_0}\|b(x(t) - z) + t(\dot{x}(t) + \beta A_{\lambda(t)}(x(t)))\|^2 < +\infty. \tag{19}$$

Moreover, the $\frac{1}{\lambda(t)}$ Lipschitz continuity of $A_{\lambda(t)}$, and $z \in S$ leads to

$$\|A_{\lambda(t)}(x(t))\| = \|A_{\lambda(t)}(x(t)) - A_{\lambda(t)}(z)\| \le \frac{1}{\lambda(t)}\|x(t) - z\|.$$

Taking into account that $\lambda(t) = \lambda t^2$ and $\|x(t) - z\|$ is bounded, we deduce that

$$\|A_{\lambda(t)}(x(t))\| = \mathcal{O}\left(\frac{1}{t^2}\right), \text{ as } t \to +\infty. \tag{20}$$

Further, from the boundedness of the trajectory $x(\cdot)$ and (19) we deduce that

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right), \text{ as } t \to +\infty.$$

In particular, we have $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$. According to Lemma 1 we have

$$\left\|\frac{d}{dt}\lambda(t)A_{\lambda(t)}(x(t))\right\| \le 2\|\dot{x}(t)\| + 2\frac{|\lambda'(t)|}{\lambda(t)}\|x(t) - z\| = \mathcal{O}\left(\frac{1}{t}\right), \text{ as } t \to +\infty. \tag{21}$$

Combining the latter relation with (20), we deduce that

$$\left\|\lambda(t)\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| = \mathcal{O}\left(\frac{1}{t}\right), \text{ as } t \to +\infty.$$

Consequently, we have

$$\left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| = \mathcal{O}\left(\frac{1}{t^3}\right), \text{ as } t \to +\infty, \tag{22}$$

which implies

$$\left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| = o\left(\frac{1}{t^{2+\eta}}\right), \text{ as } t \to +\infty, \text{ for all } 0 < \eta < 1.$$

Let us improve the estimate (20), and show that

$$\|A_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t^2}\right), \text{ as } t \to +\infty.$$

To this end, we use the techniques of [16]. We have

$$\left|\frac{d}{dt}\|\lambda(t)A_{\lambda(t)}(x(t))\|^4\right| = 4\left|\left\langle\lambda(t)A_{\lambda(t)}(x(t)), \frac{d}{dt}(\lambda(t)A_{\lambda(t)}(x(t)))\right\rangle\right| \|\lambda(t)A_{\lambda(t)}(x(t))\|^2. \tag{23}$$

According to (20) and (21) there exists $K > 0$ such that

$$\left|\left\langle \lambda(t)A_{\lambda(t)}(x(t)), \frac{d}{dt}(\lambda(t)A_{\lambda(t)}(x(t)))\right\rangle\right| \leq \|\lambda(t)A_{\lambda(t)}(x(t))\| \left\|\frac{d}{dt}(\lambda(t)A_{\lambda(t)}(x(t)))\right\| \leq \frac{K}{t}.$$

Hence, (23) leads to

$$\left|\frac{d}{dt}\|\lambda(t)A_{\lambda(t)}(x(t))\|^4\right| \leq \frac{4K}{t}\|\lambda(t)A_{\lambda(t)}(x(t))\|^2. \tag{24}$$

According to (18) the right hand side of (24) belongs to $\mathrm{L}^1(t_0, +\infty)$, which implies

$$\frac{d}{dt}\|\lambda(t)A_{\lambda(t)}(x(t))\|^4 \in \mathrm{L}^1(t_0, +\infty).$$

Therefore

$$\lim_{t\to+\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\|^4 \text{ exists.}$$

But then, $L := \lim_{t\to+\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\|^2$ also exists. Using (18) again, *i.e.*

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

we deduce that $L = 0$. Therefore, $\lim_{t\to+\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\|^2 = 0$, which gives

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

Finally, by using (DIN-AVD) we have

$$\|\ddot{x}(t)\|^2 = \left\|e(t) - \frac{\alpha}{t}\dot{x}(t) - \beta\frac{d}{dt}A_{\lambda(t)}(x(t)) - A_{\lambda(t)}(x(t))\right\|^2.$$

According to an elementary convexity inequality, we get

$$t^3\|\ddot{x}(t)\|^2 \leq 4t^3\|e(t)\|^2 + 4\alpha^2 t\|\dot{x}(t)\|^2 + 4\beta^2 t^3 \left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\|^2 + 4t^3 \left\|A_{\lambda(t)}(x(t))\right\|^2.$$

According to $\int_{t_0}^{+\infty} t^3\|e(t)\|^2 dt < +\infty$, (17), (22) and (18) we obtain that

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

Let us now prove that $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$, as $t \to +\infty$. We have

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

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

Hence,

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

According to (26) and (17) we have $t^3\|\ddot{x}(t)\|^2 + 3t\|\dot{x}(t)\|^2 \in \mathrm{L}^1(t_0, +\infty)$. Therefore, from [1, Lemma 5.1] there exists $\lim_{t\to+\infty} t^2\|\dot{x}(t)\|^2 \in \mathbb{R}$. Using (17) again, we have

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

Therefore, $\lim_{t\to+\infty} t^2\|\dot{x}(t)\|^2 = 0$, and $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ as $t \to +\infty$.

The limit. To prove the existence of the weak limit of $x(t)$, we use Opial lemma. Let us introduce the anchor function $h_z(t) = \frac{1}{2}\|x(t) - z\|^2$. The classical derivation chain rule gives

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) = \left\langle \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t), x(t) - z\right\rangle + \|\dot{x}(t)\|^2.$$

By using (DIN-AVD) we get

$$\beta\left\langle \frac{d}{dt}A_{\lambda(t)}(x(t)), x(t) - z\right\rangle = \left\langle e(t) - \ddot{x}(t) - \frac{\alpha}{t}\dot{x}(t) - A_{\lambda(t)}(x(t)), x(t) - z\right\rangle$$
$$= \left\langle e(t) - A_{\lambda(t)}(x(t)), x(t) - z\right\rangle - \left(\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t)\right) + \|\dot{x}(t)\|^2.$$

Therefore

$$\begin{aligned} &t\ddot{h}_z(t) + \alpha\dot{h}_z(t) + t\left\langle A_{\lambda(t)}(x(t)), x(t) - z\right\rangle \\ &= t\|\dot{x}(t)\|^2 + t\langle e(t), x(t) - z\rangle - \beta t\left\langle \frac{d}{dt}A_{\lambda(t)}(x(t)), x(t) - z\right\rangle. \end{aligned} \tag{27}$$

By Cauchy-Schwarz inequality

$$-\beta t\left\langle \frac{d}{dt}A_{\lambda(t)}(x(t)), x(t) - z\right\rangle \le \beta t\left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| \|x(t) - z\|$$
$$t\langle e(t), x(t) - z\rangle \le t\|e(t)\|\|x(t) - z\|.$$

According to (27), we deduce that

$$\begin{aligned} &t\ddot{h}_z(t) + \alpha\dot{h}_z(t) + t\left\langle A_{\lambda(t)}(x(t)), x(t) - z\right\rangle \\ &\le t\|\dot{x}(t)\|^2 + \beta t\left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| \|x(t) - z\| + t\|e(t)\|\|x(t) - z\|. \end{aligned}$$

According to $\|x(t) - z\|$ is bounded, $t\|e(t)\| \in L^1(t_0, +\infty)$, (17) and (22) we have

$$t\|\dot{x}(t)\|^2 + \beta t\left\|\frac{d}{dt}A_{\lambda(t)}(x(t))\right\| \|x(t) - z\| + t\|e(t)\|\|x(t) - z\| \in L^1(t_0, +\infty).$$

Moreover $t \mapsto t\left\langle A_{\lambda(t)}(x(t)), x(t) - z\right\rangle$ is a non-negative function. So, we can apply Lemma A.6 from [16], and obtain that $\lim_{t\to+\infty} h_z(t)$ exists. In other words,

$$\lim_{t\to+\infty} \|x(t) - z\| \text{ exists for all } z \in S.$$

To complete the proof via the Opial's lemma, we need to prove that every weak sequential cluster point of $x(t)$ belongs to S. To this end, we use the following property of the Yosida approximation of a maximally monotone operator:

$$A_\lambda(x) \in A(x - \lambda A_\lambda(x)), \quad \text{for all } x \in \mathcal{H} \text{ and } \lambda > 0. \tag{28}$$

Let $t_n \to +\infty$ such that $x(t_n) \rightharpoonup x^*$, $n \to +\infty$. Since the graph of A is demi-closed, that is, the graph of A is sequentially closed in the product of the weak topology of $\mathcal{H}$ and strong topology of $\mathcal{H}$, by using (25) we have

$$0 = \lim_{n\to+\infty} A_{\lambda(t_n)}(x(t_n)) \in A(\lim_{n\to+\infty}(x(t_n) - \lambda(t_n)A_{\lambda(t_n)}(x(t_n)))) = A(x^*).$$

Consequently, $x(t)$ converges weakly to an element of S. □

4 Some Numerical Experiments

Take $A : \mathbb{R}^2 \to \mathbb{R}^2,\ A(x, y) = (-y, x)$, which is a linear skew symmetric operator. Clearly, A is a maximally monotone whose single zero is $x^* = (0, 0)$. Further, A can be identified with the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and an easy computation shows that the Yosida regularization of A can be identified with the matrix $A_\lambda = \begin{pmatrix} \frac{\lambda}{1+\lambda^2} & \frac{-1}{1+\lambda^2} \\ \frac{1}{1+\lambda^2} & \frac{\lambda}{1+\lambda^2} \end{pmatrix}$.

As a standing assumption, throughout the following numerical experiments, we take the coefficients $\lambda(t)$ of the form $\lambda(t) = \lambda t^2$, and the perturbation errors of the form $e(t) = \left(\frac{e_1}{t^p}, \frac{e_2}{t^q}\right)$, $p, q > 2,\ e_1, e_2 \in \mathbb{R}$. Further, we consider the starting points $u_0 = (1, 1),\ v_0 = (1, 1)$. Note that this choice of the initial speed vector v_0 means that, at the start, the trajectory tends to move away from the origin. Obviously, a trajectory of (DIN-AVD) in this case is of the form $x(t) = (x_1(t), x_2(t))$. In order to solve the dynamical system (DIN-AVD) with starting points $x(t_0) = u_0,\ \dot{x}(t_0) = v_0$ on an interval $[t_0, T]$, we use the MATLAB function ode45. To this purpose we rewrite (DIN-AVD) as the first order system used in Proposition 1, that is,

$$\begin{cases} \dot{x}_1(t) = \left(\frac{1}{\beta} - \frac{\alpha}{t} - \frac{\beta\lambda t^2}{1+\lambda^2 t^4}\right) x_1(t) + \frac{\beta}{1+\lambda^2 t^4} x_2(t) - \frac{1}{\beta} y_1(t) \\ \dot{x}_2(t) = -\frac{\beta}{1+\lambda^2 t^4} x_1(t) + \left(\frac{1}{\beta} - \frac{\alpha}{t} - \frac{\beta\lambda t^2}{1+\lambda^2 t^4}\right) x_2(t) - \frac{1}{\beta} y_2(t) \\ \dot{y}_1(t) = \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right) x_1(t) - \frac{1}{\beta} y_1(t) - \frac{\beta e_1}{t^p} \\ \dot{y}_2(t) = \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right) x_2(t) - \frac{1}{\beta} y_2(t) - \frac{\beta e_2}{t^q} \\ (x_1(t_0), x_2(t_0), y_1(t_0), y_2(t_0)) = \\ \left(1, 1, -\beta^2 \frac{\lambda t_0^2 - 1}{1+\lambda^2 t_0^4} - \frac{\alpha\beta}{t_0} - \beta + 1, -\beta^2 \frac{\lambda t_0^2 + 1}{1+\lambda^2 t_0^4} - \frac{\alpha\beta}{t_0} - \beta + 1\right). \end{cases} \tag{29}$$

In the case $\beta = 0$, we rewrite (DIN-AVD) as in Theorem 1 (b), that is

$$\begin{cases} \dot{x}_1(t) = y_1(t) \\ \dot{x}_2(t) = y_2(t) \\ \dot{y}_1(t) = -\frac{\lambda t^2}{1+\lambda^2 t^4} x_1(t) + \frac{1}{1+\lambda^2 t^4} x_2(t) - \frac{\alpha}{t} y_1(t) + \frac{e_1}{t^p} \\ \dot{y}_2(t) = -\frac{1}{1+\lambda^2 t^4} x_1(t) - \frac{\lambda t^2}{1+\lambda^2 t^4} x_2(t) - \frac{\alpha}{t} y_2(t) + \frac{e_2}{t^q} \\ (x_1(t_0), x_2(t_0), y_1(t_0), y_2(t_0) = (1, 1, 1, 1). \end{cases} \tag{30}$$

I. In our first experiment, we are interested in the asymptotic behavior of the components x_1 and x_2, obtained by taking different values of the parameters $\alpha, \beta, \lambda, e_1, e_2, p, q$. According to Theorem 2 (i), $\lim_{t\to+\infty}(x_1(t), x_2(t)) = (0, 0)$, under the conditions

$$\alpha > 1,\ \beta \geq 0, \lambda > \frac{1}{(\alpha-1)^2}, \int_{t_0}^{+\infty} t^3\|e(t)\|^2 dt < +\infty, \int_{t_0}^{+\infty} t\|e(t)\| dt < +\infty.$$

We are interested in the gain that the term $\beta\frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right)$ brings in (DIN-AVD) (considered with error term or without error term). We consider the following cases which fit the assumptions of Theorem 2 (recalled just above).

α	β	λ	e_1	e_2	p	q	Figure
2.5	0	0.5	0	0	–	–	Fig. 1a
2.5	0.5	0.5	0	0	–	–	Fig. 1b
2.5	0	0.5	1	1	3.1	3.1	Fig. 1c
2.5	0.5	0.5	1	1	3.1	3.1	Fig. 1d

The trajectories obtained by solving (29) and (30) with the ode45 function in Matlab on the interval [0.1, 50] are depicted at Fig. 1a–d, where we represent the component $x_1(t)$ with red and $x_2(t)$ with black.

II. In our second experiment, we examine the rapid convergence of the speed towards 0. According to Theorem 2, we have $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ as $t \to +\infty$. So, next to $\|\dot{x}(t)\|$, we also represent the entity $t\|\dot{x}(t)\|$. We solve (29) and (30) with the ode45 function in Matlab on the interval [0.1, 50] by considering the following instances.

α	β	λ	e_1	e_2	p	q	Figure
2.1	0	1	0	0	–	–	Fig. 2a
2.1	0	1	1	1	3	3	Fig. 2b
2.1	0.25	1	0	0	–	–	Fig. 2c
2.1	0.75	1	1	1	3	3	Fig. 2d

Note that for these values also, the hypotheses of Theorem 2 are verified, and, therefore, its conclusions are valid. The results obtained are depicted at Fig. 2a–d, where we represent $t\|\dot{x}(t)\|$ with red and $\|\dot{x}(t)\|$ with black.

A conclusion that these experiment give, is that indeed the term $\beta\frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right)$ in (DIN-AVD) has an accelerating effect, even in the presence of an error term e.

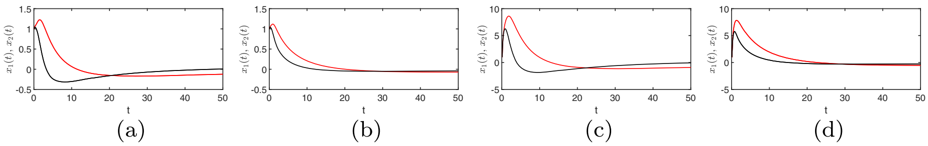

Fig. 1 Trajectories of (DIN-AVD) for different instances of the parameters $\alpha, \beta, \lambda, e$

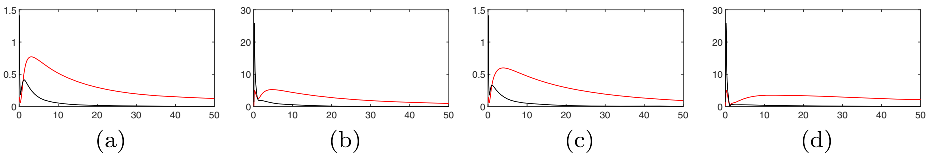

Fig. 2 Fast convergence of the velocity for different instances of α, β, λ, e

III. In our third experiment, we are interested in the behavior of the error $\|x(t) - x^*\|$. Since the operator A has a single zero at $x^* = (0, 0)$, according to Theorem 4, $\|x(t)\|$ must converge to 0. Our experiment shows the importance of the correction term $\beta \frac{d}{dt}\left(A_{\lambda(t)}(x(t))\right)$ in (DIN-AVD) as well as the sensitivity of the trajectories with respect to the error e.

Hence, we solve (29) and (30) with the ode45 function in Matlab on the interval $[0.1, 50]$ where we take at first the error $e(t) \equiv 0$ that is $e_1 = e_2 = 0$. For the values $\beta \in \{0, 0.25, 0.5, 0.75, 1\}$ we consider $(\alpha, \lambda) = (2.5, 0.5)$ depicted at Fig. 3a and $(\alpha, \lambda) = (3.1, 0.25)$ depicted at Fig. 3b, respectively. Obviously, for these values the hypotheses of Theorem 2 are satisfied, and consequently its conclusions hold.

Consider now $e_1 = e_2 = 1$ and $p = q = 3$, that is $e(t) = \left(\frac{1}{t^3}, \frac{1}{t^3}\right)$. For the values $\beta \in \{0, 0.25, 0.5, 0.75, 1\}$ we consider $(\alpha, \lambda) = (2.5, 0.5)$ depicted at Fig. 4a and $(\alpha, \lambda) = (3.1, 1)$ depicted at Fig. 4b, respectively.

Besides the importance of the correction term ($\beta > 0$), these numerical experiments show that the trajectories of (DIN-AVD) are quite sensitive to the parameters α, β, λ as well as to the error $e(t)$.

5 The Convex Case

Let us specialize the previous results to the case of convex minimization, and show the rapid convergence of values. Given a lower semicontinuous convex and proper function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ such that $\operatorname{argmin} f \neq \emptyset$, we consider the minimization problem

$$(\mathcal{P}) \quad \inf_{x \in \mathcal{H}} f(x).$$

Fermat's rule states that x is a global minimum of f if and only if

$$0 \in \partial f(x). \tag{31}$$

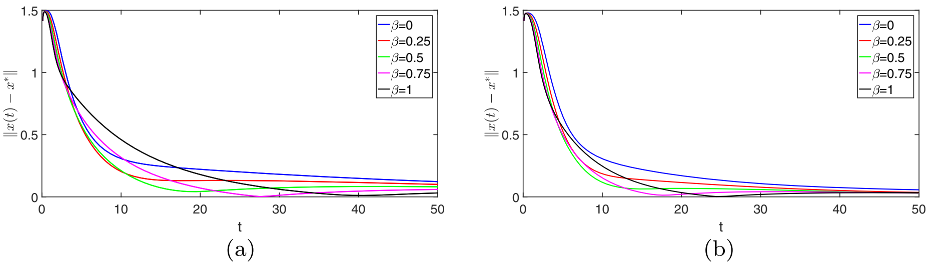

Fig. 3 Comparison of the iteration error $\|x(t) - x^*\|$ for different instances of (DIN-AVD) without error

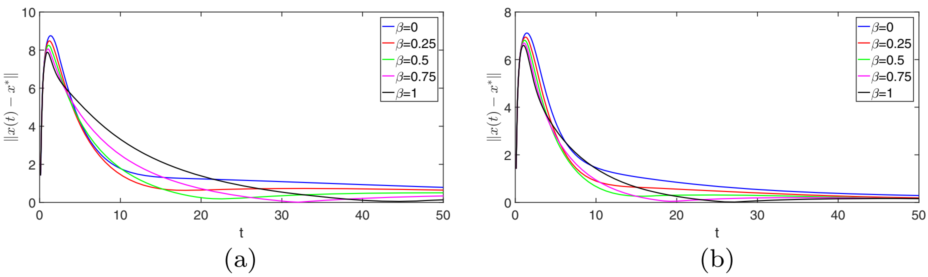

Fig. 4 Comparison of the iteration error $\|x(t) - x^*\|$ for different instances of (DIN-AVD) with error

Therefore, $(\mathcal{P})$ is equivalent to the monotone inclusion problem (1) with $A = \partial f$. Moreover, $\operatorname{argmin} f = (\partial f)^{-1}(0)$. Recall that the convex subdifferential of the function f at a point $x \in \mathcal{H}$ is defined by

$$\partial f(x) = \{x^* \in \mathcal{H} : \langle x^*, y - x\rangle \leq f(y) - f(x), \text{ for all } y \in \mathcal{H}\}.$$

The Yosida approximation of ∂f is equal to the gradient of the Moreau envelope of f: for any $\lambda > 0$

$$(\partial f)_\lambda = \nabla f_\lambda. \tag{32}$$

Recall that $f_\lambda : \mathcal{H} \to \mathbb{R}$ is a $\mathcal{C}^{1,1}$ function, which is defined by: for any $x \in \mathcal{H}$

$$f_\lambda(x) = \inf_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{1}{2\lambda}\|x - \xi\|^2 \right\}.$$

Further, the proximal point operator of λf at a point $x \in \mathcal{H}$ is defined by

$$\operatorname{prox}_{\lambda f}(x) = \operatorname{argmin}_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{1}{2\lambda}\|x - \xi\|^2 \right\}.$$

We deduce at once that $\min_{x\in\mathcal{H}} f_\lambda(x) = \min_{x\in\mathcal{H}} f(x)$. In this context, (DIN-AVD) reads as follows: for $t \geq t_0 > 0$

$$\text{(DIN-AVD)-convex} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\left(\nabla f_{\lambda(t)}(x(t))\right) + \nabla f_{\lambda(t)}(x(t)) = e(t),$$

where $\alpha > 1$, $\beta \geq 0$ and $e \in L^1_{loc}(t_0, +\infty; \mathcal{H})$.

As a direct consequence of Theorem 2 we have the following result.

Theorem 3 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous convex proper function such that $S = \operatorname{argmin} f \neq \emptyset$. Assume further that $\lambda(t) = \lambda t^2$ with $\lambda > \dfrac{1}{(\alpha - 1)^2}$ and*

$$\int_{t_0}^{+\infty} t^3\|e(t)\|^2 dt < +\infty \text{ and } \int_{t_0}^{+\infty} t\|e(t)\|dt < +\infty.$$

Then, for any trajectory $x : [t_0, +\infty[\to \mathcal{H}$ of (DIN-AVD)-convex the following properties are satisfied:

(i) *(convergence) $x(t)$ is bounded, and $x(t)$ converges weakly, as $t \to +\infty$, to an element of S.*

(ii) *(integral estimates)* $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty$, $\int_{t_0}^{+\infty} t^3\|\ddot{x}(t)\|^2 dt < +\infty$ *and* $\int_{t_0}^{+\infty} t^3\|\nabla f_{\lambda(t)}(x(t))\|^2 dt < +\infty$.

(iii) *(pointwise estimates)* $\lim_{t\to+\infty}\|\dot{x}(t)\| = 0$, $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ *as* $t \to +\infty$ *and, for all* $0 < \eta < 1$, *as* $t \to +\infty$ *we have*

$$\|\nabla f_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t^2}\right), \quad \left\|\frac{d}{dt}\nabla f_{\lambda(t)}(x(t))\right\| = o\left(\frac{1}{t^{2+\eta}}\right).$$

(iv) *(fast convergence of the values) As* $t \to +\infty$

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right) \text{ and } f(\mathrm{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right).$$

In addition, $\|\mathrm{prox}_{\lambda(t)f}(x(t)) - x(t)\| \to 0$ *as* $t \to +\infty$.

Proof (i)-(iii) follow directly from Theorem 2 applied to ∂f and using (32).
(iv) Take $x^* \in \mathrm{argmin}\, f$. From the gradient inequality, and $x(t)$ bounded, we have

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

where $M := \sup_{t\geq t_0}\|x(t) - x^*\|$. Combining the above relation with the estimate obtained in (25)), $\|\nabla f_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t^2}\right)$ as $t \to +\infty$, we obtain

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right) \text{ as } t \to +\infty. \tag{33}$$

By definition of $f_{\lambda(t)}$ and of the proximal mapping, we have

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = f(\mathrm{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f + \frac{1}{2\lambda(t)}\|x(t) - \mathrm{prox}_{\lambda(t)f}(x(t))\|^2. \tag{34}$$

Combining (33) with (34), we obtain, as $t \to +\infty$

$$f(\mathrm{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right), \quad \lim_{t\to+\infty} t^2\frac{1}{2\lambda(t)}\|x(t) - \mathrm{prox}_{\lambda(t)f}(x(t))\|^2 = 0.$$

Therefore, $\lim_{t\to+\infty}\|x(t) - \mathrm{prox}_{\lambda(t)f}(x(t))\| = 0$, which completes the proof. □

Remark 1 When $A = \partial f$, f convex, we have additional tools, such as the gradient inequality. We will show in the following theorem that, in this case, some assumptions can be weakened.

Theorem 4 *Let* $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ *be a lower semicontinuous convex, proper function such that* $S = \mathrm{argmin}_{\mathcal{H}} f \neq \emptyset$. *Assume that the parameters of (DIN-AVD) satisfy*

$$\alpha > 3,\ \beta \geq 0 \text{ and } \lambda(t) = \lambda t^r, \text{ with } \lambda > 0,\ r \geq 0.$$

Suppose that the error terms satisfy the integrability properties

$$\int_{t_0}^{+\infty} t^3\|e(t)\|^2 dt < +\infty \text{ and } \int_{t_0}^{+\infty} t\|e(t)\| dt < +\infty.$$

Then, for any trajectory $x : [t_0, +\infty[\to \mathcal{H}$ *of (DIN-AVD)-convex the following properties are satisfied.*

a) *In the general case of* $\beta \geq 0, r \geq 0$ *we have the following properties:*
(pointwise estimates) *The trajectory* $x(t)$ *is bounded, and as* $t \to +\infty$

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right).$$

$f(\text{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$, *and* $\|x(t) - \text{prox}_{\lambda(t)f}(x(t))\| = \mathcal{O}\left(\frac{\sqrt{\lambda(t)}}{t}\right)$.

Further we have $\|\nabla f_{\lambda(t)}(x(t))\| = \mathcal{O}\left(\frac{1}{t\sqrt{\lambda(t)}}\right)$, *and* $\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right)$.

Whenever $r = 2$, *one has* $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ *as* $t \to +\infty$.

(integral estimates) $\int_{t_0}^{+\infty} t(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f)dt < +\infty$, $\int_{t_0}^{\infty} t\|\dot{x}(t)\|^2 dt < +\infty$, $\int_{t_0}^{\infty} t\lambda(t)\|\nabla f_{\lambda(t)}(x(t))\|^2 dt < +\infty$ *and* $\beta \int_{t_0}^{\infty} t^2\|\nabla f_{\lambda(t)}(x(t))\|^2 dt < +\infty$.

Further, if $r \leq 2$ $\int_{t_0}^{+\infty} t\lambda(t)\|\ddot{x}(t)\|^2 dt < +\infty$.

b) *In the case* $\beta = 0$ *or* $\beta > 0$ *and* $r > 1$ *we have the following properties:*
$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right)$, *as* $t \to +\infty$, $f(\text{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right)$,
$\|x(t) - \text{prox}_{\lambda(t)f}(x(t))\| = o\left(\frac{\sqrt{\lambda(t)}}{t}\right)$ *as* $t \to +\infty$. *Further we have* $\|\nabla f_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t\sqrt{\lambda(t)}}\right)$ *as* $t \to +\infty$. *Moreover,* $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ *as* $t \to +\infty$.

c) *In the case* $\beta = 0$ *and* $r \in [0, 2]$ *or* $\beta > 0$ *and* $r \in]1, 2]$ *the trajectory* $x(t)$ *converges weakly, as* $t \to +\infty$, *to an element of* S.

Proof **Lyapunov analysis.** Take $z \in S$. Since $\alpha > 3$ we can take $2 < b < \alpha - 1$. In this particular case, the energy functional (3) becomes

$$\mathcal{E}_b(t) = \frac{1}{2}\|b(x(t) - z) + t(\dot{x}(t) + \beta\nabla f_{\lambda(t)}(x(t)))\|^2 + \frac{b(\alpha - 1 - b)}{2}\|x(t) - z\|^2. \quad (35)$$

Using the same arguments as in the proof of Theorem 2, we obtain (9)), which in this particular case reads as

$$\begin{aligned}\dot{\mathcal{E}}_b(t) \leq\ & b(\beta - t)\left\langle\nabla f_{\lambda(t)}(x(t)), x(t) - z\right\rangle + \left(b + 1 - \alpha + \frac{1}{4p_1}\right)t\|\dot{x}(t)\|^2 \\ &+ (-t^2 + \beta(b + 2 - \alpha)t)\left\langle\nabla f_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle + \left(\beta(\beta - t)t + \frac{\beta}{2}t\right)\|\nabla f_{\lambda(t)}(x(t))\|^2 \\ &+ bt\|e(t)\|\|x(t) - z\| + \left(p_1 + \frac{\beta}{2}\right)t^3\|e(t)\|^2. \end{aligned} \quad (36)$$

Further we have

$$(-t^2+\beta(b+2-\alpha)t)\left\langle \nabla f_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle = \frac{d}{dt}\left((-t^2+\beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f)\right) + (2t-\beta(b+2-\alpha))(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f). \tag{37}$$

Take $0<\epsilon<b-2$. Then there exists $t_1 \geq t_0$ such that $b(\beta-t)+\epsilon t<0$ for all $t\geq t_1$. So, by the gradient inequality we get, for all $t \geq t_1$

$$(b(\beta-t)+\epsilon t)\left\langle \nabla f_{\lambda(t)}(x(t)), x(t)-z\right\rangle \leq (b(\beta-t)+\epsilon t)(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f). \tag{38}$$

By adding (37) with (38), we obtain, for all $t \geq t_1$

$$\begin{aligned}&b(\beta-t)\left\langle \nabla f_{\lambda(t)}(x(t)), x(t)-z\right\rangle + (-t^2+\beta(b+2-\alpha)t)\left\langle \nabla f_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle \\ &\leq \frac{d}{dt}\left((-t^2+\beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f)\right) \\ &+(b(\beta-t)+\epsilon t+2t-\beta(b+2-\alpha))(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f) \\ &-\epsilon t\left\langle \nabla f_{\lambda(t)}(x(t)), x(t)-z\right\rangle.\end{aligned} \tag{39}$$

Combining (36) and (39) we get, for all $t \geq t_1$

$$\begin{aligned}&\dot{\mathcal{E}}_b(t)+\frac{d}{dt}\left((t^2-\beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f)\right)+\epsilon t\left\langle \nabla f_{\lambda(t)}(x(t)), x(t)-z\right\rangle \\ &\leq ((-b+\epsilon+2)t-\beta(2-\alpha))(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f) \\ &+\left(b+1-\alpha+\frac{1}{4p_1}\right)t\|\dot{x}(t)\|^2+\left(\beta(\beta-t)t+\frac{\beta}{2}t\right)\|\nabla f_{\lambda(t)}(x(t))\|^2 \\ &+bt\|e(t)\|\|x(t)-z\|+\left(p_1+\frac{\beta}{2}\right)t^3\|e(t)\|^2.\end{aligned} \tag{40}$$

By using the cocoerciveness of $\nabla f_{\lambda(t)}$ and the gradient inequality, we obtain

$$\epsilon t\left\langle \nabla f_{\lambda(t)}(x(t)), x(t)-z\right\rangle \geq \frac{\epsilon}{2}t\lambda(t)\|\nabla f_{\lambda(t)}(x(t))\|^2+\frac{\epsilon}{2}t(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f).$$

Further, take

$$\epsilon_1<2(\alpha-1-b), \quad p_1=\frac{1}{\epsilon_1}, \quad 0<\epsilon_2<1.$$

Then, (40) leads to

$$\begin{aligned}&\dot{\mathcal{E}}_b(t)+\frac{d}{dt}\left((t^2-\beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f)\right) \\ &+\frac{\epsilon}{2}t\lambda(t)\|\nabla f_{\lambda(t)}(x(t))\|^2+\frac{\epsilon}{2}t(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f)+\frac{\epsilon_1}{4}t\|\dot{x}(t)\|^2 \\ &+\beta\epsilon_2 t^2\|\nabla f_{\lambda(t)}(x(t))\|^2 \\ &\leq ((-b+\epsilon+2)t-\beta(2-\alpha))(f_{\lambda(t)}(x(t))-\min_{\mathcal{H}} f) \\ &+\left(b+1-\alpha+\frac{\epsilon_1}{2}\right)t\|\dot{x}(t)\|^2+\left(\beta(\beta-t)t+\beta\epsilon_2 t^2+\frac{\beta}{2}t\right)\|\nabla f_{\lambda(t)}(x(t))\|^2 \\ &+bt\|e(t)\|\|x(t)-z\|+\left(\frac{1}{\epsilon_1}+\frac{\beta}{2}\right)t^3\|e(t)\|^2, \text{ for all } t\geq t_1.\end{aligned} \tag{41}$$

Now, obviously there exists $t_2 \geq t_1$ such that for all $t \geq t_2$ one has

$$t^2 - \beta(b+2-\alpha)t > 0,$$
$$(-b+\epsilon+2)t - \beta(2-\alpha) \leq 0$$
$$\beta(\beta - t)t + \beta\epsilon_2 t^2 + \frac{\beta}{2}t \leq 0.$$

Hence, for all $t \geq t_2$, it holds

$$\begin{aligned}\dot{\mathcal{E}}_b(t) \;+\; &\frac{d}{dt}\left((t^2 - \beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f)\right) + \frac{\epsilon}{2}t\lambda(t)\|\nabla f_{\lambda(t)}(x(t))\|^2\\ &+ \frac{\epsilon}{2}t(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) + \frac{\epsilon_1}{4}t\|\dot{x}(t)\|^2 + \beta\epsilon_2 t^2\|\nabla f_{\lambda(t)}(x(t))\|^2 \leq\\ &+ bt\|e(t)\|\|x(t)-z\| + \left(\frac{1}{\epsilon_1} + \frac{\beta}{2}\right)t^3\|e(t)\|^2. \qquad (42)\end{aligned}$$

Estimates. By integrating (42) on an interval $[t_2, t]$, and by denoting

$$C_0 = \left(\frac{1}{\epsilon_1} + \frac{\beta}{2}\right)\int_{t_2}^{+\infty} t^3\|e(t)\|^2 dt + \mathcal{E}_b(t_2) + (t_2^2 - \beta(b+2-\alpha)t_2)(f_{\lambda(t_2)}(x(t_2)) - \min_{\mathcal{H}} f)$$

we obtain that, for all $t \geq t_2$

$$\begin{aligned}&\mathcal{E}_b(t) + \left((t^2 - \beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f)\right) + \frac{\epsilon}{2}\int_{t_2}^t s\lambda(s)\|\nabla f_{\lambda(s)}(x(s))\|^2 ds\\ &+\frac{\epsilon}{2}\int_{t_2}^t s(f_{\lambda(s)}(x(s)) - \min_{\mathcal{H}} f)ds + \frac{\epsilon_1}{4}\int_{t_2}^t s\|\dot{x}(s)\|^2 ds + \beta\epsilon_2\int_{t_2}^t s^2\|\nabla f_{\lambda(s)}(x(s))\|^2 ds\\ &\leq C_0 + b\int_{t_2}^t s\|e(s)\|\|x(s)-z\|ds. \qquad (43)\end{aligned}$$

Taking into account the form of the energy functional $\mathcal{E}_b(t)$, (43) leads to

$$\frac{b(\alpha-1-b)}{2}\|x(t)-z\|^2 \leq C_0 + b\int_{t_2}^t s\|e(s)\|\|x(s)-z\|ds.$$

More precisely, we have

$$\frac{1}{2}\|x(t)-z\|^2 \leq C_1 + \frac{1}{\alpha-1-b}\int_{t_2}^t s\|e(s)\|\|x(s)-z\|ds, \qquad (44)$$

where $C_1 = \frac{C_0}{b(\alpha-1-b)}$. Now, applying the Gronwall lemma (see [22, Lemma A.5]) to (44), we obtain

$$\|x(t)-z\| \leq \sqrt{2C_1} + \frac{1}{\alpha-1-b}\int_{t_2}^t s\|e(s)\|ds.$$

Therefore, $\|x(t)-z\|$ and consequently $x(t)$ are bounded. Further, the boundedness of $\|x(t)-z\|$ leads to

$$\int_{t_2}^{+\infty} s\|e(s)\|\|x(s)-z\|ds < +\infty.$$

Consequently, (43) becomes

$$\begin{aligned}\mathcal{E}_b(t) + \left((t^2 - \beta(b+2-\alpha)t)(f_{\lambda(t)}(x(t)) - \min\nolimits_{\mathcal{H}} f)\right) + \tfrac{\epsilon}{2}\textstyle\int_{t_2}^t s\lambda(s)\|\nabla f_{\lambda(s)}(x(s))\|^2 ds\\ +\tfrac{\epsilon}{2}\textstyle\int_{t_2}^t s(f_{\lambda(s)}(x(s)) - \min\nolimits_{\mathcal{H}} f)ds + \tfrac{\epsilon_1}{4}\textstyle\int_{t_2}^t s\|\dot{x}(s)\|^2 ds\\ +\beta\epsilon_2\textstyle\int_{t_2}^t s^2\|\nabla f_{\lambda(s)}(x(s))\|^2 ds \leq C, \qquad (45)\end{aligned}$$

where $C := C_0 + b\int_{t_2}^{+\infty} s\|e(s)\|\|x(s)-z\|ds < +\infty$. From (45) we get

$$\sup_{t\geq t_0}\|b(x(t)-z)+t(\dot{x}(t)+\beta\nabla f_{\lambda(t)}(x(t)))\| < +\infty \tag{46}$$

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right), \text{ as } t\to+\infty. \tag{47}$$

$$\int_{t_0}^{+\infty} t(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f)dt < +\infty. \tag{48}$$

$$\int_{t_0}^{\infty} t\|\dot{x}(t)\|^2 dt < +\infty. \tag{49}$$

$$\int_{t_0}^{\infty} t\lambda(t)\|\nabla f_{\lambda(t)}(x(t))\|^2 dt < +\infty. \tag{50}$$

$$\beta\int_{t_0}^{\infty} t^2\|\nabla f_{\lambda(t)}(x(t))\|^2 dt < +\infty. \tag{51}$$

Note that (51) provides information only in the case $\beta > 0$. Now, (46) gives

$$\|\dot{x}(t)+\beta\nabla f_{\lambda(t)}(x(t))\| = \mathcal{O}\left(\frac{1}{t}\right), \text{ as } t\to+\infty. \tag{52}$$

Further, from (47) and

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = f(\text{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f + \frac{1}{2\lambda(t)}\|x(t)-\text{prox}_{\lambda(t)f}(x(t))\|^2,$$

we deduce that, as $t\to+\infty$

$$f(\text{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right), \ \|x(t)-\text{prox}_{\lambda(t)f}(x(t))\| = \mathcal{O}\left(\frac{\sqrt{\lambda(t)}}{t}\right) \tag{53}$$

Further we have $\nabla f_{\lambda(t)} = (\partial f)_{\lambda(t)} = \frac{1}{\lambda(t)}(I - \text{prox}_{\lambda(t)f})$, hence

$$\|\nabla f_{\lambda(t)}(x(t))\| = \mathcal{O}\left(\frac{1}{t\sqrt{\lambda(t)}}\right) \text{ as } t\to+\infty. \tag{54}$$

Combining (52) and (54) we get

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right), \text{ as } t\to+\infty. \tag{55}$$

In particular we have

$$\lim_{t\to+\infty}\|\dot{x}(t)\| = 0.$$

Now, according to Lemma 1 we have

$$\left\|\frac{d}{dt}\nabla f_{\lambda(t)}(x(t))\right\| \leq \frac{2}{\lambda(t)}\|\dot{x}(t)\| + 2\frac{|\lambda'(t)|}{\lambda(t)}\|\nabla f_{\lambda(t)}(x(t))\|$$

which yields

$$\left\|\frac{d}{dt}\nabla f_{\lambda(t)}(x(t))\right\|^2 \leq \frac{8}{\lambda^2(t)}\|\dot{x}(t)\|^2 + 8\frac{\lambda'^2(t)}{\lambda^2(t)}\|\nabla f_{\lambda(t)}(x(t))\|^2.$$

Finally, by using (DIN-AVD)-convex, we have

$$\|\ddot{x}(t)\|^2 = \left\|e(t) - \frac{\alpha}{t}\dot{x}(t) - \beta\frac{d}{dt}\nabla f_{\lambda(t)}(x(t)) - \nabla f_{\lambda(t)}(x(t))\right\|^2.$$

Therefore,

$$\begin{aligned} t\lambda(t)\|\ddot{x}(t)\|^2 &\leq 4t\lambda(t)\|e(t)\|^2 + \left(\frac{4\alpha^2\lambda(t)}{t} + \frac{32\beta^2 t}{\lambda(t)}\right)\|\dot{x}(t)\|^2 \\ &\quad + \left(\frac{32\beta^2\lambda'^2(t)t}{\lambda(t)} + 4t\lambda(t)\right)\left\|\nabla f_{\lambda(t)}(x(t))\right\|^2. \end{aligned}$$

Recall that $\lambda(t) = \lambda t^r$ and assume that $r \leq 2$. Then, according to the fact that $\int_{t_0}^{+\infty} t^3 \|e(t)\|^2 dt < +\infty$, (49) and (50) we obtain that

$$\int_{t_0}^{+\infty} t\lambda(t)\|\ddot{x}(t)\|^2 dt < +\infty. \tag{56}$$

Let us now prove that $\|\dot{x}(t)\| = o\left(\frac{1}{t^{\frac{1}{2}+\frac{r}{4}}}\right)$, as $t \to +\infty$. We have

$$\frac{d}{dt} t^{1+\frac{r}{2}} \|\dot{x}(t)\|^2 = \left(1 + \frac{r}{2}\right) t^{\frac{r}{2}} \|\dot{x}(t)\|^2 + 2t^{1+\frac{r}{2}} \langle \ddot{x}(t), \dot{x}(t)\rangle$$

and

$$2t^{1+\frac{r}{2}} \langle \ddot{x}(t), \dot{x}(t)\rangle \leq t^{1+r} \|\ddot{x}(t)\|^2 + t\|\dot{x}(t)\|^2.$$

Hence,

$$\frac{d}{dt} t^{1+\frac{r}{2}} \|\dot{x}(t)\|^2 \leq t^{1+r} \|\ddot{x}(t)\|^2 + \left(\left(1 + \frac{r}{2}\right) t^{\frac{r}{2}} + t\right) \|\dot{x}(t)\|^2.$$

According to (56) and (49), we have $t^{1+r} \|\ddot{x}(t)\|^2 + \left(\left(1 + \frac{r}{2}\right) t^{\frac{r}{2}} + t\right) \|\dot{x}(t)\|^2 \in \mathrm{L}^1[t_0, +\infty[$. Consequently, from [1, Lemma 5.1] we obtain that there exists

$$\lim_{t\to+\infty} t^{1+\frac{r}{2}} \|\dot{x}(t)\|^2 \in \mathbb{R}.$$

Now, using (49) again, we have

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

hence,

$$\lim_{t\to+\infty} t^{1+\frac{r}{2}} \|\dot{x}(t)\|^2 = 0.$$

Consequently,

$$\|\dot{x}(t)\| = o\left(\frac{1}{t^{\frac{1}{2}+\frac{r}{4}}}\right), \text{ as } t \to +\infty.$$

The limit. To prove the existence of the weak limit of $x(t)$, we use Opial lemma, and we follow the line of proof of Theorem 2 by considering the same anchor function h_z and taking $A_{\lambda(t)} = \nabla f_{\lambda(t)}$.

We obtain

$$\begin{aligned} &t\ddot{h}_z(t) + \alpha\dot{h}_z(t) + t\left\langle \nabla f_{\lambda(t)}(x(t)), x(t) - z\right\rangle \\ &\leq t\|\dot{x}(t)\|^2 + \beta t \left\| \frac{d}{dt} \nabla f_{\lambda(t)}(x(t)) \right\| \|x(t) - z\| + t\|e(t)\| \|x(t) - z\|. \end{aligned}$$

Now, according to Lemma 1(c_2) we have

$$\left\| \frac{d}{dt} \nabla f_{\lambda(t)}(x(t)) \right\| \leq \frac{2}{\lambda(t)} \|\dot{x}(t)\| + 2\frac{|\lambda'(t)|}{\lambda(t)} \|\nabla f_{\lambda(t)}(x(t))\|.$$

Therefore,

$$\begin{aligned} &t\ddot{h}_z(t) + \alpha\dot{h}_z(t) + t\left\langle \nabla f_{\lambda(t)}(x(t)), x(t) - z\right\rangle \\ &\leq t\|\dot{x}(t)\|^2 + t\|e(t)\| \|x(t) - z\| + \left(\tfrac{2\beta t}{\lambda(t)} \|\dot{x}(t)\| + 2\beta \tfrac{|\lambda'(t)| t}{\lambda(t)} \|\nabla f_{\lambda(t)}(x(t))\|\right) \| \|x(t) - z\|. \end{aligned} \tag{57}$$

Now, if $\beta = 0$ or $\beta > 0$ and $r > 1$ then, according to the fact that $\|x(t) - z\|$ is bounded, $t\|e(t)\| \in L^1(t_0, +\infty)$, (49), (55) and (54) we have

$$t\|\dot{x}(t)\|^2 + \left(t\|e(t)\| + \frac{2\beta t}{\lambda(t)}\|\dot{x}(t)\| + 2\beta\frac{|\lambda'(t)|t}{\lambda(t)}\|\nabla f_{\lambda(t)}(x(t))\| \right)\|\|x(t) - z\| \in L^1(t_0, +\infty).$$

Moreover $t\left\langle \nabla f_{\lambda(t)}(x(t)), x(t) - z \right\rangle$ is a non-negative term. So, we can apply Lemma A.6 from [16], and obtain that $\lim_{t\to+\infty} h_z(t)$ exists. In other words,

$$\lim_{t\to+\infty} \|x(t) - z\| \text{ exists for all } z \in S.$$

Let us return to the fact that, according to (42), for all $t \geq t_2$

$$\dot{\mathcal{E}}_b(t) + \frac{d}{dt}\left((t^2 - \beta(b + 2 - \alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) \right) \leq bt\|e(t)\|\|x(t) - z\| + \left(\tfrac{1}{\epsilon_1} + \tfrac{\beta}{2}\right) t^3\|e(t)\|^2.$$

The right hand side of the latter inequality belongs to $L^1(t_0, +\infty)$. Therefore, according to [1, Lemma 5.1] we get that there exists

$$\lim_{t\to+\infty} \mathcal{E}_b(t) + (t^2 - \beta(b + 2 - \alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) \in \mathbb{R}.$$

Since $\lim_{t\to+\infty} \|x(t) - z\|$ exists for all $z \in S$, we obtain that there exists

$$\lim_{t\to+\infty} \frac{1}{2}\|t(\dot{x}(t) + \beta\nabla f_{\lambda(t)}(x(t)))\|^2 + (t^2 - \beta(b + 2 - \alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) \in \mathbb{R}.$$

On the other hand, from (48), (49) and (51) we get

$$\int_{t_0}^{+\infty} \tfrac{1}{t}\left(\tfrac{1}{2}\|t(\dot{x}(t) + \beta\nabla f_{\lambda(t)}(x(t)))\|^2 + (t^2 - \beta(b + 2 - \alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) \right) dt$$
$$\leq \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt + \int_{t_0}^{+\infty} \beta^2 t\|\nabla f_{\lambda(t)}(x(t)))\|^2 dt$$
$$+ \int_{t_0}^{+\infty} (t - \beta(b + 2 - \alpha))(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) dt < +\infty.$$

Consequently,

$$\lim_{t\to+\infty} \frac{1}{2}\|t(\dot{x}(t) + \beta\nabla f_{\lambda(t)}(x(t)))\|^2 + (t^2 - \beta(b + 2 - \alpha)t)(f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f) = 0,$$

which shows that, as $t \to +\infty$

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right), \text{ and } \|\dot{x}(t)\| = o\left(\frac{1}{t}\right).$$

From

$$f_{\lambda(t)}(x(t)) - \min_{\mathcal{H}} f = f(\text{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f + \frac{1}{2\lambda(t)}\|x(t) - \text{prox}_{\lambda(t)f}(x(t))\|^2,$$

we deduce that, as $t \to +\infty$

$$f(\text{prox}_{\lambda(t)f}(x(t))) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right), \ \|x(t) - \text{prox}_{\lambda(t)f}(x(t))\| = o\left(\frac{\sqrt{\lambda(t)}}{t}\right) \quad (58)$$

Further we have $\nabla f_{\lambda(t)} = (\partial f)_{\lambda(t)} = \frac{1}{\lambda(t)}(I - \text{prox}_{\lambda(t)f})$, hence

$$\|\nabla f_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t\sqrt{\lambda(t)}}\right) \text{ as } t \to +\infty. \quad (59)$$

It remains to show that every weak sequential cluster point of the trajectory $x(t)$ belongs to argmin f. Let x^* be a weak sequential cluster point of $x(t)$. Then, there exists a sequence

$t_n \to +\infty$, $n \to +\infty$ such that $x(t_n) \rightharpoonup x^*$, $n \to +\infty$. According to (58), if $r \leq 2$, we have $\lim_{n\to+\infty} \|x(t_n) - \text{prox}_{\lambda(t_n)f}(x(t_n))\| = 0$. Therefore,

$$\text{prox}_{\lambda(t_n)f}(x(t_n)) \rightharpoonup x^*, \ n \to +\infty.$$

Since f is lower semicontinuous and convex, it is weakly lower semicontinuous. Combined with $\lim_{n\to+\infty}(f(\text{prox}_{\lambda(t_n)f}(x(t_n))) - \min_{\mathcal{H}} f) = 0$, it yields

$$0 = \liminf_{n\to+\infty}(f(\text{prox}_{\lambda(t_n)f}(x(t_n))) - \min_{\mathcal{H}} f) \geq f(x^*) - \min_{\mathcal{H}} f.$$

The latter relation shows that $x^* \in \text{argmin}\, f$. Consequently, according to Opial lemma, $x(t)$ converges weakly to an element $\hat{x} \in \text{argmin}\, f$ as $t \to +\infty$. □

6 Conclusion, Perspective

Recent developments in convex optimization show the importance of the introduction of the Hessian driven damping in the continuous versions of the Nesterov accelerated gradient method. It allows to control and attenuate the oscillations, resulting in faster methods. The extension of these results to general monotone inclusions is an important and non-trivial question. Our study of the continuous dynamic (DIN-AVD) gives a solid mathematical basis to the algorithmic results obtained by the authors for these questions, and confirm them. Dealing with these issues in the context of general maximally monotone operators offers a wide range of applications. It is a natural idea for further study to specialize this study in the case of convex-concave saddle value problems. The convergence results are valid in the presence of perturbations or errors. This is an important step to study other instances. Among them, the introduction of a Tikhonov regularization term with vanishing coefficient, in order to asymptotically obtain the solution of minimum norm. It also suggests developing stochastic versions of (DIN-AVD) and corresponding algorithms in the context of general maximally monotone operators. To deal with concrete examples, it would be very interesting to develop corresponding splitting methods to solve structured monotone inclusions. Finally, it would also be interesting to consider the closed loop version of these dynamics and algorithms where the coefficient λ of the Yosida regularization is taken as a feedback control of the state or the velocity of the system.

Appendix: Auxiliary results

In the proof of Theorem 2, we use the following straightforward result.

Lemma 2 *Let $A, B, C \in \mathbb{R}$. The inequality*

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

is satisfied for all $X, Y \in \mathcal{H}$, if and only if $C^2 - AB \leq 0$ and $A, B \geq 0$.

References

1. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **161**(2), 331–360 (2014)

2. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set-Valued Analysis **9**(1-2), 3–11 (2001)
3. Alvarez, 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)
4. Apidopoulos, V., Aujol, J.-F., Dossal, C.h.: Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule. Math. Program. **180**, 137–156 (2020)
5. Attouch, H., Cabot, A.: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators. Math Program. https://doi.org/10.1007/s10107-019-01412-0 (2019)
6. Attouch, H., Cabot, A.: Convergence rates of inertial forward-backward algorithms. SIAM J. Optim. **28**(1), 849–874 (2018)
7. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order algorithms via inertial systems with Hessian driven damping. HAL-02193846 (2019)
8. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. B **168**, 123–175 (2018)
9. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. SIAM J. Optim. **29**(3), 2227–2256 (2019)
10. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. ESAIM-COCV 25, Article number 2 (2019)
11. 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)
12. Attouch, H., László, S.C.: Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. HAL-02549730 (2020)
13. Attouch, H., Maingé, P.E.: Asymptotic behavior of second order dissipative evolution equations combining potential with non-potential effects. ESAIM Control Optim. Calc. Var. **17**(3), 836–857 (2011)
14. Attouch, H., Maingé, P.E., Redont, P.: A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws. Differ. Equ. Appl. **4**(1), 27–65 (2012)
15. Attouch, H., Marques Alves, M., Svaiter, B.F.: A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert Spaces, with complexity $\mathcal{O}(1/n^2)$. J. Convex Anal. **23**(1), 139–180 (2016)
16. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators. Math. Program. **174**(1-2), 391–432 (2019)
17. Attouch, H., Peypouquet, J.: The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. SIAM J. Optim. **26**(3), 1824–1834 (2016)
18. Attouch, H., Peypouquet, J., Redont, P.: Fast convex minimization via inertial dynamics with Hessian driven damping. J. Differ. Equ. **261**(10), 5734–5783 (2016)
19. Attouch, H., Redont, P., Svaiter, B.F.: Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **157**(3), 624–650 (2013)
20. Attouch, H., Svaiter, B.F.: A continuous dynamical Newton-Like approach to solving monotone inclusions. SIAM J. Control Optim. **49**(2), 574–598 (2011)
21. Bauschke, H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert spaces. CMS Books in Mathematics. Springer (2011)
22. Brézis, H.: Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution. Lecture Notes 5 North Holland (1972)
23. Boţ, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control Optim. **54**, 1423–1443 (2016)
24. Boţ, R.I., Csetnek, E.R., László, S.C.: A second-order dynamical approach with variable damping to nonconvex smooth minimization. Appl. Anal. **99**(3), 361–378 (2020)
25. Boţ, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian damping. Mathematical Programming. https://doi.org/10.1007/s10107-020-01528-8 (2020)
26. Castera, C., Bolte, J., Févotte, C., Pauwels, E.: An Inertial Newton Algorithm for Deep Learning. HAL-02140748 (2019)
27. Chambolle, A., Dossal, C.h.: On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm. J. Optim. Theory Appl. **166**, 968–982 (2015)
28. Haraux, A.: Systèmes dynamiques dissipatifs et applications. RMA 17 Masson (1991)
29. Kim, D.: Accelerated Proximal Point Method for Maximally Monotone Operators. arXiv:1905.05149v3 (2020)
30. László, S.C.: Convergence rates for an inertial algorithm of gradient type associated to a smooth nonconvex minimization. Mathematical Programming. https://doi.org/10.1007/s10107-020-01534-w (2020)

31. Lin, T., Jordan, M.I.: A Control-Theoretic Perspective on Optimal High-Order Optimization. arXiv:1912.07168v1 (2019)
32. May, R.: Asymptotic for a second-order evolution equation with convex potential and vanishing damping term. Turkish J. Math. **41**(3), 681–685 (2017)
33. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $o(1/k^2)$. (Russian) Dokl. Akad. Nauk SSSR **269**(3), 543–547 (1983)
34. Polyak, B.T.: Some methods of speeding up the convergence of iterative methods. Z. Vylist. Math. Fiz. **4**, 1–17 (1964)
35. Shi, B., Du, S.S., Jordan, M.I., Su, W.J.: Understanding the acceleration phenomenon via high-resolution differential equations. arXiv:submit/2440124[cs.LG] (2018)
36. Su, W.J., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. Neural Information Processing Systems **27**, 2510–2518 (2014)