



HAL
open science

Asymptotic behavior of Newton-like inertial dynamics involving the sum of potential and nonpotential terms

Samir Adly, Hedy Attouch, van Nam Vo

► **To cite this version:**

Samir Adly, Hedy Attouch, van Nam Vo. Asymptotic behavior of Newton-like inertial dynamics involving the sum of potential and nonpotential terms. 2021. hal-03213925

HAL Id: hal-03213925

<https://hal.science/hal-03213925v1>

Preprint submitted on 30 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Asymptotic behavior of Newton-like inertial dynamics involving the sum of potential and nonpotential terms.

Samir ADLY*, Hedy ATTOUCH† and Van Nam VO*

April 2, 2021

ABSTRACT. In a Hilbert space \mathcal{H} , we study a dynamic inertial Newton method which aims to solve additively structured monotone equations involving the sum of potential and nonpotential terms. Precisely, we are looking for the zeros of an operator $A = \nabla f + B$ where ∇f is the gradient of a continuously differentiable convex function f , and B is a nonpotential monotone and cocoercive operator. Besides a viscous friction term, the dynamic involves geometric damping terms which are controlled respectively by the Hessian of the potential f and by a Newton-type correction term attached to B . Based on a fixed point argument, we show the well-posedness of the Cauchy problem. Then we show the weak convergence as $t \rightarrow +\infty$ of the generated trajectories towards the zeros of $\nabla f + B$. The convergence analysis is based on the appropriate setting of the viscous and geometric damping parameters. The introduction of these geometric dampings makes it possible to control and attenuate the known oscillations for the viscous damping of inertial methods. Rewriting the second-order evolution equation as a first-order dynamical system enables us to extend the convergence analysis to nonsmooth convex potentials. These results open the door to the design of new first-order accelerated algorithms in optimization taking into account the specific properties of potential and nonpotential terms. The proofs and techniques are original and differ from the classical ones due to the presence of the nonpotential term.

Mathematics Subject Classifications:

Key words and phrases: proximal-gradient algorithms; inertial methods; Hessian driven damping; nonpotential term; cocoercive operators.

Contents

| | | |
|----------|--|----------|
| 1 | Introduction and preliminary results | 2 |
| 1.1 | Dynamical inertial Newton method for additively structured monotone problems | 2 |
| 1.2 | Historical aspects of the inertial systems with Hessian-driven damping | 4 |
| 1.3 | Inertial dynamics involving cocoercive operators | 4 |
| 1.4 | Link with Newton-like methods for solving monotone inclusions | 5 |
| 1.5 | Contents | 5 |
| 2 | Well posedness of the Cauchy-Lipschitz problem | 5 |
| 2.1 | First-order in time and space equivalent formulation | 5 |

*Laboratoire XLIM, Université de Limoges, 123, avenue Albert Thomas, 87060 Limoges, France. E-mail: samir.adly@unilim.fr, van-nam.vo@etu.unilim.fr

†IMAG, Université Montpellier, CNRS, Place Eugène Bataillon, 34095 Montpellier CEDEX 5, France. E-mail: hedy.attouch@umontpellier.fr

| | | |
|----------|---|-----------|
| 2.2 | Well-posedness of the evolution equation (DINAM) | 7 |
| 3 | Asymptotic convergence properties of (DINAM) | 8 |
| 3.1 | General case | 8 |
| 3.2 | Case $\beta_b = \beta_f$ | 15 |
| 4 | Numerical illustrations | 15 |
| 5 | The nonsmooth case | 20 |
| 6 | Conclusion, perspectives | 23 |
| 7 | Appendix | 24 |
| 7.1 | Technical lemmas | 24 |

1 Introduction and preliminary results

Let \mathcal{H} be a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$. Many situations coming from physics, biology, human sciences involve equations containing both potential and nonpotential terms. In human sciences, this comes from the presence of both cooperative and noncooperative aspects. To describe such situations we will focus on solving additively structured monotone equations of the type

$$\text{Find } x \in \mathcal{H} : \nabla f(x) + B(x) = 0. \quad (1.1)$$

In the above equation, ∇f is the gradient of a convex continuously differentiable function $f : \mathcal{H} \rightarrow \mathbb{R}$ (that's the potential part), and $B : \mathcal{H} \rightarrow \mathcal{H}$ is a nonpotential operator¹ which is supposed to be monotone and cocoercive. To this end, we will consider continuous inertial dynamics whose solution trajectories converge as $t \rightarrow +\infty$ to solutions of (1.1). Our study is part of the active research stream that studies the close relationship between continuous dissipative dynamical systems and optimization algorithms which are obtained by their temporal discretization. To avoid lengthening the paper, we limit our study to the analysis of the continuous dynamic. The analysis of the algorithmic part and its link with first-order numerical optimization will be carried out in a second companion paper. From this perspective, damped inertial dynamics offers a natural way to accelerate these systems. As the main feature of our study, we will introduce into the dynamic geometric dampings which are respectively driven by the Hessian for the potential part, and by the corresponding Newton term for the nonpotential part. In addition to improving the convergence rate, this will considerably reduce the oscillatory behavior of the trajectories. We will pay particular attention to the minimal assumptions which guarantee convergence of the trajectories, and which highlight the asymmetric role played by the two operators involved in the dynamic. We will see that many results can be extended to the case where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function, which makes it possible to broaden the field of applications.

1.1 Dynamical inertial Newton method for additively structured monotone problems

Let us introduce the following second-order differential equation which will form the basis of our analysis:

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) + B(x(t)) + \beta_f \nabla^2 f(x(t)) \dot{x}(t) + \beta_b B'(x(t)) \dot{x}(t) = 0, \quad t \geq t_0. \quad (\text{DINAM})$$

We use (DINAM) as an abbreviation for Dynamical Inertial Newton method for Additively structured Monotone problems. We call $t_0 \in \mathbb{R}$ the origin of time. Since we are considering autonomous systems,

¹*i.e.* B is not supposed to be equal to the gradient of a given function.

we can take any arbitrary real number for t_0 . For simplicity, we set $t_0 = 0$. When considering the corresponding Cauchy problem, we add the initial conditions: $x(0) = x_0 \in \mathcal{H}$ and $\dot{x}(0) = x_1 \in \mathcal{H}$. The term $B'(x(t))\dot{x}(t)$ is interpreted as $\frac{d}{dt}(B(x(t)))$ taken in the distribution sense. Likewise the term $\nabla^2 f(x(t))\dot{x}(t)$ is interpreted as $\frac{d}{dt}(\nabla f(x(t)))$ taken also in the distribution sense. Because of the assumptions made below these terms are indeed measurable functions which are bounded on the bounded time intervals. So, we will consider strong solutions of the above equation (DINAM). Throughout the paper we make the following standing assumptions:

$$\left\{ \begin{array}{l} \text{(A1) } f : \mathcal{H} \rightarrow \mathbb{R} \text{ is convex, of class } \mathcal{C}^1, \nabla f \text{ is Lipschitz continuous on the bounded sets;} \\ \text{(A2) } B : \mathcal{H} \rightarrow \mathcal{H} \text{ is a } \lambda\text{-cocoercive operator for some } \lambda > 0; \\ \text{(A3) } \gamma > 0, \beta_f \geq 0, \beta_b \geq 0 \text{ are given real damping parameters.} \end{array} \right.$$

We emphasize the fact that we do not assume the gradient of f to be globally Lipschitz continuous. Developing our analysis without using any bound on the gradient of f is a key to further extend the theory to the nonsmooth case. As a specific property, the inertial system (DINAM) combines two different types of driving forces associated respectively with the potential operator ∇f and the nonpotential operator B . It also involves three different types of friction:

- (a) The term $\gamma\dot{x}(t)$ models viscous damping with a positive coefficient $\gamma > 0$.
- (b) The term $\beta_f \nabla^2 f(x(t))\dot{x}(t)$ is the so-called Hessian driven damping, which allows to attenuate the oscillations that naturally occur with the inertial gradient dynamics.
- (c) The term $\beta_b B'(x(t))\dot{x}(t)$ is the nonpotential version of the Hessian driven damping. It can be interpreted as a Newton-type correction term.

Note that each driving force term enters (DINAM) with its temporal derivative. In fact, we have

$$\nabla^2 f(x(t))\dot{x}(t) = \frac{d}{dt}(\nabla f(x(t))) \text{ and } B'(x(t))\dot{x}(t) = \frac{d}{dt}(B(x(t))).$$

This is a crucial observation which makes (DINAM) equivalent to a first order system in time and space, and makes the corresponding Cauchy problem well posed. This will be proved later (see subsection 2.1 for more details). The cocoercivity assumption on the operator B plays an important role in the analysis of (DINAM), not only to ensure the existence of solutions, but also to analyze their asymptotic behavior as time $t \rightarrow +\infty$.

Recall that the operator $B : \mathcal{H} \rightarrow \mathcal{H}$ is said to be λ -cocoercive for some $\lambda > 0$ if

$$\langle By - Bx, y - x \rangle \geq \lambda \|By - Bx\|^2, \forall x, y \in \mathcal{H}.$$

Note that B is λ -cocoercive is equivalent to B^{-1} is λ -strongly monotone, *i.e.* cocoercivity is a dual notion of strong monotonicity. It is easy to check that B is λ -cocoercive implies that B is $1/\lambda$ -Lipschitz continuous. The reverse implication holds true in the case where the operator is the gradient of a convex and differentiable function. Indeed, according to Baillon-Haddad's Theorem [19], ∇f is L -Lipschitz continuous implies that ∇f is a $1/L$ -cocoercive operator (we refer to [20, Corollary 18.16] for more details).

1.2 Historical aspects of the inertial systems with Hessian-driven damping

The following inertial system with Hessian-driven damping

$$\ddot{x}(t) + \gamma\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0,$$

was first considered by Alvarez-Attouch-Peypouquet-Redont in [6]. Then, according to the continuous interpretation by Su-Boyd-Candès [31] of the accelerated gradient method of Nesterov, Attouch-Peypouquet-Redont [16] replaced the fixed viscous damping parameter γ by an asymptotic vanishing damping parameter $\frac{\alpha}{t}$, with $\alpha > 0$. 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 these dynamics provides first-order algorithms of the form

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

As a specific feature, and by comparison with the classical accelerated gradient methods, these algorithms contain a correction term which is equal to the difference of the gradients at two consecutive steps. While preserving the convergence properties of the accelerated gradient 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 [8], Boţ-Csetnek-László [22], Kim [27], Lin-Jordan [28], Shi-Du-Jordan-Su [30], and Alesca-Lazlo-Pinta [4] for an implicit version of the Hessian driven damping. Application to deep learning has been recently developed by Castera-Bolte-Févotte-Pauwels [25]. In [3], Adly-Attouch studied the finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping.

1.3 Inertial dynamics involving cocoercive operators

Let's come to the transposition of these techniques to the case of maximally monotone operators. Álvarez-Attouch [5] and Attouch-Maingé [12] studied the equation

$$\ddot{x}(t) + \gamma\dot{x}(t) + A(x(t)) = 0, \tag{1.2}$$

when $A : \mathcal{H} \rightarrow \mathcal{H}$ is a cocoercive (and hence maximally monotone) operator, (see also [21]). The cocoercivity assumption plays an important role in the study of (1.2), 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 $\lambda\gamma^2 > 1$, Attouch-Maingé [12] showed that each trajectory of (1.2) converges weakly to a zero of A , i.e. $x(t) \rightharpoonup x_\infty \in A^{-1}(0)$ as $t \rightarrow +\infty$. Moreover, the condition $\lambda\gamma^2 > 1$ is sharp.

For general maximally monotone operators this property has been further exploited by Attouch-Peypouquet [15], and by Attouch-Laszlo [10, 11]. The key property is that for $\lambda > 0$, the Yosida approximation A_λ of A is λ -cocoercive and $A_\lambda^{-1}(0) = A^{-1}(0)$. So the idea is to replace the operator A by its Yosida approximation, and adjust the Yosida regularization parameter. Another related work has been done by Attouch-Maingé [12] who first consider the asymptotic behavior of the second order dissipative evolution equation with $f : \mathcal{H} \rightarrow \mathbb{R}$ convex and $B : \mathcal{H} \rightarrow \mathcal{H}$ cocoercive

$$\ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) + B(x(t)) = 0, \tag{1.3}$$

combining potential with nonpotential effects. Our study will therefore consist initially in introducing the Hessian term and the Newton-type correcting term into this dynamic.

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 (1.1). 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 [18],

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

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 assumption on the regularization parameter $\gamma(t)$, this system is well posed and generates trajectories that converge weakly to equilibria (zeroes of A). Parallel results have been obtained for the associated proximal algorithms obtained by implicit temporal discretization, see [2], [14], [17]. Formally, this system is written as

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

Thus (DINAM) can be considered as an inertial version of this dynamical system for structured monotone operator $A = \nabla f + B$. Our study is also linked to the recent works by Attouch-Laszlo [10, 11] who considered the general case of monotone equations. By contrast with [10, 11], according to the cocoercivity of B , we don't use the Yosida regularization, and exhibit minimal assumptions involving only the nonpotential component.

1.5 Contents

The paper is organized as follows. Section 1 introduces (DINAM) with some historical perspective. In section 2, based on the first order equivalent formulation of (DINAM), we show that the Cauchy problem is well-posed (in the sense of existence and uniqueness of solutions). In section 3, we analyze the asymptotic convergence properties of the trajectories generated by (DINAM). Using appropriate Lyapunov functions, we show that any trajectory of (DINAM) converges weakly as $t \rightarrow +\infty$, and that its limit belongs to $S = (\nabla f + B)^{-1}(0)$. The interplay between the damping parameters β_f, β_b, γ and the cocoercivity parameter λ will play an important role in our Lyapounov analysis. In Section 4, we perform numerical experiments showing that the well-known oscillations in the case of the heavy ball with friction, are damped with the introduction of the geometric (Hessian-like) damping terms. An application to the LASSO problem with a nonpotential operator as well as a coupled system in dynamical games are considered. Section 5 deals with the extension of the study to the nonsmooth and convex case. Section 6 contains some concluding remarks and perspectives.

2 Well posedness of the Cauchy-Lipschitz problem

We first show the existence and the uniqueness of the solution trajectory for the Cauchy problem associated with (DINAM) for any given initial condition data $(x_0, x_1) \in \mathcal{H} \times \mathcal{H}$.

2.1 First-order in time and space equivalent formulation

The following first-order equivalent formulation of (DINAM) was first considered by Alvarez-Attouch-Bolte-Redont [6] and Attouch-Peypouquet-Redont [16] in the framework of convex minimization. Specifically, in our context, we have the following equivalence, which follows from a simple differential and algebraic calculation.

Proposition 2.1 Suppose that $\beta_f > 0$. Then, the following problems are equivalent: (i) \iff (ii)

$$(i) \quad \ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) + B(x(t)) + \beta_f \nabla^2 f(x(t)) \dot{x}(t) + \beta_b B'(x(t)) \dot{x}(t) = 0.$$

$$(ii) \quad \begin{cases} \dot{x}(t) + \beta_f \nabla f(x(t)) + \beta_b B(x(t)) + \left(\gamma - \frac{1}{\beta_f}\right) x(t) + y(t) = 0; \\ \dot{y}(t) - \left(1 - \frac{\beta_b}{\beta_f}\right) B(x(t)) + \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f}\right) x(t) + \frac{1}{\beta_f} y(t) = 0. \end{cases}$$

Proof. (i) \implies (ii). For $t \geq 0$, set

$$y(t) := -\dot{x}(t) - \beta_f \nabla f(x(t)) - \beta_b B(x(t)) - \left(\gamma - \frac{1}{\beta_f}\right) x(t), \quad (2.1)$$

which gives the first equation of (ii). By differentiating $y(\cdot)$ and using (i), we get

$$\begin{aligned} \dot{y}(t) &= -\ddot{x}(t) - \beta_f \nabla^2 f(x(t)) \dot{x}(t) - \beta_b B'(x(t)) \dot{x}(t) - \left(\gamma - \frac{1}{\beta_f}\right) \dot{x}(t) \\ &= \gamma \dot{x}(t) + \nabla f(x(t)) + B(x(t)) - \left(\gamma - \frac{1}{\beta_f}\right) \dot{x}(t) \\ &= \nabla f(x(t)) + B(x(t)) + \frac{1}{\beta_f} \dot{x}(t). \end{aligned} \quad (2.2)$$

By combining (2.1) and (2.2), we obtain

$$\dot{y}(t) + \frac{1}{\beta_f} y(t) = \left(1 - \frac{\beta_b}{\beta_f}\right) B(x(t)) - \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f}\right) x(t). \quad (2.3)$$

This gives the second equation of (ii).

(ii) \implies (i). By differentiating the first equation of (ii), we obtain

$$\ddot{x}(t) + \beta_f \nabla^2 f(x(t)) \dot{x}(t) + \beta_b B'(x(t)) \dot{x}(t) + \left(\gamma - \frac{1}{\beta_f}\right) \dot{x}(t) + \dot{y}(t) = 0. \quad (2.4)$$

Let us eliminate y from this equation to obtain an equation involving only x . For this, we successively use the second equation in (ii), then the first equation in (ii) to obtain

$$\begin{aligned} \dot{y}(t) &= \left(1 - \frac{\beta_b}{\beta_f}\right) B(x(t)) - \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f}\right) x(t) - \frac{1}{\beta_f} y(t) \\ &= \left(1 - \frac{\beta_b}{\beta_f}\right) B(x(t)) - \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f}\right) x(t) + \frac{1}{\beta_f} \dot{x}(t) \\ &\quad + \nabla f(x(t)) + \frac{\beta_b}{\beta_f} B(x(t)) + \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f}\right) x(t). \end{aligned}$$

Therefore,

$$\dot{y}(t) = \nabla f(x(t)) + B(x(t)) + \frac{1}{\beta_f} \dot{x}(t). \quad (2.5)$$

From (2.4) and (2.5), we obtain (i). ■

2.2 Well-posedness of the evolution equation (DINAM)

The following theorem shows the well-posedness of the Cauchy problem for the evolution equation (DINAM).

Theorem 2.1 *Suppose that $\beta_f > 0$ and $\beta_b \geq 0$. Then, for any $(x_0, x_1) \in \mathcal{H} \times \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\rightarrow \mathcal{H}$ of the continuous dynamic (DINAM) which satisfies the Cauchy data $x(0) = x_0, \dot{x}(0) = x_1$.*

Proof. The system (ii) in Proposition 2.1 can be written equivalently as

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

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

$$F(x, y) = \beta_f(\nabla f(x), 0) + \left(\beta_b B(x) + \left(\gamma - \frac{1}{\beta_f} \right) x + y, - \left(1 - \frac{\beta_b}{\beta_f} \right) B(x) + \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f} \right) x + \frac{1}{\beta_f} y \right),$$

$$y_0 = -x_1 - \beta_f \nabla f(x_0) - \beta_b B(x_0) - \left(\gamma - \frac{1}{\beta_f} \right) x_0.$$

Therefore, $F = \nabla \Phi + G$ where $\Phi : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ is the convex differentiable function

$$\Phi(x, y) := \beta_f f(x)$$

and $G : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H}$

$$G(x, y) := \left(\beta_b B(x) + \left(\gamma - \frac{1}{\beta_f} \right) x + y, - \left(1 - \frac{\beta_b}{\beta_f} \right) B(x) + \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f} \right) x + \frac{1}{\beta_f} y \right)$$

is a Lipschitz continuous map. Indeed, the Lipschitz continuity of G is a direct consequence of the Lipschitz continuity of B . The existence of a classical solution to

$$\dot{Z}(t) + \nabla \Phi(Z(t)) + G(Z(t)) = 0, \quad Z(0) = (x_0, y_0)$$

follows from Brézis [23, Proposition 3.12]. In fact, the proof of this result relies on a fixed point argument. It consists in finding a fixed point of the mapping $u \in \mathcal{C}([0, T], \mathcal{H}) \mapsto K(u) \in \mathcal{C}([0, T], \mathcal{H})$ where $K(u) = w$ is the solution of

$$\dot{w}(t) + \nabla \Phi(w(t)) = -G(u(t)), \quad w(0) = (x_0, y_0).$$

It is proved that the sequence of iterates (w_n) generated by the corresponding Picard iteration

$$\dot{w}_{n+1}(t) + \nabla \Phi(w_{n+1}(t)) = -G(w_n(t)), \quad w_{n+1}(0) = (x_0, y_0),$$

converges uniformly on $[0, T]$ to a fixed point of K . When returning to (DINAM), that's equation (i) of Proposition 2.1, we recover a strong solution. Precisely, \dot{x} is Lipschitz continuous on the bounded time intervals, and \ddot{x} taken in the distribution sense is locally essentially bounded. ■

Remark 2.1 Note that when ∇f is supposed to be globally Lipschitz continuous, the above proof can be notably simplified, by just applying the classical Cauchy-Lipschitz theorem.

3 Asymptotic convergence properties of (DINAM)

In this section, we study the asymptotic behavior of the solution trajectories of (DINAM). For each solution trajectory $t \mapsto x(t)$ of (DINAM) we show that the weak limit, $w\text{-}\lim_{t \rightarrow +\infty} x(t) = x_\infty$ exists, and satisfies $x_\infty \in S$, where

$$S := \{p \in \mathcal{H} : \nabla f(p) + B(p) = 0\}.$$

Before stating our main result, notice that $B(p)$ is uniquely defined for $p \in S$.

Lemma 3.1 *$B(p)$ is uniquely defined for $p \in S$, i.e.*

$$p_1 \in S, p_2 \in S \implies B(p_1) = B(p_2).$$

Proof. Since $p_1 \in S, p_2 \in S$ we have

$$\nabla f(p_1) + B(p_1) = \nabla f(p_2) + B(p_2) = 0.$$

By monotonicity of ∇f we have

$$\langle \nabla f(p_2) - \nabla f(p_1), p_2 - p_1 \rangle \geq 0.$$

Replacing $\nabla f(p_1)$ by $-B(p_1)$ and $\nabla f(p_2)$ by $-B(p_2)$, we get

$$\langle B(p_2) - B(p_1), p_2 - p_1 \rangle \leq 0,$$

which by cocoercivity of B gives $\lambda \|B(p_2) - B(p_1)\|^2 \leq 0$, and hence $B(p_2) = B(p_1)$. ■

3.1 General case

The general line of the demonstration is close to one given by Attouch-Laszlo in [10, 11]. A first major difference with the approach developed in [10, 11] is that in our context thanks to the hypothesis of cocoercivity on the nonpotential part, we do not need to go through the Yosida regularization of the operators. A second difference is that we treat the potential and nonpotential operators in a differentiated way. These points are crucial for applications to numerical algorithms, because the computation of the Yosida regularization of the sum of the two operators is often out of reach numerically.

The following Theorem states the asymptotic convergence properties of (DINAM).

Theorem 3.1 *Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a λ -cocoercive operator and $f : \mathcal{H} \rightarrow \mathbb{R}$ be a C^1 convex function whose gradient is Lipschitz continuous on the bounded sets. Suppose that $S = (\nabla f + B)^{-1}(0) \neq \emptyset$. Consider the evolution equation (DINAM) where the involved parameters satisfy the following conditions: $\beta_f > 0$ and*

$$4\lambda\gamma > \frac{(\beta_b - \beta_f)^2}{\beta_f} + 2 \left(\beta_b + \frac{1}{\gamma} \right) + 2\sqrt{\left(\beta_b + \frac{1}{\gamma} \right)^2 + \frac{(\beta_b - \beta_f)^2}{\gamma\beta_f}}.$$

Then, for any solution trajectory $x : [0, +\infty[\rightarrow \mathcal{H}$ of (DINAM) the following properties are satisfied:

- (i) (convergence) $x(t)$ converges weakly, as $t \rightarrow +\infty$, to an element of S .

(ii) (integral estimates) Set $A := B + \nabla f$ and $p \in S$. Then,

$$\begin{aligned} \int_0^{+\infty} \|\dot{x}(t)\|^2 dt &< +\infty, \int_0^{+\infty} \|\ddot{x}(t)\|^2 dt < +\infty, \\ \int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt &< +\infty, \int_0^{+\infty} \left\| \frac{d}{dt} B(x(t)) \right\|^2 dt < +\infty, \\ \int_0^{+\infty} \|A(x(t))\|^2 dt &< +\infty, \text{ and } \int_0^{+\infty} \left\| \frac{d}{dt} A(x(t)) \right\|^2 dt < +\infty. \end{aligned}$$

(iii) (pointwise estimates)

$$\lim_{t \rightarrow +\infty} \|\dot{x}(t)\| = 0, \lim_{t \rightarrow +\infty} \|B(x(t)) - B(p)\| = 0, \lim_{t \rightarrow +\infty} \|A(x(t))\| = 0,$$

where $B(p)$ is uniquely defined for $p \in S$.

Proof. Lyapunov analysis. Set $A := B + \nabla f$ and $A_\beta := \beta_b B + \beta_f \nabla f$. Take $p \in S$. Consider the function $t \in [0, +\infty[\mapsto \mathcal{V}_p(t) \in \mathbb{R}_+$ defined by

$$\mathcal{V}_p(t) := \frac{1}{2} \|x(t) - p + c(\dot{x}(t) + A_\beta(x(t)) - A_\beta(p))\|^2 + \frac{\delta}{2} \|x(t) - p\|^2, \quad (3.1)$$

where c and δ are coefficients to adjust. Using the derivation chain rule for absolutely continuous functions (see [24, Corollary VIII.10]) and (DINAM), we get

$$\dot{\mathcal{V}}_p(t) = \left\langle \dot{x}(t) - c(\gamma \dot{x} + A(x(t))), x(t) - p + c(\dot{x}(t) + A_\beta(x(t)) - A_\beta(p)) \right\rangle + \delta \langle \dot{x}(t), x(t) - p \rangle. \quad (3.2)$$

Setting $\delta := c\gamma - 1 > 0$, from (3.2) we obtain

$$\dot{\mathcal{V}}_p(t) = \langle -cA(x(t)), x(t) - p \rangle + c \langle (1 - c\gamma)\dot{x}(t) - cA(x(t)), \dot{x}(t) + A_\beta(x(t)) - A_\beta(p) \rangle. \quad (3.3)$$

We have

$$\begin{aligned} &c \langle (1 - c\gamma)\dot{x}(t) - cA(x(t)), \dot{x}(t) + A_\beta(x(t)) - A_\beta(p) \rangle \\ &= c(1 - c\gamma) \|\dot{x}(t)\|^2 + c(1 - c\gamma) \langle \dot{x}(t), A_\beta(x(t)) - A_\beta(p) \rangle \\ &\quad - c^2 \langle A(x(t)), \dot{x}(t) \rangle - c^2 \langle A(x(t)), A_\beta(x(t)) - A_\beta(p) \rangle, \\ &= c(1 - c\gamma) \|\dot{x}(t)\|^2 - c^2 \beta_b \|B(x(t)) - B(p)\|^2 - c^2 \beta_f \|\nabla f(x(t)) - \nabla f(p)\|^2 \\ &\quad + [c(1 - c\gamma)\beta_b - c^2] \langle \dot{x}(t), B(x(t)) - B(p) \rangle + [c(1 - c\gamma)\beta_f - c^2] \langle \dot{x}(t), \nabla f(x(t)) - \nabla f(p) \rangle \\ &\quad - c^2(\beta_b + \beta_f) \langle B(x(t)) - B(p), \nabla f(x(t)) - \nabla f(p) \rangle. \end{aligned} \quad (3.4)$$

Using the fact that $p \in S$, ∇f is monotone, and B is λ -cocoercive, we have

$$\begin{aligned} -c \langle A(x(t)), x(t) - p \rangle &= -c \langle A(x(t)) - A(p), x(t) - p \rangle \\ &= -c \langle \nabla f(x(t)) - \nabla f(p), x(t) - p \rangle - c \langle B(x(t)) - B(p), x(t) - p \rangle \\ &\leq -c\lambda \|B(x(t)) - B(p)\|^2. \end{aligned} \quad (3.5)$$

From (3.3)-(3.5), we deduce that

$$\begin{aligned} \dot{\mathcal{V}}_p(t) &\leq -c\delta \|\dot{x}(t)\|^2 - [c^2 \beta_b + c\lambda] \|B(x(t)) - B(p)\|^2 - c^2 \beta_f \|\nabla f(x(t)) - \nabla f(p)\|^2 \\ &\quad - [c\delta \beta_b + c^2] \langle \dot{x}(t), B(x(t)) - B(p) \rangle - [c\delta \beta_f + c^2] \langle \dot{x}(t), \nabla f(x(t)) - \nabla f(p) \rangle \\ &\quad - c^2(\beta_b + \beta_f) \langle B(x(t)) - B(p), \nabla f(x(t)) - \nabla f(p) \rangle. \end{aligned} \quad (3.6)$$

Let $\Gamma : [0, +\infty[\rightarrow \mathbb{R}$ be the function defined by

$$\Gamma(t) := f(x(t)) - f(p) - \langle \nabla f(p), x(t) - p \rangle,$$

and $\mathcal{E}_p : [0, +\infty[\rightarrow \mathbb{R}$ be the energy function given by

$$\mathcal{E}_p(t) := \mathcal{V}_p(t) + [c\delta\beta_f + c^2]\Gamma(t).$$

Since f is convex, we have $\Gamma(t) \geq 0$, for all $t \geq 0$. This implies $\mathcal{E}_p(t) \geq 0$ for all $t \geq 0$ as well.

We have,

$$\dot{\Gamma}(t) = \langle \dot{x}(t), \nabla f(x(t)) - \nabla f(p) \rangle, \quad (3.7)$$

$$\dot{\mathcal{E}}_p(t) = \dot{\mathcal{V}}_p(t) + [c\delta\beta_f + c^2]\dot{\Gamma}(t). \quad (3.8)$$

By using (3.7) and (3.8), the equation (3.6) can be rewritten as

$$\begin{aligned} & \dot{\mathcal{E}}_p(t) + c\delta\|\dot{x}(t)\|^2 + [c^2\beta_b + c\lambda]\|B(x(t)) - B(p)\|^2 + c^2\beta_f\|\nabla f(x(t)) - \nabla f(p)\|^2 \\ & + [c\delta\beta_b + c^2]\langle \dot{x}(t), B(x(t)) - B(p) \rangle + c^2(\beta_b + \beta_f)\langle B(x(t)) - B(p), \nabla f(x(t)) - \nabla f(p) \rangle \leq 0. \end{aligned} \quad (3.9)$$

Let us eliminate the term $\nabla f(x(t)) - \nabla f(p)$ from this relation by using the elementary algebraic inequality

$$\begin{aligned} & c^2\beta_f\|\nabla f(x(t)) - \nabla f(p)\|^2 + c^2(\beta_b + \beta_f)\langle B(x(t)) - B(p), \nabla f(x(t)) - \nabla f(p) \rangle \\ & \geq -\frac{c^2(\beta_b + \beta_f)^2}{4\beta_f}\|B(x(t)) - B(p)\|^2. \end{aligned}$$

We obtain

$$\begin{aligned} & \dot{\mathcal{E}}_p(t) + c\delta\|\dot{x}(t)\|^2 + [c^2\beta_b + c\lambda - \frac{c^2(\beta_b + \beta_f)^2}{4\beta_f}]\|B(x(t)) - B(p)\|^2 \\ & + [c\delta\beta_b + c^2]\langle \dot{x}(t), B(x(t)) - B(p) \rangle \leq 0. \end{aligned}$$

Equivalently

$$\dot{\mathcal{E}}_p(t) + c\mathcal{S}(t) \leq 0, \quad (3.10)$$

where

$$\begin{aligned} \mathcal{S}(t) & := \delta\|\dot{x}(t)\|^2 + [\delta\beta_b + c]\langle \dot{x}(t), B(x(t)) - B(p) \rangle \\ & + [c\beta_b + \lambda - \frac{c(\beta_b + \beta_f)^2}{4\beta_f}]\|B(x(t)) - B(p)\|^2. \end{aligned}$$

Set $X(t) = \dot{x}(t)$ and $Y(t) = B(x(t)) - B(p)$. We have $\mathcal{S}(t) = q(X(t), Y(t))$ where $q : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ is the quadratic form

$$q(X, Y) := a\|X\|^2 + b\langle X, Y \rangle + g\|Y\|^2$$

with $a = \delta$, $b = \delta\beta_b + c$, and $g = c\beta_b + \lambda - \frac{c(\beta_b + \beta_f)^2}{4\beta_f} = \lambda - \frac{c(\beta_b - \beta_f)^2}{4\beta_f}$.

According to Lemma 7.3, and since $a = \delta = c\gamma - 1 > 0$, we have that q is positive definite if and only if $4ag - b^2 > 0$. Equivalently

$$4\delta\left(\lambda - \frac{c(\beta_b - \beta_f)^2}{4\beta_f}\right) - [\delta\beta_b + c]^2 > 0. \quad (3.11)$$

Our aim is to find c such that $c\gamma - 1 > 0$ and such that (3.11) is satisfied. Take $\delta := c\gamma - 1 > 0$ as a new variable. Equivalently, we must find $\delta > 0$ such that

$$4\delta \left(\lambda - \frac{\delta + 1}{\gamma} \cdot \frac{(\beta_b - \beta_f)^2}{4\beta_f} \right) - \left(\delta\beta_b + \frac{\delta + 1}{\gamma} \right)^2 > 0.$$

After development and simplification we obtain

$$4\lambda > \left[\frac{(\beta_b - \beta_f)^2}{\gamma\beta_f} + \frac{2}{\gamma} \left(\beta_b + \frac{1}{\gamma} \right) \right] + \frac{1}{\gamma^2\delta} + \left[\left(\beta_b + \frac{1}{\gamma} \right)^2 + \frac{(\beta_b - \beta_f)^2}{\gamma\beta_f} \right] \delta.$$

Therefore, we just need to assume that

$$4\lambda > \left[\frac{(\beta_b - \beta_f)^2}{\gamma\beta_f} + \frac{2}{\gamma} \left(\beta_b + \frac{1}{\gamma} \right) \right] + \inf_{\delta > 0} \left(\frac{1}{\gamma^2\delta} + \left[\left(\beta_b + \frac{1}{\gamma} \right)^2 + \frac{(\beta_b - \beta_f)^2}{\gamma\beta_f} \right] \delta \right).$$

Elementary optimization argument gives that

$$\inf_{\delta > 0} \left(\frac{C}{\delta} + D\delta \right) = 2\sqrt{CD}.$$

Therefore we end up with the condition

$$4\lambda > \left[\frac{(\beta_b - \beta_f)^2}{\gamma\beta_f} + \frac{2}{\gamma} \left(\beta_b + \frac{1}{\gamma} \right) \right] + \frac{2}{\gamma} \sqrt{\left(\beta_b + \frac{1}{\gamma} \right)^2 + \frac{(\beta_b - \beta_f)^2}{\gamma\beta_f}}.$$

Equivalently

$$4\lambda\gamma > \left[\frac{(\beta_b - \beta_f)^2}{\beta_f} + 2 \left(\beta_b + \frac{1}{\gamma} \right) \right] + 2\sqrt{\left(\beta_b + \frac{1}{\gamma} \right)^2 + \frac{(\beta_b - \beta_f)^2}{\gamma\beta_f}}. \quad (3.12)$$

When $\beta_b = \beta_f = \beta$ we recover the condition

$$\lambda\gamma > \beta + \frac{1}{\gamma}.$$

Note that $c\gamma = 1 + \delta$ and $\delta > 0$ implies $c > 0$. Therefore, there exist positive real numbers c, μ such that

$$\dot{\mathcal{E}}_p(t) + c\mu \|\dot{x}(t)\|^2 + c\mu \|B(x(t)) - B(p)\|^2 \leq 0. \quad (3.13)$$

Estimates. We have shown that there exist positive real numbers c, μ such that, for all $t \geq 0$

$$\dot{\mathcal{E}}_p(t) + c\mu \|\dot{x}(t)\|^2 + c\mu \|B(x(t)) - B(p)\|^2 \leq 0. \quad (3.14)$$

By integrating (3.14) on an interval $[0, t]$, we obtain that for all $t \geq 0$,

$$\mathcal{E}_p(t) + c\mu \int_0^t \|\dot{x}(s)\|^2 ds + c\mu \int_0^t \|B(x(s)) - B(p)\|^2 ds \leq \mathcal{E}_p(0). \quad (3.15)$$

From (3.15) and the definition of \mathcal{E}_p we immediately deduce

$$\sup_{t \geq 0} \|x(t) - p\| < +\infty, \quad (3.16)$$

$$\sup_{t \geq 0} \|x(t) - p + c(\dot{x}(t) + A_\beta(x(t)) - A_\beta(p))\| < +\infty, \quad (3.17)$$

$$\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty, \quad (3.18)$$

$$\int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt < +\infty. \quad (3.19)$$

Let us return to (3.9). We recall that

$$\begin{aligned} \dot{\mathcal{E}}_p(t) + c\delta\|\dot{x}(t)\|^2 + [c^2\beta_b + c\lambda]\|B(x(t)) - B(p)\|^2 + c^2\beta_f\|\nabla f(x(t)) - \nabla f(p)\|^2 \\ + [c\delta\beta_b + c^2]\langle \dot{x}(t), B(x(t)) - B(p) \rangle + c^2(\beta_b + \beta_f)\langle B(x(t)) - B(p), \nabla f(x(t)) - \nabla f(p) \rangle \leq 0. \end{aligned} \quad (3.20)$$

After integration on $[0, t]$, and by using the integral estimates $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$, and $\int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt < +\infty$, we obtain the existence of a constant $C > 0$ such that

$$c^2\beta_f \int_0^t \|\nabla f(x(s)) - \nabla f(p)\|^2 ds \leq C + c^2(\beta_b + \beta_f) \int_0^t \|B(x(s)) - B(p)\| \|\nabla f(x(s)) - \nabla f(p)\| ds.$$

Therefore, for any $\epsilon > 0$, we have

$$\begin{aligned} c^2\beta_f \int_0^t \|\nabla f(x(s)) - \nabla f(p)\|^2 ds \\ \leq C + c^2(\beta_b + \beta_f) \int_0^t \left(\frac{1}{4\epsilon} \|B(x(s)) - B(p)\|^2 + \epsilon \|\nabla f(x(s)) - \nabla f(p)\|^2 \right) ds. \end{aligned}$$

By taking $\epsilon > 0$ such that $\beta_f > \epsilon(\beta_b + \beta_f)$ which is always possible since $\beta_f > 0$, we conclude

$$\int_0^{+\infty} \|\nabla f(x(t)) - \nabla f(p)\|^2 dt < +\infty.$$

Combining this with $\int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt < +\infty$, it follows immediately

$$\int_0^{+\infty} \|A(x(t)) - A(p)\|^2 dt < +\infty. \quad (3.21)$$

Moreover, we also have

$$\begin{aligned} \int_0^{+\infty} \|A_\beta(x(t)) - A_\beta(p)\|^2 dt &= \int_0^{+\infty} \|\beta_f(\nabla f(x(t)) - \nabla f(p)) + \beta_b(B(x(t)) - B(p))\|^2 dt \\ &\leq (\beta_f^2 + \beta_b^2) \int_0^{+\infty} \|\nabla f(x(t)) - \nabla f(p)\|^2 + \|B(x(t)) - B(p)\|^2 dt < +\infty. \end{aligned} \quad (3.22)$$

According to (3.16) the trajectory $x(\cdot)$ is bounded. Set $R := \sup_{t \geq 0} \|x(t)\|$. By assumption, ∇f is Lipschitz continuous on the bounded sets. Let $L_R < +\infty$ be the Lipschitz constant of ∇f on $B(0, R)$. Since B is λ -cocoercive, it is $\frac{1}{\lambda}$ -Lipschitz continuous. Therefore A is L -Lipschitz continuous on the trajectory with $L := L_R + \frac{1}{\lambda}$. Therefore

$$\frac{d}{dt} \|A(x(t))\| \leq \left\| \frac{d}{dt} A(x(t)) \right\| \leq L \|\dot{x}(t)\|, \text{ for all } t \geq 0. \quad (3.23)$$

Using (3.21) and (3.23) we deduce that $u(t) := \|A(x(t))\|$ satisfies the condition of Lemma 7.2 (with $p = 2$ and $r = 2$). Therefore,

$$\lim_{t \rightarrow +\infty} \|A(x(t))\| = 0. \quad (3.24)$$

Likewise, according to (3.22) we have

$$\lim_{t \rightarrow +\infty} \|A_\beta(x(t)) - A_\beta(p)\| = 0. \quad (3.25)$$

Returning to (3.17)-(3.18), we deduce that $\dot{x}(\cdot)$ is bounded and $\lim_{t \rightarrow +\infty} \|\dot{x}(t)\| = 0$. By using the same argument as in (3.23), we obtain that $\frac{d}{dt}A_\beta(x(t))$ is bounded. From (3.23) we also get that

$$\int_0^{+\infty} \left\| \frac{d}{dt}A(x(t)) \right\|^2 dt < +\infty.$$

Similarly, we also have

$$\int_0^{+\infty} \left\| \frac{d}{dt}B(x(t)) \right\|^2 dt < +\infty.$$

By using (DINAM) we have

$$\begin{aligned} \ddot{x}(t) &= -\gamma\dot{x}(t) - A(x(t)) - \frac{d}{dt}A_\beta(x(t)) \\ &= -\gamma\dot{x}(t) - A(x(t)) - \beta_f \frac{d}{dt}A(x(t)) - (\beta_b - \beta_f) \frac{d}{dt}B(x(t)). \end{aligned}$$

Since the second member of the above equality belongs to $L^2(0, +\infty; \mathcal{H})$, we finally get

$$\int_0^{+\infty} \|\ddot{x}(t)\|^2 dt < +\infty.$$

The limit. To prove the existence of the weak limit of $x(t)$, we use Opial's lemma (see [29] for more details). Given $p \in S$, let us consider the anchor function defined by, for every $t \in [0, +\infty[$

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

From $\dot{q}_p(t) = \langle \dot{x}(t), x(t) - p \rangle$ and $\ddot{q}_p(t) = \|\dot{x}(t)\|^2 + \langle \ddot{x}(t), x(t) - p \rangle$, we obtain

$$\begin{aligned} \ddot{q}_p(t) + \gamma\dot{q}_p(t) &= \|\dot{x}(t)\|^2 + \langle \ddot{x}(t) + \gamma\dot{x}(t), x(t) - p \rangle \\ &= \|\dot{x}(t)\|^2 - \langle A(x(t)) + \frac{d}{dt}A_\beta(x(t)), x(t) - p \rangle \\ &\leq \|\dot{x}(t)\|^2 - \langle \frac{d}{dt}A_\beta(x(t)), x(t) - p \rangle. \end{aligned}$$

Equivalently,

$$\ddot{q}_p(t) + \gamma\dot{q}_p(t) + \langle \frac{d}{dt}A_\beta(x(t)), x(t) - p \rangle \leq \|\dot{x}(t)\|^2. \quad (3.26)$$

According to the derivation formula for a product, we can rewrite (3.26) as follows

$$\ddot{q}_p(t) + \gamma\dot{q}_p(t) + \frac{d}{dt} \langle A_\beta(x(t)) - A_\beta(p), x(t) - p \rangle \leq \|\dot{x}(t)\|^2 + \langle A_\beta(x(t)) - A_\beta(p), \dot{x}(t) \rangle.$$

By Cauchy-Lipschitz inequality we get

$$\ddot{q}_p(t) + \gamma\dot{q}_p(t) + \frac{d}{dt} \langle A_\beta(x(t)) - A_\beta(p), x(t) - p \rangle \leq \|\dot{x}(t)\|^2 + \|A_\beta(x(t)) - A_\beta(p)\| \|\dot{x}(t)\|. \quad (3.27)$$

Then note that the second member of (3.27)

$$g(t) := \|\dot{x}(t)\|^2 + \|A_\beta(x(t)) - A_\beta(p)\| \|\dot{x}(t)\|$$

is nonnegative and belongs to $L^1(0, +\infty)$. Indeed, we have

$$\int_0^{+\infty} \|A_\beta(x(t)) - A_\beta(p)\| \|\dot{x}(t)\| dt \leq \frac{1}{2} \int_0^{+\infty} \|A_\beta(x(t)) - A_\beta(p)\|^2 dt + \frac{1}{2} \int_0^{+\infty} \|\dot{x}(t)\|^2 dt.$$

Using (3.18) and (3.22), we deduce that

$$\int_0^{+\infty} g(t) dt < +\infty.$$

Note that the left member of (3.27) can be rewritten as a derivative of a function, precisely

$$\ddot{q}_p(t) + \gamma \dot{q}_p(t) + \frac{d}{dt} \langle A_\beta(x(t)) - A_\beta(p), x(t) - p \rangle = \dot{h}(t)$$

with

$$h(t) = \dot{q}_p(t) + \gamma q_p(t) + \langle A_\beta(x(t)) - A_\beta(p), x(t) - p \rangle. \quad (3.28)$$

So we have

$$\dot{h}(t) \leq g(t), \text{ for every } t \geq 0.$$

Let us prove that the function h given in (3.28) is bounded below by some constant. Indeed, since the terms $q_p(t)$ and $\langle A_\beta(x(t)) - A_\beta(p), x(t) - p \rangle$ are nonnegative, we have

$$h(t) \geq \dot{q}_p(t) \geq -\|\dot{x}(t)\| \|x(t) - p\|.$$

Using (3.16) and the fact that $\dot{x}(\cdot)$ is bounded, we deduce that there exists some $m \in \mathbb{R}$ such that

$$h(t) \geq m \text{ for every } t \geq 0.$$

Let us introduce the real-valued function $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$, $t \mapsto \varphi(t)$ defined by

$$\varphi(t) = h(t) - \int_0^t g(s) ds.$$

We have $\varphi'(t) = \dot{h}(t) - g(t) \leq 0$. Hence, the function φ is nonincreasing on $[0, +\infty[$. This classically implies that the limit of φ exists as $t \rightarrow +\infty$. Since $g \in L^1(0, +\infty)$, we deduce that $\lim_{t \rightarrow +\infty} h(t)$ exists.

Using the fact that $\langle A_\beta(x(t)) - A_\beta(p), x(t) - p \rangle$ tends to zero as $t \rightarrow +\infty$ (a consequence of (3.25) and $x(\cdot)$ bounded), we obtain

$$\dot{q}_p(t) + \gamma q_p(t) = \theta(t)$$

with limit of $\theta(t)$ exists as $t \rightarrow +\infty$. The existence of the limit of q_p then follows from a classical general result concerning the convergence of evolution equations governed by strongly monotone operators (here γId , see Theorem 3.9 page 88 in [23]). This means that for all $p \in S$

$$\lim_{t \rightarrow +\infty} \|x(t) - p\| \text{ exists.}$$

To complete the proof via the Opial's lemma, we need to show that every weak sequential cluster point of $x(t)$ belongs to S . Let $t_n \rightarrow +\infty$ such that $x(t_n) \rightharpoonup x^*$, $n \rightarrow +\infty$. We have

$$A(x(t_n)) \rightarrow 0 \text{ strongly in } \mathcal{H} \text{ and } x(t_n) \rightharpoonup x^* \text{ weakly in } \mathcal{H}.$$

From the closedness property of the graph of the maximally monotone operator A in $w - \mathcal{H} \times s - \mathcal{H}$, we deduce that $A(x^*) = 0$, that is $x^* \in S$.

Consequently, $x(t)$ converges weakly to an element of S as t goes to $+\infty$. The proof of Theorem 3.1 is thereby completed. ■

3.2 Case $\beta_b = \beta_f$

Let us specialize the previous results in the case $\beta_b = \beta_f$. We set $\beta_b = \beta_f := \beta > 0$, and $A := \nabla f + B$. We thus consider the evolution system

$$\text{(DINAM)} \quad \ddot{x}(t) + \gamma \dot{x}(t) + A(x(t)) + \beta \frac{d}{dt}(A(x(t))) = 0, \quad t \geq 0.$$

The existence of strong global solutions to this system is guaranteed by Theorem 2.1. The convergence properties as $t \rightarrow +\infty$ of the solution trajectories generated by this system is a consequence of Theorem 3.1 and are given below.

Corollary 3.1 *Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a λ -cocoercive operator and $f : \mathcal{H} \rightarrow \mathbb{R}$ be a \mathcal{C}^1 convex function whose gradient is Lipschitz continuous on the bounded sets. Suppose that the solution set $S = (\nabla f + B)^{-1}(0) \neq \emptyset$. Consider the evolution equation (DINAM) where $A = \nabla f + B$, $\beta_b = \beta_f := \beta > 0$ and where the involved parameters satisfy the following conditions*

$$\gamma > 0, \beta > 0, \text{ and } \lambda\gamma > \beta + \frac{1}{\gamma}. \quad (3.29)$$

Then, for any trajectory solution $x : [0, +\infty[\rightarrow \mathcal{H}$ of (DINAM) the following properties are satisfied:

(i) (convergence) *The trajectory $x(\cdot)$ is bounded, and $x(t)$ converges weakly, as $t \rightarrow +\infty$, to an element $x^* \in S$.*

(ii) (integral estimate)

$$\begin{aligned} \int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_0^{+\infty} \|\ddot{x}(t)\|^2 dt < +\infty, \\ \int_0^{+\infty} \|A(x(t))\|^2 dt < +\infty, \text{ and } \int_0^{+\infty} \left\| \frac{d}{dt} A(x(t)) \right\|^2 dt < +\infty. \end{aligned}$$

(iii) (pointwise estimate)

$$\lim_{t \rightarrow +\infty} \|\dot{x}(t)\| = 0, \text{ and } \lim_{t \rightarrow +\infty} \|A(x(t))\| = 0.$$

Remark 3.1 It is worth stating the result of Corollary 3.1 apart because this is an important case. This also makes it possible to highlight this result compared to the existing literature for second-order dissipative evolution systems involving cocoercive operators. Indeed, letting β go to zero in (3.29) gives the condition

$$\lambda\gamma^2 > 1, \quad (3.30)$$

introduced by Attouch-Maingé in [12] to study the second order dynamic (1.3) without geometric damping. With respect to [12], the introduction of the geometric damping, *i.e.*, taking $\beta > 0$, provides some useful additional estimates.

4 Numerical illustrations

In this section, we give some numerical illustrations by using a temporal discretization of the dynamic (DINAM). Let us recall the condensed formulation of (DINAM)

$$\ddot{x}(t) + \gamma \dot{x}(t) + A(x(t)) + \frac{d}{dt}(A_\beta(x(t))) = 0, \quad \text{(DINAM)}$$

where $A := \nabla f + B$ and $A_\beta := \beta_b B + \beta_f \nabla f$. Take a fixed time step $h > 0$, and consider the following implicit finite-difference scheme for (DINAM):

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \frac{1}{h}(A_\beta(x_{k+1}) - A_\beta(x_k)) + A(x_{k+1}) = 0. \quad (4.1)$$

After expanding (4.1), we obtain

$$x_{k+1} + \frac{h^2}{1 + \gamma h} A(x_{k+1}) + \frac{h}{1 + \gamma h} A_\beta(x_{k+1}) = x_k + \frac{1}{1 + \gamma h}(x_k - x_{k-1}) + \frac{h}{1 + \gamma h} A_\beta(x_k). \quad (4.2)$$

Set $s := \frac{h}{1 + \gamma h}$ and $\alpha := \frac{1}{1 + \gamma h}$. So we have

$$x_{k+1} + s\mathcal{A}_h(x_{k+1}) = y_k, \quad (4.3)$$

where

$$\mathcal{A}_h = (h + \beta_b)B + (h + \beta_f)\nabla f, \quad (4.4)$$

$$y_k = x_k + \alpha(x_k - x_{k-1}) + sA_\beta(x_k). \quad (4.5)$$

From (4.3) we get

$$x_{k+1} = (\text{Id} + s\mathcal{A}_h)^{-1}(y_k). \quad (4.6)$$

Combining (4.3) and (4.6), we obtain the following algorithm

(DINAAM):

Initialize: $x_0 \in \mathcal{H}, x_1 \in \mathcal{H}$

$$\alpha = \frac{1}{1 + \gamma h},$$

$$s = \frac{h}{1 + \gamma h},$$

$$y_k = x_k + \alpha(x_k - x_{k-1}) + sA_\beta(x_k),$$

$$x_{k+1} = (\text{Id} + s\mathcal{A}_h)^{-1}(y_k).$$

(4.7)

Remark 4.1 (i) The convergence analysis of the algorithm (DINAAM) will be postponed to an other research investigation. In the current version, we focus only on the continuous dynamic (DINAM) and its asymptotic convergence. The numerical experiments below are given for illustrative purposes.

(ii) A general method to generate monotone cocoercive operators which are not gradients of convex functions is to start from a linear skew symmetric operator A and then take its Yosida approximation A_λ . As a model situation, take $\mathcal{H} = \mathbb{R}^2$ and start from A equal to the rotation of angle $\frac{\pi}{2}$. We have

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

An elementary computation gives that, for any $\lambda > 0$

$$A_\lambda = \frac{1}{1 + \lambda^2} \begin{pmatrix} \lambda & -1 \\ 1 & \lambda \end{pmatrix},$$

which is therefore λ -cocoercive. As a consequence, for $\lambda = 1$ we obtain that the matrix

$$B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

is $\frac{1}{2}$ -cocoercive. With these basic blocks, one can easily construct many other cocoercive operators which are not potential operators. For that, use Lemma 7.1 which gives that the sum of two cocoercive operators is still cocoercive, and therefore the set of cocoercive operators is a convex cone.

Example 4.1 Let us start this section by a simple illustrative example in \mathbb{R}^2 . We take $\mathcal{H} = \mathbb{R}^2$ equipped with the usual Euclidean structure. Let us consider B as a linear operator whose matrix in the canonical basis of \mathbb{R}^2 is defined by $B = A_\lambda$ for $\lambda = 5$. According to Remark 4.1, we can check that B is λ -cocoercive with $\lambda = 5$ and that B is a nonpotential operator. To observe the classical oscillations, in the heavy ball with friction, we take $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x_1, x_2) = 50x_2^2.$$

We set $\gamma = 0.9$. It is clear that f is convex but not strongly convex. We study 3 cases: (1) $\beta_b = 1, \beta_f = 0.5$, (2) $\beta_b = 0.5, \beta_f = 1$ and (3) $\beta_b = \beta_f = 0.5$. As a straight application of Theorem 3.1, we obtain that the trajectory $x(t)$ generated by (DINAM) converges to x_∞ , where $x_\infty \in S = (B + \nabla f)^{-1}(0) = \{0\}$. The trajectory obtained by using Matlab is depicted in Figure 1, where we represent the components $x_1(t)$ and $x_2(t)$ in red and blue respectively.

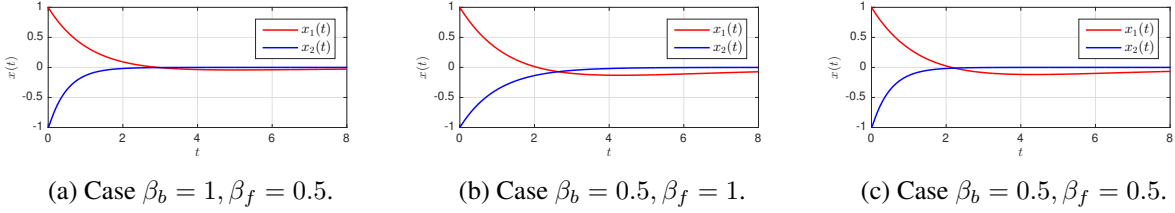


Figure 1: Trajectories of (DINAM) for different values of the parameters β_b, β_f .

Now we study the behaviour of the trajectories by considering more different values of β_b and β_f . We study four cases in Figures 2. The plots of the second variable of the solutions have been depicted in Figure 2(a) while in Figure 2(b) is plotted the number of iterations k versus $\|B(x_k) + \nabla f(x_k)\|$. Through Figures 1 and 2, we can conclude that by introducing the Hessian damping ($\beta_f > 0$), the oscillations of the trajectories in Fig. 2 are attenuated. The oscillations of the solutions appear whenever β_f goes to 0.

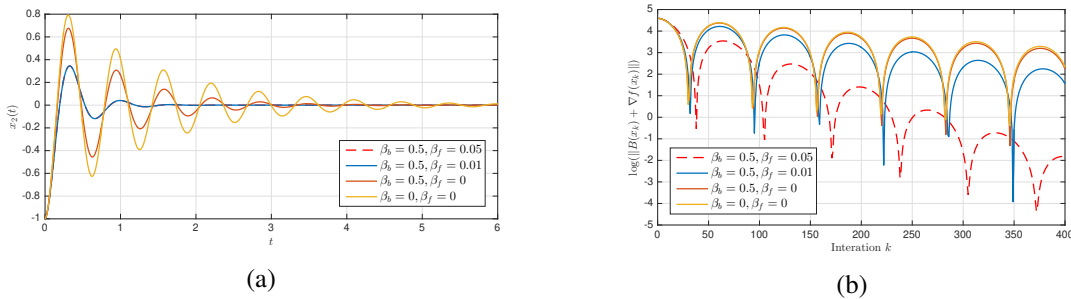


Figure 2: Oscillation of the trajectories of (DINAM) for different values of β_b, β_f .

Example 4.2 Now we are looking at another higher dimensional example. Let us consider $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by $f(x) = \frac{1}{2}\|Mx - b\|^2$, where $M \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. We have

$$\nabla f(x) = M^\top(Mx - b), \quad \nabla^2 f(x) = M^\top M.$$

Since $M^\top M$ is positive semidefinite for any matrix M , the quadratic function f is convex. Furthermore, if M has full column rank, *i.e.* $\text{rank}(M) = n$, then $M^\top M$ is positive definite. Therefore f is strongly convex. Take B defined as below

$$B = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \vdots \\ 0 & 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Then, B is cocoercive. Indeed, for any $x, y \in \mathbb{R}^n$, we have

$$\begin{aligned} \langle Bx - By, x - y \rangle &= \|x_1 - y_1\|^2 + \|x_2 - y_2\|^2 + \cdots + \|x_n - y_n\|^2 \\ &\geq \frac{1}{2} [2(\|x_1 - y_1\|^2 + \|x_2 - y_2\|^2) + \|x_3 - y_3\|^2 + \cdots + \|x_n - y_n\|^2] \\ &= \frac{1}{2} \|Bx - By\|^2. \end{aligned}$$

If the matrix M has not full column rank with $M^\top M + B$ nonsingular, then

$$B(x) + \nabla f(x) = 0 \text{ if and only if } x = (M^\top M + B)^{-1} M^\top b.$$

In our experiment, we pick M a random 10×100 matrix which has not full column rank. Set $\gamma = 3, \beta_b = 1, \beta_f = 1$ and the operator B as presented above. Thanks to Corollary 3.1, we conclude that the trajectory $x(t)$ generated by the system (DINAM) converges to $x_\infty = (M^\top M + B)^{-1} M^\top b$. Implementing the algorithm (DINAAM) in Matlab, we obtain the plot of k versus the norm of $B(x_k) + \nabla f(x_k)$. Similarly, we study several cases by changing the parameters β_b, β_f . This is depicted in Fig 3.

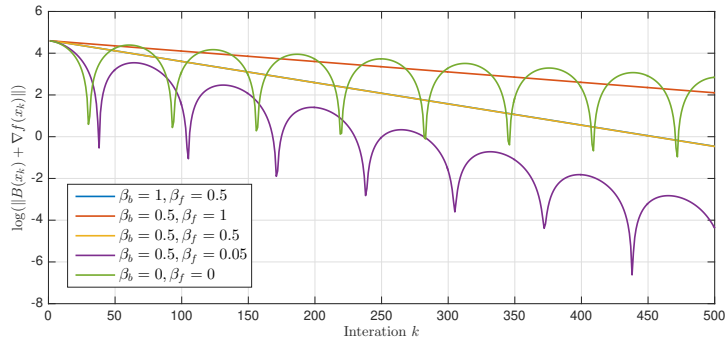


Figure 3: The behaviour of (DINAAM) for high dimension problem.

Before ending this part, we discuss an application of our model to dynamical games. The following example is taken from Attouch-Maingé [12] and adapted to our context.

Example 4.3 We make the following standing assumptions:

- (i) $\mathcal{H} = \mathcal{X}_1 \times \mathcal{X}_2$ is the Cartesian product of two Hilbert spaces equipped with norms $\|\cdot\|_{\mathcal{X}_1}$ and $\|\cdot\|_{\mathcal{X}_2}$ respectively. In which, $x = (x_1, x_2)$, with $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, stands for an element in \mathcal{H} ;
- (ii) $f : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathbb{R}$ is convex function whose gradient is Lipschitz continuous on bounded sets;
- (iii) $B = (\nabla_{x_1} \mathcal{L}, -\nabla_{x_2} \mathcal{L})$ is the maximally monotone operator which is attached to a smooth convex-concave function $\mathcal{L} : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathbb{R}$. The operator B is assumed to be λ -cocoercive with $\lambda > 0$.

In our setting, with $x(t) = (x_1(t), x_2(t))$ the system (DINAM) is written

$$\begin{cases} \ddot{x}_1(t) + \gamma \dot{x}_1(t) + \nabla_{x_1} f(x_1(t), x_2(t)) + \nabla_{x_1} \mathcal{L}(x_1(t), x_2(t)) \\ \quad + \beta_f \frac{d}{dt} (\nabla_{x_1} f(x_1(t), x_2(t))) + \beta_b \frac{d}{dt} (\nabla_{x_1} \mathcal{L}(x_1(t), x_2(t))) = 0 \\ \ddot{x}_2(t) + \gamma \dot{x}_2(t) + \nabla_{x_2} f(x_1(t), x_2(t)) - \nabla_{x_2} \mathcal{L}(x_1(t), x_2(t)) \\ \quad + \beta_f \frac{d}{dt} (\nabla_{x_2} f(x_1(t), x_2(t))) - \beta_b \frac{d}{dt} (\nabla_{x_2} \mathcal{L}(x_1(t), x_2(t))) = 0. \end{cases} \quad (4.8)$$

According to Theorem 3.1, $x(t) \rightharpoonup x_\infty = (x_{1,\infty}, x_{2,\infty})$ weakly in \mathcal{H} , where $(x_{1,\infty}, x_{2,\infty})$ is solution of

$$\begin{cases} \nabla_{x_1} f(x_1(t), x_2(t)) + \nabla_{x_1} \mathcal{L}(x_1(t), x_2(t)) = 0 \\ \nabla_{x_2} f(x_1(t), x_2(t)) - \nabla_{x_2} \mathcal{L}(x_1(t), x_2(t)) = 0. \end{cases} \quad (4.9)$$

Structured systems such as (4.9) contain both potential and nonpotential terms which are often present in decision sciences and physics. In game theory, (4.9) describes Nash equilibria of the normal form game with two players 1, 2 whose static loss functions are respectively given by

$$\begin{cases} F_1 : (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow F_1(x_1, x_2) = f(x_1, x_2) + \mathcal{L}(x_1, x_2) \\ F_2 : (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow F_2(x_1, x_2) = f(x_1, x_2) - \mathcal{L}(x_1, x_2). \end{cases} \quad (4.10)$$

$f(\cdot, \cdot)$ is their joint convex payoff, and \mathcal{L} is a convex-concave payoff with zero-sum rule. For more details, we refer the reader to [12]. As an example, take $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{R}$, and $\mathcal{L} : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by $\mathcal{L}(x) = \frac{1}{2}(x_1^2 - 2x_1x_2 - x_2^2)$. Then $B = (\nabla_{x_1} \mathcal{L}, -\nabla_{x_2} \mathcal{L}) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Pick $f(x) = \frac{1}{2}(3x_1^2 - 2x_1x_2 + x_2^2) - x_1 - 2x_2$. The Nash equilibria described in (4.9) can be solved by using (DINAAM). Take $\gamma = 3, \beta_b = 0.5, \beta_f = 0.5$ and $x_0 = (1, -1), \dot{x}_0 = (-10, 10)$ as initial conditions, then the numerical solution for (DINAM) converges to $x_\infty = (\frac{3}{4}, 1)$ which is the solution of (4.9) as well. The numerical trajectories and phase portrait of our model applied to dynamical games are depicted in Figure 4.

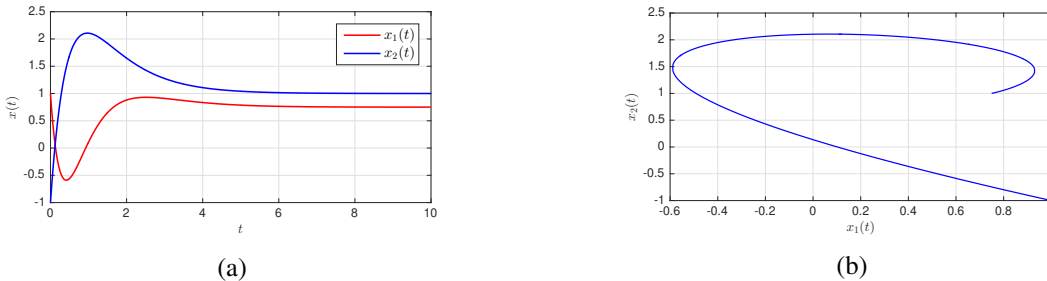


Figure 4: An application of (DINAAM) to dynamical games: trajectories (a) and phase portrait (b).

5 The nonsmooth case

The equivalence obtained in Proposition 2.1 between (DINAM) and a first-order evolution system in time and space allows a natural extension of both our theoretical and numerical results to the case of a convex, lower semicontinuous and proper function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. It suffices to replace the gradient of f by the convex subdifferential ∂f . We recall that the subdifferential of f at $x \in \mathcal{H}$ is defined by

$$\partial f(x) = \{z \in \mathcal{H} : \langle z, \xi - x \rangle \leq f(\xi) - f(x) \text{ for every } \xi \in \mathcal{H}\},$$

and the domain of f is equal to $\text{dom } f = \{x \in \mathcal{H} : f(x) < +\infty\}$. This leads to consider the system

$$\text{(g-DINAM)} \begin{cases} \dot{x}(t) + \beta_f \partial f(x(t)) + \beta_b B(x(t)) + \left(\gamma - \frac{1}{\beta_f}\right) x(t) + y(t) \ni 0; \\ \dot{y}(t) - \left(1 - \frac{\beta_b}{\beta_f}\right) B(x(t)) + \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta_f}\right) x(t) + \frac{1}{\beta_f} y(t) = 0. \end{cases}$$

The prefix g in front of (DINAM) stands for generalized. Note that the first equation of (g-DINAM) is now a differential inclusion, because of the possibility for $\partial f(x(t))$ to be multivalued. By taking $f = f_0 + \delta_C$, where δ_C is the indicator function of a constraint set C , the system (g-DINAM) allows to model damped inelastic shocks in mechanics and decision sciences, see [13]. The original aspect comes from the fact that (g-DINAM) now involves both potential driven forces (attached to f_0) and nonpotential driven forces (attached to B). As we will see, taking into account shocks created by nonpotential driving forces is a source of difficulties.

Let us first establish the existence and uniqueness of the solution trajectory of the Cauchy problem.

Theorem 5.1 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous and proper function. Suppose that $\beta_f > 0$ and $\beta_b \geq 0$. Then, for any $(x_0, x_1) \in \text{dom } f \times \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\rightarrow \mathcal{H}$ of (g-DINAM) which satisfies the Cauchy data $x(0) = x_0, \dot{x}(0) = x_1$.*

Proof. The proof is parallel to that of Theorem 2.1. The system (g-DINAM) can be equivalently written as

$$\dot{Z}(t) + \partial \Phi(Z(t)) + G(Z(t)) \ni 0, \quad Z(0) = (x_0, y_0), \quad (5.1)$$

where the function $\Phi(Z) = \Phi(x, y) = \beta_f f(x)$ is now convex lower semicontinuous proper on $\mathcal{H} \times \mathcal{H}$. The operator G is unchanged, and is globally Lipschitz continuous. The above equation falls under the setting of the Lipschitz perturbation of an evolution system governed by the subdifferential of a convex function. The existence and uniqueness of the strong solution to (5.1) follows from Brézis [23, Proposition 3.12], and the fact that $(x_0, y_0) \in \text{dom } \Phi$. ■

The solution set S is defined by

$$S := \{p \in \mathcal{H} : \partial f(p) + B(p) \ni 0\}.$$

Before stating our main result, notice that $B(p)$ is uniquely defined for $p \in S$.

Lemma 5.1 *$B(p)$ is uniquely defined for $p \in S$, i.e.*

$$p_1 \in S, p_2 \in S \implies B(p_1) = B(p_2).$$

Proof. The proof is similar to that of Lemma 3.1. It is based on the monotonicity of the subdifferential of f and the cocoercivity of the operator B . ■

For sake of simplicity, we give a detailed proof in the case $\beta_f = \beta_b = \beta > 0$. The system (g-DINAM) takes the simpler form:

$$\text{(g-DINAM)} \begin{cases} \dot{x}(t) + \beta \partial f(x(t)) + \beta B(x(t)) + \left(\gamma - \frac{1}{\beta}\right) x(t) + y(t) \ni 0; \\ \dot{y}(t) + \frac{1}{\beta_f} \left(\gamma - \frac{1}{\beta}\right) x(t) + \frac{1}{\beta} y(t) = 0. \end{cases}$$

To formulate the convergence results and the corresponding estimates, we write the first equation of (g-DINAM) as follows

$$\dot{x}(t) + \beta \xi(t) + \beta B(x(t)) + \left(\gamma - \frac{1}{\beta}\right) x(t) + y(t) = 0, \quad (5.2)$$

where $\xi(t) \in \partial f(x(t))$ and we set $A(x(t)) = \xi(t) + B(x(t))$.

Theorem 5.2 *Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a λ -cocoercive operator. Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, proper function. Suppose that $S = \{p \in \mathcal{H} : 0 \in \partial f(p) + B(p)\} \neq \emptyset$. Consider the evolution equation (g-DINAM) where the parameters satisfy the conditions: $\beta_f = \beta_b = \beta > 0$ and*

$$\gamma > 0, \beta > 0 \text{ and } \lambda\gamma > \beta + \frac{1}{\gamma}. \quad (5.3)$$

Then, for any solution trajectory $x : [0, +\infty[\rightarrow \mathcal{H}$ of (g-DINAM) the following properties are satisfied:

(i) *(integral estimates)* Set $A(x(t)) := \xi(t) + B(x(t))$, with $\xi(t) \in \partial f(x(t))$ as defined in (5.2) and $p \in S$. Then,

$$\begin{aligned} \int_0^{+\infty} \|\dot{x}(t)\|^2 dt &< +\infty, \int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt < +\infty, \\ \int_0^{+\infty} \|A(x(t))\|^2 dt &< +\infty, \int_0^{+\infty} \langle A(x(t)), x(t) - p \rangle dt < +\infty. \end{aligned}$$

(ii) *(convergence)* For any $p \in S$,

1. $\lim_{t \rightarrow +\infty} \|x(t) - p\|$ exists.
2. $\lim_{t \rightarrow +\infty} \|B(x(t)) - B(p)\| = 0$, where $B(p)$ is uniquely defined for $p \in S$.

Proof. Let us adapt the Lyapunov analysis developed in the previous sections to the case where f is nonsmooth. We have to pay attention to the following points. First, we must invoke the (generalized) chain rule for derivatives over curves (see [23, Lemme 3.3]), that is, for a.e $t \geq 0$

$$\frac{d}{dt} f(x(t)) = \langle \xi(t), \dot{x}(t) \rangle.$$

The second ingredient is the validity of the subdifferential inequality for convex functions. So, let us consider the function $t \in [0, +\infty[\mapsto \mathcal{E}_p(t) \in \mathbb{R}_+$ defined by

$$\mathcal{E}_p(t) := \frac{1}{2} \|x(t) - p + c(\dot{x}(t) + A(x(t)))\|^2 + \frac{\delta}{2} \|x(t) - p\|^2 + [c\delta\beta_f + c^2]\Gamma(t). \quad (5.4)$$

When derivating $\mathcal{E}_p(t)$, we use the formulation (g-DINAM) which gives

$$\dot{x}(t) + \beta A(x(t)) = - \left(\gamma - \frac{1}{\beta} \right) x(t) - y(t)$$

which allows to derivate $\dot{x}(t) + \beta A(x(t))$, and obtain similar formulas as in the smooth case. Then a close examination of the Lyapunov analysis shows that we can obtain the additional estimate

$$\int_0^\infty \langle A(x(t)), x(t) - p \rangle dt < +\infty. \quad (5.5)$$

Set $0 \in \partial f(p) + B(p)$. To obtain (5.5), we return to (3.5), and consider the following minorization that we split into a sum with coefficients ϵ' and $1 - \epsilon'$ (where $\epsilon' > 0$ will be taken small enough)

$$\begin{aligned} c \langle A(x(t)), x(t) - p \rangle &= c\epsilon' \langle A(x(t)), x(t) - p \rangle + c(1 - \epsilon') \langle A(x(t)) - Ap, x(t) - p \rangle \\ &\geq c\epsilon' \langle A(x(t)), x(t) - p \rangle + c(1 - \epsilon') \langle B(x(t)) - B(p), x(t) - p \rangle \\ &\geq c\epsilon' \langle A(x(t)), x(t) - p \rangle + c(1 - \epsilon') \lambda \|B(x(t)) - B(p)\|^2. \end{aligned} \quad (5.6)$$

Note that in the second above inequality we have used the monotonicity of ∂f . So the proof continues with λ replaced by $(1 - \epsilon')\lambda$. This does not change the conditions on the parameters since in our assumptions, the inequality $\lambda\gamma > \beta + \frac{1}{\gamma}$ is strict, it is still satisfied by $(1 - \epsilon')\lambda$ when ϵ' is taken small enough. So, after integrating the resulting strict Lyapunov inequality, we obtain the supplementary property (5.5). Until (3.22) the proof is essentially the same as in the case of a smooth function f . We obtain the integral estimates

$$\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty, \int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt < +\infty, \int_0^{+\infty} \|A(x(t))\|^2 dt < +\infty.$$

But then we can no longer invoke the Lipschitz continuity on the bounded sets of ∇f . To overcome this difficulty, we modify the end of the proof as follows. Recall that given $p \in S$, the anchor function is defined by, for every $t \in [0, +\infty[$

$$q_p(t) := \frac{1}{2} \|x(t) - p\|^2,$$

and that we need to prove that the limit of the anchor functions exists, as $t \rightarrow +\infty$. The idea is to play on the fact that we have in hand a whole collection of Lyapunov functions, parametrized by the coefficient c . Recall that we have obtained that the limit of $\mathcal{E}_p(t)$ exists as $t \rightarrow +\infty$, and this is satisfied for a whole interval of values of c . So, for such c , the limit of $W_c(t) := \frac{1}{c\delta\beta + c^2} \mathcal{E}_p(t)$ as $t \rightarrow +\infty$ exists, where

$$W_c(t) = \frac{1}{2(c\delta\beta + c^2)} \|x(t) - p + c(\dot{x}(t) + \beta A(x(t)))\|^2 + \frac{\delta}{2(c\delta\beta + c^2)} \|x(t) - p\|^2 + \Gamma(t).$$

Take two such values of c , let c_1 and c_2 , and make the difference (recall that $\delta = c\gamma - 1$). We obtain

$$W_{c_1}(t) - W_{c_2}(t) = \left[\frac{1}{(c_1\gamma - 1)\beta + c_1} - \frac{1}{(c_2\gamma - 1)\beta + c_2} \right] W(t)$$

where

$$W(t) := \frac{\gamma}{2} \|x(t) - p\|^2 + \frac{1}{2} \|\dot{x}(t) + \beta A(x(t))\|^2 + \langle \dot{x}(t) + \beta A(x(t)), x(t) - p \rangle.$$

So, we obtain the existence of the limit as $t \rightarrow +\infty$ of $W(t)$. Then note that $W(t) = \gamma q_p(t) + \frac{d}{dt} w(t)$

where

$$w(t) := q_p(t) + \beta \int_0^t \langle A(x(s)), x(s) - p \rangle ds + \frac{1}{2} \int_0^t \|\dot{x}(s) + \beta A(x(s))\|^2 ds.$$

Reformulate $W(t)$ in terms of $w(t)$ as follows

$$W(t) = \gamma w(t) + \frac{d}{dt}w(t) - \left(\gamma\beta \int_0^t \langle A(x(s), x(s) - p) \rangle ds + \frac{\gamma}{2} \int_0^t \|\dot{x}(s) + A(x(s))\|^2 ds \right).$$

As a consequence of (5.5) and of the previous estimates, we have that the limit of the two above integrals exists as $t \rightarrow +\infty$. Therefore, according to the convergence of $W(t)$ we obtain that

$$\lim_{t \rightarrow +\infty} \left(\gamma w(t) + \frac{d}{dt}w(t) \right) \text{ exists.}$$

The existence of the limit of w follows from a classical general result concerning the convergence of evolution equations governed by strongly monotone operators (here γId , see Theorem 3.9 page 88 in [23]). In turn, using the same argument as above, we obtain that for all $p \in S$

$$\lim_{t \rightarrow +\infty} \|x(t) - p\| \text{ exists.}$$

As in the smooth case, the strong convergence of $B(x(t))$ to $B(p)$ is a direct consequence of the integral estimates $\int_0^{+\infty} \|B(x(t)) - B(p)\|^2 dt < +\infty$, $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$ and of the fact that B is Lipschitz continuous. The proof of Theorem 3.1 is thereby completed. ■

Remark 5.1 (i) A natural question is to know if the weak limit of the trajectory exists. Indeed we are not far from this result since $\int_0^{+\infty} \|A(x(t))\|^2 dt < +\infty$, which implies that $A(x(t))$ converges strongly to zero in an "essential" way. According to Opial's lemma, this allows to complete the convergence proof as in the smooth case. This is a seemingly difficult question to examine in the future.

(ii) A particular situation is the case $\gamma = \frac{1}{\beta}$, in which case the system (g-DINAM) can be written in an equivalent way

$$\dot{u}(t) + \gamma u(t) = 0$$

where

$$\dot{x}(t) + \beta A(x(t)) \ni u(t).$$

The convergence of the trajectory $t \mapsto x(t)$ is then a consequence of the convergence of the semigroup generated by the sum of a cocoercive operator with the subdifferential of a convex lower semicontinuous and proper function, see Abbas-Attouch [1]. Note that in this case the condition for the convergence of the trajectories generated by (g-DINAM) does not depend any more on the cocoercivity parameter λ .

6 Conclusion, perspectives

In this paper, in a general real Hilbert space setting, we investigated a dynamic inertial Newton method for solving additively structured monotone problems. The dynamic is driven by the sum of two monotone operators with distinct properties: the potential component is the gradient of a continuously differentiable convex function f , and the nonpotential is a monotone and cocoercive operator B . The geometric damping is controlled by the Hessian of the potential f and by a Newton-type correction term attached to B . The well-posedness of the Cauchy problem is shown as well as the asymptotic convergence properties of the trajectories generated by the continuous dynamic. The convergence analysis is carried out through the parameters β_f and β_b attached to the geometric dampings as well as the parameters γ and λ (the viscous damping and the coefficient of cocoercivity respectively). The introduction of geometric damping makes it

possible to control and attenuate the oscillations known for viscous damping of inertial systems, giving rise to faster numerical methods. It would be interesting to extend the analysis for both the continuous dynamic and its discretization to the case of an asymptotic vanishing damping $\gamma(t) = \frac{\alpha}{t}$, with $\alpha > 0$ as in [31]. This is a decisive step towards proposing faster algorithms for solving structured monotone inclusions, which are connected to the accelerated gradient method of Nesterov. The study of the corresponding splitting methods is also an important topic which needs further investigations. In fact, replacing ∇f by a general maximally monotone operator A , whose resolvent can be computed easily, would be an interesting problem to study an inertial forward-backward algorithm with Hessian-driven damping for solving structured monotone inclusions of the form: $Ax + Bx \ni 0$. This is beyond the scope of the current manuscript and will be the subject of a future separate work.

7 Appendix

7.1 Technical lemmas

Let us show that the sum of two cocercive operators is still cocoercive. For further properties concerning cocoercive operators see [20].

Lemma 7.1 *Let $T_1, T_2 : \mathcal{H} \rightarrow \mathcal{H}$ be two cocoercive operators with respective cocoercivity coefficients $\lambda_1, \lambda_2 > 0$. Then, $T := T_1 + T_2 : \mathcal{H} \rightarrow \mathcal{H}$ is λ -cocoercive with $\lambda = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$.*

Proof. According to the cocoercivity assumptions of T_1 and T_2 , we have

$$\begin{aligned} \langle T_1 y - T_1 x, y - x \rangle &\geq \lambda_1 \|T_1 y - T_1 x\|^2, \quad \forall x, y \in \mathcal{H}, \\ \langle T_2 y - T_2 x, y - x \rangle &\geq \lambda_2 \|T_2 y - T_2 x\|^2, \quad \forall x, y \in \mathcal{H}. \end{aligned}$$

Let us show that the sum $T = T_1 + T_2$ is still cocoercive. Using elementary computation in Hilbert spaces, for all $x, y \in \mathcal{H}$, we have

$$\begin{aligned} \|Ty - Tx\|^2 &= \|T_1 y - T_1 x + T_2 y - T_2 x\|^2 \\ &= \|T_1 y - T_1 x\|^2 + \|T_2 y - T_2 x\|^2 + 2\langle T_1 y - T_1 x, T_2 y - T_2 x \rangle \\ &\leq \|T_1 y - T_1 x\|^2 + \|T_2 y - T_2 x\|^2 + \frac{\lambda_1}{\lambda_2} \|T_1 y - T_1 x\|^2 + \frac{\lambda_2}{\lambda_1} \|T_2 y - T_2 x\|^2 \\ &= (\lambda_1^{-1} + \lambda_2^{-1}) (\lambda_1 \|T_1 y - T_1 x\|^2 + \lambda_2 \|T_2 y - T_2 x\|^2). \end{aligned}$$

Since T_1, T_2 are cocoercive, we deduce that

$$\begin{aligned} \|Ty - Tx\|^2 &\leq (\lambda_1^{-1} + \lambda_2^{-1}) (\langle T_1 y - T_1 x, y - x \rangle + \langle T_2 y - T_2 x, y - x \rangle) \\ &= (\lambda_1^{-1} + \lambda_2^{-1}) \langle Ty - Tx, y - x \rangle. \end{aligned}$$

Equivalently,

$$\langle Ty - Tx, y - x \rangle \geq \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \|Ty - Tx\|^2, \quad \forall x, y \in \mathcal{H}.$$

So, T is still λ -cocoercive with $\lambda = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} > 0$.

Let us show that this estimate is sharp. Take $T_1 : \mathcal{H} \rightarrow \mathcal{H}, x \mapsto \lambda_1^{-1}x$ and $T_2 : \mathcal{H} \rightarrow \mathcal{H}, x \mapsto \lambda_2^{-1}x$. It is easy to check that T_1, T_2 are two cocoercive operators with cocoercivity coefficients λ_1, λ_2 respectively. Then their sum operator is equal to $Tx = (\lambda_1^{-1} + \lambda_2^{-1})x = \lambda^{-1}x$ with $\lambda = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$, and hence is λ cocoercive. This shows that we cannot obtain a better estimate. ■

The next lemma is a classical result in integration theory.

Lemma 7.2 *Let $1 \leq p < \infty$ and $1 \leq r \leq \infty$. Suppose $u \in L^p([0, \infty[; \mathbb{R})$ is a locally absolutely continuous nonnegative function, $g \in L^r([0, \infty[; \mathbb{R})$ and*

$$\dot{u}(t) \leq g(t)$$

for almost every $t > 0$. Then $\lim_{t \rightarrow \infty} u(t) = 0$.

In the proof of Theorem 3.1, we use the following elementary result concerning positive quadratic forms.

Lemma 7.3 *Let a, b, c be three real numbers. The quadratic form $q : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$*

$$q(X, Y) := a\|X\|^2 + 2b\langle X, Y \rangle + c\|Y\|^2$$

is positive definite if and only if $ac - b^2 > 0$ and $a > 0$. Moreover

$$q(X, Y) \geq \mu(\|X\|^2 + \|Y\|^2) \quad \text{for all } X, Y \in \mathcal{H}$$

where the positive real number $\mu := \frac{1}{2} \left(a + c - \sqrt{(a - c)^2 + 4b^2} \right)$ is the smallest eigenvalue of the positive symmetric matrix associated with q .

References

- [1] B. 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] 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)
- [3] Adly, S., Attouch, H.: Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping, (2020), *SIAM J. Optim.*, hal-02423584
- [4] Alecsa, C.D., László, S., Pinta, T.: An extension of the second order dynamical system that models Nesterov’s convex gradient method, *Applied Mathematics and Optimization*, (2020), arXiv:1908.02574v1.
- [5] 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)
- [6] 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)
- [7] Attouch, H., Cabot, A.: Convergence of a relaxed inertial proximal algorithm for maximally monotone operators, *Math. Program.* **184**, 243–287 (2020).
- [8] Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order algorithms via inertial systems with Hessian driven damping. *Mathematical Programming*, (2020) <https://doi.org/10.1007/s10107-020-01591-1>, preprint available at hal-02193846.
- [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., László, S.C.: Continuous Newton-like Inertial Dynamics for Monotone Inclusions, Set-Valued and Variational Analysis, 2020, <https://doi.org/10.1007/s11228-020-00564-y>
- [11] Attouch, H., László, S.C.: Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. SIAM J. Optim. (2020), hal-02549730
- [12] Attouch, H., Maingé, P.E.: Asymptotic behavior of second order dissipative evolution equations combining potential with nonpotential effects. ESAIM Control Optim. Calc. of Var. **17**(3), 836–857 (2011)
- [13] Attouch, H., Maingé, P.E., Redont, P.: A second-order differential system with Hessian-driven damping; Application to nonelastic shock laws. Differential Equations and Applications **4**(1), 27–65 (2012)
- [14] Attouch, H., Marques Alves, M., Svaiter, B.F.: A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert Spaces, with complexity $\mathcal{O}(1/n^2)$. Journal of Convex Analysis **23**(1), 139–180 (2016)
- [15] Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators. Mathematical Programming **174**(1-2), 391–432 (2019)
- [16] Attouch, H., Peypouquet, J., Redont, P.: Fast convex minimization via inertial dynamics with Hessian driven damping. J. Differential Equations **261**(10), 5734–5783 (2016)
- [17] 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)
- [18] Attouch, H., Svaiter, B.F.: A continuous dynamical Newton-Like approach to solving monotone inclusions. SIAM J. Control Optim. **49**(2), 574–598 (2011)
- [19] J.-B. Baillon and G. Haddad, *Quelques propriétés des opérateurs angles-bornés et n-cycliquement monotones*. Israel J. Math. 26 (1977) 137-150.
- [20] Bauschke, H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert spaces. CMS Books in Mathematics. Springer (2011)
- [21] Boj, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control Optim. **54**, 1423–1443 (2016)
- [22] Boj, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian damping. Mathematical Programming, (2020), <https://doi.org/10.1007/s10107-020-01528-8>
- [23] Brézis, H.: Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution. Lecture Notes 5. North Holland (1972)
- [24] Brézis, H., Analyse fonctionnelle, Collection Mathématiques Appliquées pour le Maîtrise, Masson, (1983).
- [25] Castera, C., Bolte, J., Févotte, C., Pauwels, E.: An Inertial Newton Algorithm for Deep Learning. (2019), HAL-02140748

- [26] Constantin P. Niculescu, Lars-Erik Persson: *Convex functions and their applications: A contemporary approach*. Springer (2014).
- [27] Kim, D.: Accelerated Proximal Point Method for Maximally Monotone Operators. (2020), preprint available at [arXiv:1905.05149v3](https://arxiv.org/abs/1905.05149v3)
- [28] Lin, T., Jordan, M.I.: A Control-Theoretic Perspective on Optimal High-Order Optimization. (2019), [arXiv:1912.07168v1](https://arxiv.org/abs/1912.07168v1)
- [29] Peypouquet, J., Sorin, S., : Evolution Equations for Maximal Monotone Operators: Asymptotic Analysis in Continuous and Discrete Time *Journal of Convex Analysis*, 17(3-4) (2010), pp. 1113–1163.
- [30] Shi, B., Du, S.S., Jordan, M.I., Su, W.J.: Understanding the acceleration phenomenon via high-resolution differential equations. (2018), [arXiv:submit/2440124](https://arxiv.org/abs/submit/2440124)[cs.LG]
- [31] W. Su, S. Boyd, E. J. Candès, A differential equation for modeling Nesterov’s accelerated gradient method, *Journal of Machine Learning Research*, 17 (2016), pp. 1–43.