Convergence of inertial dynamics driven by sums of potential and nonpotential operators and with implicit Newton-like damping

Samir Adly, Hedy Attouch, Van Nam Vo

Convergence of inertial dynamics driven by sums of potential and nonpotential operators and with implicit Newton-like damping

Samir Adly* Hedy Attouch† Van Nam Vo‡

Abstract

We propose and study the convergence properties of the trajectories generated by a damped inertial dynamic which is driven by the sum of potential and nonpotential operators. Precisely, we seek to reach asymptotically the zeros of sums of potential term (the gradient of a continuously differentiable convex function) and nonpotential monotone and cocoercive operator. As an original feature, in addition to viscous friction, the dynamic involves implicit Newton-type damping. This contrasts with the authors' previous study where explicit Newton-type damping was considered, which, for the potential term, corresponds to Hessian-driven damping. We show the weak convergence, as time goes to infinity, of the generated trajectories towards the zeros of the sum of the potential and nonpotential operators. Our results are based on Lyapunov analysis and appropriate setting of damping parameters. The introduction of geometric dampings allows to control and attenuate the oscillations known for the viscous damping of inertial methods. Rewriting the second-order evolution equation as a system involving only first order derivative in time and space allows us to extend the convergence analysis to nonsmooth convex potentials. Our study concerns the autonomous case with positive fixed parameters. These results open the door to their extension to the nonautonomous case and 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 due to the presence of the nonpotential term.

1 Introduction

With the explosion of digital information, the Gradient Descent Method (GDM) is one of the most popular methods used in data science, image and statistical processing to minimize a function, due to its simplicity. First-order methods have gained popularity in recent years due to their importance in solving large scale optimization problems in Machine Learning and Data Science by only having access to the gradient of the function. One of the drawbacks of the Gradient Descent Method is its slowness (zig-zag pattern convergence on quadratic functions). An improvement of the Gradient Descent Method was proposed in 1964 by B. Polyak [38] where he considered a momentum term associated with a gradient descent step. The associated continuous Ordinary Differential Equation (ODE) surrogate of the Polyak momentum is known as the heavy ball with friction (HBF), an inertial system with a fixed viscous damping coefficient. From a mechanical point of view, it could be interpreted as the motion of a material point subject to viscous friction damping and conservative potential forces. The (HBF) is a second order (in time) dissipative system where the presence of inertia allows the system to overcome some known drawbacks of the (GDM) and acts to accelerate the convergence. We note that the (HBF) is not a descent method and the convergence of the trajectories towards a critical point of the potential to be minimized is well-known under various

*Laboratoire XLIM, Université de Limoges, 123 avenue A. Thomas, 87060 Limoges, France. E-mail: samir.adly@unilim.fr
†IMAG, CNRS, Université Montpellier, France. E-mail: hedy.attouch@umontpellier.fr
‡Laboratoire XLIM, Université de Limoges, 123 av. A. Thomas, 87060 Limoges, France. E-mail: van-nam.vo@etu.unilim.fr

assumptions like convexity or analyticity of the potential to be minimized. For a strongly convex function and a viscous damping coefficient judiciously chosen, (HBF) provides convergence at exponential rate. For a general convex function, 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. An other momentum method was introduced by Nesterov [39] in 1983, known in the literature as Nesterov Accelerated Gradient (NAG). To obtain a continuous ODE surrogate of the Nesterov Accelerated Gradient algorithm, a decisive step was taken by Su-Boyd-Candès [44] with the introduction of an Asymptotic Vanishing Damping (AVD) coefficient of the form $\frac{\alpha}{t}$, with $\alpha > 0$ and $t > 0$ represents the time variable. In particular, for a general convex function f, the condition $\alpha > 3$ guarantees the asymptotic convergence rate of the values with a rate of ordre $o\left(1/t^2\right)$, as well as the weak convergence of the trajectories towards optimal solutions. The subcritical case $\alpha \leq 3$ has been examined in [7] and [12]. In line with the founding article by Beck and Teboulle [28] devoted to Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) method, an abundant literature has been devoted to the extension of these results to the inertial proximal gradient algorithms for solving additively structured "smooth + nonsmooth" optimization problems by splitting methods, see [9], [11], [33], [42] and references therein. The introduction of the Hessian-driven damping in [10] allows to damp the transversal oscillations which can occur with (HBF). Recent studies have been devoted to inertial dynamics that combines asymptotic vanishing damping with Hessian-driven damping. In fact, the corresponding algorithms involve a correcting term in the Nesterov method which reduces the oscillatory aspects [23, 13, 44].
The Ravine method was introduced by Gelfand and Tsetlin [34] in 1961. It is closely related to the Nesterov method. It mimics the flow of water in the mountains which first flows rapidly downhill through small, steep ravines and then flows along the main river into the valley. It was put forward by B. Polyak and more recently in [17] and [41]. It has been shown in [17] that the Ravine and the Nesterov methods have the same dynamic interpretation and they benefit from similar fast convergence properties. In fact, the low resolution ODE (in the sense of [43]) of both Nesterov Accelerated Gradient and the Ravine method is given by the Su-Boyd-Candès dynamic. The high-resolution ODE of Nesterov's and Ravine's accelerated gradient methods shows the Hessian-driven damping, giving a more accurate dynamic interpretation of both methods. The explicit form of the Hessian-driven damping was introduced in [13] and [43], while the implicit form was considered by Alesca, Laszlo and Pinta in [5].
Equally important is the study of additively structured monotone problems involving the sum of potential and nonpotential operators. Indeed, many situations coming from physics, biology, decision sciences involve equations containing both potential and nonpotential terms. For example, in decision sciences and game theory, it comes from the presence of both cooperative and noncooperative aspects. In physics, this is the case when the phenomena of diffusion and convection both occur. The Lagrangian approach to linear constrained optimization problems also gives rise to similar structures. Our main concern in this paper is the analysis of the convergence properties of the trajectories generated by a damped inertial dynamic, called (iDINAM), driven by the sum of a potential (the gradient of a continuously differentiable convex function) and a nonpotential monotone operators. The originality of this model lies in the fact that it contains an implicit Newton-type damping in addition to the viscous friction. Our approach is based on the Lyapunov analysis combined with an adequate tuning of the parameters involved in the dynamic. We note that the explicit Newton-type damping was considered by the authors in [3, 4]. Our main results are Theorems 4.1 and 5.1 which show that a judicious adjustment of the damping parameters ensures the weak convergence of the trajectories generated by (iDINAM) and the associated proximal-gradient algorithms, obtained by temporal discretization.

The content of the paper is as follows. After the introductory Section 1, in Section 3, we show the well-posedness of the Cauchy problem for (iDINAM). In Section 4, we analyze the convergence properties of the solution trajectories generated by the continuous dynamics (iDINAM). We highlight the interplay between the damping parameters β_f, β_b, γ and the cocoercivity parameter λ, which plays a significant role in our Lyapunov analysis. In Section 5, we analyze various inertial proximal-gradient splitting algorithms which come naturally from the temporal discretization of (iDINAM). We also examine the effect of errors, perturbations in these algorithms. In Section 6, we perform numerical experiments which show that the oscillations are considerably reduced with the introduction of geometric damping. Applications to structured monotone equations involving a nonpotential operator are considered.

2 Problem statement and related works

2.1 General presentation

Let $\mathcal{H}$ be a real Hilbert space endowed with the scalar product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$. Our study focuses on the dynamic approach to solving the additively structured monotone problem

$$\text{Find } x \in \mathcal{H} \ : \ \nabla f(x) + B(x) = 0, \tag{2.1}$$

where ∇f is the gradient of a continuously differentiable convex function $f : \mathcal{H} \to \mathbb{R}$ (this is the potential part), and $B : \mathcal{H} \to \mathcal{H}$ is an operator which is supposed to be monotone and cocoercive (this is the nonpotential part). Specifically, our study concerns the convergence properties when $t \to +\infty$ of the trajectories generated by the second-order evolution equation

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f\Big(x(t) + \beta_f \dot{x}(t)\Big) + B\Big(x(t) + \beta_b \dot{x}(t)\Big) = 0, \tag{iDINAM}$$

whose stationary points are solutions of (2.1). We will see that the nonnegative coefficients β_f and β_b in (iDINAM) can be interpreted as geometric damping parameters. The terminology (iDINAM) in short stands for implicit Dynamic Inertial Newton method for Additively structured Monotone problems. In addition to the modeling aspects described above, this system is part of the rich family of inertial systems that have been considered in recent years to design fast first-order optimization algorithms. In the potential case (*i.e.*$B = 0$) this system study was considered by Alesca, Laszlo and Pinta in [5], see also [37] for a related autonomous system in the case of a strongly convex function f. The dynamic (iDINAM) is closely related to its explicit version

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

previously studied by the authors in [3]. (DINAM) is an autonomous dynamic which involves geometric dampings which are respectively controlled by the Hessian of the potential function f, and by a Newton-type correction term attached to B. The link between the two dynamics above, and the justification of their respective explicit and implicit qualification is explained by the following. When $t \to +\infty$ we have $\dot{x}(t) \to 0$. Therefore, using the Taylor expansion, we get, when $t \to +\infty$

$$\begin{aligned} \nabla f\left(x(t) + \beta_f \dot{x}(t)\right) &\approx \nabla f(x(t)) + \beta_f \nabla^2 f(x(t))\dot{x}(t), \\ B\left(x(t) + \beta_b \dot{x}(t)\right) &\approx B(x(t)) + \beta_b B'(x(t))(\dot{x}(t)). \end{aligned}$$

The replacement of these terms in (iDINAM) by their equivalent expressions gives (DINAM). Therefore, both systems can be expected to behave similarly when $t \to +\infty$. It is our main objective in this paper to study the new system (iDINAM) and to compare it to (DINAM). In the potential case, (*i.e.*$B = 0$), such a comparative study was carried out in [16] from the point of view of the stability of the dynamics with respect to disturbances, errors.

Our main motivation for the study of these dynamical systems comes from the fact that the geometric damping makes it possible to control and attenuate the oscillations known for the viscous damping of

the inertial methods. This is crucial to develop corresponding fast optimization algorithms obtained by temporal discretization.
Throughout the paper we make the following standing assumptions: [1]

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

The cocoercivity assumption on the operator B plays a central role in our analysis. Recall that the operator $B : \mathcal{H} \to \mathcal{H}$ is said to be λ-cocoercive for some $\lambda > 0$ if

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

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 [26], ∇f is L-Lipschitz continuous implies that ∇f is a $1/L$-cocoercive operator (see [27, Corollary 18.16] for more details).

2.2 Related works

Some of the material presented in this section, which refers to the existing literature on the subject, is taken from the authors' previous articles [3, 4]. We reproduce it for the convenience of the reader.

2.2.1 Potential case

Let us first recall some classical results concerning the potential case ($B = 0$). The following inertial system with Hessian-driven damping

$$\ddot{x}(t) + \gamma \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0, \tag{2.2}$$

was considered by Alvarez-Attouch-Peypouquet-Redont in [10]. Then, according to the continuous interpretation by Su-Boyd-Candès [44] of the accelerated gradient method of Nesterov, Attouch-Peypouquet-Redont [23] 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 equation 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

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

As a specific feature, and by comparison with the 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 [13, 14], Boţ-Csetnek-László [30], Kim [35], Lin-Jordan

[1] At several places the assumption (A1) will be relaxed, just assuming ∇f to be Lipschitz continuous on the bounded sets

[36], Shi-Du-Jordan-Su [43]. Application to deep learning has been recently developed by Castera-Bolte-Févotte-Pauwels [32]. In [2], Adly-Attouch studied the finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping.

In (2.2), the Hessian appears explicitly. A closely related ODE is obtained by considering an approach where the Hessian driven damping appears in an implicit form. This was initiated by Alesca-Lazlo-Pinta in [5], see also [37] for a related autonomous system in the case of a strongly convex function f. This ODE, coined (ISIHD) for Inertial System with Implicit Hessian Damping, takes the form

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\Big(x(t) + \beta(t)\dot{x}(t)\Big) = 0, \qquad \text{(ISIHD)}$$

where $\alpha \geq 3$ and $\beta(t) = \gamma + \frac{\beta}{t}$, $\gamma,\ \beta \geq 0$. As mentioned above, the rationale justifying the use of the term "implicit" comes from the observation that by a Taylor expansion (as $t \to +\infty$ we have $\dot{x}(t) \to 0$ which justifies using Taylor expansion), one has

$$\nabla f\Big(x(t) + \beta(t)\dot{x}(t)\Big) \approx \nabla f(x(t)) + \beta(t)\nabla^2 f(x(t))\dot{x}(t),$$

hence making the Hessian damping appear indirectly in (ISIHD). As for (2.2), this ODE was found to have a smoothing effect on the oscillations.

2.2.2 Non potential case

Let us now examine how these techniques can be transposed to the case of maximally monotonic operators. The first studies carried out by Álvarez-Attouch [8] and Attouch-Maingé [20] concerned the equation

$$\ddot{x}(t) + \gamma\dot{x}(t) + A(x(t)) = 0, \qquad (2.3)$$

when $A : \mathcal{H} \to \mathcal{H}$ is a cocoercive (and hence maximally monotone) operator, (see also [29]). The cocoercivity assumption plays a crucial role in the study of (2.3), 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 $\lambda\gamma^2 > 1$, it is proved in [20] that each trajectory of (2.3) converges weakly to a zero of A, as $t \to +\infty$.

Then this approach has been adapted to the case of general maximally monotone operators by Attouch-Peypouquet [22], and by Attouch-Laszlo [18, 19]. 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.

The "potential + nonpotential" structured monotone case was first considered by Attouch-Maingé [20] who studied the asymptotic behavior of the second-order dissipative evolution equation

$$\ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) + B(x(t)) = 0, \qquad (2.4)$$

combining potential (f convex) with nonpotential effects (B λ-cocoercive). It was shown that the condition insuring convergence of the trajectories to equilibria is still $\lambda\gamma^2 > 1$, *i.e.* the potential term does not enter into this condition. (DINAM) is obtained by introducing the Hessian term and the corrector term of the Newton type into this dynamic.

2.2.3 Regularized Newton methods for solving monotone inclusions

As can be expected, the geometric damping related to the Hessian, has a natural link with the method of Newton for solving (2.1). To overcome the ill-posed character of the continuous Newton method for solving the equation governed by a general maximally monotone operator A, the following first order evolution system was studied by Attouch-Svaiter [25],

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

Taking $\gamma(t) > 0$, this system can be considered as a continuous version of the Levenberg-Marquardt method, which acts as a regularization of the Newton method. Under a fairly general assumption on the regularization parameter $\gamma(t)$, this system is well posed and generates trajectories that converge weakly to equilibria (zeros of A). Parallel results have been obtained for the associated proximal algorithms obtained by implicit temporal discretization, and for the corresponding forward-backward algorithms in the case of structured monotone problems, see [1], [21], [24]. Formally, this system is written as

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

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 [18, 19] who considered the general case of monotone equations. By contrast with [18, 19], according to the cocoercivity of B, we don't use the Yosida regularization, and exhibit minimal assumptions involving only the nonpotential component.

3 Well-posedness of the Cauchy problem for (iDINAM)

We are going to show an existence and uniqueness result for the Cauchy problem associated with the dynamical system (iDINAM). We will present two different approaches and results, depending on the hypothesis on the potential function f. The first, relatively simple, concerns the case where f is differentiable with ∇f globally continuous Lipschitz on $\mathcal{H}$. It is based on a direct application of the Cauchy-Lipschitz theorem to the Hamiltonian formulation of (iDINAM). The second, more complicated proof concerns the case where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semi-continuous proper function. In both cases we will use the notion of strong solution, as presented below.

Definition 3.1 *The function $x : [0, +\infty[\to \mathcal{H}$ is called a strong global solution of the dynamical system* (iDINAM) *if it satisfies the following properties:*

(i) $x, \dot{x} : [0, +\infty[\to \mathcal{H}$ *are locally absolutely continuous;*

(ii) $\ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t) + \beta_f\dot{x}(t)) + B(x(t) + \beta_b\dot{x}(t)) = 0$ *for almost every $t \geq 0$;*

Recall that a map $x : [t_0, +\infty[\to \mathcal{H}$ is said to be locally absolutely continuous if it is absolutely continuous on any compact interval $[t_0, T]$, where $T > t_0$. Moreover, we have the following equivalent characterizations of an absolutely continuous function $x : [t_0, T] \to \mathcal{H}$, (see, for example [1, 25]):

(a) there exists $y : [t_0, T] \to \mathcal{H}$ a Lebesgue-integrable function, such that

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

(b) x is a continuous and its distributional derivative is Lebesgue integrable on the interval $[0, T]$;

(c) for every $\epsilon > 0$, there exists $\eta > 0$ such that for every finite family $I_k = (a_k, b_k)$ from $[0, T]$, the following implication is valid:

$$\left[I_k \cap I_j = \emptyset \text{ and } \sum_k |b_k - a_k| < \eta \right] \Longrightarrow \left[\sum_k \|x(b_k) - x(a_k)\| < \epsilon \right].$$

3.1 Existence and uniqueness: the smooth case

Theorem 3.1 *Suppose that $f : \mathcal{H} \to \mathbb{R}$ is differentiable with ∇f globally continuous Lipschitz on $\mathcal{H}$. Suppose that $\beta_f > 0$ and $\beta_b > 0$. Then, for any $(x_0, x_1) \in \mathcal{H} \times \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\to \mathcal{H}$ of the continuous dynamic* (iDINAM) *which satisfies the Cauchy data $x(0) = x_0$, $\dot{x}(0) = x_1$.*

Proof Let us reformulate (iDINAM) as a first order evolution equation. According to its Hamiltonian formulation, the system (iDINAM) can be rewritten as

$$\begin{cases} \dot{Z}(t) = F(Z(t)) \\ Z(0) = (x_0, x_1), \end{cases} \tag{3.1}$$

where $Z(t) = (x(t), y(t))$ and $F : \mathcal{H}^2 \to \mathcal{H}^2$ is given by

$$F(x, y) = \begin{pmatrix} y \\ -\gamma y - \nabla f(x + \beta_f y) - B(x + \beta_b y) \end{pmatrix}.$$

According to the Lipschitz continuity properties of ∇f and B, it is immediate to verify that F is a Lipschitz continuous map. By applying the classical Cauchy-Lipschitz theorem, we obtain the existence and uniqueness of the solution of (3.1), and hence of the Cauchy problem for (iDINAM). Note that, without any other assumption, we obtain a strong solution, and not a classical $\mathcal{C}^2$ solution, because the vector field F is only Lipschitz continuous.

3.2 Existence and uniqueness: the nonsmooth case

Let us introduce another first order formulation of (iDINAM) which is based on the new function

$$y(t) := x(t) + \beta_f \dot{x}(t). \tag{3.2}$$

Equivalently,

$$\dot{x}(t) = \frac{1}{\beta_f}(y(t) - x(t)). \tag{3.3}$$

Elementary algebra gives

$$x(t) + \beta_b \dot{x}(t) = \frac{\beta_b}{\beta_f} y(t) + \left(1 - \frac{\beta_b}{\beta_f}\right) x(t). \tag{3.4}$$

According to the above formula, and the constitutive equation (iDINAM), the time derivation of $y(t)$ gives

$$\begin{aligned} \dot{y}(t) &= \dot{x}(t) + \beta_f \ddot{x}(t) & (3.5) \\ &= \dot{x}(t) - \beta_f \Big(\gamma \dot{x}(t) + \nabla f(y(t)) + B\Big(\frac{\beta_b}{\beta_f} y(t) + \left(1 - \frac{\beta_b}{\beta_f}\right) x(t) \Big) \Big) & (3.6) \\ &= (1 - \gamma \beta_f) \dot{x}(t) - \beta_f \nabla f(y(t)) - \beta_f B\Big(\frac{\beta_b}{\beta_f} y(t) + \left(1 - \frac{\beta_b}{\beta_f}\right) x(t) \Big). & (3.7) \end{aligned}$$

Replacing $\dot{x}(t)$ with $\frac{1}{\beta_f}(y(t) - x(t))$, as given by (3.3), gives

$$\dot{y}(t) = \frac{1 - \gamma \beta_f}{\beta_f}(y(t) - x(t)) - \beta_f \nabla f(y(t)) - \beta_f B\Big(\frac{\beta_b}{\beta_f} y(t) + \left(1 - \frac{\beta_b}{\beta_f}\right) x(t) \Big). \tag{3.8}$$

The reverse transformation which consists in passing from (3.3), (3.8) to (iDINAM) is obtained in a similar way. Let us summarize the results.

Theorem 3.2 *Let $f \in \mathcal{C}^1(\mathcal{H})$. Suppose that $\beta_f > 0$. The following statements are equivalent:*

1. $x : [0, +\infty[\to \mathcal{H}$ *is a solution trajectory of* (iDINAM) *with initial conditions* $x(0) = x_0$, $\dot{x}(0) = x_1$.
2. $(x, y) : [0, +\infty[\to \mathcal{H} \times \mathcal{H}$ *is a solution trajectory of the first-order system*

$$\begin{cases} \dot{x}(t) + \dfrac{1}{\beta_f}x(t) - \dfrac{1}{\beta_f}y(t) = 0. \\ \dot{y}(t) + \beta_f \nabla f(y(t)) + \beta_f B\Big(\dfrac{\beta_b}{\beta_f}y(t) + \Big(1 - \dfrac{\beta_b}{\beta_f}\Big)x(t)\Big) + \dfrac{1-\gamma\beta_f}{\beta_f}(x(t) - y(t)) = 0. \end{cases}$$

with initial conditions $x(0) = x_0$, $y(0) = x_0 + \beta_f x_1$.

We can naturally extend the above formulation to the case where $f \in \Gamma_0(\mathcal{H})$, by replacing the gradient ∇f with the subdifferential ∂f.

Definition 3.2 *Let $\beta_f > 0$, $f \in \Gamma_0(\mathcal{H})$. Given $(x_0, y_0) \in \mathcal{H} \times \mathrm{dom}(f)$, the Cauchy problem associated with the generalized* (iDINAM) *system is defined by*

$$\begin{cases} \dot{x}(t) + \dfrac{1}{\beta_f}x(t) - \dfrac{1}{\beta_f}y(t) = 0 \\ \dot{y}(t) + \beta_f \partial f(y(t)) + \beta_f B\Big(\dfrac{\beta_b}{\beta_f}y(t) + \Big(1 - \dfrac{\beta_b}{\beta_f}\Big)x(t)\Big) + \dfrac{1-\gamma\beta_f}{\beta_f}(x(t) - y(t)) \ni 0. \\ x(0) = x_0, y(0) = y_0. \end{cases} \tag{3.9}$$

The existence and uniqueness of a global strong solution of the Cauchy problem (3.9) is established in the following theorem.

Theorem 3.3 *Let $f \in \Gamma_0(\mathcal{H})$. Suppose that $\beta_f > 0$. Then, for any Cauchy data $(x_0, y_0) \in \mathcal{H} \times \mathrm{dom}(f)$, there exists a unique global strong solution $(x, y) : [0, +\infty[\to \mathcal{H}\times\mathcal{H}$ of the generalized* (iDINAM) *system* (3.9) *satisfying the initial condition $x(0) = x_0$, $y(0) = y_0$. Moreover when $f \in \mathcal{C}^1(\mathcal{H})$, $x(\cdot)$ is a classical (i.e. $\mathcal{C}^2$) global solution of the Cauchy problem associated with* (iDINAM).

Proof We reformulate (3.9) in the product space $\mathcal{H} \times \mathcal{H}$ by setting $Z(t) = (x(t), y(t)) \in \mathcal{H} \times \mathcal{H}$, and thus (3.9) can be equivalently written as

$$\dot{Z}(t) + \beta_f \partial \mathcal{G}(Z(t)) + \mathcal{D}(Z(t)) \ni 0, \tag{3.10}$$

where $\mathcal{G} \in \Gamma_0(\mathcal{H} \times \mathcal{H})$ is the function defined as $\mathcal{G}(Z) = f(y)$, and operator $\mathcal{D} : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is given by

$$\mathcal{D}(Z) = \Big(\frac{1}{\beta_f}(x - y), \beta_f B\Big(\frac{\beta_b}{\beta_f}y + \Big(1 - \frac{\beta_b}{\beta_f}\Big)x\Big) + \frac{1-\gamma\beta_f}{\beta_f}(x - y)\Big). \tag{3.11}$$

The differential inclusion (3.10) is governed by the sum of the convex subdifferential operator $\beta_f \partial \mathcal{G}$ and the Lipschitz continuous operator $\mathcal{D}(\cdot)$. The existence and uniqueness of a global strong solution for the Cauchy problem (3.10), and hence for (3.9), follows from a direct application of [31, Proposition 3.12]. In turn, if $f \in \mathcal{C}^1(\mathcal{H})$, then (3.9) admits a unique $\mathcal{C}^1([0, +\infty[)$ global solution (x, y). It then follows from the first equation in (3.9) that $\dot{x}$ is a $\mathcal{C}^1([0, +\infty[)$ function, and hence $x \in \mathcal{C}^2([0, +\infty[)$. Existence and uniqueness of a classical global solution to the Cauchy problem associated to (iDINAM) is then obtained thanks to the equivalence in Theorem 3.2.

4 Asymptotic convergence properties of (iDINAM)

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

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

We complete these results by producing integral and pointwise convergence rates.

4.1 Preliminary results

The following result relies on the cocoercivity of B.

Lemma 4.1 *$B(p)$ and $\nabla f(p)$ are uniquely defined for $p \in S$, i.e.,*

$$p_1 \in S, \; p_2 \in S \Longrightarrow B(p_1) = B(p_2) \text{ and } \nabla f(p_1) = \nabla f(p_2).$$

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 the monotonicity of ∇f we have

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

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

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

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

The following lemma is a classic result from integration theory, often called Barlabat's theorem in control theory.

Lemma 4.2 *Let $1 \leq p < +\infty$ and $1 \leq r \leq +\infty$. Suppose that $u \in L^p([0,+\infty[;\mathcal{H})$ is a locally absolutely function, such that $\dot{u} \in L^r([0,+\infty[;\mathcal{H})$.*
Then $\lim_{t\to\infty} u(t) = 0$.

The following lemma will play a central role in the proof of our main convergence theorem. The proof can be found in [6, 20].

Lemma 4.3 *([6]) If $w \in C^2([0,+\infty[,\mathbb{R})$ is bounded from below and satisfies the following inequality*

$$\ddot{w}(t) + \gamma \dot{w}(t) \leq g(t),$$

where γ is a positive constant and $g \in L^1([0,+\infty[,\mathbb{R})$, then $w(t)$ converges as $t \to +\infty$.

4.2 Main result

The following result will be obtained by using Lyapunov analysis. Take $p \in S$. Let $x(\cdot)$ be a solution trajectory of the dynamical system (iDINAM). To analyze the convergence properties of $x(\cdot)$, we introduce the function $\mathcal{E}_p : [0, +\infty[\to \mathbb{R}+$ defined by

$$\begin{aligned}\mathcal{E}_p(t) &:= a\Big(f(x(t) + \beta_f \dot{x}(t)) - f(p) - \langle \nabla f(p), x(t) + \beta_f \dot{x}(t)) - p\rangle\Big) \\ &\quad + \frac{1}{2}\|x(t) - p + \beta_f \dot{x}(t)\|^2 + \frac{d}{2}\|x(t) - p\|^2, \end{aligned} \tag{4.1}$$

and that will serve us as a Lyapunov function. According to the convexity of f, $\mathcal{E}_p(\cdot)$ is a nonnegative function. Our goal is to adjust the constant $a > 0$ and $d > 0$ so that we have $\dot{\mathcal{E}}_p(t) \le 0$ for every $t \ge 0$.

Theorem 4.1 *Let $B : \mathcal{H} \to \mathcal{H}$ be a λ-cocoercive operator and $f : \mathcal{H} \to \mathbb{R}$ a $\mathcal{C}^1$ 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* (iDINAM) *where the involved parameters satisfy the following conditions:*

$$\gamma\beta_f > 1 \;\; \textit{and} \;\; \lambda > \frac{(\beta_b - \beta_f)^2}{4(\gamma\beta_f - 1)}. \tag{4.2}$$

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

(i) *(convergence)*

$$\begin{aligned}&x(t) \textit{ converges weakly , as } t \to +\infty, \textit{ to an element of } S.\\ &\lim_{t\to+\infty} \|\dot{x}(t)\| = 0,\\ &\lim_{t\to+\infty} \|B(x(t)) - B(p)\| = 0, \lim_{t\to+\infty} \|\nabla f(x(t)) - \nabla f(p)\| = 0,\end{aligned}$$

where $B(p)$ and $\nabla f(p)$ are uniquely defined for $p \in S$.

(ii) *(integral estimates)*

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

Proof **Lyapunov analysis.** Let us derivate the function $\mathcal{E}_p(\cdot)$ defined in (4.1). The derivation chain rule gives

$$\begin{aligned}\dot{\mathcal{E}}_p(t) &= a\langle \nabla f(x(t) + \beta_f \dot{x}(t)) - \nabla f(p), \dot{x}(t) + \beta_f \ddot{x}(t)\rangle\\ &\quad + \langle x(t) - p + \beta_f \dot{x}(t), \dot{x}(t) + \beta_f \ddot{x}(t)\rangle + d\langle x(t) - p, \dot{x}(t)\rangle.\end{aligned}$$

According to the constitutive equation (iDINAM) we have

$$\ddot{x}(t) = -\gamma \dot{x}(t) - \nabla f\left(x(t) + \beta_f \dot{x}(t)\right) - B(x(t) + \beta_b \dot{x}(t)).$$

Therefore,

$$\begin{aligned}\dot{\mathcal{E}}_p(t) =& a\langle \nabla f(x(t)+\beta_f \dot{x}(t)) - \nabla f(p), \dot{x}(t) + \beta_f(-\gamma\dot{x}(t) - \nabla f\left(x(t)+\beta_f\dot{x}(t)\right) - B(x(t)+\beta_b\dot{x}(t)))\rangle \\ &+\langle x(t)-p+\beta_f\dot{x}(t), \dot{x}(t)+\beta_f(-\gamma\dot{x}(t) - \nabla f\left(x(t)+\beta_f\dot{x}(t)\right) - B(x(t)+\beta_b\dot{x}(t)))\rangle \\ &+d\langle x(t)-p, \dot{x}(t)\rangle.\end{aligned}$$

Let us denote shortly

$$X(t) := \nabla f\Big(x(t)+\beta_f\dot{x}(t)\Big) - \nabla f(p),$$
$$Y(t) := B\Big(x(t)+\beta_b\dot{x}(t)\Big) - B(p).$$

Since $p \in S$, we have $\nabla f(p) + B(p) = 0$. So, we can arrange $\dot{\mathcal{E}}_p(t)$ as follows

$$\begin{aligned}\dot{\mathcal{E}}_p(t) &= a\langle X(t), \dot{x}(t) + \beta_f(-\gamma\dot{x}(t) - X(t) - Y(t))\rangle \\ &\quad + \langle x(t)-p+\beta_f\dot{x}(t), \dot{x}(t)+\beta_f(-\gamma\dot{x}(t) - X(t) - Y(t))\rangle + d\langle x(t)-p, \dot{x}(t)\rangle \\ &= -a\beta_f\|X(t)\|^2 + a(1-\gamma\beta_f)\langle X(t)), \dot{x}(t)\rangle - a\beta_f\langle X(t), Y(t)\rangle \qquad (4.3)\\ &\quad + \beta_f(1-\gamma\beta_f)\|\dot{x}(t)\|^2 + (d+1-\gamma\beta_f)\langle x(t)-p, \dot{x}(t)\rangle - \beta_f\langle x(t)-p+\beta_f\dot{x}(t), X(t)+Y(t)\rangle.\end{aligned}$$

By convexity of f, we have that ∇f is monotone. By definition of $X(t)$ this gives

$$\langle x(t)-p+\beta_f\dot{x}(t), X(t)\rangle \geq 0.$$

Moreover, since B is λ-cocoercive, we have

$$\begin{aligned}\langle x(t)-p+\beta_f\dot{x}(t), Y(t)\rangle &= \langle x(t)-p+\beta_b\dot{x}(t), Y(t)\rangle + (\beta_f-\beta_b)\langle \dot{x}(t), Y(t)\rangle \\ &\geq \lambda\|Y(t)\|^2 + (\beta_f-\beta_b)\langle \dot{x}(t), Y(t)\rangle.\end{aligned}$$

Combining the above results, and taking $d = \gamma\beta_f - 1 > 0$, we deduce from (4.3) that

$$\begin{aligned}\dot{\mathcal{E}}_p(t) \leq& -a\beta_f\|X(t)\|^2 + a(1-\gamma\beta_f)\langle X(t)), \dot{x}(t)\rangle - a\beta_f\langle X(t), Y(t)\rangle \\ &+ \beta_f(1-\gamma\beta_f)\|\dot{x}(t)\|^2 - \lambda\beta_f\|Y(t)\|^2 - \beta_f(\beta_f-\beta_b)\langle \dot{x}(t), Y(t)\rangle. \qquad (4.4)\end{aligned}$$

Let us majorize the scalar products that enter (4.4) with the help of the following elementary inequalities: for any $\rho > 0$ and $r > 0$, which are parameters that will be adjusted (recall that $\gamma\beta_f > 1$)

$$a(1-\gamma\beta_f)\langle X(t)), \dot{x}(t)\rangle \leq \frac{1}{2}\rho a(\gamma\beta_f-1)\|X(t)\|^2 + \frac{1}{2\rho}a(\gamma\beta_f-1)\|\dot{x}(t)\|^2 \qquad (4.5)$$
$$-a\beta_f\langle X(t), Y(t)\rangle \leq \frac{1}{2}ar\beta_f\|X(t)\|^2 + \frac{1}{2r}a\beta_f\|Y(t)\|^2. \qquad (4.6)$$

Combining (4.4) with (4.5) and (4.6), we get

$$\begin{aligned}\dot{\mathcal{E}}_p(t) \leq& -a\beta_f\|X(t)\|^2 + \frac{1}{2}\rho a(\gamma\beta_f-1)\|X(t)\|^2 + \frac{1}{2\rho}a(\gamma\beta_f-1)\|\dot{x}(t)\|^2 \\ &+ \frac{1}{2}ar\beta_f\|X(t)\|^2 + \frac{1}{2r}a\beta_f\|Y(t)\|^2 \\ &+ \beta_f(1-\gamma\beta_f)\|\dot{x}(t)\|^2 - \lambda\beta_f\|Y(t)\|^2 - \beta_f(\beta_f-\beta_b)\langle \dot{x}(t), Y(t)\rangle. \qquad (4.7)\end{aligned}$$

After rearranging the terms, we get

$$\begin{aligned}\dot{\mathcal{E}}_p(t) \leq& -a\left(\beta_f - \frac{1}{2}\rho(\gamma\beta_f-1) - \frac{1}{2}r\beta_f\right)\|X(t)\|^2 - (\gamma\beta_f-1)\left(\beta_f - \frac{a}{2\rho}\right)\|\dot{x}(t)\|^2 \\ &- \beta_f\left(\lambda - \frac{a}{2r}\right)\|Y(t)\|^2 - \beta_f(\beta_f-\beta_b)\langle \dot{x}(t), Y(t)\rangle. \qquad (4.8)\end{aligned}$$

Equivalently,

$$\dot{\mathcal{E}}_p(t) + a\left(\beta_f - \frac{1}{2}\rho(\gamma\beta_f - 1) - \frac{1}{2}r\beta_f\right)\|X(t)\|^2 + \beta_f\mathcal{S}(t) \leq 0, \tag{4.9}$$

where

$$\mathcal{S}(t) := \left(\lambda - \frac{a}{2r}\right)\|Y(t)\|^2 + (\beta_f - \beta_b)\langle \dot{x}(t), Y(t)\rangle + (\gamma\beta_f - 1)\left(1 - \frac{a}{2\rho\beta_f}\right)\|\dot{x}(t)\|^2.$$

We have $\mathcal{S}(t) = q(Y(t), \dot{x}(t))$ where $q : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is the quadratic form

$$q(Y, Z) := \left(\lambda - \frac{a}{2r}\right)\|Y\|^2 + (\beta_f - \beta_b)\langle Y, Z\rangle + (\gamma\beta_f - 1)\left(1 - \frac{a}{2\rho\beta_f}\right)\|Z\|^2.$$

The following system of conditions on the positive parameters a, r, ρ ensures that the coefficient of $\|X(t)\|^2$ in (4.9) is positive and the quadratic form q is positive definite:

$$\beta_f - \frac{1}{2}\rho(\gamma\beta_f - 1) - \frac{1}{2}r\beta_f > 0; \tag{4.10}$$

$$\lambda - \frac{a}{2r} > 0; \tag{4.11}$$

$$1 - \frac{a}{2\rho\beta_f} > 0; \tag{4.12}$$

$$4\left(\lambda - \frac{a}{2r}\right)(\gamma\beta_f - 1)\left(1 - \frac{a}{2\rho\beta_f}\right) - (\beta_f - \beta_b)^2 > 0. \tag{4.13}$$

Conditions (4.11) and (4.12) are respectively equivalent to $r > \frac{a}{2\lambda}$ and $\rho > \frac{a}{2\beta_f}$. So they are satisfied by taking $r = \frac{\tau a}{2\lambda}$ and $\rho = \frac{\tau a}{2\beta_f}$ with $\tau > 1$. Reinjecting these values in (4.10) we get the following condition

$$\tau a \leq \frac{4\lambda\beta_f^2}{\lambda(\gamma\beta_f - 1) + \beta_f^2}. \tag{4.14}$$

Let us now examine the last condition (4.13) which, due to the choice of r and ρ, simplifies as follows

$$\Delta(\tau) := 4\lambda\left(1 - \frac{1}{\tau}\right)^2(\gamma\beta_f - 1) - (\beta_f - \beta_b)^2 > 0.$$

We have

$$\lim_{\tau + \infty} \Delta(\tau) = 4\lambda(\gamma\beta_f - 1) - (\beta_f - \beta_b)^2$$

which is positive by our assumption (4.2) on the parameters. So, by taking τ large enough, and adjusting a small enough according to (4.14), we get that that the coefficient of $\|X(t)\|^2$ in (4.9) is positive, and that the quadratic form q is positive definite. We infer the existence of positive real numbers η and μ such that

$$\dot{\mathcal{E}}_p(t) + \eta\|X(t)\|^2 + \mu\beta_f\|\dot{x}(t)\|^2 + \mu\beta_f\|Y(t)\|^2 \leq 0. \tag{4.15}$$

Estimates. We rely on the estimate (4.15) that we integrate on $[0, t], t \geq 0$. We obtain

$$\mathcal{E}_p(t) + \eta\int_0^t \|X(s)\|^2 ds + \mu\beta_f\int_0^t \|\dot{x}(s)\|^2 ds + \mu\beta_f\int_0^t \|Y(s)\|^2 ds \leq \mathcal{E}_p(0). \tag{4.16}$$

From this we immediately obtain that $\mathcal{E}_p(t) \leq \mathcal{E}_p(0)$, *i.e.* $\mathcal{E}_p(t)$ is bounded from above. According to the definition of $\mathcal{E}_p(\cdot)$ we deduce that

$$\sup_{t\geq 0} \|x(t)-p\| < +\infty, \tag{4.17}$$

$$\sup_{t\geq 0} \|x(t)-p+\beta_f \dot{x}(t)\| < +\infty. \tag{4.18}$$

From (4.17)-(4.18) and $\beta_f > 0$, by using the triangle inequality we infer

$$\sup_{t\geq 0} \|\dot{x}(t)\| < +\infty. \tag{4.19}$$

Moreover, we immediately deduce from (4.16) and $\mathcal{E}_p(t)$ nonnegative the following integral estimates

$$\int_0^{+\infty} \|X(t)\|^2 dt < +\infty, \ \int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty, \ \int_0^{+\infty} \|Y(t)\|^2 dt < +\infty. \tag{4.20}$$

Let us rewrite (iDINAM) equivalently as follows (recall that $\nabla f(p) + Bp = 0$)

$$\ddot{x}(t) = -\gamma\dot{x}(t) - X(t) - Y(t).$$

According to (4.20) the second member of the above equality belongs to $L^2(0,+\infty;\mathcal{H})$. Therefore

$$\int_0^{+\infty} \|\ddot{x}(t)\|^2 \, dt < +\infty. \tag{4.21}$$

From (4.20) and (4.21) we have $\dot{x} \in L^2([0,+\infty[;\mathcal{H})$ and $\ddot{x} \in L^2([0,+\infty[;\mathcal{H})$. By Lemma 4.2 applied to $u = \dot{x}$ with $p = r = 2$ we deduce that

$$\lim_{t\to+\infty} \dot{x}(t) = 0. \tag{4.22}$$

Furthermore, since B is λ-cocoercive, it is $\dfrac{1}{\lambda}$-Lipschitz continuous. Therefore,

$$\left\| \frac{d}{dt} B(x(t) + \beta_b \dot{x}(t)) \right\| \leq \frac{1}{\lambda} \|\dot{x}(t) + \beta_b \ddot{x}(t)\|, \text{ for all } t \geq 0. \tag{4.23}$$

Hence,

$$\begin{aligned}
\int_0^{+\infty} \left\| \frac{d}{dt} B(x(t) + \beta_b \dot{x}(t)) \right\|^2 dt &\leq \frac{1}{\lambda^2} \int_0^{+\infty} \|\dot{x}(t) + \beta_b \ddot{x}(t)\|^2 dt \\
&\leq \frac{2}{\lambda^2} \int_0^{+\infty} \|\dot{x}(t)\|^2 dt + \frac{2\beta_b^2}{\lambda^2} \int_0^{+\infty} \|\ddot{x}(t)\|^2 dt < +\infty.
\end{aligned}$$

Similarly, we have

$$\int_0^{+\infty} \left\| \frac{d}{dt} \nabla f(x(t) + \beta_f \dot{x}(t)) \right\|^2 dt < +\infty$$

where have used that $x(t) + \beta_f \dot{x}(t)$ remains bounded (according to (4.17) and (4.19)) and that ∇f is Lipschitz continuous on the bounded sets. So, according to the definition of $X(t)$ and $Y(t)$ we have

$$\int_0^{+\infty} \left\| \frac{d}{dt} X(t) \right\|^2 dt < +\infty, \ \int_0^{+\infty} \left\| \frac{d}{dt} Y(t) \right\|^2 dt < +\infty. \tag{4.24}$$

From (4.20)-(4.24), by applying Lemma 4.2 we deduce that $\lim_{t\to+\infty} X(t) = \lim_{t\to+\infty} Y(t) = 0$, that is

$$\lim_{t\to+\infty} \|B((x(t) + \beta_b \dot{x}(t)) - B(p)\| = 0, \lim_{t\to+\infty} \|\nabla f(x(t) + \beta_f \dot{x}(t)) - \nabla f(p)\| = 0 \tag{4.25}$$

According to the Lipschitz continuity of B, and the Lipschitz continuity of ∇f on the bounded sets (recall that $x(t)$ and $\dot{x}(t)$ are bounded) we immediately deduce from (4.25) and $\lim_{t\to+\infty} \dot{x}(t) = 0$, that

$$\lim_{t\to+\infty} \|B(x(t)) - B(p)\| = 0, \lim_{t\to+\infty} \|\nabla f(x(t)) - \nabla f(p)\| = 0. \tag{4.26}$$

Convergence of the trajectory. To prove the existence of the weak limit of $x(t)$ as $t \to +\infty$, we use Opial's lemma (see [40] 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

$$\begin{aligned}\ddot{q}_p(t) + \gamma \dot{q}_p(t) &= \|\dot{x}(t)\|^2 + \langle \ddot{x}(t) + \gamma \dot{x}(t), x(t) - p\rangle \\ &= \|\dot{x}(t)\|^2 - \langle \nabla f(x(t) + \beta_f \dot{x}(t)) + B(x(t) + \beta_b \dot{x}(t)), x(t) - p\rangle.\end{aligned}$$

According to the monotonicity of ∇f and B, we have

$$\begin{aligned}&\langle \nabla f(x(t) + \beta_f \dot{x}(t)) + B(x(t) + \beta_b \dot{x}(t)), x(t) - p\rangle \\ &= \langle X(t) + Y(t), x(t) - p\rangle \\ &\geq -\beta_f \langle X(t), \dot{x}(t)\rangle - \beta_b \langle Y(t), \dot{x}(t)\rangle.\end{aligned}$$

Therefore,

$$\ddot{q}_p(t) + \gamma \dot{q}_p(t) \leq \|\dot{x}(t)\|^2 + \beta_f \langle X(t), \dot{x}(t)\rangle + \beta_b \langle Y(t), \dot{x}(t)\rangle. \tag{4.27}$$

Applying the Cauchy-Schwarz inequality, we get

$$\ddot{q}_p(t) + \gamma \dot{q}_p(t) \leq \|\dot{x}(t)\|^2 + \beta_f \|X(t)\|\|\dot{x}(t)\| + \beta_b \|Y(t)\|\|\dot{x}(t)\|. \tag{4.28}$$

Then note that the second member of (4.28)

$$g(t) := \|\dot{x}(t)\|^2 + \beta_f \|X(t)\|\|\dot{x}(t)\| + \beta_b \|Y(t)\|\|\dot{x}(t)\|$$

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

$$\begin{aligned}\int_0^{+\infty} \|X(t)\|\|\dot{x}(t)\| dt &\leq \frac{1}{2}\int_0^{+\infty} \|X(t)\|^2 dt + \frac{1}{2}\int_0^{+\infty} \|\dot{x}(t)\|^2 dt, \\ \int_0^{+\infty} \|Y(t)\|\|\dot{x}(t)\| dt &\leq \frac{1}{2}\int_0^{+\infty} \|Y(t)\|^2 dt + \frac{1}{2}\int_0^{+\infty} \|\dot{x}(t)\|^2 dt.\end{aligned}$$

Using (4.20), we deduce that

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

Since q_p is nonnegative, Lemma 4.3 shows that $\lim_{t\to+\infty} q_p(t)$ exists. To complete the proof via Opial's lemma, we need to show that every weak sequential cluster point of $x(t)$ belongs to S. Let $t_n \to +\infty$ such that $x(t_n) \rightharpoonup x^*, n \to +\infty$. According to (4.26)

$$\nabla f(x(t_n)) \to \nabla f(p);\ B(x(t_n)) \to B(p) \text{ strongly in } \mathcal{H}$$

and

$$x(t_n) \rightharpoonup x^* \text{ weakly in } \mathcal{H}.$$

From the closedness property of the graph of the maximally monotone operators ∇f and B in $w-\mathcal{H}\times s-\mathcal{H}$, we deduce that $\nabla f(x^*) = \nabla f(p)$, and $B(x^*) = B(p)$. Therefore $\nabla f(x^*)+B(x^*) = \nabla f(p)+B(p) = 0$, that is $x^* \in S$. Consequently, $x(t)$ converges weakly to an element of S as t goes to $+\infty$. The proof of Theorem 4.1 is thus completed.

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

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

The existence of strong global solutions to this system is guaranteed by Theorem 3.1. The convergence properties as $t \to +\infty$ of the solution trajectories generated by this system is a consequence of Theorem 4.1 and are given below.

Corollary 4.1 *Let $B : \mathcal{H} \to \mathcal{H}$ be a λ-cocoercive operator and $f : \mathcal{H} \to \mathbb{R}$ 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* (iDINAM)*, where $A = \nabla f + B$, $\beta_b = \beta_f := \beta > 0$ and where the involved parameters satisfy the following condition $\gamma\beta > 1$. Then, for any solution trajectory $x : [0, +\infty[\to \mathcal{H}$ of* (iDINAM)*, the following properties are satisfied:*

(i) *(convergence) The trajectory $x(t)$ converges weakly, as $t \to +\infty$, to an element $x^* \in S$. Moreover* $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0,$ *and* $\lim_{t\to+\infty} \|A(x(t) + \beta\dot{x}(t))\| = 0.$

(ii) *(integral estimate)*

$$\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty, \int_0^{+\infty} \|\ddot{x}(t)\|^2 dt < +\infty,$$
$$\int_0^{+\infty} \|A(x(t) + \beta\dot{x}(t))\|^2 dt < +\infty, \textit{ and } \int_0^{+\infty} \left\| \frac{d}{dt} A(x(t) + \beta\dot{x}(t)) \right\|^2 dt < +\infty.$$

4.3 Comparison of the dynamics with explicit and implicit Newton-type damping

For simplicity, let us compare the dynamics in the case $\beta_f = \beta_b = \beta > 0$. According to the previous study of the authors in [3] concerning the dynamic (DINAM) with explicit Newton-type damping, the condition on the parameters ensuring the convergence of the trajectories is

$$\lambda\gamma > \beta + \frac{1}{\gamma} \tag{4.29}$$

On the other hand, the corresponding condition for (iDINAM), as given by Corollary 4.1 is

$$\gamma\beta > 1. \tag{4.30}$$

As a striking result, we can observe that, contrary to (DINAM), the cocoercivity parameter λ no longer enters the condition relative to (iDINAM). This suggests in particular that it would be interesting to consider the case of an asymptotic vanishing damping coefficient $\gamma(t) = \frac{\alpha}{t}$ which is in accordance with the Nesterov accelerated scheme. By adjusting accordingly the coefficient $\beta(t)$ which now tends to infinity, this would make it possible to obtain fast convergence results for general monotone inclusions. In fact, first results in this direction have been obtained for the ADMM algorithm, see [15].

5 Inertial proximal algorithms associated with (iDINAAM)

We are interested in the convergence properties of several splitting algorithms with inertial features obtained by temporal discretization of the second-order (in time) evolution equation:

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f\left(x(t) + \beta_f \dot{x}(t)\right) + B\left(x(t) + \beta_b \dot{x}(t)\right) = 0. \tag{iDINAM}$$

We aim to obtain, under an appropriate adjustment of the parameters and the discretization step, convergence results of the same type as those obtained in the previous section, in the continuous case.

5.1 An inertial proximal-gradient algorithm

In this section, f is a $\mathcal{C}^1$ convex function whose gradient is L-Lipschitz continuous. Take a fixed time step $h > 0$, and consider the following finite-difference scheme for (iDINAM):

$$\begin{aligned}\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right)& \\ + B\left(x_{k+1} + \frac{\beta_b}{h}(x_{k+1} - x_k)\right) = 0.&\end{aligned} \tag{5.1}$$

This scheme is implicit with respect to the nonpotential B and explicit with respect to the potential operator ∇f. According to the gradient-like structure of the algorithm when $B = 0$, we expect to obtain convergence results by assuming that the step size h is taken small enough. After expanding (5.1), we obtain

$$\begin{aligned}(1 + \gamma h)(x_{k+1} - x_k) + h^2 B\left(x_{k+1} + \frac{\beta_b}{h}(x_{k+1} - x_k)\right)& \\ = (x_k - x_{k-1}) - h^2 \nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right).&\end{aligned} \tag{5.2}$$

Set $\alpha := 1 + \frac{\beta_b}{h}$. After arranging (5.2), we obtain equivalently

$$x_{k+1} = \frac{\alpha - 1}{\alpha} x_k + \frac{1}{\alpha}(\mathrm{Id} + \frac{\alpha h^2}{1 + \gamma h} B)^{-1}(\xi_k),$$

with

$$\xi_k = x_k + \frac{\alpha}{1 + \gamma h}\left((x_k - x_{k-1}) - h^2 \nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right)\right).$$

We thus obtain the following algorithm.

(iDINAAM-split):

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

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

$\xi_k = x_k + \frac{\alpha}{1 + \gamma h}(x_k - x_{k-1}) - \frac{\alpha h^2}{1 + \gamma h} \nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right)$,

$x_{k+1} = \frac{\alpha - 1}{\alpha} x_k + \frac{1}{\alpha}\left(\mathrm{Id} + \frac{\alpha h^2}{1 + \gamma h} B\right)^{-1}(\xi_k)$.

Theorem 5.1 *Let $B : \mathcal{H} \to \mathcal{H}$ be a λ-cocoercive operator and $f : \mathcal{H} \to \mathbb{R}$ a differentiable convex function whose gradient is L-Lipschitz continuous. Suppose the positive parameters $\lambda, \gamma, \beta_b, \beta_f$ satisfy*

$$0 < h < \frac{2}{L\beta_f},\ \gamma\beta_f > 1 \text{ and } \lambda > \frac{(\beta_b - \beta_f)^2}{4(\gamma\beta_f - 1)}. \tag{5.3}$$

Then, the sequence (x_k) generated by the algorithm (iDINAAM-split) *has the following properties:*

(i) (x_k) *converges weakly to an element in S;*

(ii) $\lim_{k\to\infty} \|\nabla f(x_k) - \nabla f(p)\| = 0, \quad \lim_{k\to\infty} \|B(x_k) - B(p)\| = 0$.

(iii) $\sum_{k=1}^{\infty} \|x_k - x_{k-1}\|^2 < +\infty,\ \sum_{k=1}^{\infty} \|\nabla f(x_k) - \nabla f(p)\|^2 < +\infty, \sum_{k=1}^{\infty} \|B(x_k) - B(p)\|^2 < +\infty$ *where $\nabla f(p)$, $B(p)$ do not depend on the choice of $p \in S$.*

Proof **The discrete energy** Take $p \in S$. Let us consider the sequence (E_k) defined for all $k \geq 1$ by

$$E_k := \frac{1}{2}\|(x_k - p) + \frac{\beta_f}{h}(x_k - x_{k-1})\|^2 + \frac{\delta}{2}\|x_k - p\|^2,$$

where δ is a positive coefficient to adjust.
For each $k \geq 1$, let us briefly write E_k as follows:

$$E_k = \frac{1}{2}\|v_k\|^2 + \frac{\delta}{2}\|x_k - p\|^2,$$

with

$$v_k := x_k - p + \frac{\beta_f}{h}(x_k - x_{k-1}).$$

By definition of v_k and (5.1), we have

$$\begin{aligned}
&v_{k+1} - v_k = x_{k+1} - x_k + \frac{\beta_f}{h}(x_{k+1} - 2x_k + x_{k-1}) \\
&= (1 - \gamma\beta_f)(x_{k+1} - x_k) - h\beta_f\nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - h\beta_f B\left(x_{k+1} + \frac{\beta_b}{h}(x_{k+1} - x_k)\right) \\
&= (1 - \gamma\beta_f)(x_{k+1} - x_k) - h\beta_f\nabla f(y_k) - h\beta_f B(z_k),
\end{aligned}$$

where we write shortly

$$\begin{aligned}
y_k &:= x_k + \frac{\beta_f}{h}(x_k - x_{k-1}), \\
z_k &:= x_{k+1} + \frac{\beta_b}{h}(x_{k+1} - x_k).
\end{aligned}$$

Therefore, for $k \geq 1$, we have

$$\begin{aligned}
&\frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2 = -\frac{1}{2}\|v_{k+1} - v_k\|^2 + \langle v_{k+1} - v_k, v_{k+1}\rangle \\
&= -\frac{1}{2}(\gamma\beta_f - 1)^2\|x_{k+1} - x_k\|^2 - \frac{1}{2}h^2\beta_f^2\|\nabla f(y_k) + B(z_k)\|^2 \\
&-h\beta_f(\gamma\beta_f - 1)\langle x_{k+1} - x_k, \nabla f(y_k) + B(z_k)\rangle \\
&-\langle x_{k+1} - p + \frac{\beta_f}{h}(x_{k+1} - x_k), (\gamma\beta_f - 1)(x_{k+1} - x_k) + h\beta_f\nabla f(y_k) + h\beta_f B(z_k)\rangle.
\end{aligned} \tag{5.4}$$

Then use the elementary identity

$$\frac{1}{2}\|x_{k+1}-p\|^2-\frac{1}{2}\|x_k-p\|^2=-\frac{1}{2}\|x_{k+1}-x_k\|^2+\langle x_{k+1}-x_k,x_{k+1}-p\rangle. \tag{5.5}$$

Take $\delta=\gamma\beta_f-1$. Thus, as the first condition on the parameters, we ask

$$\gamma\beta_f>1. \tag{5.6}$$

From (5.4) and (5.5), we deduce that

$$\begin{aligned}E_{k+1}-E_k=&-\left(\frac{1}{2}\delta^2+\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)\|x_{k+1}-x_k\|^2-\frac{1}{2}h^2\beta_f^2\|\nabla f(y_k)+B(z_k)\|^2\\&-h\beta_f\delta\langle x_{k+1}-x_k,\nabla f(y_k)+B(z_k)\rangle\\&-\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k),h\beta_f\nabla f(y_k)+h\beta_f B(z_k)\rangle.\end{aligned}$$

According to $\nabla f(p)+B(p)=0$, we can rewrite the previous relation as follows

$$\begin{aligned}E_{k+1}-E_k=&-\left(\frac{1}{2}\delta^2+\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)\|x_{k+1}-x_k\|^2-\frac{1}{2}h^2\beta_f^2\|Y_k+Z_k\|^2\\&-h\beta_f\delta\langle x_{k+1}-x_k,Y_k+Z_k\rangle-h\beta_f\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k),Y_k+Z_k\rangle,\end{aligned} \tag{5.7}$$

where $Y_k=\nabla f(y_k)-\nabla f(p)$ and $Z_k=B(z_k)-B(p)$.
Since B is λ-cocoercive we have

$$\begin{aligned}\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k),Z_k\rangle&=\langle z_k-p+\frac{1}{h}(\beta_f-\beta_b)(x_{k+1}-x_k),B(z_k)-B(p)\rangle\\&\geq\lambda\|B(z_k)-B(p)\|^2+\frac{1}{h}(\beta_f-\beta_b)\langle(x_{k+1}-x_k),B(z_k)-B(p)\rangle\\&=\lambda\|Z_k\|^2+\frac{1}{h}(\beta_f-\beta_b)\langle(x_{k+1}-x_k),Z_k\rangle,\end{aligned}$$

Similarly, since ∇f is $1/L$-cocoercive, and by using the constitutive equation (5.1), we get

$$\begin{aligned}&\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k),Y_k\rangle\\&=\langle y_k-p+x_{k+1}-x_k+\frac{\beta_f}{h}(x_{k+1}-2x_k+x_{k-1}),\nabla f(y_k)-\nabla f(p)\rangle\\&\geq\frac{1}{L}\|Y_k\|^2+\langle x_{k+1}-x_k+\frac{\beta_f}{h}(x_{k+1}-2x_k+x_{k-1}),\nabla f(y_k)-\nabla f(p)\rangle,\\&=\frac{1}{L}\|Y_k\|^2+\langle x_{k+1}-x_k-\gamma\beta_f(x_{k+1}-x_k)-h\beta_f\nabla f(y_k)-h\beta_f B(z_k),\nabla f(y_k)-\nabla f(p)\rangle\\&=\frac{1}{L}\|Y_k\|^2-\langle\delta(x_{k+1}-x_k)+h\beta_f Y_k+h\beta_f Z_k,Y_k\rangle\end{aligned}$$

Combining the above relations with (5.7), we get

$$\begin{aligned}E_{k+1}-E_k\leq&\left(\frac{1}{2}h^2\beta_f^2-\frac{h\beta_f}{L}\right)\|Y_k\|^2-\left(\frac{1}{2}\delta^2+\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)\|x_{k+1}-x_k\|^2\\&-(h\beta_f\delta+\beta_f(\beta_f-\beta_b))\langle x_{k+1}-x_k,Z_k\rangle-\left(\frac{1}{2}h^2\beta_f^2+h\beta_f\lambda\right)\|Z_k\|^2.\end{aligned} \tag{5.8}$$

Equivalently,

$$E_{k+1} - E_k + \mathcal{S}_k \leq \left(\frac{1}{2}h^2\beta_f^2 - \frac{h\beta_f}{L}\right)\|Y_k\|^2, \tag{5.9}$$

where

$$\begin{aligned} S_k &= \left(\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta\right)\|x_{k+1} - x_k\|^2 + (h\beta_f\delta + \beta_f(\beta_f - \beta_b))\langle x_{k+1} - x_k, Z_k\rangle \\ &+ \left(\frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda\right)\|Z_k\|^2. \end{aligned}$$

Thus, as second conditions on the parameters, we ask $\frac{1}{2}h^2\beta_f^2 - \frac{h\beta_f}{L} < 0$, that is

$$0 < h < \frac{2}{L\beta_f}. \tag{5.10}$$

Then note that $S_k = q(x_{k+1} - x_k, Z_k) > 0$ if $4ag - b^2 > 0$ where $q : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is the quadratic form

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

where

$$\begin{aligned} a &= \frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta, \\ b &= h\beta_f\delta + \beta_f(\beta_f - \beta_b), \\ g &= \frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda. \end{aligned}$$

The third and last condition on the parameters will be obtained by asking the quadratic form q to be positive definite. Since a and g are positive this is equivalent to having $4ag - b^2 > 0$. We have

$$\begin{aligned} 4ag - b^2 &= 4\left(\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta\right)\left(\frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda\right) - (h\beta_f\delta + \beta_f(\beta_f - \beta_b))^2 \\ &= \beta_f^2\Big(4\lambda\delta - (\beta_f - \beta_b)^2\Big) + 2h\delta\beta_f\Big(\lambda(\delta + 1) + \beta_f\beta_b\Big) + h^2\beta_f^2\delta \\ &\geq \beta_f^2\Big(4\lambda\delta - (\beta_f - \beta_b)^2 > 0, \end{aligned} \tag{5.11}$$

where the last inequality comes from our assumptions. Therefore, q is positive definite, and there exist positive real numbers μ and η such that for any $k \geq 1$,

$$E_{k+1} - E_k + \mu\|x_{k+1} - x_k\|^2 + \mu\|B(z_k) - B(p)\|^2 + \eta\|\nabla f(y_k) - \nabla f(p)\|^2 \leq 0. \tag{5.12}$$

Note that μ depends on all the damping coefficients involved in the algorithm and on the step size h. Its precise estimation is an interesting subject for numerical purpose.

Estimates. According to (5.12), the sequence of nonnegative numbers (E_k) is nonincreasing, and therefore converges. In particular, it is bounded. From this, we immediately deduce that

$$\sup_k \|(x_k - p) + \frac{\beta_f}{h}(x_k - x_{k-1})\| < +\infty, \tag{5.13}$$

$$\sup_k \|x_k - p\| < +\infty. \tag{5.14}$$

Moreover, by summing the inequalities (5.12), we deduce that

$$\sum_{k=1}^{\infty} \|x_k - x_{k-1}\|^2 < \infty, \quad \sum_{k=1}^{\infty} \|\nabla f(y_k) - \nabla f(p)\|^2 < \infty, \quad \sum_{k=1}^{\infty} \|B(z_k) - B(p)\|^2 < \infty. \tag{5.15}$$

Elementary algebra and the Lipschitz continuity of ∇f give, for each $k \geq 1$

$$\begin{aligned} \|\nabla f(x_k) - \nabla f(p)\|^2 &\leq (\|\nabla f(y_k) - \nabla f(p)\| + \|\nabla f(x_k) - \nabla f(y_k)\|)^2 \\ &\leq 2\|\nabla f(y_k) - \nabla f(p)\|^2 + 2\|\nabla f(x_k) - \nabla f(y_k)\|^2 \\ &\leq 2\|\nabla f(y_k) - \nabla f(p)\|^2 + 2L^2\|x_k - y_k\|^2 \\ &\leq 2\|\nabla f(y_k) - \nabla f(p)\|^2 + \frac{2L^2\beta_f^2}{h^2}\|x_k - x_{k-1}\|^2. \end{aligned} \tag{5.16}$$

According to (5.15) we get

$$\sum_{k=1}^{\infty} \|\nabla f(x_k) - \nabla f(p)\|^2 < +\infty.$$

Similarly, since B is $1/\lambda$-Lipschitz, we get

$$\sum_{k=1}^{\infty} \|B(x_k) - B(p)\|^2 < +\infty.$$

Since the general term of a convergent series goes to zero, we deduce (ii).

Convergence of (x_k)**.** Let us first show that every weak cluster point x^* of the sequence (x_k) belongs to S. Consider a subsequence (x_{k_n}) of (x_k), such that $x_{k_n} \rightharpoonup x^*$, as $n \to +\infty$. According to the item (ii) already proved we have

$$\nabla f\left(x_{k_n}\right) \to \nabla f(p), \quad B\left(x_{k_n}\right) \to B(p) \text{ strongly in } \mathcal{H},$$

and

$$x_{k_n} \rightharpoonup x^* \text{ weakly in } \mathcal{H}.$$

From the closedness property of the graph of the maximally monotone operators ∇f and B in $w-\mathcal{H} \times s-\mathcal{H}$, we deduce that $\nabla f(x^*) = \nabla f(p)$, and $B(x^*) = B(p)$. Therefore $\nabla f(x^*) + B(x^*) = \nabla f(p) + B(p) = 0$, that is $x^* \in S$.

According to the estimate (iii) we have $\sum_{k=1}^{\infty} \|x_k - x_{k-1}\|^2 < +\infty$. Since the general term of a convergent series goes to zero, we deduce that $\lim_k \|x_k - x_{k-1}\| = 0$. According to the definition of E_k, and since $\lim_k E_k$ exists (indeed it is nonincreasing) , we deduce that, for any $p \in S$

$$\lim_{k\to\infty} \|x_k - p\| \quad \text{exists.}$$

So, the two conditions of the Opial's lemma are satisfied, which completes the proof of the convergence of the sequence (x_k).

5.2 Errors, perturbations

Now we will examine the effect of the introduction of perturbations, errors in the algorithm (iDINAAM). Let us start from the perturbed version of (iDINAM):

$$\ddot{x}(t) + \gamma\dot{x}(t) + \nabla f\left(x(t) + \beta_f \dot{x}(t)\right) + B\left(x(t) + \beta_b \dot{x}(t)\right) = e(t), \qquad \text{(iDINAM)}$$

where the right-handside $e(\cdot)$ takes into account perturbations, errors. A similar discretization as before gives

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_k - x_{k-1}) + \nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) + B\left(x_{k+1} + \frac{\beta_b}{h}(x_{k+1} - x_k)\right) = e_k. \qquad (5.17)$$

Hence, we obtain the following algorithm.

(iDINAAM-pert):

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

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

$\xi_k = x_k + \frac{\alpha}{1+\gamma h}(x_k - x_{k-1}) - \frac{\alpha h^2}{1+\gamma h}\nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) + \frac{\alpha h^2}{1+\gamma h} e_k,$

$x_{k+1} = \frac{\alpha - 1}{\alpha} x_k + \frac{1}{\alpha}\left(\mathrm{Id} + \frac{\alpha h^2}{1+\gamma h} B\right)^{-1}(\xi_k).$

Theorem 5.2 *Let us make the assumptions of Theorem 5.1, and suppose that the sequence (e_k) of perturbations, errors satisfies:*

$$\sum_{k=1}^{\infty} \|e_k\| < +\infty.$$

Then, the sequence (x_k) generated by the algorithm (iDINAAM-pert) *has the following properties (where $p \in S$):*

(i) (x_k) *converges weakly to an element in S;*

(ii) $\sum_{k=1}^{\infty} \|x_k - x_{k-1}\|^2 < +\infty, \quad \sum_{k=1}^{\infty} \left\|\nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - \nabla f(p)\right\|^2 < +\infty,$

$\sum_{k=1}^{\infty} \left\|B\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - B(p)\right\|^2 < +\infty, \sum_{k=1}^{\infty} \|\nabla f(x_k) - \nabla(p)\|^2 < +\infty,$

$\sum_{k=1}^{\infty} \left\|B\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - B(p)\right\|^2 < +\infty, \sum_{k=1}^{\infty} \|B(x_k) - B(p)\|^2 < +\infty,$

$\sum_{k=1}^{\infty} \|\nabla f(x_k) - \nabla f(x_{k-1})\|^2 < +\infty,$ *and* $\sum_{k=1}^{\infty} \|B(x_k) - B(x_{k-1})\|^2 < +\infty;$

(iii) $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0, \ \lim_{k\to\infty} \|B(x_k) - B(p)\| = 0, \ \lim_{k\to\infty} \|\nabla f(x_k) - \nabla f(p)\| = 0.$

Passing from the Lyapunov analysis in the unperturbed case to the perturbed case is a classical procedure, see [11] for example. It is based on a similar Lyapunov analysis and the use of the following discrete version of the Gronwall Lemma, see [11, Lemma A.9.] for a proof.

Lemma 5.1 *Let a be a positive real number and $(y_k), (g_k)$ be nonnegative sequences such that for all $k \geq 0$, we have*

$$\frac{1}{2}y_k^2 \leq \frac{1}{2}a^2 + \sum_{0\leq i<k} g_i y_i.$$

Then, the following inequality holds for all $k \geq 0$: $\quad y_k \leq a + \sum_{0\leq i<k} g_i.$

Proof (of Theorem 5.2) The proof is similar to that of Theorem 5.1. It uses the following sequence (E_k) as a discrete energy function

$$E_k := \frac{1}{2}\|x_k - p + \frac{\beta_f}{h}(x_k - x_{k-1})\|^2 + \frac{\delta}{2}\|x_k - p\|^2,$$

where δ are positive coefficient to adjust.
By setting $\delta = \gamma\beta_f - 1, Y_k = \nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - \nabla f(p), Z_k = B\left(x_{k+1} + \frac{\beta_b}{h}(x_{k+1} - x_k)\right) - B(p)$ for $k \geq 1$ and using the same argument as in the proof of Theorem 5.1, we have

$$E_{k+1} - E_k + \mathcal{S}_k + \left(\frac{h\beta_f}{L} - \frac{1}{2}h^2\beta_f^2\right)\|Y_k\|^2 \leq \varepsilon_k, \tag{5.18}$$

where

$$\begin{aligned} S_k = &\left(\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta\right)\|x_{k+1} - x_k\|^2 + (h\beta_f\delta + \beta_f(\beta_f - \beta_b))\langle x_{k+1} - x_k, Z_k\rangle \\ &+ \left(\frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda\right)\|Z_k\|^2, \end{aligned}$$

and

$$\varepsilon_k = h\beta_f\langle x_{k+1} - p + \frac{\beta_f}{h}(x_{k+1} - x_k), e_k\rangle.$$

According to an elementary inequality, we have that

$$\langle x_{k+1} - x_k, e_k\rangle \leq \frac{1}{2\eta}\|x_{k+1} - x_k\|^2 + \frac{\eta}{2}\|e_k\|^2, \tag{5.19}$$

holds for any $\eta > 0$. Moreover, by using Cauchy-Schwarz inequality, we have

$$\langle x_{k+1} - p, e_k\rangle \leq \|x_{k+1} - p\|\|e_k\|. \tag{5.20}$$

Combining (5.18)-(5.20), we obtain

$$E_{k+1} - E_k + S_k + \left(\frac{h\beta_f}{L} - \frac{1}{2}h^2\beta_f^2\right)\|Y_k\|^2 \leq \varepsilon'_k, \tag{5.21}$$

where

$$\begin{aligned} S_k = &\left(\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta - \frac{\beta_f^2}{2\eta}\right)\|x_{k+1} - x_k\|^2 + (h\beta_f\delta + \beta_f(\beta_f - \beta_b))\langle x_{k+1} - x_k, Z_k\rangle \\ &+ \left(\frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda\right)\|Z_k\|^2, \end{aligned}$$

and

$$\varepsilon_k' = \frac{\eta\beta_f^2}{2}\|e_k\|^2 + h\beta_f\|x_{k+1} - p\|\|e_k\|.$$

We choose $\eta > 0$ such that $\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta - \frac{\beta_f^2}{2\eta} > 0$.
Since S_k is a quadratic form, $S_k > 0$ if

$$4\left(\frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda\right)\left(\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta - \frac{\beta_f^2}{2\eta}\right) - (h\beta_f\delta + \beta_f(\beta_f - \beta_b))^2 > 0. \tag{5.22}$$

Notice that

$$\begin{aligned}&\lim_{h\to 0^+} 4\left(\frac{1}{2}h^2\beta_f^2 + h\beta_f\lambda\right)\left(\frac{1}{2}\delta^2 + \frac{\delta\beta_f}{h} + \frac{1}{2}\delta - \frac{\beta_f^2}{2\eta}\right) - (h\beta_f\delta + \beta_f(\beta_f - \beta_b))^2 \\ &= 4\beta_f^2\left[\lambda - \frac{(\beta_b - \beta_f)^2}{4\delta}\right] > 0\end{aligned} \tag{5.23}$$

thanks to the assumption on the parameters. This guarantees the existence of $h > 0$ satisfying (5.22). Thus, there exists a positive real number μ such that for any $k \geq 1$,

$$E_{k+1} - E_k + \mu\|x_{k+1} - x_k\|^2 + \mu\|B(z_k) - B(p)\|^2 + \left(\frac{h\beta_f}{L} - \frac{1}{2}h^2\beta_f^2\right)\|\nabla f(y_k) - \nabla f(p)\|^2 \leq \varepsilon_k'. \tag{5.24}$$

From (5.24) we deduce that

$$E_{k+1} \leq E_1 + \sum_{1\leq i<k+1} \varepsilon_i'.$$

Taking into account the form of the energy sequence (E_k), we obtain

$$\frac{\delta}{2}\|x_{k+1} - p\|^2 \leq E_1 + \sum_{1\leq i<k+1} \varepsilon_i'. \tag{5.25}$$

According to the assumption $\sum_{k=1}^{\infty}\|e_k\| < +\infty$, this implies that $\sum_{k=1}^{\infty}\|e_k\|^2 < +\infty$. Therefore, there exists $C > 0$ such that

$$\sum_{1\leq i<k+1} \varepsilon_i' \leq h\beta_f \sum_{1\leq i<k+1} \|x_{k+1} - p\|\|e_k\| + C. \tag{5.26}$$

From (5.25) and (5.26), we conclude that

$$\frac{\delta}{2}\|x_{k+1} - p\|^2 \leq E_1 + h\beta_f \sum_{1\leq i<k+1} \|x_{k+1} - p\|\|e_k\| + C.$$

More precisely, we can rewrite this estimate as follows

$$\frac{1}{2}\|x_{k+1} - p\|^2 \leq \frac{1}{2}C_0^2 + c_0 \sum_{1\leq i<k+1} \|x_{k+1} - p\|\|e_k\|, \tag{5.27}$$

where

$$C_0 = \sqrt{\frac{E_1 + C}{\delta}}, \quad c_0 = \frac{h\beta_f}{\delta}.$$

Now, by applying Lemma 5.1 to (5.27), we obtain

$$\|x_{k+1} - p\| \le C_0 + c_0 \sum_{1 \le i < k+1} \|e_i\| < +\infty. \tag{5.28}$$

Therefore, $(\|x_k - p\|)$ and consequently $(\|x_k\|)$ is a bounded sequence.
Returning to (5.26), according to the boundedness of $(\|x_k - p\|)$ and the assumption of (e_k), we obtain

$$\sum_{k=1}^{\infty} \epsilon'_k < +\infty.$$

The rest of the proof is similar to that of Theorem 5.1, so we omit here.

5.3 A variant of the proximal-gradient algorithm

In this section, we consider a variant of the previous proximal-gradient algorithm, where the role of the operators is reversed. We consider the following semi-implicit finite-difference scheme for (iDINAM):

$$\begin{aligned}\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \nabla f\left(x_{k+1} + \frac{\beta_f}{h}(x_{k+1} - x_k)\right)& \\ + B\left(x_k + \frac{\beta_b}{h}(x_k - x_{k-1})\right) = 0,&\end{aligned} \tag{5.29}$$

where $h > 0$ is a fixed time step.
After expanding (5.29), we obtain the following algorithm.

(iDINAAM-var):

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

$\alpha = 1 + \dfrac{\beta_f}{h}$,

$y_k = x_k + (h^2 - \gamma h)(x_k - x_{k-1}) - h^2 B\left(x_k + \dfrac{\beta_b}{h}(x_k - x_{k-1})\right)$,

$z_k = (\mathrm{Id} + \alpha h^2 \nabla f)^{-1}(\alpha y_k - (\alpha - 1)x_k)$,

$x_{k+1} = \dfrac{1}{\alpha}(\alpha - 1)x_k + \dfrac{1}{\alpha} z_k$.

Theorem 5.3 *Let $B : \mathcal{H} \to \mathcal{H}$ be a λ-cocoercive operator and $f : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function whose gradient is L-Lipschitz continuous. Suppose that the positive parameters $\lambda, \gamma, \beta_b, \beta_f$ satisfy*

$$\gamma\beta_f > 1 \text{ and } \lambda > \frac{(\beta_b - \beta_f)^2}{4(\gamma\beta_f - 1)}. \tag{5.30}$$

Then, there exists h^* such that for all $0 < h < h^*$, the sequence (x_k) generated by the algorithm (iDINAAM-var) has the following properties (where $p \in S$):

(i) (x_k) *converges weakly to an element in S;*

(ii) $\displaystyle\sum_{k=1}^{\infty} \|x_k - x_{k-1}\|^2 < +\infty, \quad \sum_{k=1}^{\infty} \left\|\nabla f\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - \nabla f(p)\right\|^2 < +\infty,$

$\displaystyle\sum_{k=1}^{\infty} \left\|B\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - B(p)\right\|^2 < +\infty, \sum_{k=1}^{\infty} \|\nabla f(x_k) - \nabla(p)\|^2 < +\infty,$

$$\sum_{k=1}^{\infty} \|B\left(x_k + \frac{\beta_f}{h}(x_k - x_{k-1})\right) - B(p)\|^2 < +\infty, \sum_{k=1}^{\infty} \|B\left(x_k\right) - B(p)\|^2 < +\infty,$$

$$\sum_{k=1}^{\infty} \|\nabla f(x_k) - \nabla f(x_{k-1})\|^2 < +\infty, \text{ and } \sum_{k=1}^{\infty} \|B(x_k) - B(x_{k-1})\|^2 < +\infty;$$

(iii) $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0, \ \lim_{k\to\infty} \|B(x_k) - B(p)\| = 0, \ \lim_{k\to\infty} \|\nabla f(x_k) - \nabla f(p)\| = 0.$

Proof **The discrete energy** Take $p \in S$. Consider the sequence (E_k) defined for all $k \geq 1$ by the formula

$$E_k := \frac{1}{2}\|x_k - p + \frac{\beta_f}{h}(x_k - x_{k-1})\|^2 + \frac{\delta}{2}\|x_k - p\|^2,$$

where δ is a positive coefficient to adjust.
For each $k \geq 1$, let us briefly write E_k as follows:

$$E_k = \frac{1}{2}\|v_k\|^2 + \frac{\delta}{2}\|x_k - p\|^2,$$

with

$$v_k := x_k - p + \frac{\beta_f}{h}(x_k - x_{k-1}).$$

By definition of v_k and the formula (5.29), we have

$$\begin{aligned} v_{k+1} - v_k &= x_{k+1} - x_k + \frac{\beta_f}{h}(x_{k+1} - 2x_k + x_{k-1}) \\ &= (1 - \gamma\beta_f)(x_{k+1} - x_k) - h\beta_f \nabla f\left(x_{k+1} + \frac{\beta_f}{h}(x_{k+1} - x_k)\right) - h\beta_f\left(B(x_k + \frac{\beta_b}{h}(x_k - x_{k-1})\right) \\ &= (1 - \gamma\beta_f)(x_{k+1} - x_k) - h\beta_f \nabla f\,(y_k) - h\beta_f B(z_k)\rangle, \end{aligned}$$

in which

$$\begin{aligned} y_k \quad &= x_{k+1} + \frac{\beta_f}{h}(x_{k+1} - x_k), \\ z_k \quad &= x_k + \frac{\beta_b}{h}(x_k - x_{k-1}). \end{aligned}$$

Therefore, for $k \geq 1$, we have

$$\begin{aligned} \frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2 &= -\frac{1}{2}\|v_{k+1} - v_k\|^2 + \langle v_{k+1} - v_k, v_{k+1}\rangle \\ &\leq -(\gamma\beta_f - 1)\langle x_{k+1} - p + \frac{\beta_f}{h}(x_{k+1} - x_k), x_{k+1} - x_k\rangle \\ &\quad - \langle x_{k+1} - p + \frac{\beta_f}{h}(x_{k+1} - x_k), h\beta_f \nabla f\,(y_k) + h\beta_f B(z_k)\rangle. \end{aligned} \tag{5.31}$$

Using the elementary identity, one has

$$\frac{1}{2}\|x_{k+1} - p\|^2 - \frac{1}{2}\|x_k - p\|^2 = -\frac{1}{2}\|x_{k+1} - x_k\|^2 + \langle x_{k+1} - x_k, x_{k+1} - p\rangle. \tag{5.32}$$

Take $\delta = \gamma\beta_f - 1$. Then, from (5.31) and (5.32), we deduce that

$$\begin{aligned} E_{k+1} - E_k \leq &-\left(\frac{\delta\beta_f}{h} + \frac{1}{2}\delta\right)\|x_{k+1} - x_k\|^2 \\ &- \langle x_{k+1} - p + \frac{\beta_f}{h}(x_{k+1} - x_k), h\beta_f \nabla f\,(y_k) + h\beta_f B(z_k)\rangle. \end{aligned}$$

Notice that $\nabla f(p)+B(p)=0$. Thus, we can rewrite the previous relation as follows

$$E_{k+1}-E_k \leq -\left(\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)\|x_{k+1}-x_k\|^2 - h\beta_f\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k), Y_k+Z_k\rangle,$$

where $Y_k=\nabla f(y_k)-\nabla f(p)$ and $Z_k=B(z_k)-B(p)$.
Since B is λ-cocoercive, we have

$$\begin{aligned}&\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k), Z_k\rangle = \langle z_k-p+(1+\frac{1}{h}(\beta_f-\beta_b))(x_{k+1}-x_k), B(z_k)-B(p)\rangle\\ &\geq \lambda\|B(z_k)-B(p)\|^2+(1+\frac{1}{h}(\beta_f-\beta_b))\langle (x_{k+1}-x_k), B(z_k)-B(p)\rangle\\ &= \lambda\|Z_k\|^2+(1+\frac{1}{h}(\beta_f-\beta_b))\langle (x_{k+1}-x_k), Z_k\rangle,\end{aligned}$$

Moreover, due to ∇f is $1/L$-cocoercive, we deduce that

$$\langle x_{k+1}-p+\frac{\beta_f}{h}(x_{k+1}-x_k), Y_k\rangle = \langle y_k-p, \nabla f(y_k)-\nabla f(p)\rangle \geq \frac{1}{L}\|\nabla f(y_k)-\nabla f(p)\|^2. \tag{5.33}$$

$$E_{k+1}-E_k \leq -\left(\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)\|x_{k+1}-x_k\|^2 \tag{5.34}$$

$$+\left(-h\beta_f-\beta_f(\beta_f-\beta_b)\right)\langle x_{k+1}-x_k, Z_k\rangle - h\beta_f\lambda\|Z_k\|^2-\frac{h\beta_f}{L}\|Y_k\|^2. \tag{5.35}$$

Equivalently,

$$E_{k+1}-E_k+\frac{h\beta_f}{L}\|Y_k\|^2+\mathcal{S}_k \leq 0, \tag{5.36}$$

where $S_k=\left(\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)\|x_{k+1}-x_k\|^2+\left(h\beta_f+\beta_f(\beta_f-\beta_b)\right)\langle x_{k+1}-x_k, Z_k\rangle + h\beta_f\lambda\|Z_k\|^2$.
Our goal here is to find $h>0$ such that $S_k>0$. Let us observe that $q:\mathcal{H}\times\mathcal{H}\to\mathbb{R}$ is a quadratic form

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

with

$$\begin{aligned}a &= \frac{\delta\beta_f}{h}+\frac{1}{2}\delta,\\ b &= h\beta_f+\beta_f(\beta_f-\beta_b),\\ g &= h\beta_f\lambda.\end{aligned}$$

Then, $S_k=q(x_{k+1}-x_k, Z_k)>0$ if $4ag-b^2>0$. One has,

$$\begin{aligned}4ag-b^2 &= 4\left(\frac{\delta\beta_f}{h}+\frac{1}{2}\delta\right)h\beta_f\lambda-\left(h\beta_f+\beta_f(\beta_f-\beta_b)\right)^2\\ &= 4\left(\delta\beta_f+\frac{1}{2}h\delta\right)\beta_f\lambda-\left(h\beta_f+\beta_f(\beta_f-\beta_b)\right)^2.\end{aligned}$$

Hence, $\lim_{h\to 0^+}(4ag-b^2)=\beta_f^2(4\lambda\delta-(\beta_f-\beta_b)^2)>0$ since $4\lambda\delta>(\beta_f-\beta_b)^2$. This implies there exists $h^*>0$ such that for all $h\in(0,h^*)$, we have $S_k>0$.
Therefore, under the above condition, and by taking h sufficiently small, there exist positive real numbers μ and η such that for all $k\geq 1$,

$$E_{k+1}-E_k+\mu\|x_{k+1}-x_k\|^2+\mu\|B(z_k)-B(p)\|^2+\eta\|\nabla f(y_k)-\nabla f(p)\|^2 \leq 0. \tag{5.37}$$

The rest of the proof is analogous to Theorem 5.1's one, so we omit it.

6 Numerical illustrations

The main purpose of this section is to implement our algorithms to numerically compute the trajectory of the dynamical system (iDINAM). For further applications, we refer the reader to [3], [4]. Before we start, let us recall a useful remark.

Remark 6.1 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.

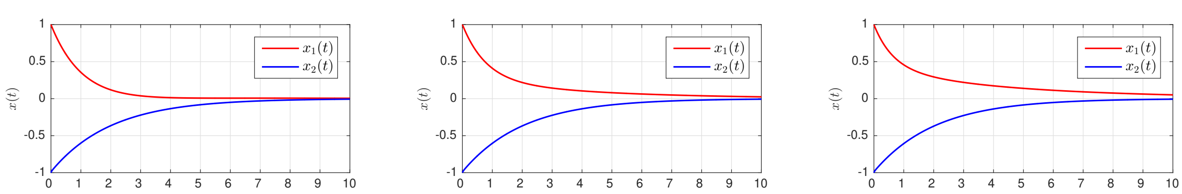
Example 6.1 Let us start this section by a simple illustrative example in $\mathbb{R}^2$. We take $\mathcal{H} = \mathbb{R}^2$ equipped with the usual Euclidean structure. Let us consider B as a linear operator whose matrix in the canonical basis of $\mathbb{R}^2$ is defined by $B = A_\lambda$ for $\lambda = 5$. According to the above remark, 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 $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x_1, x_2) = 10x_2^2.$$

It is clear that f is convex but not strongly convex. We set $\gamma = 0.9$ and consider the dynamical system (iDINAM) which γ, f, and B defined as before. As a straight application of Theorem 4.1, we obtain that the trajectory $x(t)$ generated by (iDINAM) converges to x_∞, where $x_\infty \in S = (B + \nabla f)^{-1}(0) = \{0\}$ whenever the positive parameters β_b, β_f satisfy

$$\gamma\beta_f > 1 \text{ and } \lambda > \frac{(\beta_b - \beta_f)^2}{4(\gamma\beta_f - 1)}.$$

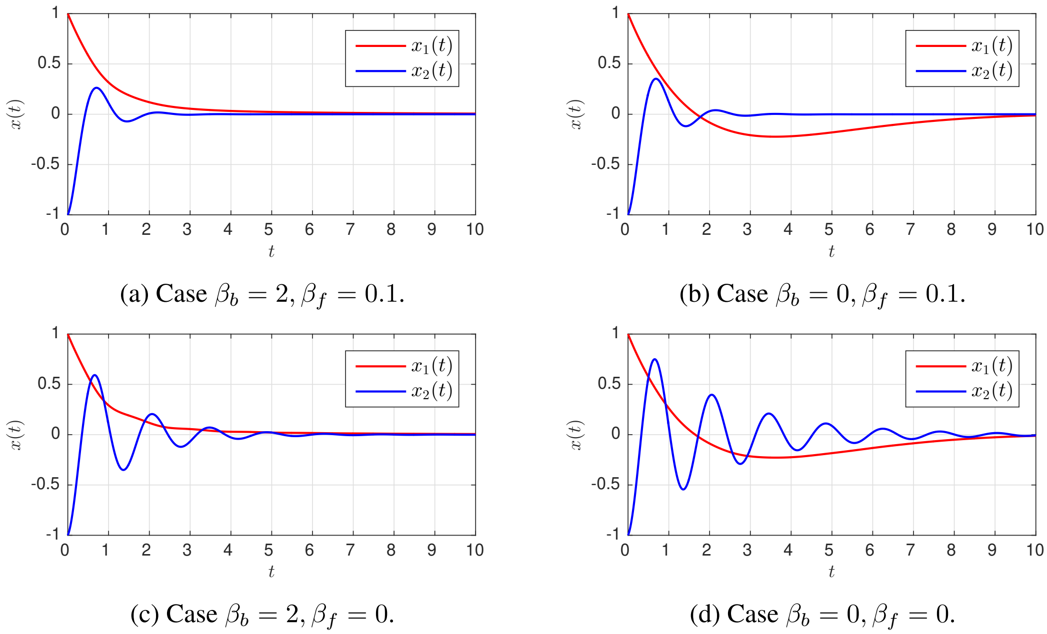
The trajectory obtained by using Matlab is depicted in Figure 1, where we represent the components $x_1(t)$ and $x_2(t)$ in red and blue respectively.



(a) Case $\beta_b = 1, \beta_f = 2$. (b) Case $\beta_b = 2, \beta_f = 2$. (c) Case $\beta_b = 3, \beta_f = 2$.

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

Now we study the behavior of the trajectories by considering more different values of β_b and β_f. We study four more different cases where the plots of the solutions have been depicted in Figure 2. Through Figures 1 and 2, we can conclude that by introducing the Hessian damping ($\beta_f > 0$), the oscillations of the trajectories in Figure 2 are attenuated. The oscillations of the solutions appear whenever β_f goes to 0. It is depicted clearly in Figure 3.



(a) Case $\beta_b = 2, \beta_f = 0.1$. (b) Case $\beta_b = 0, \beta_f = 0.1$.

(c) Case $\beta_b = 2, \beta_f = 0$. (d) Case $\beta_b = 0, \beta_f = 0$.

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

Example 6.2 In [3], it is considered the dynamical system

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

It is shown that under certain conditions on the parameters, namely $\beta_f > 0$ and

$$4\lambda\gamma > \frac{(\beta_b - \beta_f)^2}{\beta_f} + 2\left(\beta_b + \frac{1}{\gamma}\right) + 2\sqrt{\left(\beta_b + \frac{1}{\gamma}\right)^2 + \frac{(\beta_b - \beta_f)^2}{\gamma\beta_f}}, \tag{6.1}$$

then any trajectory generated by (DINAM) converges weakly, and its limit belongs to the solution set $S = (\nabla f + B)^{-1}$. Morover, in [4], the authors proposed some algorithms to find the zeros of $\nabla f + B$. Since our article provides similar results, it is interesting to compare these different types of algorithms. Following the same setting on B and γ as in the previous example and replacing f by $f(x) = 5x_1^2 + 10x_2^2$, let us compare their numerical performance.

In Figure 4, we show the objective function for each iteration k when we apply our algorithms including 2 new ones and (DINAAM-split) proposed in [4]. We can see that (iDIAAM-split) and (iDIAAM-var) gave the same numerical results while (DINAAM-split) did better in the long term in this case. A comparison between explicit algorithm (iDINAAM) and implicit one (DINAAM), is done in Figure 4. We note that, by introducing the implicit terms in both operators ∇f and B, we obtain a new algorithm for finding the zeros of $\nabla f + B$.

Example 6.3 Let us return to Example 6.1 and consider the effect of the introduction of perturbations, errors. With the same numerical values of the involved parameters, we just add the errors $e_k = \frac{1}{k^2}$. Clearly, the errors (e_k) satisfy the assumptions of Theorem 5.2. Running algorithm (iDINAAM-pert) in Matlab, the plot of $\|\nabla f(x_k) + B(x_k)\|$ versus k is depicted in Figure 5. To give a link with the analogous algorithm presented in [4], we applied (DINAAM-split-pert) in our numerical experiment. From Figure 5, we can see that algorithm (iDINAAM-pert) behaves as well as the nonperturbed version and gives almost the same numerical results as (DINAAM-split-pert) does.

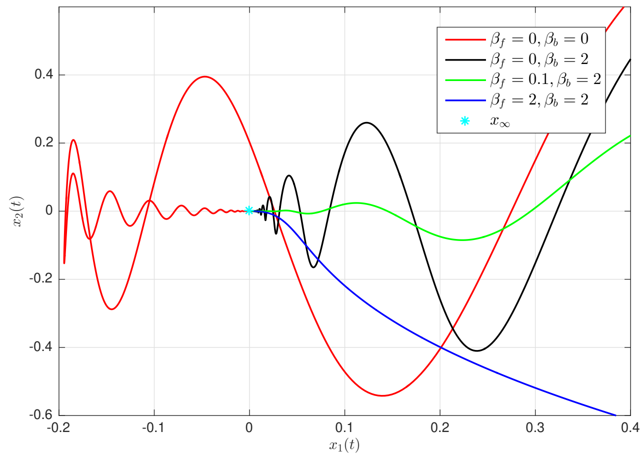
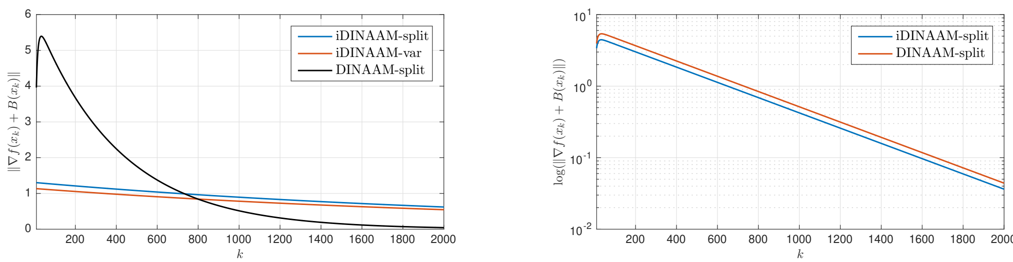


Figure 3: The attenuation of the oscillation by introducing the Hessian damping ($\beta_f > 0$).



(a) Numerical results obtained by algorithms with $h = 5.10^{-3}$

(b) The equivalance of (iDINAAM) and (DINAAM).

Figure 4: The numerical performance of algorithms to find the zeros of $\nabla f + B$.

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).

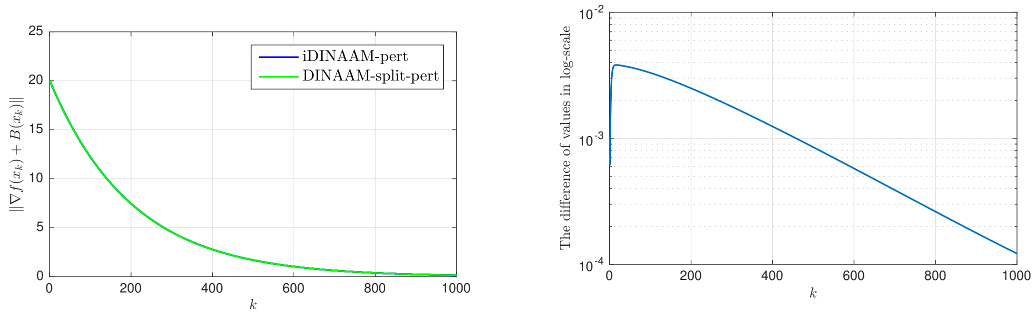
[2] Adly, S., Attouch, H. Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping. SIAM J. Optim. 30 (2020), no. 3, 2134-2162.

[3] Adly, S., Attouch, H., Vo V.N. Asymptotic behavior of Newton-like inertial dynamics involving the sum of potential and nonpotential terms. Fixed Point Theory Algorithms Sci. Eng. , Paper No. 17, 30 pp, (2021).

[4] Adly, S., Attouch, H., Vo V.N. Newton-type inertial algorithms for solving monotone equations governed by sums of potential and nonpotential operators. Appl. Math. Optim. 85, no. 3, Paper No. 31, (2022).

[5] Alecsa, C.D., László, S., Pinta, T. An extension of the second order dynamical system that models Nesterov's convex gradient method. Appl. Math. Optim. 84, no. 2, 1687-1716, (2021).

[6] Alvarez, F., On the minimizing property of a second order dissipative system in Hilbert space. SIAM J. Control Optim. 38 , no. 4, 1102-1119, (2000).



(a) Norm of the objective function.

(b) The difference of the objective function obtained by two algorithms.

Figure 5: The effect of perturbations, errors in the algorithm (iDINAAM).

[7] Apidopoulos V., Aujol J.-F. , Dossal Ch., The differential inclusion modeling FISTA algorithm and optimality of convergence rate in the case $b \leq 3$, SIAM J. Optim., **28** (2018), 551-574.

[8] 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).

[9] Attouch H., Cabot A. Convergence rates of inertial forward-backward algorithms, SIAM J. Optim., **28** (1) (2018), 849–874.

[10] 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).

[11] Attouch, H., Chbani, Z., Peypouquet, J., Redont, P. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, Math. Program. Ser. B, **168**, 123–175 (2018).

[12] Attouch, H., Chbani, Z., Riahi, H., Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$, ESAIM Control Optim. Calc. Var. 25 (2019), Paper No. 2, 34 pp.

[13] Attouch, H., Chbani, Z., Fadili, J., Riahi, H. First-order algorithms via inertial systems with Hessian driven damping. Math. Program. 193 (2022), no. 1, Ser. A, 113-155.

[14] Attouch, H., Chbani, Z., Fadili, J., Riahi, H. Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping, Optimization (2021), https//doi.org/10.1080/02331934.2021.2009828. preprint available at HAL Id hal-03285272

[15] Attouch H., Chbani Z., Fadili J., Riahi H. Fast convergence of dynamical ADMM via time scaling of damped inertial dynamics, accepted in JOTA, doi 10.1007/s10957-021-01859-2, preprint available at arXiv2103.12675v1 [math.OC] 23 Mar 2021.

[16] Attouch, H., Fadili, J., Kungurtsev, V. On the effect of perturbations in first-order optimization methods with inertia and Hessian driven damping, EECT, (2022), doi 10.3934/eect.2022022, preprint available at arXiv2106.16159v2 [math.OC].

[17] Attouch H., Fadili J. From the Ravine method to the Nesterov method and vice versa A dynamical system perspective, arXiv2201.11643v1 [math.OC] (2022).

[18] Attouch, H., László, S.C. Continuous Newton-like Inertial Dynamics for Monotone Inclusions, Set-Valued Var. Anal. 29 (2021), no. 3, 555-581.

[19] Attouch, H., László, S.C. Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. SIAM J. Optim. 30 (2020), no. 4, 3252-3283.

[20] 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).

[21] 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. of Convex Analysis **23**(1), 139–180 (2016).

[22] Attouch, H., Peypouquet, J. Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators. Math. Program. **174**(1-2), 391–432 (2019).

[23] Attouch, H., Peypouquet, J., Redont, P. Fast convex minimization via inertial dynamics with Hessian driven damping. J. Differential Equations **261**(10), 5734–5783 (2016).

[24] 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).

[25] Attouch, H., Svaiter, B.F. A continuous dynamical Newton-like approach to solving monotone inclusions. SIAM J. Control Optim. **49**(2), 574–598 (2011).

[26] Baillon, J.-B, Haddad, G., Quelques propriétés des opérateurs angles-bornés et n-cycliquement monotones. Israel J. Math. 26 (1977) 137-150.

[27] Bauschke, H., Combettes, P.L. Convex Analysis and Monotone Operator Theory in Hilbert spaces. CMS Books in Mathematics. Springer (2011).

[28] Beck A., Teboulle M.. A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci., **2** (1) (2009), 183–202.

[29] Boţ, R.I., Csetnek, E.R. Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control Optim. **54**, 1423–1443 (2016).

[30] Boţ, R.I., Csetnek, E.R., László, S.C. Tikhonov regularization of a second order dynamical system with Hessian damping. Math. Program. 189 (2021), no. 1-2, Ser. B, 151-186.

[31] Brézis, H. Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution. Lecture Notes 5. North Holland (1972).

[32] Castera, C., Bolte, J., Févotte, C., Pauwels, E. An Inertial Newton Algorithm for Deep Learning. (2019), HAL-02140748.

[33] Chambolle A., Dossal Ch. On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm, J. Optim. Theory Appl., **166** (2015), 968–982.

[34] Gelfand I.M. , Tsetlin M. Printszip nelokalnogo poiska v sistemah avtomatich, Optimizatsii, Dokl. AN SSSR, 137 (1961), pp. 295–298 (in Russian).

[35] Kim, D. Accelerated proximal point method for maximally monotone operators. Math. Program., Series A, (accepted)(2020), preprint available at arXiv1905.05149v3

[36] Lin, T., Jordan, M.I. A Control-Theoretic Perspective on Optimal High-Order Optimization. (2019), arXiv1912.07168v1

[37] Muehlebach, M., Jordan, M. I. A Dynamical Systems Perspective on Nesterov Acceleration. Proc. of the International Conference on Machine Learning, 2019, https//arxiv.org/abs/1905.07436

[38] Polyak B.T. Some methods of speeding up the convergence of iteration methods, U.S.S.R. Comput. Math. Math. Phys., **4** (1964), 1–17.

[39] 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)

[40] 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.

[41] Suvrit Sra. Optimization for Machine Learning Subgradient method; Accelerated gradient. Massachusetts Institute of Technology, March 2021.

[42] Villa S., Salzo S. , Baldassarres L., Verri A. Accelerated and inexact forward-backward, SIAM J. Optim., **23** (3) (2013), 1607–1633.

[43] Shi, B., Du, S.S., Jordan, M.I., Su, W.J. Understanding the acceleration phenomenon via high-resolution differential equations. (2018), arXivsubmit/2440124[cs.LG]

[44] Su, W., Boyd, S., Candès, E. J., A differential equation for modeling Nesterov's accelerated gradient method. J. Mach. Learn. Res. 17 (2016), Paper No. 153, 43 pp.