Finite time stabilization of continuous inertial dynamics combining dry friction with Hessian-driven damping

Samir Adly, Hedy Attouch

Finite time stabilization of continuous inertial dynamics combining dry friction with Hessian-driven damping

Samir ADLY[*] and Hedy ATTOUCH[†]

June 8, 2020

This paper is dedicated to Professor Umberto Mosco on the occasion of his 80th birthday.

ABSTRACT. In a Hilbert space $\mathcal{H}$, we study the stabilization in finite time of the trajectories generated by a continuous (in time t) damped inertial dynamic system. The potential function $f : \mathcal{H} \to \mathbb{R}$ to be minimized is supposed to be differentiable, not necessarily convex. It enters the dynamic via its gradient. The damping results from the joint action of dry friction, viscous friction, and a geometric damping driven by the Hessian of f. The dry friction damping function $\phi : \mathcal{H} \to \mathbb{R}_+$, which is convex and continuous with a sharp minimum at the origin (typically $\phi(x) = r\|x\|$ with $r > 0$), enters the dynamic via its subdifferential. It acts as a soft threshold operator on the velocities, and is at the origin of the stabilization property in finite time. The Hessian driven damping, which enters the dynamics in the form $\nabla^2 f(x(t))\dot{x}(t)$, permits to control and attenuate the oscillations which occur naturally with the inertial effect. We give two different proofs, in a finite dimensional setting, of the existence of strong solutions of this second-order differential inclusion. One is based on a fixed point argument and Leray-Schauder theorem, the other one uses the Yosida approximation technique together with the Mosco convergence. We also give an existence and uniqueness result in a general Hilbert framework by assuming that the Hessian of the function f is Lipschitz continuous on the bounded sets of $\mathcal{H}$. Then, we study the convergence properties of the trajectories as $t \to +\infty$, and show their stabilization property in finite time. The convergence results tolerate the presence of perturbations, errors, under the sole assumption of their asymptotic convergence to zero. The study is extended to the case of a nonsmooth convex function f by using Moreau's envelope.

Mathematics Subject Classifications: 37N40, 34A60, 34G25, 49K24, 70F40.

Key words and phrases: damped inertial dynamics; differential inclusion; dry friction; Hessian-driven damping; finite time stabilization;

1 Introduction and preliminary results

Throughout the paper $\mathcal{H}$ is a real Hilbert space, with the scalar product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$, and $f : \mathcal{H} \to \mathbb{R}$ is a $\mathcal{C}^1$ function (not necessarily convex) whose gradient is Lipschitz continuous. In the introduction, when we consider continuous dynamics in which the Hessian intervenes, to simplify the presentation, we can assume that f is a $\mathcal{C}^2$ function.
We call $t_0 \in \mathbb{R}$ the origin of time. Since we consider autonomous systems, we can take an arbitrary real number for t_0.

[*]Laboratoire XLIM, Université de Limoges, 87060 Limoges, France. E-mail: samir.adly@unilim.fr
[†]IMAG, Université Montpellier, CNRS, 34095 Montpellier CEDEX 5, France. E-mail: hedy.attouch@umontpellier.fr

1.1 Presentation of the dynamic

We will analyze the finite time stabilization of the following second-order differential inclusion

$$\text{(IGDH)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,\ t \in [t_0, +\infty[. \tag{1.1}$$

Here, (IGDH) stands shortly for Inertial Gradient system with Dry friction and Hessian-driven damping.

Let us briefly describe the three types of damping that occur in (IGDH), and which provide stabilization:

- $\gamma > 0$ is a positive viscous damping coefficient.
- The dry friction damping function $\phi : \mathcal{H} \to \mathbb{R}_+$ is convex and continuous with a sharp minimum at the origin, typically $\phi(x) = r\|x\|$ with $r > 0$. It is at the origin of the finite time stabilization. It enters the dynamic via its subdifferential, which is defined by

$$\partial\phi(x) = \{p \in \mathcal{H} \ : \ \langle p, y - x\rangle \leq \phi(y) - \phi(x),\ \ \forall y \in \mathcal{H}\}.$$

- The geometrical damping driven by the Hessian of f enters the dynamics in the form $\nabla^2 f(x(t))\dot{x}(t)$. It allows to attenuate and control the oscillations which naturally occur with inertial systems.
  In the dynamic (IGDH), $\beta > 0$ is the corresponding damping coefficient.

Let us mention that, from an abstract point of view, the viscous friction $\gamma\dot{x}(t)$ could enter the dynamic (IGDH) into two ways:

- since $\gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) = \partial(\frac{\gamma}{2}\|\cdot\|^2 + \phi)(\dot{x}(t))$, it is possible to introduce the new dry friction function $\psi = \frac{\gamma}{2}\|\cdot\|^2 + \phi$.
- since $\gamma\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) = \frac{d}{dt}\Big(\nabla\Big(\frac{\gamma}{2}\|\cdot\|^2 + \beta f\Big)\Big)(x(t))$, it is possible to combine the viscous friction with the damping driven by the Hessian.

As we have in mind optimization algorithms, and in order to place our results in the context of the literature on the subject, we will keep the three frictions separated in the dynamic (IGDH).

1.1.1 Link with optimization

Our main motivation for studying the continuous dynamic (IGDH) comes from optimization. Its temporal discretization provides algorithms which share its good convergence properties. As a key property, we use that $\nabla^2 f(x(t))\dot{x}(t)$ is the time derivative of $\nabla f(x(t))$. This gives, in discretized form, first-order algorithms, which we describe below. Given a constant time step $h > 0$, we consider the following temporal discretization of (IGDH)

$$\begin{aligned} &\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\gamma}{h}(x_{k+1} - x_k) + \partial\phi\left(\frac{1}{h}(x_{k+1} - x_k)\right) \\ &+\frac{\beta}{h}\left(\nabla f(x_k) - \nabla f(x_{k-1})\right) + \nabla f(x_k) \ni 0. \end{aligned} \tag{1.2}$$

In (1.2), the terms $\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1})$ and $\frac{1}{h}(x_{k+1} - x_k)$ respectively represent the discrete acceleration, and the discrete velocity. According to $\nabla^2 f(x(t))\dot{x}(t) = \frac{d}{dt}\nabla f(x(t))$, the correcting term $\frac{1}{h}(\nabla f(x_k) - \nabla f(x_{k-1}))$ is directly linked to the temporal discretization of the Hessian-driven damping term. By solving (1.2) with respect to x_{k+1}, we obtain the following first-order algorithm where the dry friction enters via the proximal operator of ϕ, and the function to minimize f enters via its gradient:

$$\text{(IPAHDD)} \begin{cases} z_k = \frac{1}{h(1+h\gamma)}(x_k - x_{k-1}) - \frac{\beta}{1+h\gamma}(\nabla f(x_k) - \nabla f(x_{k-1})) - \frac{h}{1+h\gamma}\nabla f(x_k) \\ x_{k+1} = x_k + h \operatorname{prox}_{\frac{h}{1+h\gamma}\phi}(z_k). \end{cases}$$

In the above algorithm, (IPAHDD) stands shortly for Inertial Proximal-gradient Algorithm with Hessian-Damping and Dry friction. Note that the temporal discretization (1.2) of (IGDH) is implicit with respect to the nonsmooth function ϕ, and explicit with respect to the smooth function f. In the above algorithm (IPAHDD), prox_ϕ denotes the proximal mapping associated with the convex function ϕ. Recall that, for any $x \in \mathcal{H}$, for any $\lambda > 0$

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

The proximal mapping can be equivalently formulated as the resolvent of the maximally monotone operator $\partial\phi : \mathcal{H} \to 2^{\mathcal{H}}$, that is, for any $\lambda > 0$

$$\operatorname{prox}_{\lambda\phi} = (I + \lambda\partial\phi)^{-1}.$$

In [3], it is proved the finite convergence of the sequences (x_k) generated by the (IPAHDD) algorithm. Indeed, for continuous optimization algorithms, convergence in a finite number of steps, with the estimation of this number, is a very favorable property. The limit x_∞ of the sequence (x_k) satisfies

$$\nabla f(x_\infty) + \partial\phi(0) \ni 0.$$

Thus, x_∞ is an “approximate” critical point of f. In practice, we choose ϕ with $\partial\phi(0)$ ”small”. By taking $\phi(x) = r\|x\|$, this means taking a small $r > 0$. This amounts to solving the optimization problem $\min_{\mathcal{H}} f$ with the variational principle of Ekeland, instead of the classical Fermat rule.

(IPAHDD) is a splitting algorithm in the sense that the two constitutive potential functions f and ϕ are treated separately, via the gradient and the proximal mapping respectively. Since the proximal mapping of ϕ can be easily calculated in most practical situations, this makes (IPAHDD) a first-order algorithm that can be used to deal with problems of large size.

The continuous dynamic (IGDH) is the basis of these algorithmic developments. Its mechanical interpretation naturally suggests Lyapunov functions. This therefore leads us to study (IGDH) in depth, and explore some of its extensions. This is the main purpose of this article. For the numerical aspects and the convergence analysis of (IPAHDD), as well as several other algorithms, we refer to [3].

1.1.2 Link with mechanics, damped shocks

Another motivation for the study of (IGDH) comes from mechanics, and the modeling of damped shocks. In [18], Attouch-Maingé-Redont consider the second-order differential system with Hessian-driven damping

$$\ddot{x}(t) + \gamma\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) + \nabla g(x(t)) = 0, \tag{1.3}$$

where $g : \mathcal{H} \to \mathbb{R}$ is a smooth real-valued function. An interesting property of this system is that, after the introduction of an auxiliary variable y , it can be equivalently written as a first-order system involving only the time derivatives $\dot{x}(t)$, $\dot{y}(t)$ and the gradient terms $\nabla f(x(t))$, $\nabla g(x(t))$. More precisely, the system (1.3) is equivalent to the following first-order differential equation

$$\begin{cases} \dot{x}(t) + \beta\nabla f(x(t)) + ax(t) + by(t) = 0, \\ \dot{y}(t) - \beta\nabla g(x(t)) + ax(t) + by(t) = 0, \end{cases} \tag{1.4}$$

where a and b are real numbers such that: $a + b = \gamma$ and $\beta b = 1$. Note that (1.4) is different from the classical Hamiltonian formulation, which would still involve the Hessian of f. In contrast, the formulation (1.4) uses only first-order information from the function f (no occurence of the Hessian of f). Replacing ∇f by ∂f in (1.4) allows us to extend the analysis to the case of a convex lower semicontinuous function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, and so to introduce constraints in the model. When $f = \delta_K$ is the indicator function of a closed convex set $K \subset \mathcal{H}$, the subdifferential operator ∂f takes account of the contact forces, while ∇g takes account of the driving forces. In this setting, by playing with the geometrical damping parameter β, one can describe nonelastic shock laws with restitution coefficient (for more details we refer to [18] and references therein). Introducing the dry friction into this dynamic would allow to obtain finite type stabilization of damped oscillating systems with obstacle constraint.

1.2 Some historical facts

Let us explain the role and the importance of each of the three damping terms which enter into the continuous dynamics (IGDH). Our main contribution in this article is to show how to combine them to obtain fast stabilization properties.

1.2.1 Viscous friction

The use of inertial dynamics to accelerate the gradient method in optimization was first considered by B. Polyak in [35]. Based on the inertial system with a fixed viscous damping coefficient $\gamma > 0$

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

he introduced the Heavy Ball with Friction method. This system was further developed by Attouch-Goudou-Redont [17] as a tool to explore the local minima of f. For a strongly convex function f, and γ judiciously chosen, (HBF) provides convergence at exponential rate of $f(x(t))$ to $\min_{\mathcal{H}} f$. For a general convex function f, the asymptotic convergence rate of (HBF) is $\mathcal{O}(\frac{1}{t})$ (in the worst case). This is however not better than the steepest descent. A decisive step to obtain a faster asymptotic convergence was taken by Su-Boyd-Candès [39] with the introduction of an Asymptotic Vanishing Damping coefficient of the form $\gamma(t) = \frac{\alpha}{t}$, with $\alpha > 0$. More precisely, they considered the dynamic

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

As a specific feature, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes (tends to zero) as time t goes to infinity, hence the terminology. For a general convex function f, it provides a continuous version of the accelerated gradient method of Nesterov. For $\alpha \geq 3$, each trajectory $x(\cdot)$ of (AVD)_α satisfies the asymptotic convergence rate of the values $f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(1/t^2\right)$ as $t \to +\infty$. The convergence properties of the dynamic (AVD)_α have been the subject of many recent studies, see [7, 10, 11, 12, 14, 16, 19, 22, 23, 32, 39]. The case $\alpha = 3$, which corresponds to Nesterov's historical algorithm, is critical. In the case $\alpha = 3$, the question of the convergence of the trajectories remains an open problem (except in one dimension where convergence holds [16]). For $\alpha > 3$, it has been shown by Attouch-Chbani-Peypouquet-Redont [14] that each trajectory converges weakly to a minimizer. The corresponding algorithmic result has been obtained by Chambolle-Dossal [30]. For $\alpha > 3$, it is shown in [19] and [32] that the asymptotic convergence rate of the values is actually $o(1/t^2)$. The subcritical case $\alpha \leq 3$ has been examined by Apidopoulos-Aujol-Dossal [7] and Attouch-Chbani-Riahi [16], with the convergence rate of the objective values $\mathcal{O}\left(t^{-\frac{2\alpha}{3}}\right)$. These rates are optimal, that is, they can be reached, or approached arbitrarily close.

1.2.2 Dry friction

Some first results concerning the finite convergence property under the action of dry friction have been obtained by Adly-Attouch-Cabot [4] for the continuous dynamics

$$\ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni 0, \ t \in [t_0, +\infty[. \tag{1.5}$$

Assuming that the potential friction function ϕ has a sharp minimum at the origin (dry friction), they showed that, generically with respect to the initial data, the solution trajectories converge in finite time to equilibria. Similar results for the corresponding proximal-based algorithms have been obtained by Baji-Cabot [24] and Adly-Attouch [2].
Let's make precise the tools that will be useful for the mathematical analysis of the set-valued term $\partial\phi(\dot{x}(t))$ in (1.5) which models dry friction. We say that the friction potential function ϕ has the Dry Friction property (denoted by (DF)) if it satisfies the following properties

$$\text{(DF)} \qquad \begin{cases} \phi : \mathcal{H} \to \mathbb{R} \text{ is convex continuous;} \\ \min\limits_{\xi \in \mathcal{H}} \phi(\xi) = \phi(0) = 0; \\ 0 \in \operatorname{int}(\partial\phi(0)). \end{cases}$$

The particular case $\phi = r\|\cdot\|$, with $r > 0$, models dry friction (also called Coulomb friction) in mechanics. The key assumption $0 \in \operatorname{int}(\partial\phi(0))$ expresses that ϕ has a sharp minimum at the origin. This is specified in the following elementary lemma, see [1, Lemma 4.1 page 83], where, in item (iv), ϕ^* is the Fenchel conjugate of ϕ.

Lemma 1.1 *Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function such that $\min_{\xi\in\mathcal{H}} \phi(\xi) = \phi(0) = 0$. Then, the following formulations of the dry friction are equivalent:*

(i) $0 \in \operatorname{int}(\partial\phi(0))$*;*

(ii) *there exists some $r > 0$ such that $B(0, r) \subset \partial\phi(0)$;*

(iii) *there exists some $r > 0$ such that, for all $\xi \in \mathcal{H}$, $\phi(\xi) \geq r\|\xi\|$.*

(iv) *there exists some $r > 0$ such that, $\|g\| \leq r \Longrightarrow \partial\phi^*(g) \ni 0$.*

The positive parameter $r > 0$ will play a crucial role in the analysis of the corresponding numerical optimization algorithms (we refer to [2, 3] for more details). To enlighten its role, we will say that the friction potential function ϕ satisfies the property (DF)_r if ϕ satisfies the Dry Friction property (DF) with $B(0, r) \subset \partial\phi(0)$. The property (iv) above expresses that, when the force g exerted on the system is less than a threshold $r > 0$, then the system stabilizes, *i.e.* the velocity $v = 0 \in \partial\phi^*(g)$. This contrasts with the viscous damping that can asymptotically produce many small oscillations.

1.2.3 Hessian-driven damping

The inertial system

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

was introduced in [6]. In the same spirit as (HBF), the dynamic $\text{(DIN)}_{\gamma,\beta}$ contains a *fixed* positive friction coefficient $\gamma > 0$. The introduction of the Hessian-driven damping allows to damp the transversal oscillations that might arise with (HBF), as observed in [6] in the case of the Rosenbrook function. The need to take a geometric damping adapted to f had already been observed by Alvarez [5] who considered the inertial system

$$\ddot{x}(t) + \Gamma\dot{x}(t) + \nabla f(x(t)) = 0,$$

where $\Gamma : \mathcal{H} \to \mathcal{H}$ is a linear positive anisotropic operator. But still this damping operator is fixed. For a general convex function, the Hessian-driven damping in (DIN)$_{\gamma,\beta}$ performs a similar operation in a closed-loop adaptive way. The terminology (DIN) stands shortly for Dynamical Inertial Newton. It refers to the natural link between this dynamic and the continuous Newton method, see Attouch-Svaiter [21]. Recent studies have been devoted to the study of the inertial dynamic

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

which combines asymptotic vanishing damping with Hessian-driven damping. The corresponding algorithms involve a correcting term in the Nesterov accelerated gradient method which reduces the oscillatory aspects, see Attouch-Peypouquet-Redont [20], Attouch-Chbani-Fadili-Riahi [13], Shi-Du-Jordan-Su [38].

1.3 Mosco convergence and Attouch theorem

In 1969, U. Mosco [33] introduced a fundamental concept of convergence for sequence of closed convex sets in reflexive Banach spaces, that can be transposed to functions via their epigraphs. This concept, called the Mosco convergence, is widely applicable in convex optimization, stability of variational inequalities and optimal control. Precisely, a sequence $(\varphi_n) \in \Gamma_0(\mathcal{H})$ of convex proper and lower semicontinuous functions, is declared to be Mosco-convergent to $\varphi \in \Gamma_0(\mathcal{H})$ provided, (i) and (ii) are satisfied:

(i) for all $x \in \mathrm{dom}\varphi$, there exists a sequence (x_n) strongly convergent to x s.t. $\lim_{n\to+\infty} \varphi_n(x_n) = \varphi(x)$

(ii) whenever (x_n) converges weakly to x, we have $\liminf_{n\to+\infty} \varphi_n(x_n) \geq \varphi(x)$.

An other important result was given by U. Mosco in [34] where the bicontinuity of the Fenchel conjugate is established: for any sequence of convex functions (φ_n) we have the equivalence:

$$\varphi_n \text{ is Mosco-convergent to } \varphi \Longleftrightarrow (\varphi_n^*) \text{ is Mosco-convergent to } \varphi^*.$$

This notion was used in 1977 by H. Attouch (see Theorem 3.66 in [8] page 373) to establish that a sequence $(\varphi_n) \subset \Gamma_0(\mathcal{H})$ converges in the sense of Mosco to $\varphi \in \Gamma_0(\mathcal{H})$ if and only if the sequence $(\partial\varphi_n)$ converges in the sense of Painlevé-Kuratowski to $\partial\varphi$, provided some normalization condition is satisfied. Here, $\partial\varphi_n$ (resp. $\partial\varphi$) is the convex subdifferential of φ_n (resp. φ), and the operators are identified with their graphs in $\mathcal{H} \times \mathcal{H}$.

1.4 Contents

The paper is organized as follows. In Sections 2 and 3, we give two different proofs of the existence of solutions to (IGDH) in a finite dimensional setting: one is based on a fixed point argument and Leray-Schauder theorem, the other one uses the Yosida approximation together with the Mosco convergence. Section 4 is devoted to an existence and uniqueness result for the Cauchy problem associated with (IGDH) in an infinite dimensional Hilbert space. In Section 5, we present our main results, which concern the convergence in finite time of the trajectories generated by the inertial dynamic (IGDH). In Section 6, we examine the effect of the introduction of perturbations, errors in the dynamic (IGDH). In Section 7, based on the variational properties of Moreau's envelope, we extend these results to the case where $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous and proper function such that $\inf f > -\infty$. Finally, Section 8 concludes and gives some perspectives about future work.

2 Existence and Uniqueness for the Cauchy problem

2.1 The notion of strong solution of (IGDH)

Unless specified, we make the following assumptions on the potential functions f and ϕ:

- $f : \mathcal{H} \to \mathbb{R}$ is a differentiable function whose gradient is Lipschitz continuous;
- $\phi : \mathcal{H} \to \mathbb{R}$ is a convex continuous function which satisfies (DF) as well as the following property: there exists a constant $c > 0$ such that for all $x \in \mathcal{H}$

$$\|\partial\phi(x)\| \leq c(1 + \|x\|),$$

where the above formula means: for any $\xi \in \partial\phi(x)$, we have $\|\xi\| \leq c(1 + \|x\|)$.

Definition 2.1 *For a given $(t_0, x_0, \dot{x}_0) \in \mathbb{R} \times \mathcal{H} \times \mathcal{H}$, we say that $x : [t_0, +\infty[\to \mathcal{H}$ is a strong global solution of* (IGDH) *with Cauchy data $x(t_0) = x_0$, and $\dot{x}(t_0) = \dot{x}_0$, if the following properties 1 and 2 are satisfied:*

1. *For all $T > t_0$, $x \in \mathbb{H}^2(t_0, T; \mathcal{H})$. Equivalently, $x \in \mathbb{L}^2(t_0, T; \mathcal{H})$, $\dot{x} \in \mathbb{L}^2(t_0, T; \mathcal{H})$ and $\ddot{x} \in \mathbb{L}^2(t_0, T; \mathcal{H})$.*
2. *The following inclusion is satisfied: for almost all $t \geq t_0$*

$$\ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\frac{d}{dt}\Big(\nabla f(x(t))\Big) + \nabla f(x(t)) \ni 0, \tag{2.1}$$

with $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$.

Let us comment the above definition. First note that, since ∇f is Lipschitz continuous (say with constant L), we have for all $s, t \geq t_0$

$$\|\nabla f(x(t)) - \nabla f(x(s))\| \leq L\|x(t) - x(s)\|. \tag{2.2}$$

Since $x \in \mathbb{H}^1(t_0, T; \mathcal{H})$, x is absolutely continuous. According to the inequality (2.2), this implies that $t \mapsto \nabla f(x(t))$ is also absolutely continuous. Therefore, $t \mapsto \nabla f(x(t))$ is almost everywhere differentiable. Using again (2.2), we get

$$\|\frac{d}{dt}\Big(\nabla f(x(t))\Big)\| \leq L\|\dot{x}(t)\|.$$

According to the fact that $\dot{x} \in \mathbb{L}^2(t_0, T; \mathcal{H})$, it results

$$\frac{d}{dt}\Big(\nabla f(x(\cdot))\Big) \in \mathbb{L}^2(t_0, T; \mathcal{H}),$$

where $\frac{d}{dt}\Big(\nabla f(x(t))\Big)$ can be taken indifferently in the distribution or pointwise sense (the two notions coincide for an absolutely continuous function). Without ambiguity, we will write shortly

$$\frac{d}{dt}\Big(\nabla f(x(t))\Big) = \nabla^2 f(x(t))\dot{x}(t)$$

which is justified when f is twice differentiable.
In addition, we have

$$\|\partial\phi(\dot{x}(t))\| \leq c(1 + \|\dot{x}(t)\|),$$

where we briefly write $\partial\phi(\dot{x}(t))$ to designate a measurable selection $\xi(t)$ with $\xi(t) \in \partial\phi(\dot{x}(t))$ for almost every $t \geq t_0$. According to $x \in \mathbb{H}^1(t_0, T; \mathcal{H})$ this gives

$$\partial\phi(x(\cdot)) \in \mathbb{L}^2(t_0, T; \mathcal{H}).$$

Therefore, for all $T > t_0$, all the constitutive elements of (IGDH) belong to $\mathbb{L}^2(t_0, T; \mathcal{H})$. Finally note that since for all $T > t_0,\ x \in \mathbb{H}^2(t_0, T; \mathcal{H})$, we have that x and $\dot{x}$ are continuous functions of t, which allow to give a sense to the Cauchy data.

2.2 The case without the Hessian driven damping term, and with right handside

As a preliminary result, we consider (IGDH) in the case $\beta = 0$, *i.e.* without the Hessian driven damping term, but with a second member $e(\cdot)$. This is a preliminary step to study (IGDH) in the general case, using a fixed point argument. More precisely, we have the following equivalence, which is immediate to obtain, and which is just the Hamiltonian formulation of (IGDH).

Proposition 2.1 *The following are equivalent:* $(i) \iff (ii)$

$$(i) \quad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni e(t).$$

$$(ii) \quad \begin{cases} \dot{x}(t) = u(t); \\ \dot{u}(t) \in -\nabla f(x(t)) - \gamma(t)u(t) + e(t) - \partial\phi(u(t)) \end{cases}$$

Proof. $(i) \implies (ii)$. Set $u(t) := \dot{x}(t)$. We have $\dot{u}(t) = \ddot{x}(t)$, which, by the constitutive equation, gives

$$\dot{u}(t) + \gamma(t)\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni e(t). \tag{2.3}$$

This gives the first-order system (ii).

$(ii) \implies (i)$. By differentiating the first equation of (ii) we obtain $\ddot{x}(t) = \dot{u}(t)$. According to the second equation of (ii) we obtain (i). ■

We can now state the following existence and uniqueness result.

Theorem 2.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a differentiable function whose gradient is Lipschitz continuous, and let $\phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function. Suppose that, for any $T > t_0$, the function $\gamma : [t_0, +\infty[\to \mathbb{R}_+$ belongs to $L^1(t_0, T; \mathbb{R})$, and $e : [t_0, +\infty[\to \mathcal{H}$ belongs to $L^2(t_0, T; \mathcal{H})$. Then, for any Cauchy data $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$, there exists a unique strong global solution of*

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni e(t) \tag{2.4}$$

satisfying $x(t_0) = x_0$, and $\dot{x}(t_0) = \dot{x}_0$.

Proof. According to Proposition 2.1, it is equivalent to solve the first-order system (ii) with the Cauchy data $x(t_0) = x_0,\ \ u(t_0) = \dot{x}_0$. Set

$$Z(t) = (x(t), u(t)) \in \mathcal{H} \times \mathcal{H}.$$

The system (ii) can be written equivalently as

$$\dot{Z}(t) + F(t, Z(t)) \ni 0, \quad Z(t_0) = (x_0, \dot{x}_0),$$

where $F : [t_0, +\infty[\times \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}, \;\; (t,x,u) \mapsto F(t,x,u)$ is defined by

$$F(t,x,u) = \Big(0, \partial\phi(u)\Big) + \Big(-u, \nabla f(x) + \gamma(t)u - e(t)\Big).$$

Hence F splits as follows

$$F(t,x,u) = F_1(x,u) + F_2(t,x,u)$$

where

$$F_1(x,u) = \partial\Phi(x,u)$$

is the subdifferential of the convex lower semicontinuous function (which depends only on u)

$$\Phi(x,u) = \phi(u)$$

and

$$F_2(t,x,u) = \Big(-u, \nabla f(x) + \gamma(t)u - e(t)\Big).$$

Therefore, the second-order dynamic (2.4) is equivalent to the following first-order differential inclusion

$$\dot{Z}(t) + \partial\Phi(Z(t)) + F_2(t, Z(t)) \ni 0, \quad Z(t_0) = (x_0, \dot{x}_0), \tag{2.5}$$

Since ∇f is Lipschitz continuous, so is F_2 with respect to $Z = (x,u)$. According to [28, Proposition 3.12] which considers the evolution equation governed by a Lipschitz perturbation (which can be time dependent as here) of a convex subdifferential, we deduce the existence and uniqueness of the solution to the Cauchy problem formulated as (ii), which proves the claim of Theorem 2.1. ■

2.3 The general case of (IGDH)

In this section, we assume that $\mathcal{H}$ is a finite dimensional Hilbert space.

Theorem 2.2 *Let $f : \mathcal{H} \to \mathbb{R}$ be a differentiable function whose gradient is L-Lipschitz continuous. Suppose that the dry potential friction $\phi : \mathcal{H} \to \mathbb{R}$ is a convex continuous function that satisfies* $(\mathrm{DF})_r$*, and $\|\partial\phi(x)\| \leq c(1 + \|x\|)$. Suppose that the damping parameters satisfy*

$$\gamma > \beta L.$$

Then, for any Cauchy data $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$, there exists a strong global solution $x : [t_0, +\infty[\to \mathcal{H}$ of

$$\text{(IGDH)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0 \tag{2.6}$$

satisfying the Cauchy data $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$.

Proof. Since no geometrical assumption is made on the function f (it is not assumed to be convex), we will treat the Hessian driven damping term as a perturbation. According to the growth conditions made on this term, we will apply a fixed point argument. To this end, let us write the equation (IGDH) in the following form

$$\ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \ni -\beta\nabla^2 f(x(t))\dot{x}(t). \tag{2.7}$$

Then, to solve it, we will apply the Leray-Schauder fixed point theorem in the space $\mathbb{H}^1(t_0, T; \mathcal{H})$, with $T \geq t_0$. Recall the definition of the Sobolev space

$$\mathbb{H}^1(t_0, T; \mathcal{H}) = \left\{ x \in \mathbb{L}^2(t_0, T; \mathcal{H}) \text{ such that } \dot{x} \in \mathbb{L}^2(t_0, T; \mathcal{H}) \right\},$$

where $\dot{x}$ is taken in the distribution sense. $\mathbb{H}^1(t_0,T;\mathcal{H})$ is equipped with the norm

$$\|x\|_{\mathbb{H}^1(t_0,T;\mathcal{H})} = \left(\|x\|^2_{\mathbb{L}^2(t_0,T;\mathcal{H})} + \|\dot{x}\|^2_{\mathbb{L}^2(t_0,T;\mathcal{H})}\right)^{\frac{1}{2}},$$

which makes it a Hilbert space, see [28, Appendix] for a detailed presentation of these spaces.
We can formulate equation (2.7) as

$$\mathcal{T}(x) = x$$

where $\mathcal{T} : \mathbb{H}^1(t_0,T;\mathcal{H}) \to \mathbb{H}^1(t_0,T;\mathcal{H})$ is the mapping which associates to $x \in \mathbb{H}^1(t_0,T;\mathcal{H})$ the unique solution $z = \mathcal{T}x$ of the evolution equation

$$\ddot{z}(t) + \gamma\dot{z}(t) + \partial\phi(\dot{z}(t)) + \nabla f(z(t)) \ni -\beta\nabla^2 f(x(t))\dot{x}(t), \tag{2.8}$$

which satisfies the Cauchy data $z(t_0) = x_0$, and $\dot{z}(t_0) = \dot{x}_0$. Let us first verify that this equation is well posed. For $x \in \mathbb{H}^1(t_0,T;\mathcal{H})$ and according to the L-Lipschitz continuity of ∇f, we have

$$\|\nabla^2 f(x(t))\dot{x}(t)\| \leq L\|\dot{x}(t)\|,$$

which belongs to $\mathbb{L}^2(t_0,T;\mathcal{H})$. So, we can apply Theorem 2.1 which gives the existence and uniqueness of the solution of the solution z of (2.8) with the given Cauchy data.
Let us show that $\mathcal{T} : \mathbb{H}^1(t_0,T;\mathcal{H}) \to \mathbb{H}^1(t_0,T;\mathcal{H})$ is a compact operator. Suppose that $\|x\|_{\mathbb{H}^1(t_0,T;\mathcal{H})} \leq C$, where C is a positive number. Equivalently $\|x\|_{\mathbb{L}^2(t_0,T;\mathcal{H})} \leq C$ and $\|\dot{x}\|_{\mathbb{L}^2(t_0,T;\mathcal{H})} \leq C$. By definition of $z = \mathcal{T}(x)$, we have that (2.8) is verified. Take the scalar product of (2.8) with $\dot{z}(t)$. We obtain

$$\langle \ddot{z}(t), \dot{z}(t)\rangle + \gamma\|\dot{z}(t)\|^2 + \langle \partial\phi(\dot{z}(t)), \dot{z}(t)\rangle + \langle \nabla f(z(t)), \dot{z}(t)\rangle = \left\langle -\beta\nabla^2 f(x(t))\dot{x}(t), \dot{z}(t)\right\rangle. \tag{2.9}$$

Since $\phi : \mathcal{H} \to \mathbb{R}$ is a convex continuous function that satisfies $(\mathrm{DF})_{\mathrm{r}}$, we obtain

$$\frac{1}{2}\frac{d}{dt}\|\dot{z}(t)\|^2 + \gamma\|\dot{z}(t)\|^2 + r\|\dot{z}(t)\| + \frac{d}{dt}f(z(t) \leq \left\langle -\beta\nabla^2 f(x(t))\dot{x}(t), \dot{z}(t)\right\rangle.$$

According to the L-Lipschitz continuity of ∇f, and the Cauchy-Schwarz inequality, we have

$$|\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{z}(t)\right\rangle| \leq L\|\dot{x}(t)\|\|\dot{z}(t)\| \leq \frac{L}{2}(\|\dot{x}(t)\|^2 + \|\dot{z}(t)\|^2).$$

Collecting the above results, we obtain

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{z}(t)\|^2 + f(z(t))\right) + \left(\gamma - \frac{\beta L}{2}\right)\|\dot{z}(t)\|^2 + r\|\dot{z}(t)\| \leq \frac{\beta L}{2}\|\dot{x}(t)\|^2. \tag{2.10}$$

After integration, we get

$$\left(\gamma - \frac{\beta L}{2}\right)\int_{t_0}^{T} \|\dot{z}(t)\|^2 dt \leq \frac{1}{2}\|\dot{x}_0\|^2 + f(x_0) - \inf_{\mathcal{H}} f + \frac{\beta L}{2}\int_{t_0}^{T} \|\dot{x}(t)\|^2 dt.$$

According to the hypothesis $\gamma > \beta L$, and since $\|\dot{x}\|_{\mathbb{L}^2(t_0,T;\mathcal{H})} \leq C$, we deduce that

$$\sup\left\{\|\mathcal{T}(x)\|_{\mathbb{H}^1(t_0,T;\mathcal{H})} : \quad \|\dot{x}\|_{\mathbb{L}^2(t_0,T;\mathcal{H})} \leq C\right\} < +\infty. \tag{2.11}$$

Returning to equation (2.8), and according to the growth condition on $\partial\phi$ it results that

$$\sup\left\{\|\mathcal{T}(x)\|_{\mathbb{H}^2(t_0,T;\mathcal{H})} : \quad \|\dot{x}\|_{\mathbb{L}^2(t_0,T;\mathcal{H})} \leq C\right\} < +\infty. \tag{2.12}$$

Since $\mathcal{H}$ is a finite dimensional space, we use the compact embedding of $\mathbb{H}^2(t_0, T; \mathcal{H})$ into $\mathbb{H}^1(t_0, T; \mathcal{H})$, and so conclude that $\mathcal{T}$ is a compact mapping from $\mathbb{H}^1(t_0, T; \mathcal{H})$ into itself. To conclude to the existence of a fixed point to $\mathcal{T}$ we will use the Schaefer theorem (see [37]), which, besides the fact that $\mathcal{T}$ is a compact mapping, requires the following property: the set

$$K := \left\{x \in \mathbb{H}^1(t_0, T; \mathcal{H}) : \quad \exists \lambda \in]0,1[: \; x = \lambda \mathcal{T}(x)\right\} \quad \text{is bounded.}$$

So suppose that $x = \lambda \mathcal{T}(x)$ for some $\lambda \in]0,1[$. This is equivalent to $\mathcal{T}(x) = \frac{1}{\lambda} x$, that is

$$\frac{1}{\lambda}\ddot{x}(t) + \frac{1}{\lambda}\gamma \dot{x}(t) + \partial\phi(\frac{1}{\lambda}\dot{x}(t)) + \nabla f(\frac{1}{\lambda}x(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) \ni 0. \tag{2.13}$$

After multiplication by λ, we get

$$\ddot{x}(t) + \gamma \dot{x}(t) + \lambda\partial\phi(\frac{1}{\lambda}\dot{x}(t)) + \lambda\nabla f(\frac{1}{\lambda}x(t)) + \beta\lambda\nabla^2 f(x(t))\dot{x}(t) \ni 0. \tag{2.14}$$

By taking the scalar product of (2.14) with $\dot{x}(t)$, and using a similar computation as above, we obtain

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2 + \lambda^2 f(\frac{1}{\lambda}x(t))\right) + (\gamma - \beta L)\,\|\dot{x}(t)\|^2 + r\|\dot{x}(t)\| \leq 0. \tag{2.15}$$

After integration on $[t_0, T]$, we get

$$(\gamma - \beta L)\int_{t_0}^{T} \|\dot{x}(t)\|^2 dt \leq \frac{1}{2}\|\dot{x}_0\|^2 + \lambda^2 f(\frac{1}{\lambda}x_0) - \inf_{\mathcal{H}} f.$$

Since ∇f is L-Lipschitz continuous, the classical gradient descent lemma gives

$$f(x) \leq f(0) + \langle \nabla f(0), x\rangle + \frac{L}{2}\|x\|^2.$$

Hence,

$$\sup_{\lambda\in]0,1[} \lambda^2 f\left(\frac{1}{\lambda}x_0\right) \leq \sup_{\lambda\in]0,1[}\left(\lambda^2 f(0) + \lambda\,\langle \nabla f(0), x_0\rangle + \frac{L}{2}\|x_0\|^2\right) < +\infty.$$

Therefore the set K is bounded in $\mathbb{H}^1(t_0, T; \mathcal{H})$, which completes the proof of the existence of a fixed point to $\mathcal{T}$, thanks to the Schaefer theorem. ■

3 Another approach to existence using Mosco convergence

The major difficulty in (IGDH) is the presence of the term $\partial\phi(\dot{x}(t))$, which is attached to dry friction, and which involves the nonsmooth operator $\partial\phi$. A natural idea is to regularize this operator, and thus obtain a classical evolution equation. To this end, we assume that $\mathcal{H}$ is a finite dimensional Hilbert space, and we will use the Moreau-Yosida regularization (this technique was used in [4] but without the Hessian term). Let us recall some basic facts concerning this regularization procedure. For any $\lambda > 0$, the Moreau envelope of ϕ of index λ is the function $\phi_\lambda : \mathcal{H} \to \mathbb{R}$ defined by: for all $x \in \mathcal{H}$,

$$\phi_\lambda(x) = \min_{\xi\in\mathcal{H}}\left\{\phi(\xi) + \frac{1}{2\lambda}\|x - \xi\|^2\right\}.$$

The function ϕ_λ is convex, of class $\mathcal{C}^{1,1}$, and satisfies $\inf_{\mathcal{H}} \phi_\lambda = \inf_{\mathcal{H}} \phi$, $\operatorname{argmin}_{\mathcal{H}} \phi_\lambda = \operatorname{argmin}_{\mathcal{H}} \phi$. One can consult [9, section 17.2.1], [25], [28] for an in-depth study of the properties of the Moreau envelope in a Hilbert framework. In our context, since $\phi : \mathcal{H} \to \mathbb{R}$ is a convex function that satisfies $(\mathrm{DF})_{\mathrm{r}}$, we will have that ϕ_λ is convex with $\phi_\lambda(0) = \inf_{\mathcal{H}} \phi_\lambda = 0$. This implies that, for all $x \in \mathcal{H}$

$$\langle \nabla\phi_\lambda(x), x\rangle \geq 0. \tag{3.1}$$

This inequality will be very useful later. For each $\lambda > 0$, we consider the following regularized ordinary evolution equation

$$(\mathrm{IGDH})_\lambda \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla\phi_\lambda(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0,\ t \in [t_0, +\infty[. \tag{3.2}$$

As a key property, this equation can be written in an equivalent way as a first-order system in time and space. The following writing is different from the Hamiltonian formulation used in the previous section. More precisely, in our context, we have the following equivalence, which follows from elementary differential calculus.

Proposition 3.1 *The following are equivalent:* $(i) \Longleftrightarrow (ii)$

$$(i) \quad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla\phi_\lambda(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0.$$

$$(ii) \quad \begin{cases} \dot{x}(t) = u(t) - \beta\nabla f(x(t)); \\ \dot{u}(t) = -(\gamma I + \nabla\phi_\lambda)\Big(u(t) - \beta\nabla f(x(t))\Big) - \nabla f(x(t)). \end{cases}$$

Proof. $(i) \Longrightarrow (ii)$. Set

$$u(t) := \dot{x}(t) + \beta\nabla f(x(t)). \tag{3.3}$$

According to the classical derivation chain rule, we have

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

By using (i), we obtain

$$\dot{u}(t) + \gamma\dot{x}(t) + \nabla\phi_\lambda(\dot{x}(t)) + \nabla f(x(t)) = 0. \tag{3.4}$$

Putting together (3.3) and (3.4) we obtain the first-order system in time and space, as given in (ii).

$(ii) \Longrightarrow (i)$. Suppose that ∇f is differentiable (that is f is twice differentiable). By differentiating the first equation of (ii) we obtain

$$\ddot{x}(t) = \dot{u}(t) - \beta\nabla^2 f(x(t))\dot{x}(t).$$

According to the second equation of (ii), we deduce that

$$\begin{aligned} \ddot{x}(t) &= -(\gamma I + \nabla\phi_\lambda)(u(t) - \beta\nabla f(x(t))) - \nabla f(x(t)) - \beta\nabla^2 f(x(t))\dot{x}(t) \\ &= -(\gamma I + \nabla\phi_\lambda)(\dot{x}(t)) - \nabla f(x(t)) - \beta\nabla^2 f(x(t))\dot{x}(t) \end{aligned}$$

which gives (i). ■

This approach leads to a slightly more general existence result than that obtained in Theorem 2.2.

Theorem 3.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function that satisfies $(\mathrm{DF})_{\mathrm{r}}$ and the following growth condition*

$$\|\,(\partial\phi)^o\,(x)\| \leq c(1 + \|x\|),$$

where $(\partial\phi)^o\,(x)$ denotes the element of minimal norm of the set $\partial\phi(x)$ and $c > 0$. Suppose that

$$\gamma > \beta L.$$

Then, for any Cauchy data $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$, there exists a strong global solution of

$$(\mathrm{IGDH}) \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0. \tag{3.5}$$

satisfying $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$.

Proof. For each $\lambda > 0$, $(\text{IGDH})_\lambda$ is relevant of the classical Cauchy-Lipschitz theorem. Indeed, according to Proposition 3.1, we can formulate it equivalently as follows. Set

$$Z(t) = (x(t), u(t)) \in \mathcal{H} \times \mathcal{H}.$$

The system (ii) of Proposition 3.1 can be written equivalently as

$$\dot{Z}(t) + F_\lambda(Z(t)) = 0, \quad Z(t_0) = (x_0, \dot{x}_0)$$

where

$$F_\lambda(x,u) = \Big(\beta\nabla f(x) - u, (\gamma I + \nabla\phi_\lambda)\Big(u - \beta\nabla f(x)\Big) + \nabla f(x)\Big).$$

The Lipschitz continuity property is stable by sum and composition. Since ∇f and $\nabla\phi_\lambda$ are Lipschitz continuous, the components of F_λ are Lipschitz continuous with respect to (x,u), as well as F_λ. Therefore, according to the Cauchy-Lipschitz theorem, there exists a unique global classical solution (x_λ, u_λ) of the above system, which gives that x_λ is solution of

$$(\text{IGDH})_\lambda \qquad \ddot{x}_\lambda(t)+\gamma\dot{x}_\lambda(t)+\nabla\phi_\lambda(\dot{x}_\lambda(t))+\beta\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)+\nabla f(x_\lambda(t)) = 0,\ t \in [t_0, +\infty[. \quad (3.6)$$

Let us first establish energy estimates on the sequence (x_λ). By taking the scalar product of $(\text{IGDH})_\lambda$ with $\dot{x}_\lambda(t)$, we obtain

$$\langle \ddot{x}_\lambda(t), \dot{x}_\lambda(t)\rangle+\gamma\|\dot{x}_\lambda(t)\|^2+\langle\nabla\phi_\lambda(\dot{x}_\lambda(t)), \dot{x}_\lambda(t)\rangle+\beta\left\langle\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t), \dot{x}_\lambda(t)\right\rangle+\langle\nabla f(x_\lambda(t)), \dot{x}_\lambda(t)\rangle = 0.$$

According to the classical derivation chain rule and inequality (3.1), we obtain

$$\frac{1}{2}\frac{d}{dt}\|\dot{x}_\lambda(t)\|^2 + \gamma\|\dot{x}_\lambda(t)\|^2 + \beta\left\langle\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t), \dot{x}_\lambda(t)\right\rangle + \frac{d}{dt}(f(x_\lambda(t)) = 0.$$

According to the L-Lipschitz continuity of ∇f, and the Cauchy-Schwarz inequality, we have

$$|\left\langle\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t), \dot{x}_\lambda(t)\right\rangle| \leq L\|\dot{x}_\lambda(t)\|^2.$$

Therefore,

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}_\lambda(t)\|^2 + f(x_\lambda(t))\right) + (\gamma - \beta L)\|\dot{x}_\lambda(t)\|^2 \leq 0. \quad (3.7)$$

According to the hypothesis $\gamma > \beta L$, we deduce that the global energy

$$E_\lambda(t) = \frac{1}{2}\|\dot{x}_\lambda(t)\|^2 + f(x_\lambda(t))$$

is nonincreasing. From this, and by using that f is minorized we deduce that

$$\sup_{t\geq t_0,\, \lambda>0} \|\dot{x}_\lambda(t)\| < +\infty. \quad (3.8)$$

According to the mean value theorem, and $x_\lambda(t_0) = x_0$ fixed, we deduce that, for each $T \geq t_0$

$$\sup_{t\in[t_0,T],\, \lambda>0} \|x_\lambda(t)\| < +\infty. \quad (3.9)$$

Since ∇f is Lipschitz continuous, this immediately implies that, for each $T \geq t_0$

$$\sup_{t\in[t_0,T],\lambda>0} \|\nabla f(x_\lambda(t))\| < +\infty. \quad (3.10)$$

In addition, by integrating (3.7), we deduce that

$$\sup_{\lambda>0}\int_{t_0}^{t}\|\dot{x}_\lambda(t)\|^2dt<+\infty. \tag{3.11}$$

As a classical property of the Yosida approximation of a maximally monotone operator A, we have

$$\|A_\lambda x\|\leq\|A^o(x)\|$$

where $A^o(x)$ is the element of minimal norm of $A(x)$, see Brézis [28, Proposition 2.6]. According to the assumption $\|\left(\partial\phi\right)^o(x)\|\leq c(1+\|x\|)$, we obtain

$$\|\nabla\phi_\lambda(x)\|\leq c(1+\|x\|).$$

Taking into account (3.8), for each $T\geq t_0$, we have

$$\sup_{t\in[t_0,T],\,\lambda>0}\|\nabla\phi_\lambda(\dot{x}_\lambda(t))\|<+\infty. \tag{3.12}$$

Finally, for each $T\geq t_0$

$$\sup_{t\in[t_0,T],\,\lambda>0}\|\nabla^2f(x_\lambda(t))\dot{x}_\lambda(t)\|\leq L \text{ and } \sup_{t\in[t_0,T],\,\lambda>0}\|\dot{x}_\lambda(t)\|<+\infty. \tag{3.13}$$

Let us now return to $(\mathrm{IGDH})_\lambda$ that we write as follows:

$$\ddot{x}_\lambda(t)=-\gamma\dot{x}_\lambda(t)-\nabla\phi_\lambda(\dot{x}_\lambda(t))-\beta\nabla^2f(x_\lambda(t))\dot{x}_\lambda(t)-\nabla f(x_\lambda(t)). \tag{3.14}$$

Using the above estimates, we deduce that, for each $T\geq t_0$

$$\sup_{\lambda>0}\int_{t_0}^{T}\|\ddot{x}_\lambda(t)\|^2dt<+\infty. \tag{3.15}$$

Therefore, for each $T\geq t_0$, the sequence (x_λ) is bounded in $\mathbb{H}^2(t_0,T;\mathcal{H})$. Since $\mathcal{H}$ has been supposed to be a finite dimensional Hilbert space, this implies that the sequence (x_λ) is relatively compact in $\mathbb{H}^1(t_0,T;\mathcal{H})$ for all $T\geq t_0$. After extraction of a subsequence, still denoted (x_λ), we deduce the existence of x such that

1. $x\in\mathbb{H}^2(t_0,T;\mathcal{H})$ for all $T\geq t_0$.
2. $x_\lambda\to x$ uniformly on $[t_0,T]$ for all $T\geq t_0$.
3. $\nabla f(x_\lambda)\to\nabla f(x)$ uniformly on $[t_0,T]$ for all $T\geq t_0$.
4. $\dot{x}_\lambda\to\dot{x}$ uniformly on $[t_0,T]$ for all $T\geq t_0$.
5. $\ddot{x}_\lambda\to\ddot{x}$ weakly in $\mathbb{L}^2(t_0,T;\mathcal{H})$ for all $T\geq t_0$.
6. $\nabla^2f(x_\lambda)\dot{x}_\lambda\to\nabla^2f(x)\dot{x}$ weakly in $\mathbb{L}^2(t_0,T;\mathcal{H})$ for all $T\geq t_0$.

The above property 6. comes from the fact that

$$\nabla^2f(x_\lambda)\dot{x}_\lambda=\frac{d}{dt}\nabla f(x_\lambda),$$

and the continuity of the derivation in the distribution sense.
To pass to the limit in $(\mathrm{IGDH})_\lambda$, we write is as follows

$$\nabla\phi_\lambda(\dot{x}_\lambda(t)) = \xi_\lambda(t), \tag{3.16}$$

where

$$\xi_\lambda(t) = -\ddot{x}_\lambda(t) - \gamma\dot{x}_\lambda(t) - \beta\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t) - \nabla f(x_\lambda(t)).$$

We now rely on the variational convergence properties of the Yosida approximation. Since ϕ_λ converges increasingly to ϕ, the sequence of integral functionals

$$\Phi^\lambda(\xi) := \int_{t_0}^T \phi_\lambda(\xi(t))dt$$

converges increasingly to

$$\Phi(\xi) = \int_{t_0}^T \phi(\xi(t))dt.$$

Therefore it converges in the Mosco sense in $\mathbb{L}^2(t_0, T; \mathcal{H})$, see Mosco [33]. According to the theorem which makes the link between the Mosco convergence of a sequence of convex lower semicontinuous functions and the graph convergence of their subdifferentials, see Attouch [8, Theorem 3.66], we have that

$$\partial\Phi^\lambda \to \partial\Phi$$

with respect to the topology $\mathrm{strong} - \mathbb{L}^2(t_0, T; \mathcal{H}) \times \mathrm{weak} - \mathbb{L}^2(t_0, T; \mathcal{H})$. We have

$$\xi_\lambda = \nabla\Phi^\lambda(\dot{x}_\lambda).$$

Since

$$\dot{x}_\lambda \to \dot{x} \ \text{ strongly in}\, \mathbb{L}^2(t_0, T; \mathcal{H})$$

and ξ_λ converges weakly in $\mathbb{L}^2(t_0, T; \mathcal{H})$ to ξ given by

$$\xi(t) = -\ddot{x}(t) - \gamma\dot{x}(t) - \beta\nabla^2 f(x(t))\dot{x}(t) - \nabla f(x(t)), \tag{3.17}$$

we deduce that $\xi \in \partial\Phi(\dot{x})$, that is

$$\xi(t) \in \partial\phi(\dot{x}(t)).$$

According to the formulation (3.17) of ξ, we finally obtain that x is a solution of (IGDH). ■

4 Existence and uniqueness results in an infinite dimensional setting

We are going to prove existence and uniqueness results for (IGDH) in a general Hilbert framework. For this, we will need to make further regularity properties on the potential function f. We assume that

$$x \mapsto \nabla^2 f(x) \ \text{ is Lipschitz continuous on the bounded subsets of } \mathcal{H}.$$

Theorem 4.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^2$ function which satisfies:*

(i) ∇f is L-Lipschitz continuous on $\mathcal{H}$;

(i) $\nabla^2 f$ is Lipschitz continuous on the bounded subsets of $\mathcal{H}$.

Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function that satisfies (DF) *and the following growth condition: there exists some $c > 0$ such that for all $x \in \mathcal{H}$*

$$\| (\partial\phi)^o (x)\| \leq c(1 + \|x\|).$$

Suppose that

$$\gamma \geq \beta L.$$

Then, for any Cauchy data $(x_0, \dot{x}_0) \in \mathcal{H}\times\mathcal{H}$, there exists a unique strong global solution $x : [t_0, +\infty[\to \mathcal{H}$ of

$$(\mathrm{IGDH}) \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0, \tag{4.1}$$

satisfying $x(t_0) = x_0$, and $\dot{x}(t_0) = \dot{x}_0$.

Proof. According to Proposition 2.1, and the Hamiltonian formulation of (IGDH), it is equivalent to solve the following first-order system

$$\begin{cases} \dot{x}(t) - u(t) = 0; \\ \dot{u}(t) + \partial\phi(u(t)) + \gamma u(t) + \nabla f(x(t)) + \beta\nabla^2 f(x(t))u(t) \ni 0, \end{cases}$$

with the Cauchy data $x(t_0) = x_0, \;\; u(t_0) = \dot{x}_0$. Set

$$Z(t) = (x(t), u(t)) \in \mathcal{H} \times \mathcal{H}.$$

The above system can be rewritten equivalently as

$$\dot{Z}(t) + F(Z(t)) \ni 0, \quad Z(t_0) = (x_0, \dot{x}_0),$$

where $F : \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}, \;\; (x, u) \mapsto F(x, u)$ is defined by

$$F(x, u) = \Big(0, \partial\phi(u)\Big) + \Big(- u, \gamma u + \nabla f(x) + \beta\nabla^2 f(x)u\Big).$$

Hence F splits as follows

$$F(x, u) = \partial\Phi(x, u) + G(x, u),$$

where

$$\Phi(x, u) = \phi(u)$$

and

$$G(x, u) = \Big(- u, \; \gamma u + \nabla f(x) + \beta\nabla^2 f(x)u\Big). \tag{4.2}$$

Consequently, the dynamic problem (4.1) is equivalent to the resolution of the following first-order differential inclusion with Cauchy data

$$\dot{Z}(t) + \partial\Phi(Z(t)) + G(Z(t)) \ni 0, \quad Z(t_0) = (x_0, \dot{x}_0). \tag{4.3}$$

We cannot apply [28, Proposition 3.12] which considers the evolution equation governed by a Lipschitz perturbation of a convex subdifferential, since the mapping $(x, u) \mapsto \nabla^2 f(x)u$ is not Lipschitz continuous. In fact, we are going to prove that the mapping $(x, u) \mapsto G(x, u)$, defined in (4.2), is locally Lipschitz continuous. Indeed, this property, combined with a priori estimates, will allow us to conclude.

1. Local Lipschitz continuity of G. In our framework, local Lipschitz continuity is taken in the sense of Lipschitz continuity on the bounded sets.

For any $(x,u) \in \mathcal{H} \times \mathcal{H}$, set $K(x,u) := \gamma u + \nabla f(x) + \beta \nabla^2 f(x) u$, so that $G(x,u) = (-u, K(x,u))$. For any $(x_i, u_i) \in \mathcal{H} \times \mathcal{H}$, $i = 1, 2$ we have

$$K(x_2,u_2) - K(x_1,u_1) = \gamma(u_2 - u_1) + (\nabla f(x_2) - \nabla f(x_1)) + \beta(\nabla^2 f(x_2)u_2 - \nabla^2 f(x_1)u_1).$$

According to the triangle inequality and the L-Lipschitz continuity assumption of ∇f, we have

$$\begin{aligned}
\|K(x_2,u_2) - K(x_1,u_1)\| &\leq \gamma\|u_2 - u_1\| + \|\nabla f(x_2) - \nabla f(x_1)\| \\
&+ \beta\|\nabla^2 f(x_2)u_2 - \nabla^2 f(x_1)u_2\| + \beta\|\nabla^2 f(x_1)u_2 - \nabla^2 f(x_1)u_1\| \\
&\leq \gamma\|u_2 - u_1\| + L\|x_2 - x_1\| \\
&+ \beta\|\nabla^2 f(x_2) - \nabla^2 f(x_1)\|\|u_2\| + \beta L\|u_2 - u_1\|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|G(x_2,u_2) - G(x_1,u_1)\| &\leq (\gamma + 1)\|u_2 - u_1\| + L\|x_2 - x_1\| \\
&+ \beta\|\nabla^2 f(x_2) - \nabla^2 f(x_1)\|\|u_2\| + \beta L\|u_2 - u_1\|. \qquad (4.4)
\end{aligned}$$

Using the fact that $x \mapsto \nabla^2 f(x)$ is Lipschitz continuous on bounded sets of $\mathcal{H}$, it is easy to deduce from (4.4) that the mapping $(x,u) \mapsto G(x,u)$ is locally Lipschitz continuous on $\mathcal{H} \times \mathcal{H}$.

2. A priori estimates.

Let x be a strong global solution of (IGDH). Taking the dot product of (IGDH) with $\dot{x}(t)$ gives

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t))\right) + (\gamma - \beta L)\|\dot{x}\|^2 \leq 0. \qquad (4.5)$$

According to $\gamma - \beta L \geq 0$, we deduce that

$$\|\dot{x}(t)\|^2 \leq \|\dot{x}_0\|^2 + 2(f(x_0) - \inf_{\mathcal{H}} f), \qquad (4.6)$$

which gives

$$\|\dot{x}(t)\| \leq \|\dot{x}_0\| + \sqrt{2(f(x_0) - \inf_{\mathcal{H}} f)}. \qquad (4.7)$$

After integration we obtain that, for all $t \in [t_0, T]$

$$\|x(t)\| \leq \|x_0\| + T\left(\|\dot{x}_0\| + \sqrt{2(f(x_0) - \inf_{\mathcal{H}} f)}\right). \qquad (4.8)$$

Therefore (recall that $Z(t) = (x(t), \dot{x}(t))$

$$\|Z(t)\| \leq C_T := \|x_0\| + (1+T)\|\dot{x}_0\| + T\sqrt{2(f(x_0) - \inf_{\mathcal{H}} f)}. \qquad (4.9)$$

Note that, for given $T \geq t_0$, C_T only depends on the Cauchy data and the infimum of f.

3. Uniqueness

Let x_1 and x_2 be two solutions of (IGDH) satisfying the given Cauchy data. Set $Z_1(t) = (x_1(t), u_1(t))$ and $Z_2(t) = (x_2(t), u_2(t))$, which satisfy

$$\dot{Z}_1(t) + \partial\Phi(Z_1(t)) + G(Z_1(t)) \ni 0, \quad Z_1(t_0) = (x_0, \dot{x}_0) \qquad (4.10)$$

$$\dot{Z}_2(t) + \partial\Phi(Z_2(t)) + G(Z_2(t)) \ni 0, \quad Z_2(t_0) = (x_0, \dot{x}_0). \qquad (4.11)$$

According to the local Lipschitz behavior of G, see (4.4), we have

$$\begin{array}{rcl}\|G(Z_1(t)) - G(Z_2(t))\| & \leq & (\gamma+1)\|u_2(t) - u_1(t)\| + L\|x_2(t) - x_1(t)\| \\ & + & \beta\|\nabla^2 f(x_2(t)) - \nabla^2 f(x_1(t))\|\|u_2(t)\| + \beta L\|u_2(t) - u_1(t)\|.\end{array}$$

According to the apriori estimates (4.9), we have $\|x_i(t)\| \leq C_T$, and $\|u_i(t)\| = \|\dot{x}_i(t)\| \leq C_T$ for $i = 1, 2$ and all $t \in [t_0, T]$. Let L_2 be the Lipschitz constant of $\nabla^2 f$ on the ball centered at the origin and of radius C_T. We deduce from the above inequality that

$$\begin{array}{rcl}\|G(Z_1(t)) - G(Z_2(t))\| & \leq & (\gamma+1)\|u_2(t) - u_1(t)\| + L\|x_2(t) - x_1(t)\| \\ & + & \beta L_2 C_T\|x_2(t) - x_1(t)\| + \beta L\|u_2(t) - u_1(t)\|.\end{array}$$

Therefore, there exists a positive constant M_T such that, for all $t \in [t_0, T]$

$$\|G(Z_1(t)) - G(Z_2(t))\| \leq M_T\|Z_1(t) - Z_2(t)\|.$$

Precisely, we can take

$$M_T = \sqrt{2}\max\{\gamma + 1 + \beta L;\ L + \beta L_2 C_T\}.$$

Then, the uniqueness follows from a standard monotonicity argument. Make the difference of the two equations (4.10) and (4.11), and take the scalar product with $Z_1(t) - Z_2(t)$. According to the monotonicity property of $\partial\Phi$ and the Lipschitz continuity of G we infer

$$\frac{1}{2}\frac{d}{dt}\|Z_1(t) - Z_2(t)\|^2 \leq M_T\|Z_1(t) - Z_2(t)\|^2.$$

After integration , we get

$$\|Z_1(t) - Z_2(t)\|^2 \leq \|Z_1(t_0) - Z_2(t_0)\|^2 + 2M_T\int_{t_0}^{t}\|Z_1(\tau) - Z_2(\tau)\|^2 d\tau.$$

According to the classical Gronwall lemma (see Lemma A.4 [28]), we infer

$$\|Z_1(t) - Z_2(t)\| \leq \|Z_1(t_0) - Z_2(t_0)\|e^{M_T(t-t_0)}.$$

Since $Z_1(t_0) = Z_2(t_0)$, this immediately implies that $Z_1(t) = Z_2(t)$ for all $t \in [t_0, T]$.

4. Existence. We now examine the question of existence. We follow a parallel approach to that used for the uniqueness result, but now we are working with the approximate dynamics

$$(\text{IGDH})_\lambda \qquad \ddot{x}_\lambda(t)+\gamma\dot{x}_\lambda(t)+\nabla\phi_\lambda(\dot{x}_\lambda(t))+\beta\nabla^2 f(x_\lambda(t))\dot{x}_\lambda(t)+\nabla f(x_\lambda(t)) = 0,\ t \in [t_0, +\infty[\quad (4.12)$$

considered in Section 3, and which uses the Moreau-Yosida approximates (ϕ_λ) of ϕ. We will prove that the filtered sequence (x_λ) converges uniformly as $\lambda \to 0$ over the bounded time intervals towards a solution of (IGDH). According to Proposition 2.1, and the Hamiltonian formulation of $(\text{IGDH})_\lambda$, it is equivalent to consider the first-order (in time) ordinary system

$$\begin{cases} \dot{x}_\lambda(t) - u_\lambda(t) = 0; \\ \dot{u}_\lambda(t) + \nabla\phi_\lambda(u_\lambda(t)) + \gamma u_\lambda(t) + \nabla f(x_\lambda(t)) + \beta\nabla^2 f(x_\lambda(t))u_\lambda(t) = 0, \end{cases}$$

with the Cauchy data $x_\lambda(t_0) = x_0,\ \ u_\lambda(t_0) = \dot{x}_0$. Set

$$Z_\lambda(t) = (x_\lambda(t), u_\lambda(t)) \in \mathcal{H} \times \mathcal{H}.$$

The above system can be rewritten equivalently as

$$\dot{Z}_\lambda(t) + F_\lambda(Z_\lambda(t)) \ni 0, \quad Z_\lambda(t_0) = (x_0, \dot{x}_0),$$

where $F_\lambda : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}, \ (x, u) \mapsto F_\lambda(x, u)$ is defined by

$$F_\lambda(x, u) = \Big(0, \nabla\phi_\lambda(u)\Big) + \Big(-u, \gamma u + \nabla f(x) + \beta\nabla^2 f(x)u\Big).$$

Hence F_λ splits as follows

$$F(x, u) = \nabla\Phi_\lambda(x, u) + G(x, u)$$

where

$$\Phi(x, u) = \phi(u), \quad \Phi_\lambda(x, u) = \phi_\lambda(u)$$

and

$$G(x, u) = \Big(-u,\ \gamma u + \nabla f(x) + \beta\nabla^2 f(x)u\Big).$$

Therefore, it is equivalent to consider the following first-order differential equation with Cauchy data

$$\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(t_0) = (x_0, \dot{x}_0). \tag{4.13}$$

To prove the uniform convergence of the filtered sequence (Z_λ) on the bounded time intervals we proceed in a similar way as in the proof of Brezis [28, Theorem 3.1]. Take $T > t_0$, and $\lambda, \mu > 0$. Consider the corresponding solutions on $[t_0, T]$

$$\begin{aligned} \dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(t_0) = (x_0, \dot{x}_0) \\ \dot{Z}_\mu(t) + \nabla\Phi_\mu(Z_\mu(t)) + G(Z_\mu(t)) = 0, \quad Z_\mu(t_0) = (x_0, \dot{x}_0). \end{aligned}$$

Let's make the difference between the two equations, and take the scalar product with $Z_\lambda(t) - Z_\mu(t)$. We obtain

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 \quad &+ \quad \langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \\ &+ \quad \langle G(Z_\lambda(t)) - G(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle = 0. \end{aligned} \tag{4.14}$$

We now use the following basic ingredients:

a) According to the general properties of the Yosida approximation (see [28, Theorem 3.1]) , we have

$$\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \geq -\frac{\lambda}{4}\|\nabla\Phi_\mu(Z_\mu(t))\|^2 - \frac{\mu}{4}\|\nabla\Phi_\lambda(Z_\lambda(t))\|^2.$$

According to the proof of Theorem 3.1, the sequence (Z_λ) is uniformly bounded on $[t_0, T]$, let

$$\|Z_\lambda(t)\| \leq C_T.$$

From these properties we immediately infer

$$\|\nabla\Phi_\lambda(Z_\lambda(t))\| \leq \sup_{\|\xi\| \leq C_T} \|(\partial\phi)^0(\xi)\| = M_T < +\infty.$$

Thanks to our assumption on ϕ, $\partial\phi$ is bounded on the bounded sets. Therefore

$$\langle \nabla\Phi_\lambda(Z_\lambda(t)) - \nabla\Phi_\mu(Z_\mu(t)), Z_\lambda(t) - Z_\mu(t)\rangle \geq -\frac{1}{4}M_T(\lambda + \mu).$$

b) According to the assumptions on ∇f and $\nabla^2 f$, the mapping $G : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is Lipschitz continuous on the bounded sets (see the proof of the uniqueness). Using again the fact that the sequence (Z_λ) is uniformly bounded on $[t_0, T]$, we deduce that there exists a constant L_2 such that

$$\|G(Z_\lambda(t)) - G(Z_\mu(t))\| \leq L_2 \|Z_\lambda(t) - Z_\mu(t)\|.$$

Combining the above results, and using Cauchy-Schwarz inequality, we deduce from (4.14) that

$$\frac{1}{2}\frac{d}{dt}\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{1}{4} M_T(\lambda + \mu) + L_2 \|Z_\lambda(t) - Z_\mu(t)\|^2.$$

We now proceed with the integration of this differential inequality. According to the fact that $Z_\lambda(t_0) - Z_\mu(t_0) = 0$, elementary calculus gives

$$\|Z_\lambda(t) - Z_\mu(t)\|^2 \leq \frac{M_T}{4L_2}(\lambda + \mu)\Big(e^{2L_2(t-t_0} - 1\Big).$$

Therefore, the filtered sequence (Z_λ) is a Cauchy sequence for the uniform convergence on $[t_0, T]$, and hence it converges uniformly. This means the uniform convergence of x_λ and $\dot{x}_\lambda$ to x and $\dot{x}$ respectively. Proving that x a is solution of (IGDH) can be obtained in a similar way as in Theorem 3.1. ■

We have the following stability property with respect to the initial condition of the system (IGDH).

Corollary 4.1 Let x_1 and x_2 be two solutions of the system. Then, for all $t \in [t_0, T]$

$$\|(x_1(t), \dot{x}_1(t)) - (x_2(t), \dot{x}_2(t))\| \leq \|(x_1(t_0), \dot{x}_1(t_0)) - (x_2(t_0), \dot{x}_2(t_0))\| e^{M_T(t-t_0)},$$

where

$$M_T = \sqrt{2}\max\{\gamma + 1 + \beta L;\ L + \beta L_2 C_T\},$$

and L_2 is the Lipschitz constant of $\nabla^2 f$ on the ball centered at the origin and of radius $C_T = \|x_0\| + (1 + T)\|\dot{x}_0\| + T\sqrt{2(f(x_0) - \inf_{\mathcal{H}} f)}$.

Proof. The proof is straightforward and is a consequence of that given in Theorem 4.1. ■

5 Finite time convergence of the trajectories

Let's analyze the asymptotic behavior as $t \to +\infty$, and the finite convergence property, of the solution trajectories of the second-order differential inclusion

$$\text{(IGDH)} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,\ t \in [t_0, +\infty[.$$

Theorem 5.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function that satisfies* $(\mathrm{DF})_{\mathrm{r}}$ *and which is bounded on the bounded sets. Suppose that*

$$\gamma > \beta L.$$

Then, for any solution trajectory $x(\cdot)$ of (IGDH) *we have:*

(i) $\|\dot{x}\| \in \mathbb{L}^1(t_0, +\infty; \mathbb{R})$, *and therefore the strong limit $x_\infty := \lim_{t\to+\infty} x(t)$ exists.*

(ii) The limit point x_∞ is an equilibrium point of (IGDH), *i.e.* $-\nabla f(x_\infty) \in \partial\phi(0)$.

(iii) If $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$, *then there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$.*

Proof. (i) Take the scalar product of (IGDH) with $\dot{x}(t)$. We obtain

$$\langle \ddot{x}(t), \dot{x}(t)\rangle + \gamma\|\dot{x}(t)\|^2 + \langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle + \beta\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle + \langle \nabla f(x(t)), \dot{x}(t)\rangle = 0, \quad (5.1)$$

which gives

$$\frac{1}{2}\frac{d}{dt}\|\dot{x}(t)\|^2 + \gamma\|\dot{x}(t)\|^2 + \langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle + \beta\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle + \frac{d}{dt}(f(x(t) - \inf_{\mathcal{H}} f) = 0.$$

According to the L-Lipschitz continuity of ∇f, and the Cauchy-Schwarz inequality, we have

$$|\left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t)\right\rangle| \leq L\|\dot{x}(t)\|^2.$$

According to the assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1,

$$\langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle \geq \phi\left(\dot{x}(t)\right) \geq r\|\dot{x}(t)\|.$$

Collecting the above results, we obtain

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t)) - \inf_{\mathcal{H}} f\right) + (\gamma - \beta L)\|\dot{x}(t)\|^2 + r\|\dot{x}(t)\| \leq 0. \quad (5.2)$$

According to the hypothesis $\gamma > \beta L$, we deduce that the global energy

$$E(t) = \frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t) - \inf_{\mathcal{H}} f,$$

is nonincreasing. Moreover, by integrating (5.2) we obtain

$$\int_{t_0}^{\infty} \|\dot{x}(t)\|^2 dt < +\infty \quad \text{and} \quad \int_{t_0}^{\infty} \|\dot{x}(t)\| dt < +\infty. \quad (5.3)$$

This last property expresses that the trajectory has a finite length, and hence $\lim_{t\to+\infty} x(t) := x_\infty$ exists.

(ii) Since $E(\cdot)$ is nonincreasing, we have that the velocity $\dot{x}(t)$ remains bounded. Since ϕ is bounded on the bounded sets, so is $\partial\phi$. Therefore, from equation (IGDH) we deduce that the acceleration $\ddot{x}(t)$ remains bounded. This combined with $\int_{t_0}^{\infty} \|\dot{x}(t)\| dt < +\infty$ implies that the velocity $\dot{x}(t)$ converges strongly to zero, as $t \to +\infty$. Let us now pass to the limit in (IGDH). Set $u(t) = \dot{x}(t)$. Let us write (IGDH) equivalently as

$$\dot{u}(t) + (\gamma I + \partial\phi)(u(t)) = h(t)$$

with $h(t) := -\beta\nabla^2 f(x(t))\dot{x}(t) - \nabla f(x(t))$. The operator $A = \gamma I + \partial\phi$ is strongly monotone since $\gamma > 0$. According to the above results, we have that $h(t)$ converges strongly to $-\nabla f(x_\infty)$. We now apply Theorem 3.9 of Brezis [28], which tells us that the strong limit of $u(t)$, that's zero, satisfies $A(0) \ni -\nabla f(x_\infty)$. Equivalently

$$\partial\phi(0) \ni -\nabla f(x_\infty).$$

(iii) The assumption $-\nabla f(x_\infty) \in \mathrm{int}(\partial\phi(0))$ implies the existence of $\varepsilon > 0$ such that

$$-\nabla f(x_\infty) + B(0, 2\varepsilon) \subset \partial\phi(0).$$

On the other hand, since $\lim_{t\to+\infty} \nabla f(x(t)) = \nabla f(x_\infty)$, there exists $t_1 \geq t_0$ such that for all $t \geq t_1$

$$\nabla f(x(t)) \in \nabla f(x_\infty) + B(0, \varepsilon).$$

Hence,

$$-\nabla f(x(t)) + B(0,\varepsilon) \subset -\nabla f(x_\infty) + B(0, 2\,\varepsilon) \subset \partial\phi(0).$$

Equivalently, for every $t \geq t_1$ and for every $u \in B(0,1)$, we have:

$$-\nabla f(x(t))) + \varepsilon\, u \in \partial\phi(0).$$

Let's write the corresponding subdifferential inequality at the origin (recall that $\phi(0) = 0$)). For every $t \geq t_1$

$$\forall u \in B(0,1), \quad \phi(\dot{x}(t)) \geq \langle -\nabla f(x(t)) + \varepsilon\, u, \dot{x}(t)\rangle.$$

Taking the supremum over $u \in B(0,1)$, we obtain that, for every $t \geq t_1$,

$$\phi(\dot{x}(t)) + \langle \nabla f(x(t)), \dot{x}(t)\rangle \ \geq \varepsilon\, \|\dot{x}(t)\|. \tag{5.4}$$

Let's return to (5.1). According to the above results, we obtain

$$\frac{d}{dt}\frac{1}{2}\|\dot{x}(t)\|^2 + (\gamma - \beta L)\|\dot{x}(t)\|^2 + \varepsilon\|\dot{x}(t)\| \leq 0. \tag{5.5}$$

a) Neglecting the nonnegative term $\varepsilon\|\dot{x}(t)\|$ we obtain

$$\frac{d}{dt}\frac{1}{2}\|\dot{x}(t)\|^2 + (\gamma - \beta L)\|\dot{x}(t)\|^2 \leq 0, \tag{5.6}$$

whose integration gives

$$\|\dot{x}(t)\| \leq \|\dot{x}(t_0)\| e^{-(\gamma-\beta L)t}.$$

b) Neglecting the nonnegative term $(\gamma - \beta L)\|\dot{x}(t)\|^2$ we obtain

$$\frac{d}{dt}\|\dot{x}(t)\|^2 + 2\varepsilon\|\dot{x}(t)\| \leq 0, \tag{5.7}$$

Set $v(t) = \|\dot{x}(t)\|^2$. We have $\dot{v}(t) + 2\varepsilon\sqrt{v(t)} \leq 0$. As long as $v(t) > 0$, we will have $\frac{d}{dt}\sqrt{v(t)} \leq -\epsilon$. This forces $v(t)$ to be equal to zero after some finite time. ■

Remark 5.1 With the condition $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$, the finite time convergence of the trajectory to a stationary point of the dynamic (IGDH) is ensured, i.e. there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$. In addition, an estimate of the final time can be given. Thanks to a detailed analysis of the end of the previous demonstration, we can show that

$$t_1 \leq \tau_0 + \frac{2\|\dot{x}(\tau_0)\|}{\text{dist}\Big(-\nabla f(x_\infty), \text{boundary}(\partial\phi(0)\Big)},$$

where τ_0 is the first time instant such that

$$\nabla f(x(t)) \in \nabla f(x_\infty) + B(0,\varepsilon), \text{ for all } t \geq t_0, \text{ with } \varepsilon = \tfrac{1}{2}\text{dist}\Big(-\nabla f(x_\infty), \text{boundary}(\partial\phi(0))\Big).$$

Soft thresholding of the velocities As a model situation for dry friction, take $\phi : \mathcal{H} \to \mathbb{R}$ given by $\phi(x) = r\|x\|$, with $r > 0$. We have

$$\partial\phi(x) = \begin{cases} r\frac{x}{\|x\|} & \text{if } x \neq 0; \\ B(0,r) & \text{if } x = 0. \end{cases} \tag{5.8}$$

A direct application of Theorem 5.1 gives the following result:

Corollary 5.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous. Assume that the potential friction function ϕ is given by $\phi(x) = r\|x\|$. Suppose that*

$$\gamma > \beta L.$$

Then, for any solution trajectory $x(\cdot)$ of (IGDH) *we have:*

- *(i) $\|\dot{x}\| \in \mathbb{L}^1(t_0, +\infty; \mathbb{R})$, and therefore the strong limit $x_\infty := \lim_{t\to+\infty} x(t)$ exists.*
- *(ii) The limit point x_∞ satisfies*
$$\|\nabla f(x_\infty)\| \leq r.$$
- *(iii) If $\|\nabla f(x_\infty)\| < r$, then there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$.*

Remark 5.2 Taking $r > 0$ small is the most interesting situation for optimization. It should be noted that one of the stopping criteria used in numerical optimization, when minimizing a smooth function f, is

$$\|\nabla f(x_k)\| < \text{ tolerance},$$

where (x_k) is the sequence generated by the algorithm and the tolerance $= r$ is to be chosen by the user. In doing so, the condition, $\|\nabla f(x_\infty)\| < r$, is implicitly used.

6 Errors, perturbations

Let's analyze the asymptotic behavior as $t \to +\infty$, and the finite convergence property, of the solution trajectories of the second-order differential inclusion

$$\text{(IGDH)-pert} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni e(t)$$

which comes from the introduction of a second member $e : [t_0, +\infty[\to \mathcal{H}$ in (IGDH). Depending on the context, e can be interpreted as an external action, a control term, or coming from perturbations or errors.

Theorem 6.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ function whose gradient is L-Lipschitz continuous, and let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function that satisfies* $(\mathrm{DF})_\mathrm{r}$ *and which is bounded on the bounded sets. Suppose that*

$$\gamma > \beta L.$$

Suppose that the external action $e : [t_0, +\infty[\to \mathcal{H}$ satisfies

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

Then, for any solution trajectory $x(\cdot)$ of (IGDH)-pert *we have:*

- *(i) $\|\dot{x}\| \in \mathbb{L}^1(t_0, +\infty; \mathbb{R})$, and therefore the strong limit $x_\infty := \lim_{t\to+\infty} x(t)$ exists.*
- *(ii) The limit point x_∞ satisfies: $-\nabla f(x_\infty) \in \partial\phi(0)$.*
- *(iii) If $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$, then there exists $t_1 \geq 0$ such that $x(t) = x_\infty$ for every $t \geq t_1$.*

Proof. (i) By taking the scalar product of (IGDH)-pert with $\dot{x}(t)$, we obtain

$$\langle \ddot{x}(t), \dot{x}(t) \rangle + \gamma \|\dot{x}(t)\|^2 + \langle \partial\phi(\dot{x}(t)), \dot{x}(t) \rangle + \\ \beta \left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t) \right\rangle + \langle \nabla f(x(t)), \dot{x}(t) \rangle = \langle e(t), \dot{x}(t) \rangle, \tag{6.1}$$

which gives

$$\frac{1}{2}\frac{d}{dt}\|\dot{x}(t)\|^2 + \gamma\|\dot{x}(t)\|^2 + \langle \partial\phi(\dot{x}(t)), \dot{x}(t) \rangle + \beta \left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t) \right\rangle + \frac{d}{dt}(f(x(t)) = \langle e(t), \dot{x}(t) \rangle .$$

According to the L-Lipschitz continuity of ∇f, and the Cauchy-Schwarz inequality, we get

$$| \left\langle \nabla^2 f(x(t))\dot{x}(t), \dot{x}(t) \right\rangle | \leq L\|\dot{x}(t)\|^2.$$

Using assumption $(\mathrm{DF})_r$ on ϕ and Lemma 1.1, we have

$$\langle \partial\phi(\dot{x}(t)), \dot{x}(t) \rangle \geq \phi\left(\dot{x}(t)\right) \geq r\|\dot{x}(t)\|.$$

By Cauchy-Schwarz inequality

$$| \langle e(t), \dot{x}(t) \rangle | \leq \|e(t)\|\|\dot{x}(t)\|.$$

Collecting the above results, we obtain

$$\frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t)) - \inf_{\mathcal{H}} f\right) + (\gamma - \beta L)\|\dot{x}(t)\|^2 + (r - \|e(t)\|)\|\dot{x}(t)\| \leq 0. \tag{6.2}$$

According to the hypothesis $\gamma > \beta L$, and $\lim_{t\to+\infty} \|e(t)\| = 0$, we deduce that there exists $t_1 \geq t_0$ such that, for $t \geq t_1$, the global energy

$$E(t) = \frac{1}{2}\|\dot{x}(t)\|^2 + f(x(t) - \inf_{\mathcal{H}} f$$

is nonincreasing. Moreover, by integrating (6.2) we obtain

$$\int_{t_0}^{\infty} \|\dot{x}(t)\|^2 dt < +\infty \quad \text{and} \quad \int_{t_0}^{\infty} \|\dot{x}(t)\| dt < +\infty. \tag{6.3}$$

This last property expresses that the trajectory has a finite length, and hence $\lim_{t\to+\infty} x(t) := x_\infty$ exists.

(ii) Since $E(\cdot)$ is nonincreasing, the velocity $\dot{x}(t)$ remains bounded. Since ϕ is bounded on the bounded sets, so is $\partial\phi$. Therefore, from equation (IGDH)-pert we deduce that the acceleration $\ddot{x}(t)$ remains bounded. This combined with $\int_{t_0}^{\infty} \|\dot{x}(t)\| dt < +\infty$ implies that the velocity $\dot{x}(t)$ converges strongly to zero, as $t \to +\infty$. Let us now pass to the limit in (IGDH)-pert . Set $u(t) = \dot{x}(t)$. Let us write (IGDH)-pert equivalently as

$$\dot{u}(t) + (\gamma I + \partial\phi)(u(t)) = h(t)$$

with $h(t) := -\beta\nabla^2 f(x(t))\dot{x}(t) - \nabla f(x(t)) + e(t)$. The operator $A = \gamma I + \partial\phi$ is strongly monotone since $\gamma > 0$. According to the above results, we have that $h(t)$ converges strongly to $-\nabla f(x_\infty)$. We now apply Theorem 3.9 of Brezis [28], which gives us that the strong limit of $u(t)$, that is zero, satisfies $A(0) \ni -\nabla f(x_\infty)$. Equivalently

$$\partial\phi(0) \ni -\nabla f(x_\infty).$$

(iii) The assumption $-\nabla f(x_\infty) \in \mathrm{int}\left(\partial\phi(0)\right)$ implies the existence of $\varepsilon > 0$ such that

$$-\nabla f(x_\infty) + B(0, 2\,\varepsilon) \subset \partial\phi(0).$$

On the other hand, since $\lim_{t\to+\infty} \nabla f(x(t)) = \nabla f(x_\infty)$, there exists $t_1 \geq t_0$ such that for all $t \geq t_1$

$$\nabla f(x(t)) \in \nabla f(x_\infty) + B(0, \varepsilon).$$

Hence,

$$-\nabla f(x(t)) + B(0, \varepsilon) \subset -\nabla f(x_\infty) + B(0, 2\,\varepsilon) \subset \partial\phi(0).$$

Equivalently, for every $t \geq t_1$ and for every $u \in B(0,1)$, we have:

$$-\nabla f(x(t))) + \varepsilon\, u \in \partial\phi(0).$$

Let's write the corresponding subdifferential inequality at the origin (recall that $\phi(0) = 0$)). For every $t \geq t_1$

$$\forall u \in B(0,1), \quad \phi(\dot{x}(t)) \geq \langle -\nabla f(x(t)) + \varepsilon\, u, \dot{x}(t)\rangle.$$

Taking the supremum over $u \in B(0,1)$, we obtain that, for every $t \geq t_1$,

$$\phi(\dot{x}(t)) + \langle \nabla f(x(t)), \dot{x}(t)\rangle \ \geq \varepsilon\, \|\dot{x}(t)\|. \tag{6.4}$$

Let's return to (6.1). According to the above results, we obtain

$$\frac{d}{dt}\frac{1}{2}\|\dot{x}(t)\|^2 + (\gamma - \beta L)\|\dot{x}(t)\|^2 + (\varepsilon - \|e(t)\|)\|\dot{x}(t)\| \leq 0. \tag{6.5}$$

a) Neglecting the nonnegative term $(\varepsilon - \|e(t)\|)\|\dot{x}(t)\|$ (recall that $\varepsilon > 0$ and $\|e(t)\| \to 0$), we obtain

$$\frac{d}{dt}\frac{1}{2}\|\dot{x}(t)\|^2 + (\gamma - \beta L)\|\dot{x}(t)\|^2 \leq 0, \tag{6.6}$$

whose integration gives

$$\|\dot{x}(t)\| \leq \|\dot{x}(t_0)\|e^{-(\gamma-\beta L)t}.$$

b) Neglecting the nonnegative term $(\gamma - \beta L)\|\dot{x}(t)\|^2$ we obtain that for t large enough

$$\frac{d}{dt}\|\dot{x}(t)\|^2 + \varepsilon\|\dot{x}(t)\| \leq 0. \tag{6.7}$$

Set $v(t) = \|\dot{x}(t)\|^2$. We have $\dot{v}(t) + \varepsilon\sqrt{v(t)} \leq 0$. As long as $v(t) > 0$, we will have $\frac{d}{dt}\sqrt{v(t)} \leq -\frac{\epsilon}{2}$. This forces $v(t)$ to be equal to zero after some finite time. ■

7 Nonsmooth case

In this section, we assume that $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous and proper function such that $\inf_{\mathcal{H}} f > -\infty$. The preceding sections deal with a differentiable function f, without any convexity assumption on f. Now, when considering nonsmooth functions, we assume the convexity of f. This allows us to use the regularity properties of the Moreau envelope in the convex case. Indeed, to reduce to the previous situation, where $f : \mathcal{H} \to \mathbb{R}$ is a $\mathcal{C}^1$ function whose gradient is Lipschitz continuous, the idea is to replace f by its Moreau envelope $f_\lambda : \mathcal{H} \to \mathbb{R}$ which is defined by: for all $x \in \mathcal{H}$,

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

The properties of Moreau envelopes useful for our study were recalled in Section 3. Since the infimal value and the set of minimizers are preserved by taking the Moreau envelope, the idea is to replace f by

f_λ in the previous study, and take advantage of the fact that f_λ is continuously differentiable. The dynamic (IGDH) becomes formally

$$\text{(IGDH)-regularized} \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f_\lambda(x(t))\dot{x}(t) + \nabla f_\lambda(x(t)) \ni 0,\ t \in [t_0, +\infty[.$$

In fact f_λ is only a $\mathcal{C}^{1,1}$ function. The term $\nabla^2 f_\lambda(x(t))\dot{x}(t)$ has to be understood as the derivative in the distribution sense of the Lipschitz continuous function $t \mapsto \nabla f_\lambda(x(t))$. Based on the properties of the Moreau envelope, a direct adaptation of Theorem 2.2 gives the following existence result for the regularized dynamic (IGDH)-regularized .

Theorem 7.1 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous and proper function such that $\inf_{\mathcal{H}} f > -\infty$. Let $\phi : \mathcal{H} \to \mathbb{R}$ be a convex continuous function that satisfies* $(\mathrm{DF})_{\mathrm{r}}$ *and* $\| (\partial\phi)^o (x)\| \leq c(1 + \|x\|)$. *Take $\lambda > 0$ and consider the corresponding differential equation* (IGDH)-regularized .
Suppose that

$$\gamma\lambda > \beta.$$

Then, for any Cauchy data $(x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$, there exists a strong global solution of the differential equation (IGDH)-regularized *satisfying $x(t_0) = x_0$, and $\dot{x}(t_0) = \dot{x}_0$.*

Proof. We have that ∇f_λ is L-Lipschitz continuous with $L = \frac{1}{\lambda}$. So, the condition $\gamma > \beta L$ of Theorems 2.2 and 3.1 becomes $\gamma\lambda > \beta$. Then, just apply the conclusion of these theorems. ■

In the following, we denote by x_λ a solution of (IGDH)-regularized . Note that λ is fixed, it is not intended to go to zero.

Theorem 7.2 *Under the assumptions of Theorem 7.1, let x_λ be a solution of* (IGDH)-regularized . *Then, the following properties are satisfied:*

(i) $\|\dot{x}_\lambda\| \in \mathbb{L}^1(t_0, +\infty; \mathbb{R})$*, and therefore the strong limit $x_{\lambda,\infty} := \lim_{t\to+\infty} x_\lambda(t)$ exists.*

(ii) The limit point $x_{\lambda,\infty}$ is an equilibrium point of (IGDH)-regularized *, i.e.* $-\nabla f_\lambda(x_{\lambda,\infty}) \in \partial\phi(0)$.

(iii) If $-\nabla f_\lambda(x_{\lambda,\infty}) \notin \text{boundary}(\partial\phi(0))$*, then there exists $t_1 \geq 0$ such that $x_\lambda(t) = x_{\lambda,\infty}$ $\forall t \geq t_1$.*

(iv) Set $p(t) := \operatorname{prox}_{\lambda f}(x_\lambda(t))$. We have that $p(\cdot)$ has also a finite length, and $p(t)$ converges strongly to $p_\infty := \operatorname{prox}_{\lambda f}(x_{\lambda,\infty})$. In addition

$$\partial f(p_\infty) + \partial\phi(0) \ni 0.$$

Moreover, if $t \mapsto x_\lambda(t)$ is finitely convergent, then $t \mapsto p(t) := \operatorname{prox}_{\lambda f}(x_\lambda(t))$ is also finitely convergent.

Proof. The proof consists in replacing f by f_λ in Theorem 5.1, and in using that ∇f_λ is $\frac{1}{\lambda}$-Lipschitz continuous. Since the proximal mapping is nonexpansive, we immediately deduce that for any $s, t \geq t_0$

$$\|p(t) - p(s)\| \leq \|x_\lambda(t) - x_\lambda(s)\|. \tag{7.1}$$

This implies that p is Lipschitz continuous, hence almost everywhere differentiable and, for a.e. $t \geq t_0$

$$\|\dot{p}(t)\| \leq \|\dot{x}_\lambda(t))\|.$$

Since $\|\dot{x}_\lambda\| \in \mathbb{L}^1(t_0, +\infty; \mathbb{R})$, we also have $\|\dot{p}\| \in \mathbb{L}^1(t_0, +\infty; \mathbb{R})$. Therefore, the trajectory $p(\cdot)$ has a finite length. It converges strongly towards $p_\infty = \operatorname{prox}_{\lambda f} x_{\lambda,\infty}$. Using the relation

$$\nabla f_\lambda(x_{\lambda,\infty}) \in \partial f(\operatorname{prox}_{\lambda f} x_{\lambda,\infty}) = \partial f(p_\infty),$$

we obtain the approximate optimality property:

$$\partial f(p_\infty) + \partial\phi(0) \ni 0.$$

According to (7.1), if $t \mapsto x_\lambda(t)$ is finitely convergent, then $t \mapsto p(t) := \operatorname{prox}_{\lambda f}(x_\lambda(t))$ is also finitely convergent. ■

8 Perspectives

As a major property, the inertial system (IGDH) provides strong global solutions which converge in finite time. It would be interesting to consider whether these results can be extended to PDE's, nonlinear wave equations, and the modeling of damped shocks. From the optimization point of view, the main challenge is to obtain an accurate estimate of the finite time stabilization. This is based on the study of the condition $-\nabla f(x_\infty) \notin \text{boundary}(\partial\phi(0))$, which is a non-trivial subject. Thus, this paper opens the door to several interesting problems arising in various domains.

References

[1] S. Adly, *A variational approach to nonsmooth dynamics: applications in unilateral mechanics and electronics*, Springer Briefs in Mathematics, 2017.

[2] S. Adly, H. Attouch, *Finite convergence of proximal-gradient inertial algorithms with dry friction damping*, (2019) https://hal.archives-ouvertes.fr/hal-02388038.

[3] S. Adly, H. Attouch, *Finite convergence of proximal-gradient inertial algorithms combining dry friction with Hessian-driven damping*, (2019) https://hal.archives-ouvertes.fr/hal-02423584.

[4] S. Adly, H. Attouch, A. Cabot, *Finite time stabilization of nonlinear oscillators subject to dry friction*, Nonsmooth Mechanics and Analysis, Adv. Mech. Math., 12 (2006), Springer, New York, pp. 289–304.

[5] F. Alvarez, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM J. Control Optim., **38** (4) (2000), pp. 1102–1119.

[6] F. Álvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping*, J. Math. Pures Appl., 81 (8) (2002), pp. 747–779.

[7] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule*, HAL-01551873, (2017), to appear in Mathematical Programming.

[8] H. Attouch, Variational convergence for functions and operators, *Pitman Advanced Publishing Program, Applicable Mathematics Series*, 1984.

[9] H. Attouch, G. Buttazzo, G. Michaille, *Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization.* Second Edition, MOS/SIAM Series on Optimization, MO 17, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, (2014).

[10] H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, 263 (2017), pp. 5412–5458.

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

[12] H. Attouch, A. Cabot, Z. Chbani, H. Riahi, *Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient*, Evolution Equations and Control Theory, 7 (3) (2018), pp. 353–371.

[13] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, 2019. hal-02193846.

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

[15] H. Attouch, Z. Chbani, H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim., 29 (3) 2019, pp. 2227–2256.

[16] H. Attouch, Z. Chbani, H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case* $\alpha \leq 3$, arXiv:1706.05671v1 [math.OC] 2017, ESAIM-COCV (2019) published electronically.

[17] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system*, Commun. Contemp. Math., 2 (1) (2000), pp. 1–34.

[18] H. Attouch, P.E. Maingé, P. Redont, *A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws*, Differential Equations and Applications, 4 (1) (2012), pp. 27–65.

[19] H. Attouch, J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$, SIAM J. Optim., 26 (3) (2016), pp. 1824–1834.

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

[21] H. Attouch, B. F. Svaiter, *A continuous dynamical Newton-Like approach to solving monotone inclusions*, SIAM J. Control Optim., 49 (2011), No. 2, pp. 574–598.

[22] J.-F. Aujol, Ch. Dossal, *Stability of over-relaxations for the Forward-Backward algorithm, application to FISTA*, SIAM J. Optim., 25 (4) (2015), pp. 2408–2433.

[23] J.-F. Aujol, Ch. Dossal, *Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for* $b > 0$, 2017, https://hal.inria.fr/hal-01547251v2.

[24] B. Baji, A. Cabot, *An inertial proximal algorithm with dry friction: finite convergence results*, Set-Valued Analysis, 9 (1) (2006), pp. 1–23.

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

[26] R. I. Bot, E. R. Csetnek, S.C. László, *A second order dynamical approach with variable damping to nonconvex smooth minimization*, to appear in Applicable Analysis, (2018).

[27] R. I. Bot, E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim., 54 (3) (2016), pp. 1423–1443.

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

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

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

[31] A. Chambolle, T. Pock, *An introduction to continuous optimization for imaging*, Acta Numerica, 25 (2016), pp. 161-319.

[32] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Math., 41 (3) (2017), pp. 681–685.

[33] U. Mosco, *Convergence of convex sets and of solutions of variational inequalities* Advances in Math., 3 (1969), pp. 510-585.

[34] U. Mosco, *On the continuity of the Young-Fenchel transform*, J. Math. Anal. Appl. **35** (1971).

[35] B.T. Polyak, *Some methods of speeding up the convergence of iterative methods*, Z. Vylist Math. Fiz., 4 (1964), pp. 1–17.

[36] B.T. Polyak, *Introduction to optimization*. New York: Optimization Software. (1987).

[37] H. Schaefer, *Uber die method der a priori-schranken*, Math. Ann., 126 (1955), pp. 415–416.

[38] B. Shi, S. S. Du, M. I. Jordan, W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, arXiv:submit/2440124[cs.LG] 21 Oct 2018.

[39] 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.