On a second order differential inclusion modeling the FISTA algorithm

Vassilis Apidopoulos, Jean-François Aujol, Charles Dossal

On a second order differential inclusion modeling the FISTA algorithm

Apidopoulos Vassilis
Université de Bordeaux
IMB, UMR 5251, F-33400 Talence, France
vasileios.apidopoulos@math.u-bordeaux.fr

Aujol Jean-François
Université de Bordeaux
IMB, UMR 5251, F-33400 Talence, France
jean-francois.aujol@math.u-bordeaux.fr

Dossal Charles
Université de Bordeaux
IMB, UMR 5251, F-33400 Talence, France
Charles.Dossal@math.u-bordeaux1.fr

May 3, 2017

Abstract

In this paper we are interested in the differential inclusion $0 \in \ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \partial F(x(t))$ in a finite-dimensional Hilbert space $\mathcal{H}$, where F is a sum of two convex, lower semi-continuous functions with one being differentiable with Lipschitz gradient. The motivation of this study is that the differential inclusion models an accelerated version of proximal gradient algorithm called FISTA. In particular we prove existence of a global solution for this inclusion. Furthermore we show that under the condition $b > 3$, the convergence rate of $F(x(t))$ towards the minimimum of F is of order of $o\big(t^{-2}\big)$ and that the solution-trajectory converges to a minimizer of F. These results generalize the ones obtained in the differential setting (where F is differentiable) in [6].

Keywords : Convex optimization, differential inclusion, FISTA algorithm, fast minimization, asymptotic behavior

1 Introduction

In this paper we are interested in the following second order differential inclusion.

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \partial F(x(t)) \ni 0 \tag{DI}$$

with some initial conditions $x(t_0) = x_0 \in \mathcal{H}$ and $\dot{x}(t_0) = v_0 \in \mathcal{H}$. We make the following hypotheses :

H 1. $\mathcal{H}$ is a finite dimensional Hilbert space (e.g. $\mathcal{H} = \mathbb{R}^d, d \geq 1$)

H 2. $t_0 > 0$

H 3. $b > 1$

H 4. $F(= f + g) : \mathcal{H} \longrightarrow \mathbb{R}$ where f, g are lower semi-continuous and convex functions such that $f \in \mathcal{C}^1(\mathcal{H})$ with L-Lipschitz gradient and $F = f + g$ is coercive.

Remark 1. We point out that the hypotheses made on F, ensure the existence of a minimizer of F (which may not be necessarily unique)

The interest of studying this inclusion comes from the fact that it models the FISTA algorithm. In other words the numerical scheme that one can obtain by discretizing (DI) is FISTA. The FISTA algorithm (Fast Iterative Shrinkage-Thresholding Algorithm) consists of an accelerated version of the classical proximal algorithm (Forward-Backward algorithm). It was introduced by Beck and Teboulle in [15], based on ideas of Nesterov in [29] (see also [30]) and Güler [25]. Several versions of FISTA have been studied (depending on the over-relaxation parameters). In our case we focus on the one studied in [21] and will use the name FISTA for this version. Some more details on this algorithm are presented in paragraph 2.2.

In the seminal works of Alvarez [3] and Attouch and al [7], the authors study the following second order differential equation often called "Heavy Ball with Friction" (HBF).

$$\ddot{x}(t) + \gamma \dot{x}(t) + \nabla F(x(t)) = 0 \tag{HBF}$$

where $\gamma \geq 0$ is a non-negative parameter and F is a convex and continuously differentiable function. The interest on studying this differential equation, is that its solution describes the motion of a mass rolling over the graph of F, allowing to explore the different minimum points. It turns out that the values of F over the trajectory, converge asymptotically to its minimum (if one exists for the non-convex case), as well as that the trajectory itself converges to minimizer of F.

Further investigations concerning the asymptotic properties of the solution-trajectory of (HBF), had also been carried out, when the constant term $\gamma \geq 0$ is replaced by a general aymptotical vanishing viscosity term $\gamma(t) \geq 0$ verifying some integrability conditions (see for example [18], [19] and [27])

By extending the analysis for the semi-differential case (when F is not necessarily differentiable), in [5] and [20] the authors study the corresponding differential inclusion :

$$\ddot{x}(t) + \gamma \dot{x}(t) + \partial F(x(t)) \ni 0 \tag{1.1}$$

where $\gamma \geq 0$ and $\mathrm{dom} F$ possibly different from $\mathcal{H}$ (this allows for example g be an indicator function of a closed convex set). This leads to consider new types of solutions to (1.1), other than the classical ones (see definition 2.1 in [5] and in [20]), due to the fact that $\ddot{x}$ can be a Radon-measure. For these solutions it is shown that the same asymptotical properties as the ones obtained in the completely differential setting in [3], hold.

In [2] the authors study a differential inclusion in the same setting as the one that we treat in this work, with the viscosity term $\frac{b}{t}\dot{x}(t)$ replaced by the term $\partial g(\dot{x}(t))$. The authors show the existence and uniqueness of a solution. Moreover by some additional hypotheses, they obtain a finite-time stabilization result concerning the generated trajectory.

In [32] and [6] the authors study, in a possible infinite dimensional Hilbert space $\mathcal{H}$, the differential equation modeling the FISTA algorithm :

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \nabla F(x(t)) = 0 \tag{1.2}$$

where $b > 0$ and F is a continuously differentiable and convex function. They show that under some additional hypotheses on $b > 0$, the solution-trajectory of (1.2) enjoys fast convergence minimization properties over F of a quadratic order. Furthermore in [6] they establish the weak convergence of the trajectory to a minimizer of F.

In a recent paper [10], the authors study the ODE (1.2) with the additional geometrical damping-term $\gamma \nabla^2 F(x(t))\dot{x}(t)$.

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \gamma \nabla^2 F(x(t))\dot{x}(t) + \nabla F(x(t)) = 0 \tag{1.3}$$

where $b, \gamma > 0$. Interestingly, despite the presence of the Hessian of F-term (which demands a higher order of regularity over F), the ODE (1.3) can be written equivalently as a coupled system of two first-order differential equations which makes also sense even when the function F is not differentiable (by considering its subdifferential instead of its gradient) (see also [4] and [8] in the same spirit). For this system they show the existence and uniqueness of a (coupled) solution, for which the same fast convergence properties as in the completely differential setting holds.

Motivated by these works, in this paper we study the differential inclusion (DI). More precisely, the main result of this paper is that (DI) admits a solution to some suitable space. In addition we obtain the same results concerning the asymptotical behavior of this solution, to the ones obtained for the solution of (1.2) in the completely differential setting, (where F is continuously differentiable), as presented in [6].

Our approach consists in discretizing the differential inclusion (DI) which leads to a discrete scheme corresponding to the FISTA algorithm (version considered in [21]). Then we reformulate the problem by defining some suitable interpolate functions. By using some good properties of the FISTA algorithm, we are able to deduce a-priori estimates for the interpolate functions. We then conclude by a compactness argument in some suitable space. This method is also used in [16], [11] and [24] (see Chapter 2 and 3) for approximation of first order dynamical systems.

Remark 2. The inclusion (DI) can be written equivalently as

$$\dot{X}(t) + A(t, X(t)) + H(X(t)) \ni 0 \tag{1.4}$$

with $X(t) = (x(t), \dot{x}(t))^T$, $A(t, (a_1, a_2)) = (-a_2, \frac{b}{t}a_2 + \nabla f(a_1))^T$ and $H(((a_1, a_2))) = (0, \partial g(a_1))^T$ for all $t \geq t_0$ and $a = (a_1, a_2) \in \mathcal{H}^2$. Nevertheless, under this reformulation, the operator H is not necessarily maximal monotone, hence the classical theory for monotone inclusions for existence and uniqueness of a solution of (1.4), can not be applied directly (for more information in this topic, we address the reader to Chapter 3 in [17]).

The organization of this paper is the following. In Section 2, we introduce some standard notions that we use in our analysis. We also recall the FISTA algorithm and some of its basic properties, useful for the proof of existence of a solution for (DI). In Section 3, we present the main Theorem of this paper, concerning the existence of a solution of (DI). Finally in Section 4 we show that the regularity of the solution of (DI) is sufficient to obtain the fast convergence properties of $F(x(t))$ to the minimimum of F and the convergence of the trajectory to a minimizer.

2 Preliminary material

2.1 Basic Notions

We start by recalling some basic tools that will be used in this paper.

Given an interval $I \subset \mathbb{R}^+$, $p \in [1, +\infty]$ and $m \in \mathbb{N}$, we denote as $W^{m,p}(I; \mathcal{H})$ the classical Sobolev space with values on $\mathcal{H}$, i.e. the space of functions in $L^p(I)$ whose distributional derivatives up to order m belong to $L^p(I)$. For a detailed presentation of some properties of these spaces, we address the reader to [1] and [23]

Given a function $G : \mathcal{H} \to \mathbb{R}$, we define its subdifferential, as the multi-valued operator $\partial G : \mathcal{H} \to 2^{\mathcal{H}}$, such that for all $x \in \mathcal{H}$:

$$\partial G(x) = \{z \in \mathcal{H} : \forall y \in \mathcal{H}, G(x) \leq G(y) + \langle z, x - y \rangle\}$$

We also recall the definition of the proximal operator which is the basic tool for FISTA algorithm. If G is a lower semi-continuous, proper and convex function, the proximal operator of G is the operator $\text{Prox}_G \mathcal{H} \longrightarrow \mathbb{R}$, such that :

$$\text{Prox}_G(x) = \underset{y \in \mathcal{H}}{\arg\min} \{G(y) + \frac{\|x - y\|^2}{2}\} \quad , \forall x \in \mathcal{H} \tag{2.1}$$

Here we must point out that the proximal operator is well-defined, since by the hypothesis made on G, for every $x \in \mathcal{H}$, the function $y \to G(y) + \frac{\|x-y\|^2}{2}$, admits a unique minimizer.

Equivalently the proximal operator can be also seen as the resolvent of the maximal monotone operator ∂G, i.e. for all $x \in \mathcal{H}$ and γ a positive parameter we have that :

$$Prox_{\gamma G}(x) = (Id + \gamma \partial G)^{-1}(x) \tag{2.2}$$

For a detailed study concerning the subdifferential and the proximal operator and their properties, we address the reader to [14].

Finally, for a sequence of $\{f_n\}_{n \in \mathbb{N}}$ defined in X^* (the dual of a Banach space X), we will use the classical notation for weak-star convergence to f with the symbol $\rightharpoonup^*$, ($f_n \rightharpoonup^* f$ in X^*) i.e.

$$\langle f_n, \phi \rangle \to \langle f, \phi \rangle \quad \forall \phi \in X \tag{2.3}$$

2.2 The FISTA algorithm

In this paragraph we present different results concerning the FISTA algorithm. The setting is the same as the one settled in the introduction. We furthermore denote with x^* a minimizer of F (we recall that the hypotheses made on F ensure the existence of such a minimizer). For ease of reading the proofs of the results presented here are postponed in the Appendix. For a detailed presentation of some of these results we also address the reader to [21].

Firstly, we recall the FISTA algorithm (version considered in [21]) :

Algorithm 1 FISTA

Let $0 < \gamma \leq \frac{1}{L}$, $\{t_n\}_{n\in\mathbb{N}^*}$ and a real number $a > 0$. We consider the sequences $\{t_n\}_{n\in\mathbb{N}}$ $\{x_n\}_{n\in\mathbb{N}}$, $\{y_n\}_{n\in\mathbb{N}}$, $\{u_n\}_{n\in\mathbb{N}}$, such that $t_0 = 1$, $x_0, x_1 \in \mathcal{H}$ and for every $n \in \mathbb{N}^*$ we set :

$$t_n = \frac{n+a-1}{a} \tag{2.4}$$

$$u_n = (1-t_n)x_{n-1} + t_n x_n \tag{2.5}$$

$$y_n = \left(1 - \frac{1}{t_{n+1}}\right)x_n + \frac{1}{t_{n+1}}u_n \tag{2.6}$$

$$= x_n + a_n(x_n - x_{n-1}) \quad \text{with} \quad a_n = \frac{t_n - 1}{t_{n+1}} \tag{2.7}$$

$$x_{n+1} = T(y_n) \tag{2.8}$$

$$\text{where} \qquad T(x) = Prox_{\gamma g}(x - \gamma\nabla f(x))$$

For our analysis we also consider the following sequences : $\{\delta_n\}_{n\in\mathbb{N}^*}$, $\{v_n\}_{n\in\mathbb{N}^*}$, $\{w_n\}_{n\in\mathbb{N}}$ and $\{E_n\}_{n\in\mathbb{N}^*}$ such that:

$$\delta_n = \frac{1}{2}\|x_n - x_{n-1}\|^2 \tag{2.9}$$

$$v_n = \frac{1}{2}\|u_n - x^*\|^2 \tag{2.10}$$

$$w_n = F(x_n) - F(x^*) \tag{2.11}$$

$$E_n = a_n\delta_n + \gamma w_n \tag{2.12}$$

As already mentioned, the FISTA algorithm consists of accelerating the classical proximal (F-B) algorithm (i.e. where every iteration is generated by evaluating the proximal operator on the previous point), by adding an extra over-relaxation term in the previous point (this is the term $a_n(x_n - x_{n-1})$ in (2.7)). Remarkably, this leads to some fast convergence properties of the sequence $\{w_n\}_{n\in\mathbb{N}}$ to 0 which is of order $O(t^{-2})$, in comparison to the FB algorithm which is of order $O(t^{-1})$ (recently it was shown in [9] that when $a > 2$, the convergence rate of $\{w_n\}_{n\in\mathbb{N}}$, is actually $o(t^{-2})$ asymptotically).

The first versions of FISTA algorithm that were considered, correspond to different choices of the sequence $\{t_n\}_{n\in\mathbb{N}^*}$ such as $t_{n+1} = \frac{n+1}{2}$ (that is (2.4) for $a = 2$) in [29] and $t_{n+1} = \frac{1+\sqrt{4t_n^2+1}}{2}$ in [15]. In a first view the choice on a seems anodyne, nevertheless it turns out that a smart tuning of $a > 2$ allow to deduce some additional properties such as the weak convergence of $\{x_n\}_{n\in\mathbb{N}}$ to a minimizer of F (see Theorem 3 in [21]), or the improvement of the asymptotic convergence rate of $\{w_n\}_{n\in\mathbb{N}}$ up to $o(t^{-2})$ (see Theorem 1 in [9]).

In order to study FISTA we use the following classical Lemma which is based on the L-Lipschitz character of ∇f and the definition of the proximal operator. A proof for this Lemma can also be found in [15] (see Lemma 2.3).

Lemma 2.1. *For any $y \in \mathcal{H}$ and $0 < \gamma \leq \frac{1}{L}$ we have that for every $x \in \mathcal{H}$:*

$$2\gamma(F(x) - F(T(y))) \geq \|T(y) - x\|^2 - \|y - x\|^2 \tag{2.13}$$

Proof. See A.1 in Appendix □

The next Corollary shows the decreasing property of $\{E_n\}_{n\in\mathbb{N}^*}$ and therefore the strong connection between the sequences $\{\delta_n\}_{n\in\mathbb{N}^*}$ and $\{w_n\}_{n\in\mathbb{N}^*}$.

Corollary 2.1. *Let $0<\gamma\leq\frac{1}{L}$ and $\{x_n\}_{n\in\mathbb{N}}$ a sequence generated by FISTA. The sequence $\{E_n\}_{n\in\mathbb{N}^*}$ is decreasing*

Proof. By applying (2.13) with $y=y_n$ and $x=x_n$, we find that for every $n\geq 1$ it holds :

$$\delta_{n+1}-a_n^2\delta_n\leq\gamma(w_n-w_{n+1}) \tag{2.14}$$

and since $a_n\leq 1 \quad \forall n\geq 1$, we deduce that for all $n\geq 1$ it holds :

$$a_{n+1}^2\delta_{n+1}+\gamma w_{n+1}\leq a_n^2\delta_n+\gamma w_n$$

which concludes this corollary.

□

The next Proposition contains some basic results for the different sequences generated by FISTA algorithm. In our setting they provide some suitable bound-estimates for the interpolate functions that we use later. For a detailed presentation of these results and their proofs, we also adress the reader to [21].

Proposition 2.2. *Let $0<\gamma\leq\frac{1}{L}$, $a\geq 2$, $t_n=\frac{n+a-1}{a}$ and the sequence $\{x_n\}_{n\in\mathbb{N}}$ defined by FISTA algorithm. Then there exists a positive constant $C_1>0$ which does not depend on γ, such that for all $n\in\mathbb{N}^*$:*

$$\delta_n\leq C_1\gamma \tag{2.15}$$

Remark 3. Here we must point out that, if in addition we suppose that $a>2$, then there exists some positive constants C_2, C_3 which do not depend on γ such that for all $n\geq 1$ it holds :

$$\|x_n\|\leq C_2 \quad \text{and} \quad \sum_{k=1}^{n}k\delta_k\leq C_3 \tag{2.16}$$

Proof. See A.2 in Appendix □

3 Existence of solution for (DI)

In this section we present the main Theorem of this paper, concerning the existence of a solution of (DI).

Theorem 3.1. *Under the hypotheses **H1.**, **H2.**, **H.3.** and **H.4.** made on F, for $\lambda\in(0,1)$ and any compact set K of $[t_0,+\infty)$, the inclusion (DI) admits a solution x in $W^{2,\infty}_{loc}((t_0,+\infty);\mathcal{H})\cap\mathscr{C}^{1,\lambda}([t_0,T];\mathcal{H})$, such that $x(t_0)=x_0$ and $\dot{x}(t_0)=v_0$, in the sense that for every $T>t_0$ and $\phi\in\mathscr{C}^1_c((t_0,T);\mathbb{R}_+)$, the following inequality holds:*

$$\int_{t_0}^{T}\big(g(x(t))-g(z)\big)\phi(t)dt\leq\int_{t_0}^{T}\langle\ddot{x}(t)+\frac{b}{t}\dot{x}(t)+\nabla f(x(t)),z-x(t)\rangle\phi(t)dt \tag{WFDI}$$

In fact we have that (DI) holds almost everywhere in $[t_0,+\infty)$

Remark 4. In fact, for $\lambda\in(0,1)$and any compact set $K\subset[t_0,+\infty)$, we have that $W^{2,\infty}_{loc}((t_0,+\infty);\mathcal{H})\cap\mathscr{C}^{1,\lambda}(K;\mathcal{H})=W^{2,\infty}_{loc}((t_0,+\infty);\mathcal{H})$, since $W^{2,\infty}_{loc}((t_0,+\infty);\mathcal{H})$ is compactly embedded in $\mathscr{C}^{1,\lambda}(K;\mathcal{H})$. The notation in the Theorem and in the proof stands for pedagogical reasons.

3.1 Discretization of (DI)

As already mentioned, in order to prove the existence of a solution of the differential inclusion (DI), we will study it in a discrete setting. Let $T>t_0$ We discretize (DI) in $[t_0,T]$ (with a time step denoted as $h>0$ and $t_{n-1}=(n-1)h+t_0$) explicitly with respect to f (with $y_n=x_n+\frac{n-1+\frac{t_0}{h}-b}{n-1+\frac{t_0}{h}}(x_n-x_{n-1})$) and implicitly with respect to g.

Let $h\in(0,m)$, where $m=\min\{1,\frac{1}{\sqrt{L}}\}$. We obtain :

$$\frac{x_{n+1}-2x_n+x_{n-1}}{h^2}+\frac{b}{(n-1)h+t_0}\frac{x_n-x_{n-1}}{h}+\nabla f(y_n)+\partial g(x_{n+1})\ni 0 \qquad \text{with} \quad x_0=x(t_0) \quad \text{and} \quad x_1=hv_0+x_0 \tag{3.1}$$

We remark that the scheme (3.1) is equivalent to the FISTA algorithm as considered in paragraph 2.1, with $a=b-1\geq 2$ and $\gamma=h^2$. In other words, given the starting points $x_0,x_1\in\mathcal{H}$, $\forall n\geq 1$, we have :

$$\begin{aligned} y_n &= x_n+\frac{n-1}{n+a}(x_n-x_{n-1}) \qquad \text{and} \\ x_{n+1} &= \mathrm{Prox}_{\gamma g}(y_n-\gamma\nabla f(y_n)) \end{aligned} \tag{3.2}$$

Indeed by (3.1) we have that :

$$x_{n+1}+h^2\partial g(x_{n+1})\ni x_n+\frac{n-1+\frac{t_0}{h}-b}{n-1+\frac{t_0}{h}}(x_n-x_{n-1})-h^2\nabla f(y_n)=y_n-h^2\nabla f(y_n) \tag{3.3}$$

which by characterization of the proximal operator (2.2) for all $n\geq 1$, gives :

$$\begin{aligned} y_n &= x_n+\frac{n-1+\frac{t_0}{h}-b}{n-1+\frac{t_0}{h}}(x_n-x_{n-1}) \qquad \text{and} \\ x_{n+1} &= \mathrm{Prox}_{h^2 g}(y_n-h^2\nabla f(y_n)) \end{aligned} \tag{3.4}$$

which (up to a re-indexation $n\rightsquigarrow n-1-\frac{t_0}{h}+b-1,$) is also equivalent to (3.2) with $\gamma=h^2$ and $b=a+1$.

As mentioned before, the strategy that we follow, is to define some suitable interpolate functions that model (3.1), which by using some compactness arguments will converge to a limit-function solution of (DI) in the sense as given in Theorem 3.1.

In fact, the discrete scheme (3.4) of (DI) (or equivalently the FISTA algorithm), provides some suitable a-priori estimates for the corresponding generated sequences $\{x_n\}_{n\geq 1}$. Then we conclude with some compactness argument in some suitable space.

We first define the interpolate functions which model (3.1).

3.2 Interpolate functions

By discretization with a time-step $h\in(0,m)$, for all $t\in[t_0,T]$, we set $n=\lfloor\frac{t-t_0}{h}\rfloor+1$ and $t_{n-1}=(n-1)h+t_0$. Let $\{x_n\}_{n\in\mathbb{N}}$ the corresponding sequences generated by FISTA (in what follows, we do not denote the dependency of the generated sequences $\{x_n\}_{n\in\mathbb{N}}$ on the time-step h). For all $t\in[t_0,T]$, we introduce the following interpolate functions :

$$\tilde{x}_h(t)=x_{n+1} \quad \forall t\in[t_{n-1},t_n) \tag{3.5}$$

$$\text{and} \quad \tilde{y}_h(t)=y_n=x_n+a_n(x_n-x_{n-1}) \quad \forall t\in[t_{n-1},t_n) \tag{3.6}$$

with $\tilde{x}_h(t_0)=\tilde{y}_h(t_0)=x_1$ (we recall that $a_n=\frac{n-1}{n+a}$). These functions are both piecewise constant.

We also define :

$$\hat{x}_h(t)=\frac{x_n+x_{n-1}}{2}+\frac{x_n-x_{n-1}}{h}(t-t_{n-1})+\frac{x_{n+1}-2x_n+x_{n-1}}{2h^2}(t-t_{n-1})^2 \quad \forall t\in[t_{n-1},t_n) \tag{3.7}$$

with $\hat{x}_h(t_0)=\frac{x_1+x_0}{2}$ and $\dot{\hat{x}}_h(t_0)=\frac{x_1-x_0}{h}=v_0$.

In fact, by construction, the function $\hat{x}_h$ is $\mathscr{C}^1([t_0,T];\mathcal{H})$ and piecewise $\mathscr{C}^2([t_0,T];\mathcal{H})$ with:

$$\dot{\hat{x}}_h(t)=\frac{x_n-x_{n-1}}{h}+\frac{x_{n+1}-2x_n+x_{n-1}}{h^2}(t-t_{n-1}) \quad , \quad t_{n-1}\leq t<t_n \tag{3.8}$$

$$\text{and} \quad \ddot{\hat{x}}_h(t)=\frac{x_{n+1}-2x_n+x_{n-1}}{h^2} \quad , \quad t_{n-1}<t<t_n \tag{3.9}$$

By definition of these functions and relation (3.1), it follows that

$$\left(1-\frac{b}{(n-1)h+t_0}(t-t_{n-1})\right)\ddot{\tilde{x}}_h(t)+\frac{b}{(n-1)h+t_0}\dot{\tilde{x}}_h(t)+\nabla f(\tilde{y}_h(t))+\partial g(\tilde{x}_h(t))\ni 0 \quad \text{a.e. in } [t_0,T] \tag{3.10}$$

where $n=\lfloor\frac{t-t_0}{h}\rfloor+1$, $\forall t\in[t_0,T]$ and $t_{n-1}=(n-1)h+t_0$.

By definition of ∂g and by multiplying by a (positive) test-function $\phi\in\mathscr{C}_c^1((t_0,T);\mathbb{R}^+)$ and then integrating in (t_0,T), we obtain that for all $z\in\mathcal{H}$ it holds :

$$\int_{t_0}^{T}\big(g(\tilde{x}_h(t))-g(z)\big)\phi(t)dt\le\int_{t_0}^{T}\langle(1-\frac{b}{t_{n-1}}(t-t_{n-1}))\ddot{\tilde{x}}_h(t)+\frac{b}{t_{n-1}}\dot{\tilde{x}}_h(t)+\nabla f(\tilde{y}_h(t)),z-\tilde{x}_h(t)\rangle\phi(t)dt \tag{3.11}$$

We now turn our attention into finding some suitable a-priori estimates for these interpolate functions.

3.3 A-priori estimates

In this paragraph we derive some useful estimates for the interpolate functions considered before, in order to show their convergence in some suitable space, to a limit-function solution to (DI) by some compactness argument.

We recall that from FISTA's analysis in Section 2 and in particular from Proposition 2.2, there exist a positive constant $C_1>0$ which do not depend on h, such that :

$$\|x_n-x_{n-1}\|\le C_1h \quad ,\forall n\in\mathbb{N}^* \tag{3.12}$$

In fact by (3.12) and the discretization scheme, for all $t\in[t_0,T]$, we have :

$$\|x_n\|\le C_1nh=C_1(\lfloor\frac{t-t_0}{h}\rfloor+1)h\le C_1(T-t_0)+C_1h\le C_T \tag{3.13}$$

for a positive constant $C_T>0$,depending on T and independent of h, given that $h<m$.

In particular we have the following Lemma showing the control level according to the step h that we obtain for the family of the interpolate functions $\{\tilde{x}_h\}_{0<h<m}$, $\{\tilde{y}_h\}_{0<h<m}$ and $\{\hat{x}_h\}_{0<h<m}$.

Lemma 3.1. *For any $h\in(0,m)$ there exists a positive constant C_T (depending on T) which does not depend on h such that :*

$$\sup_{h\in(0,m)}\{\{\|\tilde{x}_h\|_\infty,\|\tilde{y}_h\|_\infty,\|\hat{x}_h\|_\infty\}<C_T \tag{3.14}$$

Proof. In fact by definition of function $\tilde{x}_h$ and relation (3.13), it follows that for all $h\in(0,m)$ and $t\in[t_0,T]$ we have :

$$\|\tilde{x}_h(t)\|=\|x_{n+1}\|\le C_T<+\infty$$

So that $\|\tilde{x}_h\|_\infty\le C_T$.

By definition of function $\hat{x}_h$, for all $h\in(0,m)$ and $t\in[t_0,T]$, we also have :

$$\begin{aligned}\|\hat{x}_h(t)\|&\le\frac{\|x_n\|+\|x_{n-1}\|}{2}+\frac{\|x_n-x_{n-1}\|}{h}|t-t_{n-1}|+\frac{(\|x_{n+1}-x_n\|+\|x_n-x_{n-1}\|)}{2h^2}|t-t_{n-1}|^2\\&\le C_T+2C_1h\le C_T\end{aligned} \tag{3.15}$$

where C_T is a (renamed) positive constant which does not depend on h.

Similarly for $\tilde{y}_h$, since $a_n\le 1$, for all $h\in(0,m)$ and $t\in[t_0,T]$, we have :

$$\|\tilde{y}_h(t)\|\le\|x_n\|+\|x_n-x_{n-1}\|\le C_T+C_1h\le C_T \tag{3.16}$$

In other words, for the families of functions $\{\tilde{x}_h\}_{0<h<m}$, $\{\tilde{y}_h\}_{0<h<m}$ and $\{\hat{x}_h\}_{0<h<m}$, we deduce that they are uniformly bounded on $[t_0,T]$ with respect to h, i.e. :

$$\sup_{h\in(0,m)}\{\|\tilde{x}_h\|_\infty,\|\tilde{y}_h\|_\infty,\|\hat{x}_h\|_\infty\}<+\infty \tag{3.17}$$

□

The next Lemma shows that a uniform bound according to h, also holds for the family of the first derivative of the interpolate functions $\{\dot{\hat{x}}_h\}_{0<h<m}$.

Lemma 3.2. *For any $h \in (0, m)$ there exists a positive constant c_1 which do not depend on h and T such that :*

$$\sup_{0<h<m} \{\|\dot{\hat{x}}_h\|_\infty\} \leq c_1 \tag{3.18}$$

Proof. For all $h \in (0, m)$ and $t \in [t_0, T]$ we have :

$$\|\dot{\hat{x}}_h(t)\| \leq \frac{\|x_n - x_{n-1}\|}{h} + \frac{(\|x_{n+1} - x_n\| + \|x_n - x_{n-1}\|)}{h^2}|t - t_{n-1}| \leq 3\frac{C_1 h}{h} \leq 3C_1 \tag{3.19}$$

Since the last inequality holds true for all $t \in [t_0, T]$, for all $h \in (0, m)$, we deduce that the family of functions $\{\dot{\hat{x}}_h\}_{0<h<m}$ is uniformly bounded with respect to h (and T), i.e. :

$$\sup_{0<h<m} \{\{\|\dot{\hat{x}}_h\|_\infty\}\} < +\infty \tag{3.20}$$

□

Remark 5. In fact by Remark 3 we have that for $a > 2$, there exist some positive constants C_2 and C_3 which do not depend on γ such that for all $n \geq 1$ it holds $\sum_{k=1}^{n} k\delta_k \leq C_3$. From these bounds one can deduce that $\hat{x}_h$ is uniformly bounded in $L^\infty((t_0, +\infty); \mathcal{H})$ and $\dot{\hat{x}}_h$ is uniformly bounded in $L^2((t_0, +\infty); \mathcal{H})$ with respect to h. In the present work we do not make use of these estimates, nevertheless it may be a useful estimation for future studies.

Finally by exploiting the structure of (3.10) and the properties of the subdifferential, we find some uniform estimates for the second derivative of x_h.

Corollary 3.1. *For $h \in (0, m)$ we have*

$$\sup_{0<h<m} \{\{\|\ddot{\hat{x}}_h\|_\infty\}\} < +\infty \tag{3.21}$$

Proof. By Lemma 3.1 as the family $\{\tilde{x}_h\}_{0<h<m}$ is uniformly bounded with respect to h, $\partial g(\tilde{x}_h)$ is also uniformly bounded on h (see Lemma (A.1) in Appendix). Since by Lemma (3.2) $\{\dot{\hat{x}}_h\}_{0<h<m}$ is also uniformly bounded on h, it follows from (3.10) that $\{\ddot{\hat{x}}_h\}_{0<h<m}$ is also uniformly bounded on h, which concludes the proof of Corollary 3.1 □

Remark 6. Here we must stress out the importance of the hypothesis that $\text{dom}F = \mathcal{H}$. This allows to use Lemma A.1. When $\text{dom}F \subsetneqq \mathcal{H}$ we can not expect a uniform bound for the second derivative. Such a case corresponds for example to the one when g is the indicator function of a closed convex set. In that case as the trajectory must be in $\text{dom}F$, a shock may occur on the boundary of $\text{dom}F$, which explains why we can not expect that the acceleration stays uniformly bounded (for more details on this phenomenon, see [5] and [20]).

3.4 Proof of Theorem 3.1

We are now ready to give the proof of Theorem 3.1. As mentioned before, we exploit the different estimates obtained in the previous section, in order to deduce convergence of the interpolate functions to a solution of the problem (DI).

Proof. Let $T > t_0$. By Corollary 3.1, we deduce that $\hat{x}_h$ is uniformly bounded in $W^{2,\infty}((t_0, T); \mathcal{H})$. Since for any $\lambda \in (0, 1)$, $W^{2,\infty}((t_0, T); \mathcal{H})$ is compactly embedded in $\mathscr{C}^{1,\lambda}([t_0, T]; \mathcal{H})$ (see Theorem 6.3 in [1]), up to a subsequence still denoted by $\{\hat{x}_h\}_{0<h<m}$, we have that :

$$\hat{x}_h \underset{h \to 0}{\longrightarrow} \hat{x} \quad \text{in } \mathscr{C}^{1,\lambda}([t_0, T]; \mathcal{H}) \quad \text{where} \quad 0 < \lambda < 1 \tag{3.22}$$

Furthermore, as $\ddot{\hat{x}}_h$ is bounded in $L^\infty((t_0, T); \mathcal{H})$ and $L^\infty((t_0, T); \mathcal{H})$ can be identified with the dual space of $L^1((t_0, T); \mathcal{H})$, we also have (that is the Banach-Alaoglu Theorem) that, up to a subsequence (here we extract from the subsequence considered before) still denoted by $\{\ddot{\hat{x}}_h\}_{0<h<m}$, :

$$\ddot{\hat{x}}_h \rightharpoonup^* u \quad \text{in } L^\infty((t_0, T); \mathcal{H}) \tag{3.23}$$

where by uniqueness of the limit (in the distributional sense) we have that $\ddot{\hat{x}} \equiv u \in L^\infty((t_0,T);\mathcal{H})$. Hence for $0<\lambda<1$, we have that $\hat{x} \in \mathscr{C}^{1,\lambda}([t_0,T];\mathcal{H}) \cap W^{2,\infty}((t_0,T);\mathcal{H})$.

We also have that :

$$\hat{x}_h(t_0) = \frac{x_1+x_0}{2} = x_0 + \frac{v_0}{2}h \underset{h\to 0}{\longrightarrow} x_0 \tag{3.24}$$

which shows that $\hat{x}(t_0)=x_0$.

Furthermore by construction we have that :

$$\dot{\hat{x}}_h(t_0) = \frac{x_1-x_0}{h} = v_0 \underset{h\to 0}{\longrightarrow} v_0 \tag{3.25}$$

which shows that $\dot{\hat{x}}(t_0)=v_0$.

In fact for all $i \in \mathbb{N}^*$, one can construct the sequences (of sequences) of functions $\{\{x^i_{h(n)}\}_{n\in\mathbb{N}}\}_{i\in\mathbb{N}}$ as follows :

$$\begin{aligned}
&\hat{x}^1_{h(n)} \underset{n\to\infty}{\longrightarrow} \hat{x}^1 && \in W^{2,\infty}([t_0,t_0+1])\\
&\hat{x}^2_{h(n)} \underset{n\to\infty}{\longrightarrow} \hat{x}^2 && \in W^{2,\infty}([t_0,t_0+2])\\
&\vdots\\
&\hat{x}^i_{h(n)} \underset{n\to\infty}{\longrightarrow} \hat{x}^i && \in W^{2,\infty}([t_0,t_0+i])
\end{aligned} \tag{3.26}$$

in a way that every time we extract a subsequence $\{\hat{x}^{i+1}_{h(n)}\}_{n\in\mathbb{N}}$ from the subsequence considered before $\{\hat{x}^{i}_{h(n)}\}_{n\in\mathbb{N}}$, for every $i\in\mathbb{N}^*$. By diagonal extraction we consider the sequence of functions $\{\hat{x}^{n}_{h(n)}\}_{n\in\mathbb{N}}$. We then define the sequence of functions $\{w_n\}_{n\in\mathbb{N}}$ in $[t_0,+\infty)$, as the $W^{2,\infty}([t_0+n,+\infty))$ extensions of $\hat{x}^n_{h(n)}$, for all $n\in\mathbb{N}$. By this construction there exists a function $\hat{x}:[t_0,+\infty)\longrightarrow\mathcal{H}$ such that the sequence of functions $\{w_n\}_{n\in\mathbb{N}}$ converges to with respect to the $W^{2,\infty}_{loc}([t_0,+\infty))$ norm. This shows that for $\lambda\in(0,1)$ and a compact set $K\subset[t_0+\infty)$, $\hat{x}\in W^{2,\infty}_{loc}((t_0,+\infty);\mathcal{H})\cap\mathscr{C}^{1,\lambda}(K;\mathcal{H})$.

In addition by definition of $\tilde{x}_h$, $\tilde{y}_h$ and $\hat{x}_h$ for all $h\in(0,m)$, for all $t\in[t_0,T]$, we have :

$$\begin{aligned}
\|\tilde{x}_h(t)-\hat{x}_h(t)\| &= \|x_{n+1} - \frac{x_n+x_{n-1}}{2} - \frac{x_n-x_{n-1}}{h}(t-t_n) - \frac{x_{n+1}-2x_n+x_{n-1}}{2h^2}(t-t_n)^2\|\\
&\le \|x_{n+1}-x_n\| + 2\|x_n-x_{n-1}\| + \frac{\|x_{n+1}-x_n\|}{2}\\
&\le \frac{7C_1h}{2}
\end{aligned} \tag{3.27}$$

$$\text{so that} \quad \|\tilde{x}_h-\hat{x}_h\|_\infty \le \frac{7C_1h}{2} \underset{h\to 0}{\longrightarrow} 0 \tag{3.28}$$

In the same way :

$$\|\tilde{x}_h-\tilde{y}_h\|_\infty \le 2C_1h \underset{h\to 0}{\longrightarrow} 0 \tag{3.29}$$

which implies that :

$$\tilde{x}_h \underset{h\to 0}{\longrightarrow} \hat{x} \quad \text{and} \quad \tilde{y}_h \underset{h\to 0}{\longrightarrow} \hat{x} \quad \text{in } L^\infty((t_0,T);\mathcal{H}) \tag{3.30}$$

In addition as ∇f is L-Lipschitz in every bounded set, we also have :

$$\nabla f(y_h) \underset{h\to 0}{\longrightarrow} \nabla f(\hat{x}) \quad \text{in } L^\infty((t_0,T);\mathcal{H}) \tag{3.31}$$

By (3.22), (3.23), (3.30) and (3.31), when $h\to 0$, we deduce that the Right-Hand-Side of (3.11) converges to:

$$\int_{t_0}^T \langle \ddot{\hat{x}}(t) + \frac{b}{t}\dot{\hat{x}}(t) + \nabla f(\hat{x}(t)), z-\hat{x}(t)\rangle\phi(t)dt \tag{3.32}$$

In addition by lower semi-continuity of g we also have that for all $t\in[t_0,T]$:

$$g(\hat{x}(t)) \le \liminf_{h\to 0} g(\tilde{x}_h(t)) \tag{3.33}$$

Finally by taking $h \to 0$ in (3.11), for all $\phi \in \mathscr{C}_c^1((t_0,T);\mathbb{R}^+)$ and $z \in \mathcal{H}$, we obtain :

$$\int_{t_0}^{T} (g(\hat{x}(t)) - g(z))\phi(t)dt \le \int_{t_0}^{T} \langle \ddot{\hat{x}}(t) + \frac{b}{t}\dot{\hat{x}}(t) + \nabla f(\hat{x}(t)), z - \hat{x}(t)\rangle \phi(t)dt \tag{3.34}$$

Since this last inequality holds true for all $\phi \in \mathscr{C}_c^1((t_0,T);\mathbb{R}^+)$ and all the integrated functions are in $L^1((t_0,T);\mathcal{H})$, for all $z \in \mathcal{H}$, we have that :

$$g(\hat{x}(t)) - g(z) \le \langle \ddot{\hat{x}}(t) + \frac{b}{t}\dot{\hat{x}}(t) + \nabla f(\hat{x}(t)), z - \hat{x}(t)\rangle \quad \text{a.e. in } [t_0,T] \tag{3.35}$$

Since the last inequality stands for all $T > t_0$, it holds almost everywhere in $[t_0,+\infty)$. By definition of the subdifferential we finally obtain :

$$\ddot{\hat{x}}(t) + \frac{b}{t}\dot{\hat{x}}(t) + \nabla f(\hat{x}(t)) + \partial g(\hat{x}(t)) \ni 0 \quad \text{a.e. in } [t_0,+\infty) \tag{3.36}$$

where $\hat{x} \in \mathscr{C}^{1,\lambda}([t_0,T];\mathcal{H}) \cap W^{2,\infty}((t_0,T);\mathcal{H})$ for all $T > t_0$, which is equivalent to (DI) and concludes the proof of the Theorem (3.1).

□

4 Asymptotic behavior of the trajectory

4.1 Energy estimates

In this section for a given a solution $x \in W^{2,\infty}_{loc}((t_0,+\infty);\mathcal{H})$ of the inclusion (DI) and a minimizer of F which we denote as x^*, we are interested in asymptotic properties of the trajectory $\{x(t) : t \in [t_0,+\infty)\}$. W denote $W(t) = F(x(t)) - F(x^*)$.

For $\lambda \ge 0$ and $\xi \ge 0$ we define the following energy-function :

$$\mathcal{E}_{\lambda,\xi}(t) = t^2 W(t) + \frac{1}{2}\|\lambda(x(t)-x^*) + t\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t)-x^*\|^2 \tag{4.1}$$

This function can be seen as the negative entropy up to the balanced distance $\frac{\xi}{2}\|x(t)-x^*\|^2$. This functional was considered in [32] and in [6] in order to deduce some fast convergence asymptotic behavior for $W(t)$ and $\|\dot{x}(t)\|$ as well as the convergence of the trajectory to a minimizer x^* . Here in the same way, one can obtain the same fast convergence properties for a solution of (DI). The difficulty comes from the fact that the solution is not everywhere differentiable, hence we can not differentiate directly $\mathcal{E}$. Nevertheless the regularity of the solution is sufficient to deduce the same bound estimates for $W(t)$ and $\|\dot{x}(t)\|$ as in [6].

First, we recall the definition of an absolutely continuous function that we will use later (see for instance Example 1.13 in [22]).

Definition 4.1. Let $[a,b]$ be an interval in $[t_0,+\infty)$. A function $G : [a,b] \longrightarrow \mathbb{R}$ is said to be absolutely continuous if for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every finite collection $\{[a_i,b_i]\}_{i\in J}$ of disjoint subintervals of $[a,b]$, we have

$$\sum_{i\in J}\big(b_i - a_i\big) < \delta \quad \Longrightarrow \quad \sum_{i\in J}|G(b_i) - G(a_i)| < \varepsilon \tag{4.2}$$

Equivalently a function $G : [a,b] \longrightarrow \mathbb{R}$ is absolutely continuous if there exists a function $v \in L^1(a,b)$, such that

$$G(t) = G(s) + \int_s^t v(\tau)d\tau \quad \forall a \le s \le t \le b \tag{4.3}$$

and in that case we have that G is differentiable a.e. in (a,b) with $\dot{G}(t) = v(t)$ a.e. in (a,b).

Remark 7. From the definition of absolute continuity (in particular (4.3)), it follows that an absolutely continuous function with non-positive derivative a.e. in (a,b) is non-increasing.

Next we give the following Lemma which can be found in [17] (Lemme 3.3) and allows us to " use the chain rule for differentiation" and show the non-increasing property of the energy $\mathcal{E}$ in $[t_0,+\infty)$.

Lemma 4.1. *Let $T > t_0$ and F be a convex, lower semi-continuous, proper function and $x \in W^{1,2}((t_0, T); \mathcal{H})$. Let also $h \in L^2((t_0, T); \mathcal{H})$, such that $h \in \partial F(x(t))$ a.e. in (t_0, T). Then the function $F \circ x : [t_0, T] \longrightarrow \mathbb{R}$ is absolutely continuous in $[t_0, T]$ with :*

$$\frac{d}{dt}\big(F(x(t))\big) = \langle z, \dot{x}(t)\rangle \quad \forall z \in \partial F(x(t)) \qquad a.e.\ in\ (t_0, T) \tag{4.4}$$

In fact, for any $T > t_0$, if x is a solution of (DI) in $W^{2,\infty}((t_0, T); \mathcal{H})$, we have in particular that $x \in W^{1,2}((t_0, T))$ and that the function $h(t) = -\ddot{x}(t) - \frac{b}{t}\dot{x}(t)$ is in $L^2((t_0, T); \mathcal{H})$.

In view of Lemma 4.1, $W(t)$ is absolutely continuous in $[t_0, T]$ with :

$$\dot{W}(t) = \langle z, \dot{x}(t)\rangle \quad \forall z \in \partial F(x(t)) \qquad \text{a.e. in } (t_0, T)$$

In addition as $\dot{x} \in W^{1,\infty}((t_0, T); \mathcal{H})$, it is in particular Lipschitz continuous in (t_0, T) (see characterization of $W^{1,\infty}$ space, for example Theorem 4.1 in [26] or Theorem 4.5 in [23]), therefore it is absolutely continuous in $[t_0, T]$ (this follows from definition of absolute continuity). As a consequence we have the following proposition.

Proposition 4.1. *Let $T > t_0$. The energy $\mathcal{E}_{\lambda,\xi}$ is absolutely continuous on $[t_0, T]$ with :*

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \le (2-\lambda)tW(t) + \big(\xi + \lambda(\lambda+1-b)\big)\langle \dot{x}(t), x(t) - x^*\rangle + (\lambda+1-b)t\|\dot{x}(t)\|^2 \quad a.e\ in\ (t_0, T) \tag{4.5}$$

Proof. By definition $\mathcal{E}_{\lambda,\xi}$ is an absolutely continuous as sum of absolutely continuous functions. In addition, by Lemma 4.1, let $z \in \partial F(x(t))$ such that (DI) holds. We have

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) &= 2tW(t) + t^2\langle z, \dot{x}(t)\rangle + \langle (\lambda+1)\dot{x}(t) + t\ddot{x}(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle \\ &\quad + \xi\langle x(t) - x^*, \dot{x}(t)\rangle \quad \text{a.e in } (t_0, T)\end{aligned} \tag{4.6}$$

By using that $x(t)$ is a solution of (DI), we obtain :

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) &= 2tW(t) - \lambda t\langle z, x(t) - x^*\rangle + \big(\xi + \lambda(\lambda+1-b)\big)\langle \dot{x}(t), x(t) - x^*\rangle \\ &\quad + (\lambda+1-b)t\|\dot{x}(t)\|^2 \quad \text{a.e in } (t_0, T)\end{aligned} \tag{4.7}$$

Finally by using that $z \in \partial F(x(t))$ by definition of the subdifferential, we deduce that :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \le (2-\lambda)tW(t) + \big(\xi + \lambda(\lambda+1-b)\big)\langle \dot{x}(t), x(t) - x^*\rangle + (\lambda+1-b)t\|\dot{x}(t)\|^2 \quad \text{a.e in } (t_0, T) \tag{4.8}$$

which concludes the proof of Proposition 4.1 □

Corollary 4.2. *For any $\xi = \lambda(b - \lambda - 1)$ and $2 \le \lambda \le b - 1$, the energy-function $\mathcal{E}_{\lambda,\xi}$ is non-increasing in $[t_0, +\infty)$*

Proof. By relation (4.5) of Proposition 4.1, if we choose $\xi = \lambda(b - \lambda - 1)$ and $2 \le \lambda \le b - 1$, as $b \ge 3$, we have that :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \le 0 \qquad \text{a.e in } (t_0, T) \tag{4.9}$$

Since $\mathcal{E}$ is absolutely continuous on $[t_0, T]$ with non-positive derivative a.e. in (t_0, T), we deduce that $\mathcal{E}$ is non-increasing in $[t_0, T]$. Since this is true for every $T > t_0$, in view of continuity of $\mathcal{E}$, we have that $\mathcal{E}$ is non-increasing in $[t_0, +\infty)$. □

Remark 8. Here we must point out that the absolute continuity of $\mathcal{E}$ is essential, since by (4.9), and supposing only continuity of $\mathcal{E}$, one cannot conclude that $\mathcal{E}$ is non-increasing.

In view of the previous Lemma and the non-increasing property of $\mathcal{E}$, as in [6], we have the following Theorem.

Theorem 4.1. *Let $x \in W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H})$ a solution of (DI) and x^* a minimizer of F. Then if $b \ge 3$, there exist some positive constants $C_1, C_2 > 0$, such that for all $t \in [t_0, +\infty)$, it holds :*

$$W(t) \le \frac{C_1}{t^2} \qquad and \qquad \|\dot{x}(t)\| \le \frac{C_2}{t} \tag{4.10}$$

In addition if $b > 3$, we have :

$$\int_{t_0}^{+\infty} tW(t)dt < +\infty \quad \textit{and} \quad \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty \tag{4.11}$$

Proof. Let $b \geq 3$. As mentioned before for $\lambda = 2$ and $\xi = \lambda(b - \lambda - 1) = 2(b-3) \geq 0$, the energy $\mathcal{E}_{2,2(b-3)}$ is non-increasing in $[t_0, +\infty)$, hence for all $t \in [t_0, +\infty)$ we have :

$$t^2 W(t) \leq \mathcal{E}_{2,2(b-3)}(t) \leq \mathcal{E}_{2,2(b-3)}(t_0) \tag{4.12}$$

This shows that $W(t)$ converges asymptotically to 0 which is equivalent to the convergence of $F(x(t))$ to $F(x^*)$. In view of this convergence property, together with coercivity of F we deduce that $\|x(t) - x^*\|$ is bounded. Hence by definition of $\mathcal{E}$, we have also that for all $t \in [t_0, +\infty)$, it holds :

$$t^2\|\dot{x}(t)\|^2 \leq \mathcal{E}_{2,2(b-3)}(t_0) + C\sup_{t\geq t_0}\|x(t) - x^*\|^2 \tag{4.13}$$

for a positive constant $C > 0$. This concludes the first part of the Theorem.

For the second part, if $T > t_0$ and $b > 3$, choosing $\lambda = b - 1$ and $\xi = \lambda(b - \lambda - 1) = 0$, by integrating (4.5) in $[t_0, T]$, we obtain :

$$(b-3)\int_{t_0}^{T} tW(t)dt \leq \mathcal{E}_{b-1,0}(t_0) - \mathcal{E}_{b-1,0}(T) \leq \mathcal{E}_{b-1,0}(t_0) \tag{4.14}$$

In a similar way if we choose $\lambda = 2$ and $\xi = \lambda(b - \lambda - 1) = 2(b-3) > 0$, by integrating (4.5) in $[t_0, T]$, we obtain :

$$(b-3)\int_{t_0}^{T} t\|\dot{x}(t)\|^2 dt \leq \mathcal{E}_{2,2(b-3)}(t_0) - \mathcal{E}_{2,2(b-3)}(T) \leq \mathcal{E}_{2,2(b-3)}(t_0) \tag{4.15}$$

Since (4.14) and (4.15) hold for every $T > t_0$, we can conclude the second part of Theorem 4.1. □

Corollary 4.3. *Let $x \in W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H})$ be a solution of* (DI) *and x^* a minimizer of F. If $b \geq 3$ then $x \in W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H}) \cap W^{1,\infty}((t_0, +\infty); \mathcal{H})$*

Proof. As mentioned before from Theorem 4.1 we have that $F(x(t))$ converges asymptotically to $F(x^*)$, so it is bounded in $[t_0, +\infty)$. Since it is coercive we deduce that $\|x(t)\|$ is also bounded in $[t_0, +\infty)$.In addition from (4.10) of Theorem 4.1, we have that $\|\dot{x}(t)\|$ is also bounded in $[t_0, +\infty)$, which concludes the proof. □

4.2 Fast asymptotic convergence to a minimimum

The last Theorem asserts that for $b \geq 3$, $W(t)$ and $\|\dot{x}(t)\|^2$ is of order of $O(t^2)$ asymptotically. Nevertheless for $b > 3$, this order can be improved to one of $o(t^{-2})$. In fact, as before, the regularity of the solution allows to proceed as the analysis carried out in the differential case(where F is differentiable) in [28].

Theorem 4.2. *Let $b > 3$, $x \in W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H})$ a solution of* (DI) *and x^* a minimizer of F. Then*

$$\lim_{t\to\infty} t^2 W(t) = 0 \quad \textit{and} \quad \lim_{t\to\infty} t\|\dot{x}(t)\| = 0 \tag{4.16}$$

In other words : $W(t) = o(t^{-2})$ and $\|\dot{x}(t)\| = o(t^{-1})$

Proof. First of all, we consider the following energy function :

$$U(t) = t^2 W(t) + \frac{t^2}{2}\|\dot{x}(t)\|^2 \geq 0 \quad , \forall t \in [t_0, +\infty) \tag{4.17}$$

Let $T > t_0$. U is absolutely continuous on $[t_0, T]$ as sum of absolutely continuous functions. In addition by Lemma 4.1 for a $z \in \partial F(x(t))$ such that (DI) holds, we have :

$$\frac{d}{dt}U(t) = t^2\langle z, \dot{x}(t)\rangle + t^2\langle \ddot{x}, \dot{x}(t)\rangle + 2tW(t) + t\|\dot{x}(t)\|^2 \quad \text{a.e. in } (t_0, T) \tag{4.18}$$

By using (DI) and $b > 3$, we find

$$\frac{d}{dt}U(t) = 2tW(t) + (1-b)t\|\dot{x}(t)\|^2 \leq 2tW(t) \quad \text{a.e. in } (t_0, T) \tag{4.19}$$

If we consider the positive part of $\frac{d}{dt}U(t)$, i.e. $[\frac{d}{dt}U(t)]_+(t) = max\{0, \frac{d}{dt}U(t)(t)\}$, for all $t \geq t_0$, we obtain :

$$\left[\frac{d}{dt}U(t)\right]_+ \leq 2tW(t) \quad \text{a.e. in } (t_0, T) \tag{4.20}$$

By Theorem 4.1 for $b > 3$, the term $2tW(t)$ is integrable on $[t_0, T)$ for all $T > t_0$, and so is $\left[\frac{d}{dt}U(t)\right]_+$.

We define the function $\Theta(t) = U(t) - \int_{t_0}^{t}\left[\frac{d}{ds}U(s)\right]_+ ds$. By definition, Θ is an absolutely continuous function on $[t_0, T]$ with negative derivative a.e. in (t_0, T), hence it is non-increasing in $[t_0, T]$. Since this is true for every $T > t_0$, in view of continuity of Θ, we deduce that it is non-increasing in $[t_0, +\infty)$. This together with the fact that the function Θ is bounded from below on $[t_0, +\infty)$, implies that when $t \to +\infty$, $\Theta(t)$ converges to its infimum $\Theta_\infty = \underset{t \geq t_0}{inf}\{\Theta(t)\}$

As a consequence we have that $U(t)$ is also convergent when $t \to +\infty$ with :

$$\lim_{t\to+\infty} U(t) = \lim_{t\to+\infty} \Theta(t) + \int_{t_0}^{+\infty} \left[\dot{U}(t)\right]_+ dt \in \mathbb{R} \tag{4.21}$$

Finally since $b > 3$, by Theorem 4.1 on a :

$$\int_{t_0}^{+\infty} \frac{1}{t}U(t)\, dt = \int_{t_0}^{+\infty} tW(t)\, dt + \frac{1}{2}\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2\, dt < +\infty \tag{4.22}$$

As $\int_{t_0}^{+\infty} \frac{1}{t}\, dt = +\infty$ and $U(t)$ is convergent when $t \to +\infty$, we deduce that $U(t)$ converge to zero, when $t \to +\infty$. This together with the positivity of $t^2W(t)$ and $\frac{t^2}{2}\|\dot{x}(t)\|^2$ allow us to conclude the Theorem. □

4.3 Convergence of the trajectory to a minimizer

Lastly, the $W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H})$ regularity of the solution allows us also to prove the convergence of the trajectory towards a minimizer, using the same argument as in [6]. More precisely we have the following Theorem.

Theorem 4.3. *Let $x \in W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H})$ a given solution to* (DI). *For $b > 3$, we have that the trajectory $\{x(t)\}_{t\geq t_0}$ converges asymptotically to a minimizer x^* of F.*

For the proof of Theorem 4.3 we use the continuous version of Opial's Lemma for which we omit the proof (for more details see [31] or Lemma 4.1 in [7]) :

Lemma 4.2. *Let $S \subset \mathcal{H}$ be a non-empty set and a function $x : [t_0, +\infty)$, such that the following conditions hold:*

1. $\lim_{t\to+\infty} \|x(t) - x^*\| \in \mathbb{R}$ *, for all $x^* \in S$*
2. *Every weak-cluster point of $x(t)$ belongs to S*

Then we have that $x(t)$ converges weakly to a point of S as $t \to +\infty$.

Remark 9. We will invoke the previous Lemma with $S = \arg\min F$. In fact Opial's Lemma holds true for a general separable Hilbert space $\mathcal{H}$, but in our case as $\mathcal{H}$ is finite-dimensional, we also deduce strong convergence of $x(t)$ to a point of S.

Proof. In order to apply Opial's Lemma, we define : $\psi(t) = \frac{1}{2}\|x(t) - x^*\|$. Let $T > t_0$. As $x \in W^{2,\infty}_{loc}((t_0, +\infty); \mathcal{H})$ we have in particular that $\dot{x}$ is differentiable almost everywhere in (t_0, T) with derivative $\ddot{x}$, so that :

$$\dot{\psi}(t) = \langle \dot{x}(t), x(t) - x^* \rangle \quad \text{and} \quad \ddot{\psi}(t) = \|\dot{x}(t)\|^2 + \langle \ddot{x}(t), x(t) - x^* \rangle \quad \text{a.e. in } (t_0, T)$$

By using (DI), for a $z(t) \in \partial F(x(t))$ such that (DI) holds, we obtain :

$$\begin{aligned} \ddot{\psi}(t) + \frac{b}{t}\dot{\psi}(t) &= \|\dot{x}(t)\|^2 + \langle \ddot{x}(t) + \frac{b}{t}\dot{x}(t), x(t) - x^* \rangle \\ &= \|\dot{x}(t)\|^2 - \langle z(t), x(t) - x^* \rangle \\ &\leq \|\dot{x}(t)\|^2 - W(t) \leq \|\dot{x}(t)\|^2 \quad \text{a.e. in } (t_0, T) \end{aligned} \tag{4.23}$$

where in the first inequality we used that $z(t) \in \partial F(x(t))$ and in the second that $W(t) \geq 0$.

Hence by multiplying both sides by t^b we obtain :

$$t^b\ddot{\psi}(t) + bt^{b-1}\dot{\psi}(t) \leq t^b\|\dot{x}(t)\|^2 \quad \text{a.e. in } (t_0, T) \tag{4.24}$$

By integrating over $[t_0, s] \subset (t_0, T)$ we find :

$$\dot{\psi}(s) \leq \frac{t_0^b\dot{\psi}(t_0)}{s^b} + \frac{1}{s^b}\int_{t_0}^{s} t^b\|\dot{x}(t)\|^2 dt \leq \frac{C_0}{s^b} + \frac{1}{s^b}\int_{t_0}^{s} t^b\|\dot{x}(t)\|^2 dt \tag{4.25}$$

where C_0 is a positive constant. If we consider the positive part of $\dot{\psi}$, we have :

$$[\dot{\psi}]_+(s) \leq \frac{C_0}{s^b} + \frac{1}{s^b}\int_{t_0}^{s} t^b\|\dot{x}(t)\|^2 dt \tag{4.26}$$

Hence, by applying Fubini's Theorem, by integrating over $[t_0, T]$, we have that :

$$\begin{aligned} \int_{t_0}^{T} [\dot{\psi}]_+(s)ds &\leq C_0\int_{t_0}^{T} \frac{1}{s^b} + \int_{t_0}^{T} \frac{1}{s^b}\int_{t_0}^{s} t^b\|\dot{x}(t)\|^2 dtds \\ &= (b-1)C_0(t_0^{1-b} - T^{1-b}) + \int_{t_0}^{T} t^b\|\dot{x}(t)\|^2\Big(\int_{t}^{T} s^{-b}ds\Big)dt \\ &\leq C_0 + (b-1)\int_{t_0}^{T} t\|\dot{x}(t)\|^2 dt \end{aligned} \tag{4.27}$$

Since, by Theorem 4.1, for $b > 3$, the right-hand member of this inequality is finite for every $T > t_0$, we deduce that :

$$\int_{t_0}^{+\infty} [\dot{\psi}]_+(s)ds < +\infty \tag{4.28}$$

Hence if we consider the function $\theta(t) = \psi(t) - \int_{t_0}^{t} [\dot{\psi}]_+(s)ds \quad \forall t \in [t_0, +\infty)$, we have that θ is non-increasing and bounded from below on $[t_0, +\infty)$, so it converges to its infimum $\theta_\infty = \inf_{t \geq t_0}\{\theta(t)\}$.

As a consequence we obtain that :

$$\lim_{t\to\infty} \psi(t) = \theta_\infty + \int_{t_0}^{+\infty} [\dot{\psi}]_+(s)ds \in \mathbb{R} \tag{4.29}$$

This shows that the first condition of Opial's Lemma is satisfied.

For the second condition, let $\tilde{x}$ be a weak-cluster point of the trajectory $x(t)$, when $t \to +\infty$. By lower semi-continuity of F, we have that :

$$F(\tilde{x}) \leq \liminf_{t\to\infty} F(x(t)) \tag{4.30}$$

By Theorem 4.1 we have that $\lim_{t\to\infty} F(x(t)) = F(x^*)$, where x^* is a minimizer, so that $\tilde{x} \in \arg\min F$, which shows that the second condition of Opial's Lemma is satisfied, therefore we can conclude the proof by applying Opial's Lemma.

□

5 Conclusion and future directions

In this work we showed that the differential inclusion (DI) admits a solution which is regular enough to ensure that the analysis carried out in [6] can be extended to the case when F is not differentiable. This framework corresponds to the one considered in the discrete setting (i.e when the function g is not differentiable.) for FISTA algorithm (see for example [15] and [21]).

Apart from establishing a uniqueness-result for the solution of (DI), another arising question is if similar results of fast asymptotical convergence of Theorem 4.1 still hold true, if we "allow" the function F having a possible domain $\mathrm{dom}F \subsetneq \mathcal{H}$. This case comes naturally by considering the function g as the indicator function of a closed-convex set. As mentioned before in that case the trajectory exhibits " shocks " which technically correspond to non-boundedness of its second derivative. As a by product we cannot expect that the solution enjoys a $W^{2,+\infty}_{loc}((t_0,+\infty);\mathcal{H})$ regularity. Nevertheless as shown in [5] a solution of the (HBF)-inclusion (1.1), has some convergence properties to a minimizer of F. It would be interesting if this analysis could be extended, in order to obtain the fast convergence properties of a solution of (DI)

Another challenging question is the tuning of the parameter b. In particular in Corollary 4.3 we showed that for $b \geq 3$, the solution of (DI) admit some further global regularity ($W^{1,+\infty}((t_0,+\infty);\mathcal{H})$) other that the local one ($W^{2,+\infty}_{loc}((t_0,+\infty);\mathcal{H})$). It would be interesting to know if such global regularity also holds for $1 < b \leq 3$.

Finally it would be interesting to extend these results for a perturbed version of (DI), by a perturbation term $p(t)$ to which we have to impose some integrability properties. In the same (perturbed) framework, it is also interesting to study a slightly different differential inclusion by replacing the viscosity term $\frac{b}{t}$ by a function which behaves asymptotically as $\frac{1}{t^d}$ where $d \in (0,1)$. This would be the continuous counter-part of the inertial forward-backward algorithm introduced in [12]. The interest of this case is the investigation of a better trade-off between the speed of the convergence and the integrability of the perturbed term $p(t)$ (for more information on this issue we address the reader to a recent work of Balti and al [13])

A Appendix

A.1. *Proof of Lemma 2.1*

Proof. (see proof of Lemma 2.2 in [15])

In order to prove this lemma, we define the quadratic approximation of F in a point $y \in \mathcal{H}$ with step $0 < \gamma \leq \frac{1}{L}$ such that for all $x \in \mathcal{H}$:

$$S_{\gamma,y}(x) = f(y) + \langle x - y, \nabla f(y)\rangle + \frac{1}{2\gamma}\|x-y\|^2 + g(x) \tag{A.1}$$

We can observe that $S_{\gamma,y}$ is a function majorating F and is equal to F in point $y \in \mathcal{H}$

In particular for any $y \in \mathcal{H}$, we have :

$$F(T(y)) \leq S_{\gamma,y}(T(y)) \tag{A.2}$$

so that for any $x \in \mathcal{H}$ it follows that :

$$\begin{aligned}
F(x) - F(T(y)) &\geq F(x) - S_{\gamma,y}(T(y)) \\
&= f(x) - f(y) - \langle T(y) - y, \nabla f(y)\rangle - \frac{1}{2\gamma}\|T(y)-y\|^2 + g(x) - g(T(y)) \\
&\geq \langle x - T(y), \nabla f(y) + z\rangle - \frac{1}{2\gamma}\|T(y)-y\|^2 \quad (\text{ where } z \in \partial g(T(y))) \\
&= \frac{1}{\gamma}\langle x - T(y), y - T(y)\rangle - \frac{1}{2\gamma}\|T(y)-y\|^2 \\
&= \frac{1}{2\gamma}(-\|T(y)-y\|^2 + 2\langle x - y, T(y) - y\rangle) \\
&= \frac{1}{2\gamma}(\|T(y)-x\|^2 - \|y-x\|^2)
\end{aligned}$$

where in the second inequality we used the convexity of f and the definition of $\partial g(T(y)) \neq \{\emptyset\}$ (because g is proper and convex.)
For the second equality we used the following characterization :

$$T(y) = Prox_{\gamma g}(y - \gamma \nabla f(y)) \quad \text{iff} \quad y - \gamma \nabla f(y) - T(y) = \gamma z \quad \text{with } z \in \partial g(T(y))$$
$$\text{iff} \quad z = \frac{1}{\gamma}(y - T(y)) - \nabla f(y)$$

□

A.2. *Proof of Proposition 2.2*

Proof. (see proof of Corollary 2 in [21])
By Corollary 2.1 we have that for all $n \geq 1$:

$$a_n \delta_n \leq a_1 \delta_1 + \gamma(w_1 - w_n) \leq \gamma w_1 \tag{A.3}$$

since $a_1 = 0$. Hence from (A.3) we deduce that there exists a constant C_1 which do not depend in γ such that for all $n \in \mathbb{N}$:

$$\delta_n \leq C_1 \gamma \tag{A.4}$$

which concludes the first part of this Proposition with $C_1 = w_1 + w_0 > 0$.

For the second part, if we apply the Lemma 2.1 to the points $y = y_n$ and $x = (1 - \frac{1}{t_{n+1}})x_n + \frac{1}{t_{n+1}}x^*$, with $\gamma \in (0, 1/L]$ so that we have :

$$N^2 w_{N+1} + \sum_{n=1}^{N} \big[(a-2)(n+a) + 1\big] w_n \leq \frac{(v_1 - v_N)}{\gamma} \tag{A.5}$$

From this inequality as $a > 2$ it follows that $v_n \leq v_1$, for all $n \geq 1$. On the other hand, by applying 2.1 to $x = x_n$ and $y = y_n$, as in Corollary 2.1 we have that :

$$\delta_{n+1} - \left(\frac{n-1}{n+a}\right)^2 \delta_n \leq \gamma(w_n - w_{n+1}) \tag{A.6}$$

By multiplying by $(n+a)^2$ and summing on $n \in \{1, \dots, N\}$ we find :

$$(N+a)^2 \delta_{N+1} + \sum_{n=1}^{N} \big((n+a-1)^2 - (n-1)^2\big)\delta_n \leq \gamma\left((a+1)w_1 + \sum_{n=2}^{N} \big((n+a)^2 - (n+a-1)^2\big) w_n\right) \tag{A.7}$$

By (A.5) and as $a > 2$, it follows from (A.7) that :

$$N^2 \delta_{N+1} + \sum_{n=1}^{N} n\delta_n \leq \gamma\big((a+1)w_1 + \sum_{n=2}^{N}(2n + 2a - 1)w_n\big) \leq \gamma(c_1 + \frac{c_2}{\gamma}) = \gamma c_1 + c_2 \leq C < +\infty \tag{A.8}$$

where c_1, c_2 and C are positive constants which do not depend on the step γ.

Finally by relation (A.8) and (A.5), the sequences $n(x_{n+1} - x_n)$ and v_n are bounded, so, by definition of $\{u_n\}_{n \in \mathbb{N}}$, it follows that $\{u_n\}_{n \in \mathbb{N}}$ is also bounded, hence by definition of $\{u_n\}_{n \in \mathbb{N}}$, we deduce that $\{x_n\}_{n \in \mathbb{N}}$ is also bounded by a constant (noted as) $C_2 > 0$, which is independent of the step γ, given that $\gamma < \frac{1}{L}$.

□

A.3. *Subdifferential properties*

The following Lemma shows that the subdifferential of a convex function defined in $\mathbb{R}^d$, preserves the boundedness of sets.

Lemma A.1. *(see Proposition 4.14 in [22])*
Let $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ be convex function and let K be a bounded set in $\mathbb{R}^d$. Then the set :

$$A = \bigcup_{x \in K} \partial g(x)$$

is bounded.

Proof. By contradiction we assume that there exists a subsequence in A noted as $\{z_n\}_{n\in\mathbb{N}}$ such that $z_n \in \partial f(x_n)$ for all $n \in \mathbb{N}$ and $z_n \longrightarrow +\infty$, where $\{x_n\}_{n\in\mathbb{N}}$ is bounded ($x_n \in K$ for all $n \in \mathbb{N}$).

From boundedness of $\{x_n\}_{n\in\mathbb{N}}$ we deduce that up to a subsequence still noted as $\{x_n\}_{n\in\mathbb{N}}$ we have that $x_n \longrightarrow x \in \overline{K}$. For all $n \in \mathbb{N}$ we define the sequence $\{e_n\}_{n\in\mathbb{N}}$ as

$$e_n = \begin{cases} \frac{z_n}{\|z_n\|} & \text{if} \quad z_n \neq 0 \\ 1 & \text{otherwise} \end{cases}$$

It is clear that $\|e_n\| \leq 1$, hence there exists a subsequence noted again as $\{e_n\}_{n\in\mathbb{N}}$ such that $e_n \longrightarrow e \in \mathbb{R}$.

From the definition of subdifferential, as $z_n \in \partial f(x_n)$, we have that :

$$g(x_n + e_n) - g(x_n) \geq \langle z_n, e_n \rangle = \|z_n\| \quad \forall n \in \mathbb{N} \tag{A.9}$$

By taking the limit to $n \to +\infty$ from continuity of g (since it is convex on an open set in a finite dimensional space) we obtain that the Left-Hand-Side of the previous inequality converges to $g(x+e) - g(x)$ which is finite. On the other side by hypothesis we have that $\|z_n\|$ diverges to infinity, which leads to a contradiction.

□

Acknowledgements : This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR GOTMI) (ANR-16-CE33-0010-01). Jean-François Aujol is a member of Institut Universitaire de France (IUF).

References

[1] Robert A Adams and John JF Fournier. *Sobolev spaces*, volume 140. Academic press, 2003.

[2] Samir Adly, Hedy Attouch, and Alexandre Cabot. Finite time stabilization of nonlinear oscillators subject to dry friction. In *Nonsmooth mechanics and analysis*, pages 289–304. Springer, 2006.

[3] Felipe Alvarez. On the minimizing property of a second order dissipative system in hilbert spaces. *SIAM Journal on Control and Optimization*, 38(4):1102–1119, 2000.

[4] Felipe Alvarez, Hedy Attouch, Jérôme Bolte, and P Redont. A second-order gradient-like dissipative dynamical system with hessian-driven damping.: Application to optimization and mechanics. *Journal de mathématiques pures et appliquées*, 81(8):747–779, 2002.

[5] Hedy Attouch, Alexandre Cabot, and Patrick Redont. The dynamics of elastic shocks via epigraphical regularization of a differential inclusion. barrier and penalty approximations. *Advances in Mathematical Sciences and Applications*, 12(1):273–306, 2002.

[6] Hedy Attouch, Zaki Chbani, Juan Peypouquet, and Patrick Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, pages 1–53, 2016.

[7] Hedy Attouch, Xavier Goudou, and Patrick Redont. The heavy ball with friction method, i. the continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Communications in Contemporary Mathematics*, 2(01):1–34, 2000.

[8] Hedy Attouch, Paul-Emile Maingé, and Patrick Redont. A second-order differential system with hessian-driven damping; application to non-elastic shock laws. *Differential Equations & Applications*, 4(1):27–65, 2012.

[9] Hedy Attouch and Juan Peypouquet. The rate of convergence of nesterov's accelerated forward-backward method is actually faster than 1/kˆ2. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016.

[10] Hedy Attouch, Juan Peypouquet, and Patrick Redont. Fast convex optimization via inertial dynamics with hessian driven damping. *Journal of Differential Equations*, 261(10):5734–5783, 2016.

[11] Gilles Aubert and Jean-Francois Aujol. A variational approach to removing multiplicative noise. *SIAM Journal on Applied Mathematics*, 68(4):925–946, 2008.

[12] J-F Aujol and Ch Dossal. Stability of over-relaxations for the forward-backward algorithm, application to fista. *SIAM Journal on Optimization*, 25(4):2408–2433, 2015.

[13] Mounir Balti and Ramzi May. Asymptotic for the perturbed heavy ball system with vanishing damping term. *arXiv preprint arXiv:1609.00135*, 2016.

[14] Heinz H Bauschke and Patrick L Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer Science & Business Media, 2011.

[15] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[16] Abdelmounim Belahmidi and Antonin Chambolle. Time-delay regularization of anisotropic diffusion and image processing. *ESAIM: Mathematical Modelling and Numerical Analysis*, 39(2):231–251, 2005.

[17] Haim Brezis. *Opeérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, volume 5. Elsevier, 1973.

[18] Alexandre Cabot, Hans Engler, and Sébastien Gadat. On the long time behavior of second order differential equations with asymptotically small dissipation. *Transactions of the American Mathematical Society*, 361(11):5983–6017, 2009.

[19] Alexandre Cabot, Hans Engler, Sébastien Gadat, et al. Second-order differential equations with asymptotically small dissipation and piecewise flat potentials. *Electronic Journal of Differential Equation*, 17:33–38, 2009.

[20] Alexandre Cabot and Laetitia Paoli. Asymptotics for some vibro-impact problems with a linear dissipation term. *Journal de mathématiques pures et appliquées*, 87(3):291–323, 2007.

[21] A Chambolle and Ch Dossal. On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *Journal of Optimization Theory and Applications*, 166(3):968–982, 2015.

[22] Francis Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264. Springer Science & Business Media, 2013.

[23] Lawrence Craig Evans and Ronald F Gariepy. *Measure theory and fine properties of functions*. CRC press, 2015.

[24] N Gigli, Luigi Ambrosio, Giuseppe Savare, et al. Gradient flows: in metric spaces and in the space of probability measures. *LECTURE NOTES IN MATHEMATICS*, pages 1–333, 2005.

[25] Osman Güler. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization*, 2(4):649–664, 1992.

[26] Juha Heinonen. *Lectures on Lipschitz analysis*. University of Jyväskylä, 2005.

[27] Mohamed Ali Jendoubi and Ramzi May. Asymptotics for a second-order differential equation with nonautonomous damping and an integrable source term. *Applicable Analysis*, 94(2):435–443, 2015.

[28] Ramzi May. Asymptotic for a second order evolution equation with convex potential and vanishing damping term. *arXiv preprint arXiv:1509.05598*, 2015.

[29] Yurii Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983.

[30] Yurii Nesterov. Introductory lectures on convex optimization: A basic course, 2013.

[31] Zdzisław Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bulletin of the American Mathematical Society*, 73(4):591–597, 1967.

[32] Weijie Su, Stephen Boyd, and Emmanuel J Candes. A differential equation for modeling nesterov's accelerated gradient method: theory and insights. *Journal of Machine Learning Research*, 17(153):1–43, 2016.