The Differential Inclusion Modeling FISTA Algorithm and Optimality of Convergence Rate in the Case b

Vassilis Apidopoulos, Jean-François Aujol, Charles H 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

September 14, 2017

Contents

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 $\mathbb{R}^d$, where F is a proper, convex, lower semi-continuous function. The motivation of this study is that the differential inclusion models the FISTA algorithm as considered in [18]. In particular we investigate the different asymptotic properties of solutions for this inclusion for $b > 0$. We show that the convergence rate of $F(x(t))$ towards the minimum of F is of order of $O\left(t^{-\frac{2b}{3}}\right)$ when $0 < b < 3$, while for $b > 3$ this order is of $o(t^{-2})$ and 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], [7], [11] and [31]. In addition we show that the order of the

convergence rate $O\left(t^{-\frac{2b}{3}}\right)$ of $F(x(t))$ towards the minimum is optimal, in the case of low friction $b < 3$, by making a particular choice of F.

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 $x(t_0) = x_0 \in \mathbb{R}^d$ 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. $b > 0$

H 4. $F : \mathbb{R}^d \longrightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ a lower semi-continuous, convex and coercive function

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). Its basic tool is the proximal operator, which we recall in Section 2. It was introduced by Beck and Teboulle in [13], based on ideas of Nesterov in [26] (see also [27]) and Güler [22].

In the seminal works of Alvarez [3] and Attouch and al [8], 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) with an order of convergence rate $O\left(t^{-1}\right)$, 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 asymptotical vanishing viscosity term $\gamma(t) \geq 0$ verifying some integrability conditions (see for example [15], [16] and [24])

By extending the analysis for the semi-differential case (when F is not necessarily differentiable), in [5] and [17] 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 $\mathrm{dom}F$ is possibly different from the whole space $\mathbb{R}^d$ (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 [17]), 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 true.

In [2] the authors study a differential inclusion in the same setting as the one that we also treat in this work (where $\mathcal{D}(F) = \mathbb{R}^d$), 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 [31], [6], [25], [7] and [11] the authors study, in a possible infinite dimensional Hilbert space $\mathbb{R}^d$, the differential equation modeling Nesterov's accelerated algorithm (see [26]) :

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

where $b > 0$ and F is a continuously differentiable and convex function.

The importance of studying this particular equation compared to (HBF) is the "acceleration effect" due to the viscosity term $\frac{b}{t}$. In particular, in these works, it is shown that under some additional hypothesis $b \geq 3$, the solution-trajectory of (1.2) enjoys fast convergence minimization properties over F of inverse quadratic order $O(t^{-2})$. Furthermore in [6] the authors establish the weak convergence of the trajectory to a minimizer of F. Further investigation of this ODE concerning the convergence rate of F to its minimum, has been made recently in the case when $b < 3$, in [7] and [11].

Let us recall here that the FISTA algorithm (version considered in [18]), consists in the following scheme. Let $x_0 = y_0 \in \mathbb{R}^d$, $b > 0$ and the step $\gamma > 0$. For all $n \geq 1$, define :

$$\begin{aligned} x_{n+1} &= \text{Prox}_{\gamma g}\big(y_n - \gamma \nabla f(y_n)\big) \\ y_n &= x_n + \frac{n}{n+b}(x_n - x_{n-1}) \end{aligned} \tag{FISTA}$$

where f is a convex differentiable function with Lipschitz derivative and g is a proper, lower semi-continuous convex function (for a definition of the proximal operator Prox see (2.1) in Section 2).

In the case when the function g is also differentiable, FISTA is equivalent to Nesterov's accelerated algorithm, i.e. for all $n \geq 1$:

$$\begin{aligned} x_{n+1} &= y_n - \gamma \nabla\big(f(y_n) + g(y_n)\big) = y_n - \gamma \nabla F(y_n) \\ y_n &= x_n + \frac{n}{n+b}(x_n - x_{n-1}) \end{aligned} \tag{NS}$$

For Nesterov's accelerated algorithm (NS) and FISTA algorithm , it is proven that the objective function $F(x_n) - \min F$ enjoys an inverse quadratic order of convergence rate $O(n^{-2})$ asymptotically when $b \geq 3$, as also this order becomes a $o(n^{-2})$ and the sequence $\{x_n\}_{n \geq 1}$ converge to a minimizer of F, when $b > 3$ (see [26],[27], [13], [18], [6] and [9]). For $b < 3$ it is recently proven (see [7] and [4]) that the order of the convergence rate for $F(x_n) - \min F$ is $O\left(n^{-\frac{2b}{3}}\right)$ asymptotically.

As mentioned before, a time discretization of the ODE (1.2) with time step $\sqrt{\gamma}$ and $F = f + g$ corresponds to (NS) algorithm (see for example [31] and [6]) and in the same way the same discretization of the differential inclusion (DI) corresponds to FISTA algorithm.

Motivated by these works, in this paper we study the differential inclusion (DI) where F is a proper, convex and lower semi-continuous function (not necessarily differentiable) with domain $\mathcal{D}(F)$ which is the same setting as the one considered for the study of the FISTA algorithm.

In fact, concerning the existence of a solution of (DI), the inclusion (DI) falls into the generalized one, studied in [29] (see also [30]) which is the following :

$$\ddot{x}(t) + \partial F(x(t)) \ni h(t, x(t), \dot{x}(t)) \tag{1.3}$$

where $h : \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is a continuous function and Lipschitz continuous in its last two arguments uniformly with respect to the first one and F is a proper, convex and lower semi-continuous function.

The contributions of this paper are the followings. We extend the study in the differential setting [31], [6], [25], [7] and [11] for the ODE (1.2), to the non-differential setting, which is the same as the one considered for the study of FISTA algorithm (see for example [18], [4], [10]). In particular, for a shock solution x of (DI) (see Definition 3.1) we obtain "almost" the same fast asymptotic properties as the ones obtained in the differential setting when $b \geq 3$, i.e. :

$$(F(x(t)) - \min F) \sim O(t^{-2}) \text{ and } \|\dot{x}(t)\| \sim O(t^{-1}) \text{ almost everywhere when } t \to +\infty$$

and the convergence of the trajectory $\{x(t)\}_{t \geq t_0}$ to a minimizer of F, as also the improvement of the previous orders to $o(t^{-2})$ and $o(t^{-1})$ respectively, when $b > 3$. In the same spirit, we show that for $0 < b < 3$, almost the same asymptotic properties hold true as the ones settled in the differential setting in [7] and [11], i.e.

$$(F(x(t)) - \min F) \sim O\left(t^{-\frac{2b}{3}}\right) \text{ and } \|\dot{x}(t)\| \sim O\left(t^{-\frac{b}{3}}\right) \text{ a.e. when } t \to +\infty$$

3

In the case when the domain of F is the whole space $\mathbb{R}^d$, we show that the regularity of a solution x of (DI) is sufficient to obtain exactly the same results concerning the asymptotical behavior of this solution, to the ones obtained for the solution of the ODE (1.2) in the differential setting.

Finally, we show that in the particular case when F is the absolute value function and $b < 3$, the convergence rate $O\left(t^{-\frac{2b}{3}}\right)$ of $F(x(t))$ to the minimum can not be improved (it is achieved), therefore this order is optimal. Here we must stress out that the example of the absolute value, is only valid in the non-differential setting (since the absolute value is not everywhere differentiable).

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 $a = (a_1, a_2) \in \mathbb{R}^d \times \mathbb{R}^d$. 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 [14]).

The organization of this paper is the following. In Section 2, we introduce some standard notions that we use in our analysis. In Section 3, we present some results, concerning the existence of a solution of (DI). In Section 4 we present the results concerning the asymptotic properties of a solution of (DI) in the case when $b \geq 3$ and when $0 < b < 3$. Finally in Section 5 we show the optimality of the order of convergence rate of $F(x(t))$ to $\min F$ in the case when $b < 3$ by making a special choice for F.

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 $W^{m,p}(I; \mathbb{R}^d)$ the classical Sobolev space with values on $\mathbb{R}^d$, i.e. the space of functions in $L^p(I)$ whose distributional derivatives up to order m belong to $L^p(I)$ and with $BV(I; \mathbb{R}^d)$ the space of functions with bounded variation. We also denote with $\mathcal{C}^{m,1}(I; \mathbb{R}^d)$ the class of continuously differentiable functions up to order m with Lipschitz m-th derivative and $\mathcal{M}(I; \mathbb{R}^d)$ the class of all Radon measures with values in $\mathbb{R}^d$. For a detailed presentation of some properties of these spaces, we address the reader to [1] [14], [20] and [21].

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

$$\partial G(x) = \{z \in \mathbb{R}^d : \forall y \in \mathbb{R}^d, G(x) \leq G(y) + \langle z, x - y \rangle\}$$

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 :

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

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

Equivalently the proximal operator can be also seen as the resolvent of the maximal monotone operator ∂G, i.e. for all $x \in \mathbb{R}^d$ 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 [12].

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.

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

3 Existence of a solution of (DI)

3.1 Shock solutions

In this section we will present the results concerning the existence of a solution of (DI). As mentioned before for most of these results we address the reader to [5] and [29].

Let us recall the system (DI) with some initial condition

$$\begin{cases} \ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \partial F(x(t)) \ni 0 \\ x(t_0) = x_0 \in \mathcal{D}(F) \quad \text{and} \quad \dot{x}(t_0) = v_0 \in T_{\overline{\mathcal{D}(F)}}(x_0) \end{cases} \tag{DI}$$

where T_K denotes the tangent cone of a closed convex set K ,i.e. for all $x \in K$.

$$T_K(x) = \overline{\{\frac{u-x}{s} : s > 0, u \in K\}}$$

As already mentioned the system (DI) falls into the one studied in [29] : $\ddot{x} + \partial F(x) \ni h(t, x, \dot{x})$ for a general continuous function h from $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}^d$ and Lipschitz in its last two arguments with respect to the first one.

Here we recall the basic results concerning the existence of a solution fo (DI). For a detailed presentation and proofs of these results, we address the reader in [29].

Definition 3.1. Let $I = [t_0, +\infty)$. A function $x : I \longrightarrow \mathbb{R}^d$ is an energy-conserving shock solution of (DI) if the following conditions hold :

1. $x \in \mathcal{C}^{0,1}([t_0, T]; \mathbb{R}^d)$, for all $T > t_0$ i.e. x is a Lipschitz continuous function

2. $\dot{x} \in BV([t_0, T]; \mathbb{R}^d)$, for all $T > t_0$

3. $x(t) \in \mathcal{D}(F)$, for all $t \in I$,

4. For all $\phi \in \mathcal{C}^1_c(I, \mathbb{R}^+)$ and $v \in \mathcal{C}(I, \mathcal{D}(F))$, it holds :

$$\int_{t_0}^{T} (F(x(t)) - F(v(t))\phi(t)dt \leq \langle \ddot{x} + \frac{b}{t}\dot{x}, (v-x)\phi \rangle_{\mathcal{M} \times \mathcal{C}} \tag{3.1}$$

In fact we have that (DI) holds almost everywhere in I

5. x satisfies the following energy equation

$$F(x(t)) - F(x_0) + \frac{1}{2}\|\dot{x}(t)\|^2 - \frac{1}{2}\|v_0\|^2 + \int_{t_0}^{t} \frac{b}{s}\|\dot{x}(s)\|^2 ds = 0 \tag{3.2}$$

almost everywhere in I

We have the following existence result (see Theorem 3.1 in [29]).

Theorem 3.1. *Under the hypotheses **H1.**, **H2.**, **H.3.** and **H.4.** made on F, the inclusion (DI) admits a solution x in the sense of Definition 3.1 In fact we have that (DI) holds a.e. in I*

Following [29] we consider the Moreau-Yosida approximation of ∂F, which we denote as ∇F_γ and for all $\gamma > 0$ we consider the following approximating ODE :

$$\begin{cases} \ddot{x}_\gamma(t) + \frac{b}{t}\dot{x}_\gamma(t) + \nabla F_\gamma(x_\gamma(t)) \ni 0 \\ x_\gamma(t_0) = x_0 \in \mathcal{D}(F) \quad \text{and} \quad \dot{x}_\gamma(t_0) = v_0 \end{cases} \tag{ADE}$$

We will give a sketch of the proof of Theorem 3.1 since we will use some of its elements in the following section. For a full proof we address the reader to [29].

The schema of the proof is classic. Find some a-priori estimates for the family of solutions $\{x_\gamma\}_{\gamma > 0}$ of (ADE) and its derivatives $\{\dot{x}_\gamma\}_{\gamma > 0}$, $\{\ddot{x}_\gamma\}_{\gamma > 0}$ and then conclude by extracting a subsequence which converge to a solution of (DI) in some suitable space.

In particular we have the following Theorem (see proof of Theorem 3.1 in [29]) :

Theorem 3.2. *Let* $\{F_\gamma\}_{\gamma>0}$ *a family of functions such that* ∇F_γ *is the Moreau - Yosida approximation of* ∂F *for all* $\gamma > 0$. *There exists a subsequence* $\{x_\gamma\}_{\gamma>0}$ *of solutions of* (ADE) *that converge to a shock solution* x *of* (DI) *in the following sense :*

A.1 $x_\gamma \xrightarrow[\gamma\to 0]{} x$ *uniformly on* $[t_0, T]$ *for all* $T > t_0$

A.2 $\dot{x}_\gamma \xrightarrow[\gamma\to 0]{} \dot{x}$ *in* $L^p([t_0, T]; \mathbb{R}^d)$, *for all* $p \in [1, +\infty)$, *for all* $T > t_0$

A.3 $F_\gamma(x_\gamma) \xrightarrow[\gamma\to 0]{} F(x)$ *in* $L^p([t_0, T]; \mathbb{R}^d)$, *for all* $p \in [1, +\infty)$, *for all* $T > t_0$

$\left.\rule{0pt}{5em}\right\}$ **AS**

In order to prove Theorem 3.1 and 3.2 we make use of the following a-priori estimates for the approximations $\{x_\gamma\}_{\gamma>0}$. In particular we have the following :

Lemma 3.1. *Let* $\{x_\gamma\}_{\gamma>0}$ *be a family of solutions of* (ADE) *for any* $\gamma > 0$. *Then :*

$$\sup_{\gamma>0}\{\|x_\gamma\|_\infty, \|\dot{x}_\gamma\|_\infty\} < +\infty \tag{3.3}$$

Lemma 3.2. *Let* $\{x_\gamma\}_{\gamma>0}$ *be a family of solutions of* (ADE) *for any* $\gamma > 0$. *Then :*

$$\sup_{\gamma>0}\{\|\nabla F_\gamma(x_\gamma)\|_1, \{\|\ddot{x}_\gamma\|_1\} < +\infty \tag{3.4}$$

From Lemmas 3.1 and 3.2, one can extract a subsequence -still- denoted as $\{x_\gamma\}_{\gamma>0}$ which converges according to the approximate scheme (**AS**) to a solution of (DI) in the sense of Definition 3.1.

3.2 The case of $\mathcal{D}(F) = \mathbb{R}^d$

In the case when $\mathcal{D}(F) = \mathbb{R}^d$, one can expect more regularity over the solution x of (DI). In particular we have the following corollary.

Corollary 3.1. *Under the hypotheses* **H1.**, **H2.**, **H.3**. **H.4**., *if we suppose additionally that* $\mathcal{D}(F) = \mathbb{R}^d$, *then the differential inclusion* (DI) *admits a solution* x *in the sense of Definition 3.1, such that :* $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$, *for all* $T > t_0$, *i.e.* x *is defined everywhere in* $[t_0, +\infty)$ *and differentiable with locally Lipschitz gradient.*

Remark 3. Notice that when $\mathcal{D}(F) = \mathbb{R}^d$, the function is continuous (as it is convex in $\mathbb{R}^d$), hence the lower semi-continuity property of hypothesis **H.4** is automatically satisfied.

In order to achieve this supplementary regularity for the solution x, we use the following Lemma :

Lemma 3.3. *Let* $\{x_\gamma\}_{\gamma>0}$ *be a family of solutions of* (ADE) *for* $\gamma > 0$. *Then :*

$$\sup_{\gamma>0}\{\|\dddot{x}_\gamma\|_\infty\} < +\infty \tag{3.5}$$

Proof. In fact by Lemma 3.1 we have that $\{\|x_\gamma\|_\infty\}_{\gamma>0}$ and $\{\|\dot{x}_\gamma\|_\infty\}_{\gamma>0}$ are uniformly bounded with respect to γ. By using Lemma A.2 we deduce that the family $\{\|\nabla F_\gamma(x_\gamma)\|_\infty\}_{\gamma>0}$ is also uniformly bounded with respect to γ sequences. Finally by invoking equation (ADE), we obtain that $\{\ddot{x}_\gamma\}_{\gamma>0}$ is uniformly bounded with respect to γ. $\qquad\square$

Proof. By estimations (3.3) and (3.5) we deduce that $\{x_\gamma\}_{\gamma>0}$ is bounded in $W^{2,\infty}(t_0, T); \mathbb{R}^d)$. By using the fact that $W^{2,\infty}((t_0, T); \mathbb{R}^d) \subset \mathcal{C}^{1,1}([t_0, T]; \mathbb{R}^d) \Subset \mathcal{C}^1([t_0, T]; \mathbb{R}^d)$ (see Theorem 4.5 in [20] and Theorem 1.34 in [1]), we deduce the existence of a subsequence (still denoted as) $\{x_\gamma\}_{\gamma>0}$ that converges to a function x in $\mathcal{C}^1([t_0, T]; \mathbb{R}^d)$.

Furthermore, as $\ddot{x}_\gamma$ is bounded in $L^\infty((t_0, T); \mathbb{R}^d)$ and $L^\infty((t_0, T); \mathbb{R}^d)$ can be identified with the dual space of $L^1((t_0, T); \mathbb{R}^d)$, 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 $\{\ddot{x}_\gamma\}_{\gamma>0}$, :

$$\ddot{x}_\gamma \rightharpoonup^* u \quad \text{in } L^\infty((t_0, T); \mathbb{R}^d) \tag{3.6}$$

where by uniqueness of the limit (in the distributional sense) we have that $\ddot{x} \equiv u \in L^\infty((t_0, T); \mathbb{R}^d)$. Hence we have that $x \in \mathscr{C}^1([t_0, T]; \mathbb{R}^d) \cap W^{2,\infty}((t_0, T); \mathbb{R}^d)$.

In fact for all $i \in \mathbb{N}^*$, one can construct the sequences (of sequences) of functions $\{\{x^i_{h(\gamma)}\}_{\gamma>0}\}_{i\in\mathbb{N}}$ as follows :

$$
\begin{aligned}
\hat{x}^1_{h(\gamma)} &\xrightarrow[\gamma\to 0]{} \hat{x}^1 &&\in W^{2,\infty}([t_0, t_0+1]) \\
\hat{x}^2_{h(\gamma)} &\xrightarrow[\gamma\to 0]{} \hat{x}^2 &&\in W^{2,\infty}([t_0, t_0+2]) \\
&\vdots \\
\hat{x}^i_{h(\gamma)} &\xrightarrow[\gamma\to 0]{} \hat{x}^i &&\in W^{2,\infty}([t_0, t_0+i])
\end{aligned}
\tag{3.7}
$$

in a way that every time we extract a subsequence $\{x^{i+1}_{h(\gamma)}\}_{\gamma>0}$ from the subsequence considered before $\{x^i_{h(\gamma)}\}_{\gamma>0}$, for every $i \in \mathbb{N}^*$. By diagonal extraction we consider the sequence of functions $\{x^i_{h(1/i)}\}_{i\in\mathbb{N}}$. We then define the sequence of functions $\{w_i\}_{i\in\mathbb{N}}$ in $[t_0,+\infty)$, as the $W^{2,\infty}([t_0+i,+\infty))$ extensions of $x^i_{h(1/i)}$, for all $i \in \mathbb{N}$. By this construction there exists a function $x : [t_0, +\infty) \longrightarrow \mathbb{R}^d$ such that the sequence of functions $\{w_i\}_{i\in\mathbb{N}}$ converges to, with respect to the $W^{2,\infty}_{loc}([t_0,+\infty))$ norm. This shows that $x \in W^{2,\infty}((t_0,T);\mathbb{R}^d) \cap \mathcal{C}^1([t_0,+\infty);\mathbb{R}^d)$, for all $T > t_0$. $\qquad\square$

4 Asymptotic behavior of the trajectory

In this section we are interested in the asymptotic properties of a solution of (DI). As the regularity of such a solution depends on the domain of F, we will split the presentation into two parts, one which treats the case of a shock solution and the other concerning the case when $\mathcal{D}(F) = \mathbb{R}^d$. In what follows we denote x^* a minimizer of F and $W(t) = F(x(t)) - F(x^*)$.

4.1 Energy estimates for shock solutions

For $\lambda \geq 0$ and $\xi \geq 0$ 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 [31] 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 shock solution of (DI). The difficulty comes from the fact that the solution is not everywhere differentiable, hence we can not differentiate directly $\mathcal{E}_{\lambda,\xi}$. Nevertheless by an approximating scheme we deduce the same bound estimates for $W(t)$ and $\|\dot{x}(t)\|$ as in [6] hold for almost every $t \geq t_0$.

For the asymptotic properties of a shock solution we will systematically make use of its approximative scheme in the spirit of the study made in [5]. Let $\{x_\gamma\}_{\gamma>0}$ a suitable subsequence of solutions of (ADE) such that the approximating scheme ($\mathcal{AS}$) holds, i.e. :

A.1 $x_\gamma \xrightarrow[\gamma\to 0]{} x$ uniformly on [0,T] for all $T > 0$

A.2 $\dot{x}_\gamma \xrightarrow[\gamma\to 0]{} \dot{x}$ in $L^p([t_0,T];\mathbb{R}^d)$, for all $p \in [1,+\infty)$, for all $T > t_0$

A.3 $F_\gamma(x_\gamma) \xrightarrow[\gamma\to 0]{} F(x)$ in $L^p([t_0,T];\mathbb{R}^d)$, for all $p \in [1,+\infty)$, for all $T > t_0$

We will also use the following notation $W_\gamma(t) = F_\gamma(x_\gamma(t)) - F_\gamma(x^*)$, for all $\gamma > 0$.

$$
\mathcal{E}^\gamma_{\lambda,\xi}(t) = t^2 W_\gamma(t) + \frac{1}{2}\|\lambda(x_\gamma(t) - x^*) + t\dot{x}_\gamma(t)\|^2 + \frac{\xi}{2}\|x_\gamma(t) - z\|^2
\tag{4.2}
$$

By differentiating we find :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) = 2tW_{\gamma}(t) + t^2\langle\nabla F_{\gamma}(x_{\gamma}(t)), \dot{x}_{\gamma}(t)\rangle + \xi\langle\dot{x}_{\gamma}(t), x_{\gamma}(t) - x^*\rangle$$
$$+ \langle(\lambda+1)\dot{x}_{\gamma}(t) + t\ddot{x}_{\gamma}(t), \lambda(x_{\gamma}(t) - x^*) + t\dot{x}_{\gamma}(t)\rangle$$
$$(\text{ADE}) = 2tW_{\gamma}(t) + t^2\langle\nabla F_{\gamma}(x_{\gamma}(t)), \dot{x}_{\gamma}(t)\rangle + \xi\langle\dot{x}_{\gamma}(t), x_{\gamma}(t) - x^*\rangle \tag{4.3}$$
$$+ \langle(\lambda+1-b)\dot{x}_{\gamma}(t) - t\nabla F_{\gamma}(x(t)), \lambda(x_{\gamma}(t) - x^*) + t\dot{x}_{\gamma}(t)\rangle$$
$$= 2tW_{\gamma}(t) - t\langle\nabla F_{\gamma}(x_{\gamma}(t)), x_{\gamma}(t) - x^*\rangle + (\lambda+1-b)t\|\dot{x}_{\gamma}(t)\|^2$$
$$+ \big(\xi + \lambda(\lambda+1-b)\big)\langle\dot{x}_{\gamma}(t), x_{\gamma}(t) - x^*\rangle$$

By choosing $\xi = \lambda(b-\lambda-1)$ and the convexity of F_{γ}, we find :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) = 2tW_{\gamma}(t) - \langle\nabla F_{\gamma}(x_{\gamma}(t)), x_{\gamma}(t) - x^*\rangle + (\lambda+1-b)t\|\dot{x}_{\gamma}(t)\|^2$$
$$\leq (2-\lambda)tW_{\gamma}(t) + (\lambda+1-b)t\|\dot{x}_{\gamma}(t)\|^2 \tag{4.4}$$

4.1.1 The case of high friction $b \geq 3$

In this paragraph we study the case where the friction parameter b in (DI) is high i.e. we consider $b \geq 3$. We have the following Lemma :

Lemma 4.1. *Let x be a shock solution of* (DI) *and $b \geq 3$. For $\xi = \lambda(b-\lambda-1)$ and $2 \leq \lambda \leq b-1$ the function $\mathcal{E}_{\lambda,\xi}$ is essentially non-increasing, i.e.*

$$\mathcal{E}_{\lambda,\xi}(t) \leq \mathcal{E}_{\lambda,\xi}(s) \quad \text{for a.e. } t_0 \leq s \leq t \tag{4.5}$$

In particular $\mathcal{E}_{\lambda,\xi}(t) \leq \mathcal{E}_{\lambda,\xi}(t_0)$ for a.e. $t \geq t_0$

Proof. Following [6], as $b \geq 3$ by choosing $2 \leq \lambda \leq b-1$ in (4.4), we have that $\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) \leq 0$ for all $\gamma > 0$. Hence $\mathcal{E}_{\lambda,\xi}^{\gamma}$ is non-increasing in $[t_0, +\infty)$. In particular for $2 \leq \lambda \leq b-1$ and $\xi = \lambda(b-\lambda-1)$, we have :

$$\mathcal{E}_{\lambda,\xi}^{\gamma}(t) \leq \mathcal{E}_{\lambda,\xi}^{\gamma}(s) \quad \text{for a.e. } t_0 \leq s \leq t \tag{4.6}$$

Let $T > t_0$. By extracting a suitable subsequence when $\gamma \to 0$ in (4.6), thanks to the approximation scheme $(\mathcal{AS})$, we obtain :

$$\mathcal{E}_{\lambda,\xi}(t) \leq \mathcal{E}_{\lambda,\xi}(s) \quad \text{for a.e. } t_0 \leq s \leq t \leq T \tag{4.7}$$

Since $T > t_0$ is arbitrary, we deduce that $\mathcal{E}_{\lambda,\xi}(t) \leq \mathcal{E}_{\lambda,\xi}(s)$ for a.e. $t_0 \leq s \leq t$ and in particular $\mathcal{E}_{\lambda,\xi}(t) \leq \mathcal{E}_{\lambda,\xi}(t_0)$ for a.e. $t \geq t_0$, which concludes the proof of this Lemma. □

Theorem 4.1. *Let $x :$ be a shock solution of* (DI) *and x^* a minimizer of F. Then, there exist some positive constants $C_1, C_2 > 0$, such that for almost every $t \geq t_0$, it holds :*

$$W(t) \leq \frac{C_1}{t^2} \quad \text{and} \quad \|\dot{x}(t)\| \leq \frac{C_2}{t} \quad \text{for a.e. } t \geq t_0 \tag{4.8}$$

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

$$\int_{t_0}^{+\infty} tW(t)dt < +\infty \quad \text{and} \quad \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty \tag{4.9}$$

Proof. Since $\mathcal{E}_{\lambda,\xi}$ is essentially non-increasing, for $2 \leq \lambda \leq b-1$ we have :

$$t^2 W(t) \leq \mathcal{E}_{\lambda,\xi}(t) \leq \mathcal{E}_{\lambda,\xi}(t_0) < +\infty \quad \text{for a.e. } t \geq t_0 \quad \text{and}$$
$$t\|\dot{x}(t)\| \leq \sqrt{\mathcal{E}_{\lambda,\xi}(t)} + \underset{t\geq t_0}{ess\sup}\{\|x(t) - x^*\|\} \leq \sqrt{\mathcal{E}_{\lambda,\xi}(t_0)} + \underset{t\geq t_0}{ess\sup}\{\|x(t) - x^*\|\} \tag{4.10}$$
$$< +\infty$$

which concludes the first point of the Theorem 4.1, with $C_1 = \mathcal{E}_{\lambda,\xi}(t_0)$ and $C_2 = \sqrt{\mathcal{E}_{\lambda,\xi}(t_0)} + \underset{t\geq t_0}{ess\sup}\{\|x(t) - x^*\|\}$

8

Here we must stress out that since F is coercive and $F(x(t))$ essentially bounded, it follows that $\|x(t) - x^*\|$, is also essentially bounded.

For the second point, for $b > 3$, by choosing $\lambda = b - 1$ in (4.4) we obtain :

$$\frac{d}{dt}\mathcal{E}^\gamma_{\lambda,\xi}(t) \leq (3 - b)tW_\gamma(t) \tag{4.11}$$

By integrating in $[t_0, T]$, we have :

$$\int_{t_0}^{T} tW_\gamma(t)dt \leq \mathcal{E}^\gamma_{\lambda,\xi}(t_0) - \mathcal{E}^\gamma_{\lambda,\xi}(T) \leq \mathcal{E}^\gamma_{\lambda,\xi}(t_0) < +\infty \tag{4.12}$$

By passing to the limit (up to a subsequence) when $\gamma \to 0$ thanks to the convergence scheme ($\mathcal{AS}$) we deduce that : $\int_{t_0}^{T} tW(t)dt \leq \mathcal{E}^\gamma_{\lambda,\xi}(t_0) < +\infty$ Since the last inequality hold for all $T > t_0$, we obtain $\int_{t_0}^{\infty} tW(t)dt < +\infty$.

In the same way for $\lambda = 2$ and $b > 3$ in (4.4), we find :

$$\frac{d}{dt}\mathcal{E}^\gamma_{\lambda,\xi}(t) \leq (3 - b)t\|\dot{x}_\gamma(t)\|^2 \tag{4.13}$$

where by integrating and passing to the limit when $\gamma \to 0$ we find $\int_{t_0}^{+\infty} \|\dot{x}_\gamma(t)\|^2 dt < +\infty$ which concludes the proof of Theorem 4.1. $\qquad\qquad\square$

Fast asymptotic convergence to a minimum

The last Theorem asserts that for $b \geq 3$, $W(t)$ and $\|\dot{x}(t)\|^2$ is of order of $O(t^{-2})$ a.e. asymptotically. Nevertheless for $b > 3$, this order can be improved to one of $o(t^{-2})$ a.e.

Theorem 4.2. *Let $b > 3$, x a shock solution of* (DI) *and x^* a minimizer of F. Then*

$$ess \lim_{t\to\infty} t^2 W(t) = 0 \quad and \quad ess \lim_{t\to\infty} t\|\dot{x}(t)\| = 0 \tag{4.14}$$

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

$$U(t) = t^2 W(t) + \frac{t^2}{2}\|\dot{x}(t)\|^2 \geq 0 \tag{4.15}$$

and its approximation, for all $\gamma > 0$ and x_γ solution to (ADE) :

$$U_\gamma(t) = t^2 W_\gamma(t) + \frac{t^2}{2}\|\dot{x}_\gamma(t)\|^2 \geq 0 \quad, \forall t \in [t_0, +\infty) \tag{4.16}$$

By differentiating, we have :

$$\frac{d}{dt}U_\gamma(t) = t^2\langle z, \dot{x}_\gamma(t)\rangle + t^2\langle \ddot{x}_\gamma, \dot{x}(t)\rangle + 2tW_\gamma(t) + t\|\dot{x}(t)\|^2 \tag{4.17}$$

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

$$\frac{d}{dt}U_\gamma(t) = 2tW_\gamma(t) + (1 - b)t\|\dot{x}_\gamma(t)\|^2 \leq 2tW_\gamma(t) \tag{4.18}$$

We now define the function $\Theta_\gamma(t) = U_\gamma(t) - \int_{t_0}^{t} 2sW_\gamma(s)ds$. By definition, Θ has non-positive derivative, hence it is non-increasing, i.e.

$$\Theta_\gamma(t) \leq \Theta_\gamma(s) \quad \forall t_0 \leq s \leq t \tag{4.19}$$

By passing to the limit in (4.19) up to a subsequence when $\gamma \to 0$, thanks to the convergence scheme ($\mathcal{AS}$), we obtain that the function $\Theta(t) = U(t) - \int_{t_0}^{t} 2sW(s)ds$ is essentially non-increasing. In addition from Theorem 4.1 for $b > 3$ we have that $tW(t)$ is integrable, therefore the function Θ is essentially bounded from below. Since it is also essentially non-increasing, it is essentially convergent i.e. : $ess \lim_{t\to\infty} \Theta(t) = l \in \mathbb{R}$

9

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

$$ess \lim_{t \to +\infty} U(t) = ess \lim_{t \to +\infty} \Theta(t) + \int_{t_0}^{+\infty} 2tW(t)\, dt \in \mathbb{R} \qquad (4.20)$$

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

$$\int_{t_0}^{+\infty} \frac{1}{t} U(t)\, dt = \int_{t_0}^{+\infty} tW(t)\, dt + \frac{1}{2} \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2\, dt < +\infty \qquad (4.21)$$

As $\int_{t_0}^{+\infty} \frac{1}{t}\, dt = +\infty$ and $U(t)$ is essentially convergent when $t \to +\infty$, we deduce that $ess \lim_{t \to \infty} U(t) = 0$. This together with the positivity of $t^2 W(t)$ and $\frac{t^2}{2}\|\dot{x}(t)\|^2$ a.e. allow us to conclude the Theorem. $\quad\square$

Convergence of the trajectory to a minimizer

Lastly, for $b > 3$ we show that the trajectory $\{x(t)\}_{t \geq t_0}$ converges to a minimizer. More precisely we have the following Theorem.

Theorem 4.3. *Let x be a shock 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 [28] or Lemma 4.1 in [8]) :

Lemma 4.2. *Let $S \subset \mathbb{R}^d$ be a non-empty set and $x : [t_0, +\infty)$ 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 4. We will invoke the previous Lemma with $S = \arg\min F$. In fact Opial's Lemma holds true for a general separable Hilbert space , but in our case as $\mathbb{R}^d$ is finite-dimensional, we also deduce strong convergence of $x(t)$ to a point of S.

Proof. By Theorem 4.1 for $b > 3$ and suitable λ and ξ the energy function $\mathcal{E}_{\lambda,\xi}$ is essentially non-increasing and bounded from below (at least by zero), so it is essentially convergent. By developing the term $\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2$ in the definition of $\mathcal{E}_{\lambda,\xi}$, we have :

$$\mathcal{E}_{\lambda,\xi}(t) = t^2 W(t) + t^2 \|\dot{x}(t)\|^2 + \lambda t \langle x(t) - x^*, \dot{x}(t) \rangle + \frac{\lambda^2 + \xi}{2} \|x(t) - x^*\|^2 \qquad (4.22)$$

Since by Theorem 4.2, for $b > 3$, $ess \lim_{t \to \infty} t^2 W(t) = 0$ and $ess \lim_{t \to \infty} t\|\dot{x}(t)\| = 0$, from (4.22), we deduce that $\|x(t) - x^*\|$ essentially converges with :

$$ess \lim_{t \to \infty} \|x(t) - x^*\| = ess \lim_{t \to \infty} \sqrt{\frac{2\mathcal{E}_{\lambda,\xi}(t)}{\lambda^2 + \xi}} \qquad (4.23)$$

Since x is Lipschitz continuous function we deduce that $\lim_{t \to \infty} \|x(t) - x^*\| \in \mathbb{R}$. 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 :

$$F(\tilde{x}) \leq \liminf_{t \to \infty} F(x(t)) \qquad (4.24)$$

By Theorem 4.1 we have that $ess \lim_{t \to \infty} F(x(t)) = F(x^*)$, 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. $\quad\square$

4.1.2 The case of low friction $0 < b < 3$

In this paragraph we investigate the asymptotic properties of a shock solution when the friction parameter is low, i.e. $b \in (0,3)$. Our analysis follows the one made for the solutions of the differential equation (1.2) (see [11] and [7]). Here we extend this analysis into the non-differential case. For this purpose, for $\lambda = \frac{2b}{3}$, $\xi = \lambda(\lambda + 1 - b) = \frac{2b(3-b)}{9} > 0$ and $c = 2 - \frac{2b}{3} > 0$, for all $t \geq t_0$, we consider the following energy-function :

$$\mathcal{H}(t) = t^{-c}\mathcal{E}_{\lambda,\xi}(t) = t^{-c}\left(W(t) + \frac{1}{2}\|\frac{2b}{3}(x(t) - x^*) + t\dot{x}(t)\|^2 + \frac{2b(3-b)}{3}\|x(t) - x^*\|^2\right) \qquad (4.25)$$

Proposition 4.1. $\mathcal{H}$ *is essentially non-increasing function, i.e. for almost every $s, t \geq t_0$, such that $s \leq t$, we have $\mathcal{H}(t) \leq \mathcal{H}(s)$ (see also Lemma 3.1 in [5]).*

As before in order to prove we will consider the approximating scheme ($\mathcal{AS}$)

Proof. We recall that from (4.3), for all $\gamma > 0$ we have

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) = 2tW_{\gamma}(t) - \lambda t\langle \nabla F_{\gamma}(x_{\gamma}(t)), x_{\gamma}(t) - x^* \rangle + (\lambda + 1 - b)t\|\dot{x}_{\gamma}(t)\|^2$$
$$+ \big(\xi + \lambda(\lambda + 1 - b)\big)\langle \dot{x}_{\gamma}(t), x_{\gamma}(t) - x^* \rangle \qquad (4.26)$$

By convexity of F_{γ} we have :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) \leq (2 - \lambda)tW_{\gamma}(t) + (\lambda + 1 - b)t\|\dot{x}_{\gamma}(t)\|^2$$
$$+ \big(\xi + \lambda(\lambda + 1 - b)\big)\langle \dot{x}_{\gamma}(t), x_{\gamma}(t) - x^* \rangle \qquad (4.27)$$

In addition by definition of $\mathcal{E}_{\lambda,\xi}^{\gamma}$, by developing the term $\|\lambda(x_{\gamma}(t) - x^*) + t\dot{x}_{\gamma}(t)\|^2$, we find :

$$t\|\dot{x}_{\gamma}(t)\|^2 = \frac{2}{t}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) - 2tW_{\gamma}(t) - 2\lambda\langle x_{\gamma}(t) - x^*, \dot{x}_{\gamma}(t)\rangle - \frac{(\lambda^2 + \xi)}{t}\|x_{\gamma}(t) - x^*\|^2 \qquad (4.28)$$

By injecting the last equality into (4.27), we obtain :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) \leq +\frac{2(\lambda + 1 - b)}{t}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) - \frac{(\lambda^2 + \xi)(\lambda + 1 - b)}{t}\|x_{\gamma}(t) - x^*\|^2$$
$$+ (2b - 3\lambda)tW_{\gamma}(t) + \big(\xi - \lambda(\lambda + 1 - b)\big)\langle \dot{x}_{\gamma}(t), x_{\gamma}(t) - x^* \rangle \qquad (4.29)$$

For $\lambda = \frac{2b}{3}$, $\xi = \lambda(\lambda + 1 - b) > 0$ and $c = 2 - \frac{2b}{3}$, we obtain :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) \leq \frac{c}{t}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) - \frac{2b(9 - b^2)}{27t}\|x_{\gamma}(t) - x^*\|^2 \qquad (4.30)$$

which is equivalent (by multiplying both sides by t^{-c}) to

$$t^{-c}\frac{d}{dt}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) - ct^{-c-1}\mathcal{E}_{\lambda,\xi}^{\gamma}(t) \leq -\frac{2b(9 - b^2)}{27}t^{-c-1}\|x_{\gamma}(t) - x^*\|^2 \leq 0 \qquad (4.31)$$

If we name $\mathcal{H}_{\gamma} = \mathcal{E}_{\lambda,\xi}^{\gamma}$, for $\lambda = \frac{2b}{3}$, $\xi = \lambda(\lambda + 1 - b)$, the inequality (4.31) shows that $\mathcal{H}_{\gamma}$ has a non positive derivative, for all $\gamma > 0$. Hence $\mathcal{H}_{\gamma}$ is a non-increasing function for all $\gamma > 0$, i.e. :

$$\mathcal{H}_{\gamma}(t) \leq \mathcal{H}_{\gamma}(s) \quad \text{for all } t_0 \leq s \leq t \qquad (4.32)$$

By passing to the limit (up to a subsequence) and using the approximate scheme ($\mathcal{AS}$), we conclude the proof of the Proposition. $\qquad \square$

As a result of the previous proposition, in the same spirit as in the proof of Theorem 4.1 we have the following theorem :

Theorem 4.4. *Let x be a shock solution of (DI) and x^* a minimizer of F. If $0 < b < 3$, there exist some positive constants $C_1, C_2 > 0$, such that for almost every $t \geq t_0$, it holds :*

$$W(t) \leq \frac{C_1}{t^{\frac{2b}{3}}} \quad \text{and} \quad \|\dot{x}(t)\| \leq \frac{C_2}{t^{\frac{b}{3}}} \quad \text{for a.e. } t \geq t_0 \qquad (4.33)$$

with $C_1 = \mathcal{E}_{\lambda,\xi}(t_0)$ and $C_2 = \sqrt{\mathcal{E}_{\lambda,\xi}(t_0)} + \underset{t \geq t_0}{ess\sup}\{\|x(t) - x^*\|\}$, where $\lambda = \frac{2b}{3}$ and $\xi = \frac{2b(3-b)}{9}$

11

4.2   The case of $\mathcal{D}(F) = \mathbb{R}^d$

In this section we present the results concerning the asymptotic analysis in the case when $\mathcal{D}(F) = \mathbb{R}^d$. In that case the regularity of a solution of (DI) allow to have finer results than in the previous paragraph. In fact given the regularity $W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$ of a solution x of (DI), most of the results presented here can be presented as direct corollaries from Theorems 4.1, 4.2, 4.3 and 4.4 (remark that when $\mathcal{D}(F) = \mathbb{R}^d$, $W(t)$ and $\|\dot{x}(t)\|$ are defined for all $t \geq t_0$). Nevertheless we give full proofs to stress out the importance of this regularity of the solution in the case where $\mathcal{D}(F)$ is the whole space $\mathbb{R}^d$.

In particular, as we will see, these proofs can be well adapted to the ones made in the differential setting (see ,[31], [6], [25], [11] and [7]) and there is no need to consider the Moreau-Yosida approximation and pass through the different approximation schemes.

First, we recall the definition of an absolutely continuous function that we will use later (see for instance Example 1.13 in [19]).

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

$$\sum_{i \in J} (b_i - a_i) < \delta \quad \Longrightarrow \quad \sum_{i \in J} |G(b_i) - G(a_i)| < \varepsilon \tag{4.34}$$

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

$$G(t) = G(s) + \int_s^t v(\tau) d\tau \quad \forall a \leq s \leq t \leq b \tag{4.35}$$

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 5. From the definition of absolute continuity (in particular (4.35)), 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 [14] (Lemme 3.3) and allows us to " use the chain rule for differentiation".

Lemma 4.3. *Let $T > t_0$ and F be a convex, lower semi-continuous, proper function and $x \in W^{1,2}((t_0, T); \mathbb{R}^d)$. Let also $h \in L^2((t_0, T); \mathbb{R}^d)$, 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 :*

$$\frac{d}{dt}\big(F(x(t))\big) = \langle z, \dot{x}(t) \rangle \quad \forall z \in \partial F(x(t)) \quad a.e. \ in \ (t_0, T) \tag{4.36}$$

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

In view of Lemma 4.3, $W(t)$ is absolutely continuous in $[t_0, T]$ with :

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

In addition as $\dot{x} \in W^{1,\infty}((t_0, T); \mathbb{R}^d)$, it is in particular Lipschitz continuous in (t_0, T) (see characterization of $W^{1,\infty}$ space, Theorem 4.1 in [23] or Theorem 4.5 in [20]), therefore it is absolutely continuous in $[t_0, T]$. As a consequence we have the following proposition.

Proposition 4.2. *Let $T > t_0$. The energy $\mathcal{E}_{\lambda, \xi}$ is absolutely continuous on $[t_0, T]$ with :*

$$\frac{d}{dt}\mathcal{E}_{\lambda, \xi}(t) \leq (2 - \lambda)tW(t) + \big(\xi + \lambda(\lambda + 1 - b)\big)\langle \dot{x}(t), x(t) - x^* \rangle + (\lambda + 1 - b)t\|\dot{x}(t)\|^2 \tag{4.37}$$
$$a.e \ in \ (t_0, T)$$

Proof. By definition $\mathcal{E}_{\lambda, \xi}$ is an absolutely continuous as sum of absolutely continuous functions. In addition, by Lemma 4.3, let $z \in \partial F(x(t))$ such that (DI) holds. We have

$$\frac{d}{dt}\mathcal{E}_{\lambda, \xi}(t) = 2tW(t) + t^2\langle z, \dot{x}(t) \rangle + \langle (\lambda + 1)\dot{x}(t) + t\ddot{x}(t), \lambda(x(t) - x^*) + t\dot{x}(t) \rangle$$
$$+ \xi\langle x(t) - x^*, \dot{x}(t) \rangle \quad \text{a.e in } (t_0, T) \tag{4.38}$$

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

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) = 2tW(t) - \lambda t\langle z, x(t) - x^* \rangle + \big(\xi + \lambda(\lambda + 1 - b)\big)\langle \dot{x}(t), x(t) - x^* \rangle$$
$$+ (\lambda + 1 - b)t\|\dot{x}(t)\|^2 \quad \text{a.e in } (t_0, T) \tag{4.39}$$

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

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \leq (2 - \lambda)tW(t) + \big(\xi + \lambda(\lambda + 1 - b)\big)\langle \dot{x}(t), x(t) - x^* \rangle + (\lambda + 1 - b)t\|\dot{x}(t)\|^2$$
$$\text{a.e in } (t_0, T) \tag{4.40}$$

which concludes the proof of Proposition 4.2 $\qquad\square$

4.2.1 The case of high friction $b \geq 3$

Corollary 4.3. *For any $\xi = \lambda(b - \lambda - 1)$ and $2 \leq \lambda \leq b - 1$, the energy-function $\mathcal{E}_{\lambda,\xi}$ is non-increasing in $[t_0, +\infty)$*

Proof. By relation (4.37) of Proposition 4.2, if we choose $\xi = \lambda(b - \lambda - 1)$ and $2 \leq \lambda \leq b - 1$, as $b \geq 3$, we have that :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \leq 0 \quad \text{a.e in } (t_0, T) \tag{4.41}$$

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

Remark 6. Here we must point out that the absolute continuity of $\mathcal{E}_{\lambda,\xi}$ is essential, since by (4.41), and supposing only continuity of $\mathcal{E}_{\lambda,\xi}$, one cannot conclude directly that $\mathcal{E}_{\lambda,\xi}$ is non-increasing.

In view of the previous Lemma and the non-increasing property of $\mathcal{E}_{\lambda,\xi}$, as in [6], we have the following Theorem. Its proof -which we omit- is similar to the one made before for Theorem (4.1) without need to consider the approximating energy function $\{\mathcal{E}_{\lambda,\xi}^\gamma\}_{\gamma > 0}$, due to the regularity of $\mathcal{E}_{\lambda,\xi}$.

Theorem 4.5. *Let $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$, $\forall T > t_0$, be a solution of (DI) and x^* a minimizer of F. Then, there exist some positive constants $C_1, C_2 > 0$, such that for all $t \geq t_0$, it holds :*

$$W(t) \leq \frac{C_1}{t^2} \quad and \quad \|\dot{x}(t)\| \leq \frac{C_2}{t} \tag{4.42}$$

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

$$\int_{t_0}^{+\infty} tW(t)dt < +\infty \quad and \quad \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty \tag{4.43}$$

Fast asymptotic convergence to a minimum As in the case of shock solutions, for $b > 3$, we can expect a slightly better convergence rate for $W(t)$ and $\|\dot{x}(t)\|$ than the ones given in Theorem 4.5. 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 [25].

Theorem 4.6. *Let $b > 3$, $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$, $\forall T > t_0$ a solution of (DI) and x^* a minimizer of F. Then*

$$\lim_{t \to \infty} t^2 W(t) = 0 \quad and \quad \lim_{t \to \infty} t\|\dot{x}(t)\| = 0 \tag{4.44}$$

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

Proof. The proof follows the one of Theorem 4.2 without need to pass from the approximate scheme.

Let $T > t_0$. Since U is absolutely continuous on $[t_0, T]$ as sum of absolutely continuous functions, by Lemma 4.3 for a $z \in \partial F(x(t))$ such that (DI) holds, we have :

$$\frac{d}{dt}U(t) = t^2\langle z, \dot{x}(t) \rangle + t^2\langle \ddot{x}, \dot{x}(t) \rangle + 2tW(t) + t\|\dot{x}(t)\|^2$$
$$= 2tW(t) + (1 - b)t\|\dot{x}(t)\|^2 \leq 2tW(t) \quad \text{a.e. in } (t_0, T) \tag{4.45}$$

13

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 :

$$\left[\frac{d}{dt}U(t)\right]_+ \leq 2tW(t) \quad \text{a.e. in } (t_0, T) \tag{4.46}$$

By Theorem 4.5 for $b > 3$, the term $2tW(t)$ is integrable on $[t_0, T)$ for all $T > t_0$, and so is $\left[\frac{d}{dt}U(t)\right]_+$.

The function $\Theta(t) = U(t) - \int_{t_0}^t \left[\frac{d}{ds}U(s)\right]_+ ds$ 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 $[t_0, +\infty)$. From this point the proof is exactly the same as the one made before for Theorem 4.2. $\qquad \square$

Convergence of the trajectory to a minimizer Lastly, we establish the convergence of the trajectory towards a minimizer. In fact the result is already established in the case of shock solutions, nevertheless -as already mentioned- here we give a proof in the spirit of the one made in [6], by exploiting the regularity of the solution. More precisely we have the following Theorem.

Theorem 4.7. *Let Let $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$ for all $T > t_0$, 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.*

Proof. As in proof of Theorem 4.3, we will use Opial's Lemma. In order to apply Opial's Lemma, we define : $\psi(t) = \frac{1}{2}\|x(t) - x^*\|$. Let $T > t_0$. As $x \in W^{2,\infty}((t_0, T); \mathcal{H})$ 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 :

$$\begin{aligned}
\ddot{\psi}(t) + \frac{b}{t}\dot{\psi}(t) &= \|\dot{x}(t)\|^2 + \langle \ddot{x}(t) + \frac{b}{t}\dot{x}(t), x(t) - x^* \rangle \\
&= \|\dot{x}(t)\|^2 - \langle z(t), x(t) - x^* \rangle \\
&\leq \|\dot{x}(t)\|^2 - W(t) \leq \|\dot{x}(t)\|^2 \quad \text{a.e. in } (t_0, T)
\end{aligned} \tag{4.47}$$

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 :

$$t^b\ddot{\psi}(t) + bt^{b-1}\dot{\psi}(t) \leq t^b\|\dot{x}(t)\|^2 \quad \text{a.e. in } (t_0, T) \tag{4.48}$$

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

$$\dot{\psi}(s) \leq \frac{t_0^b \dot{\psi}(t_0)}{s^b} + \frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2 dt \leq \frac{C_0}{s^b} + \frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2 dt \tag{4.49}$$

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

$$[\dot{\psi}]_+(s) \leq \frac{C_0}{s^b} + \frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2 dt \tag{4.50}$$

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

$$\begin{aligned}
\int_{t_0}^T [\dot{\psi}]_+(s)ds &\leq C_0 \int_{t_0}^T \frac{1}{s^b} + \int_{t_0}^T \frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2 dtds \\
&= (b-1)C_0(t_0^{1-b} - T^{1-b}) + \int_{t_0}^T t^b\|\dot{x}(t)\|^2 \left(\int_t^T s^{-b}ds\right)dt \\
&\leq C_0 + (b-1)\int_{t_0}^T t\|\dot{x}(t)\|^2 dt
\end{aligned} \tag{4.51}$$

14

Since, by Theorem 4.5, for $b > 3$, the right-hand member of this inequality is finite for every $T > t_0$, we deduce that :

$$\int_{t_0}^{+\infty} [\dot{\psi}]_+(s)ds < +\infty \tag{4.52}$$

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 :

$$\lim_{t \to \infty} \psi(t) = \theta_\infty + \int_{t_0}^{+\infty} [\dot{\psi}]_+(s)ds \in \mathbb{R} \tag{4.53}$$

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 :

$$F(\tilde{x}) \leq \liminf_{t \to \infty} F(x(t)) \tag{4.54}$$

By Theorem 4.5 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. $\qquad \square$

4.2.2 The case of low friction $0 < b < 3$

Let $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$ for all $T > t0$, be a solution to (DI). As in paragraph 4.1.2, for $0 < b < 3$, we study the energy function $\mathcal{H}(t) = t^{-c}\mathcal{E}_{\lambda,\xi}(t)$ for all $t \geq t_0$, where $c = 2 - \frac{2b}{3}$, $\lambda = \frac{2b}{3}$ and $\xi = \frac{2b(3-b)}{9}$. The function $\mathcal{H}$ is absolutely continuous on every interval $[t_0, T]$, $\forall T > t_0$, as a product of absolutely continuous function.

Proposition 4.4. $\mathcal{H}$ *is a non-increasing function.*

Proof. The proof follows the same arguments as in paragraph 4.1.2 with some simplifications. From (4.40) we have

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \leq (2-\lambda)tW(t) + \left(\xi + \lambda(\lambda+1-b)\right)\langle \dot{x}(t), x(t) - x^*\rangle + (\lambda+1-b)t\|\dot{x}(t)\|^2 \tag{4.55}$$
$$\text{a.e in } (t_0, T)$$

By developing the term $\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2$ in definition of $\mathcal{E}$ and substituting in (4.55) (as in proof of Proposition 4.1), for $\lambda = \frac{2b}{3}$, $\xi = \lambda(\lambda+1-b) > 0$ and $c = 2 - \frac{2b}{3}$, we obtain :

$$\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) \leq \frac{c}{t}\mathcal{E}_{\lambda,\xi}(t) - \frac{2b(9-b^2)}{27t}\|x(t) - x^*\|^2 \quad \text{a.e. } (t_0, T) \tag{4.56}$$

Hence, as $\mathcal{H}$ is absolutely continuous, we have :

$$\frac{d}{dt}\mathcal{H}(t) = t^{-c-1}\left(t\frac{d}{dt}\mathcal{E}_{\lambda,\xi}(t) - c\mathcal{E}_{\lambda,\xi}(t)\right) \leq -\frac{2b(9-b^2)}{27}t^{-c-1}\|x(t) - x^*\|^2 \leq 0 \tag{4.57}$$
$$\text{a.e. in } (t_0, T)$$

Since $\mathcal{H}$ is absolutely continuous, after (4.57), it is non-increasing in $[t_0, T]$ for all $T > t_0$. By continuity of $\mathcal{H}$, it is non-increasing in whole $[t_0, +\infty)$ $\qquad \square$

As a direct consequence of the non-increasing property of $\mathcal{H}$, we have the following Theorem.

Theorem 4.8. *Let $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$, $\forall T > t_0$ be a solution of* (DI) *and x^* a minimizer of F. If $0 < b < 3$, there exist some positive constants $C_1, C_2 > 0$, such that for all $t \in [t_0, +\infty)$, it holds :*

$$W(t) \leq \frac{C_1}{t^{\frac{2b}{3}}} \quad \text{and} \quad \|\dot{x}(t)\| \leq \frac{C_2}{t^{\frac{b}{3}}} \tag{4.58}$$

with $C_1 = \mathcal{E}_{\lambda,\xi}(t_0)$ and $C_2 = \sqrt{\mathcal{E}_{\lambda,\xi}(t_0)} + \sup_{t \geq t_0}\{\|x(t) - x^*\|\}$, where $\lambda = \frac{2b}{3}$ and $\xi = \frac{2b(3-b)}{9}$

Finally we present that when $0 < b < 3$, the convergence rate $O\left(t^{-\frac{2b}{3}}\right)$ of $W(t)$, that Theorem 4.8 asserts is optimal.

5   Optimality of convergence rate for $0 < b < 3$

In this section we will study the differential inclusion (DI), for $0 < b < 3$ when $F(x) = |x|$. This function enters in the framework studied before and in particular $\mathcal{D}(F) = \mathbb{R}$.

In this case, Theorem 3.1 asserts that (DI) admits a solution x such that $x \in W^{2,\infty}((t_0, T); \mathbb{R}^d) \cap \mathcal{C}^1([t_0, +\infty); \mathbb{R}^d)$, for all $T > t_0$. In addition, Theorem 4.8 asserts that when $0 < b < 3$, the convergence rate of $|x(t)|$ to zero is of order of $O\left(t^{-\frac{2b}{3}}\right)$ asymptotically. We show that this order is optimal. In particular we have the following Theorem.

Theorem 5.1. *Let x be a solution of* (DI) *with $F(x) = |x|$ and $0 < b < 3$ such that $x(t_0) \neq 0$. Then there exists a constant $K_1 > 0$, such that for any $T > 0$, there exists $t > T$ such that :*

$$|x(t)| \geq \frac{K_1}{t^{\frac{2b}{3}}} \tag{5.1}$$

Before proceeding to the proof, we must stress out some facts concerning the particular example of $F(x) = |x|$.

Since the minimizer of F is clearly zero (i.e. $x^* = 0$) and F is a convex, positively $1-$homogeneous function, we have :

$$W(t) = F(x(t)) - F(x^*) = |x(t)| = \langle z, x(t)\rangle \quad \text{with } z \in \partial F(x(t)) \tag{5.2}$$

In addition for any $\lambda, \xi \geq 0$:

$$\mathcal{E}_{\lambda,\xi}(t) = t^2 |x(t)| + \frac{1}{2}|\lambda x(t) + t\dot{x}(t)|^2 + \frac{\xi}{2}|x(t)|^2 \tag{5.3}$$

and for $\lambda = \frac{2b}{3}$, $\xi = \frac{2b(3-b)}{9} > 0$ and $c = 2 - \frac{2b}{3}$

$$\mathcal{H}(t) = t^{-c}\mathcal{E}_{\lambda,\xi}(t) = t^{2-c}|x(t)| + \frac{t^{-c}}{2}|\lambda x(t) + t\dot{x}(t)|^2 + \frac{\xi t^{-c}}{2}|x(t)|^2 \tag{5.4}$$

In order to prove Theorem 5.1, we will make use of the following Lemma :

Lemma 5.1. *Let $0 < b < 3$ and x a solution to* (DI) *such that $x(t_0) = x_0 > 0$. Then $\lim\limits_{t \to \infty} \mathcal{H}(t) = l > 0$*.

Proof. Let $T > t_0$. From (4.39) and (5.2) we have :

$$\dot{\mathcal{E}}_{\lambda,\xi}(t) = (2-\lambda)tW(t) + \big(\xi + \lambda(\lambda + 1 - b)\big)\langle\dot{x}(t), x(t) - x^*\rangle + (\lambda + 1 - b)t\|\dot{x}(t)\|^2 \tag{5.5}$$
$$\text{a.e in } (t_0, T)$$

Since $\lambda = \frac{2b}{3}$ and $\xi = \frac{2b(3-b)}{9} > 0$, by substituting the term $t\|\dot{x}(t)\|^2$ as exactly done in the previous paragraph, we find :

$$\dot{\mathcal{E}}_{\lambda,\xi}(t) = \frac{c}{t}\mathcal{E}_{\lambda,\xi}(t) - \frac{2b(9 - b^2)}{27t}\|x(t) - x^*\|^2 \quad \text{a.e. in } (t_0, T) \tag{5.6}$$

where $c = 2 - \frac{2b}{3}$

By rewriting the previous equation in terms of $\mathcal{H}$, we have :

$$\dot{\mathcal{H}}(t) = -\frac{2b(9 - b^2)}{27}t^{-c-1}\|x(t) - x^*\|^2 \quad \text{a.e. in } (t_0, T) \tag{5.7}$$

which in our framework can also be written as :

$$\dot{\mathcal{H}}(t) = -\frac{2b(9 - b^2)}{27}t^{-c-1}|x(t)|^2 \quad \text{a.e. in } (t_0, T) \tag{5.8}$$

By definition of $\mathcal{H}$ (5.4) and its non-increasing property, for all $t \geq t_0$, we have :

$$t^{2-c}|x(t)| \leq \mathcal{H}(t) \leq \mathcal{H}(t_0) \tag{5.9}$$

16

By injecting the last inequality into (5.8), we find :

$$\dot{\mathcal{H}}(t) = -\frac{2b(9-b^2)}{27}t^{c-5}t^{c-2}|x(t)|t^{c-2}|x(t)| \geq -\frac{2b(9-b^2)}{27}t^{c-5}\mathcal{H}(t_0)\mathcal{H}(t) \tag{5.10}$$
$$\text{a.e. in } (t_0, T)$$

Hence if we set the functions $\psi(t) = \frac{2b(9-b^2)\mathcal{H}(t_0)}{27(c-4)}t^{c-4}$ and $\Psi(t) = \mathcal{H}(t)e^{\psi(t)}$ for all $t \geq t_0$, we have that ψ and Ψ are absolutely continuous with :

$$\dot{\Psi}(t) = e^{\psi(t)}\big(\dot{\mathcal{H}}(t) + \xi\mathcal{H}(t_0)\mathcal{H}(t)\big) \geq 0 \quad \text{a.e. in } (t_0, T) \tag{5.11}$$

where we used the relation (5.10) for the last inequality.

From (5.11), and the absolute continuity of Ψ, we deduce that it is non-decreasing on every interval (t_0, T) and since it is continuous we have that Ψ is non-decreasing function for all $t \geq t_0$.

Hence for all $t \geq t_0$, we obtain :

$$\mathcal{H}(t) \geq \mathcal{H}(t_0)e^{\psi(t_0)-\psi(t)} \geq \mathcal{H}(t_0)e^{\psi(t_0)} > 0 \tag{5.12}$$

Since $\mathcal{H}$ is non increasing function and bounded from below, with $\inf_{t \geq t_0}\{\mathcal{H}(t)\} \geq \mathcal{H}(t_0)e^{\psi(t_0)} > 0$, we have that $\lim_{t \to \infty}\mathcal{H}(t) = l \geq \mathcal{H}(t_0)e^{\psi(t_0)} > 0$. $\qquad\square$

We are now ready to give the proof of Theorem 5.1.

Proof. From relation 5.12, we have that :

$$\mathcal{E}_{\lambda,\xi}(t) = \mathcal{H}(t)t^c \geq K_1 t^c \tag{5.13}$$

where $K_1 = \mathcal{H}(t_0)e^{\psi(t_0)}$.

Let $T > t_0$. We distinguish four cases :

1. There exist some $t_1 > T$, such that :

$$\frac{1}{2}|\lambda x(t_1) + t\dot{x}(t_1)|^2 + \frac{\xi}{2}|x(t_1)|^2 \leq \frac{K_1}{2}t_0^c \tag{5.14}$$

   Then from definition of $\mathcal{E}_{\lambda,\xi}(t)$ and (5.13), we deduce that :

$$t_1^2|x(t_1)| \geq K_1 t_1^c - \frac{K_1}{2}t_0^c \geq \frac{K_1}{2}t_1^c \tag{5.15}$$

   which concludes the proof.

2. There exists some $t_2 \geq T$ such that $\dot{x}(t) = 0$ for all $t \geq t_2$. By using the fact that $E(t) = t^2x(t) + \frac{\lambda^2+\xi}{2}|x(t)|^2$ and (5.13), we have :

$$t^2|x(t)| \geq K_1 t^c - \frac{\lambda^2 + \xi}{2}|x(t)|^2 \tag{5.16}$$

   Since $\lim_{t \to \infty}|x(t)|^2 = 0$, there exists some $t \geq t_2$, such that $|x(t)|^2 \leq \frac{K_1}{2}t_0^c$, hence we can conclude as in the first point.

3. There exists some $t_3 > T$ such that $x(t_3) = 0$. Since $\lim_{t \to \infty}|x(t)| = 0$, there exists $t > t_3$ such that $\dot{x}(t) = 0$ thus we can use the previous point to conclude.

4. Finally we suppose that $x(T) > 0$ and that the sign of $\dot{x}$ is constant for all $t \geq T$. Since $\lim_{t \to \infty}|x(t)| = 0$ we deduce that $sign(\dot{x}(t)) < 0$, for all $t \geq T$. In addition for all $t \geq T$, we have :

$$x(t) - x(T) = \int_T^t \dot{x}(s)ds \tag{5.17}$$

   Since $x(t)$ converges to 0, we deduce that for any $\eta > 0$, there exists $t_\eta \geq T$ such that $|t_\eta\dot{x}(t)| < \eta$. Hence for any ϵ, there exists $t_\epsilon \geq T$, such that :

$$\frac{1}{2}|\lambda x(t_\epsilon) + t\dot{x}(t_\epsilon)|^2 + \frac{\xi}{2}|x(t_\epsilon)|^2 < \epsilon \tag{5.18}$$

   thus we can conclude as in the first point.

This concludes the proof of Theorem 5.1. $\qquad\square$

A Appendix

A.1. *The Yosida approximation*

For $\gamma > 0$ and a maximally monotone operator A one can define the resolvent of A and the Yosida approximation of A by $J_\gamma(A)$ and A_γ respectively as follows :

$$J_{\gamma A} = (Id + A)^{-1} \quad \text{and} \quad A_\gamma = \frac{1}{\gamma}(Id - J_{\gamma A}) \tag{A.1}$$

Let $\Phi : \mathbb{R}^d \longrightarrow \overline{\mathbb{R}}$ be a proper lower semi-continuous and convex function and $\partial\Phi$ its subdifferential. Then $\partial\Phi$ is a maximally monotone operator and for $\gamma > 0$, one can define :

$$\nabla\Phi_\gamma = \frac{1}{\gamma}(Id - J_\gamma) \quad \text{where} \quad J_{\gamma\Phi} = (Id + \partial\Phi)^{-1} \tag{A.2}$$

In particular we have :

$$\Phi_\gamma(x) = \min_{y \in \mathbb{R}^d}\{\Phi(y) + \frac{\|x - y\|^2}{2\gamma}\} \quad \text{and} \quad J_{\gamma\Phi}(x) = \underset{y \in \mathbb{R}^d}{\arg\min}\left\{\Phi(y) + \frac{\|x - y\|^2}{2\gamma}\right\} \tag{A.3}$$

Lemma A.1. *The following convergence property holds (see Proposition 2.11 in [14]) :*

$$\Phi_\gamma(x) \underset{\gamma \to 0}{\nearrow} \Phi(x) \quad \forall x \in \mathbb{R}^d \tag{A.4}$$

A.2. *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.2. *(see Proposition 4.14 in [19])*
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 :

$$g(x_n + e_n) - g(x_n) \geq \langle z_n, e_n \rangle = \|z_n\| \quad \forall n \in \mathbb{N} \tag{A.5}$$

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. $\square$

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] Vassilis Apidopoulos, Jean-François Aujol, and Charles Dossal. Convergence rate of inertial forward-backward algorithm beyond nesterov's rule. 2017.

[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, Zaki Chbani, and Hassan Riahi. Rate of convergence of the nesterov accelerated gradient method in the subcritical case $a > 3$. *arXiv preprint arXiv:1706.05671*, 2017.

[8] 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.

[9] Hedy Attouch and Juan Peypouquet. The rate of convergence of nesterov's accelerated forward-backward method is actually faster than 1/k^2. *SIAM Journal on Optimization*, 26(3):1824–1834, 2016.

[10] 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.

[11] JF Aujol and Ch Dossal. Optimal rate of convergence of an ode associated to the fast gradient descent schemes for $b > 0$. 2017.

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

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

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

[21] 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.

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

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

[24] 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.

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

[26] 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.

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

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

[29] Laetitia A Paoli. An existence result for vibrations with unilateral constraints: case of a nonsmooth set of constraints. *Mathematical Models and Methods in Applied Sciences*, 10(06):815–831, 2000.

[30] Michelle Schatzman. A class of nonlinear differential equations of second order in time. *Nonlinear Analysis: Theory, Methods & Applications*, 2(3):355–373, 1978.

[31] 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.