# Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for b>0

Jf Aujol, Ch Dossal

# Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for $b > 0$

J-F. Aujol[*†] and Ch. Dossal [‡§]
Université de Bordeaux, IMB, UMR 5251, F-33400 Talence, France

June 2017

**Abstract** This paper is devoted to the study of an ODE associated to FISTA. New convergence results are presented for convex functions and a class of non convex functions. The asymptotic rates are proven to be optimal with the study of particular instances. Robustness to noise is also investigated.

**keyword** Lyapunov functions, rate of convergence, ODE

## 1 Introduction

We study the following equation

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \nabla F(x(t)) = 0 \tag{1}$$

where $F$ is a differentiable function and $b > 0$, having at least one minimizer. We denote by $x^*$ such a minimizer and $W(t) = F(x(t)) - F(x^*)$.

This ODE is associated to FISTA [6] or Accelerated Gradient Method [11] :

$$x_{n+1} = y_n - \gamma\nabla F(y_n) \text{ and } y_n = x_n + \frac{n}{n+b}(x_n - x_{n-1}) \tag{2}$$

with $\gamma$ and $b$ positive parameters. This equation, including or not a perturbation term, has been widely studied in the literature [4, 12, 7, 5, 10]. It was proved in

[*]Jean-Francois.Aujol@math.u-bordeaux.fr
[†]J-F. Aujol is a member of Institut Universitaire de France.
[‡]Charles.Dossal@math.u-bordeaux.fr
[§]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).

[2] that if $F$ is convex with Lipschitz gradient and if $b > 0$, the trajectory $F(x(t))$, where $x$ is solution of (1), converges to the minimum $F^* := F(x^*)$ of $F$. It is also known that for $b \geqslant 3$ and $F$ convex we have

$$W(t) := F(x(t)) - F(x^*) = 0\left(\frac{1}{t^2}\right) \tag{3}$$

Extending to the continuous setting the proof of Chambolle et al. [8] of the convergence of iterates of FISTA, Attouch et al. [2] proved that for $b > 3$ the trajectory $x$ weakly converges to a minimizer $\tilde{x}$ of $F$.
Su et al. [12] proposed some new results, proving the integrability of $t \mapsto tW(t)$ if $b > 3$, and they gave more accurate bounds on $W(t)$ in the case of strong convexity and if $g = (F - F^*)^\beta$ is convex. More recently several studies including a perturbation term [2] have been proposed but always in the setting when $b \geqslant 3$.

The main contribution of this article is to study the decay of $W(t)$ for any $b > 0$, by giving bounds and showing their optimality. The optimal decay was known when $F$ is convex for $b \geqslant 3$ but not for $b < 3$.
More precise results on the convergence of the trajectory of $x$ are given depending on hypotheses made on $F$. We also study the robustness to perturbations. The results are available for any function $F$ such that $g = (F - F^*)^\beta$ is convex for $\beta > 0$ and they depend on $\beta$. These results extend the known results about solutions of (1) to a class of non-convex functions and they complete the known results on convex functions. Such hypotheses with $\beta < 1$ were also used by Su et al. [12] to show that the $\frac{1}{t^2}$ rate may hold for such functions even for some $b < 3$. Moreover we prove that the given bounds on the decay of $W(t)$ are optimal, by building examples where they are achieved.

The plan of the paper is the following. In Section 2, we present the first main result of the paper, which gives bounds on $W(t)$ for $b > 0$. We give some comments on their implications. Section 3 details the proof of Theorem 1. In Section 4 we prove the optimality of the bounds given in Section 2. Robustness to noise is investigated in Section 5. The proofs of technical lemmas is postponed to the appendix.

# 2 Decay of $F - F^*$ and convergence of the trajectory $x$ for $b > 0$.

In this paper, we consider a function $F$ with values in $\mathbb{R}$, and defined on some finite dimensional Hilbert space $H$ (e.g. $H = \mathbb{R}^n$). Notice that in this case, if $F$ is continous, then there exists locally a solution $x$ for (1) (thanks to the Cauchy-Peano-Arzela theorem, see e.g p. 137 in [9]).

The main contribution of this paper is the following Theorem :

**Theorem 1.** *Let $F$ a differentiable function from a (finite dimension) Hilbert space $H$ to $\mathbb{R}$ having a minimizer $x^* \in H$ and $\beta > 0$ such that the function $g(x) = (F(x) - F(x^*))^\beta$ is convex. Then*

1. *If $b \in (0, 1+2\beta)$ then for any solution $x(t)$ of* (1)

$$W(t) = O\left(\frac{1}{t^{\frac{2b}{2\beta+1}}}\right) \tag{4}$$

2. *if $b \geqslant 2\beta + 1$ then for any solution $x(t)$ of* (1)

$$W(t) = O\left(\frac{1}{t^2}\right) \tag{5}$$

3. *If $b > 2\beta + 1$ then for any solution $x(t)$ of* (1)

   (a) *The trajectory $x(t)$ converges to a minimizer $\tilde{x}$ of $F$.*

   (b)

$$W(t) = o\left(\frac{1}{t^2}\right) \tag{6}$$

Several comments about this theorem:

- Points 2. and 3. are known for the convex case, ie when $\beta = 1$, see for example [10, 2].
- The two last points generalize what is known on convex functions. If the function $F$ is convex and if $\beta$ can be chosen smaller than 1, the proposition ensures that the optimal decay $\frac{1}{t^2}$ is achieved even for some $b \leqslant 3$. If $F$ is not convex but there exists $\beta > 1$ such that $g = (F - F(x^*))^\beta$ is convex, the solution of (1) may converge and ensure good decay if $b$ is chosen sufficiently large.
- The first point ensures that if $b$ is too small, for example $b < 3$ for the convex case, one also has some decay of $W$, which is not as fast as what can be expected for large $b$.
  This decay was proved by Su et al. [12] for strongly convex functions.
- The decay in $O\left(\frac{1}{t^{\frac{2b}{2\beta+1}}}\right)$ is optimal in the sense that we show in Proposition 6 that it can be achieved for some explicit functions $F$ for any $\beta > 1$. As a consequence one cannot expect a $o\left(\frac{1}{t^{\frac{2b}{2\beta+1}}}\right)$ decay if $b = (0, 1+2\beta]$

- For $\beta = 1$ and $b = 3$, we cannot exhibit such a function $F$ as differential inclusions are beyond the scope of this paper. But the same analysis shows that this optimal rate is reached for the differential inclusion associated to (1) with $F(x) = |x|$. Hence for this function we cannot have $W(t) = o(\frac{1}{t^2})$ with $b = 3$.

It is worth rewriting Theorem 1 in the particular case when $\beta = 1$, ie when $F$ is a convex function.

**Corollary 1.** *Let $F$ be a convex lower semi-continuous function (lsc) function from a (finite dimension) Hilbert space $H$ to $\mathbb{R}$ having a minimizer $x^* \in H$. Then*

1. *If $b \in (0,3)$ then for any solution $x(t)$ of* (1)

$$W(t) = O\left(\frac{1}{t^{\frac{2b}{3}}}\right) \tag{7}$$

2. *if $b \geqslant 3$ then for any solution $x(t)$ of* (1)

$$W(t) = O\left(\frac{1}{t^2}\right) \tag{8}$$

3. *If $b > 3$ then for any solution $x(t)$ of* (1)

   (a) *The trajectory $x(t)$ converges to a minimizer $\tilde{x}$ of $F$.*

   (b)

$$W(t) = o\left(\frac{1}{t^2}\right) \tag{9}$$

The two first points of the corollary are well known. Our results show that the same decay for $W$ cannot be expected for small $b > 0$, but that some control of the $W$ nevertheless remains.

While submitting this paper we were informed that in a parallel but independent way, Attouch and al. worked on the same problem with $\beta = 1$ and had just submitted the article "Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$" ( see [1], https://arxiv.org/ abs/ 1706. 05671 ). Nevertheless our present work contains important materials which are not in [1], such as the optimality of the rate of point 1. of Theorem 1 (thanks to Proposition 6), and the fact that our results can be applied to a class of non convex functions.

# 3 Proof of Theorem 1

This section is devoted to the proof of Teorem 1.

The points 2. and 3. of Theorem 1 are generalizations of classical results to $\beta > 0$. The main contribution is the point 1.

## 3.1 Preliminary material and proof of point 2. of Theorem 1

The proofs of all the following results rely on the next lemma, bounding the value of $F(x)$ using $\nabla F(x)$, if $g$ is convex.

**Lemma 1.** *Let $\beta \in ]0,1[$, if $g(x) = (F(x) - F(x^*))^\beta$ is convex then*

$$\frac{1}{\beta}(F(x(t)) - F(x^*)) \leqslant \langle \nabla F(x(t)), x(t) - x^* \rangle \tag{10}$$

**Proof** Since $g$ is convex we have

$$g(x(t)) \leqslant \langle \nabla g(x(t)), x(t) - x^* \rangle \tag{11}$$

and $\nabla g(x(t)) = \beta(F(x(t) - F(x^*))^{\beta-1}\nabla F(x(t))$. Replacing $g(x(t))$ by $(F(x(t)) - F(x^*))^\beta$ we get the result. □

**Notations** In the following we use the notation :

$$W(t) := F(x(t)) - F(x^*) \tag{12}$$

where $x^*$ is a minimizer of $F$. If $x^*$ is a fixed minimizer of $F$, $\lambda \in \mathbb{R}$ and $x$ a solution of (1) we denote :

$$u(t) := tW(t) + \frac{1}{2t}\left\|\lambda(x(t) - x^*) + t\dot{x}(t)\right\|^2 \tag{13}$$

and

$$v(t) := \frac{1}{2t}\left\|x(t) - x^*\right\|^2 \tag{14}$$

The analysis is based on Lyapunov functions. Following Su, Boyd and Candès [12], for any $(\lambda, \xi) \in \mathbb{R}^2$ we define

$$\mathcal{E}_{\lambda,\xi}(t) = t^2 W(t) + \frac{1}{2}\left\|\lambda(x(t) - x^*) + t\dot{x}(t)\right\|^2 + \frac{\xi}{2}\left\|x(t) - x^*\right\|^2. \tag{15}$$

One can observe that this energy can be defined using the functions $u$ and $v$

$$\mathcal{E}(t) = tu(t) + \xi tv(t). \tag{16}$$

As usual, we need to compute the derivative $\mathcal{E}'_{\lambda,\xi}$. We will need two different expressions of the derivative.

**Lemma 2.**

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t(\lambda+1-b)\left\|\dot{x}(t)\right\|^2 & (17)\\ &+ \lambda t\langle -\nabla F(x(t)), x(t)-x^*\rangle + (\lambda(\lambda+1) - b\lambda + \xi)\langle \dot{x}(t), x(t)-x^*\rangle & (18)\end{aligned}$$

The proof of Lemma 2 is given in Appendix and is only a sequence of calculations.

Under the hypotheses of Proposition 1 we have the following Lemma :

**Lemma 3.** *If $g$ is convex, then:*

$$\mathcal{E}'_{\lambda,\xi}(t) \leqslant (2-\frac{\lambda}{\beta})tW(t) + (\xi+\lambda^2+\lambda-b\lambda)\langle \dot{x}(t), x(t)-x^*\rangle + (\lambda+1-b)t\left\|\dot{x}(t)\right\|^2 \quad (19)$$

*Proof.* To prove Lemma 3 we only apply the inequality of Lemma 1 in the equality of Lemma 2. □

The following result will give the proof of point 2. of Theorem 1.

**Proposition 1.** *If $b \geqslant 2\beta+1$, if $\lambda = 2\beta$ and $\xi = \lambda(b-\lambda-1) \geqslant 0$, then*

$$\mathcal{E}'_{\lambda,\xi}(t) \leqslant 0 \quad (20)$$

*Proof.* Let us suppose $b \geqslant 2\beta+1$. If we apply Lemma 3 with $\lambda = 2\beta$ and $\xi = \lambda(b-\lambda-1) \geqslant 0$ we have

$$\mathcal{E}'_{\lambda,\xi}(t) \leqslant (2\beta+1-b)t\left\|\dot{x}(t)\right\|^2 \leqslant 0 \quad (21)$$

□

**Proof of point 2. of Theorem 1** It follows from Proposition 1 that the function $\mathcal{E}_{\lambda,\xi}$ is non increasing and thus bounded. As a consequence, we see that the second point of Theorem 1 holds true. □

## 3.2 Proof of point 1. of Theorem 1.

In this section suppose that $b \in (0, 2\beta+1)$. To prove this point we use another Lyapunov energy. We need another expression of $\mathcal{E}_{\lambda,\xi}(t)$ :

**Lemma 4.**

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + \lambda t\langle -\nabla F(x(t)), x(t)-x^*\rangle & (22)\\ &+ (\xi - \lambda(\lambda+1-b))\langle \dot{x}(t), x(t)-x^*\rangle & (23)\\ &+ (\lambda+1-b)\frac{1}{t}\left\|\lambda(x(t)-x^*) + t\dot{x}(t)\right\|^2 - \frac{\lambda^2(\lambda+1-b)}{t}\left\|x(t)-x^*\right\|^2 & (24)\end{aligned}$$

The proof of Lemma 4 is given in Appendix.

Under the hypotheses of Proposition 1 we have the following Lemma :

**Lemma 5.** *If $g$ is convex, then:*

$$\mathcal{E}'_{\lambda,\xi}(t) \leqslant (2-\frac{\lambda}{\beta})tW(t) + (\xi - \lambda(\lambda+1-b))\langle \dot{x}(t), x(t)-x^*\rangle \tag{25}$$

$$+ (\lambda+1-b)\frac{1}{t}\left\|\lambda(x(t)-x^*)+t\dot{x}(t)\right\|^2 - \frac{\lambda^2(\lambda+1-b)}{t}\left\|x(t)-x^*\right\|^2 \tag{26}$$

*Proof.* To prove Lemma 5 we only apply the inequality of Lemma 1 in the equality of Lemma 4. □

**Notations** Now, we choose $\lambda = \frac{2b\beta}{2\beta+1}$ and $\xi = \lambda(\lambda+1-b)$ and we denote $\mathcal{E}_{\lambda,\xi} = \mathcal{E}$. For these parameters one can observe that $\lambda+1-b = 1-\frac{b}{2\beta+1} \geqslant 0$. We can define

$$c := 2 - \frac{2b}{2\beta+1} \tag{27}$$

Notice that $c > 0$. We introduce the new Lyapunov energy:

$$\mathcal{H}(t) := t^{-c}\mathcal{E}(t). \tag{28}$$

We can rewrite Lemma 5 in a more compact way:

**Lemma 6.** *If $g$ is convex, then:*

$$\mathcal{E}'(t) \leqslant cu(t) - dv(t). \tag{29}$$

*where $d > 0$ is defined as:*

$$d := \left(\frac{2b\beta}{2\beta+1}\right)^2\left(1-\frac{b}{2\beta+1}\right) \tag{30}$$

*Proof.*

$$\mathcal{E}'(t) \leqslant (2-\frac{2b}{2\beta+1})tW(t) + (1-\frac{b}{2\beta+1})\frac{1}{t}\left\|\frac{2b\beta}{2\beta+1}(x(t)-x^*)+t\dot{x}(t)\right\|^2 \tag{31}$$

$$-\left(\frac{2b\beta}{2\beta+1}\right)^2(1-\frac{b}{2\beta+1})\frac{1}{t}\left\|x(t)-x^*\right\|^2 \tag{32}$$

Using

$$u(t) = tW(t) + \frac{1}{2t}\left\|\frac{2b\beta}{2\beta+1}(x(t)-x^*)+t\dot{x}(t)\right\|^2, \tag{33}$$

the definition of $v(t)$, $c$ and $d$, we get the result. □

The following result will give the proof of point 1. of Theorem 1.

**Proposition 2.** *If $b \in (0, 2\beta+1)$, if $\lambda = \frac{2b\beta}{2\beta+1}$ and $\xi = \lambda(\lambda+1-b)$, then*

$$\mathcal{H}'(t) \leqslant 0. \tag{34}$$

*Proof.* We have

$$\mathcal{H}'(t) = t^{-c-1}(t\mathcal{E}'(t) - c\mathcal{E}(t)). \tag{35}$$

Using Lemma 6, we deduce that:

$$\mathcal{H}'(t) \leqslant -\left(2 - \frac{2b(1-\beta)}{2\beta+1}\right) t^{-c} v(t) \leqslant 0. \tag{36}$$

□

**Proof of point 1. of Theorem 1** It follows from Proposition 2 that the function $\mathcal{H}'$ is non increasing and thus bounded. We thus get that for all $t \geqslant t_0$, we have

$$t^2 W(t) \leqslant \mathcal{E}(t) \leqslant \frac{t^c}{t_0^c}\mathcal{E}(t_0) \tag{37}$$

ie for any $t \geqslant t_0$

$$W(t) \leqslant \frac{\mathcal{E}(t_0)}{t_0^{2-\frac{2b}{2\beta+1}}} \frac{1}{t^{\frac{2b}{2\beta+1}}}. \tag{38}$$

As a consequence, we see that point 1. of Theorem 1 holds true.

□

## 3.3 Convergence of the trajectory and proof of point 3. of Theorem 1

We start with some classical energy estimates.

**Proposition 3.** *If $b > 1 + 2\beta$, then we have:*

$$\int_{t_0}^{+\infty} tW(t).,dt < +\infty \tag{39}$$

*and*

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2.,dt < +\infty \tag{40}$$

*Proof.* if $T > t_0$ and $b > 1 + 2\beta$, choosing $\lambda = b - 1$ and $\xi = \lambda(b - \lambda - 1) = 0$, by integrating (25) in $[t_0, T]$, we obtain (notice that we thus have $\lambda > 2\beta$, i.e. $2 - \lambda/\beta < 0$):

$$\left(\frac{\lambda}{\beta} - 2\right) \int_{t_0}^{T} tW(t)dt \leq \mathcal{E}_{b-1,0}(t_0) - \mathcal{E}_{b-1,0}(T) \leq \mathcal{E}_{b-1,0}(t_0) \tag{41}$$

In a similar way if we choose $\lambda = 2\beta$ and $\xi = \lambda(b - \lambda - 1) > 0$, by integrating (25) in $[t_0, T]$, we obtain :

$$(b - 1 - 2\beta) \int_{t_0}^{T} t \|\dot{x}(t)\|^2 dt \leq \mathcal{E}_{2\beta,\xi}(t_0) - \mathcal{E}_{2\beta,\xi}(T) \leq \mathcal{E}_{2\beta,\xi}(t_0) \tag{42}$$

Since (41) and (42) hold for every $T > t_0$, we can conclude.

□

Thanks to the energy estimates of Proposition 3, we can establish the convergence of the trajectory.

**Proposition 4.** *If $b > 1 + 2\beta$, then the trajectory $x(t)$ converges to $x^*$.*

For the proof of Proposition 4 we use the continuous version of Opial's Lemma for which we omit the proof

**Lemma 7.** *Let $S \subset \mathcal{H}$ 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:** We will use the previous Lemma with $S =$argmin $F$. In fact Opial's Lemma holds true for a general separable Hilbert space $\mathcal{H}$, but in our case as $\mathcal{H}$ has a finite dimension, we also deduce strong convergence of $x(t)$ to a point of $S$.

**Proof of Proposition 4** In order to apply Opial's Lemma, we define : $\psi(t) = \frac{1}{2} \|x(t) - x^*\|$.

We have

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

By using (1), we obtain

$$\begin{aligned}\ddot{\psi}(t)+\frac{b}{t}\dot{\psi}(t)&=\|\dot{x}(t)\|^2+\langle\ddot{x}(t)+\frac{b}{t}\dot{x}(t),x(t)-x^*\rangle\\&=\|\dot{x}(t)\|^2-\langle z(t),x(t)-x^*\rangle\\&\leq\|\dot{x}(t)\|^2-W(t)\leq\|\dot{x}(t)\|^2\end{aligned}\tag{43}$$

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

$$t^b\ddot{\psi}(t)+bt^{b-1}\dot{\psi}(t)\leq t^b\|\dot{x}(t)\|^2\tag{44}$$

By integrating over $[t_0,s]$ we find

$$\dot{\psi}(s)\leq\frac{t_0^b\dot{\psi}(t_0)}{s^b}+\frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2dt\leq\frac{C_0}{s^b}+\frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2dt\tag{45}$$

where $C_0$ is a positive constant. If we consider the positive part of $\dot{\psi}$, i.e. $[\dot{\psi}]_+(t)=max\{0,\psi(t)\}\quad\forall t\geq t_0$, we have

$$[\dot{\psi}]_+(s)\leq\frac{C_0}{s^b}+\frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2dt\tag{46}$$

Hence, by applying Fubini's Theorem, for $T>t_0$, we have that :

$$\begin{aligned}\int_{t_0}^T[\dot{\psi}]_+(s)ds&\leq C_0\int_{t_0}^T\frac{1}{s^b}+\int_{t_0}^T\frac{1}{s^b}\int_{t_0}^s t^b\|\dot{x}(t)\|^2dtds\\&=(b-1)C_0(t_0^{1-b}-T^{1-b})+\int_{t_0}^T t^b\|\dot{x}(t)\|^2\Big(\int_t^T s^{-b}ds\Big)dt\\&\leq C_0+(b-1)\int_{t_0}^T t\|\dot{x}(t)\|^2dt\end{aligned}\tag{47}$$

Since, by Proposition 3, for $b>1+2\beta$, the right-hand side of this inequality is finite for every $T>t_0$, we deduce that :

$$\int_{t_0}^{+\infty}[\dot{\psi}]_+(s)ds<+\infty\tag{48}$$

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 $\theta$ 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 :

$$\lim_{t\to\infty}\psi(t)=\theta_\infty+\int_{t_0}^{+\infty}[\dot{\psi}]_+(s)ds\in\mathbb{R}\tag{49}$$

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 :

$$F(\tilde{x}) \leq \liminf_{t\to\infty} F(x(t)) \tag{50}$$

From point 2. of Theorem 1, we have that $\lim_{t\to\infty} F(x(t)) = F(x^*)$, where $x^*$ is a minimizer, so that $\tilde{x} \in \text{argmin } F$, which shows that the second condition of Opial's Lemma is satisfied, therefore we can conclude the proof by applying Opial's Lemma. □

It is now classical to establish that a better convergence rate can be reached [2] .

**Proposition 5.** *Let f $b > 1 + 2\beta$, $x$ a solution of* (1) *and $x^*$ a minimizer of $F$. Then*

$$\lim_{t\to\infty} t^2 W(t) = 0 \quad \text{and} \quad \lim_{t\to\infty} t \left\|\dot{x}(t)\right\| = 0 \tag{51}$$

*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

$$U(t) = t^2 W(t) + \frac{t^2}{2} \left\|\dot{x}(t)\right\|^2 \geq 0 \quad , \forall t \in [t_0, +\infty) \tag{52}$$

Let $T > t_0$. We have

$$\frac{d}{dt} U(t) = t^2 \langle z, \dot{x}(t) \rangle + t^2 \langle \ddot{x}, \dot{x}(t) \rangle + 2tW(t) + t \left\|\dot{x}(t)\right\|^2 \tag{53}$$

By using (1) and $b > 1 + 2\beta$, we find

$$\frac{d}{dt} U(t) = 2tW(t) + (1 - b)t \left\|\dot{x}(t)\right\|^2 \leq 2tW(t) \tag{54}$$

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

$$\left[\frac{d}{dt} U(t)\right]_+ \leq 2tW(t) \tag{55}$$

By Proposition 3 for $b > 1 + 2\beta$ the term $2tW(t)$ is integrable on $[t_0, +\infty)$ , 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, $\Theta$ has negative derivative on $(t_0, +\infty)$; hence it is non-increasing in $(t_0, +\infty)$. This together with the fact that the function $\Theta$ 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 $U(t)$ also converges when $t \to +\infty$ with :

$$\lim_{t\to+\infty} U(t) = \lim_{t\to+\infty} \Theta(t) + \int_{t_0}^{+\infty} \left[\dot{U}(t)\right]_+ dt \in \mathbb{R} \tag{56}$$

Finally since $b > 1 + 2\beta$, by Proposition 3, we have:

$$\int_{t_0}^{+\infty} \frac{1}{t} U(t)\, dt = \int_{t_0}^{+\infty} tW(t)\, dt + \frac{1}{2}\int_{t_0}^{+\infty} t\left\|\dot{x}(t)\right\|^2\, dt < +\infty \tag{57}$$

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

**Proof of point 3. of Theorem 1** Propositions 4 and 5 give the results of point 3. of Theorem 1.

□

# 4 Optimality of the power $\frac{2b}{2\beta+1}$.

The goal of this section is to show that the asymptotic rate of point 1. of Theorem 1 is optimal, in the sense that it can be obtained as a limit case for class of functions $F$.

If $F(x) = |x|^\alpha$ with $\alpha > 1$ we can consider the following ordinary differential equation (ODE):

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \alpha\frac{|x|^\alpha}{x} = 0 \tag{58}$$

Of course, since $\alpha > 1$, we have $\lim_{x\to 0} \frac{|x|^\alpha}{x} = 0$ and the ODE is well defined.

**Notations** As before, we choose $\lambda = \frac{2b\beta}{2\beta+1}$ and $\xi = \lambda(\lambda + 1 - b)$ and we denote $\mathcal{E}_{\lambda,\xi} = \mathcal{E}$. For these parameters one can observe that $\lambda + 1 - b = 1 - \frac{b}{2\beta+1} \geqslant 0$. We recall that

$$c = 2 - \frac{2b}{2\beta + 1} \tag{59}$$

Notice that $c \geqslant 0$. We introduce the new Lyapunov energy:

$$\mathcal{H}_\varepsilon(t) := t^{-c+\varepsilon}\mathcal{E}(t) = t^\varepsilon \mathcal{H}(t). \tag{60}$$

**Proposition 6.** *Let $\beta \in (0,1)$ and $b \leqslant 1+2\beta$, if $x_0 \neq 0$, for any solution $x$ of (58) with $\beta = \frac{1}{\alpha}$,*

1. *if $b = 2\beta+1$, $\mathcal{E}$ is constant,*
2. *if $b \in (0, 2\beta+1)$, the Lyapunov energy satisfies $\mathcal{E}(t) \sim t^{2-\frac{2b}{2\beta+1}}$ when $t \to +\infty$.*
3. *There exists $K > 0$ such that for any $T > 0$, there exists $t > T$ such that*

$$W(t) \geqslant \frac{K}{t^{\frac{2b}{2\beta+1}}} \tag{61}$$

*Remarks :*

- We can observe that for the function $F(x) = |x|^{\frac{1}{\beta}}$, $g(x) = (F(x) - F(x^*))^{\beta}$ is convex and thus we can apply Theorem 1. This Proposition ensures that the decay stated in point 1. of Theorem 1 is actually optimal and is reached for an explicit function.
- For the function $F(x) = |x|^{\frac{1}{\beta}}$ the inequality (10) is actually an equality for any $x \in \mathbb{R}$, ie:

$$W(t) := F(x(t)) - F(x^*) = |x(t)|^{\frac{1}{\beta}} = \beta\langle \nabla F(x(t)), x(t) - x^* \rangle \tag{62}$$

**Proof of Proposition 6** To prove the first point of Proposition 6, using equation (62) we have that for this specific ODE (58) the inequality (25) is actually an equality :

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& (2 - \frac{\lambda}{\beta})tW(t) + (\xi - \lambda(\lambda+1-b))\langle \dot{x}(t), x(t) - x^* \rangle & (63)\\ &+ (\lambda+1-b)\frac{1}{t}\left\|\lambda(x(t)-x^*) + t\dot{x}(t)\right\|^2 - \frac{\lambda^2(\lambda+1-b)}{t}\left\|x(t)-x^*\right\|^2 & (64)\end{aligned}$$

and thus choosing $\lambda = 2\beta$ and $\xi = 0$ this derivative is zero.
To prove the two last points of Proposition 6 we need the following lemmas whose proofs are more technical and are given in Appendix.

**Lemma 8.** *If $b \in (0, 2\beta+1)$ for any $\varepsilon > 0$, there exists $K > 0$ and $T_\varepsilon > t_0$ such that for any $t \geqslant T_\varepsilon$*

$$\mathcal{H}(t) \geqslant \frac{K}{t^{\varepsilon}}. \tag{65}$$

*when $t \to +\infty$.*

**Lemma 9.** *If $b \in (0, 2\beta+1)$ there exists $A > 0$ such that $\mathcal{H}(t) \to A$ when $t \to +\infty$.*

The second point of Proposition 6 is a direct consequence of Lemmas 9, 8, point 1. of Theorem 1, and the definition of $\mathcal{H}$.

To prove the last point of Proposition 6 we observe that in this specific case

$$\mathcal{E}(t) = t^2 W(t) + \frac{1}{2}|\lambda x(t) + t\dot{x}(t)|^2 + \frac{\xi}{2}|x(t)|^2. \tag{66}$$

Let $K_1 > 0$ such that for any $t > T$, $\mathcal{E}(t) \geqslant K_1 t^{2-\frac{2b}{2\beta+1}}$. We can notice here that $2 - \frac{2b}{2\beta+1} \geqslant 0$. There are some cases where the conclusion holds directly.

1. If for $t_2 > T$, $\frac{1}{2}|\lambda x(t_2) + t\dot{x}(t_2)|^2 + \frac{\xi}{2}|x(t_2)|^2 \leqslant \frac{K_1}{2} t_0^{2-\frac{2b}{2\beta+1}}$ we have $t_2^2 |x(t_2)|^{\frac{2}{\beta}} \geqslant K_1 t_2^{2-\frac{2b}{2\beta+1}} - \frac{K_1}{2} t_0^{2-\frac{2b}{2\beta+1}} \geqslant \frac{K_1}{2} t_2^{2-\frac{2b}{2\beta+1}}$ and thus we can conclude.

2. If there exists $t > T$ such that $\dot{x}(t) = 0$, using the fact that $\mathcal{E}(t) = t^2|x(t)|^{\frac{2}{\beta}} + \frac{\lambda^2+\xi}{2}|x(t)|^2$ and the fact that $\lim_{t\to\infty} \frac{\lambda^2+\xi}{2}|x(t)|^2 = 0$, we can conclude using the previous point.

3. If there exists $t > T$ such that $x(t) = 0$, since $\lim_{u\to\infty} |x(u)| = 0$, there exists $t_1 \geqslant t$ such that $\dot{x}(t_1) = 0$ and we can use the previous point.

We now suppose that $x(T) > 0$ and the sign of $\dot{x}$ is constant on $[T, +\infty)$. Since $\lim_{t\to\infty} x(t) = 0$ we deduce that $\forall t \in (T, +\infty)$, $\dot{x}(t) < 0$.
For any $t_1 > T$ we have

$$x(t_1) - x(T) = \int_{u=T}^{t_1} \dot{x}(t)dt \tag{67}$$

Since $x(t_1)$ converges to 0, we deduce that for any $\varepsilon' > 0$, there exists $t_2 > T$ such that $|t_2 \dot{x}(t_2)| \leqslant \varepsilon'$. Hence for any $\varepsilon > 0$, there exists $t_2 > T$ such that $\frac{1}{2}|\lambda x(t_2) + t\dot{x}(t_2)|^2 + \frac{\xi}{2}|x(t_2)|^2 \leqslant \varepsilon$ and we can conclude since we are back to case 1. This conludes the proof of the Proposition. $\square$

# 5 Robustness to noise

We focus in this section on the perturbed ODE :

$$\ddot{x}(t) + \frac{b}{t}\dot{x}(t) + \nabla F(x(t)) = g(t) \tag{68}$$

where $g(t)$ can be seen as a perturbation term.
In [3], Attouch et al. proved that $W(t) = 0(\frac{1}{t^2})$ when $F$ is convex and $b \geqslant 3$ and $\displaystyle\int_{t_1}^{+\infty} t \|g(t)\|_2 \, dt < +\infty$.
Moreover under weaker integrability conditions on $g$ (Theorem 4.1), the authors of [3] proved that $F(x(t))$ tends to 0 when $t$ goes to infinity.

In this section we propose an extension of these results when $b < 2\beta + 1$ :

**Theorem 2.** *Let $x$ be any solution of* (68) *with $b \in (0, 2\beta + 1)$ and* $\displaystyle\int_{t_0}^{+\infty} t^{\frac{b}{2\beta+1}} \|g(t)\| < +\infty$. *Then*

$$W(t) = 0\left(\frac{1}{t^{\frac{2b}{2\beta+1}}}\right) \tag{69}$$

The proof of this last theorem relies on Lyapunov functions.
We define

$$p(t) = \int_{t_0}^{t} u^{\frac{2b}{2\beta+1}-1} \langle g(u), \lambda(x(u) - x^*) + u\dot{x}(u) \rangle du \tag{70}$$

and

$$\tilde{\mathcal{H}}(t) = t^{-c}\mathcal{E}(t) + p(t) \tag{71}$$

where $\mathcal{E} = \mathcal{E}_{\lambda,\xi}$ with $\lambda = \frac{2b\beta}{2\beta+1}$ and $\xi = \lambda(\lambda + 1 - b)$.
We state the following lemma (whose proof is given in Appendix):

**Lemma 10.** $\tilde{\mathcal{H}}'(t) \leqslant 0$

We also need the classical Gronwall-Bellman Lemma (which can be found in [], Lemma 2.1)

**Lemma 11.** *Let $f \in L^1$ and $c \geqslant 0$. Suppose that $w$ is a continue function from $[a, b]$ to $\mathbb{R}$ that satisfies : for all $t \in [a, b]$*

$$\frac{1}{2}w^2(t) \leqslant \frac{1}{2}c^2 + \int_a^t f(s)w(s)ds \tag{72}$$

*Then, for all $t \in [a, b]$*

$$w(t) \leqslant c + \int_a^t f(s)ds. \tag{73}$$

We are now in position to prove Theorem 2.

**Proof of Theorem 2** Thanks to Lemma 10, we have for $t \geqslant t_0$, that $\mathcal{H}(t) \leqslant \mathcal{H}(t_0)$. Denoting $w(t) := \left\| t^{1-\frac{c}{2}} \left(\lambda(x(t)-x^*) + t\dot{x}(t)\right)\right\|_2$ with $c = 2 - \frac{2b}{2\beta+1}$, that is $1 - \frac{c}{2} = \frac{b}{2\beta+1}$, we get

$$w(t) \leqslant \mathcal{H}(t_0) + \int_{t_0}^{t} w(u) \left\| u^{1-\frac{c}{2}} g(u) \right\|_2 du \tag{74}$$

and we apply Lemma 11 to conclude. □

# A Proofs of lemmas

This appendix give the detailed proofs of the techical lemmas used in the paper.

## A.1 Lemmas used in Section 3

**Proof of Lemma 2** We differentiate :

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t^2\langle \nabla F(x(t)), \dot{x}(t)\rangle && (75)\\ &+ \langle \lambda\dot{x}(t) + t\ddot{x}(t) + \dot{x}(t), \lambda(x(t)-x^*) + t\dot{x}(t)\rangle + \xi\langle \dot{x}(t), x(t)-x^*\rangle && (76)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t^2\langle \nabla F(x(t)) + \ddot{x}(t), \dot{x}(t)\rangle && (77)\\ &+ (\lambda+1)t\left\|\dot{x}(t)\right\|^2 + \lambda t\langle \ddot{x}(t), x(t)-x^*\rangle + (\lambda(\lambda+1)+\xi)\langle \dot{x}(t), x(t)-x^*\rangle && (78)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t^2\langle -\frac{b}{t}\dot{x(t)}, \dot{x}(t)\rangle && (79)\\ &+ (\lambda+1)t\left\|\dot{x}(t)\right\|^2 + \lambda t\langle \ddot{x}(t), x(t)-x^*\rangle + (\lambda(\lambda+1)+\xi)\langle \dot{x}(t), x(t)-x^*\rangle && (80)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t(\lambda+1-b)\left\|\dot{x}(t)\right\|^2 && (81)\\ &+ \lambda t\langle \ddot{x}(t), x(t)-x^*\rangle + (\lambda(\lambda+1)+\xi)\langle \dot{x}(t), x(t)-x^*\rangle && (82)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t(\lambda+1-b)\left\|\dot{x}(t)\right\|^2 && (83)\\ &+ \lambda t\langle -\nabla F(x(t)) - \frac{b}{t}\dot{x(t)}, x(t)-x^*\rangle + (\lambda(\lambda+1)+\xi)\langle \dot{x}(t), x(t)-x^*\rangle && (84)\end{aligned}$$

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + t(\lambda+1-b)\left\|\dot{x}(t)\right\|^2 && (85)\\ &+ \lambda t\langle -\nabla F(x(t)), x(t)-x^*\rangle + (\lambda(\lambda+1)-b\lambda+\xi)\langle \dot{x}(t), x(t)-x^*\rangle && (86)\end{aligned}$$

□

**Proof of Lemma 4** We start from the result of Lemma 2. Observing that

$$\frac{1}{t}\left\|\lambda(x(t)-x^*)+t\dot{x}(t)\right\|^2 = t\left\|\dot{x}(t)\right\|^2 + 2\lambda\langle\dot{x}(t),x(t)-x^*\rangle + \frac{\lambda^2}{t}\left\|x(t)-x^*\right\|^2 \quad (87)$$

we can write

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& 2tW(t) + \lambda t\langle -\nabla F(x(t)), x(t)-x^*\rangle && (88)\\ &+ (\xi-\lambda(\lambda+1-b))\langle\dot{x}(t),x(t)-x^*\rangle && (89)\\ &+ (\lambda+1-b)\frac{1}{t}\left\|\lambda(x(t)-x^*)+t\dot{x}(t)\right\|^2 - \frac{\lambda^2(\lambda+1-b)}{t}\left\|x(t)-x^*\right\|^2 && (90)\end{aligned}$$

□

## A.2 Lemmas used in Section 4

**Proof of Lemma 8** Using equation (62) we have that for this specific ODE (58) the inequality (25) is actually an equality :

$$\begin{aligned}\mathcal{E}'_{\lambda,\xi}(t) =& (2-\frac{\lambda}{\beta})tW(t) + (\xi-\lambda(\lambda+1-b))\langle\dot{x}(t),x(t)-x^*\rangle && (91)\\ &+ (\lambda+1-b)\frac{1}{t}\left\|\lambda(x(t)-x^*)+t\dot{x}(t)\right\|^2 - \frac{\lambda^2(\lambda+1-b)}{t}\left\|x(t)-x^*\right\|^2 && (92)\end{aligned}$$

and thus choosing $\lambda = \frac{2b\beta}{2\beta+1}$ and $\xi = \lambda(\lambda+1-b)$ and denoting $c := 2-\frac{2b}{2\beta+1} > 0$ and $d := -(\frac{2b\beta}{2\beta+1})^2(1-\frac{b}{2\beta+1}) > 0$,we have

$$\mathcal{E}'(t) = cu(t) - dv(t). \quad (93)$$

which leads to

$$\mathcal{H}'(t) = t^{-c-1}(t\mathcal{E}'(t) - c\mathcal{E}(t)) = -rt^{-c}v(t) \leqslant 0. \quad (94)$$

where $r := (2-\frac{2b(1-\beta)}{2\beta+1}) > 0$. Hence $\mathcal{H}$ is a non increasing function, and thus there exists $K > 0$ such that

$$\mathcal{H}(t) \leqslant \frac{K}{t^c} \quad (95)$$

Let us now compute the derivative of $\mathcal{H}_\varepsilon(t) = t^\varepsilon\mathcal{H}(t) = t^{-c+\varepsilon}\mathcal{E}(t)$:

$$\mathcal{H}_\varepsilon'(t) = t^\varepsilon \mathcal{H}'(t) + \varepsilon t^{\varepsilon-1}\mathcal{H}(t) \tag{96}$$
$$= -rt^{-c+\varepsilon}v(t) + \varepsilon t^{-c+\varepsilon-1}\mathcal{E}(t) \tag{97}$$

Since $\mathcal{E}(t) = tu(t) + \xi tv(t)$, we deduce that:

$$\mathcal{H}_\varepsilon'(t) = -rt^{-c+\varepsilon}v(t) + \varepsilon t^{-c+\varepsilon-1}(tu(t) + \xi tv(t)) \tag{98}$$

ie:

$$\mathcal{H}_\varepsilon'(t) = t^{-c+\varepsilon}\left(\varepsilon u(t) + v(t)(\xi\varepsilon - r)\right) \tag{99}$$

We now define:

$$T_\varepsilon := \sup_{t\geqslant t_0}\{u(t) \leqslant \delta v(t)\}. \tag{100}$$

with $\delta = \frac{r}{\varepsilon} - \xi$.

If for any $t \geqslant t_0$, $\delta v(t) < u(t)$ we define $T_\varepsilon := t_0$.

We have $u(t) \geq tW(t)$, and $W(t) = F(x(t) - F(x^*) = |x(t)|$. Hence $u(t) \geq t|x(t)|$. Since

$$v(t) = \frac{1}{2t}\|x(t) - x^*\|^2 = \frac{1}{2t}|x(t)|^2 \tag{101}$$

we deduce that for $t > 0$:

$$0 \leq v(t) \leqslant \frac{1}{2t^2}|x(t)|u(t) \tag{102}$$

Since $|x(t)| = W(t)^\beta$ tends to 0, we deduce that $T_\varepsilon < +\infty$.
Using the continuity of $u$ and $v$ and inequality (102) we deduce that $v(T_\varepsilon) > 0$. Hence for $t \geqslant T_\varepsilon$

$$0 \leq v(t) \leqslant \frac{1}{\delta}u(t) \tag{103}$$

which implies that for any $t \geqslant T_\varepsilon$,

$$\mathcal{H}_\varepsilon'(t) = \varepsilon t^{-c+\varepsilon}\left(u(t) - \delta v(t)\right) \geqslant 0 \tag{104}$$

We deduce that for all $t \geqslant T_\varepsilon$

$$t^{-c+\varepsilon}\mathcal{E}(t) \geqslant T_\varepsilon^{-c+\varepsilon}\mathcal{E}(T_\varepsilon) = K > 0 \tag{105}$$

From the definition of $\mathcal{H}$ we deduce that for all $t \geqslant T_\varepsilon$,

$$\mathcal{H}(t) \geqslant \frac{K}{t^\varepsilon}. \tag{106}$$

**Proof of Lemma 9** We proved that

$$\mathcal{H}'(t) = -rt^{-c-1}v(t) \leqslant 0. \tag{107}$$

with $r = (2 - \frac{2b(1-\beta)}{2\beta+1}) > 0$. Since $W(t)$ tends to 0, the function $t \mapsto v(t)$ is bounded by a real number $M$ and we deduce that for all pair $(t_1, t_2) \in \mathbb{R}^2$ such that $t_1 < t_2$

$$\mathcal{H}(t_2) \geqslant \mathcal{H}(t_1) - \frac{Mr}{c}(t_1^{-c} - t_2^{-c}). \tag{108}$$

Let us consider $\varepsilon < c$ , we proved that there exists $K > 0$ and $T_\varepsilon$ such that for all $t \geqslant T_\varepsilon$,

$$\mathcal{H}(t) \geqslant \frac{K}{t^\varepsilon}. \tag{109}$$

We deduce that for all $(t_1, t_2)$ such that $T_\varepsilon \leqslant t_1 \leqslant t_2$

$$\mathcal{H}(t_2) - \frac{Mr}{c}t_2^{-c} \geqslant \frac{K}{t_1^\varepsilon} - \frac{Mr}{c}t_1^{-c} = t_1^{-\varepsilon}\left(K - \frac{Mr}{c}t_1^{-c+\varepsilon}\right) \tag{110}$$

Since $\varepsilon < c$, we can choose $t_1$ such that the right hand side of the inequality is strictly positive which ensures that $H$ has a strictly positive limit and concludes the proof of the Lemma.

□

## A.3 Lemmas used in Section 5

**Proof of Lemma 10** To prove that $\tilde{\mathcal{H}}'(t) \leqslant 0$ in the pertubed case as in the unpertubrted case we differentiate $\mathcal{E}_{\lambda,\xi}$

$$\begin{aligned}\mathcal{E}_{\lambda,\xi}'(t) =& 2tW(t) + t^2\langle \nabla F(x(t)), \dot{x}(t)\rangle & (111)\\ &+ \langle \lambda\dot{x}(t) + t\ddot{x}(t) + \dot{x}(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle + \xi\langle \dot{x}(t), x(t) - x^*\rangle & (112)\end{aligned}$$

Previously, when $g = 0$ we used the fact that $\ddot{x}(t) = -\frac{b}{t}\dot{x}(t) - \nabla F(x(t))$. When $g \neq 0$, we replace $\ddot{x}(t)$ using

$$\ddot{x}(t) = -\frac{b}{t}\dot{x}(t) - \nabla F(x(t)) - g(t). \tag{113}$$

This difference makes that the expression of $\mathcal{E}_{\lambda,\xi}'(t)$ is equal to the one without perturbation up to a term equal to $-\langle tg(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle$. Using the same calculation that the ones of Lemma 2 we get

$$\begin{aligned}\mathcal{E}_{\lambda,\xi}'(t) =& 2tW(t) + t(\lambda + 1 - b)\left\|\dot{x}(t)\right\|^2 - \langle tg(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle & (114)\\ &+ \lambda t\langle -\nabla F(x(t)), x(t) - x^*\rangle + (\lambda(\lambda + 1) - b\lambda + \xi)\langle \dot{x}(t), x(t) - x^*\rangle & (115)\end{aligned}$$

Using the same calculations that in Lemma 4 we have

$$\mathcal{E}'_{\lambda,\xi}(t) = 2tW(t) + \lambda t\langle -\nabla F(x(t)), x(t) - x^*\rangle - \langle tg(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle \tag{116}$$

$$+ (\xi - \lambda(\lambda + 1 - b))\langle \dot{x}(t), x(t) - x^*\rangle \tag{117}$$

$$+ (\lambda + 1 - b)\frac{1}{t}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 - \frac{\lambda^2(\lambda + 1 - b)}{t}\|x(t) - x^*\|^2 \tag{118}$$

It follows that

$$\mathcal{E}'_{\lambda,\xi}(t) \leqslant cu(t) - dv(t) - \langle tg(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle \tag{119}$$

where $u$, $v$ and $c$ and $d$ are defined as in the unperturbed case (see (13) and (14)). Using the definitions $\tilde{\mathcal{H}}$ and $p$ we get the inequality (36)

$$\tilde{\mathcal{H}}'(t) \leqslant -\left(2 - \frac{2b(1-\beta)}{2\beta + 1}\right) t^{-c} v(t) \leqslant 0.$$

which concludes the proof of the lemma.

# References


[1] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$". *Archiv Preprint*, June 2017.

[2] Hedy Attouch and Chbani. Fast inertial dynamics and fista algorithms in convex optimization. perturbation aspects. *arXiv preprint arXiv:1507.01367*, 2015.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

[9] J-P. Demailly. Analyse numérique et équations aux dérivées partielles. *EDP Sciences*, 2006.

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

[11] 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.

[12] 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.