HAL open science

Nesterov's acceleration and Polyak's heavy ball method in continuous time: convergence rate analysis under geometric conditions and perturbations

Othmane Sebbouh, Charles H Dossal, Aude Rondepierre

▶ **To cite this version:**

Othmane Sebbouh, Charles H Dossal, Aude Rondepierre. Nesterov's acceleration and Polyak's heavy ball method in continuous time: convergence rate analysis under geometric conditions and perturbations. 2019. hal-02173978v1

HAL Id: hal-02173978

https://hal.science/hal-02173978v1

Preprint submitted on 4 Jul 2019 (v1), last revised 20 May 2020 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nesterov's acceleration and Polyak's heavy ball method in continuous time: convergence rate analysis under geometric conditions and perturbations

Othmane Sebbouh[1], Charles Dossal[1], Aude Rondepierre[1,2]

[1] IMT, Univ. Toulouse, INSA Toulouse, France.
[3] LAAS, Univ. Toulouse, CNRS, Toulouse, France.

othmane.sebbouh@gmail.com, {Charles.Dossal,Aude.Rondepierre}@insa-toulouse.fr

Abstract

In this article a family of second order ODEs associated to inertial gradient descend is studied. These ODEs are widely used to build trajectories converging to a minimizer x^* of a function F, possibly convex. This family includes the continuous version of the Nesterov inertial scheme and the continuous heavy ball method. Several damping parameters, not necessarily vanishing, and a perturbation term g are thus considered. The damping parameter is linked to the inertia of the associated inertial scheme and the perturbation term g is linked to the error that can be done on the gradient of the function F. This article presents new asymptotic bounds on $F(x(t)) - F(x^*)$ where x is a solution of the ODE, when F is convex and satisfies local geometrical properties such as Łojasiewicz properties and under integrability conditions on g. Even if geometrical properties and perturbations were already studied for most ODEs of these families, it is the first time they are jointly studied. All these results give an insight on the behavior of these inertial and perturbed algorithms if F satisfies some Łojasiewicz properties especially in the setting of stochastic algorithms.

Keywords Lyapunov functions, rate of convergence, ODEs, optimization, Łojasiewicz property.

Contents

1 Introduction **2**

2 Preliminaries: local geometry of convex functions **3**

3 Contributions **4**
3.1 Convergence rates for sharp geometries . . . 4
3.2 Convergence rates for flat geometries . . . 5
3.3 Strategies of proofs . . . 6
3.3.1 State of the art strategies . . . 6
3.3.2 Sketch of proof . . . 7

4 Proofs **7**
4.1 Proof of Theorem 3.2 . . . 8
4.2 Proof of Theorem 3.3 . . . 9
4.3 Proof of Theorem 3.4 . . . 13
4.4 Proof of Theorem 3.5 . . . 15
4.5 Proof of Corollary 3.1 . . . 16

A Appendix **17**
A.1 Proof of Lemma 4.1: differentiating the energy function for Theorem 3.2 . . . 17
A.2 Proof of Lemma 4.2: differentiating the energy function of Theorem 3.3 . . . 18
A.3 Proof of Lemma 4.3: differentiating the energy function of Theorem 3.3 for the classical heavy ball system . . . 19
A.4 Proof of Lemma 4.4: differentiating the energy function of Theorem 3.4 . . . 20

1 Introduction

Let $F : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function admitting at least one minimizer. In this paper we study the asymptotic behavior of the trajectories of the perturbed second-order ordinary differential equation (ODE):

$$\ddot{x}(t) + \beta(t)\dot{x}(t) + \nabla F(x(t)) = g(t), \tag{1.1}$$

where $t_0 > 0$, $\beta(t) = \dfrac{\alpha}{t^\theta}$, with $\alpha > 0$ and $\theta \in [0,1]$, is a viscous damping coefficient and $g : [t_0, +\infty[\to \mathbb{R}$ an integrable source term that can be interpreted as a small external perturbation exerted on the system. Throughout the paper, we assume that, for any initial conditions $(x_0, v_0) \in \mathbb{R}^n \times \mathbb{R}^n$, the Cauchy problem associated with the differential equation (1.1), has a unique global solution satisfying $(x(t_0), \dot{x}(t_0)) = (x_0, v_0)$. This is guaranteed for instance when the gradient function ∇F is Lipschitz on bounded subsets of $\mathbb{R}^n$ [19, 20].

During the last five years many articles study these ODEs, the convergence of the trajectory $x(t)$ or the decay rate of $F(x(t))$ to its minimum value F^*, see for example [8, 25, 23] and reference therein. In [25] Su et al. proved the Nesterov acceleration scheme can be seen as a discretization scheme of the ODE (1.1) with $\theta = 1$ and $g(t) = 0$. Moreover the convergence properties of the solution $x(t)$ of (1.1) are directly linked with the ones of the sequence defined by the Nesterov scheme and the Lyapunov analysis used in both cases to prove the convergence are very similar. As another example, the choice $\theta = 0$ corresponds to the heavy ball damping. It turns out that this family of ODEs is related to inertial optimization algorithms, with various inertia, depending on the choice of the damping function $\frac{\alpha}{t^\theta}$ and including perturbation or error terms defined by g. Many results concerning inertial algorithms have been transposed to the continuous setting such as the convergence of FISTA iterates in [17] by Chambolle et al. which has been transposed by Attouch et al. [6] to the weak convergence of the trajectory of the solution of (1.1) with $\theta = 1$ and $\alpha > 3$. Conversely May in [23] and Attouch et al. in [6] proved that for $\theta = 1$ and $\alpha > 3$, if F is convex, the solution x of (1.1) satisfies $F(x(t)) - F^* = o\left(\frac{1}{t^2}\right)$ and this result has been extended to the sequence generated by FISTA by Attouch et al. in [8]. Consequently, studying (1.1) is also a first step to have a better understanding of general and perturbed inertial schemes to minimize convex functions.

In [15] Cabot et al. consider a general damping term and a vanishing perturbation term $g = 0$. Their study gives decay rates on $F(x(t)) - F^*$ when $\theta \in [0,1)$. If only a convexity assumption is made on F, Su et al [25] proved that if $\theta = 1$ and $\alpha \geqslant 3$, we can get $F(x(t)) - F^* = O\left(\frac{1}{t^2}\right)$. In [6, 3] authors complete these first results for $\theta = 1$, when $\alpha > 3$, proving the weak convergence of trajectory x and showing that $F(x(t)) - F^* = o\left(\frac{1}{t^2}\right)$. In [9, 7] authors give some optimal bound on $F(x(t)) - F^*$ in the subcritical case $\alpha < 3$. More general damping functions $\beta(t)$ have been studied by Cabot et al., Jendoubi et al., Attouch et al. see [16, 20, 3] for complete results. In particular, if F is convex and $\theta \in [0,1)$, we can get: $F(x(t)) - F^* = O\left(\frac{1}{t^{1+\theta}}\right)$.

Several works extend these previous results with a non vanishing perturbation term g proposing some integrability conditions on g, see for example Balti et al. [11] for $\theta \in [0,1)$ and Attouch et al. [6] for $\theta = 1$ and $\alpha \geqslant 3$. In these two settings, the condition on g ensuring the optimal decay rate $F(x(t)) - F^* = O\left(\frac{1}{t^{1+\theta}}\right)$ is the following:

$$\int_{t_0}^{+\infty} t^{\frac{1+\theta}{2}} \|g(t)\| dt < +\infty. \tag{1.2}$$

For $\theta = 1$ and $\alpha < 3$, Attouch et al. and Aujol et al. [7, 9] proved that this condition can be weakened to

$$\int_{t_0}^{+\infty} t^{\frac{\alpha}{3}} \|g(t)\| dt < +\infty \tag{1.3}$$

to ensure that $F(x(t)) - F^* = O\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$. In [3] Attouch et al. for $\theta \in (0,1]$ and in [10] Aujol et al. for $\theta = 1$ proved that these decay rates can be improved if more geometrical properties are known on F when the perturbation term g vanishes. These geometrical properties describe the growth of F around the set of minimizers and are linked with Łojasiewicz properties when F is convex.

The goal of this work is to generalize all the previous works providing accurate rates on $F(x(t)) - F^*$ for any $\alpha > 0$, for any $\theta \in [0,1]$, depending on the geometrical properties of F such as Łojasiewicz properties and integrability conditions on g. To our best knowledge, this is the first work combining geometrical properties on F and integrability on g to provide decays on $F(x(t)) - F^*$. More precisely, we will always consider that F is convex, has a unique minimizer and we always assume integrability conditions on g that ensure the convergence of $F(x(t)) - F^*$ to 0. Consequently, the convergence of the trajectory $(x(t))_{t \geqslant t_0}$ to the unique minimizer is always ensured. That is why, in all theorems, the geometrical assumptions are only made on a neighborhood of the minimizer and are not necessarily global.

The paper is organized as follows. In Section 2, we introduce the geometrical hypotheses we consider on the function F, and their relation with Łojasiewicz property. We then present the contributions of the paper in Section 3: depending on the geometry of the function F and the value of the damping parameters α and θ, we show that combining a flatness condition and a sharpness condition such as the Łojasiewicz property provides new and better convergence rates for the values $F(x(t)) - F^*$. The proofs of the theorems are given in Section 4. Some technical proofs are postponed to Appendix A.

2 Preliminaries: local geometry of convex functions

In this section we recall some definitions and results concerning the local geometry of convex functions around their set of minimizers, see [10] for more details.

Throughout the paper, we assume that the ODE (1.1) is defined in $\mathbb{R}^n$ equipped with the euclidean scalar product $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$. As usual $B(x^*, r)$ denotes the open euclidean ball with center $x^* \in \mathbb{R}^n$ and radius $r > 0$. We now introduce on the one hand a flatness assumption that ensures that the function is not too sharp in the neighborhood of its minimizers, and on the other hand a sharpness assumption ensuring that the magnitude of the gradient is not too low in the neighborhood of the minimizers.

Definition 2.1. *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function with $X^* = \operatorname{argmin} F \neq \emptyset$, and $F^* = \inf f$.*

1. *Let $\gamma \geqslant 1$. The function F satisfies the condition $\mathbf{H_1}(\gamma)$ if, for any minimizer $x^* \in X^*$, there exists $\eta > 0$ such that:*
$$\forall x \in B(x^*, \eta),\ F(x) - F^* \leqslant \frac{1}{\gamma} \langle \nabla F(x), x - x^* \rangle.$$
2. *Let $r \geqslant 1$. The function F satisfies the growth condition $\mathbf{H_2}(r)$ if for any minimizer $x^* \in X^*$, there exist $K_r > 0$ and $\epsilon > 0$ such that:*
$$\forall x \in B(x^*, \epsilon),\ K_r d(x, X^*)^r \leqslant F(x) - F^*.$$

The assumption $\mathbf{H}_1(\gamma)$ has been already used in [15, 25, 9, 10, 1]. Note that any convex differentiable function satisfies $\mathbf{H}_1(1)$ and that any differentiable function such that $(F - F^*)^{\frac{1}{\gamma}}$ is convex for some $\gamma \geqslant 1$, satisfies $\mathbf{H_1}(\gamma)$. More precisely, the hypothesis $\mathbf{H}_1(\gamma)$ can be seen as a flatness condition on the geometry of a convex function around its sets of minimizers [10, Lemma 2.4]: any convex differentiable function F satisfying $\mathbf{H}_1(\gamma)$ for some $\gamma \geqslant 1$, also satisfies: for any minimizer $x^* \in X^*$, there exist $M > 0$ and $\eta > 0$ such that:

$$\forall x \in B(x^*, \eta),\ F(x) - F(x^*) \leq M \|x - x^*\|^\gamma. \tag{2.1}$$

The hypothesis $\mathbf{H}_2(r)$ with $r \geq 1$, is a growth condition on the function F around its set of minimizers (critical points in the non-convex case) ensuring that F is sufficiently sharp (at least as sharp as $x \mapsto \|x - x^*\|^r$) in the neighborhood of X^*. It is also called r-conditioning [18] or Hölderian error bounds [13]. In the convex setting, this growth condition is equivalent to the Łojasiewicz inequality [21, 22], a key tool in the mathematical analysis of continuous and discrete dynamical systems, with exponent $\theta = 1 - \frac{1}{r} \in (0, 1]$:

Definition 2.2. *A differentiable function $F : \mathbb{R}^n \to \mathbb{R}$ is said to have the Łojasiewicz property with exponent $\theta \in [0, 1)$ if, for any critical point x^*, there exist $c > 0$ and $\varepsilon > 0$ such that:*

$$\forall x \in B(x^*, \varepsilon),\ \|\nabla F(x)\| \geqslant c \left(F(x) - F(x^*)\right)^\theta.$$

where: $0^0 = 0$ when $\theta = 0$ by convention.

Typical examples of functions having the Łojasiewicz property are real-analytic functions and C^1 subanalytic functions, or semi-algebraic functions [21, 22]. Strongly convex functions satisfy a global Łojasiewicz property with exponent $\theta = \frac{1}{2}$ [2], or equivalently a global version of the growth condition, namely:

$$\forall x \in \mathbb{R}^n, F(x) - F^* \geqslant \frac{\mu}{2} d(x, X^*)^2,$$

where $\mu > 0$ denotes the parameter of strong convexity. Likewise, convex functions having a strong minimizer in the sense of [4, Section 3.3], also satisfy a global version of $\mathbf{H}_2(2)$. By extension, uniformly convex functions of order $p \geqslant 2$ satisfy the global version of the hypothesis $\mathbf{H}_2(p)$ [18].

Finally, observe that any convex differentiable function F satisfying both hypothesis $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(r)$, has to be at least as flat as $\|x - x^*\|^\gamma$ and as sharp as $\|x - x^*\|^r$ in the neighborhood of its minimizers. More precisely, combining (2.1) and $\mathbf{H}_2(r)$, we have:

Lemma 2.1 ([10, Lemma 2.5])**.** *If a convex differentiable function F satisfies both $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(r)$, with $\gamma, r \geqslant 1$, then necessarily: $r \geqslant \gamma$.*

3 Contributions

In this section, we state convergence rates for the values $F(x(t)) - F^*$ along the trajectory $x(t)$ solution of (1.1), depending on geometrical properties $\mathbf{H}_1$ and $\mathbf{H}_2$ of the function F and on integrability conditions on the perturbation term g. Geometry and perturbations have been studied separately in several papers for this family of ODEs, and the specificity of this work is to study both aspects jointly.

Consider first the case when $\theta = 1$. If F only satisfies $\mathbf{H}_1(\gamma)$ for some $\gamma \geqslant 1$ then for low friction parameter α, we have the following result:

Theorem 3.1 ([9, Theorem 2])**.** *Let $\alpha > 0$ and $t_0 > 0$. Let $x(\cdot)$ be any solution of the ODE (1.1) with $\theta = 1$ and $(x(t_0), \dot{x}(t_0)) = (x_0, v_0)$. Assume that: $\int_{t_0}^{+\infty} t^{\frac{\gamma\alpha}{\gamma+2}} \|g(t)\| dt < +\infty$. If F satisfies $\mathbf{H}_1(\gamma)$ for some $\gamma \geqslant 1$ and if $\alpha \leqslant 1 + \frac{2}{\gamma}$ then:*

$$F(x(t)) - F^* = \mathcal{O}\Big(t^{-\frac{2\gamma\alpha}{\gamma+2}}\Big). \tag{3.1}$$

This result has been first stated and proved in the unpublished report [9, Theorem 2] by Aujol and Dossal in 2017 for convex differentiable functions satisfying $(F - F^*)^{\frac{1}{\gamma}}$ convex.

For large friction parameters α, the sole assumption $\mathbf{H}_1(\gamma)$ on F is not sufficient anymore to obtain a decay faster than $\mathcal{O}\big(\frac{1}{t^2}\big)$ which is the uniform rate that can be achieved for $\alpha \geqslant 3$ [25]. The contribution of this paper is to show that a flatness condition $\mathbf{H}_1$ associated to a sharpness condition such as the Łojasiewicz property provides new and better convergence rates for the values $F(x(t)) - F^*$. We can thus compare these results with classical bounds that can be achieved with geometrical assumptions on F, such as convexity or Łojasiewicz properties without any perturbation term or with results dealing with a non vanishing perturbation term g but with simple assumptions on F.

3.1 Convergence rates for sharp geometries

In this section, we state convergence rates on the values $F(x(t)) - F^*$ along the trajectory $x(t)$, that can be achieved for functions satisfying geometrical hypothesis such as $\mathbf{H}_1(\gamma)$ and/or $\mathbf{H}_2(2)$. The cases $\theta = 1$ and $\theta \in [0, 1)$ are treated separately.

Let us first consider the case when $\theta = 1$ i.e. the ODE:

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla F(x(t)) = g(t). \tag{3.2}$$

Theorem 3.2. *Let $\alpha > 0$ and $t_0 > 0$. Let $x(\cdot)$ be any solution of the ODE (3.2) with $\theta = 1$ and $(x(t_0), \dot{x}(t_0)) = (x_0, v_0)$. Assume that:*

$$\int_{t_0}^{+\infty} t^{\frac{\gamma\alpha}{\gamma+2}} \|g(t)\| dt < +\infty. \tag{3.3}$$

If F satisfies $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(2)$, for some $\gamma \leqslant 2$, if F has a unique minimizer and if $\alpha > 1 + \frac{2}{\gamma}$, then

$$F(x(t)) - F^* = \mathcal{O}\Big(t^{-\frac{2\gamma\alpha}{\gamma+2}}\Big). \tag{3.4}$$

Note that Theorem 3.2 whose proof is detailed in Section 4.1, only applies for $\gamma \leqslant 2$ since according to Lemma 2.1, there exists no function satisfying both $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(2)$ for $\gamma > 2$. Moreover the integrability condition given in Theorem 3.2 generalizes the integrability condition given in [9] to any $\alpha > 0$ under the growth condition $\mathbf{H}_2(2)$ and coincides in the limit case $\alpha = 3$, which was expected. Theorems 3.2 can be seen as an extension of former results with a non vanishing perturbation term g, see [10, Theorem 4.2].

Let $\alpha > 0$ and $\theta \in [0, 1)$. We now consider the heavy ball system with a general friction term:

$$\ddot{x}(t) + \frac{\alpha}{t^\theta}\dot{x}(t) + \nabla F(x(t)) = g(t). \tag{3.5}$$

Theorem 3.3. *Let $\gamma \in [1, 2]$, $m \in \Big(0, \frac{2\gamma}{\gamma+2}\Big)$ and $t_0 > 0$. Note: $\Gamma(t) = \int_{t_0}^{t} \frac{\alpha}{s^\theta} ds$. Let $x(\cdot)$ be any solution of the ODE (3.5) with $\theta \in [0, 1)$ and $(x(t_0), \dot{x}(t_0)) = (x_0, v_0)$. Assume that:*

$$\int_{t_0}^{+\infty} e^{m\Gamma(t)} \|g(t)\| dt < +\infty.$$

If F satisfies $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(2)$, and admits a unique minimizer x^* then:

$$F(x(t)) - F^* = \mathcal{O}\Big(e^{-m\Gamma(t)}\Big), \quad \|x(t) - x^*\|^2 = \mathcal{O}\Big(e^{-m\Gamma(t)}\Big), \quad \|\dot{x}(t)\|^2 = \mathcal{O}\Big(e^{-m\Gamma(t)}\Big).$$

Theorem 3.3 when $\theta \in (0,1)$ can be seen as an extension of [3, Theorem 6.1] to functions with some geometrical properties as $\mathbf{H}_1$ and $\mathbf{H}_2$. Its proof is detailed in Section 4 and is an extension of the proof of [3, Theorem 3.12] to a non-vanishing perturbation term $g \neq 0$. Note that [3] deals only with vanishing damping, that is $\theta > 0$, while Theorem 3.3 deals also with the case $\theta = 0$.

Observe also that Theorem 3.3 also applies in the case $\theta = 1$ and provides convergence rates in $\mathcal{O}(t^{-m\alpha})$ for any $m \in \left(0, \frac{2\gamma}{\gamma+2}\right)$, which is slower but infinitely close to the convergence rate $\mathcal{O}\left(t^{-\frac{2\gamma\alpha}{\gamma+2}}\right)$ provided by Theorem 3.2.

Finally, observe that the integrability conditions given in these three theorems are always stronger than the condition $\int_{t_0}^{+\infty} t^p \|g(t)\| dt < +\infty$, with $p = \min(1, \frac{\alpha}{3})$, given in [7, Theorem 5.1] and ensuring the convergence of the values $F(x(t)) - F^*$ to 0. Consequently the trajectory $x(t)$ actually converges to the unique minimizer of F so that the geometrical assumptions $\mathbf{H}_1$ and $\mathbf{H}_2$ can be used locally that is in the neighborhood of the unique minimizer of F.

3.2 Convergence rates for flat geometries

In this section, we state new convergence rates on the values $F(x(t)) - F^*$ along the trajectory $x(t)$, that can be achieved for functions satisfying geometrical hypothesis such as $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(\gamma)$ for any $\gamma > 2$. The cases $\theta = 1$ and $\theta \in [0,1)$ are treated jointly.

Let us first consider the unperturbed case ($g = 0$). We announce new results on the convergence of the function values along the trajectory $x(t)$ with additional geometrical assumptions, but without perturbations:

Theorem 3.4. *Let $\gamma_1 > 2$, $\gamma_2 \geqslant \gamma_1$. Note $r = \frac{1+\theta}{2}$. Suppose x is a solution to the ODE (1.1) with $g = 0$. If F is coercive and satisfies $\mathbf{H_1}(\gamma_1)$ and $\mathbf{H_2}(\gamma_2)$, and*

1. *if $\theta = 1$ and $\alpha \geqslant \frac{\gamma_1+2}{\gamma_1-2}$ or*
2. *if $\theta < 1$,*

then:

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\frac{2r\gamma_2}{\gamma_2-2}}}\right). \tag{3.6}$$

Note that Theorem 3.4 is a generalization of Theorem 4.3 in [10] available for $\theta = 1$ to any $\theta \in (0,1]$. We now prove that the convergence rates provided by Theorem 3.4 remain valid in the perturbed case ($g \neq 0$):

Theorem 3.5. *Let $\alpha > 0$, $\theta \in [0,1]$ and $t_0 > 0$. Let x be the solution of the ODE (1.1) for given initial conditions $(x(t_0), \dot{x}(t_0)) = (x_0, v_0)$. Let $\gamma_1 > 2$, $\gamma_2 \geqslant \gamma_1$ and $r = \frac{1+\theta}{2}$. Assume that:*

$$\int_{t_0}^{+\infty} t^{\frac{r\gamma_2}{\gamma_2-2}} \|g(t)\| dt < +\infty. \tag{3.7}$$

If F satisfies $\mathbf{H_1}(\gamma_1)$ and $\mathbf{H_2}(\gamma_2)$ and admits a unique minimizer then:

1. *if $\theta = 1$ and $\alpha \geqslant \frac{\gamma_1+2}{\gamma_1-2}$ or*
2. *if $\theta < 1$,*

then we have

$$F(x(t)) - F^* = \mathcal{O}\left(\frac{1}{t^{\frac{2r\gamma_2}{\gamma_2-2}}}\right). \tag{3.8}$$

As in [10] with a non vanishing perturbation term g, if $\gamma_1 = \gamma_2$, we have furthermore the convergence of the trajectory:

Corollary 3.1. *Let $\alpha > 0$, $\theta \in [0,1]$ and $t_0 > 0$. Let x be the solution of the ODE (1.1) for given initial conditions $(x(t_0), \dot{x}(t_0)) = (x_0, v_0)$. Let $\gamma > 2$. Note $r = \frac{1+\theta}{2}$.Assume that:*

$$\int_{t_0}^{+\infty} t^{\frac{r\gamma}{\gamma-2}} \|g(t)\| dt < +\infty.$$

If F satisfies $\mathbf{H}_1(\gamma)$ and $\mathbf{H}_2(\gamma)$ and admits a unique minimizer then:

1. *if* $\theta = 1$ *and* $\alpha \geqslant \frac{\gamma+2}{\gamma-2}$ *or*

2. *if* $\theta < 1$,

then we have:

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^{\frac{r\gamma}{\gamma-2}}}\right). \tag{3.9}$$

Note that in the case of the classical heavy ball ($\theta = 0$), Theorem 3.5 can be seen as an extension of [12, Corollary 5.1] using a different approach: indeed in [12], the authors proved a similar convergence rate under Łojasiewicz properties, but without any convexity assumption on F.

Observe also that to deal with a non-vanishing perturbation term, the uniqueness of the minimizer seems to be crucial despite the fact we can avoid this assumption when $g = 0$, see [10, Theorem 4.5 and Corollary 4.6].

Finally observe that in the integrability condition given in Theorem 3.5, the exponent is $\frac{r\gamma_2}{\gamma_2-2} = \frac{(1+\theta)\gamma_2}{2(\gamma_2-2)}$. Since $\frac{r\gamma_2}{\gamma_2-2} \geqslant r$, the integrability condition (1.2) from [23] is automatically satisfied and ensures that the trajectory $x(t)$ converges to the unique minimizer of F. The geometrical assumptions $\mathbf{H}_1$ and $\mathbf{H}_2$ thus can be used locally. Moreover:

$$\inf_{\gamma_2 \in (2+\infty)} \frac{(1+\theta)\gamma_2}{2(\gamma_2-2)} = \frac{1+\theta}{2},$$

so that we get the same exponent as in the integrability condition in (1.2) which is also what we expected.

3.3 Strategies of proofs

The guideline of the proofs of our results is the same: Lyapunov functions (or energies). In this section we present the state of the art strategies using Lyapunov functions and a sketch of the strategies used in this paper.

Note that in the four Theorems 3.1, 3.2, 3.3 and 3.5 dealing with a non vanishing perturbation term, we modified Lyapunov functions used when $g = 0$, and proposed integrability conditions on g that ensure the same decay that the one that can be achieved with $g = 0$. That is the reason why the Lyapunov functions we use are closed to those that can be found in the literature. Note also that the Grönwall-Bellman Lemma is a key lemma in each proof, but the exact way to deal with a non vanishing perturbation term g is different in each theorem.

3.3.1 State of the art strategies

To prove each decay of $F(x(t)) - F^*$, the main idea is to define a function of t which will be denoted by $\mathcal{E}$, $\mathcal{H}$ or $\mathcal{G}$, which involves the term $F(x(t)) - F^*$ and which is bounded. The choice of the Lyapunov function depends on the ODE (α and θ) and on assumptions on function F (the flatness hypothesis $\mathbf{H}_1$). A simple Lyapunov function to study solutions of (1.1) when F is convex with $g = 0$ is:

$$\mathcal{E}(t) = F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2. \tag{3.10}$$

Indeed, $\mathcal{E}$ is a sum of positive terms and $\mathcal{E}'(t) = -\beta(t)\|\dot{x}(t)\|^2 \leqslant 0$. This simple Lyapunov function ensures that $\mathcal{E}$ is non increasing which implies that $F(x(t))$ is bounded. Many Lyapunov have been proposed to study (3.2) when F is convex, see for example [25, 5, 7, 9] or for more general friction term, see [15, 11]. A simple example to study the specific case of Nesterov damping (i.e. $\theta = 1$) when F is convex and $g = 0$, is:

$$\mathcal{E}(t) = t^2(F(x(t)) - F^*) + \frac{1}{2}\|(\alpha-1)(x(t) - x^*) + t\dot{x}(t)\|_2^2. \tag{3.11}$$

Indeed, a simple calculation shows that:

$$\mathcal{E}'(t) \leqslant -\alpha t\|\dot{x}(t)\|^2 + (3-\alpha)t(F(x(t)) - F^*). \tag{3.12}$$

From this Lyapunov function we deduce that if $\alpha \geqslant 3$, $\mathcal{E}$ is non increasing and thus that

$$F(x(t)) - F^* \leqslant \frac{\mathcal{E}(t_0)}{t^2}. \tag{3.13}$$

In [10], we propose to extend this Lyapunov approach to deal with geometrical properties of F for the Nesterov damping i.e (3.2) with $g = 0$ using Lyapunov functions $\mathcal{H}$:

$$\mathcal{H}(t) = t^p(F(x(t)) - F^*) + R(x(t)), \tag{3.14}$$

where R is non necessarily positive and p depends on properties of the damping parameter α and on the geometric properties of F to get $F(x(t)) - F^* = O\left(\frac{1}{t^p}\right)$. To get such a bound we first bound the energy $\mathcal{H}$ and thus use the Łojasiewicz properties of F ($\mathbf{H}_2$) to deduce the bound on $F(x(t)) - F^*$.

For other choices of function $\beta(t)$, especially when $\theta < 1$ in (1.1), the decay may be faster than polynomial if F satisfies some Łojasiewicz properties, see [24, 3, 13]. A way to prove these faster decay is to build Lyapunov energies satisfying some differential inequalities for suitable function γ :

$$\mathcal{E}'(t) \leqslant -\gamma(t)\mathcal{E}(t) \tag{3.15}$$

which implies that

$$\mathcal{E}(t) \leqslant \mathcal{E}(t_0)e^{-\int_{t_0}^{t}\gamma(s)ds}. \tag{3.16}$$

Moreover, a simple way to deal with the perturbation term g in (1.1) is to add an integral term in the Lyapunov energy $\mathcal{E}$ or $\mathcal{H}$ depending on x and g, see for example [11, 5, 9] and references therein and to use a Grönwall-Bellman lemma to conclude.

3.3.2 Sketch of proof

For each theorem, we define a Lyapunov Energy $\mathcal{G}$ (or $\mathcal{H}$) defined by parameters that are set depending on the hypotheses of each theorem. This function $\mathcal{G}$ also depends on the perturbation term g. The first step of the proof consists in proving that $\mathcal{G}$ is bounded. The second step use a Grönwall-Bellman lemma and hypothesis $\mathbf{H}_2$ to conclude. More precisely, Theorems 3.2 and 3.3 deal with sharp functions, that is functions satisfying $\mathbf{H}_2(2)$, or Łojasiewicz properties with parameter equal to $\frac{1}{2}$. The first one is dedicated to Nesterov damping ($\theta = 1$). The polynomial rate was known in the case where $g = 0$, see [10], and we propose to modify the Lyapunov function that is used by adding an integral term :

$$\mathcal{G}(t) = t^p\left(t^2(F(x(t)) - F^*) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|_2^2 + \frac{\xi}{2}\|x(t) - x^*\|^2\right) + \int_t^T s^\delta\langle\lambda(x(s)-x^*)+s\dot{x}(s), g(s)\rangle ds$$

and prove that $\mathcal{G}$ is bounded using differential inequalities. The Grönwall-Bellman lemma and integrability hypotheses on g are used to conclude.

Theorem 3.3 deals with the case $\theta < 1$. In this case it is known [24, 3, 13] that the decay of $F(x(t)) - F^*$ is faster than polynomial. We propose to use a Lyapunov function similar to the previous one. This time λ and ξ may be functions of t :

$$\mathcal{G}(t) = F(x(t)) - F^* + \frac{1}{2}\|\lambda(t)(x(t) - x^*) + \dot{x}(t)\|_2^2 + \frac{\xi(t)}{2}\|x(t) - x^*\|^2 + \int_t^T s^\delta\langle\lambda(x(s) - x^*) + s\dot{x}(s), g(s)\rangle ds.$$

For suitable choices of λ, ξ and δ we can get some differential inequalities like (3.15) satisfied by $\mathcal{G}$ and conclude using Grönwall-Bellman lemma.

Theorems 3.4 and 3.5 deal with *flat functions* that is functions F satisfying $\mathbf{H}_2(\gamma_2)$ with $\gamma_2 > 2$, i.e Łojasiewicz properties with an exponent greater than $\frac{1}{2}$. Theorem 3.4 focuses on the case $g = 0$ and Theorem 3.5 is the general perturbed case. Both theorems provide results for $\theta \in [0, 1]$ including Heavy Ball ($\theta = 0$) and Nesterov ($\theta = 1$). In both theorems, the bound given on $F(x(t)) - F^*$ is polynomial. For Theorem 3.4, inspired by [10] we define an energy function

$$\mathcal{H}(t) = t^p\left(t^2(F(x(t)) - F^*) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|_2^2 + \frac{\xi(t)}{2}\|x(t) - x^*\|^2\right)$$

for a suitable choice of parameters λ and p, and function ξ. And we prove that it exists $t_1 \geqslant t_0$ such that for any $t \geqslant t_1$, $\mathcal{H}'(t) \leqslant 0$. The function ξ may be negative but using Łojasiewicz properties of F, we can deduce bounds on $F(x(t)) - F^*$. In Theorem 3.5, we consider:

$$\mathcal{G}(t) = t^p\left(t^2(F(x(t)) - F^*) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|_2^2 + \frac{\xi(t)}{2}\|x(t) - x^*\|^2\right) + \int_t^T s^\delta\langle\lambda(x(s)-x^*)+s\dot{x}(s), g(s)\rangle ds$$

and prove that $\mathcal{G}$ is bounded. Combining approaches developed in Theorem 3.4, Grönwall-Bellman lemma and integrability hypotheses on g, we are able to conclude.

4 Proofs

In this section, we detail the proofs of Theorem 3.2, Theorem 3.3, Theorem 3.4, Theorem 3.5 and Corollary 3.1.

4.1 Proof of Theorem 3.2

In this section, we prove Theorem 3.2. For a complete proof of Theorem 3.1, we refer the reader to [9].

Let x^* be a minimizer of F and λ, ξ and T three real numbers. The proof of Theorem 3.2 relies on the following energy:

$$\mathcal{G}(t) = t^p\mathcal{E}(t) + \int_t^T \langle s^{\frac{p}{2}}(\lambda(x(s) - x^*) + s\dot{x}(s)), s^{\frac{p+2}{2}} g(s)\rangle ds \tag{4.1}$$

where the energy $\mathcal{E}$ is defined by:

$$\mathcal{E}(t) = t^2(F(x(t)) - F^*) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t) - x^*\|^2. \tag{4.2}$$

Using the following notations:

$$\begin{aligned} a(t) &= t(F(x(t)) - F^*), \\ b(t) &= \frac{1}{2t}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2, \\ c(t) &= \frac{1}{2t}\|x(t) - x^*\|^2, \end{aligned} \tag{4.3}$$

we then have:

$$\mathcal{E}(t) = t(a(t) + b(t) + \xi c(t)). \tag{4.4}$$

Note that the functions $\mathcal{E}(t)$ and $\mathcal{H}(t) = t^p\mathcal{E}(t)$ denote the same Lyapunov functions as those used in the proof of Theorem 4.1. in [10] in the non perturbed case ($g = 0$). Our proofs are based on the following lemma:

Lemma 4.1. *Let $\gamma \geqslant 1$. If F satisfies the hypothesis $\mathbf{H}_1(\gamma)$ and if $\xi = \lambda(\lambda + 1 - \alpha)$, then*

$$\mathcal{G}'(t) \leqslant t^p\left((2 + p - \gamma\lambda)a(t) + (p + 2\lambda + 2 - 2\alpha)b(t) + \lambda(\lambda + 1 - \alpha)(p - 2\lambda)c(t)\right) \tag{4.5}$$

This lemma whose proof is detailed in Appendix A.1, is the generalization of [10, Lemma 5.1] to the perturbed case: the integral term in (4.1) was chosen to cancel the terms in $\mathcal{G}'(t)$ coming from the perturbation $g(t)$: following the exact same calculation steps as in the proof of [10, Lemma 5.1], we observe that all the terms coming from the perturbation $g(t)$ cancel each other out, so that we obtain the same formula as in the non perturbed version.

Choosing now $p = \frac{2\gamma\alpha}{\gamma+2} - 2$ and $\lambda = \frac{2\alpha}{\gamma+2}$ and thus:

$$\xi = \frac{2\alpha}{(\gamma + 2)^2}(2 + \gamma(1 - \alpha)),$$

we get:

$$\mathcal{G}'(t) \leqslant K_1 t^p c(t) \tag{4.6}$$

where $K_1 = \xi(p - 2\lambda) = \frac{2\xi}{\gamma+2}\left((\gamma - 2)\alpha - (\gamma + 2)\right)$. Since $\alpha > 1 + \frac{2}{\gamma}$ and $\gamma \leqslant 2$, we necessarily have: $\xi < 0$, and thus $K_1 > 0$. Consequently, the energy $\mathcal{E}(t)$ is not a sum of non-negative terms and we cannot conclude that the energy function $\mathcal{G}$ is decreasing. To get the expected estimate on the energy, we need an additional growth condition $\mathbf{H}_2(2)$ to bound the term $\|x(t) - x^*\|^2$ as done in [10].

Using the uniqueness of the minimizer and the fact that F satisfies $\mathbf{H}_2(2)$, there exists $K > 0$ such that:

$$Kt\|x(t) - x^*\|^2 \leqslant t(F(x(t)) - F^*) = a(t), \tag{4.7}$$

hence:

$$c(t) \leqslant \frac{1}{2Kt^2}a(t). \tag{4.8}$$

Since $\xi < 0$ with our choice of parameters, we get:

$$\mathcal{H}(t) \geqslant t^{p+1}(a(t) + \xi c(t)) \geqslant t^{p+1}(1 + \frac{\xi}{2Kt^2})a(t), \tag{4.9}$$

so that there exists $t_1 \geqslant t_0$ such that for all $t \geqslant t_1$, we have:

$$\mathcal{H}(t) \geqslant \frac{1}{2}t^{p+1}a(t) \geqslant 0. \tag{4.10}$$

Now from (4.6), (4.8) and (4.10), we have:

$$\forall t \geqslant t_1,\ \mathcal{G}'(t) \leqslant \frac{K_1}{K}\frac{\mathcal{H}(t)}{t^3}. \tag{4.11}$$

Observe now that: $\mathcal{H}'(t) = \mathcal{G}'(t) + \langle \lambda(x(t) - x^*) + t\dot{x}(t), t^{p+1}g(t)\rangle$ so that we get the following differential inequality on the energy $\mathcal{H}$:

$$\begin{aligned}
\mathcal{H}'(t) &\leqslant \frac{K_1}{K}\frac{\mathcal{H}(t)}{t^3} + \langle t^{\frac{p}{2}}(\lambda(x(t) - x^*) + t\dot{x}(t)), t^{\frac{p+2}{2}}g(t)\rangle \\
&\leqslant \frac{K_1}{K}\frac{\mathcal{H}(t)}{t^3} + t^{\frac{p}{2}}\|\lambda(x(t) - x^*) + t\dot{x}(t)\| t^{\frac{p+2}{2}}\|g(t)\| \\
&= \frac{K_1}{K}\frac{\mathcal{H}(t)}{t^3} + \sqrt{2}(t^{p+1}b(t))^{\frac{1}{2}} t^{\frac{\gamma\alpha}{\gamma+2}}\|g(t)\|
\end{aligned} \tag{4.12}$$

Using again the fact that $c(t) = o(a(t))$ (see (4.8)), there exists $t_2 \geqslant t_1$ such that, for all $t \geqslant t_2$, $a(t) + \xi c(t) \geqslant \frac{1}{2}a(t)$, which implies that:

$$\forall t \geqslant t_2,\ \mathcal{H}(t) = t^{p+1}b(t) + t^{p+1}(a(t) + \xi c(t)) \geqslant \frac{1}{2}t^{p+1}a(t) + t^{p+1}b(t) \geqslant t^{p+1}b(t).$$

Hence:

$$\forall t \geqslant t_2,\ \mathcal{H}'(t) \leqslant \frac{K_1}{K}\frac{\mathcal{H}(t)}{t^3} + \sqrt{2}\mathcal{H}(t)^{\frac{1}{2}} t^{\frac{\gamma\alpha}{\gamma+2}}\|g(t)\|. \tag{4.13}$$

Dividing both sides of the inequality by $2\mathcal{H}(t)^{\frac{1}{2}}$ and integrating between t_2 and t, we get:

$$\forall t \geqslant t_2,\ \mathcal{H}(t)^{\frac{1}{2}} \leqslant \frac{K_1}{2K}\int_{t_2}^{t}\frac{\mathcal{H}(s)^{\frac{1}{2}}}{s^3}ds + \frac{\sqrt{2}}{2}\int_{t_2}^{t} s^{\frac{\gamma\alpha}{\gamma+2}}\|g(s)\|ds. \tag{4.14}$$

Since $\int_{t_2}^{+\infty} s^{\frac{\gamma\alpha}{\gamma+2}}\|g(s)\|ds < +\infty$:

$$\begin{aligned}
\forall t \geqslant t_2,\ \mathcal{H}(t)^{\frac{1}{2}} &\leqslant \frac{K_1}{2K}\int_{t_2}^{t}\frac{\mathcal{H}(s)^{\frac{1}{2}}}{s^3}ds + \frac{\sqrt{2}}{2}\int_{t_2}^{+\infty} s^{\frac{\gamma\alpha}{\gamma+2}}\|g(s)\|ds \\
&\leqslant \beta + \int_{t_1}^{t}\frac{K_1}{2Ks^3}\mathcal{H}(s)^{\frac{1}{2}}ds,
\end{aligned}$$

where $\beta = \frac{\sqrt{2}}{2}\displaystyle\int_{t_2}^{+\infty} s^{\frac{\gamma\alpha}{\gamma+2}}\|g(s)\|ds$. Applying the Grönwall Lemma, we finally get:

$$\forall t \geqslant t_2,\ \mathcal{H}(t) \leqslant \beta^2\ \exp\left(\frac{K_1}{K}\int_{t_2}^{t}\frac{1}{s^3}ds\right) \leqslant \beta^2\ \exp\left(\frac{K_1}{K}\int_{t_2}^{+\infty}\frac{1}{s^3}ds\right) = \beta^2\exp\left(\frac{K_1}{2Kt_2^2}\right) \tag{4.15}$$

In other words, we found a constant $A > 0$ such that for all $t \geqslant t_2$, $\mathcal{H}(t) \leqslant A$. According to (4.10), we conclude that $\frac{1}{2}t^{p+2}(F(x(t)) - F^*) = \frac{1}{2}t^{p+1}a(t)$ is bounded which ends the proof of Theorem 3.2.

4.2 Proof of Theorem 3.3

Let λ, ξ and T three real numbers. Let x^* be a minimizer of F and $x(\cdot)$ any trajectory solution of:

$$\ddot{x}(t) + \beta(t)\dot{x}(t) + \nabla F(x) = g(t),$$

where: $\beta(t) = \frac{\alpha}{t^\theta}$ with $\alpha > 0$ and $\theta \in [0, 1)$. The proof of Theorem 3.3 relies on the following energy:

$$\mathcal{G}(t) = \mathcal{E}(t) + \int_t^T \langle \lambda(x(s) - x^*) + \dot{x}(s)), g(s)\rangle ds \tag{4.16}$$

where:

$$\mathcal{E}(t) = F(x(t)) - F^* + \frac{1}{2}\|\lambda(x(t) - x^*) + \dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t) - x^*\|^2. \tag{4.17}$$

Case $\theta = 0$. Remember that in that case, the friction coefficient is constant: $\forall t,\ \beta(t) = \alpha$. Using the following notations:

$$\begin{aligned} a(t) &= F(x(t)) - F^*,\ b(t) = \frac{1}{2}\|\lambda(x(t) - x^*) + \dot{x}(t)\|^2, \\ c(t) &= \frac{1}{2}\|x(t) - x^*\|^2, \end{aligned} \tag{4.18}$$

the energy $\mathcal{E}(t)$ can be rewritten as:

$$\mathcal{E}(t) = a(t) + b(t) + \xi c(t). \tag{4.19}$$

The proof of Theorem 3.3 when $\theta = 0$ relies on the following differential inequality whose proof in detailed in appendix:

Lemma 4.2. *Let $\gamma \geqslant 1$, $\theta = 0$ and $\lambda \in \mathbb{R}$. If F satisfies the hypothesis $\mathbf{H}_1(\gamma)$ and $\xi = \lambda(\lambda - \alpha)$, then:*

$$\begin{aligned} \forall t \geq t_0,\ \mathcal{E}'(t) &\leqslant -\lambda\gamma a(t) + 2(\lambda - \alpha)b(t) - 2\lambda\xi c(t) + \langle g(t), \dot{x}(t) + \lambda(x(t) - x^*)\rangle \\ &\leqslant -\lambda\gamma\left(a(t) + 2\frac{\xi}{\gamma}c(t)\right) + 2(\lambda - \alpha)b(t) + \langle g(t), \dot{x}(t) + \lambda(x(t) - x^*)\rangle. \end{aligned}$$

The scheme of the rest of the proof is quite standard: we first need to control the terms in $b(t)$ and $c(t)$ in Lemma 4.2 to deduce some differential inequality on the energy $\mathcal{E}$.

Let us choose $\lambda < \alpha$. In that case, the energy $\mathcal{E}$ is not a sum of non-negative terms anymore:

$$\mathcal{E}(t) = a(t) + b(t) + \xi c(t)$$

since: $\xi = \lambda(\lambda - \alpha) < 0$. Using the growth condition $\mathbf{H}_2(2)$ combined by the uniqueness of the minimizer of F, to bound $\|x(t) - x^*\|$, we get the following inequality: there exists $t_1 \geqslant t_0$ such that:

$$\forall t \geqslant t_1,\ a(t) + 2\frac{\xi}{\gamma}c(t) = a(t) - 2\frac{|\xi|}{\gamma}c(t) \geqslant (1 - \frac{|\xi|}{K_2\gamma})a(t) \tag{4.20}$$

From now on, we choose: $\lambda = \frac{\gamma K_2}{2\alpha}$. Observe that the constant K_2 appearing in the growth condition $\mathbf{H}_2(2)$ can be chosen as small as needed to get: $\lambda \leq \frac{2}{\gamma+2}\alpha < \alpha$. With that choice, we then have:

$$|\xi| = \lambda(\alpha - \lambda) \leqslant \alpha\lambda \leqslant \frac{\gamma K_2}{2}, \tag{4.21}$$

and $2(\lambda - \alpha) \leqslant -\lambda\gamma$. It follows from (4.20) that, for all $t \geq t_1$:

$$a(t) + 2\frac{\xi}{\gamma}c(t) \geqslant \frac{1}{2}a(t),$$

and, noticing that $a(t) + b(t) \geqslant \mathcal{E}(t)$, we finally get for all $t \geq t_1$:

$$\begin{aligned} \mathcal{E}'(t) &\leqslant -\frac{\lambda\gamma}{2}(a(t) + 2b(t)) + \langle g(t), \dot{x}(t) + \lambda(x(t) - x^*)\rangle \\ \mathcal{E}'(t) &\leqslant -\frac{\lambda\gamma}{2}\mathcal{E}(t) + \langle g(t), \dot{x}(t) + \lambda(x(t) - x^*)\rangle. \end{aligned} \tag{4.22}$$

or equivalently:

$$\mathcal{G}'(t) \leqslant -\frac{\lambda\gamma}{2}\mathcal{E}(t) \leqslant 0. \tag{4.23}$$

The rest of the proof is quite standard: before integrating the differential inequality (4.22) between t_1 and t, we first need to control the term $\|\dot{x}(t) + \lambda(x(t) - x^*)\|$. To that end, observe that, according to (4.23), the energy $\mathcal{G}$ is non-increasing, hence: $\forall t \geqslant t_1,\ \mathcal{G}(t) \leq \mathcal{G}(t_1)$, i.e.:

$$\begin{aligned} \forall t \geqslant t_1,\ \mathcal{E}(t) &\leqslant \mathcal{E}(t_1) + \int_{t_1}^{t} \langle g(s), \dot{x}(s) + \lambda(x(s) - x^*)\rangle ds \\ &\leqslant \mathcal{E}(t_1) + \int_{t_1}^{t} \|g(s)\|\|\dot{x}(s) + \lambda(x(s) - x^*)\| ds \end{aligned}$$

With our choice of parameters, the energy $\mathcal{E}$ is not a sum of non-negative terms, so that the term $\|\dot{x}(s) + \lambda(x(s) - x^*)\|$ can not be directly controlled by $\mathcal{E}(t)$. But according to the growth condition $\mathbf{H}_2(2)$ and to the uniqueness of the minimizer, we have for all $t \geqslant t_1$: $c(t) \leqslant \frac{1}{2K_2} a(t)$. Using (4.21) and $\gamma \leqslant 2$, we deduce:

$$|\xi| c(t) \leqslant \frac{|\xi|}{2K_2} a(t) \leqslant \frac{\gamma}{4} \leqslant \frac{1}{2},$$

hence: $\forall t \geqslant t_1,\ \mathcal{E}(t) \geqslant \frac{1}{2} a(t) + b(t) \geqslant b(t)$. It follows:

$$\forall t \geqslant t_1,\ b(t) = \|\dot{x}(t) + \lambda(x(t) - x^*)\|^2 \leqslant \mathcal{E}(t_1) + \int_{t_1}^{t} \|g(s)\| \|\dot{x}(s) + \lambda(x(s) - x^*)\| ds$$

Applying the Grönwall-Bellman Lemma [14, Lemma A.5], we obtain:

$$\forall t \geqslant t_1,\ \|\dot{x}(t) + \lambda(x(t) - x^*)\| \leqslant c + \int_{t_1}^{t} \|g(s)\| ds,$$

where: $c = \sqrt{2\mathcal{E}(t_1)}$. Since $\int_{t_1}^{+\infty} \|g(s)\| ds < +\infty$ by assumption, we can conclude that:

$$A = \sup_{t \geqslant t_1} \|\dot{x}(t) + \lambda(x(t) - x^*)\| \leqslant c + \int_{t_1}^{+\infty} \|g(s)\| ds < +\infty.$$

Coming back to (4.22), we obtain the following differential inequality:

$$\forall t \geqslant t_1,\ \mathcal{E}'(t) + \frac{\lambda\gamma}{2} \mathcal{E}(t) \leqslant A \|g(t)\|.$$

Integrating between t_1 and t , we finally obtain:

$$\begin{aligned} \forall t \geqslant t_1,\ e^{\frac{\lambda\gamma}{2} t} \mathcal{E}(t) &\leqslant e^{\frac{\lambda\gamma}{2} t_1} \mathcal{E}(t_1) + \int_{t_1}^{t} e^{\frac{\lambda\gamma}{2} s} \|g(s)\| ds, \\ &\leqslant e^{\frac{\lambda\gamma}{2} t_1} \mathcal{E}(t_1) + \int_{t_1}^{+\infty} e^{\frac{\lambda\gamma}{2} s} \|g(s)\| ds < +\infty. \end{aligned}$$

Hence: $\mathcal{E}(t) = \mathcal{O}\left(e^{-\frac{\lambda\gamma}{2} t}\right)$. Since: $F(x(t)) - F^* = a(t) \leqslant 2\mathcal{E}(t)$ for all $t \geq t_1$, we finally get the expected result.

Case where $\theta \in (0, 1)$. Following the strategy proposed by H. Attouch and A. Cabot in the proof of [3, Theorem 3.12] in the unperturbed case, we choose in time-dependent parameters λ and ξ:

$$\lambda(t) = \frac{2\beta(t)}{\gamma + 2},\ \xi(t) = -\lambda(t)^2,$$

so that the energy $\mathcal{E}$ can be rewritten as:

$$\begin{aligned} \mathcal{E}(t) &= F(x(t)) - F^* + \frac{1}{2} \|\lambda(x(t) - x^*) + \dot{x}(t)\|^2 + \frac{\xi}{2} \|x(t) - x^*\|^2 \\ &= F(x(t)) - F^* + \frac{1}{2} \|\dot{x}(t)\|^2 + \frac{2\beta(t)}{\gamma + 2} \langle x(t) - x^*, \dot{x}(t) \rangle. \end{aligned}$$

The case $\theta = 0$ is excluded in the proof detailed hereafter since, as in [3, Theorem 3.12], we need that: $\lim_{t \to +\infty} \beta(t) = 0$. Extending the proof of [3, Theorem 3.12] to our setting, we obtain the following differential inequality whose proof is detailed in appendix:

Lemma 4.3. *Let $\gamma \geq 1$ and $\beta(t) = \frac{\alpha}{t^\theta}$. If F satisfies the hypothesis $\mathbf{H}_1(\gamma)$, then:*

$$\begin{aligned} \mathcal{E}'(t) + \frac{2\gamma}{\gamma + 2} \beta(t) \mathcal{E}(t) \leq \frac{2}{\gamma + 2} \left(\dot{\beta}(t) + \frac{\gamma - 2}{\gamma + 2} \beta(t)^2 \right) \langle x(t) - x^*, \dot{x}(t) \rangle \\ + \left\langle g(t), \dot{x}(t) + \frac{2}{\gamma + 2} \beta(t) (x(t) - x^*) \right\rangle. \end{aligned}$$

According to Lemma 4.3 and noticing that for all $t > 0$: $\dot{\beta}(t) + \frac{\gamma-2}{\gamma+2}\beta(t)^2 \leqslant 0$, we then obtain:

$$\begin{aligned}\mathcal{G}'(t) &\leqslant -\frac{2\gamma}{\gamma+2}\beta(t)\mathcal{E}(t) + \frac{2}{\gamma+2}\left(\dot{\beta}(t) + \frac{\gamma-2}{\gamma+2}\beta(t)^2\right)\langle x(t)-x^*, \dot{x}(t)\rangle, \\ &\leqslant -\frac{2\gamma}{\gamma+2}\beta(t)\mathcal{E}(t) + \frac{2}{\gamma+2}\beta(t)\left(\frac{2-\gamma}{\gamma+2}\beta(t) - \frac{\dot{\beta}(t)}{\beta(t)}\right)|\langle x(t)-x^*, \dot{x}(t)\rangle|. \end{aligned} \tag{4.24}$$

To prove that the energy $\mathcal{G}$ is non increasing and thus bounded, we have now to control the scalar product $\langle x(t)-x^*, \dot{x}(t)\rangle$. Assuming that F satisfies the growth condition $\mathbf{H}_2(2)$ and admits a unique minimizer, we first have:

$$|\langle x(t)-x^*, \dot{x}(t)\rangle| \leq \frac{1}{2}\|x(t)-x^*\|^2 + \frac{1}{2}\|\dot{x}(t)\|^2 \leqslant C\left(F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2\right) \tag{4.25}$$

where $C = \max(1, \frac{1}{2K_2})$. Since $\lim_{t\to+\infty}\beta(t) = 0$, it follows:

$$\begin{aligned}\mathcal{E}(t) &= F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2 + \frac{2\beta(t)}{\gamma+2}\langle x(t)-x^*, \dot{x}(t)\rangle \\ &= F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2 + o\left(F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2\right),\end{aligned}$$

so that there exists $t_1 \geqslant t_0$ such that for all $t \geqslant t_1$,

$$\mathcal{E}(t) \geqslant \frac{2}{\gamma+2}\left(F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2\right) (\geqslant 0). \tag{4.26}$$

Combining (4.24), (4.25) and (4.26), we then obtain:

$$\mathcal{G}'(t) \leqslant -\frac{2\gamma}{\gamma+2}\left(1 - C\left(\frac{2-\gamma}{\gamma+2}\beta(t) - \frac{\dot{\beta}(t)}{\beta(t)}\right)\right)\beta(t)\mathcal{E}(t).$$

Observe now that by definition, we have: $\beta(t)^2 = o(\beta(t))$ and $\dot{\beta}(t) = o(\beta(t))$. Hence, for any constant $m \in (0, \frac{2\gamma}{\gamma+2})$, there exists $t_2 \geq t_1$ such that for all $t \geqslant t_2$:

$$\mathcal{G}'(t) \leqslant -m\beta(t)\mathcal{E}(t) \leqslant 0, \tag{4.27}$$

or equivalently:

$$\mathcal{E}'(t) + m\beta(t)\mathcal{E}(t) \leqslant \langle g(t), \dot{x}(t) + \frac{2}{\gamma+2}\beta(t)(x(t)-x^*)\rangle. \tag{4.28}$$

The rest of the proof is quite standard: before integrating the differential inequality (4.28) between t_2 and t, we first need to control the term $\|\dot{x}(t) + \frac{2}{\gamma+2}\beta(t)(x(t)-x^*)\|$. To that end, observe that, according to (4.27), the energy $\mathcal{G}$ is non-increasing and that: $\forall t \geqslant t_2,\ \mathcal{G}(t) \leq \mathcal{G}(t_2)$, i.e.:

$$\begin{aligned}\forall t \geqslant t_2,\ \mathcal{E}(t) &\leqslant \mathcal{E}(t_2) + \int_{t_2}^{t}\langle g(s), \dot{x}(s) + \frac{2\beta(s)}{\gamma+2}(x(s)-x^*)\rangle ds \\ &\leqslant \mathcal{E}(t_2) + \int_{t_2}^{t}\|g(s)\|\|\dot{x}(s) + \frac{2\beta(s)}{\gamma+2}(x(s)-x^*)\| ds.\end{aligned}$$

Moreover, with our choice of parameters, the energy $\mathcal{E}(t)$ is not a sum of non negative terms:

$$\mathcal{E}(t) = F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t)-x^*)\|^2 - \frac{2\beta(t)^2}{(\gamma+2)^2}\|x(t)-x^*\|^2, \tag{4.29}$$

so that the term $\|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t)-x^*)\|^2$ can not be directly controlled by the energy $\mathcal{E}(t)$. But, according to the growth condition $\mathbf{H}_2(2)$ combined with the uniqueness of the minimizer, we have:

$$\begin{aligned}F(x(t)) - F^* - \frac{2\beta(t)^2}{(\gamma+2)^2}\|x(t)-x^*\|^2 &\geqslant \left(1 - \frac{2\beta(t)^2}{K_2(\gamma+2)^2}\right)(F(x(t)) - F^*) \\ &\geqslant \frac{1}{2}(F(x(t)) - F^*).\end{aligned}$$

for t large enough, and: $\mathcal{E}(t) \geqslant \frac{1}{2}(F(x(t)) - F^*) + \|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t) - x^*)\|^2$. Hence:

$$\begin{aligned}\frac{1}{2}\|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t) - x^*)\|^2 &\leqslant \mathcal{E}(t) \leqslant \mathcal{E}(t_2) + \int_{t_2}^{t} \langle g(s), \dot{x}(s) + \frac{2\beta(s)}{\gamma+2}(x(s) - x^*)\rangle ds \\ &\leqslant \frac{1}{2}c^2 + \int_{t_2}^{t} \|g(s)\|\|\dot{x}(s) + \frac{2\beta(s)}{\gamma+2}(x(s) - x^*)\| ds\end{aligned}$$

where: $c = \sqrt{2\mathcal{E}(t_2)}$. Applying the Grönwall-Bellman Lemma [14, Lemma A.5], we obtain:

$$\|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t) - x^*)\| \leq c + \int_{t_2}^{t} \|g(s)\| ds.$$

Assuming that $\int_{t_2}^{+\infty} \|g(s)\| ds < +\infty$, we can so conclude that:

$$A = \sup_{t \geqslant t_2} \|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t) - x^*)\| \leqslant c + \int_{t_2}^{+\infty} \|g(s)\| ds < +\infty.$$

Coming back to (4.28), we obtain the following differential inequality:

$$\begin{aligned}\forall t \geqslant t_2,\ \mathcal{E}'(t) + m\beta(t)\mathcal{E}(t) &\leqslant \|g(t)\|\|\dot{x}(t) + \frac{2\beta(t)}{\gamma+2}(x(t) - x^*)\| \\ &\leqslant A\|g(t)\|.\end{aligned}$$

Integrating between t_2 and t and stating: $\Gamma(t) = \displaystyle\int_{t_2}^{t} \beta(s) ds$, we finally obtain: for all $t \geqslant t_2$,

$$\begin{aligned}e^{m\Gamma(t)}\mathcal{E}(t) &\leqslant e^{m\Gamma(t_2)}\mathcal{E}(t_2) + A\int_{t_2}^{t} e^{m\Gamma(s)}\|g(s)\| ds, \\ &\leqslant e^{m\Gamma(t_2)}\mathcal{E}(t_2) + A\int_{t_2}^{+\infty} e^{m\Gamma(s)}\|g(s)\| ds = B < +\infty.\end{aligned}$$

Hence: $\mathcal{E}(t) = \mathcal{O}\big(e^{-m\Gamma(t)}\big)$. Since: $F(x(t)) - F^* = a(t) \leqslant 2\mathcal{E}(t)$ for all $t \geq t_2$, we finally get the expected result and using $\mathbf{H}_2(2)$, the other estimates follow directly.

4.3 Proof of Theorem 3.4

The proof of Theorem 3.4 relies on almost the same energy as that used in [10]:

$$\mathcal{E}(t) = t^2(F(x(t)) - F^*) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - x^*\|^2 \tag{4.30}$$

where λ is a non-negative real constant as in [10] and $\xi(.)$ is here a real-valued function. Noting:

$$\begin{aligned}a(t) &= t(F(x(t)) - F^*) \\ b(t) &= \frac{1}{2t}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 \\ c(t) &= \frac{1}{2t}\|x(t) - x^*\|^2\end{aligned} \tag{4.31}$$

we have: $\mathcal{E}(t) = t(a(t) + b(t) + \xi(t)c(t))$. We also define:

$$\mathcal{H}(t) = t^p\mathcal{E}(t) \tag{4.32}$$

The proofs of our theorems rely on the following lemma whose proof is detailed in appendix A.4:

Lemma 4.4. *Let $\gamma_1 \geqslant 1$. If F satisfies the hypothesis $\mathbf{H}_1(\gamma_1)$ and if $\xi(t) = \lambda(\lambda + 1 - \alpha t^{1-\theta})$, then:*

$$\mathcal{H}'(t) \leqslant t^p\Big((2 + p - \gamma_1\lambda)a(t) + (2\lambda + 2 + p - 2\alpha t^{1-\theta})b(t) + \lambda((\lambda+1)(p - 2\lambda) - \alpha(p + 1 - \theta - 2\lambda)t^{1-\theta})c(t)\Big) \tag{4.33}$$

Note $r = \frac{1+\theta}{2}$. Taking $\lambda = \frac{2r}{\gamma_1 - 2}$ and $p = p_1 + 2(r-1)$ with $p_1 = \frac{4r}{\gamma_1 - 2}$, we obtain:

$$\mathcal{H}'(t) \leqslant t^p \left(2\Big(\frac{\gamma_1 + 2}{\gamma_1 - 2} r - \alpha t^{-2(r-1)}\Big) b(t) + 2\lambda(\lambda+1)(r-1)c(t) \right) \tag{4.34}$$

Since $r \leqslant 1$, $2\lambda(\lambda+1)(r-1) \leqslant 0$, hence:

$$\mathcal{H}'(t) \leqslant 2t^{p_1} \Big(\frac{\gamma_1 + 2}{\gamma_1 - 2} r t^{2(r-1)} - \alpha\Big) b(t) \tag{4.35}$$

1. If $\theta < 1$, it exists t_1 depending only on γ_1, α and θ such that $\mathcal{H}'(t) \leqslant 0$ for all $t \geqslant t_1$.
2. If $\theta = 1$ and $\alpha \geqslant \frac{\gamma_1 + 2}{\gamma_1 - 2}$, $\mathcal{H}'$ is non positive for all $t \geqslant t_0$.

We will now use similar reasoning as in [10] to prove the results of our theorem. Since $\mathcal{H}'(t) \leqslant 0$, for any choice of x^* in the set of minimizers X^*, the function $\mathcal{H}$ is bounded above and since the set of minimizers is bounded because F is coercive, there exists $A > 0$ and t_0 such that for all choices of x^* in X^*:

$$\mathcal{H}(t_0) \leqslant A \tag{4.36}$$

Hence for all $x^* \in X^*$ and $t \geqslant t_0$, $\mathcal{H}(t) \leqslant A$. Hence

$$t^{\frac{2r\gamma_1}{\gamma_1 - 2}} (F(x(t)) - F^*) \leqslant \frac{|\xi(t)|}{2} t^{\frac{4r}{\gamma_1 - 2} + 2(r-1)} \|x(t) - x^*\|^2 + A. \tag{4.37}$$

We have $\xi(t) = \frac{2r}{\gamma_1 - 2}(\frac{2r}{\gamma_1 - 2} + 1 - \alpha t^{-2(r-1)})$. Hence:

- If $\theta = 1$, then $r = 1$. Since $\alpha \geqslant \frac{\gamma_1 + 2}{\gamma_1 - 2} \geqslant \frac{\gamma_1}{\gamma_1 - 2}$, we have:

$$|\xi(t)| = \frac{2}{\gamma_1 - 2}(\alpha - \frac{\gamma_1}{\gamma_1 - 2}) \leqslant \frac{2\alpha}{\gamma_1 - 2}.$$

- If $\theta \in (0, 1)$, then:

$$|\xi(t)| t^{2(r-1)} = \frac{2r}{\gamma_1 - 2}(\alpha - (\frac{2r}{\gamma_1 - 2} + 1) t^{2(r-1)} \leqslant \frac{2r\alpha}{\gamma_1 - 2}.$$

Hence, in both cases:

$$|\xi(t)| t^{2(r-1)} \leqslant \frac{2r\alpha}{\gamma_1 - 2}. \tag{4.38}$$

Therefore:

$$t^{\frac{2r\gamma_1}{\gamma_1 - 2}} (F(x(t)) - F^*) \leqslant \frac{r\alpha}{\gamma_1 - 2} t^{\frac{4r}{\gamma_1 - 2}} \|x(t) - x^*\|^2 + A \tag{4.39}$$

And since this is verified for all $x^* \in X^*$:

$$t^{\frac{2r\gamma_1}{\gamma_1 - 2}} (F(x(t)) - F^*) \leqslant \frac{r\alpha}{\gamma_1 - 2} t^{\frac{4r}{\gamma_1 - 2}} d(x(t), X^*)^2 + A \tag{4.40}$$

We set $v(t) = t^{\frac{4r}{\gamma_2 - 2}} d(x(t), X^*)^2$ Then

$$t^{\frac{2r\gamma_1}{\gamma_1 - 2}} (F(x(t)) - F^*) \leqslant \frac{r\alpha}{\gamma_1 - 2} t^{\frac{4r}{\gamma_1 - 2} - \frac{4r}{\gamma_2 - 2}} v(t) + A \tag{4.41}$$

Since F satisfies $\mathbf{H}_2(\gamma_2)$, there exists K > 0 such that

$$K(t^{-\frac{4r}{\gamma_2 - 2}} v(t))^{\frac{\gamma_2}{2}} \leqslant F(x(t)) - F^* \tag{4.42}$$

i.e

$$K v(t)^{\frac{\gamma_2}{2}} t^{\frac{-2r\gamma_2}{\gamma_2 - 2}} \leqslant F(x(t)) - F^* \tag{4.43}$$

Hence

$$K t^{\frac{2r\gamma_1}{\gamma_1 - 2}} t^{-\frac{2r\gamma_2}{\gamma_2 - 2}} v(t)^{\frac{\gamma_2}{2}} \leqslant t^{\frac{2r\gamma_1}{\gamma_1 - 2}} (F(x(t)) - F^*) \tag{4.44}$$

Back to (4.41), this yields:

$$Kt^{\frac{2r\gamma_1}{\gamma_1-2}}t^{-\frac{2r\gamma_2}{\gamma_2-2}}v(t)^{\frac{\gamma_2}{2}} \leqslant \frac{r\alpha}{\gamma_1-2}t^{\frac{4r}{\gamma_1-2}-\frac{4r}{\gamma_2-2}}v(t)+A \tag{4.45}$$

Hence

$$Kv(t)^{\frac{\gamma_2}{2}} \leqslant \frac{rc}{\gamma-2}v(t)+At^{\frac{4r}{\gamma_2-2}-\frac{4r}{\gamma_1-2}} \tag{4.46}$$

Which, since $\gamma_1 \leqslant \gamma_2$, means that v is bounded. Therefore, from (4.41) we deduce that there exists $B>0$ such that:

$$F(x(t))-F^* \leqslant Bt^{\frac{-2r\gamma_2}{\gamma_2-2}}+At^{\frac{-2r\gamma_1}{\gamma_1-2}} \tag{4.47}$$

Since $\gamma_1 \leqslant \gamma_2$, we have $\frac{-r2\gamma_2}{\gamma_2-2} \geqslant \frac{-2r\gamma_1}{\gamma_1-2}$. Hence $F(x(t))-F^* = O(t^{-\frac{2r\gamma_2}{\gamma_2-2}})$.

4.4 Proof of Theorem 3.5

The proof is inspired by the one of **Theorem** 3.4 where an additional term including the noise is considered.

First, we set parameters $r=\frac{1+\theta}{2}$, $\lambda=\frac{2r}{\gamma_1-2}$ and $p=p_1+2(r-1)$, with $p_1=2\lambda$ and functions ξ, $\mathcal{E}$ and $\mathcal{H}$ exactly as in **Theorem** 3.4. In addition to the energy functions $\mathcal{E}$ and $\mathcal{H}$ we define

$$\mathcal{G}(t)=\mathcal{H}(t)+\int_t^T \langle \lambda(x(s)-x^*)+s\dot{x}(s), s^{p+1}g(s)\rangle ds \tag{4.48}$$

We refer the reader to Appendix (A.4) for a detailed calculation of the following bound on the derivative $\mathcal{G}'$

$$\mathcal{G}'(t) \leqslant t^p\Big((2+p-\gamma_1\lambda)a(t)+(2\lambda+2+p-2\alpha t^{1-\theta})b(t)+\lambda((\lambda+1)(p-2\lambda)-\alpha(p+1-\theta-2\lambda)t^{1-\theta})c(t)\Big) \tag{4.49}$$

Actually, the integral term $\mathcal{G}(t)-\mathcal{H}(t)$ is computed such that the bound on $\mathcal{G}'(t)$ is equal to the bound (4.33) proposed in **Theorem** 3.4 for $\mathcal{H}'$. Thus we deduce once again that under the hypotheses of the Theorem 3.5, it exists $t_1 \geqslant 1$ such that the function $\mathcal{G}$ is non increasing for $t \geqslant t_1$.

We will need now the following direct lemma

Lemma 4.5. *If F satisfies the growth condition $\mathbf{H}_2(\gamma_2)$ with $\gamma_2>2$:*

$$K\|x-x^*\|^{\gamma_2} \leqslant F(x)-F^* \tag{4.50}$$

then defining $p_2=\frac{4r}{\gamma_2-2}$, with notations defined in (4.3) we have

$$t^{p_2+1}c(t) \leqslant \frac{K^{-\frac{2}{\gamma_2}}}{2}\left(t^{p_2+2r-1}a(t)\right)^{\frac{2}{\gamma_2}} \tag{4.51}$$

It follows that for any $m\in\mathbb{R}$, it exists $M\in\mathbb{R}$ such that for any $t\geqslant t_0$

$$mt^{p_2+1}c(t)-t^{p_2+2r-1}a(t) \leqslant M. \tag{4.52}$$

Since for all $t\geqslant t_1$, $\mathcal{G}(t)\leqslant \mathcal{G}(t_1)$. Then:

$$\mathcal{H}(t) \leqslant \mathcal{H}(t_1)+\int_{t_1}^t \langle(\lambda(x(s)-x^*)+s\dot{x}(s)), s^{p+1}g(s)\rangle ds. \tag{4.53}$$

Hence, we have:

$$t^{p+1}a(t)+t^{p+1}b(t) \leqslant \mathcal{H}(t_1)+|\xi(t)|t^{p+1}c(t)+\int_{t_1}^t \langle \lambda(x(s)-x^*)+s\dot{x}(s), s^{p+1}g(s)\rangle ds. \tag{4.54}$$

Using the fact that the $|\xi(t)|t^{2(r-1)}$ is uniformly bounded, see (4.38) we get

$$t^{p+1}a(t)+t^{p+1}b(t) \leqslant \mathcal{H}(t_1)+\frac{2r\alpha}{\gamma_1-2}t^{p_1+1}c(t)+\int_{t_1}^t \|\lambda(x(s)-x^*)+s\dot{x}(s)\|_2\|s^{p+1}g(s)\|_2 ds \tag{4.55}$$

Let's define $p_2 = \frac{4r}{\gamma_2 - 2}$. Since $\gamma_1 \leqslant \gamma_2$ we have $p_2 \leqslant p_1$ and for any $t \geqslant t_1$, $t^{p_1 - p_2} \geqslant 1$. Dividing the previous inequality by $t^{p_1-p_2}$ we get for any $t \geqslant t_1$

$$t^{p_2+2r-1}(a(t)+b(t)) \leqslant \mathcal{H}(t_1) + \frac{2r\alpha}{\gamma_1 - 2} t^{p_2+1} c(t) + t^{p_2-p_1} \int_{t_1}^{t} \|\lambda(x(s)-x^*) + s\dot{x}(s)\|_2 \|s^{p+1} g(s)\|_2 ds \tag{4.56}$$

which implies

$$t^{p_2+2r-1}(a(t)+b(t)) \leqslant \mathcal{H}(t_1) + \frac{2r\alpha}{\gamma_1 - 2} t^{p_2+1} c(t) + \int_{t_1}^{t} \|\lambda(x(s)-x^*) + s\dot{x}(s)\|_2 \|s^{p_2+2r-1} g(s)\|_2 ds \tag{4.57}$$

Since F satisfies the growth condition $\mathbf{H}(\gamma_2)$ we can apply Lemma 4.5 and deduce it exists $M \in \mathbb{R}$ such that for $t \geqslant t_1$

$$t^{p_2+2r-1} b(t) \leqslant M + \int_{t_1}^{t} \|(\lambda(x(s)-x^*) + s\dot{x}(s)\|_2 \|s^{p_2+2r-1} g(s)\|\rangle ds \tag{4.58}$$

which implies using the definition of $b(t)$ given in (4.3)

$$\frac{1}{2} t^{p_2+2(r-1)} \|\lambda(x(t)-x^*) + t\dot{x}(t)\|_2^2 \leqslant M + \int_{t_1}^{t} \|s^{\frac{p_2}{2}+(r-1)}(\lambda(x(s)-x^*) + s\dot{x}(s))\|_2 \|s^{\frac{p_2}{2}+r} g(s)\|_2 ds \tag{4.59}$$

Applying the Grönwall Bellman Lemma it follows that

$$t^{\frac{p_2}{2}+r-1} \|\lambda(x(t)-x^*) + t\dot{x}(t)\|_2 \leqslant \sqrt{2M} + \int_{t_1}^{t} s^{\frac{p_2}{2}+r} \|g(s)\|_2 ds \tag{4.60}$$

Under the hypotheses of the Theorem, the right member of the inequality is uniformly bounded relatively to t. It follows that it exists $M_1 \geqslant 0$ such that for any $t \geqslant t_1$

$$\|\lambda(x(t)-x^*) + t\dot{x}(t)\|_2 \leqslant M_1 t^{-\frac{p_2}{2}-r+1} \tag{4.61}$$

and thus that it exists $M_2 > 0$ such that for any $t \geqslant t_1$

$$\int_{t_1}^{t} |\langle(\lambda(x(s)-x^*) + s\dot{x}(s)), s^{p_2+2r-1} g(s)\rangle| ds \leqslant M_1 \int_{t_1}^{t} s^{\frac{p_2}{2}+r} \|g(s)\| ds \leq M_2 \tag{4.62}$$

Combining this inequality with (4.57) if follows that for any $t \geqslant t_1$

$$t^{p_2+2r-1} a(t) \leqslant \mathcal{H}(t_1) + \frac{2r\alpha}{\gamma_1 - 2} t^{p_2+1} c(t) + M_2 \tag{4.63}$$

Using once again Lemma 4.5 we deduce it exists M_3 and M_4 such that for any $t \geqslant t_1$

$$t^{p_2+2r-1} a(t) \leqslant M_3 + M_4 (t^{(p_2+2r-1)} a(t))^{\frac{\gamma_2}{2}} \tag{4.64}$$

which implies that it exists M_5 such that any $t \geqslant 1$

$$a(t) \leqslant M_5 t^{-(p_2+2r-1)} \tag{4.65}$$

that is

$$F(x(t)) - F^* \leqslant M_5 t^{-(p_2+2r)} = M_5 t^{-\frac{2r\gamma_2}{\gamma_2-2}} \tag{4.66}$$

Which is the desired result.

4.5 Proof of Corollary 3.1

In this paragraph we detail the proof of Corollary 3.1 in which it is stated that, in the flat case, the trajectory of any solution x of the ODE (3.5) is finite.

From the proof of Theorem 3.5 (see (4.61)), there exists $A_1 > 0$ and $t_1 \geqslant t_0$ such that, for all $t \geqslant t_1$:

$$\|\lambda(x(t)-x^*) + t\dot{x}(t)\| \leqslant A_1 t^{-\frac{r\gamma}{\gamma-2}+1} \tag{4.67}$$

Combining the growth condition $\mathbf{H}_2(\gamma)$ and the conclusion of Theorem 3.5, we have that there exists $t_2 \geqslant t_1$ such that:

$$\begin{aligned} \forall t \geqslant t_2,\ \|x(t) - x^*\| &\leqslant K^{-\frac{1}{\gamma}}(F(x(t)) - F^*)^{\frac{1}{\gamma}} \\ &\leqslant A_2 t^{-\frac{2r}{\gamma-2}} \end{aligned}$$

It follows that for all $t \geqslant t_2$:

$$\begin{aligned} t\|\dot{x}(t)\| &\leqslant \lambda\|x(t) - x^*\| + \|\lambda(x(t) - x^*) + t\dot{x}(t)\| \\ &\leqslant A_1 t^{-\frac{r\gamma}{\gamma-2}+1} + \lambda A_2 t^{-\frac{2r}{\gamma-2}} \\ &\leqslant t^{-\frac{r\gamma}{\gamma-2}+1}(A_1 + \lambda A_2 t^{r-1}) \end{aligned}$$

Hence:

$$\|\dot{x}(t)\| \leqslant t^{-\frac{r\gamma}{\gamma-2}}(A_1 + \lambda A_2 t^{r-1}).$$

Noticing that $r - 1 \leqslant 0$, we finally get:

$$\|\dot{x}(t)\| = \mathcal{O}\left(t^{-\frac{r\gamma}{\gamma-2}}\right).$$

Which means that $\|\dot{x}(t)\|$ is integrable and that the trajectory is finite.

Acknowledgement

This study has been carried out with financial support from the French state, managed by the French National Research Agency (ANR GOTMI - ANR-16-VCE33-0010-01) and partially supported by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02.

A Appendix

A.1 Proof of Lemma 4.1: differentiating the energy function for Theorem 3.2

1. **Differentiating $\mathcal{E}$.**

   Consider the energy $\mathcal{E}$ defined as follows by:

$$\begin{aligned} \mathcal{E}(t) &= t^2(F(x(t)) - F^*) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t) - x^*\|^2 \\ &= t(a(t) + b(t) + c(t)) \end{aligned}$$

   where:

$$a(t) = t(F(x(t)) - F^*),\ b(t) = \frac{1}{2t}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2,\ c(t) = \frac{1}{2t}\|x(t) - x^*\|^2. \tag{A.1}$$

   We then have:

$$\begin{aligned} \mathcal{E}'(t) = 2t(F(x(t)) - F^*) + t^2\langle\nabla F(x(t)), \dot{x}(t)\rangle + \xi\langle\dot{x}(t), x(t) - x^*\rangle \\ + \langle\lambda(x(t) - x^*) + t\dot{x}(t), (\lambda+1)\dot{x}(t) + t\ddot{x}(t)\rangle. \end{aligned} \tag{A.2}$$

   Since x is a solution of the ODE (3.2), we have:

$$\begin{aligned} (\lambda+1)\dot{x}(t) + t\ddot{x}(t) &= (\lambda+1)\dot{x}(t) - \alpha\dot{x}(t) - t\nabla F(x(t)) + tg(t) \\ &= (\lambda+1-\alpha)\dot{x}(t) - t\ \nabla F(x(t)) + tg(t). \end{aligned} \tag{A.3}$$

   Hence:

$$\begin{aligned} \mathcal{E}'(t) = 2a(t) - \lambda t\langle\nabla F(x(t)), x(t) - x^*\rangle + (\xi + \lambda(\lambda+1-\alpha))\langle\dot{x}(t), x(t) - x^*\rangle + t(\lambda+1-\alpha)\|\dot{x}(t)\|^2 \\ + \langle tg(t), \lambda(x(t) - x^*) + t\dot{x}(t)\rangle. \end{aligned} \tag{A.4}$$

   Noticing that:

$$\frac{1}{t}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 = t\|\dot{x}(t)\|^2 + 2\lambda\langle\dot{x}(t), x(t) - x^*\rangle + \frac{\lambda^2}{t}\|x(t) - x^*\|^2, \tag{A.5}$$

we deduce that

$$\begin{aligned}
\mathcal{E}'(t) &= 2a(t) - \lambda t\langle \nabla F(x(t)), x(t)-x^*\rangle + (\xi - \lambda(\lambda+1-\alpha))\langle \dot{x}(t), x(t)-x^*\rangle \\
&\quad + \frac{\lambda+1-\alpha}{t}\|\lambda(x(t)-x^*)+t\dot{x}(t)\|^2 - \frac{\lambda^2(\lambda+1-\alpha)}{t}\|x(t)-x^*\|^2 \\
&\quad + \langle tg(t), \lambda(x(t)-x^*)+t\dot{x}(t)\rangle \\
&= 2a(t) - \lambda t\langle \nabla F(x(t)), x(t)-x^*\rangle + (\xi - \lambda(\lambda+1-\alpha))\langle \dot{x}(t), x(t)-x^*\rangle \\
&\quad + 2(\lambda+1-\alpha)b(t) - 2\lambda^2(\lambda+1-\alpha)c(t) + \langle tg(t), \lambda(x(t)-x^*)+t\dot{x}(t)\rangle.
\end{aligned}$$

Choosing now $\underline{\xi = \lambda(\lambda+1-\alpha)}$, we get:

$$\begin{aligned}
\mathcal{E}'(t) = 2a(t) - \lambda t\langle \nabla F(x(t)), x(t)-x^*\rangle + 2(\lambda+1-\alpha)b(t) - 2\lambda^2(\lambda+1-\alpha)c(t) \\
+ \langle tg(t), \lambda(x(t)-x^*)+t\dot{x}(t)\rangle
\end{aligned}$$

Since F satisfies $\mathbf{H_1}(\gamma)$:

$$\mathcal{E}'(t) \leqslant (2-\lambda\gamma)a(t) + 2(\lambda+1-\alpha)b(t) - 2\lambda^2(\lambda+1-\alpha)c(t) + \langle tg(t), \lambda(x(t)-x^*)+t\dot{x}(t)\rangle \tag{A.6}$$

2. **Differentiating $\mathcal{H}$ and $\mathcal{G}$**

   Recall now that:

$$\mathcal{G}(t) = \mathcal{H}(t) + \int_t^T s^p\langle(\lambda(x(s)-x^*)+s\dot{x}(s)), sg(s)\rangle ds$$

   where: $\mathcal{H}(t) = t^p\mathcal{E}(t)$. Since $\mathcal{E}(t) = t(a(t)+b(t)+\xi c(t))$, we deduce from (A.6) that:

$$\begin{aligned}
\mathcal{H}'(t) &= t^{p-1}(p\mathcal{E}(t) + t\mathcal{E}'(t)) \\
&\leqslant t^p\left((2-\gamma\lambda+p)a(t) + (2\lambda+2-2\alpha+p)b(t) + \lambda(\lambda+1-\alpha)(-2\lambda+p)c(t)\right) \\
&\quad + \langle t^{p+1}g(t), \lambda(x(t)-x^*)+t\dot{x}(t)\rangle
\end{aligned}$$

   Hence the expected inequality:

$$\begin{aligned}
\mathcal{G}'(t) &= \mathcal{H}'(t) - t^p\langle(\lambda(x(t)-x^*)+t\dot{x}(t)), tg(t)\rangle \\
&\leqslant t^p((2-\gamma\lambda+p)a(t) + (2\lambda+2-2\alpha+p)b(t) + \lambda(\lambda+1-\alpha)(p-2\lambda)c(t))
\end{aligned}$$

A.2 Proof of Lemma 4.2: differentiating the energy function of Theorem 3.3

Consider the energy $\mathcal{E}$ defined as follows by:

$$\begin{aligned}
\mathcal{E}(t) &= F(x(t)) - F^* + \frac{1}{2}\|\lambda(x(t)-x^*)+\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t)-x^*\|^2 \\
&= a(t) + b(t) + c(t)
\end{aligned}$$

where:

$$a(t) = F(x(t)) - F^*,\ b(t) = \frac{1}{2}\|\lambda(x(t)-x^*)+t\dot{x}(t)\|^2,\ c(t) = \frac{1}{2}\|x(t)-x^*\|^2. \tag{A.7}$$

We then have:

$$\mathcal{E}'(t) = \langle \nabla F(x(t)), \dot{x}(t)\rangle + \langle \lambda(x(t)-x^*)+\dot{x}(t), \lambda\dot{x}(t)+\ddot{x}(t)\rangle + \xi\langle x(t)-x^*, \dot{x}(t)\rangle. \tag{A.8}$$

Since x is a solution of the ODE (3.5), we have:

$$\lambda\dot{x}(t) + \ddot{x}(t) = (\lambda-\alpha)\dot{x}(t) - \nabla F(x(t)) + g(t)$$

Hence:

$$\begin{aligned}
\mathcal{E}'(t) &= -\lambda\langle \nabla F(x(t)), x(t)-x^*\rangle + (\xi + \lambda(\lambda-\alpha))\langle \dot{x}(t), x(t)-x^*\rangle + (\lambda-\alpha)\|\dot{x}(t)\|^2 \\
&\quad + \langle g(t), \lambda(x(t)-x^*)+\dot{x}(t)\rangle.
\end{aligned}$$

Noticing that:

$$\|\lambda(x(t)-x^*)+\dot{x}(t)\|^2 = \|\dot{x}(t)\|^2 + 2\lambda\langle \dot{x}(t), x(t)-x^*\rangle + \lambda^2\|x(t)-x^*\|^2, \tag{A.9}$$

we deduce that

$$\begin{aligned}\mathcal{E}'(t) &= -\lambda\langle\nabla F(x(t)), x(t)-x^*\rangle + (\lambda-\alpha)\|\lambda(x(t)-x^*)+\dot{x}(t)\|^2 + (\xi-\lambda(\lambda-\alpha))\langle\dot{x}(t), x(t)-x^*\rangle \\ &\quad -\lambda^2(\lambda-\alpha)\|x(t)-x^*\|^2 + \langle g(t), \lambda(x(t)-x^*)+\dot{x}(t)\rangle \\ &= -\lambda\langle\nabla F(x(t)), x(t)-x^*\rangle + 2(\lambda-\alpha)b(t) + (\xi-\lambda(\lambda-\alpha))\langle\dot{x}(t), x(t)-x^*\rangle \\ &\quad -2\lambda^2(\lambda-\alpha)c(t) + \langle g(t), \lambda(x(t)-x^*)+\dot{x}(t)\rangle.\end{aligned}$$

Choosing now $\underline{\xi = \lambda(\lambda-\alpha)}$, we get:

$$\mathcal{E}'(t) = -\lambda\langle\nabla F(x(t)), x(t)-x^*\rangle + 2(\lambda-\alpha)b(t) - 2\lambda^2(\lambda-\alpha)c(t) + \langle g(t), \lambda(x(t)-x^*)+\dot{x}(t)\rangle.$$

Since F satisfies $\mathbf{H}_1(\gamma)$, we finally get:

$$\mathcal{E}'(t) \leqslant -\lambda\gamma a(t) + 2(\lambda-\alpha)b(t) - 2\lambda^2(\lambda-\alpha)c(t) + \langle g(t), \lambda(x(t)-x^*)+\dot{x}(t)\rangle.$$

A.3 Proof of Lemma 4.3: differentiating the energy function of Theorem 3.3 for the classical heavy ball system

Consider the energy:

$$\begin{aligned}\mathcal{E}(t) &= F(x(t)) - F^* + \frac{1}{2}\|\lambda(t)(x(t)-x^*)+\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t)-x^*\|^2 \\ &= F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2 + \frac{2\beta(t)}{\gamma+2}\langle x(t)-x^*, \dot{x}(t)\rangle.\end{aligned}$$

where: $\lambda(t) = \frac{2\beta}{\gamma+2}$ and $\xi(t) = -\lambda(t)^2$. Then:

$$\mathcal{E}'(t) = \langle\nabla F(x(t)), \dot{x}(t)\rangle + \langle\ddot{x}(t) + \frac{2\beta(t)}{\gamma+2}\dot{x}(t), \dot{x}(t)\rangle + \frac{2\beta(t)}{\gamma+2}\langle x(t)-x^*, \ddot{x}(t)\rangle + \frac{2\dot{\beta}(t)}{\gamma+2}\langle x(t)-x^*, \dot{x}(t)\rangle.$$

Since x satisfies the ODE (3.5) when $\theta \in (0,1)$, we have:

$$\ddot{x}(t) = g(t) - \beta(t)\dot{x}(t) - \nabla F(x(t)),$$

so that:

$$\begin{aligned}\mathcal{E}'(t) &= -\frac{2\beta(t)}{\gamma+2}\langle\nabla F(x(t)), x(t)-x^*\rangle - \frac{\gamma}{\gamma+2}\beta(t)\|\dot{x}(t)\|^2 + \frac{2}{\gamma+2}\left(\dot{\beta}(t) - \beta(t)^2\right)\langle x(t)-x^*, \dot{x}(t)\rangle \\ &\quad +\langle g(t), \dot{x}(t) + \frac{2\beta(t)}{2+\gamma}(x(t)-x^*)\rangle \\ &\leqslant -\frac{2\gamma}{\gamma+2}\beta(t)\left[F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2\right] + \frac{2\beta(t)}{\gamma+2}\left(\frac{\dot{\beta}(t)}{\beta(t)} - \beta(t)\right)\langle x(t)-x^*, \dot{x}(t)\rangle \\ &\quad +\langle g(t), \dot{x}(t) + \frac{2\beta(t)}{2+\gamma}(x(t)-x^*)\rangle\end{aligned}$$

assuming that F satisfies $\mathbf{H}_1(\gamma)$. Noticing that by definition of the energy $\mathcal{E}(t)$:

$$F(x(t)) - F^* + \frac{1}{2}\|\dot{x}(t)\|^2 = \mathcal{E}(t) - \frac{2\beta(t)}{\gamma+2}\langle x(t)-x^*, \dot{x}(t)\rangle,$$

we get:

$$\mathcal{E}'(t) \leqslant -\frac{2\gamma}{\gamma+2}\beta(t)\mathcal{E}(t) + \frac{2\beta(t)}{\gamma+2}\left(\frac{\dot{\beta}(t)}{\beta(t)} + \frac{\gamma-2}{\gamma+2}\beta(t)\right)\langle x(t)-x^*, \dot{x}(t)\rangle + \langle g(t), \dot{x}(t) + \frac{2\beta(t)}{2+\gamma}(x(t)-x^*)\rangle$$

as expected.

A.4 Proof of Lemma 4.4: differentiating the energy function of Theorem 3.4

Let λ be a non-negative real constant and $\xi(.)$ a real-valued function. Consider the energy:

$$\begin{aligned}\mathcal{E}(t) &= t^2(F(x(t))-F^*)+\frac{1}{2}\|\lambda(x(t)-x^*)+t\dot{x}(t)\|^2+\frac{\xi(t)}{2}\|x(t)-x^*\|^2\\ &= t(a(t)+b(t)+\xi c(t))\end{aligned}$$

where:

$$a(t)=t(F(x(t))-F^*),\ b(t)=\frac{1}{2t}\|\lambda(x(t)-x^*)+t\dot{x}(t)\|^2,\ c(t)=\frac{1}{2t}\|x(t)-x^*\|^2.$$

1. **Differentiating $\mathcal{E}$**

$$\begin{aligned}\mathcal{E}'(t) =& 2t(F(x(t))-F^*)+t^2\langle\nabla F(x(t)),\dot{x}(t)\rangle+\langle\lambda(x(t)-x^*)+t\dot{x}(t),(1+\lambda)\dot{x}(t)+t\ddot{x}(t)\rangle\\ &+\frac{\dot{\xi}(t)}{2}\|x(t)-x^*\|^2+\xi(t)\langle\dot{x}(t),x(t)-x^*\rangle\end{aligned} \tag{A.10}$$

Since x satisfies the ODE (3.5) with $g\equiv 0$, we have:

$$\ddot{x}(t)=-\frac{\alpha}{t^\theta}\dot{x}(t)-\nabla F(x(t)),$$

hence:

$$\begin{aligned}\mathcal{E}'(t) = 2a(t)-\lambda t\langle\nabla F(x(t)),x(t)-x^*\rangle+\left[\lambda(\lambda+1-\alpha t^{1-\theta})+\xi(t)\right]\langle x(t)-x^*,\dot{x}(t)\rangle\\ +t\left(\lambda+1-\alpha t^{1-\theta}\right)\|\dot{x}(t)\|^2+\frac{\dot{\xi}(t)}{2}\|x(t)-x^*\|^2.\end{aligned}$$

Since F satisfies $\mathbf{H}_1(\gamma_1)$, we get:

$$\begin{aligned}\mathcal{E}'(t) = (2-\lambda\gamma_1)a(t)+\left[\lambda(\lambda+1-\alpha t^{1-\theta})+\xi(t)\right]\langle x(t)-x^*,\dot{x}(t)\rangle\\ +t\left(\lambda+1-\alpha t^{1-\theta}\right)\|\dot{x}(t)\|^2+\frac{\dot{\xi}(t)}{2}\|x(t)-x^*\|^2.\end{aligned} \tag{A.11}$$

Noticing that

$$\|\lambda(x(t)-x^*)+t\dot{x}(t)\|^2=t^2\|\dot{x}(t)\|^2+2t\lambda\langle\dot{x}(t),x(t)-x^*\rangle+\lambda^2\|x(t)-x^*\|^2$$

i.e.

$$t\|\dot{x}(t)\|^2=2b(t)-2\lambda^2c(t)-2\lambda\langle x(t)-x^*,\dot{x}(t)\rangle, \tag{A.12}$$

we get:

$$\begin{aligned}\mathcal{E}'(t) = (2-\lambda\gamma_1)a(t)+2(\lambda+1-\alpha t^{1-\theta})b(t)-2\lambda^2(\lambda+1-\alpha t^{1-\theta})c(t)\\ +\left[\xi(t)-\lambda(1+\lambda-\alpha t^{1-\theta})\right]\langle x(t)-x^*,\dot{x}(t)\rangle+t\dot{\xi}(t)c(t).\end{aligned} \tag{A.13}$$

Setting $\underline{\xi(t)=\lambda(\lambda+1-\alpha t^{1-\theta})}$, we obtain:

$$\begin{aligned}\mathcal{E}'(t) = (2-\lambda\gamma_1)a(t)+2(\lambda+1-\alpha t^{1-\theta})b(t)-2\lambda^2(\lambda+1-\alpha t^{1-\theta})c(t)\\ +\left[\xi(t)-\lambda(1+\lambda-\alpha t^{1-\theta})\right]\langle x(t)-x^*,\dot{x}(t)\rangle-\lambda\alpha(1-\theta)t^{1-\theta}c(t).\end{aligned} \tag{A.14}$$

2. **Differentiating $\mathcal{H}$.**

Recall that $\mathcal{H}(t)=t^p\mathcal{E}(t)$. Hence $\mathcal{H}'(t)=t^{p-1}(p\mathcal{E}(t)+(1+t)\mathcal{E}'(t))$ Hence

$$\begin{aligned}\mathcal{H}'(t) &= t^{p-1}(p\mathcal{E}(t)+t\mathcal{E}'(t))\\ &= t^p(pa(t)+pb(t)+p\xi(t)c(t)+\mathcal{E}'(t))\\ &\leqslant t^p\left[(2-\lambda\gamma_1+p)a(t)+(2\lambda+2-2\alpha t^{1-\theta}+p)b(t)\right.\\ &\quad\left.+\lambda((\lambda+1)(p-2\lambda)-\alpha(p+1-\theta-2\lambda)t^{1-\theta})c(t)\right]\end{aligned}$$

as expected.

References

[1] V. Apidopoulos, J.-F. Aujol, Ch. Dossal, and A. Rondepierre. Convergence rates of an inertial gradient descent algorithm under growth and flatness conditions. preprint, December 2018.

[2] H. Attouch and J. Bolte. On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Mathematical Programming*, 116(1):5–16, 2009.

[3] H. Attouch and A. Cabot. Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *Journal of Differential Equations*, 263(9):5412–5458, 2017.

[4] H. Attouch and A. Cabot. Convergence rates of inertial forward-backward algorithms. *SIAM Journal on Optimization*, 28(1):849–874, 2018.

[5] H. Attouch and Z. Chbani. Fast inertial dynamics and FISTA algorithms in convex optimization. Perturbation aspects. *arXiv preprint arXiv:1507.01367*, 2015.

[6] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont. Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming*, 168(1-2):123–175, 2018.

[7] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leqslant 3$. *ESAIM: COCV*, 2019.

[8] H. Attouch and J. Peypouquet. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $\frac{1}{k^2}$. *SIAM Journal of Optimization*, 26(3):1824–1834, 2016.

[9] J-F. Aujol and C. Dossal. Optimal rate of convergence of an ODE associated to the fast gradient descent schemes for $b > 0$. *Hal Preprint hal-01547251*, June 2017.

[10] J.-F. Aujol, C. Dossal, and A. Rondepierre. Optimal convergence rates for Nesterov acceleration. *arXiv preprint arXiv:1805.05719*, 2018.

[11] M. Balti and R. May. Asymptotic for the perturbed heavy ball system with vanishing damping term. *Evolution Equations & Control Theory*, 6(2):177–186, 2017.

[12] P. Bégout, J. Bolte, and M. A. Jendoubi. On damped second order gradients systems. *Journal of Differential Equation*, 259(9):3315–3143, 2015.

[13] J. Bolte, T.P. Nguyen, J. Peypouquet, and B.W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165(2):471–507, 2017.

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

[15] A. Cabot, H. Engler, and S. 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] A. Cabot, H. Engler, S. 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] A. Chambolle and C. 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.

[18] G. Garrigos, L. Rosasco, and S. Villa. Convergence of the forward-backward algorithm: Beyond the worst case with the help of geometry. *arXiv preprint arXiv:1703.09477*, 2017.

[19] Alain Haraux and Mohamed Ali Jendoubi. On a second order dissipative ode in hilbert space with an integrable source term. *Acta Mathematica Scientia*, 32(1):155 – 163, 2012. Mathematics Dedicated to professor Constantine M. Dafermos on the occasion of his 70th birthday.

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

[21] S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. In *Les Équations aux Dérivées Partielles (Paris, 1962)*, pages 87–89. Éditions du Centre National de la Recherche Scientifique, Paris, 1963.

[22] S. Łojasiewicz. Sur la géométrie semi- et sous-analytique. *Annales de l'Institut Fourier. Université de Grenoble*, 43(5):1575–1595, 1993.

[23] R. May. Asymptotic for a second order evolution equation with convex potential and vanishing damping term. *Turkish Journal of Mathematics*, 41:681–685, 2017.

[24] B. Polyak and P. Shcherbakov. Lyapunov functions: An optimization theory perspective. *IFAC-PapersOnLine*, 50(1):7456–7461, 2017.

[25] W. Su, S. Boyd, and E. 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.