ESAIM: COCV 25 (2019) 2
https://doi.org/10.1051/cocv/2017083

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

RATE OF CONVERGENCE OF THE NESTEROV ACCELERATED GRADIENT METHOD IN THE SUBCRITICAL CASE $\alpha \leq 3$

Hedy Attouch[1,*], Zaki Chbani[2] and Hassan Riahi[2]

Abstract. In a Hilbert space setting $\mathcal{H}$, given $\Phi : \mathcal{H} \to \mathbb{R}$ a convex continuously differentiable function, and α a positive parameter, we consider the inertial dynamic system with Asymptotic Vanishing Damping

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

Depending on the value of α with respect to 3, we give a complete picture of the convergence properties as $t \to +\infty$ of the trajectories generated by (AVD)_{α}, as well as iterations of the corresponding algorithms. Indeed, as shown by Su-Boyd-Candès, the case $\alpha = 3$ corresponds to a continuous version of the accelerated gradient method of Nesterov, with the rate of convergence $\Phi(x(t)) - \min \Phi = \mathcal{O}(t^{-2})$ for $\alpha \geq 3$. Our main result concerns the subcritical case $\alpha \leq 3$, where we show that $\Phi(x(t)) - \min \Phi = \mathcal{O}(t^{-\frac{2}{3}\alpha})$. This overall picture shows a continuous variation of the rate of convergence of the values $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(t^{-p(\alpha)})$ with respect to $\alpha > 0$: the coefficient $p(\alpha)$ increases linearly up to 2 when α goes from 0 to 3, then displays a plateau. Then we examine the convergence of trajectories to optimal solutions. As a new result, in the one-dimensional framework, for the critical value $\alpha = 3$, we prove the convergence of the trajectories. In the second part of this paper, we study the convergence properties of the associated forward-backward inertial algorithms. They aim to solve structured convex minimization problems of the form $\min \{\Theta := \Phi + \Psi\}$, with Φ smooth and Ψ nonsmooth. The continuous dynamics serves as a guideline for this study. We obtain a similar rate of convergence for the sequence of iterates (x_k): for $\alpha \leq 3$ we have $\Theta(x_k) - \min \Theta = \mathcal{O}(k^{-p})$ for all $p < \frac{2\alpha}{3}$, and for $\alpha > 3$ $\;\Theta(x_k) - \min \Theta = o(k^{-2})$. Finally, we show that the results are robust with respect to external perturbations.

Mathematics Subject Classification. 49M37, 65K05, 90C25

Received June 23, 2017. Accepted December 17, 2017.

1. Introduction

Throughout the paper, $\mathcal{H}$ is a real Hilbert space which is endowed with the scalar product $\langle \cdot, \cdot \rangle$, with $\|x\|^2 = \langle x, x \rangle$ for $x \in \mathcal{H}$. Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function. In a first part, we consider the second-order

Keywords and phrases: Accelerated gradient method, FISTA, inertial forward-backward algorithms, Nesterov method, proximal-based methods; structured convex optimization, subcritical case, vanishing damping.

[1] Institut Montpelliérain A. Grothendieck, UMR CNRS 5149, Université Montpellier, 34095 Montpellier Cedex 5, France.
[2] Cadi Ayyad University, Faculty of Sciences Semlalia, Mathematics, 40000 Marrakech, Morroco.

[*] Corresponding author: hedy.attouch@univ-montp2.fr

differential equation

$$(\mathrm{AVD})_\alpha \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla \Phi(x(t)) = 0, \tag{1.1}$$

where α is a positive parameter. Depending on whether α is greater or less than the critical value 3, we analyze the convergence rate of the values $\Phi(x(t)) - \min_{\mathcal{H}} \Phi$, as $t \to +\infty$. The novelty of our results comes mainly from the study of the subcritical case $\alpha < 3$ whose convergence rate was largely unknown. In a second part, we study the corresponding inertial forward-backward algorithms in the case of structured minimization problems. The analysis of the continuous dynamics will serve as a guideline for the study of these algorithms. Let us first recall some historical facts, explaining the importance of these issues.

1.1. From the heavy ball with friction to fast optimization

The heavy ball with friction system, which involves a fixed positive damping coefficient γ

$$\ddot{x}(t) + \gamma\dot{x}(t) + \nabla \Phi(x(t)) = 0 \tag{1.2}$$

was introduced, from an optimization perspective, by Polyak [31, 32]. Its convergence in the convex case was first obtained by Álvarez in [1]. In recent years, several studies have been devoted to the study of the Inertial Gradient System $(\mathrm{IGS})_\gamma$, with a time-dependent positive damping coefficient $\gamma(\cdot)$

$$(\mathrm{IGS})_\gamma \qquad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla \Phi(x(t)) = 0. \tag{1.3}$$

A particularly interesting situation is the case $\gamma(t) \to 0$ as $t \to +\infty$, called asymptotic vanishing damping. As pointed out by Su-Boyd-Candès in [35], the $(\mathrm{IGS})_\gamma$ system with $\gamma(t) = \frac{\alpha}{t}$, that's $(\mathrm{AVD})_\alpha$ given in (1.1), can be seen as a continuous version of the accelerated gradient method of Nesterov (see [24, 25, 26]). Its adaptation to the case of structured "smooth + nonsmooth convex optimization gives the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) of Beck-Teboulle [12]. When $\alpha \geq 3$, the rate of convergence of these methods is $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$, where k is the number of iterations. Convergence of the trajectories generated by (1.1), and of the sequences generated by Nesterov's method, has been an elusive question for decades. However, when considering (1.1) with $\alpha > 3$, it was shown by Attouch-Chbani-Peypouquet-Redont [5] and May [23], that each trajectory converges weakly to an optimal solution, with the improved rate of convergence $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o(\frac{1}{t^2})$. Corresponding results for the algorithmic case have been obtained by Chambolle-Dossal [16] and Attouch-Peypouquet [6]. The case of a general time-dependent damping coefficient $\gamma(\cdot)$ has been recently considered by Attouch-Cabot in [2, 3]. The latter includes the corresponding forward-backward algorithms, and unifies previous results. These results are of great importance because they are in some sense optimal: it is well known that, for first-order methods, the rate of convergence $\frac{1}{k^2}$ is the best one can expect in the worst case.

1.2. The subcritical case $\alpha \leq 3$

Whereas the case $\alpha > 3$ has been studied in depth, the case $\alpha < 3$ has remained largely unknown. Our main contribution is to show that in the subcritical case $\alpha \leq 3$, we still have a property of rapid convergence for the values, that now depends on α, namely

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right).$$

Of course, in the case $\alpha = 3$, from this formula, we recover the well-known convergence rate $\mathcal{O}(\frac{1}{t^2})$ of the accelerated gradient method of Nesterov. This shows a continuous variation of the convergence rate for the values, when α varies along the positive real line. The following table gives a synthetic view, where α is the coefficient of the damping parameter $\frac{\alpha}{t}$ entering the definition of $(\text{AVD})_\alpha$ (defined in (1.1)).

α	$\alpha < 3$	$\alpha = 3$	$\alpha > 3$
$\Phi(x(t)) - \min_{\mathcal{H}} \Phi$	$\mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$	$\mathcal{O}\left(\frac{1}{t^2}\right)$	$o\left(\frac{1}{t^2}\right)$

1.3. From continuous dynamics to algorithms

For applications, and in order to develop fast numerical splitting methods, it is important to consider "*smooth + nonsmooth*" structured optimization problems of the form

$$\min\left\{\Phi(x) + \Psi(x) : \ x \in \mathcal{H}\right\}, \tag{1.4}$$

where $\Phi : \mathcal{H} \to \mathbb{R}$ is a continuously differentiable convex function whose gradient is Lipschitz continuous, and $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower-semicontinuous convex function. We set $\Theta := \Phi + \Psi$, which is the convex lower-semicontinuous function to minimize. To handle such non-smooth optimization problems, we are naturally led to extend $(\text{AVD})_\alpha$, and to consider the differential inclusion

$$(\text{AVD})_\alpha \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) + \partial\Psi(x(t)) \ni 0, \tag{1.5}$$

where $\partial\Psi$ is the subdifferential of Ψ. The time discretization of this system, implicit with respect to the nonsmooth operator $\partial\Psi$, and explicit with respect to the smooth operator $\nabla\Phi$, gives the Inertial Forward-Backward algorithm

$$(\text{IFB})_\alpha \quad \begin{cases} y_k = x_k + (1 - \frac{\alpha}{k})(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k)). \end{cases}$$

The first above formula can be written in an equivalent way $y_k = x_k + \alpha_k(x_k - x_{k-1})$, with the extrapolation coefficient $\alpha_k = 1 - \frac{\alpha}{k}$. This value of α_k comes naturally from the time discretization of the continuous dynamics. Of course, other commonly used choices can also be considered as $\alpha_k = \frac{k}{k+\alpha}$, or the choice of Nesterov. As a consequence of the general theory developed by Attouch-Cabot in [2, 3], they lead to similar convergence rates in the case $\alpha \geq 3$. It is likely that this is also the case when $\alpha < 3$. This is a point that deserves to be deepened. The study of the continuous dynamics will serve us as a guideline to analyze this algorithm. Indeed, we will obtain convergence results for $(\text{IFB})_\alpha$ which are very similar those of the continuous dynamics $(\text{AVD})_\alpha$, and which are summarized in the following table.

α	$\alpha < 3$	$\alpha = 3$	$\alpha > 3$
$\Theta(x_k) - \min_{\mathcal{H}} \Theta$	$\mathcal{O}\left(\frac{1}{k^p}\right) \quad \forall p < \frac{2\alpha}{3}$	$\mathcal{O}\left(\frac{1}{k^2}\right)$	$o\left(\frac{1}{k^2}\right)$

Since we submitted the article, we have been informed that, simultaneously and independently, Apidopoulos-Aujol-Dossal in [8] obtained similar results regarding the rate of convergence of the algorithm in the subcritical case. When $\alpha < 3$, in order not to have too complicated proof, we have only proved $\Theta(x_k) - \min_{\mathcal{H}} \Theta = \mathcal{O}\left(\frac{1}{k^p}\right)$ for all $p < \frac{2\alpha}{3}$. Indeed, in the above mentioned paper the authors prove the better result $\Theta(x_k) - \min_{\mathcal{H}} \Theta = \mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$. As a remarkable result, Aujol-Dossal in [10] completed this result by showing that, in the subcritical case $\alpha < 3$, this rate of convergence of the values is optimal (the best that is guarantied in the worst case).

1.4. Organization of the paper

In Section 2, based on new Lyapunov functions, we analyze in a unifying way the convergence rate of the values for the trajectories of $(\text{AVD})_\alpha$. We also study the decay property of the velocity. The main novelty concerns the subcritical case $\alpha \leq 3$. In Section 3, we study the convergence of the trajectories. In the one-dimensional framework, we prove the convergence of trajectories in the critical case $\alpha = 3$. We also analyze the convergence rate in the case of a strong minimum. Then, in Section 4, based on the results for the continuous dynamics, we study the convergence of the corresponding inertial forward-backward algorithms, and obtain very similar results. Finally, in Section 5, we complete this study by showing that the results are robust with respect to external perturbations.

2. Convergence rates of the continuous dynamics

Throughout the paper (unless otherwise stated), we assume that $\Phi : \mathcal{H} \to \mathbb{R}$ is convex and differentiable, its gradient $\nabla\Phi$ is Lipschitz continuous on bounded sets, and $S := \text{argmin}\Phi \neq \emptyset$.

We take for granted the existence and uniqueness of a global solution to the Cauchy problem associated with $(\text{AVD})_\alpha$ (defined in (1.1)). We point out that, given $t_0 > 0$, for any $x_0 \in \mathcal{H},\ v_0 \in \mathcal{H}$, the existence of a unique global solution on $[t_0, +\infty[$ for the Cauchy problem with initial condition $x(t_0) = x_0$ and $\dot{x}(t_0) = v_0$ is guaranteed by the above hypothesis, see ([18], prop. 6.2.1) . Starting from $t_0 > 0$ comes from the singularity of the damping coefficient $\gamma(t) = \frac{\alpha}{t}$ at zero. Indeed, since we are only concerned about the asymptotic behaviour of the trajectories, we do not really care about the origin of time. If one insists starting from $t_0 = 0$, then all the results remain valid taking $\gamma(t) = \frac{\alpha}{t+1}$.

Our main results concerning the continuous system are given in the following theorem.

Theorem 2.1. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex continuously differentiable function such that* argminΦ *is nonempty. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a classical global solution of $(AVD)_\alpha$.*

(i) Suppose $\alpha \geq 3$. Then, we have the following result of rapid convergence of the values

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^2}\right).$$

(ii) Suppose $0 < \alpha \leq 3$. Then the following rate of convergence of the values is verified

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right).$$

Precisely

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{t^{\frac{2\alpha}{3}}}, \quad \textit{with} \ \ C = t_0^{\frac{2\alpha}{3}} \left(\Phi(x(t_0)) - \min_{\mathcal{H}} \Phi + \|\dot{x}(t_0)\|^2\right) + \frac{\alpha(\alpha+1)}{3} \text{dist}^2(x(t_0), \text{argmin}\Phi).$$

Proof. Fix $z \in \operatorname{argmin} \Phi$, and consider the energy function $\mathcal{E}^p_{\lambda,\xi} : [t_0, +\infty[\to \mathbb{R}^+$

$$\mathcal{E}^p_{\lambda,\xi}(t) = t^{2p}\left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|\lambda(t)(x(t)-z) + t^p\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t)-z\|^2, \tag{2.1}$$

that will serve as a Lyapunov function. In the definition of $\mathcal{E}^p_{\lambda,\xi}$, p is a positive real number, and $\lambda(\cdot)$, $\xi(\cdot)$ are positive functions. They will be appropriately chosen during the proof, so as to make $\mathcal{E}^p_{\lambda,\xi}(\cdot)$ a nonincreasing function. Let us first derivate each of the constitutive elements of $\mathcal{E}^p_{\lambda}(\cdot)$. By means of the classical derivation chain rule we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} t^{2p}\left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] &= 2pt^{2p-1}\Big(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\Big) + t^{2p}\left\langle \nabla\Phi(x(t)), \dot{x}(t)\right\rangle \\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\xi(t)}{2}\|x(t)-z\|^2 &= \frac{\dot{\xi}(t)}{2}\|x(t)-z\|^2 + \xi(t)\left\langle x(t)-z, \dot{x}(t)\right\rangle \\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\|\lambda(t)(x(t)-z) + t^p\dot{x}(t)\|^2 &= \left\langle \lambda(t)(x(t)-z) + t^p\dot{x}(t), \lambda(t)\dot{x}(t) + \dot{\lambda}(t)(x(t)-z) + pt^{p-1}\dot{x}(t) + t^p\ddot{x}(t)\right\rangle \\
&= \left\langle \lambda(t)(x(t)-z) + t^p\dot{x}(t), (\lambda(t) - (\alpha-p)t^{p-1})\dot{x}(t) + \dot{\lambda}(t)(x(t)-z) - t^p\nabla\Phi(x)(t)\right\rangle \\
&\leq \left\langle \lambda(t)(x(t)-z) + t^p\dot{x}(t), (\lambda(t) - (\alpha-p)t^{p-1})\dot{x}(t) + \dot{\lambda}(t)(x(t)-z)\right\rangle \\
&\quad - t^{2p}\left\langle \nabla\Phi(x(t)), \dot{x}(t)\right\rangle - \lambda(t)t^p\Big(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\Big).
\end{aligned}$$

To obtain the two last equations we have successively used (1.1) and the convexity of $\Phi(\cdot)$. Adding the above results, we obtain, after simplification,

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \leq\ & t^p\left[2pt^{p-1} - \lambda(t))\right](\Phi(x(t)) - \min_{\mathcal{H}}\Phi) \\
&+ \left[\xi(t) + t^p\dot{\lambda}(t) - (\alpha-p)t^{p-1}\lambda(t) + \lambda(t)^2\right]\left\langle x(t)-z, \dot{x}(t)\right\rangle \\
&- t^p\left[(\alpha-p)t^{p-1} - \lambda(t)\right]\|\dot{x}(t)\|^2 + \left[\lambda(t)\dot{\lambda}(t) + \tfrac{\dot{\xi}(t)}{2}\right]\|x(t)-z\|^2.
\end{aligned}$$

Let us successively examine the differents terms entering the second member of (2.2):

Let us make the first two terms equal to zero by taking respectively

(**H**$_1$): $\lambda(t) = 2pt^{p-1}$

(**H**$_2$): $\xi(t) + t^p\dot{\lambda}(t) - (\alpha-p)t^{p-1}\lambda(t) + \lambda(t)^2 = 0.$

Combining these two equations we obtain

$\xi(t) = 2(\alpha - 4p + 1)pt^{2(p-1)}$.

To ensure that $\mathcal{E}^p_{\lambda,\xi}$ is nonnegative, we must impose $\xi(\cdot) \geq 0$. This is equivalent to assuming that

(**H**$_3$) $\alpha \geq 4p - 1$.

Recall that we want $\mathcal{E}^p_{\lambda,\xi}(\cdot)$ to be a nonincreasing function. Hence, we impose on the third term of the second member of (2.2) to satisfy $(\alpha-p)t^{p-1} - \lambda(t) \geq 0$. This is equivalent to assuming that

(**H**$_4$) $\alpha \geq 3p$.

With the above choice of the parameters $\lambda(\cdot), \xi(\cdot), p$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \leq \beta(t)\|x(t)-z\|^2, \tag{2.2}$$

where $\beta(t) = \lambda(t)\dot{\lambda}(t) + \frac{\dot{\xi}(t)}{2}$. A straightforward calculation gives $\beta(t) = -2p(1-p)(\alpha - 2p + 1)t^{2p-3}$. We want $\beta(\cdot)$ to be less than or equal to zero. Since $\alpha \geq 2p - 1$ is implied by $(\mathbf{H}_4)$ (or $(\mathbf{H}_3)$), we impose the supplementary condition:

$(\mathbf{H}_5)$ $1 \geq p$.

Let us put together the results above. By taking $p = \min(1, \frac{\alpha}{3}, \frac{\alpha+1}{4})$, and $\lambda(\cdot)$, $\xi(\cdot)$ given respectively by $(\mathbf{H}_1)$ and $(\mathbf{H}_2)$, all the conditions $(\mathbf{H}_1)$ to $(\mathbf{H}_5)$ are verified. As a consequence, $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \leq 0$, and $\mathcal{E}^p_{\lambda,\xi}$ is nonincreasing. Coming back to the definition of $\mathcal{E}^p_{\lambda}(\cdot)$, taking account that $\xi(\cdot)$ is nonegative, we deduce that, for all $t \geq t_0$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{\mathcal{E}^p_{\lambda,\xi}(t_0)}{t^{2p}} \cdot \tag{2.3}$$

A closer look at the formula $p = \min(1, \frac{\alpha}{3}, \frac{\alpha+1}{4})$ shows that this expression can be simplified as $p = \min(1, \frac{\alpha}{3})$. This leads us to distinguish the two cases:

- $\alpha \geq 3$. Then $p = \min(1, \frac{\alpha}{3}) = 1$, which, from (2.3), gives item (i) of Theorem 2.1.
- $\alpha \leq 3$. Then $p = \min(1, \frac{\alpha}{3}) = \frac{\alpha}{3}$, which, from (2.3), gives item (ii). Let us make precise the value of the constant that appears in the corresponding estimation

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{t^{\frac{2\alpha}{3}}} \cdot \tag{2.4}$$

By definition (2.1) of $\mathcal{E}^p_{\lambda,\xi}(t_0)$ and elementary computation we obtain as value of the constant C

$$C = t_0^{\frac{2\alpha}{3}} \left(\Phi(x(t_0)) - \min_{\mathcal{H}} \Phi + \|\dot{x}(t_0)\|^2\right) + \frac{\alpha(\alpha+1)}{3} \mathrm{dist}^2(x(t_0), \mathrm{argmin}\,\Phi), \tag{2.5}$$

which completes the proof of Theorem 2.1. □

Remark 2.2. The proof of the Theorem 2.1 is based on the use of the Lyapunov function $\mathcal{E}^p_{\lambda,\xi}$, which, with the specific choices of the parameters $p = \min(1, \frac{\alpha}{3})$, $\lambda(t) = 2pt^{p-1}$, $\xi(t) = 2(\alpha - 4p + 1)pt^{2(p-1)}$, is written as follows

$$\begin{aligned}\mathcal{E}^p_{\lambda,\xi}(t) = t^{2\min(1,\frac{\alpha}{3})} \left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|2\min(1, \frac{\alpha}{3})t^{\min(1,\frac{\alpha}{3})-1}(x(t) - z) + t^{\min(1,\frac{\alpha}{3})}\dot{x}(t)\|^2 \\ +(\alpha - 4\min(1, \frac{\alpha}{3}) + 1)\min(1, \frac{\alpha}{3})t^{2(\min(1,\frac{\alpha}{3})-1)}\|x(t) - z\|^2.\end{aligned}$$

Using this Lyapunov function, we were able to provide a unified proof of the convergence results in the two different cases $\alpha \leq 3$, and $\alpha \geq 3$. Obviously, the complexity of this formula explains why historically these two cases have been considered independently. In each of these cases, the formula simplifies significantly:

$i)$ In the case $\alpha \geq 3$, we have $p = 1$, $\lambda(t) = 2$, $\xi(t) = 2(\alpha - 3)$, which gives

$$\mathcal{E}^1_{\lambda,\xi}(t) = t^2 \left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|2(x(t) - z) + t\dot{x}(t)\|^2 + (\alpha - 3)\|x(t) - z\|^2.$$

This is precisely the Lyapunov function used in the proof of Theorem 2.14 in [5].

$ii)$ In the case $\alpha \leq 3$, we have $p = \frac{\alpha}{3}$, $\lambda(t) = \frac{2\alpha}{3}t^{\frac{\alpha}{3}-1}$, $\xi(t) = \frac{2\alpha}{3}(\alpha - \frac{4\alpha}{3} + 1)t^{2(\frac{\alpha}{3}-1)}$, which gives

$$\mathcal{E}^{\frac{\alpha}{3}}_{\lambda,\xi}(t) = t^{\frac{2\alpha}{3}} \left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|\frac{2\alpha}{3}t^{\frac{\alpha}{3}-1}(x(t) - z) + t^{\frac{\alpha}{3}}\dot{x}(t)\|^2 + \frac{\alpha}{3}(\alpha - \frac{4\alpha}{3} + 1)t^{2(\frac{\alpha}{3}-1)}\|x(t) - z\|^2 \cdot$$

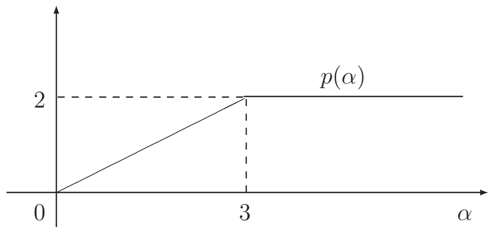

FIGURE 1. Rate of convergence of the values.

Note that it is rather difficult to postulate *a priori* these formulas: it was by an identification technique that we could find the appropriate parameters.

Remark 2.3. Let us give a synthetic view of the rate of convergence of the values for the trajectories of $(\text{AVD})_\alpha$. Given $x(\cdot)$ a global solution trajectory, we have

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{p(\alpha)}}\right) \quad \text{with} \quad p(\alpha) = \min\left(\frac{2\alpha}{3}, 2\right).$$

This formula and Figure 1 shows the two distinct regimes. On the interval $]0,3]$ the exponent $p(\alpha)$ increases linearly from zero to 2. Then, after $\alpha = 3$ there is a plateau, the exponent $p(\alpha)$ remains constant equal to 2. Of course, this is consistent with the Nesterov complexity bound [24, 25], which tells us that for first-order methods, the rate of convergence $\frac{1}{k^2}$ is the best one can expect in the worst case. See Drori-Teboulle [17] and Kim-Fessler [21] for a recent account on first-order methods that achieve best performance. In a recent paper [10], Aujol-Dossal showed the optimality of the decay rate $\mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$ in the subcritical case $\alpha < 3$, thus completing the optimality of this curve.

Let us complement the pointwise estimates concerning $\Phi(x(t)) - \min \Phi$ given in Theorem 2.1 by an integral estimate.

Theorem 2.4. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex continuously differentiable function such that* $\operatorname{argmin} \Phi$ *is nonempty. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a classical global solution of $(AVD)_\alpha$. Then, for any p satisfying $p \leq 1$ and $p < \frac{\alpha}{3}$, the following inequality is satisfied*

$$\int_{t_0}^{+\infty} t^{2p-1}(\Phi(x(t)) - \min \Phi)\mathrm{d}t \leq \frac{\mathcal{E}^p_{\lambda,\xi}(t_0)}{(\alpha - 3p)},$$

where $\lambda(t) = (\alpha - p)t^{p-1}$ and $\xi(t) = (1-p)(\alpha - p)t^{2(p-1)}$. Thus,

(i) If $\alpha > 3$, then, $\displaystyle\int_{t_0}^{+\infty} t(\Phi(x(t)) - \min \Phi)\mathrm{d}t < +\infty$.

(ii) If $\alpha \leq 3$, then for any $p < \frac{2\alpha}{3} - 1$, $\displaystyle\int_{t_0}^{+\infty} t^p(\Phi(x(t)) - \min \Phi)\mathrm{d}t < +\infty$.

In particular for $\alpha = 3$, we have $\displaystyle\int_{t_0}^{+\infty} t^p(\Phi(x(t)) - \min \Phi)\mathrm{d}t < +\infty$ for all $p < 1$.

Proof. Let us slightly modify the choices concerning the assumptions $(\mathbf{H}_1)$ to $(\mathbf{H}_5)$ that have been made in the proof of Theorem 2.1. Still we want to maintain the property $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \leq 0$. Let us assume

1. $\lambda(t) \geq 2pt^{p-1}$

2. $(\alpha - p)t^{p-1} - \lambda(t) = 0$.

Clearly, these two conditions are equivalent to

$$\lambda(t) = (\alpha - p)t^{p-1} \quad \text{with} \quad \alpha \geq 3p.$$

On the other hand, condition $(\mathbf{H}_2)$ becomes

$$\xi(t) = -t^p \dot{\lambda}(t) = (1-p)(\alpha - p)t^{2(p-1)}$$

which gives

$$\beta(t) = \lambda(t)\dot{\lambda}(t) + \frac{\dot{\xi}(t)}{2} = -(1-p)(\alpha-p)(\alpha-2p+1)t^{2p-3}.$$

Altogether taking $p \leq \min(1, \frac{\alpha}{3})$ ensures that $\xi(\cdot) \geq 0$, $\beta(\cdot) \leq 0$, and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) + \left[(\alpha - 3p)t^{2p-1}\right](\Phi(x(t)) - \min \Phi) \leq 0.$$

Then we integrate, and use $\mathcal{E}^p_{\lambda,\xi}(t) \geq 0$, $\alpha - 3p > 0$ to conclude. □

Let us supplement the asymptotic analysis by examining the rate of decay of the speed. First, we establish integral estimates, and then we build on these results to obtain sharp pointwise estimates.

Theorem 2.5. *(integral estimates of the speed) Let $x : [t_0; +\infty[\to \mathcal{H}$ be a solution of (AVD)$_\alpha$. The following integral estimates are satisfied:*

1. *If $\alpha > 3$, then*

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 \mathrm{d}t < +\infty.$$

2. *If $\alpha \leq 3$, then*

$$\int_{t_0}^{+\infty} t^p\|\dot{x}(t)\|^2 \mathrm{d}t < +\infty \quad \textit{for all} \quad p < \alpha - 2.$$

In particular for $\alpha = 3$ we have

$$\int_{t_0}^{+\infty} t^{1-\epsilon}\|\dot{x}(t)\|^2 \mathrm{d}t < +\infty \quad \textit{for all} \quad \epsilon > 0.$$

Proof. (1) Let us keep the same choice of the parameters as that done in the proof of Theorem 2.1, that is $\lambda(t) = 2pt^{p-1}$, $\xi(t) = 2(\alpha - 4p + 1)pt^{2(p-1)}$, and $p = \min(1, \frac{\alpha}{3})$. As a consequence $\mathcal{E}^p_{\lambda,\xi}(\cdot)$ is nonincreasing, and from (2.2) we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) + t^p\left[(\alpha - p)t^{p-1} - \lambda(t)\right]\|\dot{x}(t)\|^2 \leq 0.$$

for all $t \geq t_0$. Equivalently

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) + (\alpha - 3p)t^{2p-1}\|\dot{x}(t)\|^2 \leq 0. \tag{2.6}$$

This inequality gives us information about the rate of decay of the energy function $\mathcal{E}^p_{\lambda,\xi}$ only when $\alpha > 3p = \min(\alpha, 3)$. Obviously, this later condition is equivalent to $\alpha > 3$, which gives $p = 1$, and $2p-1 = 1$. By integrating (2.6) on $[t_0, t]$ we obtain

$$\mathcal{E}^p_{\lambda,\xi}(t) + (\alpha - 3)\int_{t_0}^{t} s\|\dot{x}(s)\|^2 \mathrm{d}s \leq \mathcal{E}^p_{\lambda,\xi}(t_0).$$

Thus, in the case $\alpha > 3$ we recover the well-known estimate (see for example [5, 6])

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 \mathrm{d}t < +\infty.$$

(2) Let us now assume $\alpha \leq 3$. Let us slightly modify the choices concerning the hypotheses $(\mathbf{H}_1)$ to $(\mathbf{H}_5)$ which have been made in the proof of the Theorem 2.1. We always want to maintain the property $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \leq 0$, and $\xi(\cdot) \geq 0$. Let us modify the condition $(\mathbf{H}_1)$ as follows: take $\lambda(t) = \mu t^{p-1}$, with $\mu > 0$ and

(i) $\mu \geq 2p$,

so as to have $t^p\left[2pt^{p-1} - \lambda(t)\right] \leq 0$. Then, condition (H_2) yields $\xi(t) = \mu t^{2p-2}(\alpha + 1 - 2p - \mu)$. Having $\xi(\cdot) \geq 0$ gives

(ii) $\mu \leq \alpha + 1 - 2p$.

We now have $\beta(t) = \lambda(t)\dot{\lambda}(t) + \dfrac{\dot{\xi}(t)}{2} = -\mu(1-p)(\alpha + 1 - 2p)t^{2p-3}$. We want $\beta(\cdot)$ to be less or equal than zero. By (ii) we already have $\alpha + 1 - 2p \geq \mu > 0$. Thus we only need to impose the supplementary condition on p:

(iii) $p \leq 1$.

With the above choices $(i), (ii), (iii)$ inequality (2.2) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) + t^p\left[(\alpha - p)t^{p-1} - \lambda(t)\right]\|\dot{x}(t)\|^2 \leq 0.$$

Since $\lambda(t) = \mu t^{p-1}$, we have equivalently

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) + \left[\alpha - p - \mu\right]t^{2p-1}\|\dot{x}(t)\|^2 \leq 0. \tag{2.7}$$

Hence, assuming

(iv) $\alpha - p - \mu > 0$,

by integration of (2.7) we obtain

$$\int_{t_0}^{+\infty} t^{2p-1}\|\dot{x}(t)\|^2 \mathrm{d}t < +\infty. \tag{2.8}$$

Let us now examine for which values of p the conditions (i) to (iv) are compatible. Firstly, (i) and (ii) give $2p \leq \mu \leq \alpha + 1 - 2p$. To ensure that these inequalities are compatible leads us to make the hypothesis (*v*) $p \leq \frac{\alpha+1}{4}$.

Take as a value of μ the middle point of the non void interval $[2p, \alpha+1-2p]$, that is

$$\mu = \frac{\alpha+1}{2}.$$

Then the condition $\alpha - p - \mu > 0$ become

(*vi*) $p < \frac{\alpha-1}{2}$.

Putting together the conditions (*iii*), (*v*), and (*vi*), we finally obtain the condition on p

$$0 < p < \min\left(1, \frac{\alpha+1}{4}, \frac{\alpha-1}{2}\right).$$

A close look at this formula shows that for $1 < \alpha \leq 3$ it is equivalent to $0 < p < \frac{\alpha-1}{2}\cdot$ This gives $2p-1 < \alpha-2$, which combined with (2.8) gives the result. □

Theorem 2.6 (Pointwise estimates of the speed)**.** *Let $x : [t_0; +\infty[\to \mathcal{H}$ be a solution of (AVD)$_\alpha$.*

1. *If $\alpha \geq 3$, the trajectory satisfies*

$$\sup_{t \geq t_0} \|x(t)\| < +\infty \quad \textit{and} \quad \|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t}\right).$$

2. *If $\alpha > 3$, we have*

$$\|\dot{x}(t)\| = o\left(\frac{1}{t}\right).$$

3. *If $1 \leq \alpha \leq 3$, we have for all $p < \frac{\alpha-1}{2}$*

$$\|\dot{x}(t)\| = \mathcal{O}\left(\frac{1}{t^p}\right).$$

Proof. (1) Let us keep the same choice of the different parameters as that done in the proof of Theorem 2.1, that is $\lambda(t) = 2pt^{p-1}$, $\xi(t) = 2(\alpha - 4p + 1)pt^{2(p-1)}$, and $p = \min(1, \frac{\alpha}{3})$. As a consequence $\mathcal{E}^p_{\lambda,\xi}(\cdot)$ is nonincreasing, and since $\xi(\cdot)$ is nonnegative, we deduce that, for all $t \geq t_0$

$$\mathcal{E}^p_{\lambda,\xi}(t_0) \geq \mathcal{E}^p_{\lambda,\xi}(t) \geq \frac{1}{2}\|2pt^{p-1}(x(t) - z) + t^p\dot{x}(t)\|^2. \tag{2.9}$$

After developing the above quadratic term, and neglecting the nonnegative term $t^{2p}\|\dot{x}(t)\|^2$ we obtain

$$2p^2t^{2p-2}\|x(t) - z\|^2 + 2pt^{2p-1}\langle x(t) - z, \dot{x}(t)\rangle \leq \mathcal{E}^p_{\lambda,\xi}(t_0).$$

Setting $h(t) := \|x(t) - z\|^2$, we have equivalently

$$2pt^{2p-2}h(t) + t^{2p-1}\dot{h}(t) \leq \frac{1}{p}\mathcal{E}^p_{\lambda,\xi}(t_0).$$

Then note that

$$\frac{\mathrm{d}}{\mathrm{d}t}t^{2p-1}h(t) = (2p-1)t^{2p-2}h(t) + t^{2p-1}\dot{h}(t) \leq 2pt^{2p-2}h(t) + t^{2p-1}\dot{h}(t).$$

Combining the two above inequalities we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}t^{2p-1}h(t) \leq \frac{1}{p}\mathcal{E}^p_{\lambda,\xi}(t_0),$$

which by integration gives

$$\|x(t)-z\|^2 \leq \frac{t_0^{2p-1}}{t^{2p-1}}\|x(t_0)-z\|^2 + \frac{t-t_0}{pt^{2p-1}}\mathcal{E}^p_{\lambda,\xi}(t_0).$$

Hence for $2p-1 \geq 1$, *i.e.*, $p \geq 1$ the trajectory remains bounded. Since $p = \min(1, \frac{\alpha}{3})$, this is equivalent to suppose $\alpha \geq 3$, and $p = 1$. Returning to (2.9), and by the triangle inequality we immediately infer

$$t\|\dot{x}(t)\| \leq \sqrt{2\mathcal{E}^1_{\lambda,\xi}(t_0)} + 2\|x(t)-z\|.$$

Hence

$$\|\dot{x}(t)\| \leq \frac{\sqrt{2\mathcal{E}^1_{\lambda,\xi}(t_0)}}{t} + \frac{2}{t}\sup_{t\geq t_0}\|x(t)-z\|$$

with $\lambda = 2$, $\xi = 2(\alpha-3)$. As a result, $\|\dot{x}(t)\| = \mathcal{O}(\frac{1}{t})$, which gives us item (1).

The next part of the proof is an adaptation to our framework of the technique developed by Attouch-Peypouquet in [6]. It is based on the decay properties of the scaled energy function

$$\Gamma(t) := t^{2p}W(t),$$

where W is the global energy defined by

$$W(t) := \Phi(x(t)) - \min_{\mathcal{H}}\Phi + \frac{1}{2}\|\dot{x}(t)\|^2.$$

By a direct application of the derivation chain rule, and using equation $(\mathrm{AVD})_\alpha$, we have

$$\dot{W}(t) = -\frac{\alpha}{t}\|\dot{x}(t)\|^2.$$

Hence

$$\dot{\Gamma}(t) = 2pt^{2p-1}W(t) + t^{2p}\dot{W}(t) = (p-\alpha)t^{2p-1}\|\dot{x}(t)\|^2 + 2pt^{2p-1}(\Phi(x(t)) - \min_{\mathcal{H}}\Phi).$$

Taking $p \leq \alpha$, we obtain

$$\dot{\Gamma}(t) \leq 2pt^{2p-1}(\Phi(x(t)) - \min_{\mathcal{H}}\Phi). \tag{2.10}$$

Let us successively examine the case $\alpha > 3$ and $\alpha \leq 3$.

(2) Suppose $\alpha > 3$. In that case we take $p = 1$, and hence

$$\Gamma(t) := t^2 W(t).$$

Since $p = 1 \leq 3$, the inequality (2.10) gives

$$\dot{\Gamma}(t) \leq 2t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi).$$

By Theorem 2.4 item *i*), we have

$$\int_{t_0}^{+\infty} t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \mathrm{d}t < +\infty.$$

Hence, $\left[\dot{\Gamma}\right]^+ \in L^1(t_0, +\infty)$. Since Γ is nonnegative, this implies that $\lim \Gamma(t)$ exists. On the other hand,

$$\frac{1}{t}\Gamma(t) = tW(t) = t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + \frac{t}{2}\|\dot{x}(t)\|^2.$$

By Theorem 2.5 item (1), we have $t\|\dot{x}(t)\|^2 \in L^1(t_0, +\infty)$. Applying again Theorem 2.4 item *i*) we have $t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \in L^1(t_0, +\infty)$. By combining these results, we deduce that $\frac{1}{t}\Gamma(t) \in L^1(t_0, +\infty)$. The function Γ verifies $\frac{1}{t}\Gamma(t) \in L^1(t_0, +\infty)$ and $\lim \Gamma(t)$ exists. Hence, $\lim \Gamma(t) = 0$. Equivalently $\lim t^2 \|\dot{x}(t)\|^2 = 0$, which gives the claim.

3) Now suppose $\alpha \leq 3$. In this case, we follow an argument similar to the one above but with a parameter $p \leq \alpha$ which will be chosen during the proof in a convenient way. Let us return to (2.10). By Theorem 2.4, we have $t^{2p-1}(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \in L^1(t_0, +\infty)$ for $p \leq 1$ and $p < \frac{\alpha}{3}$. As a consequence, under these two conditions on p we have $\left[\dot{\Gamma}\right]^+ \in L^1(t_0, +\infty)$. Since Γ is nonnegative, this implies that $\lim \Gamma(t)$ exists. On the other hand,

$$\frac{1}{t}\Gamma(t) = t^{2p-1}(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + \frac{t^{2p-1}}{2}\|\dot{x}(t)\|^2.$$

Applying again Theorem 2.4, we have $t^{2p-1}(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \in L^1(t_0, +\infty)$ for $p \leq 1$ and $p < \frac{\alpha}{3}$. Moreover, by Theorem 2.5 item (2), we have $t^{2p-1}\|\dot{x}(t)\|^2 \in L^1(t_0, +\infty)$ for $2p - 1 < \alpha - 2$, that is $p < \frac{\alpha-1}{2}$. Putting all these conditions together we have obtained that for $p < \min(1, \frac{\alpha}{3}, \frac{\alpha-1}{2}, \alpha)$, $\lim \Gamma(t)$ exists, and $\frac{1}{t}\Gamma(t) \in L^1(t_0, +\infty)$. Then observe that $\min(1, \frac{\alpha}{3}, \frac{\alpha-1}{2}, \alpha) = \frac{\alpha-1}{2}$ when $\alpha \leq 3$. As a consequence, when $\alpha \leq 3$ we have $\lim \Gamma(t) = 0$ for all $p < \frac{\alpha-1}{2}$. Hence $\lim t^{2p}\|\dot{x}(t)\|^2 = 0$ for all $p < \frac{\alpha-1}{2}$ which is the claim (it is equivalent to state the result with $\mathcal{O}$ or small o). □

As a direct consequence of $\lim \Gamma(t) = \lim t^2 W(t) = 0$, in the case $\alpha > 3$ we recover the following result of Attouch-Peypouquet [6].

Corollary 2.7. *For $\alpha > 3$, and for any solution trajectory $x(\cdot)$ of $(AVD)_\alpha$ we have*

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^2}\right).$$

We can now complete the table giving a synthetic view of the rate of convergence for the values and the speed of the solution trajectories of $(\mathrm{AVD})_\alpha$.

α	$\alpha < 3$	$\alpha = 3$	$\alpha > 3$
$\Phi(x(t)) - \min_{\mathcal{H}} \Phi$	$\mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$	$\mathcal{O}\left(\frac{1}{t^2}\right)$	$o\left(\frac{1}{t^2}\right)$
$I_p := \int_{t_0}^{+\infty} t^p(\Phi(x(t)) - \min \Phi)\mathrm{d}t$	$I_p < +\infty \quad \forall p < \frac{2\alpha}{3} - 1$	$I_p < +\infty \quad \forall p < 1$	$I_1 < +\infty$
$\dot{x}(t)$	$\mathcal{O}\left(\frac{1}{t^p}\right) \quad \forall p < \frac{\alpha-1}{2}$	$\mathcal{O}\left(\frac{1}{t}\right)$	$o\left(\frac{1}{t}\right)$
$J_p := \int_{t_0}^{+\infty} t^p\|\dot{x}(t)\|^2\mathrm{d}t$	$J_p < +\infty \quad \forall p < \alpha - 2$	$J_p < +\infty \quad \forall p < 1$	$J_1 < +\infty$

3. Convergence of the trajectories

3.1. Weak convergence for $\alpha > 3$

Let us recall the convergence result in the case $\alpha > 3$ obtained by Attouch-Chbani-Peypouquet-Redont [5] and May [23] in the case $\alpha > 3$. We give a brief demonstration of it, since it can be obtained as a direct consequence of our previous results, and this enlights the situation in the case $\alpha \leq 3$.

Theorem 3.1. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex continuously differentiable function such that* $\operatorname{argmin}\Phi$ *is nonempty. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a classical global solution of $(AVD)_\alpha$ with $\alpha > 3$. Then $x(t)$ converges weakly, as $t \to +\infty$ to a point in* $\operatorname{argmin}\Phi$.

Proof. The proof is based on the Opial's Lemma 6.1. By elementary calculus, convexity of Φ, and equation $(\mathrm{AVD})_\alpha$, one can first establish that for any $z \in \operatorname{argmin}\Phi$, the function $h_z(t) := \frac{1}{2}\|x(t) - z\|^2$ satisfies

$$t\ddot{h}_z(t) + \alpha\dot{h}_z(t) \leq t\|\dot{x}(t)\|^2. \tag{3.1}$$

By integrating the differential inequality (3.1), and using the estimate $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2\mathrm{d}t < +\infty$ given in Theorem 2.5 in the case $\alpha > 3$, we infer

$$\left[\dot{h}_z\right]^+ \in L^1(t_0, +\infty). \tag{3.2}$$

Since h_z is nonegative, this implies the convergence of h_z. The second item of the Opial's lemma is a direct consequence of the minimizing property of the trajectory, and of the convexity of Φ. □

3.2. Critical case

α**=3**. As we have already pointed out, the convergence of the trajectories of $(\mathrm{AVD})_\alpha$ in the case $\alpha = 3$ remains a widely open question. In the one-dimensional setting, and in the case $\alpha = 3$, as a main result, in this section, we prove the convergence of $(\mathrm{AVD})_\alpha$ trajectories, without any restrictive assumption on the convex potential Φ. This comes as a generalization of Cabot-Engler-Gadat ([15], Thm. 3.1), who obtained this result under the additional hypothesis that Φ is quadratically conditioned with respect to $S = \operatorname{argmin}\Phi$. We first establish some preliminary results of independent interest, and that will be useful in proving the convergence result in the one-dimensional framework.

Proposition 3.2. *Let $\mathcal{H}$ be a finite dimensional Hilbert space, and let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex continuously differentiable function such that $S = \operatorname{argmin}\Phi \neq \emptyset$. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a solution of $(AVD)_\alpha$ with $\alpha = 3$. Then,* *i)* x *is bounded and* $\lim_{t\to+\infty} \operatorname{dist}(x(t), S) = 0$.

ii) Moreover, if $x(t) \in S$ *for* t *large enough, then* $x(t)$ *converges to a point in* S.

Proof. By Theorem 2.1

$$\lim_{t\to+\infty} \Phi(x(t)) = \min \Phi.$$

By Theorem 2.6, for $\alpha \geq 3$, and in particular for $\alpha = 3$, we have

$$\sup_{t\geq t_0} \|x(t)\| < +\infty.$$

Then the result follows from a classical topological argument. If $\operatorname{dist}(x(t), S)$ fails to converge to zero, this implies the existence of a sequence $t_n \to +\infty$ and $\epsilon > 0$ such that, for all $n \in \mathbb{N}$, $\operatorname{dist}(x(t_n), S) \geq \epsilon$. Since $(x(t_n))$ is bounded, and $\mathcal{H}$ is a finite-dimensional Hilbert space, after extracting a convergent subsequence $x(t_{n_k}) \to \bar{x}$, we obtain $\operatorname{dist}(\bar{x}, S) \geq \epsilon$ and $\bar{x} \in S$ (a consequence of $x(t)$ minimizing), a clear contradiction.
Suppose moreover that $x(t) \in S$ for $t \geq t_1$. Hence, $\nabla\Phi(x(t)) = 0$, which by $(\text{AVD})_\alpha$ gives $\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = 0$. Equivalently, $\frac{\mathrm{d}}{\mathrm{d}t}(t^\alpha \dot{x}(t)) = 0$, which gives $\dot{x}(t) = \frac{C}{t^\alpha}$. Hence for $\alpha > 1$, and in particular for $\alpha = 3$, $\dot{x}$ is integrable, which implies the convergence of the trajectory. □

The following result is valid in a general Hilbert space, and for any $\alpha > 0$. It plays a key role in the proof of the convergence result.

Proposition 3.3. *Let* $\mathcal{H}$ *be a Hilbert space,* $\alpha \leq 3$*, and* $x(\cdot)$ *be a trajectory of* $(AVD)_\alpha$*. Suppose that for some* $t_2 \geq t_1$

$$x(t_1) = x(t_2) \in S = \operatorname{argmin} \Phi.$$

Then,

$$t_2^{\frac{\alpha}{3}} \|\dot{x}(t_2)\| \leq t_1^{\frac{\alpha}{3}} \|\dot{x}(t_1)\|.$$

In particular, for $\alpha = 3$*,*

$$t_2\|\dot{x}(t_2)\| \leq t_1\|\dot{x}(t_1)\|.$$

Proof. Set $z = x(t_1) = x(t_2) \in S = \operatorname{argmin} \Phi$, take $p = \min(1, \frac{\alpha}{3})$, and consider the function $\mathcal{E}^p_{\lambda,\xi} : [t_0, +\infty[\to \mathbb{R}^+$

$$\mathcal{E}^p_{\lambda,\xi}(t) = t^{2p}\left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|\lambda(t)(x(t) - z) + t^p\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z\|^2 \tag{3.3}$$

which serves as a Lyapunov function in the proof of Theorem 2.1. It is nonincreasing, and hence $\mathcal{E}^p_{\lambda,\xi}(t_2) \leq \mathcal{E}^p_{\lambda,\xi}(t_1)$. Because of $z = x(t_1) = x(t_2) \in S = \operatorname{argmin} \Phi$, this is equivalent to

$$\frac{1}{2}\|t_2^p \dot{x}(t_2)\|^2 \leq \frac{1}{2}\|t_1^p \dot{x}(t_1)\|^2.$$

By $p = \min(1, \frac{\alpha}{3})$, we get the announced result. □

The idea of the demonstration of the following Theorem is due to P. Redont (personal communication).

Theorem 3.4. *Take $\mathcal{H} = \mathbb{R}$, and let $\Phi : \mathbb{R} \to \mathbb{R}$ be a convex continuously differentiable function such that $S = \operatorname{argmin} \Phi \neq \emptyset$. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a solution of $(AVD)_\alpha$ with $\alpha = 3$. Then $x(t)$ converges, as $t \to +\infty$, to a point in S.*

Proof. Recall that, by Theorem 2.6, for $\alpha = 3$, the trajectory is bounded, and minimizing. As a consequence, when $\operatorname{argmin} \Phi$ is reduced to a singleton x^*, it is the unique cluster point of the trajectory, which implies the convergence of the trajectory to x^*. Thus we consider the case where $\operatorname{argmin} \Phi$ is an interval of positive length (possibly infinite), let $\operatorname{argmin} \Phi = [a, b]$. We consider the case where a and b are finite. The argument works in the same way when one of them, or both, is infinite. There are three possible cases:

- There exists $T \geq t_0$ such that $x(t) \geq b$ for all $t \geq T$. Then b is the unique cluster point of the trajectory, which implies the convergence of the trajectory to b. Symetrically, if there exists $T \geq t_0$ such that $x(t) \leq a$, for all $t \geq T$, then the trajectory converges to a.
- There exists $T \geq t_0$ such that, for all $t \geq T$, $a \leq x(t) \leq b$. Then, the convergence of the trajectory is a consequence of Proposition 3.2 (*ii*).
- Let us now observe that each time the trajectory enters the interval $[a, b]$ there are two possible cases. Either it remains in the interval $[a, b]$, then it converges, a direct consequence of the integration of $t\ddot{x}(t) + \alpha\dot{x}(t) = 0$ and $\alpha > 1$, or it ultimely goes out of the interval $[a, b]$. Note that when the trajectory enters the point a, it necessarily comes out in b. Otherwise if there is a loop around a, there would be two consecutive times t_1 (entering time) and t_2 (exit time) such that $\dot{x}(t_1) > 0$ and $\dot{x}(t_2) < 0$, which by a continuity argument would imply the existence of an intermediate time t such that $\dot{x}(t) = 0$. This forces the trajectory to stop definitively at time t, during the loop, a clear contradiction. If this scenario occurs a finite number of times, then given the last time it occurs, we conclude with the previous arguments. So, it remains to consider the case, which is the most delicate to analyze, where the trajectory passes in a and b an infinite number of times. Indeed, we will see that this is impossible. The argument is based on the analysis of the decay of the quantity $t|\dot{x}(t)|$ during a loop. Let $s_n \leq t_n \leq u_n \leq v_n$ be consecutives times such that $x(s_n) = a$, $x(t_n) = b$, and $a \leq x(t) \leq b$ for all $t \in [s_n, t_n]$, $x(u_n) = b$, $x(v_n) = a$ and $a \leq x(t) \leq b$ for all $t \in [u_n, v_n]$. For $t \in [s_n, t_n]$ we have $t\ddot{x}(t) + \alpha\dot{x}(t) = 0$. Equivalently

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(t\dot{x}(t)\right) + (\alpha - 1)\dot{x}(t) = 0. \tag{3.4}$$

  After integration of (3.4) on the interval $[s_n, t_n]$, we obtain

$$t_n\dot{x}(t_n) - s_n\dot{x}(s_n) = -(\alpha - 1)(b - a).$$

  Taking account of the sign of the derivative of x (positive at s_n, since the trajectory enters the interval $[s_n, t_n]$, positive at t_n since it leaves the interval), we have

$$|t_n\dot{x}(t_n)| = |s_n\dot{x}(s_n)| - (\alpha - 1)(b - a).$$

  Symetrically,

$$|v_n\dot{x}(v_n)| = |u_n\dot{x}(u_n)| - (\alpha - 1)(b - a).$$

  By Proposition 3.3 we have

$$|u_n\dot{x}(u_n)| \leq |t_n\dot{x}(t_n)|.$$

Combining the above relations, we obtain

$$|v_n\dot{x}(v_n)| \leq |s_n\dot{x}(s_n)| - 2(\alpha-1)(b-a).$$

Therefore, each time the trajectory returns to a after passing b, the quantity $t|\dot{x}(t)|$ decreases by a fixed positive quantity. This excludes the possibility of an infinite number of loops, and thus gives the contradiction.

□

Remark 3.5. Based on the convexity assumption on Φ, and without any further geometrical assumption on Φ, the convergence of the trajectories in the case $\alpha \leq 3$ is still an open problem. The estimates we obtained in the previous sections did not allow us to conclude. For example, let us consider the case $\alpha = 3$. Then, by Theorem 2.5

$$\int_{t_0}^{+\infty} t^{1-\epsilon}\|\dot{x}(t)\|^2 \mathrm{d}t < +\infty \quad \text{for all} \quad \epsilon > 0. \tag{3.5}$$

Let us examine how to exploit this information in the integration of inequation (3.1). First, after multiplication of (3.1) by $t^{\alpha-1}$ and integration we obtain

$$\dot{h}_z(t) \leq \frac{C}{t^\alpha} + \frac{1}{t^\alpha}\int_{t_0}^{t} s^\alpha \|\dot{x}(s)\|^2 \mathrm{d}s.$$

After division by t^ϵ, we get

$$\frac{1}{t^\epsilon}\dot{h}_z(t) \leq \frac{C}{t^{\alpha+\epsilon}} + \frac{1}{t^{\alpha+\epsilon}}\int_{t_0}^{t} s^\alpha \|\dot{x}(s)\|^2 \mathrm{d}s.$$

Integrating the above inequality, and using (3.5) with Fubini theorem gives, for all $\epsilon > 0$

$$\left[\frac{1}{t^\epsilon}\dot{h}_z\right]^+ \in L^1(t_0,+\infty). \tag{3.6}$$

Moreover, by Theorem 2.6 item (1), we know that the trajectory, and therefore h_z is bounded and nonnegative. It may be asked whether these properties are sufficient to conclude that h_z converges, at least ergodically. Unfortunately, the answer to this question is negative. Take for example

$$h(t) = 1 + \sin(\ln(t)).$$

Then, $0 \leq h(t) \leq 2$ and $\dot{h}(t) = \frac{\cos(\ln(t))}{t}$. Hence $|\frac{1}{t^\epsilon}\dot{h}_z(t)| \leq \frac{1}{t^{1+\epsilon}}$, and (3.6) is satisfied. But $h(t)$ fails to converges, even ergodically. Indeed, a direct computation gives

$$\frac{1}{t-t_0}\int_{t_0}^{t}\sin(\ln(s))\mathrm{d}s \sim \frac{1}{2}(\sin(\ln(t) - \cos(\ln(t))$$

which clearly fails to converge as t goes to $+\infty$. Indeed, the proof of Theorem 3.1 uses precisely (3.6) for $\epsilon = 0$.

Thus, it is natural to make additional geometric assumptions on Φ which guarantee the convergence of the trajectories of $(\mathrm{AVD})_\alpha$. Of particular interest is the strong convergence property which occurs for example in the presence of strong convexity. That is the situation we are now considering. For other cases of strong convergence (solution set with non-empty interior, even functions) one can consult [2, 5].

3.3. Strong convergence

In this subsection, $\mathcal{H}$ is a general Hilbert space, and we assume that the convex function Φ has a *strong minimum*, *i.e.*, there exist $x^* \in \mathcal{H}$ and $\mu > 0$ such that for every $x \in \mathcal{H}$,

$$\Phi(x) \geq \Phi(x^*) + \frac{\mu}{2}\|x - x^*\|^2. \tag{3.7}$$

This implies clearly that $\operatorname{argmin} \Phi = \{x^*\}$. Under this condition, we are able to show the strong convergence of the trajectories to this unique minimum, and determine precisely the decay rate of the energy W along the trajectories. Indeed, an interesting property that has been put recentenly to the fore in [2, 5, 35] is that the convergence rates increase indefinitely with larger values of α for these functions. The following theorem completes these results by examinating also the case $\alpha \leq 3$ and gives a synthetic view of this situation. For simplicity, we assume that the inequality (3.7) is satisfied for every $x \in \mathcal{H}$, but all that follows can be readily extended to the case where this inequality is satisfied only in a neighborhood of x^*.

An important instance for which (3.7) holds is the class of strongly convex functions. We recall that the function $\Phi : \mathcal{H} \to \mathbb{R}$ is *strongly convex*, if there exists a positive constant μ such that

$$\Phi(y) \geq \Phi(x) + \langle \nabla\Phi(x), y - x\rangle + \frac{\mu}{2}\|x - y\|^2$$

for all $x, y \in \mathcal{H}$. Of course if x^* is a minimizer of Φ, then $\nabla\Phi(x^*) = 0$ and we recover (3.7) from the above inequality.

Theorem 3.6. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex function which admits a strong minimum $x^* \in \mathcal{H}$. Let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution trajectory of $(AVD)_\alpha$ with $\alpha > 0$. Then $x(t)$ converges strongly, as $t \to +\infty$, to the unique element $x^* \in \operatorname{argmin} \Phi$. Moreover*

(i) $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$

(ii) $\|x(t) - x^*\|^2 = \mathcal{O}\left(t^{-\frac{2}{3}\alpha}\right)$,

(iii) $\|\dot{x}(t)\| = \mathcal{O}(t^{-\frac{1}{3}\alpha})$.

Proof. The case $\alpha > 3$ has already been analyzed in the articles mentioned above, see for example ([5], Thm. 3.4.) The case $\alpha \leq 3$ is a direct consequence of Theorems 2.1 and 2.5, and of the property of strong minimum of x^*. Indeed, we have

$$\|x(t) - x^*\|^2 \leq \frac{2}{\mu}\left(\Phi(x(t)) - \Phi(x^*)\right).$$

The first two items then follow from Theorem 2.1 and the above inequality. As a consequence $x(t)$ converges strongly to x^*, and so is bounded. Hence, we can apply Theorem 2.5, which gives the third item. ☐

Remark 3.7. The formulas for the rate of decay are the same in the cases $\alpha \leq 3$, and $\alpha \geq 3$. It is not such surprising, since the proof used in ([5], Thm. 3.4) in the case $\alpha > 3$ is based on the following Lyapunov function

$$\mathcal{E}^p_\lambda(t) := t^p\left(t^2(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2\right),$$

whose structure is quite similar to ours.

Remark 3.8. A model example of strongly convex function is $\Phi(x) = \frac{1}{2}\|x\|^2$, which is a positive definite quadratic function. In this case, one can compute explicitely the solution trajectories of (1.1) with the help of the Bessel functions. Indeed the solution of

$$(\mathrm{AVD})_\alpha \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + x(t) = 0$$

with Cauchy data $x(0) = x_0$, $\dot{x}(0) = 0$ is given by

$$x(t) = 2^{\frac{\alpha-1}{2}} \Gamma\left(\frac{\alpha+1}{2}\right) \frac{J_{\frac{\alpha-1}{2}}(t)}{t^{\frac{\alpha-1}{2}}} x_0.$$

In the above formula $J_{\frac{\alpha-1}{2}}(\cdot)$ is the first kind Bessel function of order $\frac{\alpha-1}{2}$. For large t,

$$J_\alpha(t) = \sqrt{\tfrac{2}{\pi t}}\left(\cos\left(t - \tfrac{\pi\alpha}{2} - \tfrac{\pi}{4}\right) + \mathcal{O}(\tfrac{1}{t})\right).$$

Hence

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(t^{-\alpha}).$$

One can compare with the rate $\mathcal{O}(t^{-\frac{2}{3}\alpha})$, which is valid for arbitrarily strongly convex functions. One can also conclude from this example that, when $0 < \alpha \leq 2$, *the parameter of the optimal rate of convergence (in the worst case) is between* $\frac{2\alpha}{3}$ *and* α. Indeed, Aujol-Dossal in [10] recently gave a complete answer to this question by showing that, in the subcritical case $\alpha < 3$, the convergence rate $\mathcal{O}(t^{-\frac{2}{3}\alpha})$ is optimal (the best guaranteed in the worst case).

Remark 3.9. The strong convexity property is a particular instance of the Kurdyka-Lojasiewicz property. It would be interesting to examine the problem of convergence for convex functions satisfying this property. One can consult Bégout-Bolte-Jendoubi [13] for a recent study of damped second-order gradient systems under this geometric assumption.

4. Inertial Forward-Backward algorithms

Our study aims at complementing in the subcritical case $\alpha < 3$ the results on the rate of convergence of the inertial forward-backward methods, and of the FISTA type algorithms. These splitting algorithms aim at solving structured optimization problems

$$\min\{\Phi(x) + \Psi(x) : \ x \in \mathcal{H}\} \tag{4.1}$$

where $\Phi : \mathcal{H} \to \mathbb{R}$ is a continuously differentiable convex function whose gradient is Lipschitz continuous, and $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper lower-semicontinuous convex function. We set $\Theta := \Phi + \Psi$, which is the convex lower-semicontinuous function to minimize. Based on the link between continuous dynamical systems and algorithms, we are naturally led to extend the dynamics studied in the previous sections, and to consider the differential inclusion

$$(\mathrm{AVD})_\alpha \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) + \partial\Psi(x(t)) \ni 0, \tag{4.2}$$

where $\partial\Psi$ is the subdifferential of Ψ in the sense of convex analysis. A detailed study of this differential inclusion goes far beyond the scope of the present article. See [4] for some results in the case of a fixed positive damping parameter, and [7] for some recent results in the case of a vanishing damping coefficient. However, given the validity of the subdifferential inequality for convex functions, the (generalized) chain rule for derivatives over curves (see [14]), most of the results presented in the previous sections can be transposed to this more general context. Thus, the results obtained in the continuous case will only serve as guidelines for the study of the associated algorithms that we present now.

As shown in [5, 35], the time discretization of this system, implicit with respect to the nonsmooth term $\partial\Psi$, and explicit with respect to the smooth term $\nabla\Phi$, gives the the Inertial Forward-Backward algorithm

$$\text{(IFB)}_\alpha \quad \begin{cases} y_k = x_k + (1-\frac{\alpha}{k})(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k)). \end{cases}$$

Note that other types of discretization give slightly different inertial forward-backward algorithms, see [22], an interesting topic for further research. In the above formula, the parameter s enters as $s = h^2$, where h is the step size of the discretization, and $\text{prox}_{\gamma\Psi} : \mathcal{H} \to \mathcal{H}$ is the classical proximal operator. Recall that, for any $\gamma > 0$, for any $x \in \mathcal{H}$,

$$\text{prox}_{\gamma\Psi}(x) = \text{argmin}_{\xi\in\mathcal{H}} \left\{ \Psi(\xi) + \frac{1}{2\gamma}\|\xi - x\|^2 \right\} = (I + \gamma\partial\Psi)^{-1}(x). \tag{4.3}$$

This last equality expresses that $\text{prox}_{\gamma\Psi}$ is the resolvent of index γ of the maximal monotone operator $\partial\Psi$. One can consult [11, 28, 29, 30] for a recent account on the proximal-based splitting methods.

a) For $\alpha = 3$, we recover the classical FISTA algorithm developed by Beck-Teboulle [12], based on the acceleration method introduced by Nesterov [24] in the smooth case, and by Güler [19] in the proximal setting. The rate of convergence of this method is $\Theta(x_k) - \min_{\mathcal{H}} \Theta = \mathcal{O}\left(\frac{1}{k^2}\right)$, where k is the number of iterations. Convergence of the sequences generated by Nesterov's accelerated gradient method and of FISTA, has been an elusive question for decades.

b) For $\alpha > 3$, it was shown by Chambolle-Dossal [16] that each sequence generated by (IFB)_α converges weakly to an optimal solution, and Attouch-Peypouquet [6] obtained the improved rate of convergence $\Theta(x_k) - \min_{\mathcal{H}} \Theta = o(\frac{1}{k^2})$.

c) The case $\alpha < 3$ has remained largely unknown. Precisely, we will study this situation and show results parallel to those obtained in the continuous framework in the previous sections.

We will systematically assume the following set of hypotheses

$$(H) \begin{cases} \bullet\mathcal{H} \text{ is a real Hilbert space;} \\ \bullet\ \Phi : \mathcal{H} \to \mathbb{R} \text{ is convex, differentiable with } L\text{-Lipschitz continuous gradient;} \\ \bullet\ \Psi : \mathcal{H} \to \mathbb{R}\cup\{+\infty\} \text{ is convex, proper and lower semicontinuous;} \\ \bullet\ \Theta := \Phi + \Psi \text{ has a nonempty set of minimizers: } S = \text{argmin}\,\Theta \neq \emptyset; \\ \bullet \text{ The parameter } \alpha \text{ is positive;} \\ \bullet \text{ The parameter } s \text{ satisfies } s \in]0, 1/L]. \end{cases}$$

Our main result in the algorithmic framework is presented below.

Theorem 4.1. *Let us make assumptions (H). Let (x_k) be a sequence of iterates generated by algorithm (IFB)_α. Suppose $0 < \alpha \leq 3$. Then the following convergence result is verified: for all $p < \frac{2\alpha}{3}$, for all $k \geq 1$*

$$(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = \mathcal{O}\left(\frac{1}{k^p}\right).$$

4.1. The inertial forward-backward algorithm: classical facts

Let us first rewrite $(\text{IFB})_\alpha$ in a more compact way. Let us define the operator $G_s : \mathcal{H} \to \mathcal{H}$ by

$$G_s(y) = \frac{1}{s}\left(y - \text{prox}_{s\Psi}(y - s\nabla\Phi(y))\right),$$

and set $\alpha_k = 1 - \frac{\alpha}{k}$. Thus, the algorithm can be written in an equivalent way as

$$(\text{IFB})_\alpha \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - sG_s(y_k). \end{cases}$$

The following classical lemma plays a crucial role in the convergence analysis of the forward-backward algorithms. It is known as the descent rule for the forward-backward methods, see ([12], Lem. 2.3) , ([16], Lem. 1). Its validity relies essentially on the stepsize limitation $s \leq \frac{1}{L}$.

Lemma 4.2. *Assume hypothesis (H). Then*

(i) For all $x, y \in \mathcal{H}$

$$\Theta(y - sG_s(y)) \leq \Theta(x) + \langle G_s(y), y - x\rangle - \frac{s}{2}\|G_s(y)\|^2. \tag{4.4}$$

(ii) The operator G_s is monotone, and the following equivalences hold true

$$z \in \operatorname{argmin}\Theta \iff G_s(z) = 0.$$

4.2. Design of the Lyapunov function

Let us give a discrete version of the Lyapunov function

$$\mathcal{E}^p_{\lambda,\xi}(t) = t^{2p}\left[\Theta(x(t)) - \min_{\mathcal{H}}\Theta\right] + \frac{1}{2}\|\lambda(t)(x(t) - z) + t^p\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t) - z\|^2, \tag{4.5}$$

that has been used in the continuous case (where now the function to minimize is Θ). To that end, let us reformulate $\mathcal{E}^p_{\lambda,\xi}$ with the help of the energy function $W(t) = \Theta(x(t)) - \min_{\mathcal{H}}\Theta + \frac{1}{2}\|\dot{x}(t)\|^2$, and the anchor function $h_z(t) = \frac{1}{2}\|x(t) - z\|^2$. After development of (4.5), we get

$$\mathcal{E}^p_{\lambda,\xi}(t) = t^{2p}\left[\Theta(x(t)) - \min_{\mathcal{H}}\Theta + \frac{1}{2}\|\dot{x}(t)\|^2\right] + \frac{1}{2}\lambda(t)^2\|x(t) - z\|^2 + \lambda(t)t^p\langle x(t) - z, \dot{x}(t)\rangle + \frac{\xi(t)}{2}\|x(t) - z\|^2.$$

Hence,

$$\mathcal{E}^p_{\lambda,\xi}(t) = t^{2p}W(t) + \lambda(t)t^p\dot{h}(t) + \left(\lambda(t)^2 + \xi(t)\right)h(t). \tag{4.6}$$

Let us introduce the global energy (potential + kinetic) at stage $k \in \mathbb{N}^*$

$$W_k := \Theta(x_k) - \min\Theta + \frac{1}{2s}\|x_k - x_{k-1}\|^2,$$

and the anchor function to the solution set (where z denotes an element of S),

$$h_k := \frac{1}{2}\|x_k - z\|^2.$$

Thus, a natural candidate for the Lyapunov function in the discrete case would be

$$\mathcal{E}_k = s^p k^{2p} W_k + s^{\frac{p-1}{2}} \lambda_k k^p (h_k - h_{k-1}) + (\lambda_k^2 + \xi_k) h_{k-1}, \tag{4.7}$$

where λ_k and ξ_k are positive parameters that will be defined further. Let us verify that $\mathcal{E}_k$ is non-negative. Indeed,

$$\begin{aligned}\mathcal{E}_k &= s^p k^{2p}(\Theta(x_k) - \min \Theta) + s^{p-1}\frac{k^{2p}}{2}\|x_k - x_{k-1}\|^2 + \frac{1}{2} s^{\frac{p-1}{2}} \lambda_k k^p (\|x_k - z\|^2 - \|x_{k-1} - z\|^2) \\ &\quad + \frac{\lambda_k^2}{2}\|x_{k-1} - z\|^2 + \frac{\xi_k}{2}\|x_{k-1} - z\|^2\end{aligned}$$

From

$$\|x_k - z\|^2 - \|x_{k-1} - z\|^2 \geq 2\langle x_{k-1} - z, x_k - x_{k-1}\rangle$$

and $\xi_k \geq 0$ we infer

$$\begin{aligned}\mathcal{E}_k &\geq s^p k^{2p}(\Theta(x_k) - \min \Theta) + \frac{1}{2}\left[s^{p-1} k^{2p}\|x_k - x_{k-1}\|^2 + 2 s^{\frac{p-1}{2}} \lambda_k k^p \langle x_k - x_{k-1}, x_{k-1} - z\rangle + \lambda_k^2 \|x_{k-1} - z\|^2\right] \\ &= s^p k^{2p}(\Theta(x_k) - \min \Theta) + \frac{1}{2}\|\lambda_k (x_{k-1} - z) + s^{\frac{p-1}{2}} k^p (x_k - x_{k-1})\|^2.\end{aligned}$$

As a result, the condition $\xi_k \geq 0$ ensures that $\mathcal{E}_k$ is non-negative, and minorized by $s^p k^{2p}(\Theta(x_k) - \min \Theta)$. Thus, if we can prove that $\mathcal{E}_k$ is bounded from above, we will get the desired result. In the continuous case, we have proved that, with judicious choices of the parameters $\lambda(t)$ and $\xi(t)$ the function $\mathcal{E}^p_{\lambda,\xi}$ was non-increasing. The discretization induces certain additional terms that we must show to be negligible, which makes the proof more technical. Indeed, we will have to argue with the slightly modified function

$$\tilde{\mathcal{E}}_k := \mathcal{E}_k - \frac{p}{s^{1-p}}(k)^{2p-1}\|x_k - x_{k-1}\|^2$$

which involves a lower order correction term.

Our first lemmas are based on the study carried out by Attouch-Cabot [3] in the case of a general coefficient α_k. They analyze successively the rate of decay of the energy W_k and of the anchor functions h_k. Then, based on formula (4.7) we will put these results together.

4.3. Decay of the energy

Given a sequence of iterates (x_k) generated by algorithm $(\mathrm{IFB})_\alpha$, let us evaluate the decay of the energy $W_k := \Theta(x_k) - \min \Theta + \frac{1}{2s}\|x_k - x_{k-1}\|^2$.

Proposition 4.3. *Under hypothesis (H), let (x_k) be a sequence generated by algorithm $(\mathrm{IFB})_\alpha$. The energy sequence (W_k) satisfies for every $k \geq 1$,*

$$W_{k+1} - W_k \leq -\frac{1 - \alpha_k^2}{2s}\|x_k - x_{k-1}\|^2. \tag{4.8}$$

As a consequence, the sequence (W_k) is nonincreasing.

Proof. By applying formula (4.4) with $y = y_k$ and $x = x_k$, we obtain

$$\Theta(x_{k+1}) = \Theta(y_k - sG_s(y_k)) \le \Theta(x_k) + \langle G_s(y_k), y_k - x_k\rangle - \frac{s}{2}\|G_s(y_k)\|^2. \tag{4.9}$$

Reformulate the second member of (4.9) using the sequence (x_k). By definition of G_s and $(\text{IFB})_\alpha$ we have

$$\begin{aligned}\langle G_s(y_k), y_k - x_k\rangle &= -\frac{1}{s}\langle x_{k+1} - y_k, y_k - x_k\rangle\\ &= -\frac{1}{s}\langle x_{k+1} - x_k - \alpha_k(x_k - x_{k-1}), \alpha_k(x_k - x_{k-1})\rangle\\ &= -\frac{\alpha_k}{s}\langle x_{k+1} - x_k, x_k - x_{k-1}\rangle + \frac{\alpha_k^2}{s}\|x_k - x_{k-1}\|^2\end{aligned}$$

and

$$\begin{aligned}\|G_s(y_k)\|^2 &= \frac{1}{s^2}\|x_{k+1} - y_k\|^2 = \frac{1}{s^2}\|x_{k+1} - x_k - \alpha_k(x_k - x_{k-1})\|^2\\ &= \frac{1}{s^2}\|x_{k+1} - x_k\|^2 + \frac{\alpha_k^2}{s^2}\|x_k - x_{k-1}\|^2 - \frac{2\alpha_k}{s^2}\langle x_{k+1} - x_k, x_k - x_{k-1}\rangle.\end{aligned}$$

In view of (4.9), we obtain

$$\Theta(x_{k+1}) \le \Theta(x_k) + \frac{\alpha_k^2}{2s}\|x_k - x_{k-1}\|^2 - \frac{1}{2s}\|x_{k+1} - x_k\|^2,$$

which can be equivalently rewritten as (4.8). □

4.4. Anchor

As a fundamental tool, we will use the distance to equilibria to anchor the trajectory to the solution set $S = \operatorname{argmin}\Theta$. To this end, given $z \in \operatorname{argmin}\Theta$, recall that $h_k = \frac{1}{2}\|x_k - z\|^2$. The next result can be found for example in [3]. It plays a central role in our Lyapunov analysis of $(\text{IFB})_\alpha$.

Proposition 4.4. *Under (H), we have*

$$h_{k+1} - h_k - \alpha_k(h_k - h_{k-1}) \le \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 - s(\Theta(x_{k+1}) - \min\Theta). \tag{4.10}$$

Proof. Observe that

$$\begin{aligned}\|y_k - z\|^2 &= \|x_k + \alpha_k(x_k - x_{k-1}) - z\|^2\\ &= \|x_k - z\|^2 + \alpha_k^2\|x_k - x_{k-1}\|^2 + 2\alpha_k\langle x_k - z, x_k - x_{k-1}\rangle\\ &= \|x_k - z\|^2 + \alpha_k^2\|x_k - x_{k-1}\|^2\\ &\quad +\alpha_k\|x_k - z\|^2 + \alpha_k\|x_k - x_{k-1}\|^2 - \alpha_k\|x_{k-1} - z\|^2\\ &= \|x_k - z\|^2 + \alpha_k(\|x_k - z\|^2 - \|x_{k-1} - z\|^2) + (\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2\\ &= 2[h_k + \alpha_k(h_k - h_{k-1})] + (\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2.\end{aligned}$$

Setting briefly $H_k = h_{k+1} - h_k - \alpha_k(h_k - h_{k-1})$, we deduce that

$$\begin{aligned} H_k &= \frac{1}{2}\|x_{k+1} - z\|^2 - \frac{1}{2}\|y_k - z\|^2 + \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 \\ &= \left\langle x_{k+1} - y_k, \frac{1}{2}(x_{k+1} + y_k) - z \right\rangle + \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2 \\ &= \langle x_{k+1} - y_k, y_k - z\rangle + \frac{1}{2}\|x_{k+1} - y_k\|^2 + \frac{1}{2}(\alpha_k^2 + \alpha_k)\|x_k - x_{k-1}\|^2. \end{aligned}$$

Using the equality $x_{k+1} = y_k - sG_s(y_k)$, we obtain (4.10). Inequality (4.4) applied with $y = y_k$ and $x = z$ yields

$$\Theta(x_{k+1}) = \Theta(y_k - sG_s(y_k)) \leq \Theta(z) + \langle G_s(y_k), y_k - z\rangle - \frac{s}{2}\|G_s(y_k)\|^2.$$

Since $\Theta(z) = \min\Theta$, we infer that

$$-s\langle G_s(y_k), y_k - z\rangle + \frac{s^2}{2}\|G_s(y_k)\|^2 \leq -s(\Theta(x_{k+1}) - \min\Theta),$$

which completes the proof of Proposition 4.4. □

4.5. Proof of Theorem 4.1

Proof. Let us evaluate from above $\mathcal{E}_{k+1} - \mathcal{E}_k$, where $\mathcal{E}_k$ is given by (4.7). We have

$$\mathcal{E}_{k+1} - \mathcal{E}_k = s^p A_k + s^{\frac{p-1}{2}} B_k + C_k, \tag{4.11}$$

where

$$\begin{aligned} A_k &= (k+1)^{2p} W_{k+1} - k^{2p} W_k \\ B_k &= \lambda_{k+1}(k+1)^p(h_{k+1} - h_k) - \lambda_k k^p(h_k - h_{k-1}) \\ C_k &= (\lambda_{k+1}^2 + \xi_{k+1})h_k - (\lambda_k^2 + \xi_k)h_{k-1}. \end{aligned}$$

Let us examine successively each of the distinctive terms A_k, B_k, C_k. We have

$$A_k = ((k+1)^{2p} - k^{2p})W_{k+1} + k^{2p}(W_{k+1} - W_k). \tag{4.12}$$

When $p \geq \frac{1}{2}$ we have $(k+1)^{2p} - k^{2p} \leq 2p(k+1)^{2p-1}$, and when $p \leq \frac{1}{2}$ we have $(k+1)^{2p} - k^{2p} \leq 2pk^{2p-1}$. These are equivalent quantities, which lead to similar computations. Thus, in the following we suppose that $p \geq \frac{1}{2}$. The analysis in the case $p \leq \frac{1}{2}$ is quite similar. Hence we have $(k+1)^{2p} - k^{2p} \leq 2p(k+1)^{2p-1}$. From Proposition 4.3 we have $W_{k+1} - W_k \leq -\frac{1-\alpha_k^2}{2s}\|x_k - x_{k-1}\|^2$. Putting these two results together, and by definition of W_{k+1}, we infer

$$A_k \leq 2p(k+1)^{2p-1}(\Theta(x_{k+1}) - \min\Theta + \frac{1}{2s}\|x_{k+1} - x_k\|^2) - \frac{1 - \alpha_k^2}{2s} k^{2p}\|x_k - x_{k-1}\|^2. \tag{4.13}$$

Let us now consider B_k.

$$B_k = \lambda_{k+1}(k+1)^p(h_{k+1} - h_k) - \alpha_k(h_k - h_{k-1})) + (\alpha_k\lambda_{k+1}(k+1)^p - \lambda_k k^p)(h_k - h_{k-1}).$$

From Proposition 4.4, and $0 \leq \alpha_k \leq 1$, we infer

$$B_k \leq \lambda_{k+1}(k+1)^p \left(\|x_k - x_{k-1}\|^2 - s(\Theta(x_{k+1}) - \min \Theta)\right) + (\alpha_k \lambda_{k+1}(k+1)^p - \lambda_k k^p)(h_k - h_{k-1}). \tag{4.14}$$

Let us finally consider C_k.

$$C_k = (\lambda_{k+1}^2 + \xi_{k+1})(h_k - h_{k-1}) + ((\lambda_{k+1}^2 + \xi_{k+1}) - (\lambda_k^2 + \xi_k))h_{k-1}. \tag{4.15}$$

Let us put together (4.13), (4.14), and (4.15). We obtain

$$\begin{aligned}\mathcal{E}_{k+1} - \mathcal{E}_k \leq\ & 2ps^p(k+1)^{2p-1}(\Theta(x_{k+1}) - \min \Theta + \frac{1}{2s}\|x_{k+1} - x_k\|^2) - \frac{1-\alpha_k^2}{2s} s^p k^{2p} \|x_k - x_{k-1}\|^2 \\ & + \lambda_{k+1}(k+1)^p s^{\frac{p-1}{2}} \left(\|x_k - x_{k-1}\|^2 - s(\Theta(x_{k+1}) - \min \Theta)\right) \\ & + s^{\frac{p-1}{2}}(\alpha_k \lambda_{k+1}(k+1)^p - \lambda_k k^p)(h_k - h_{k-1}) \\ & + (\lambda_{k+1}^2 + \xi_{k+1})(h_k - h_{k-1}) + ((\lambda_{k+1}^2 + \xi_{k+1}) - (\lambda_k^2 + \xi_k))h_{k-1}.\end{aligned}$$

Let us reorganize the above expression in a parallel way to the formula below obtained in the continuous case, and that we recall below

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \quad \leq \quad & t^p\left[2pt^{p-1} - \lambda(t))\right](\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + \left[\xi(t) + t^p\dot{\lambda}(t) - (\alpha - p)t^{p-1}\lambda(t) + \lambda(t)^2\right]\dot{h}(t) \\ & -t^p\left[(\alpha - p)t^{p-1} - \lambda(t)\right]\|\dot{x}(t)\|^2 + \left[\lambda(t)\dot{\lambda}(t) + \tfrac{\dot{\xi}(t)}{2}\right]\|x(t) - z\|^2.\end{aligned}$$

We obtain

$$\begin{aligned}\mathcal{E}_{k+1} - \mathcal{E}_k \leq\ & (k+1)^p\left[2ps^p(k+1)^{p-1} - s^{\frac{p+1}{2}}\lambda_{k+1}\right](\Theta(x_{k+1}) - \min \Theta) \\ & + \left[\lambda_{k+1}^2 + \xi_{k+1} + s^{\frac{p-1}{2}}(\alpha_k \lambda_{k+1}(k+1)^p - \lambda_k k^p)\right](h_k - h_{k-1}) \\ & + \left[\lambda_{k+1}(k+1)^p s^{\frac{p-1}{2}} - \frac{1-\alpha_k^2}{2s}s^p k^{2p}\right]\|x_k - x_{k-1}\|^2 + \frac{p}{s^{1-p}}(k+1)^{2p-1}\|x_{k+1} - x_k\|^2 \\ & + \left[(\lambda_{k+1}^2 + \xi_{k+1}) - (\lambda_k^2 + \xi_k)\right]h_{k-1}.\end{aligned} \tag{4.16}$$

Let us analyze successively the four terms which enter the second member of (4.16):

1) Let us make the two first terms of the second member of (4.16) equal to zero by taking respectively

$$\lambda_k = \frac{2p}{s^{\frac{1-p}{2}}}k^{p-1} \tag{4.17}$$

and

$$\xi_{k+1} = -\lambda_{k+1}^2 - s^{\frac{p-1}{2}}(\alpha_k \lambda_{k+1}(k+1)^p - \lambda_k k^p).$$

This last equality gives equivalently

$$\xi_{k+1} = -\lambda_{k+1}^2 - s^{\frac{p-1}{2}}\alpha_k \lambda_{k+1}(k+1)^p + s^{\frac{p-1}{2}}\lambda_k k^p.$$

By (4.17) we deduce that

$$\begin{aligned}\xi_{k+1} &= -\frac{4p^2}{s^{1-p}}(k+1)^{2p-2} - \frac{2p}{s^{1-p}}(1-\frac{\alpha}{k})(k+1)^{2p-1} + \frac{2p}{s^{1-p}}k^{2p-1}\\ &= \frac{2p}{s^{1-p}}\left[-2p(k+1)^{2p-2} - (1-\frac{\alpha}{k})(k+1)^{2p-1} + k^{2p-1}\right] := \frac{2p}{s^{1-p}}D_k.\end{aligned}$$

Let us analyze the sign of D_k by making an asymptotic development.

$$\begin{aligned}D_k &= -2p(k+1)^{2p-2} - \left[(k+1)^{2p-1} - k^{2p-1}\right] + \frac{\alpha}{k}(k+1)^{2p-1} \quad (4.18)\\ &= -2pk^{2p-2}(1+\frac{1}{k})^{2p-2} - k^{2p-1}\left[(1+\frac{1}{k})^{2p-1} - 1\right] + \alpha k^{2p-2}(1+\frac{1}{k})^{2p-1}\\ &= -2pk^{2p-2}\left[1+\frac{2p-2}{k} + \mathrm{o}(\frac{1}{k})\right] - k^{2p-1}\left[\frac{2p-1}{k} + \mathrm{o}(\frac{1}{k})\right] + \alpha k^{2p-2}\left[1+\frac{2p-1}{k} + \mathrm{o}(\frac{1}{k})\right]\\ &\sim (\alpha - 4p + 1)k^{2p-2}.\end{aligned}$$

Hence, the condition $\xi_k \geq 0$ is satisfied for

$$p < \frac{\alpha+1}{4}.$$

It is the same as in the continuous case, except that now the inequality is assumed to be strict.

2) Let us now consider

$$\begin{aligned}F_k &:= \left[\lambda_{k+1}(k+1)^p s^{\frac{p-1}{2}} - \frac{1-\alpha_k^2}{2s}s^p k^{2p}\right]\|x_k - x_{k-1}\|^2 + \frac{p}{s^{1-p}}(k+1)^{2p-1}\|x_{k+1}-x_k\|^2\\ &= \left[\lambda_{k+1}(k+1)^p s^{\frac{p-1}{2}} - \frac{1-\alpha_k^2}{2s}s^p k^{2p} + \frac{p}{s^{1-p}}(k+1)^{2p-1}\right]\|x_k - x_{k-1}\|^2\\ &\quad + \frac{p}{s^{1-p}}(k+1)^{2p-1}\left[\|x_{k+1}-x_k\|^2 - \|x_k - x_{k-1}\|^2\right]\\ &\leq \left[\lambda_{k+1}(k+1)^p s^{\frac{p-1}{2}} - \frac{1-\alpha_k^2}{2s}s^p k^{2p} + \frac{p}{s^{1-p}}(k+1)^{2p-1}\right]\|x_k - x_{k-1}\|^2\\ &\quad + \frac{p}{s^{1-p}}\left[(k+1)^{2p-1}\|x_{k+1}-x_k\|^2 - k^{2p-1}\|x_k - x_{k-1}\|^2\right],\end{aligned}$$

where in the last inequality we have used $k^{2p-1} \leq (k+1)^{2p-1}$ for $p \geq \frac{1}{2}$. By (4.17), and the definition of α_k we have

$$\begin{aligned}\lambda_{k+1}(k+1)^p s^{\frac{p-1}{2}} - \frac{1-\alpha_k^2}{2s}s^p k^{2p} + \frac{p}{s^{1-p}}(k+1)^{2p-1} &\sim \frac{2p}{s^{\frac{1-p}{2}}}(k+1)^{p-1}(k+1)^p s^{\frac{p-1}{2}}\\ &\quad -\frac{1}{2s}\frac{2\alpha}{k}s^p k^{2p} + \frac{p}{s^{1-p}}(k+1)^{2p-1}\\ &\sim \frac{1}{s^{1-p}}\left[3p-\alpha\right]k^{2p-1}.\end{aligned}$$

As a consequence, for $p < \frac{\alpha}{3}$ we have

$$F_k \leq \frac{p}{s^{1-p}}\left[(k+1)^{2p-1}\|x_{k+1}-x_k\|^2 - k^{2p-1}\|x_k - x_{k-1}\|^2\right].$$

3) Let us finally consider

$$G_k := (\lambda_{k+1}^2 + \xi_{k+1}) - (\lambda_k^2 + \xi_k) = (\lambda_{k+1}^2 - \lambda_k^2) + (\xi_{k+1} - \xi_k).$$

Let us compute an equivalent of the discrete derivative $\lambda_{k+1} - \lambda_k$. We have

$$\begin{aligned}\lambda_{k+1} - \lambda_k &= \frac{2p}{s^{\frac{1-p}{2}}}\left((k+1)^{p-1} - k^{p-1}\right)\\ &= \frac{2p}{s^{\frac{1-p}{2}}} k^{p-1}\left(\left(1+\frac{1}{k}\right)^{p-1} - 1\right)\\ &\sim \frac{2p(p-1)}{s^{\frac{1-p}{2}}} k^{p-2}.\end{aligned}$$

Hence

$$\begin{aligned}\lambda_{k+1}^2 - \lambda_k^2 &= (\lambda_{k+1} + \lambda_k)(\lambda_{k+1} - \lambda_k)\\ &\sim \frac{2p(p-1)}{s^{\frac{1-p}{2}}} k^{p-2} \times 2\frac{2p}{s^{\frac{1-p}{2}}} k^{p-1}\\ &\sim \frac{8p^2(p-1)}{s^{1-p}} k^{2p-3}\end{aligned}$$

On the other hand, from $\xi_{k+1} = \frac{2p}{s^{1-p}} D_k$ and $D_k \sim (\alpha - 4p + 1)k^{2p-2}$ (see (4.18)) we immediately infer

$$\xi_{k+1} - \xi_k \sim \frac{2p(2p-2)}{s^{1-p}}(\alpha - 4p + 1)k^{2p-3}.$$

Hence (when the equivalent is not zero!)

$$\begin{aligned}G_k &\sim \frac{8p^2(p-1)}{s^{1-p}} k^{2p-3} + \frac{2p(2p-2)}{s^{1-p}}(\alpha - 4p + 1)k^{2p-3}\\ &= 4p(p-1)(\alpha - 2p + 1)k^{2p-3}.\end{aligned}$$

The same argument as in the continuous case leads us to assume $0 < p < 1$. The condition $\alpha - 2p + 1 > 0$ is already implied by the previous assumption $p \leq \frac{\alpha+1}{4}$.
We have now all the ingredients to conclude. By taking $p < \min\{1, \frac{\alpha}{3}\}$, we have obtained that for k sufficiently large

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq \frac{p}{s^{1-p}}(k+1)^{2p-1}\|x_{k+1} - x_k\|^2 - \frac{p}{s^{1-p}} k^{2p-1}\|x_k - x_{k-1}\|^2.$$

Equivalently $\mathcal{E}_k - \frac{p}{s^{1-p}} k^{2p-1}\|x_k - x_{k-1}\|^2$ is non-increasing, and hence is bounded from above. Returning to the definition of $\mathcal{E}_k$, this gives

$$\begin{aligned}&s^p k^{2p}(\Theta(x_k) - \min \Theta) + \frac{1}{2} s^{\frac{p-1}{2}} \lambda_k k^p(\|x_k - z\|^2 - \|x_{k-1} - z\|^2) + \frac{\lambda_k^2}{2}\|x_{k-1} - z\|^2 + \frac{\xi_k}{2}\|x_{k-1} - z\|^2\\ &\quad + s^{p-1}\frac{k^{2p}}{2}\|x_k - x_{k-1}\|^2\left[1 - \frac{2p}{k}\right] \leq C\end{aligned}$$

for some positive constant C. From the definition of $\lambda_k = \frac{2p}{s^{\frac{1-p}{2}}} k^{p-1}$ and $\xi_k \sim \frac{2p}{s^{1-p}}(\alpha - 4p + 1)k^{2p-2}$ (recall that $\alpha - 4p + 1 > 0$), we deduce that, for some $\epsilon > 0$, and k large enough

$$\xi_k \geq \epsilon \lambda_k^2.$$

On the other hand, for k large enough

$$s^{p-1}\frac{k^{2p}}{2}\|x_k - x_{k-1}\|^2 \left[1 - \frac{2p}{k}\right] \geq \frac{1}{2(1+\epsilon)} s^{p-1} k^{2p} \|x_k - x_{k-1}\|^2.$$

Combining the above results, and using

$$\|x_k - z\|^2 - \|x_{k-1} - z\|^2 \geq 2\langle x_{k-1} - z, x_k - x_{k-1}\rangle,$$

we obtain

$$\begin{aligned} C &\geq s^p k^{2p}(\Theta(x_k) - \min \Theta) + \frac{1}{2}\left[\frac{s^{p-1}k^{2p}}{1+\epsilon}\|x_k - x_{k-1}\|^2 + 2s^{\frac{p-1}{2}}\lambda_k k^p \langle x_k - x_{k-1}, x_{k-1} - z\rangle \right. \\ &\quad \left. +(1+\epsilon)\lambda_k^2 \|x_{k-1} - z\|^2\right] \\ &= s^p k^{2p}(\Theta(x_k) - \min \Theta) + \frac{1}{2}\left\|\frac{1}{\sqrt{1+\epsilon}} s^{\frac{p-1}{2}} k^p (x_k - x_{k-1}) + \sqrt{1+\epsilon}\lambda_k (x_{k-1} - z)\right\|^2. \end{aligned}$$

This implies

$$\Theta(x_k) - \min \Theta = \mathcal{O}\left(\frac{1}{k^{2p}}\right).$$

Hence, for $\alpha \leq 3$, for all $p < \frac{2\alpha}{3}$ we have obtained

$$\Theta(x_k) - \min \Theta = \mathcal{O}\left(\frac{1}{k^p}\right).$$

Since p has been taken greater or equal than $\frac{1}{2}$, this corresponds to assume $\alpha \geq \frac{3}{4}$. When $p \leq \frac{1}{2}$, on the basis of the inequality $(k+1)^{2p} - k^{2p} \leq 2pk^{2p-1}$, we are led to take

$$\lambda_k = \frac{2p}{s^{\frac{1-p}{2}}}\frac{(k-1)^{2p-1}}{k^p} \sim \frac{2p}{s^{\frac{1-p}{2}}} k^{p-1}, \tag{4.19}$$

which is an equivalent expression as in the case $p \leq \frac{1}{2}$. Since the proof is based on asymptotic equivalences, it works in a similar way, which completes the proof in the case $\alpha \leq \frac{3}{4}$. ☐

Remark 4.5. Our proof uses computation based on asymptotic equivalences. This makes the proof simpler, but on the other hand we lose some information with respect to the continuous case. We have obtained the order of convergence $\mathcal{O}\left(\frac{1}{k^p}\right)$ for all $p < \frac{2\alpha}{3}$ instead of $\frac{2\alpha}{3}$. It might be tempting to pass to the limit as $p \to \frac{2\alpha}{3}$ on the inequality

$$\tilde{\mathcal{E}}_{k,p} \leq \tilde{\mathcal{E}}_{k_0,p} \quad \text{for} \quad k \geq k_0$$

where

$$\tilde{\mathcal{E}}_{k,p} = s^p k^{2p} W_k + s^{\frac{p-1}{2}} \lambda_{k,p} k^p (h_k - h_{k-1}) + (\lambda_{k,p}^2 + \xi_{k,p}) h_{k-1} - \frac{p}{s^{1-p}} (k)^{2p-1} \|x_k - x_{k-1}\|^2$$

and $\lambda_{k,p} = \frac{2p}{s^{\frac{1-p}{2}}} k^{p-1}$, $\xi_{k+1,p} = -\lambda_{k+1,p}^2 - s^{\frac{p-1}{2}} (\alpha_k \lambda_{k+1,p} (k+1)^p - \lambda_{k,p} k^p)$. But this would not be a correct argument: while $\tilde{\mathcal{E}}_{k,p}$ is actually continuous with respect to p, it is possible that at the same time, in the above inequality, the index k_0 tends to $+\infty$! However, on the basis of the continuous results, it is probable that the same order of convergence as in the continuous case is valid, that is to say $\frac{2\alpha}{3}$. [1]

4.6. Convergence of iterates. The critical case $\alpha = 3$

In the case $\alpha > 3$, as a major result, the weak convergence of the sequences generated by $(\text{IFB})_\alpha$ was obtained by Chambolle and Dossal [16], and completed by Attouch-Chbani-Peypouquet-Redont [5]. The case $\alpha = 3$ corresponds to the accelerated method of Nesterov, and Beck-Teboulle FISTA method. In this case, the convergence of iterations is still an open problem. In line with the results obtained in the continuous case, we give partial answers to this question. As a main result, we prove the convergence in the one-dimensional case. We therefore consider $(\text{IFB})_\alpha$ with $\alpha = 3$

$$(\text{IFB})_3 \quad \begin{cases} y_k = x_k + (1 - \frac{3}{k})(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k)). \end{cases}$$

In the case $\alpha = 3$, the Lyapunov function takes the simpler form

$$\mathcal{E}(k) := s\,(k+1)^2\,(\Theta(x_k) - \min\Theta) + 2\|x_k - x^* + \frac{k-1}{2}(x_k - x_{k-1})\|^2, \tag{4.20}$$

where $x^* \in S$. In [5, 6, 16, 35] it is proved that the sequence $(\mathcal{E}(k))$ is nonincreasing, and $\Theta(x_k) - \min\Theta = \mathcal{O}\left(\frac{1}{k^2}\right)$. From this we can deduce the following estimates of the sequence (x_k), which are valid in a general Hilbert space.

Proposition 4.6. *Let $\mathcal{H}$ a Hilbert space. Let (x_k) be a sequence generated by algorithm $(\text{IFB})_\alpha$ with $\alpha = 3$. Then*

$$\sup_k \|x_k\| < +\infty \textit{ and } \sup_k k\|x_k - x_{k-1}\| < +\infty.$$

Proof. Since $(\mathcal{E}(k))$ is nonincreasing, it is bounded from above, which implies that, for some positive constant C

$$\|(x_k - x^*) + \frac{k-1}{2}(x_k - x_{k-1})\|^2 \leq C. \tag{4.21}$$

After developing the above quadratic expression , and neglecting the nonnegative term $\|\frac{k-1}{2}(x_k - x_{k-1})\|^2$, we obtain

$$\|x_k - x^*\|^2 + (k-1)\,\langle x_k - x_{k-1}, x_k - x^*\rangle \leq C.$$

[1] Indeed, since the paper has been submitted, we learned that, using a different proof, the decay rate $\frac{2\alpha}{3}$ has been obtained by Apidopoulos-Aujol-Dossal in ([8], Cor. 3.2).

Let us write this expression in a recursive form. Equivalently,

$$\|x_k - x^*\|^2 + (k-1)\left\langle (x_k - x^*) - (x_{k-1} - x^*), x_k - x^* \right\rangle \le C.$$

Using the elementary convex inequality

$$\|x_{k-1} - x^*\|^2 \ge \|x_k - x^*\|^2 + 2\left\langle (x_{k-1} - x^*) - (x_k - x^*), x_k - x^* \right\rangle$$

we infer

$$\|x_k - x^*\|^2 + \frac{k-1}{2}\|x_k - x^*\|^2 - \frac{k-1}{2}\|x_{k-1} - x^*\|^2 \le C.$$

Equivalently

$$\frac{k-1}{2}\|x_k - x^*\|^2 - \frac{k-2}{2}\|x_{k-1} - x^*\|^2 + \|x_k - x^*\|^2 - \frac{1}{2}\|x_{k-1} - x^*\|^2 \le C.$$

Hence

$$\left(\frac{k-1}{2}\|x_k - x^*\|^2 + \|x_k - x^*\|^2\right) - \left(\frac{k-2}{2}\|x_{k-1} - x^*\|^2 + \|x_{k-1} - x^*\|^2\right) \le C.$$

Summing these inequalities, we infer

$$\frac{k-1}{2}\|x_k - x^*\|^2 + \|x_k - x^*\|^2 \le C + Ck,$$

which gives the announced result. Returning to (4.21), by using the triangle inequality we immediately deduce that $\sup_k k\|x_k - x_{k-1}\| < +\infty$. ☐

Theorem 4.7. *Take $\mathcal{H} = \mathbb{R}$. Let us make assumptions (H). Let (x_k) be a sequence of iterates generated by algorithm* $(\text{IFB})_3$ *with $\alpha = 3$. Then (x_k) converges, as $k \to +\infty$, to a point in $S = \operatorname{argmin} \Theta$.*

Proof. By Proposition 4.6, for $\alpha = 3$, the sequence (x_k) is bounded, and minimizing (indeed, $\Theta(x_k) - \min \Theta = \mathcal{O}\left(\frac{1}{k^2}\right)$). As a consequence, when $\operatorname{argmin} \Theta$ is reduced to a singleton x^*, it is the unique cluster point of the the sequence (x_k), which implies the convergence of (x_k) to x^*. Thus, we just need to consider the case where $S = [a, b]$ is an interval with positive length. Recall that the algorithm can be written in an equivalent way as

$$(\text{IFB})_\alpha \quad \begin{cases} y_k = x_k + (1 - \frac{3}{k})(x_k - x_{k-1}) \\ x_{k+1} = y_k - sG_s(y_k), \end{cases}$$

where the following equivalence holds true

$$z \in \operatorname{argmin} \Theta \iff G_s(z) = 0.$$

Note that $y_k - x_k \to 0$, which implies that the sequences (x_k) and (y_k) have the same cluster points. As for the continuous dynamic we have to consider the following cases:

- There exists $K \in \mathbb{N}$ such that $y_k \in [a, b]$ for all $k \ge K$. Then $G_s(y_k) = 0$, which gives

$$x_{k+1} - x_k = (1 - \frac{3}{k})(x_k - x_{k-1}).$$

Solving this induction, we obtain $0 \leq x_{k+1} - x_k \leq \frac{C}{k^3}$, which implies that the sequence (x_k) converges.

• There exists $K \in \mathbb{N}$ such that $y_k \leq a$ for all $k \geq K$. As a consequence, a is the unique cluster point of the sequence (x_k) that converges to this point. Symetrically, if $y_k \geq b$ for all $k \geq K$, then b is the unique cluster point of the sequence (x_k) that converges to this point.

• It remains to study the case where a and b are cluster points of the sequence (y_k), and consequently of the sequence (x_k). Note that $x_k - x_{k-1} \to 0$. Therefore, in parallel to the continuous case, we can find four sequences $x_{k_{n_1}}$, $x_{k_{n_2}}$, $x_{k_{n_3}}$, $x_{k_{n_4}}$, with $k_{n_1} < k_{n_2} < k_{n_3} < k_{n_4}$ such that, when "going from a to b" we have

1. $x_{k_{n_1}} \to a$, $x_{k_{n_1}} > a$, $x_{k_{n_1}+1} - x_{k_{n_1}} > 0$;
2. $x_{k_{n_2}} \to b$, $x_{k_{n_2}} < b$, $x_{k_{n_2}+1} - x_{k_{n_2}} > 0$;
3. $x_k \in [a, b]$ for all indices k such that $k_{n_1} \leq k \leq k_{n_2}$.

Symetrically when "going from b to a"

1. $x_{k_{n_3}} \to b$, $x_{k_{n_3}} < b$, $x_{k_{n_3}+1} - x_{k_{n_3}} < 0$;
2. $x_{k_{n_4}} \to a$, $x_{k_{n_4}} > a$, $x_{k_{n_4}+1} - x_{k_{n_4}} < 0$;
3. $x_k \in [a, b]$ for all indices k such that $k_{n_3} \leq k \leq k_{n_4}$.

As in the continuous case, let us analyze the decay of the quantity $|k(x_{k+1} - x_k)|$ during a loop. When $k_{n_1} \leq k \leq k_{n_2}$, we have $x_{k+1} = x_k + (1 - \frac{3}{k})(x_k - x_{k-1})$. Equivalently

$$(k+1)(x_{k+1} - x_k) - k(x_k - x_{k-1}) + 3(x_k - x_{k-1}) - (x_{k+1} - x_k) = 0.$$

Let us sum these equalities with k varying from k_{n_1} to $k_{n_2} - 1$. We obtain

$$k_{n_2}(x_{k_{n_2}} - x_{k_{n_2}-1}) - k_{n_1}(x_{k_{n_1}} - x_{k_{n_1}-1}) + 3(x_{k_{n_2}-1} - x_{k_{n_1}-1}) - (x_{k_{n_2}} - x_{k_{n_1}}) = 0.$$

Taking account of $x_{k_{n_1}} \to a$, $x_{k_{n_2}} \to b$, $x_k - x_{k-1} \to 0$, and of the sign of the above quantities, we deduce that, when "going from a to b",

$$|k_{n_2}(x_{k_{n_2}} - x_{k_{n_2}-1})| - |k_{n_1}(x_{k_{n_1}} - x_{k_{n_1}-1})| \sim -2(b-a). \tag{4.22}$$

Symetrically when "going from b to a"

$$|k_{n_4}(x_{k_{n_4}} - x_{k_{n_4}-1})| - |k_{n_3}(x_{k_{n_3}} - x_{k_{n_3}-1})| \sim -2(b-a). \tag{4.23}$$

Let us now exploit the decreasing property of

$$\mathcal{E}(k) := s\,(k+1)^2\,(\Theta(x_k) - \min \Theta) + 2\|x_k - x^* + \frac{k-1}{2}(x_k - x_{k-1})\|^2.$$

Take $x^* = x_{k_{n_2}} \in S = [a, b]$. Since $k_{n_3} \geq k_{n_2}$, we have $\mathcal{E}(k_{n_3}) \leq \mathcal{E}(k_{n_2})$. Equivalently,

$$|x_{k_{n_3}} - x_{k_{n_2}} + \frac{k_{n_3}-1}{2}(x_{k_{n_3}} - x_{k_{n_3}-1})| \leq |\frac{k_{n_2}-1}{2}(x_{k_{n_2}} - x_{k_{n_2}-1})|.$$

Hence

$$|(k_{n_3}-1)(x_{k_{n_3}} - x_{k_{n_3}-1})| \leq |(k_{n_2}-1)(x_{k_{n_2}} - x_{k_{n_2}-1})| + 2|x_{k_{n_3}} - x_{k_{n_2}}|.$$

As a consequence

$$|k_{n_3}(x_{k_{n_3}} - x_{k_{n_3}-1})| \leq |k_{n_2}(x_{k_{n_2}} - x_{k_{n_2}-1})| + |x_{k_{n_3}} - x_{k_{n_3}-1}| + 2|x_{k_{n_3}} - x_{k_{n_2}}|. \tag{4.24}$$

Let us now combine (4.22), (4.23), (4.24). Since $x_{k_{n_2}} \to b$, $x_{k_{n_3}} \to b$, and $x_k - x_{k-1} \to 0$, we obtain that during a loop, when passing from $x_{k_{n_1}} \sim a$ to $x_{k_{n_4}} \sim a$, the expression $|k(x_k - x_{k-1})|$ decreases by a fixed positive quantity. Hence, after a finite number of steps this expression would become negative, a clear contradiction. ☐

5. The perturbed case

5.1. Continuous dynamics

We consider the equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = g(t) \tag{5.1}$$

where, depending on the situation, the second member $g : [t_0, +\infty[\to \mathcal{H}$ is a forcing term, or comes from the approximation or computational errors in (1.1). We suppose that g is locally integrable to ensure existence and uniqueness for the corresponding Cauchy problem. We will show that the results of the previous sections are robust, *i.e.*, there remain satisfied if the perturbation g is not too large asymptotically.

Theorem 5.1. *Let* $\operatorname{argmin}\Phi \neq \emptyset$, *and let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution of (5.1) with* $\alpha > 0$, *and* $t_0 > 0$. *Let* $p = \min(1, \frac{\alpha}{3})$ *and suppose that* $\int_{t_0}^{+\infty} t^p g(t)\mathrm{d}t < +\infty$. *Then*

$$\Phi(x(t)) - \min\Phi = \mathcal{O}\left(\frac{1}{t^{2p}}\right).$$

Proof. Take $z \in \operatorname{argmin}\Phi$. Let us modify the energy function considering in (2.1) by introducing an additional term taking account of the perturbation g. Let us first fix some $T > t_0$, and define for $t \in [t_0, T]$ (without ambiguity we keep the same notations as in the previous sections)

$$\begin{aligned}\mathcal{E}^p_{\lambda,\xi}(t) = {} & t^{2p}\left[\Phi(x(t)) - \min\Phi\right] + \|\lambda(t)(x(t)-z) + t^p\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t)-z\|^2 \\ & + \int_t^T \langle\lambda(s)(x(s)-z) + s^p\dot{x}(s), s^p g(s)\rangle \mathrm{d}s.\end{aligned} \tag{5.2}$$

By derivation of $\mathcal{E}^p_{\lambda,\xi}(\cdot)$, one obtains

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \quad = \quad & 2pt^{2p-1}\left[\Phi(x(t)) - \min\Phi\right] + t^{2p}\langle\nabla\Phi(x(t), \dot{x}(t)\rangle + \langle\lambda(t)(x(t)-z) + t^p\dot{x}(t), t^p(\ddot{x}(t) - g(t))\rangle \\ & + \left[\lambda(t)^2 + pt^{p-1}\lambda(t) + \xi(t) + t^p\dot{\lambda}(t)\right]\langle x(t)-z, \dot{x}(t)\rangle + \left(\lambda(t)\dot{\lambda}(t) + \tfrac{\dot{\xi}(t)}{2}\right)\|x(t)-z\|^2 \\ & + \left(t^p\lambda(t) + pt^{2p-1}\right)\|\dot{x}(t)\|^2.\end{aligned}$$

Then note that the term $\ddot{x}(t) - g(t)$ is exactly the same as in the unperturbed case. By a similar argument, by taking $p = \min(1, \frac{\alpha}{3})$, we obtain $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}^p_{\lambda,\xi}(t) \leq 0$.

As a consequence, the energy function $\mathcal{E}^p_{\lambda,\xi}(\cdot)$ is nonincreasing. In particular, for $t \geq t_0$ we have $\mathcal{E}^p_{\lambda}(t) \leq \mathcal{E}^p_{\lambda}(t_0)$, which implies, by definition of $\mathcal{E}^p_{\lambda}(\cdot)$

$$t^{2p}\left[\Phi(x(t)) - \min\Phi\right] + \|\lambda(t)(x(t)-z) + t^p\dot{x}(t)\|^2 \leq C + \int_{t_0}^t \langle\lambda(s)(x(s)-z) + s^p\dot{x}(s), s^p g(s)\rangle \mathrm{d}s. \tag{5.3}$$

Applying Cauchy-Schwarz inequality we deduce that

$$\|\lambda(t)(x(t)-z)+t^p\dot{x}(t)\|^2 \leq C+\int_{t_0}^{t}\|\lambda(s)(x(s)-z)+s^p\dot{x}(s)\|\|s^pg(s)\|\mathrm{d}s.$$

Applying Lemma 6.2, we obtain

$$\|\lambda(t)(x(t)-z)+t^p\dot{x}(t)\| \leq \sqrt{C}+2\int_{t_0}^{t}\|s^pg(s)\|\mathrm{d}s.$$

By assumption $\int_{t_0}^{+\infty} t^p g(t)\mathrm{d}t < +\infty$. Hence

$$\sup_{t\geq t_0}\|\lambda(t)(x(t)-z)+t^p\dot{x}(t)\| \leq +\infty.$$

Injecting this estimate into the equation (5.3) leads to

$$t^{2p}\left[\Phi(x(t))-\min\Phi\right] \leq C+\left[\sqrt{C}+2\int_{t_0}^{\infty}\|s^pg(s)\|\mathrm{d}s\right]\int_{t_0}^{\infty}\|s^pg(s)\|\mathrm{d}s$$

which immediately gives the claim. □

Remark 5.2. Because of the numerical importance of the subject, several recent articles have been devoted the study of perturbed versions of the FISTA algorithm, with various formulation for the perturbations or errors: to cite a few of them Aujol-Dossal [9], Schmidt-Le Roux-Bach [33], Solodov-Svaiter [34], Villa-Salzo-Baldassarres-Verri [36]. Our results are among the first to address this question in the sub-critical cas $\alpha \leq 3$. A more general version of Theorem 5.1 is given in ([10], Thm. 2, page 15).

5.2. Algorithms

Let us consider the perturbed algorithm

$$(\mathrm{IFB})_{\alpha-pert}\quad\begin{cases} y_k = x_k + (1-\frac{\alpha}{k})(x_k - x_{k-1}) \\ x_{k+1} = \mathrm{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k) - sg_k). \end{cases}$$

In view of the above results, and by adapting the proof of the unperturbed case, one can prove the following result, whose detailed demonstration is left to the reader.

Theorem 5.3. *Let us make assumptions (H). Let (x_k) be a sequence of iterates generated by perturbed algorithm $(\mathrm{IFB})_{\alpha,pert}$. Suppose that $0<\alpha<3$. Then, for any $p<\frac{2\alpha}{3}$, and (g_k) such that $\sum k^p\|g_k\| < +\infty$, we have*

$$(\Phi+\Psi)(x_k) - \min_{\mathcal{H}}(\Phi+\Psi) = \mathcal{O}\left(\frac{1}{k^p}\right).$$

6. Appendix

Lemma 6.1. *([27]) Let S be a nonempty subset of $\mathcal{H}$ and let $x:[t_0,+\infty[\to\mathcal{H}$. Assume that*

(i) for every $z\in S$, $\lim_{t\to\infty}\|x(t)-z\|$ exists;
(ii) every weak sequential cluster point of $x(t)$, as $t\to\infty$, belongs to S.

Then $x(t)$ converges weakly as $t\to\infty$ to a point in S.

Lemma 6.2. *Let $m : [t_0; T] \to [0; +\infty[$ be integrable, and let $c > 0$. Suppose $w : [t_0; T] \to \mathbb{R}$ is continuous and*

$$\frac{1}{2}w(t)^2 \leq \frac{1}{2}c^2 + \int_{t_0}^{t} m(s)w(s)\mathrm{d}s$$

for all $t \in [t_0; T]$. Then, $|w(t)| \leq c + \int_{t_0}^{t} m(s)\mathrm{d}s$ for all $t \in [t_0; T]$.

References

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

[2] H. Attouch and A. Cabot, Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. *J. Differ. Equ.* **263** (2017) 5412–5458.

[3] H. Attouch and A. Cabot, Convergence rates of inertial forward-backward algorithms, HAL Preprint 01453170 (2017).

[4] H. Attouch, A. Cabot and P. Redont, The dynamics of elastic shocks via epigraphical regularization of a differential inclusion. *Adv. Math. Sci. Appl.* **12** (2002) 273–306.

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

[6] 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 J. Optim.* **26** (2016) 1824–1834.

[7] V. Apidopoulos, J.-F. Aujol and Ch. Dossal, On a second order differential inclusion modeling the FISTA algorithm. Hal:01517708 (2017).

[8] V. Apidopoulos, J.-F. Aujol and Ch. Dossal, Convergence rate of inertial forward-backward algorithm beyond Nesterov's rule. HAL Preprint 01551873 (2017).

[9] J.-F. Aujol and C. Dossal, Stability of over-relaxations for the forward-backward algorithm, application to FISTA. *SIAM J. Optim.* **25** (2015) 2408–2433.

[10] J.-F. Aujol and C. Dossal, Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for $b > 0$. hal-01547251 (2017).

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

[12] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** (2009) 183–202.

[13] P. Bégout, J. Bolte and M.A. Jendoubi, On damped second-order gradient systems. *J. Differ. Equ.* **259** (2015) 3115–3143.

[14] H. Brézis, Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution. North Holland (1972).

[15] A. Cabot, H. Engler and S. Gadat, Second-order differential equations with asymptotically small dissipation and piecewise flat potentials. *Electron. J. Differ. Eq.* **17** (2009) 33–38.

[16] A. Chambolle and Ch. Dossal, On the convergence of the iterates of the fast iterative shrinkage thresholding algorithm. *J. Optim. Theory Appl.* **166** (2015) 968–982.

[17] C.Y. Drori and M. Teboulle, Performance of first-order methods for smooth convex minimization: a novel approach. *Math. Progr. Ser. A.* **145** (2014) 451–482.

[18] A. Haraux, Systèmes dynamiques dissipatifs et applications. RMA 17, Masson (1991).

[19] O. Güler, New proximal point algorithms for convex minimization. *SIAM J. Optim.* **2** (1992) 649–664.

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

[21] D. Kim and J.A. Fessler, Optimized first-order methods for smooth convex minimization. *Math. Progr. Ser. A* **159** (2016) 81–107.

[22] J. Liang, J. Fadili and G. Peyré, Local linear convergence of forward-backward under partial smoothness. *Adv. Neural Inf. Process. Syst.* **2014** 1970–1978.

[23] R. May, Asymptotic for a second order evolution equation with convex potential and vanishing damping term. Preprint arXiv:1509.05598v2 (2015).

[24] Y. Nesterov, A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Sov. Math. Dokl.* **27** (1983) 372–376.

[25] Y. Nesterov, Introductory Lectures On Convex Optimization: A basic Course. Vol. 87 of *Applied Optimization.* Kluwer Academic Publishers, Boston, MA (2004).

[26] Y. Nesterov, Gradient methods for minimizing composite objective function. CORE Discussion Papers (2007).

[27] Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.* **73** (1967) 591–597.

[28] N. Parikh and S. Boyd, Proximal algorithms. *Found. Trends Optimiz.* **1** (2014) 123–231.

[29] J. Peypouquet, Convex optimization in normed spaces: theory, methods and examples. Springer (2015).

[30] J. Peypouquet and S. Sorin, Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Anal.* **17** (2010) 1113–1163.

[31] B.T. Polyak, Introduction to optimization. Optimization software. New York (1987).

[32] B.T. Polyak and P.S. Shcherbakova, Optimisation and asymptotic stability. *Int. J. Control* **91** (2018) 2404–2410..

[33] M. Schmidt, N. Le Roux and F. Bach, Convergence rates of inexact proximal-gradient methods for convex optimization. NIPS'11 – 25th Annual Conference on Neural Information Processing Systems, Grenada, Spain. Preprint: Hal: inria-00618152 (2011).

[34] M.V. Solodov and B.F. Svaiter, A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.* **22** (2001) 1013-35.

[35] W. Su, S. Boyd and E.J. Candès, A differential equation for modeling Nesterov's accelerated gradient method: theory and insights. *Neural Inf. Process. Syst.* **27** (2014) 2510–2518.

[36] S. Villa, S. Salzo, L. Baldassarres and A. Verri, Accelerated and inexact forward-backward. *SIAM J. Optim.* **23** (2013) 1607–1633.