



Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization

Hedy Attouch, Zaki Chbani, Hassan Riahi

► To cite this version:

Hedy Attouch, Zaki Chbani, Hassan Riahi. Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization. 2022. hal-03859665

HAL Id: hal-03859665

<https://hal.science/hal-03859665>

Preprint submitted on 18 Nov 2022

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

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

Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization

Hedy ATTOUCH · Zaki CHBANI · Hassan RIAHI

the date of receipt and acceptance should be inserted later

Abstract In a Hilbert framework, for convex differentiable optimization, we consider accelerated gradient methods obtained by combining temporal scaling and averaging techniques with Tikhonov regularization. We start from the continuous steepest descent dynamic with an additional Tikhonov regularization term whose coefficient vanishes asymptotically. We provide an extensive Lyapunov analysis of this first-order evolution equation. Then we apply to this dynamic the method of time scaling and averaging recently introduced by Attouch, Bot and Nguyen. We thus obtain an inertial dynamic which involves viscous damping associated with Nesterov's method, implicit Hessian damping and Tikhonov regularization. Under an appropriate setting of the parameters, just using Jensen's inequality, without the need for another Lyapunov analysis, we show that the trajectories have at the same time several remarkable properties: they provide a rapid convergence of values, fast convergence of the gradients to zero, and strong convergence to the minimum norm minimizer. These results complete and improve the previous results obtained by the authors.

Keywords Accelerated gradient methods; convex optimization; damped inertial dynamics; Hessian driven damping; Tikhonov approximation; time scaling and averaging.

Mathematics Subject Classification (2010) 37N40, 46N10, 49M30, 65K05, 65K10, 65K15, 65L08, 65L09, 90B50, 90C25.

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 $x \in \mathcal{H}$. Given $f : \mathcal{H} \rightarrow \mathbb{R}$ a general convex function, continuously differentiable, we will develop fast gradient methods with new favorable properties to solve the minimization problem

$$\min \{f(x) : x \in \mathcal{H}\}. \quad (1)$$

Hedy ATTOUCH
IMAG, Univ. Montpellier, CNRS, Montpellier, France
hedy.attouch@umontpellier.fr,

Zaki CHBANI · Hassan RIAHI
Cadi Ayyad University
Sémlalia Faculty of Sciences 40000 Marrakech, Morocco
chbaniz@uca.ac.ma · h-riahi@uca.ac.ma

Our approach is based on the convergence properties as $t \rightarrow +\infty$ of the trajectories $t \mapsto x(t)$ generated by the first-order evolution equation

$$\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0, \quad (2)$$

which can be interpreted as a Tikhonov regularization of the continuous steepest descent dynamics. As a basic ingredient, (2) involves a Tikhonov regularization term with positive coefficient $\varepsilon(\cdot)$ that satisfies $\lim_{t \rightarrow +\infty} \varepsilon(t) = 0$, which preserves the equilibria.

Throughout the paper, we assume that the objective function f and the Tikhonov regularization parameter $\varepsilon(\cdot)$ satisfy the following hypothesis:

$$(A) \quad \begin{cases} (1) & f : \mathcal{H} \rightarrow \mathbb{R} \text{ is convex, of class } \mathcal{C}^1, \nabla f \text{ is Lipschitz continuous on bounded sets;} \\ (2) & S := \operatorname{argmin}_{\mathcal{H}} f \neq \emptyset. \text{ We denote by } x^* \text{ the element of minimum norm of } S; \\ (3) & \varepsilon : [t_0, +\infty[\rightarrow \mathbb{R}^+ \text{ is nonincreasing, of class } \mathcal{C}^1, \text{ such that } \lim_{t \rightarrow \infty} \varepsilon(t) = 0. \end{cases}$$

We first provide an extensive Lyapunov analysis of the first-order evolution equation (2). Then we apply to (2) the recent method of time scaling and averaging introduced by Attouch, Bot and Nguyen [7]. Denoting by $t \mapsto \tau(t)$ the time scaling, we thus obtain the system

$$\ddot{x}(t) + \frac{1 + \ddot{\tau}(t)}{\dot{\tau}(t)} \dot{x}(t) + \nabla f(x(t) + \dot{\tau}(t)\dot{x}(t)) + \varepsilon(\tau(t))(x(t) + \dot{\tau}(t)\dot{x}(t)) = 0, \quad (3)$$

which involves viscous damping, implicit Hessian damping, and Tikhonov regularization.

Of particular interest is the case $\tau(s) = \frac{s^2}{2(\alpha-1)}$ for some $\alpha > 1$. Then $\frac{1+\ddot{\tau}(s)}{\dot{\tau}(s)} = \frac{\alpha}{s}$, and (3) becomes

$$\ddot{x}(s) + \frac{\alpha}{s} \dot{x}(s) + \nabla f\left[x(s) + \frac{s}{\alpha-1} \dot{x}(s)\right] + \varepsilon\left(\frac{s^2}{2(\alpha-1)}\right) \left[x(s) + \frac{s}{\alpha-1} \dot{x}(s)\right] = 0. \quad (4)$$

As a remarkable property, a judicious choice of $\varepsilon(\cdot)$ gives solution trajectories of (4) that satisfy the following three favorable properties:

- Fast convergence of the values (with rate $\frac{1}{t^2}$), therefore comparable to the dynamic model introduced by Su, Boyd, and Candès ([38]) of the Nesterov accelerated gradient method.
- Fast convergence of the gradients towards zero.
- Strong convergence of the trajectories towards the minimum norm solution.

To our knowledge, this is the first time that these three properties are verified simultaneously by the trajectories of the same dynamic.

For all the systems considered in the article, we take for granted the existence and uniqueness of the solution of the Cauchy problem. For second-order evolution equations, this is obtained by applying the classical Cauchy-Lipschitz theorem to the Hamiltonian formulation (recall that ∇f is assumed locally Lipschitz continuous). The passage from the local solution to the global solution results from the energy estimations which are established.

At the end of the paper, we will show that many results have a natural extension to the case of a convex lower semicontinuous proper function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$.

1.1 Presentation of the results

We present two model situations corresponding to cases of $\varepsilon(\cdot)$ of particular interest.

1.1.1 Case $\varepsilon(t) = \frac{\delta}{t}$

The second-order evolution equation

$$\ddot{x}(t) + \frac{\alpha + 2\delta}{t} \dot{x}(t) + \nabla f \left[x(t) + \frac{t}{\alpha - 1} \dot{x}(t) \right] + \frac{2\delta(\alpha - 1)}{t^2} x(t) = 0 \quad (5)$$

is obtained by taking in (4) $\varepsilon(t) = \frac{\delta}{t}$, with $\delta > 1$. In Theorem 6, we show that for $\alpha > 3$ and $\delta > 1$, the trajectories generated by (5) have at the same time the following properties:

- Rapid convergence of values:

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^2} \right) \text{ as } t \rightarrow +\infty.$$

According to $\alpha > 3$ and $\delta > 1$, we have $\alpha + 2\delta > 5$. This is in agreement with the fact that taking the damping coefficient γ/t with γ sufficiently large in the dynamic model of Su, Boyd and Candès ([38]) of the accelerated gradient of Nesterov method is beneficial (especially in the presence of strong convexity), see [8].

- Rapid convergence of the gradients towards zero:

$$\|\nabla f(x(t))\| = \mathcal{O} \left(\frac{1}{t} \right) \text{ as } t \rightarrow +\infty.$$

- Strong convergence of $x(t)$ to the minimum norm solution x^* , as $t \rightarrow +\infty$.

1.1.2 Case $\varepsilon(t) = \frac{1}{t^r}$

The second-order evolution equation

$$\ddot{x}(t) + \left(\frac{\alpha}{t} + \frac{2^r(\alpha - 1)^{r-1}}{t^{2r-1}} \right) \dot{x}(t) + \nabla f \left(x(t) + \frac{t}{\alpha - 1} \dot{x}(t) \right) + \frac{2^r(\alpha - 1)^r}{t^{2r}} x(t) = 0. \quad (6)$$

is obtained by taking in (4) $\varepsilon(t) = \frac{1}{t^r}$, with $0 < r < 1$. Under an appropriate setting of the parameters, just using Jensen's inequality, without the need for another Lyapunov analysis, we show in Theorem 7 that for arbitrary $0 < r < 1$, the trajectories generated by (6) have at the same time the following properties:

- Rapid convergence of values: for $\alpha > 1$, and $0 < r < 1$

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}} \right) \text{ as } t \rightarrow +\infty.$$

By taking $\alpha \geq 3$, and $r \simeq 1$, we thus have $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^{2r}} \right)$. So we can approach arbitrarily the optimal convergence rate $1/t^2$.

- Rapid convergence of the gradients towards zero:

$$\|\nabla f(x(t))\|^2 = \mathcal{O} \left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}} \right) \text{ as } t \rightarrow +\infty.$$

- Strong convergence of $x(t)$ to the minimum norm solution x^* , as $t \rightarrow +\infty$.

1.1.3 Comments

In the above results, the terminology “implicit Hessian damping” comes from the fact that by a Taylor expansion (as $t \rightarrow +\infty$ we have $\dot{x}(t) \rightarrow 0$ which justifies using Taylor expansion), we have

$$\nabla f(x(t) + \beta(t)\dot{x}(t)) \approx \nabla f(x(t)) + \beta(t)\nabla^2 f(x(t))\dot{x}(t),$$

hence making the Hessian damping appear indirectly in (5) and (6). These results improve the previous results obtained by the authors [12], [4], [5], and which were based on Lyapunov’s direct analysis of second-order damped inertial dynamics with Tikhonov regularization, see also [22]. In contrast, the Lyapunov analysis is now performed on the initial first-order evolution equation (2) which results in a significantly simplified mathematical analysis. It is also important to note that the scaling and time averaging method allows these questions to be addressed in a unified way. The numerical importance of the Hessian driven damping comes from the fact that it allows to significantly attenuates the oscillations which come naturally with inertial systems, see [9, 10], [37].

1.2 Historical perspective

Initially designed for the regularization of ill-posed inverse problems [39, 40], the field of application of the Tikhonov regularization was then considerably widened. The coupling of first-order in time gradient systems with a Tikhonov approximation whose coefficient tends asymptotically towards zero has been highlighted in a series of papers [1], [3], [13], [15], [20], [26], [29], [30]. Our approach builds on several previous works that have paved the way concerning the coupling of damped second-order in time gradient systems with Tikhonov approximation. First studies concerned the heavy ball with friction system of Polyak [35], where the damping coefficient $\gamma > 0$ is fixed. In [14] Attouch and Czarnecki considered the system

$$\ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0. \quad (7)$$

In the slow parametrization case $\int_0^{+\infty} \varepsilon(t)dt = +\infty$, they proved that any solution $x(\cdot)$ of (7) converges strongly to the minimum norm element of $\operatorname{argmin} f$, see also [16], [25], [27], [31]. This hierarchical minimization result contrasts with the case without the Tikhonov regularization term, where the convergence holds only for weak convergence, and the limit depends on the initial data. In the quest for a faster convergence, the following system with asymptotically vanishing damping

$$(\text{AVD})_{\alpha, \varepsilon} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0, \quad (8)$$

was studied by Attouch, Chbani, and Riahi in [12]. It is a Tikhonov regularization of the dynamic

$$(\text{AVD})_{\alpha} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \quad (9)$$

which was introduced by Su, Boyd and Candès in [38]. $(\text{AVD})_{\alpha}$ is a low resolution ODE of the accelerated gradient method of Nesterov [33, 34] and of the Ravine method [17], [38]. $(\text{AVD})_{\alpha}$ has been the subject of many recent studies which have given an in-depth understanding of the Nesterov accelerated gradient method, see [2], [4], [8], [11], [19], [28], [32], [38], [36], [42].

Recently, an original approach has been developed by the authors, which is based on the heavy ball with friction method of Polyak in the *strongly convex case*. In this case, this autonomous dynamic is known to provide exponential convergence rates. To take advantage of this very good convergence property, they considered the nonautonomous dynamic version of the heavy ball method which at time t is governed by the gradient of the regularized function $x \mapsto f(x) + \frac{\varepsilon(t)}{2}\|x\|^2$, where the Tikhonov regularization parameter satisfies $\varepsilon(t) \rightarrow 0$ as $t \rightarrow +\infty$. This idea which was developed in [4], [5], [18] is detailed below, and will serve us as a comparison to our results. Note that in the following statement the Hessian driven damping is now taken in the explicit form.

Theorem 1 Take $0 < r < 2$, $\delta > 2$, $\beta > 0$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\ddot{x}(t) + \frac{\delta}{t^{\frac{r}{2}}} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r} x(t) = 0. \quad (10)$$

Then, we have fast convergence of the values, fast convergence of the gradients towards zero, and strong convergence of the trajectory to the minimum norm solution, with the following rates:

- $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^r}\right)$ as $t \rightarrow +\infty$;
- $\int_{t_0}^{+\infty} t^{\frac{3r-2}{2}} \|\nabla f(x(t))\|^2 dt < +\infty$;
- $\|x(t) - x_{\varepsilon(t)}\|^2 = \mathcal{O}\left(\frac{1}{t^{\frac{2-r}{2}}}\right)$ as $t \rightarrow +\infty$,

where $\varepsilon \mapsto x_{\varepsilon} = \operatorname{argmin}_{\xi \in \mathcal{H}} \{f(\xi) + \frac{\varepsilon}{2} \|\xi\|^2\}$ is the viscosity curve, which is known to converge to the minimum norm solution.

As a major result our approach allows to improve this result by treating the limiting case $r = 2$, which corresponds to the viscous damping coefficient of the Su, Boyd and Candès dynamic version of the Nesterov accelerated gradient method.

2 Steepest descent method with vanishing Tikhonov regularization.

This section is devoted to the asymptotic analysis as $t \rightarrow +\infty$ of the continuous steepest descent with vanishing Tikhonov regularization

$$\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0. \quad (11)$$

As explained above, our results will be direct consequences of the convergence rates established for this system. We will just apply the time scaling and averaging method to obtain an inertial system with fast convergence properties.

2.1 Classical results for the continuous steepest descent

Let us recall some classic facts concerning the asymptotic behavior of the continuous steepest descent, *i.e.* with $\varepsilon(t) = 0$, see [7] for a proof of the following theorem.

Theorem 2 Suppose that $f : \mathcal{H} \rightarrow \mathbb{R}$ satisfies (\mathcal{A}) . Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\dot{x}(t) + \nabla f(x(t)) = 0. \quad (\text{SD})$$

Then the following properties are satisfied:

- i) (convergence of gradients and velocities) $\|\nabla f(x(t))\| = o\left(\frac{1}{t}\right)$ and $\|\dot{x}(t)\| = o\left(\frac{1}{t}\right)$ as $t \rightarrow +\infty$.
- ii) (integral estimate of gradients, velocities) $\int_{t_0}^{+\infty} t \|\nabla f(x(t))\|^2 dt < +\infty$ and $\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < \infty$.
- iii) (convergence of values) $f(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t}\right)$ as $t \rightarrow +\infty$.
- iv) The solution trajectory $x(t)$ converges weakly as $t \rightarrow +\infty$, and its limit belongs to $S = \operatorname{argmin} f$.

2.2 Classical results concerning the viscosity curve

For any $\varepsilon > 0$ let us define

$$x_\varepsilon = \operatorname{argmin}_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{\varepsilon}{2} \|\xi\|^2 \right\}. \quad (12)$$

Equivalently

$$\nabla f(x_\varepsilon) + \varepsilon x_\varepsilon = 0. \quad (13)$$

The mapping $\varepsilon \mapsto x_\varepsilon$ is called the viscosity curve. Its geometric properties will play a key role in our analysis. Let us recall some of its classical properties:

Lemma 1 ([3]) *Let $x^* = \operatorname{proj}_{\operatorname{argmin} f} 0$. We have*

$$(i) \quad \forall \varepsilon > 0 \quad \|x_\varepsilon\| \leq \|x^*\| ;$$

$$(ii) \quad \lim_{\varepsilon \rightarrow 0} \|x_\varepsilon - x^*\| = 0.$$

Let us now recall the differentiability properties of the viscosity curve.

Lemma 2 ([3], [13]) *The function $\varepsilon \mapsto x_\varepsilon$ is Lipschitz continuous on the compact intervals of $]0, +\infty[$, hence almost everywhere differentiable, and the following inequality holds:*

$$\left\| \frac{d}{d\varepsilon} (x_\varepsilon) \right\| \leq \frac{\|x^*\|}{\varepsilon}. \quad (14)$$

This is a sharp estimation, since, as shown by Torralba in [41], there is a convex smooth function f such that the viscosity curve has infinite length.

2.3 The steepest descent with vanishing Tikhonov regularization. Preparatory results

Let us now introduce a vanishing Tikhonov regularization in the continuous steepest descent

$$\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0. \quad (15)$$

We recall that $\varepsilon : [t_0, +\infty[\rightarrow \mathbb{R}^+$ is a nonincreasing function of class \mathcal{C}^1 , such that $\lim_{t \rightarrow +\infty} \varepsilon(t) = 0$. The function

$$\varphi_t : \mathcal{H} \rightarrow \mathbb{R}, \quad \varphi_t(x) := f(x) + \frac{\varepsilon(t)}{2} \|x\|^2 \quad (16)$$

will play a central role in the Lyapunov analysis of (15) via its strong convexity property. Thus, it is convenient to reformulate (15) with the help of the function φ_t , which gives, equivalently

$$\dot{x}(t) + \nabla \varphi_t(x(t)) = 0. \quad (17)$$

Let us introduce the real-valued function $t \in [t_0, +\infty[\mapsto E(t) \in \mathbb{R}^+$, a key ingredient of our Lyapunov analysis. It is defined by

$$E(t) := (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \frac{\varepsilon(t)}{2} \|x(t) - x_{\varepsilon(t)}\|^2 \quad (18)$$

where φ_t has been defined in (16), Hence $x_{\varepsilon(t)} = \operatorname{argmin}_{\mathcal{H}} \varphi_t$. We need the following lemmas for the proof of the main theorem of this paper.

Lemma 3 *Let $x(\cdot) : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of the continuous steepest descent with vanishing Tikhonov regularization (15). Let $t \in [t_0, +\infty[\rightarrow E(t) \in \mathbb{R}^+$ be the energy function defined in (18). Then, the following estimates are satisfied: for any $t \geq t_0$,*

$$f(x(t)) - \min_{\mathcal{H}} f \leq E(t) + \frac{\varepsilon(t)}{2} \|x^*\|^2; \quad (19)$$

$$\|x(t) - x_{\varepsilon(t)}\|^2 \leq \frac{E(t)}{\varepsilon(t)}. \quad (20)$$

Therefore, $x(t)$ converges strongly to x^* as soon as $\lim_{t \rightarrow +\infty} \frac{E(t)}{\varepsilon(t)} = 0$.

Proof (i) According to the definition of φ_t , we have

$$\begin{aligned} f(x(t)) - \min_{\mathcal{H}} f &= \varphi_t(x(t)) - \varphi_t(x^*) + \frac{\varepsilon(t)}{2} (\|x^*\|^2 - \|x(t)\|^2) \\ &= [\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})] + \underbrace{\left[\varphi_t(x_{\varepsilon(t)}) - \varphi_t(x^*) \right]}_{\leq 0} + \frac{\varepsilon(t)}{2} (\|x^*\|^2 - \|x(t)\|^2) \\ &\leq \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) + \frac{\varepsilon(t)}{2} \|x^*\|^2. \end{aligned}$$

By definition of $E(t)$ we have

$$\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \leq E(t) \quad (21)$$

which, combined with the above inequality, gives (19).

(ii) By the strong convexity of φ_t , and $x_{\varepsilon(t)} := \operatorname{argmin}_{\mathcal{H}} \varphi_t$, we have

$$\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \geq \frac{\varepsilon(t)}{2} \|x(t) - x_{\varepsilon(t)}\|^2.$$

Returning to the definition of $E(t)$, we get

$$E(t) - \frac{\varepsilon(t)}{2} \|x(t) - x_{\varepsilon(t)}\|^2 \geq \frac{\varepsilon(t)}{2} \|x(t) - x_{\varepsilon(t)}\|^2,$$

which gives (20). □

Lemma 4 *The following properties are satisfied:*

(i) For each $t \geq t_0$, $\frac{d}{dt} (\varphi_t(x_{\varepsilon(t)})) = \frac{1}{2} \dot{\varepsilon}(t) \|x_{\varepsilon(t)}\|^2$.

(ii) The function $t \mapsto x_{\varepsilon(t)}$ is Lipschitz continuous on the compact intervals of $]t_0, +\infty[$, hence almost everywhere differentiable, and the following inequality holds: for almost every $t \geq t_0$

$$\left\| \frac{d}{dt} (x_{\varepsilon(t)}) \right\|^2 \leq -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)} \left\langle \frac{d}{dt} (x_{\varepsilon(t)}), x_{\varepsilon(t)} \right\rangle.$$

Therefore, for almost every $t \geq t_0$

$$\left\| \frac{d}{dt} (x_{\varepsilon(t)}) \right\| \leq -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)} \|x_{\varepsilon(t)}\|.$$

Proof (i) We use the differentiability properties of the Moreau envelope. Recall that, given $\theta > 0$

$$f_\theta(x) = \inf_{\xi \in H} \left\{ f(\xi) + \frac{1}{2\theta} \|x - \xi\|^2 \right\} \quad (22)$$

$$= f(\text{prox}_{\theta f}(x)) + \frac{1}{2\theta} \|x - \text{prox}_{\theta f}(x)\|^2, \quad (23)$$

i.e. $\text{prox}_{\theta f}(x)$ is the unique point where the infimum in (22) is achieved. One can consult [21, section 12.4] for more details on the Moreau envelope. We thus have

$$\varphi_t(x_{\varepsilon(t)}) = \inf_{\xi \in H} \left\{ f(\xi) + \frac{\varepsilon(t)}{2} \|\xi - 0\|^2 \right\} = f_{\frac{1}{\varepsilon(t)}}(0).$$

Since $\frac{d}{d\theta} f_\theta(x) = -\frac{1}{2} \|\nabla f_\theta(x)\|^2$, (see [4, Appendix, Lemma 3]), we have:

$$\frac{d}{dt} f_{\theta(t)}(x) = -\frac{\dot{\theta}(t)}{2} \|\nabla f_{\theta(t)}(x)\|^2.$$

Therefore,

$$\frac{d}{dt} \varphi_t(x_{\varepsilon(t)}) = \frac{d}{dt} \left(f_{\frac{1}{\varepsilon(t)}}(0) \right) = \frac{1}{2} \frac{\dot{\varepsilon}(t)}{\varepsilon^2(t)} \|\nabla f_{\frac{1}{\varepsilon(t)}}(0)\|^2. \quad (24)$$

On the other hand, we have

$$\nabla \varphi_t(x_{\varepsilon(t)}) = 0 \iff \nabla f(x_{\varepsilon(t)}) + \varepsilon(t)x_{\varepsilon(t)} = 0 \iff x_{\varepsilon(t)} = \text{prox}_{\frac{1}{\varepsilon(t)}f}(0).$$

Since $\nabla f_{\frac{1}{\varepsilon(t)}}(0) = \varepsilon(t) \left(0 - \text{prox}_{\frac{1}{\varepsilon(t)}f}(0) \right)$, we get $\nabla f_{\frac{1}{\varepsilon(t)}}(0) = -\varepsilon(t)x_{\varepsilon(t)}$. This combined with (24) gives

$$\frac{d}{dt} \varphi_t(x_{\varepsilon(t)}) = \frac{1}{2} \dot{\varepsilon}(t) \|x_{\varepsilon(t)}\|^2.$$

(ii) We have

$$-\varepsilon(t)x_{\varepsilon(t)} = \nabla f(x_{\varepsilon(t)}) \quad \text{and} \quad -\varepsilon(t+h)x_{\varepsilon(t+h)} = \nabla f(x_{\varepsilon(t+h)}).$$

According to the monotonicity of ∇f , we have

$$\langle \varepsilon(t)x_{\varepsilon(t)} - \varepsilon(t+h)x_{\varepsilon(t+h)}, x_{\varepsilon(t+h)} - x_{\varepsilon(t)} \rangle \geq 0,$$

which implies

$$-\varepsilon(t) \|x_{\varepsilon(t+h)} - x_{\varepsilon(t)}\|^2 + (\varepsilon(t) - \varepsilon(t+h)) \langle x_{\varepsilon(t+h)}, x_{\varepsilon(t+h)} - x_{\varepsilon(t)} \rangle \geq 0.$$

After division by h^2 , we obtain

$$\frac{(\varepsilon(t) - \varepsilon(t+h))}{h} \left\langle x_{\varepsilon(t+h)}, \frac{x_{\varepsilon(t+h)} - x_{\varepsilon(t)}}{h} \right\rangle \geq \varepsilon(t) \left\| \frac{x_{\varepsilon(t+h)} - x_{\varepsilon(t)}}{h} \right\|^2. \quad (25)$$

We rely on the differentiability properties of the viscosity curve $\epsilon \mapsto x_\epsilon = \text{argmin} \{ f(\xi) + \frac{\epsilon}{2} \|\xi\|^2 \}$. According to [13], [30], [41], the viscosity curve is Lipschitz continuous on the compact intervals of $]0, +\infty[$. So it is absolutely continuous, and almost everywhere differentiable. Therefore, the mapping $t \mapsto x_{\varepsilon(t)}$ satisfies the same differentiability properties. By letting $h \rightarrow 0$ in (25) we obtain that, for almost every $t \geq t_0$

$$-\dot{\varepsilon}(t) \left\langle x_{\varepsilon(t)}, \frac{d}{dt} x_{\varepsilon(t)} \right\rangle \geq \varepsilon(t) \left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2,$$

which gives the claim. The last statement follows from Cauchy-Schwarz inequality. \square

2.4 The steepest descent with vanishing Tikhonov regularization: Lyapunov analysis

In the Lyapunov analysis of the continuous steepest descent with vanishing Tikhonov regularization (i.e. $\varepsilon(t) \rightarrow 0$ as $t \rightarrow +\infty$) that we recall below

$$\dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0, \quad (26)$$

the following function plays a central role

$$E(t) := (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \frac{\varepsilon(t)}{2} \|x(t) - x_{\varepsilon(t)}\|^2, \quad (27)$$

together with

$$\gamma(t) := \exp\left(\int_{t_1}^t \varepsilon(s) ds\right). \quad (28)$$

Let us state our main convergence result.

Theorem 3 *Let $x(\cdot) : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of the system (2). Define $E(t)$ and $\gamma(\cdot)$ respectively by (27) and (28).*

1. *Then, the following properties are satisfied: there exists $t_1 > 0$ such that for all $t \geq t_1$*

$$E(t) \leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} - \frac{\|x^*\|^2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds; \quad (29)$$

$$f(x(t)) - \min_{\mathcal{H}} f \leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \frac{\|x^*\|^2}{2} \left[\varepsilon(t) - \frac{2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds \right]; \quad (30)$$

$$\|x(t) - x_{\varepsilon(t)}\|^2 \leq \frac{E(t)}{\varepsilon(t)} \leq \frac{\gamma(t_1)E(t_1)}{\varepsilon(t)\gamma(t)} - \frac{\|x^*\|^2}{\varepsilon(t)\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds. \quad (31)$$

2. *Suppose that one of the two following conditions is satisfied:*

$$(i) \lim_{t \rightarrow +\infty} \frac{E(t)}{\varepsilon(t)} = 0$$

or

$$(ii) \int_{t_0}^{+\infty} \varepsilon(t)dt = +\infty.$$

Then $x(t)$ converges strongly to x^ as $t \rightarrow +\infty$.*

Proof Since the mapping $t \mapsto x_{\varepsilon(t)}$ is absolutely continuous (indeed locally Lipschitz) the classical derivation chain rule can be applied to compute the derivative of the function $E(\cdot)$, see [24, section VIII.2]. According to Lemma 4 i), for almost all $t \geq t_0$ the derivative of $E(\cdot)$ is given by:

$$\begin{aligned} \dot{E}(t) &= \langle \nabla \varphi_t(x(t)), \dot{x}(t) \rangle + \frac{1}{2} \dot{\varepsilon}(t) \|x(t)\|^2 - \frac{1}{2} \dot{\varepsilon}(t) \|x_{\varepsilon(t)}\|^2 \\ &\quad + \frac{1}{2} \dot{\varepsilon}(t) \|x(t) - x_{\varepsilon(t)}\|^2 + \varepsilon(t) \left\langle \frac{d}{dt} (x(t) - x_{\varepsilon(t)}), x(t) - x_{\varepsilon(t)} \right\rangle. \end{aligned} \quad (32)$$

According to (17) and the $\varepsilon(t)$ strong convexity of φ_t , we get

$$\begin{aligned} \dot{E}(t) &= -\|\nabla \varphi_t(x(t))\|^2 + \frac{1}{2} \dot{\varepsilon}(t) \left(\|x(t)\|^2 - \|x_{\varepsilon(t)}\|^2 \right) + \frac{1}{2} \dot{\varepsilon}(t) \|x(t) - x_{\varepsilon(t)}\|^2 \\ &\quad - \varepsilon(t) \left\langle \nabla \varphi_t(x(t)), x(t) - x_{\varepsilon(t)} \right\rangle - \varepsilon(t) \left\langle \frac{d}{dt} (x_{\varepsilon(t)}), x(t) - x_{\varepsilon(t)} \right\rangle \\ &\leq -\|\nabla \varphi_t(x(t))\|^2 + \frac{1}{2} \dot{\varepsilon}(t) \left(\|x(t)\|^2 - \|x_{\varepsilon(t)}\|^2 \right) + \frac{1}{2} \dot{\varepsilon}(t) \|x(t) - x_{\varepsilon(t)}\|^2 \\ &\quad + \varepsilon(t) \left(\varphi_t(x_{\varepsilon(t)}) - \varphi_t(x(t)) - \frac{\varepsilon(t)}{2} \|x(t) - x_{\varepsilon(t)}\|^2 \right) - \varepsilon(t) \left\langle \frac{d}{dt} (x_{\varepsilon(t)}), x(t) - x_{\varepsilon(t)} \right\rangle. \end{aligned} \quad (33)$$

According to Lemma 1 (i) and Lemma 4 (ii), we obtain

$$\begin{aligned} -\left\langle \frac{d}{dt}(x_{\varepsilon(t)}), x(t) - x_{\varepsilon(t)} \right\rangle &\leq \frac{b(t)}{2} \left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2 + \frac{1}{2b(t)} \|x(t) - x_{\varepsilon(t)}\|^2 \\ &\leq \frac{b(t)\dot{\varepsilon}^2(t)}{2\varepsilon^2(t)} \|x_{\varepsilon(t)}\|^2 + \frac{1}{2b(t)} \|x(t) - x_{\varepsilon(t)}\|^2, \end{aligned} \quad (34)$$

where $b(\cdot)$ is a continuous and positive real function to be chosen later. Combining (33) with (34), we deduce that

$$\begin{aligned} \dot{E}(t) &\leq -\|\nabla \varphi_t(x(t))\|^2 + \frac{1}{2}\dot{\varepsilon}(t)\|x(t)\|^2 - \frac{1}{2}\varepsilon(t) \left(\varepsilon(t) - \frac{\dot{\varepsilon}(t)}{\varepsilon(t)} - \frac{1}{b(t)} \right) \|x(t) - x_{\varepsilon(t)}\|^2 \\ &\quad + \varepsilon(t) (\varphi_t(x_{\varepsilon(t)}) - \varphi_t(x(t))) + \frac{\dot{\varepsilon}(t)}{2} \left(\frac{b(t)\dot{\varepsilon}(t)}{\varepsilon(t)} - 1 \right) \|x_{\varepsilon(t)}\|^2. \end{aligned}$$

Let us now build the differential inequality satisfied by $E(\cdot)$.

$$\begin{aligned} \dot{E}(t) + \varepsilon(t)E(t) &\leq -\|\nabla \varphi_t(x(t))\|^2 + \frac{1}{2}\dot{\varepsilon}(t)\|x(t)\|^2 \\ &\quad - \frac{1}{2}\varepsilon(t) \left(-\frac{\dot{\varepsilon}(t)}{\varepsilon(t)} - \frac{1}{b(t)} \right) \|x(t) - x_{\varepsilon(t)}\|^2 + \frac{\dot{\varepsilon}(t)}{2} \left(\frac{b(t)\dot{\varepsilon}(t)}{\varepsilon(t)} - 1 \right) \|x_{\varepsilon(t)}\|^2. \end{aligned} \quad (35)$$

Let us now choose $b(\cdot)$ in order to make equal to zero the coefficient of $\|x(t) - x_{\varepsilon(t)}\|^2$, that is

$$\frac{1}{b(t)} = -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}$$

Replacing $b(\cdot)$ by this expression in (35), we obtain

$$\begin{aligned} \dot{E}(t) + \varepsilon(t)E(t) &\leq -\|\nabla \varphi_t(x(t))\|^2 + \frac{1}{2}\dot{\varepsilon}(t)\|x(t)\|^2 - \dot{\varepsilon}(t)\|x_{\varepsilon(t)}\|^2 \\ &\leq -\dot{\varepsilon}(t)\|x_{\varepsilon(t)}\|^2. \end{aligned} \quad (36)$$

By taking $W(t) := \gamma(t)E(t)$, we conclude that

$$\begin{aligned} \dot{W}(t) &= \gamma(t) (\dot{E}(t) + \varepsilon(t)E(t)) \\ &\leq -\gamma(t)\dot{\varepsilon}(t)\|x^*\|^2, \end{aligned}$$

Therefore

$$\frac{d}{dt} (\gamma(t)E(t)) \leq -\gamma(t)\dot{\varepsilon}(t)\|x^*\|^2. \quad (37)$$

By integrating (37) on $[t_1, t]$, and dividing by $\gamma(t)$, we obtain our first claim (29)

$$E(t) \leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} - \frac{\|x^*\|^2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds. \quad (38)$$

Coming back to Lemma 3, and according to (38), we get, for any $t \geq t_1$,

$$\begin{aligned} f(x(t)) - \min_{\mathcal{H}} f &\leq E(t) + \frac{\varepsilon(t)}{2} \|x^*\|^2 \\ &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \frac{\|x^*\|^2}{2} \left[\varepsilon(t) - \frac{2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds \right] \end{aligned}$$

Similarly, according to Lemma 3 and (38), we get, for any $t \geq t_1$,

$$\begin{aligned} \|x(t) - x_{\varepsilon(t)}\|^2 &\leq \frac{E(t)}{\varepsilon(t)} \\ &\leq \frac{\gamma(t_1)E(t_1)}{\varepsilon(t)\gamma(t)} - \frac{\|x^*\|^2}{\varepsilon(t)\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds. \end{aligned}$$

By Lemma 1 (ii), $\|x_{\varepsilon(t)} - x^*\|$ converges to zero. Therefore, $x(t)$ converges strongly to x^* as soon as $\lim_{t \rightarrow +\infty} \frac{E(t)}{\varepsilon(t)} = 0$.

Finally, according to Cominetti-Peypouquet-Sorin [29, Theorem 2], we have that $x(t)$ converges strongly to the minimum norm solution x^* as soon as $\int_{t_0}^{+\infty} \varepsilon(t)dt = +\infty$. \square

2.5 Case $\varepsilon(t) = \frac{\delta}{t}$, $\delta > 1$

The convergence rate of the values and the strong convergence to the minimum norm solution will be obtained by particularizing Theorem 3 to this situation.

Theorem 4 Take $\varepsilon(t) = \frac{\delta}{t}$ and $\delta > 1$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\dot{x}(t) + \nabla f(x(t)) + \frac{\delta}{t}x(t) = 0. \quad (39)$$

Then, we have

$$\bullet E(t) = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty; \quad (40)$$

$$\bullet f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty; \quad (41)$$

$$\bullet \|\nabla f(x(t))\| = \mathcal{O}\left(\frac{1}{t^{\frac{1}{2}}}\right) \text{ as } t \rightarrow +\infty. \quad (42)$$

- The solution trajectory $x(\cdot)$ converges strongly to the minimum norm solution x^* .

Proof a) Take $\varepsilon(t) = \frac{\delta}{t}$ with $\delta > 1$. We get

$$\gamma(t) := \exp\left(\int_{t_1}^t \varepsilon(s)ds\right) = \exp\left(\ln\left(\frac{t}{t_1}\right)\right) = \left(\frac{t}{t_1}\right)^{\delta}.$$

By Theorem 3, we get, for t large enough

$$\begin{aligned} E(t) &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} - \frac{\|x^*\|^2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds \\ &\leq \frac{C}{t^{\delta}} + \|x^*\|^2 \left(\frac{t_1}{t}\right)^{\delta} \int_{t_1}^t \frac{\delta}{s^2} \left(\frac{s}{t_1}\right)^{\delta} ds \\ &\leq \frac{C}{t^{\delta}} + \delta \|x^*\|^2 \frac{1}{t^{\delta}} \int_{t_1}^t s^{\delta-2} ds \\ &\leq \frac{C}{t^{\delta}} + \frac{\delta}{\delta-1} \|x^*\|^2 \frac{1}{t}. \end{aligned}$$

Since $\delta > 1$ we deduce that

$$E(t) = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty.$$

According to Theorem 3 we get

$$\begin{aligned} f(x(t)) - \min_{\mathcal{H}} f &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \frac{\|x^*\|^2}{2} \left[\varepsilon(t) - \frac{2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds \right] \\ &= \frac{C}{t^\delta} + \delta \frac{\|x^*\|^2}{2t} + \frac{\delta}{\delta-1} \|x^*\|^2 \frac{1}{t}. \end{aligned}$$

Since $\delta > 1$ we deduce that

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty.$$

According to the above estimate and Lemma 5 we immediately obtain the following convergence rate of the gradients towards zero

$$\|\nabla f(x(t))\| = \mathcal{O}\left(\frac{1}{t^{\frac{1}{2}}}\right) \text{ as } t \rightarrow +\infty.$$

Finally, since $\int_{t_0}^{+\infty} \varepsilon(t)dt = +\infty$, according to Cominetti-Peypouquet-Sorin [29, Theorem 2], we have that $x(t)$ converges strongly to the minimum norm solution x^* . \square

2.6 Case $\varepsilon(t) = \frac{1}{t^r}$, $0 < r < 1$

Take $\varepsilon(t) = \frac{1}{t^r}$, $0 < r < 1$, $t_0 > 0$, and consider the system (2). The convergence rate of the values and the strong convergence to the minimum norm solution will be obtained by particularizing Theorem 3 to this situation.

Theorem 5 Take $\varepsilon(t) = \frac{1}{t^r}$ and $0 < r < 1$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r} x(t) = 0. \quad (43)$$

Then, we have

$$\bullet E(t) = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty; \quad (44)$$

$$\bullet f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^r}\right) \text{ as } t \rightarrow +\infty; \quad (45)$$

$$\bullet \|x(t) - x_{\varepsilon(t)}\|^2 = \mathcal{O}\left(\frac{1}{t^{1-r}}\right) \text{ as } t \rightarrow +\infty. \quad (46)$$

$$\bullet \|\nabla f(x(t))\| = \mathcal{O}\left(\frac{1}{t^{\frac{r}{2}}}\right) \text{ as } t \rightarrow +\infty. \quad (47)$$

$$\bullet \text{The solution trajectory } x(\cdot) \text{ converges strongly to the minimum norm solution } x^*. \quad (48)$$

Proof a) Take $\varepsilon(t) = \frac{1}{t^r}$ with $r < 1$. We get

$$\gamma(t) := \exp\left(\int_{t_1}^t \varepsilon(s)ds\right) = \exp\left(\left[\frac{t^{1-r}}{1-r} - \frac{t_1^{1-r}}{1-r}\right]\right) = K_0 \exp\left(\frac{1}{1-r} t^{1-r}\right)$$

By Theorem 3, we get, for t large enough

$$\begin{aligned} E(t) &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} - \frac{\|x^*\|^2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds \\ &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \|x^*\|^2 \exp\left(-\frac{1}{1-r}t^{1-r}\right) \int_{t_1}^t \frac{r}{s^{r+1}} \exp\left(\frac{1}{1-r}s^{1-r}\right) ds. \end{aligned}$$

To majorize this last expression, we notice that, given a positive parameter ρ

$$\begin{aligned} \frac{d}{ds} \left(\frac{1}{\rho s} \exp\left(\frac{1}{1-r}s^{1-r}\right) \right) &= \frac{1}{\rho} \left(\frac{1}{s^{1+r}} - \frac{1}{s^2} \right) \exp\left(\frac{1}{1-r}s^{1-r}\right) \\ &\geq \frac{r}{s^{r+1}} \exp\left(\frac{1}{1-r}s^{1-r}\right) \end{aligned}$$

as soon as $\rho < \frac{1}{r}$ (for s sufficiently large). Therefore

$$\begin{aligned} E(t) &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \|x^*\|^2 \exp\left(-\frac{1}{1-r}t^{1-r}\right) \int_{t_1}^t \frac{d}{ds} \left(\frac{1}{\rho s} \exp\left(\frac{1}{1-r}s^{1-r}\right) \right) ds \\ &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \frac{\|x^*\|^2}{\rho t}. \end{aligned}$$

Since the term $\frac{\gamma(t_1)E(t_1)}{\gamma(t)}$ converges to zero exponentially, we get

$$E(t) = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty;$$

Similarly, according to Theorem 3

$$\begin{aligned} f(x(t)) - \min_{\mathcal{H}} f &\leq \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \frac{\|x^*\|^2}{2} \left[\varepsilon(t) - \frac{2}{\gamma(t)} \int_{t_1}^t \dot{\varepsilon}(s)\gamma(s)ds \right] \\ &= \frac{\gamma(t_1)E(t_1)}{\gamma(t)} + \frac{\|x^*\|^2}{2t^r} + \frac{\|x^*\|^2}{\rho t}. \end{aligned}$$

Since the term $\frac{\gamma(t_1)E(t_1)}{\gamma(t)}$ converges to zero exponentially, we get

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^r}\right) \text{ as } t \rightarrow +\infty;$$

According to the above estimate and Lemma 5 we immediately obtain the following convergence rate of the gradients towards zero

$$\|\nabla f(x(t))\| = \mathcal{O}\left(\frac{1}{t^{\frac{r}{2}}}\right) \text{ as } t \rightarrow +\infty.$$

Finally, according to Theorem 3 and $E(t) = \mathcal{O}\left(\frac{1}{t}\right)$, we get

$$\|x(t) - x_{\varepsilon(t)}\|^2 \leq \frac{E(t)}{\varepsilon(t)} \leq \frac{C}{t} t^r = \frac{C}{t^{1-r}}.$$

Since $1-r > 0$ we conclude that $\|x_{\varepsilon(t)} - x(t)\|$ converges to zero. According to lemma 1 we have $\lim_{t \rightarrow +\infty} \|x_{\varepsilon(t)} - x^*\| = 0$. Therefore $x(t)$ converges strongly to x^* . Indeed, this could be obtained directly as a consequence of the Cominetti-Peypouquet-Sorin [29, Theorem 2]. Our Lyapunov analysis provides indeed a convergence rate. This completes the proof. \square

Remark 1 Let us provide another convergence rate of the gradients towards zero, of interest when $r \leq \frac{1}{2}$. Let us observe that

$$\begin{aligned} \|\nabla f(x(t))\| &\leq \|\nabla f(x(t)) - \nabla f(x_{\varepsilon(t)})\| + \|\nabla f(x_{\varepsilon(t)})\| \\ &\leq L\|x(t) - x_{\varepsilon(t)}\| + \varepsilon(t)\|x_{\varepsilon(t)}\| \\ &= \mathcal{O}\left(\frac{1}{t^{\frac{1-r}{2}}} + \frac{1}{t^r}\right). \end{aligned}$$

3 Passing from the first-order to the second-order differential equation

Having in view to develop rapid optimization methods, our study focuses on the second-order time evolution system obtained by applying the "time scaling and averaging" method of Attouch, Bot and Ngyuen [7] to the dynamic (2). In doing so, we hope to take advantage of both the fast convergence properties attached to the inertial gradient methods with vanishing damping coefficients and the strong convexity properties attached to Tikhonov's regularization. So, let us make the change of time variable $s = \tau(t)$ in the damped inertial dynamic

$$\dot{z}(s) + \nabla f(z(s)) + \varepsilon(s)z(s) = 0, \quad (49)$$

where we take z as a state variable and s as a time variable, which will end up with x and t after time scaling and averaging. The time scale $\tau(\cdot)$ is a \mathcal{C}^2 increasing function from $[t_0, +\infty[$ to $[s_0, +\infty[$, which satisfies $\lim_{t \rightarrow +\infty} \tau(t) = +\infty$. Setting $s = \tau(t)$ and multiplying (49) by $\dot{\tau}(t) > 0$, we obtain

$$\dot{\tau}(t)\dot{z}(\tau(t)) + \dot{\tau}(t)\left(\nabla f(z(\tau(t))) + \varepsilon(\tau(t))z(\tau(t))\right) = 0. \quad (50)$$

Set $v(t) := z(\tau(t))$. By the derivation chain rule, we have $\dot{v}(t) = \dot{\tau}(t)\dot{z}(\tau(t))$. So reformulating (50) in terms of $v(\cdot)$ and its derivatives, we obtain

$$\dot{v}(t) + \dot{\tau}(t)\nabla f(v(t)) + \dot{\tau}(t)\varepsilon(\tau(t))v(t) = 0. \quad (51)$$

Let us now consider the averaging process which consists in passing from v to x which is defined by means of the first-order linear differential equation

$$v(t) = x(t) + \dot{\tau}(t)\dot{x}(t). \quad (52)$$

By temporal derivation of (52) we get

$$\dot{v}(t) = \dot{x}(t) + \ddot{\tau}(t)\dot{x}(t) + \dot{\tau}(t)\ddot{x}(t). \quad (53)$$

Replacing $v(t)$ and $\dot{v}(t)$ in (51), we get

$$\dot{\tau}(t)\ddot{x}(t) + (1 + \ddot{\tau}(t))\dot{x}(t) + \dot{\tau}(t)\nabla f[x(t) + \dot{\tau}(t)\dot{x}(t)] + \dot{\tau}(t)\varepsilon(\tau(t))[x(t) + \dot{\tau}(t)\dot{x}(t)] = 0. \quad (54)$$

Dividing by $\dot{\tau}(t) > 0$, we finally obtain

$$\ddot{x}(t) + \frac{1 + \ddot{\tau}(t)}{\dot{\tau}(t)}\dot{x}(t) + \nabla f[x(t) + \dot{\tau}(t)\dot{x}(t)] + \varepsilon(\tau(t))[x(t) + \dot{\tau}(t)\dot{x}(t)] = 0. \quad (55)$$

According to the Su, Boyd and Candès [38] model for the Nesterov accelerated gradient method, we consider the case where the viscous damping coefficient in (55) satisfies $\frac{1 + \ddot{\tau}(t)}{\dot{\tau}(t)} = \frac{\alpha}{t}$ for some

$\alpha > 1$. Thus $\tau(t) = \frac{t^2}{2(\alpha - 1)}$, and the system (55) becomes

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f\left[x(t) + \frac{t}{\alpha - 1}\dot{x}(t)\right] + \varepsilon\left(\frac{t^2}{2(\alpha - 1)}\right)\left[x(t) + \frac{t}{\alpha - 1}\dot{x}(t)\right] = 0. \quad (56)$$

Equivalently

$$\ddot{x}(t) + \left(\frac{\alpha}{t} + \frac{t}{\alpha-1} \varepsilon \left(\frac{t^2}{2(\alpha-1)} \right) \right) \dot{x}(t) + \nabla f \left[x(t) + \frac{t}{\alpha-1} \dot{x}(t) \right] + \varepsilon \left(\frac{t^2}{2(\alpha-1)} \right) x(t) = 0. \quad (57)$$

We can observe that the time dependent parameter $\varepsilon(\cdot)$ enters both the damping coefficient and the Tikhonov regularization term. In parallel to the study of the first-order evolution system studied in the previous section, we consider the two cases of the function $\varepsilon(\cdot)$ which is in our hands, namely $\varepsilon(t) = \frac{\delta}{t}$ then $\varepsilon(t) = \frac{1}{t^\tau}$. We give two different proofs of convergence results, each relying on a specific technique of independent interest.

3.1 $\varepsilon(t) = \frac{\delta}{t}$ with $\delta > 1$

The system (56) becomes

$$\ddot{x}(t) + \frac{\alpha+2\delta}{t} \dot{x}(t) + \nabla f \left[x(t) + \frac{t}{\alpha-1} \dot{x}(t) \right] + \frac{2\delta(\alpha-1)}{t^2} x(t) = 0. \quad (58)$$

Let us state the main convergence properties of this system.

Theorem 6 *Take $\alpha > 3$ and $\delta > 1$, which gives $\alpha+2\delta > 5$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of (58). Then, the following properties are satisfied.*

$$(i) \ f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^2} \right) \text{ as } t \rightarrow +\infty; \quad (59)$$

$$(ii) \ \text{There is strong convergence of } x(t) \text{ to the minimum norm solution } x^*.$$

$$(iii) \ \|\nabla f(x(t))\| = \mathcal{O} \left(\frac{1}{t} \right) \text{ as } t \rightarrow +\infty. \quad (60)$$

Proof According to Theorem 4, the rescaled function $v(t) := z(\tau(t))$ satisfies

$$f(v(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^2} \right) \text{ as } t \rightarrow +\infty; \quad (61)$$

Our objective is now to obtain the corresponding convergence rate of the values $f(x(t)) - \min_{\mathcal{H}} f$ as $t \rightarrow +\infty$, and the strong convergence of $x(t)$ to x^* the minimum norm element of S . The following proof is inspired by Attouch, Bot, Ngyuen [7]. It highlights the averaging interpretation of the passage from v to x .

(i) Let us rewrite (52) as

$$t\dot{x}(t) + (\alpha-1)x(t) = (\alpha-1)v(t). \quad (62)$$

After multiplication of (62) by $t^{\alpha-2}$, we get equivalently

$$t^{\alpha-1}\dot{x}(t) + (\alpha-1)t^{\alpha-2}x(t) = (\alpha-1)t^{\alpha-2}v(t), \quad (63)$$

that is

$$\frac{d}{dt} \left(t^{\alpha-1}x(t) \right) = (\alpha-1)t^{\alpha-2}v(t). \quad (64)$$

By integrating (64) from t_0 to t , and according to $x(t_0) = x_0$, we obtain

$$x(t) = \frac{t_0^{\alpha-1}}{t^{\alpha-1}} v(t_0) + \frac{\alpha-1}{t^{\alpha-1}} \int_{t_0}^t \theta^{\alpha-2} v(\theta) d\theta \quad (65)$$

where for simplicity we take $v(s_0) = x_0$. Then, observe that $x(t)$ can be simply written as follows

$$x(t) = \int_{t_0}^t v(\theta) d\mu_t(\theta) \quad (66)$$

where μ_t is the positive Radon measure on $[t_0, t]$ defined by

$$\mu_t = \frac{t_0^{\alpha-1}}{t^{\alpha-1}} \delta_{t_0} + (\alpha-1) \frac{\theta^{\alpha-2}}{t^{\alpha-1}} d\theta.$$

We are therefore led to examining the convergence rate of $f\left(\int_{t_0}^t v(\theta) d\mu_t(\theta)\right) - \min_{\mathcal{H}} f$ towards zero as $t \rightarrow +\infty$. We have that μ_t is a positive Radon measure on $[t_0, t]$ whose total mass is equal to 1. It is therefore a probability measure, and $\int_{t_0}^t v(\theta) d\mu_t(\theta)$ is obtained by **averaging** the trajectory $v(\cdot)$ on $[t_0, t]$ with respect to μ_t . From there, we can deduce fast convergence properties of $x(\cdot)$. According to the convexity of f , and Jensen's inequality, we obtain that

$$\begin{aligned} f\left(\int_{t_0}^t v(\theta) d\mu_t(\theta)\right) - \min_{\mathcal{H}} f &= (f - \min_{\mathcal{H}} f) \left(\int_{t_0}^t v(\theta) d\mu_t(\theta)\right) \\ &\leq \int_{t_0}^t \left(f(v(\theta)) - \min_{\mathcal{H}} f\right) d\mu_t(\theta) \\ &\leq L_f \int_{t_0}^t \frac{1}{\theta^2} d\mu_t(\theta) \quad \forall t \geq t_0, \end{aligned}$$

where the last inequality above comes from (61). According to the definition of μ_t , it yields

$$\begin{aligned} f\left(\int_{t_0}^t v(\theta) d\mu_t(\theta)\right) - \min_{\mathcal{H}} f &\leq \frac{L_f t_0^{\alpha-3}}{t^{\alpha-1}} + L_f (\alpha-1) \frac{1}{t^{\alpha-1}} \int_{t_0}^t \theta^{\alpha-4} d\theta \\ &\leq \frac{L_f t_0^{\alpha-3}}{t^{\alpha-1}} + L_f \frac{\alpha-1}{\alpha-3} \left(\frac{1}{t^2} - \frac{t_0^{\alpha-3}}{t^{\alpha-1}}\right) \quad \forall t \geq t_0. \end{aligned}$$

According to (66) we get

- For $1 < \alpha < 3$,

$$f(x(t)) - \min_{\mathcal{H}} f \leq \frac{C}{t^{\alpha-1}}.$$

- For $\alpha > 3$, we have $\frac{t_0^{\alpha-3}}{t^{\alpha-1}} \leq \frac{1}{t^2}$ for every $t \geq t_0$. We therefore obtain

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right) \text{ as } t \rightarrow +\infty$$

(ii) According to Theorem 4, the rescaled function $v(\cdot)$ strongly converges to the minimum norm solution. From the interpretation of x as an average of v , and using the fact that convergence implies ergodic convergence, we deduce the strong convergence of $x(t)$ towards the minimum norm solution. This argument is detailed in the next paragraph, so we omit the details here.

(iii) We have obtained in ii) that the trajectory $x(\cdot)$ converges. Therefore, it is bounded. Let $L > 0$ be the Lipschitz constant of ∇f on a ball that contains the trajectory $x(\cdot)$. According to the convergence rate of values and Lemma 5 in the Appendix, we immediately obtain the following convergence rate of the gradients towards zero

$$\frac{1}{2L} \|\nabla f(x(t))\|^2 \leq f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right).$$

So,

$$\|\nabla f(x(t))\| = \mathcal{O}\left(\frac{1}{t}\right) \text{ as } t \rightarrow +\infty.$$

This completes the proof. \square

3.2 $\varepsilon(t) = \frac{1}{t^r}$ for $0 < r < 1$

The system (56) becomes

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f \left[x(t) + \frac{s}{\alpha-1} \dot{x}(t) \right] + \left(\frac{2(\alpha-1)}{t^2} \right)^r \left[x(t) + \frac{t}{\alpha-1} \dot{x}(t) \right] = 0. \quad (67)$$

According to Theorem 5, the rescaled function $v(t) := z(\tau(t))$ satisfies

$$f(v(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{\tau(t)^r} \right) \text{ as } t \rightarrow +\infty; \quad (68)$$

$$\|v(t) - x_{\varepsilon(\tau(t))}\|^2 = \mathcal{O} \left(\frac{1}{\tau(t)^{1-r}} \right) \text{ as } t \rightarrow +\infty. \quad (69)$$

Our objective is now to obtain the corresponding convergence rate of the values $f(x(t)) - \min_{\mathcal{H}} f$ as $t \rightarrow +\infty$, and strong convergence of $x(t)$ to x^* the minimum norm element of S .

Theorem 7 Take $0 < r < 1$ and $\alpha > 1$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\ddot{x}(t) + \left(\frac{\alpha}{t} + \frac{2^r(\alpha-1)^{r-1}}{t^{2r-1}} \right) \dot{x}(t) + \nabla f \left(x(t) + \frac{t}{\alpha-1} \dot{x}(t) \right) + \frac{2^r(\alpha-1)^r}{t^{2r}} x(t) = 0. \quad (70)$$

Then, the following properties are satisfied.

$$(i) \ f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}} \right) \text{ as } t \rightarrow +\infty; \quad (71)$$

$$(ii) \ \text{There is strong convergence of } x(t) \text{ to the minimum norm solution } x^*; \quad (72)$$

$$(iii) \ \|\nabla f(x(t))\|^2 = \mathcal{O} \left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}} \right) \text{ as } t \rightarrow +\infty. \quad (73)$$

Proof (i) Convergence rates to zero for the values $f(x(t)) - \min_{\mathcal{H}} f$. By definition of $x(\cdot)$

$$v(s) = x(s) + \frac{s}{\alpha-1} \dot{x}(s) = \frac{1}{(\alpha-1)s^{\alpha-2}} \frac{d}{ds} (s^{\alpha-1} x(s)).$$

Equivalently

$$\frac{d}{ds} (s^{\alpha-1} x(s)) = (\alpha-1) s^{\alpha-2} v(s). \quad (74)$$

From this it is easy to verify that x can be interpreted as an average of v . After integration from t to $t+h$ of (74) and division by $(t+h)^{\alpha-1}$, we get

$$\begin{aligned} x(t+h) &= \left(\frac{t}{t+h} \right)^{\alpha-1} x(t) + \frac{1}{(t+h)^{\alpha-1}} \int_t^{t+h} (\alpha-1) s^{\alpha-2} v(s) ds \\ &= \left(\frac{t}{t+h} \right)^{\alpha-1} x(t) + \left(1 - \left(\frac{t}{t+h} \right)^{\alpha-1} \right) \frac{1}{(t+h)^{\alpha-1} - t^{\alpha-1}} \int_t^{t+h} (\alpha-1) s^{\alpha-2} v(s) ds \\ &= \left(\frac{t}{t+h} \right)^{\alpha-1} x(t) + \left(1 - \left(\frac{t}{t+h} \right)^{\alpha-1} \right) \frac{1}{(t+h)^{\alpha-1} - t^{\alpha-1}} \int_{t^{\alpha-1}}^{(t+h)^{\alpha-1}} v \left(\theta^{1/(\alpha-1)} \right) d\theta, \end{aligned}$$

where the last equality comes from the change of time variable $\theta = s^{\alpha-1}$.

According to the convexity of the function $F = f - \inf_{\mathcal{H}} f$, and using Jensen's inequality, we obtain

$$\begin{aligned} F(x(t+h)) &\leq \left(\frac{t}{t+h} \right)^{\alpha-1} F(x(t)) + \left(1 - \left(\frac{t}{t+h} \right)^{\alpha-1} \right) F \left(\frac{1}{(t+h)^{\alpha-1} - t^{\alpha-1}} \int_{t^{\alpha-1}}^{(t+h)^{\alpha-1}} v \left(\theta^{1/(\alpha-1)} \right) d\theta \right) \\ &\leq \left(\frac{t}{t+h} \right)^{\alpha-1} F(x(t)) + \left(1 - \left(\frac{t}{t+h} \right)^{\alpha-1} \right) \frac{1}{(t+h)^{\alpha-1} - t^{\alpha-1}} \int_{t^{\alpha-1}}^{(t+h)^{\alpha-1}} F \left(v \left(\theta^{1/(\alpha-1)} \right) \right) d\theta. \end{aligned}$$

Using again the change of time variable $s = \theta^{1/(\alpha-1)}$, we get

$$\int_{t^{\alpha-1}}^{(t+h)^{\alpha-1}} F\left(v\left(\theta^{1/(\alpha-1)}\right)\right) d\theta = \int_t^{t+h} (\alpha-1)s^{\alpha-2} F(v(s)) ds.$$

It follows

$$F(x(t+h)) \leq \left(\frac{t}{t+h}\right)^{\alpha-1} F(x(t)) + \left(1 - \left(\frac{t}{t+h}\right)^{\alpha-1}\right) \frac{1}{(t+h)^{\alpha-1} - t^{\alpha-1}} \int_t^{t+h} (\alpha-1)s^{\alpha-2} F(v(s)) ds.$$

Using (68), with $\tau(s) = \frac{s^2}{2(\alpha-1)}$, we obtain the existence of a constant $C > 0$ such that

$$F(x(t+h)) \leq \left(\frac{t}{t+h}\right)^{\alpha-1} F(x(t)) + \left(1 - \left(\frac{t}{t+h}\right)^{\alpha-1}\right) \frac{C(\alpha-1)}{(t+h)^{\alpha-1} - t^{\alpha-1}} \int_t^{t+h} s^{\alpha-2} \left(\frac{2(\alpha-1)}{s^2}\right)^r ds.$$

Therefore, multiplying by $\frac{1}{h}(t+h)^{\alpha-1}$ and rearranging the terms of this inequality, we conclude

$$\frac{1}{h} \left((t+h)^{\alpha-1} F(x(t+h)) - t^{\alpha-1} F(x(t)) \right) \leq 2^r C(\alpha-1)^{r+1} \frac{1}{h} \int_t^{t+h} \frac{ds}{s^{2r-\alpha+2}} ds.$$

By letting $h \rightarrow 0^+$ in the above inequality, we first get that $\frac{d}{dt} \Gamma(t) \leq 0$, where

$$\Gamma(t) := t^{\alpha-1} F(x(t)) - 2^r C(\alpha-1)^{r+1} \int_{s_0}^t \frac{ds}{s^{2r-\alpha+2}} ds.$$

This implies $\Gamma(t)$ is nonincreasing. Therefore for each $0 < r < 1$ and each $t \geq s_0$, we have $\Gamma(t) \leq \Gamma(t_0)$, which gives

$$t^{\alpha-1} F(x(t)) \leq \Gamma(t_0) + 2^r C(\alpha-1)^{r+1} \int_{s_0}^t \frac{ds}{s^{2r-\alpha+2}} ds.$$

Consequently,

$$f(x(t)) - \inf_{\mathcal{H}} f = F(x(t)) \leq \frac{\Gamma(t_0)}{t^{\alpha-1}} + C_1 \frac{1}{t^{\alpha-1}} t^{\alpha-2r-1}. \quad (75)$$

We conclude that, as $t \rightarrow +\infty$

$$f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}}\right)$$

As a consequence

– When $\alpha \geq 3$ we have

$$f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{2r}}\right)$$

So we can get as close as possible to the optimal convergence rate $1/t^2$.

– When $\alpha < 3$, by taking $\frac{\alpha-1}{2} < r < 1$, we get

$$f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{\alpha-1}}\right).$$

(ii) *Strong convergence of $x(t)$ to the minimum norm solution.*

Let us go back to (74). We have

$$\frac{d}{ds}(s^{\alpha-1}x(s)) = (\alpha-1)s^{\alpha-2}v(s).$$

Integrating from s_0 to t , we get

$$\begin{aligned} x(t) &= \left(\frac{s_0}{t}\right)^{\alpha-1} x(s_0) + \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} v(\theta) d\theta \\ &= \left(\frac{s_0}{t}\right)^{\alpha-1} x(s_0) + \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} (v(\theta) - x^*) d\theta + \left(\frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} d\theta\right) x^* \\ &= \left(\frac{s_0}{t}\right)^{\alpha-1} x(s_0) + \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} (v(\theta) - x^*) d\theta + \left(1 - \left(\frac{s_0}{t}\right)^{\alpha-1}\right) x^* \end{aligned}$$

Therefore,

$$\|x(t) - x^*\| \leq \left(\frac{s_0}{t}\right)^{\alpha-1} \|x(s_0) - x^*\| + \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta. \quad (76)$$

By Theorem 5, we have, as θ goes to $+\infty$, strong convergence of $v(\theta) = z(\tau(\theta))$ to the minimum norm solution x^* .

Since $\lim_{s \rightarrow +\infty} \|v(\theta) - x^*\| = 0$, then given $a > 0$, there exists $s_a > 0$ sufficiently large so that $\|v(\theta) - x^*\| < a$, for $\theta \geq s_a$.

Then, for $t > s_a$, split the integral $\int_{s_0}^t \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta$ into two parts to obtain

$$\begin{aligned} \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta &= \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^{s_a} \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta + \frac{\alpha-1}{t^{\alpha-1}} \int_{s_a}^t \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta \\ &\leq \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^{s_a} \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta + \frac{a(\alpha-1)}{t^{\alpha-1}} \int_{s_a}^t \theta^{\alpha-2} d\theta \\ &= a + \frac{1}{t^{\alpha-1}} \left((\alpha-1) \int_{s_0}^{s_a} \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta - a s_a^{\alpha-1} \right). \end{aligned}$$

Now let $t \rightarrow +\infty$ to deduce that

$$\limsup_{t \rightarrow +\infty} \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta \leq a.$$

Since this is true for any $a > 0$, this insures

$$\lim_{t \rightarrow +\infty} \frac{\alpha-1}{t^{\alpha-1}} \int_{s_0}^t \theta^{\alpha-2} \|v(\theta) - x^*\| d\theta = 0.$$

Going back to (76), we conclude for $\alpha-1 > 0$ that $\lim_{t \rightarrow +\infty} \|x(t) - x^*\| = 0$. This means that $x(t)$ strongly converges to x^* as $t \rightarrow +\infty$.

iii) *Convergence of the gradients towards zero.* We have obtained in ii) that the trajectory $x(\cdot)$ converges. Therefore, it is bounded. Let $L > 0$ be the Lipschitz constant of ∇f on a ball that contains the trajectory $x(\cdot)$. According to the convergence rate of values and Lemma 5 in the Appendix, we immediately obtain the following convergence rate of the gradients towards zero

$$\frac{1}{2L} \|\nabla f(x(t))\|^2 \leq f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}}\right).$$

This completes the proof. \square

4 Nonsmooth case

In this section, we adapt the approach of [7] to our context. The pair of variables (v, x) defined in section 3 satisfies the following differential system, which only involves first order derivatives in time and space

$$\begin{cases} \dot{v}(t) + \frac{t}{\alpha-1} \nabla f(v(t)) + \frac{t}{\alpha-1} \varepsilon \left(\frac{t^2}{2(\alpha-1)} \right) v(t) = 0 \\ \dot{x}(t) + \frac{\alpha-1}{t} (x(t) - v(t)) = 0. \end{cases} \quad (77)$$

This differential system naturally extends to the non-smooth case, replacing the gradient of f by its subdifferential. Given $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ a convex, lower semicontinuous and proper function, this leads to consider the system of differential equation/inclusion

$$\begin{cases} \dot{v}(t) + \frac{t}{\alpha-1} \partial f(v(t)) + \frac{t}{\alpha-1} \varepsilon \left(\frac{t^2}{2(\alpha-1)} \right) v(t) \ni 0 \\ \dot{x}(t) + \frac{\alpha-1}{t} (x(t) - v(t)) = 0. \end{cases} \quad (78)$$

Solving this system gives generalized solutions to the second-order differential inclusion

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \partial f \left[x(t) + \frac{t}{\alpha-1} \dot{x}(t) \right] + \varepsilon \left(\frac{t^2}{2(\alpha-1)} \right) \left[x(t) + \frac{t}{\alpha-1} \dot{x}(ts) \right] \ni 0, \quad (79)$$

whose direct study raises several difficulties. To extend the results of the previous section to this nonsmooth case, we need to avoid the arguments using the Lipschitz continuity of ∇f . So, we are led to consider the Cauchy problem for (79) with initial data $x_0 \in \text{dom } f$ and $\dot{x}_1 = 0$, that is with initial velocity equal to zero. So doing we have $v(t_0) = x(t_0) = x_0$, which allows to interpret x as an average of v .

Then, the existence and uniqueness of a strong solution to the associated Cauchy problem relies on the equivalent formulation of (78) as a perturbation of the generalized steepest descent dynamical system in the product space $\mathcal{H} \times \mathcal{H}$. Precisely, define $F : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, which is given for any $Z = (v, x) \in \mathcal{H} \times \mathcal{H}$ by

$$F(Z) = f(v),$$

and let $G : [t_0, +\infty[\times \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H}$ be the operator defined by

$$G(t, Z) = \left(\frac{t}{\alpha-1} \varepsilon \left(\frac{t^2}{2(\alpha-1)} \right) v, \frac{\alpha-1}{t} (x - v) \right). \quad (80)$$

Then (78) is written equivalently as

$$\dot{Z}(t) + \frac{t}{\alpha-1} \partial F(Z(t)) + G(t, Z(t)) \ni 0. \quad (81)$$

The initial condition becomes $Z(t_0) = (x_0, x_0)$ which belongs to $\text{dom } F = \text{dom } f \times \mathcal{H}$. According to the classical results concerning the Lipschitz perturbation of evolution equations governed by subdifferentials of convex functions, see [23, Proposition 3.12], we obtain the existence and uniqueness of a global strong solution of the Cauchy problem associated with (78). As a major advantage of the time scaling and averaging techniques, the arguments used in the previous section still work in this more general nonsmooth situation. The rules of differential calculus are still valid for strong solutions, see [24, chapter VIII.2], and Jensen's inequality is still valid for a nonsmooth function f . Indeed Jensen's inequality is classical for a smooth convex function f . Its extension to the nonsmooth convex case can be obtained by first writing it for the Moreau-Yosida regularization f_λ of f , then passing to the limit when $\lambda \downarrow 0$. According to the monotone convergence of f_λ towards f , we can pass to the limit in the integral term thanks to the Beppo-Levy monotone convergence theorem. We so obtain the following theorems corresponding to the cases $\varepsilon(t) = \frac{\delta}{t}$, and $\varepsilon(t) = \frac{1}{t^r}$.

Theorem 8 Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, and proper function such that $S = \operatorname{argmin} f \neq \emptyset$. Take $\alpha > 3$ and $\delta > 1$, which gives $\alpha + 2\delta > 5$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\ddot{x}(t) + \frac{\alpha + 2\delta}{t} \dot{x}(t) + \partial f \left[x(t) + \frac{t}{\alpha - 1} \dot{x}(t) \right] + \frac{2\delta(\alpha - 1)}{t^2} x(t) \ni 0 \quad (82)$$

which is taken in the sense of (78), and which satisfies the initial conditions $x(t_0) \in \operatorname{dom} f$ and $\dot{x}(t_0) = 0$. Then, the following properties are satisfied.

$$(i) \quad f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^2} \right) \quad \text{as } t \rightarrow +\infty; \quad (83)$$

(ii) There is strong convergence of $x(t)$ to the minimum norm solution x^* .

Theorem 9 Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous, and proper function such that $S = \operatorname{argmin} f \neq \emptyset$. Take $0 < r < 1$ and $\alpha > 1$. Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of

$$\ddot{x}(t) + \left(\frac{\alpha}{t} + \frac{2^r(\alpha - 1)^{r-1}}{t^{2r-1}} \right) \dot{x}(t) + \partial f \left(x(t) + \frac{t}{\alpha - 1} \dot{x}(t) \right) + \frac{2^r(\alpha - 1)^r}{t^{2r}} x(t) \ni 0 \quad (84)$$

which is taken in the sense of (78), and which satisfies the initial conditions $x(t_0) \in \operatorname{dom} f$ and $\dot{x}(t_0) = 0$. Then, the following properties are satisfied.

$$(i) \quad f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} \left(\frac{1}{t^{\alpha-1}} + \frac{1}{t^{2r}} \right) \quad \text{as } t \rightarrow +\infty; \quad (85)$$

(ii) There is strong convergence of $x(t)$ to the minimum norm solution x^* .

Remark 2 As a consequence of (i), we have that $x(s)$ remains in the domain of f for all $s \geq t_0$. This viability property strongly depends on the fact that the initial position belongs to the domain of f , and that the initial velocity has been taken equal to zero.

5 Numerical illustrations

The following simple examples illustrate the properties of the trajectories generated by the dynamics studied in the paper. They show that the trajectories verify both a rapid convergence of the values, the convergence towards the minimum norm solution, and a notable attenuation of the oscillations. This highlights the additional properties obtained by the presence of the Tikhonov regularization term. A detailed numerical study of these aspects should be the subject of further work.

We consider the dynamical systems (9), (8) and (4) in \mathbb{R}^2 :

$$\begin{aligned} (\text{AVD})_{\alpha} \quad & \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0, \\ (\text{AVD})_{\alpha, \varepsilon} \quad & \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \left(\nabla f + \varepsilon(t) I \right) x(t) = 0, \\ (\text{AVD-IH})_{\alpha, \varepsilon} \quad & \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \left(\nabla f + \varepsilon \left(\frac{t^2}{2(\alpha - 1)} \right) I \right) \left(x(t) + \frac{t}{\alpha - 1} \dot{x}(t) \right) = 0. \end{aligned}$$

Let us illustrate our results with the following two examples where the function f is taken respectively strictly convex and convex with unbounded solution-set. We choose $\alpha = 3$, $\beta = 1$ and $\varepsilon(t) = t^{-r}$. Our numerical tests were implemented in Scilab version 6.1 as an open source software.

Example 1 Take $f_1 :]-1, +\infty[\rightarrow \mathbb{R}$ which is defined by $f_1(x) = \ln(1 + e^{-(x+y)}) + (x - y)^2 + y$. The function f_1 is strictly convex with $x^* = (\frac{1}{8}, -\frac{1}{8})$ the unique minimum. The trajectories corresponding to the three systems described above are shown in Fig. 1.

Example 2 Consider the non-strictly convex function $f_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f_2(x) = \frac{1}{2}(x_1 + x_2 - 1)^2$. The set of solutions is $\operatorname{argmin} f_2 = \{(x_1, 1 - x_1) : x_1 \in \mathbb{R}\}$, and $x^* = (\frac{1}{2}, \frac{1}{2})$ is the minimum norm solution. The trajectories corresponding to the systems $(\text{AVD})_3, (\text{AVD})_{3,1/t^2}$ and $(\text{AVD-IH})_{3,1/t^2}$ are represented in Fig. 2.

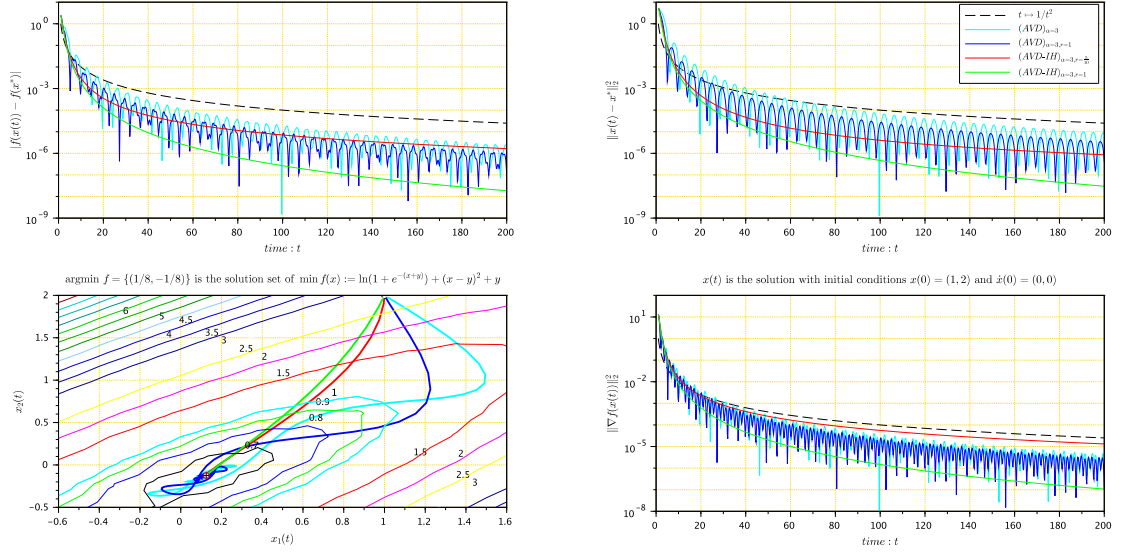


Fig. 1 Convergence rates of values $f_1(x(t)) - f_1(x^*)$, of trajectories $\|x(t) - x^*\|_2$, and gradients $\|\nabla f_1(x(t))\|_2$. Here $\text{argmin} f_1 = \{(\frac{1}{8}, -\frac{1}{8})\}$.

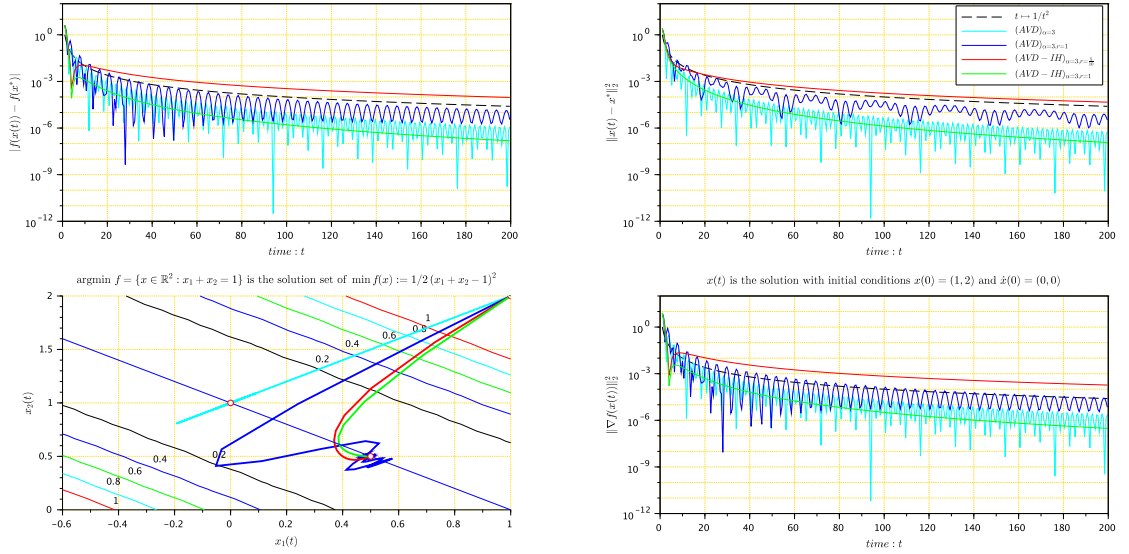


Fig. 2 Convergence rates of values $f_2(x(t)) - f_2(x^*)$, trajectories $\|x(t) - x^*\|_2^2$, and gradients $\|\nabla f_2(x(t))\|_2^2$. Here $\text{argmin} f_1 = \{(x, y) \in \mathbb{R}^2 : x + y = 1\}$.

6 Conclusion, perspective

The introduction of a Tikhonov regularization with vanishing coefficient in the optimization algorithms is beneficial in several respects. For general convex optimization, instead of weak convergence of trajectories/iterations towards an optimal solution which depends on the initial state, it provides strong convergence towards the minimum norm solution. This is especially important for inverse problems, where one seeks a solution as close as possible to a desired state. In this paper, we show that this can be achieved while preserving the fast convergence properties attached to the Nesterov accelerated gradient method. Our approach is based on the Attouch-Bot-Nguyen scaling and time averaging technique, which proves to be flexible, and allows to address these issues in a unified way. As a striking result, for the first time we obtain simultaneously the rapid convergence of the values $1/t^2$ and the strong convergence towards the minimum norm solution. Let us mention some other interesting questions to examine further:

a) Our results open the way to the study of a new class of fast algorithms in convex optimization. Lyapunov's analysis of continuous dynamics which support these algorithms will be a great help. It is probable that such algorithms share the good convergence properties of continuous dynamics, according to the results obtained in [7] concerning the time scaling and averaging method, and [5] concerning the case without the Tikhonov regularization.

b) An interesting and open question is whether a similar analysis can be developed using closed-loop Tikhonov regularization, i.e. the Tikhonov regularization coefficient is taken as a function of the current state of the system, see [6] for a survey on these questions.

c) The Tikhonov regularization and the property of convergence to the minimum norm solution is a particular case of the general hierarchical principle which is attached to the viscosity method, see [3]. In the context of numerical optimization, as an example, it is therefore natural to extend our study to the logarithmic barrier method for linear programming and to the selection of the analytical center.

7 Appendix

The following Lemma provides an extended version of the classical gradient lemma which is valid for differentiable convex functions. The following version has been obtained in [9, Lemma 1], [10]. We reproduce its proof for the convenience of the reader.

Lemma 5 *Let $f : \mathcal{H} \rightarrow \mathbb{R}$ be a convex function whose gradient is L -Lipschitz continuous. Let $s \in]0, 1/L]$. Then for all $(x, y) \in \mathcal{H}^2$, we have*

$$f(y - s\nabla f(y)) \leq f(x) + \langle \nabla f(y), y - x \rangle - \frac{s}{2} \|\nabla f(y)\|^2 - \frac{s}{2} \|\nabla f(x) - \nabla f(y)\|^2. \quad (86)$$

In particular, when $\operatorname{argmin} f \neq \emptyset$, we obtain that for any $x \in \mathcal{H}$

$$f(x) \geq \min_{\mathcal{H}} f + \frac{1}{2L} \|\nabla f(x)\|^2. \quad (87)$$

Proof Denote $y^+ = y - s\nabla f(y)$. By the standard descent lemma applied to y^+ and y , and since $sL \leq 1$ we have

$$f(y^+) \leq f(y) - \frac{s}{2} (2 - Ls) \|\nabla f(y)\|^2 \leq f(y) - \frac{s}{2} \|\nabla f(y)\|^2. \quad (88)$$

We now argue by duality between strong convexity and Lipschitz continuity of the gradient of a convex function. Indeed, using Fenchel identity, we have

$$f(y) = \langle \nabla f(y), y \rangle - f^*(\nabla f(y)).$$

L -Lipschitz continuity of the gradient of f is equivalent to $1/L$ -strong convexity of its conjugate f^* . This together with the fact that $(\nabla f)^{-1} = \partial f^*$ gives for all $(x, y) \in \mathcal{H}^2$,

$$f^*(\nabla f(y)) \geq f^*(\nabla f(x)) + \langle x, \nabla f(y) - \nabla f(x) \rangle + \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2.$$

Inserting this inequality into the Fenchel identity above yields

$$\begin{aligned} f(y) &\leq -f^*(\nabla f(x)) + \langle \nabla f(y), y \rangle - \langle x, \nabla f(y) - \nabla f(x) \rangle - \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \\ &= -f^*(\nabla f(x)) + \langle x, \nabla f(x) \rangle + \langle \nabla f(y), y - x \rangle - \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \\ &= f(x) + \langle \nabla f(y), y - x \rangle - \frac{1}{2L} \|\nabla f(x) - \nabla f(y)\|^2 \\ &\leq f(x) + \langle \nabla f(y), y - x \rangle - \frac{s}{2} \|\nabla f(x) - \nabla f(y)\|^2. \end{aligned}$$

Inserting the last bound into (88) completes the proof.

References

1. F. ALVAREZ, A. CABOT, Asymptotic selection of viscosity equilibria of semilinear evolution equations by the introduction of a slowly vanishing term, *Discrete Contin. Dyn. Syst.* 15 (2006), 921–938.
2. V. APIDOPOULOS, J.-F. AUJOL, CH. DOSSAL, The differential inclusion modeling the FISTA algorithm and optimality of convergence rate in the case $b \leq 3$, *SIAM J. Optim.*, 28 (1) (2018), 551–574.
3. H. ATTOUCH, Viscosity solutions of minimization problems, *SIAM J. Optim.* 6 (3) (1996), 769–806.
4. H. ATTOUCH, A. BALHAG, Z. CHBANI, H. RIAHI, Damped inertial dynamics with vanishing Tikhonov regularization: Strong asymptotic convergence towards the minimum norm solution, *J. Differential Equations*, 311 (2022), 29–58.
5. H. ATTOUCH, A. BALHAG, Z. CHBANI, H. RIAHI, Accelerated gradient methods combining Tikhonov regularization with geometric damping driven by the Hessian, *arXiv:2203.05457v1 [math.OC]* 10 Mar 2022.
6. H. ATTOUCH, R.I. BOŦ, E.R. CSETNEK, Fast optimization via inertial dynamics with closed-loop damping, *Journal of the European Mathematical Society (JEMS)*, 2021, hal-02910307.
7. H. ATTOUCH, R.I. BOŦ, D.-K. NGUYEN, Fast convex optimization via time scale and averaging of the steepest descent, *arXiv:2208.08260v1 [math.OC]* Aug 2022.
8. H. ATTOUCH, A. CABOT, Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity, *J. Differential Equations*, 263 (9), (2017), 5412–5458.
9. H. ATTOUCH, Z. CHBANI, J. FADILI, H. RIAHI, First order optimization algorithms via inertial systems with Hessian driven damping, *Math. Program.*, 193 (2022), 113–155.
10. H. ATTOUCH, Z. CHBANI, J. FADILI, H. RIAHI, Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping, *Optimization*, <https://doi.org/10.1080/02331934.2021.2009828>. (2021).
11. H. ATTOUCH, Z. CHBANI, J. PEYPOUQUET, P. REDONT, Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity, *Math. Program.*, 168 (1-2) (2018), 123–175.
12. H. ATTOUCH, Z. CHBANI, H. RIAHI, Combining fast inertial dynamics for convex optimization with Tikhonov regularization, *J. Math. Anal. Appl.* 457 (2018), 1065–1094.
13. H. ATTOUCH, R. COMINETTI, A dynamical approach to convex minimization coupling approximation with the steepest descent method, *J. Differential Equations*, 128 (2) (1996), 519–540.
14. H. ATTOUCH, M.-O. CZARNECKI, Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria, *J. Differential Equations* 179 (2002), 278–310.
15. H. ATTOUCH, M.-O. CZARNECKI, Asymptotic behavior of coupled dynamical systems with multiscale aspects, *J. Differential Equations* 248 (2010), 1315–1344.
16. H. ATTOUCH, M.-O. CZARNECKI, Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects, *J. Differential Equations*, 262 (3) (2017), 2745–2770.
17. H. ATTOUCH, J. FADILI, From the Ravine method to the Nesterov method and vice versa: a dynamic system perspective, *arXiv:2201.11643v1 [math.OC]* 27 Jan 2022.
18. H. ATTOUCH, S. LÁSZLÓ, Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution, *arXiv:2104.11987v1 [math.OC]* 24 Apr 2021.
19. H. ATTOUCH, J. PEYPOUQUET, The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than $1/k^2$, *SIAM J. Optim.*, 26 (3) (2016), 1824–1834.

20. J.-B. BAILLON, R. COMINETTI, A convergence result for non-autonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming, *J. Funct. Anal.* 187 (2001) 263–273.
21. H. BAUSCHKE, P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
22. R. I. BOŢ, E. R. CSETNEK, S.C. LÁSZLÓ, Tikhonov regularization of a second order dynamical system with Hessian damping, *Math. Program.*, 189 (2021), 151–186.
23. H. BRÉZIS, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).
24. H. BRÉZIS, *Analyse fonctionnelle*, Masson, 1983.
25. A. CABOT, Inertial gradient-like dynamical system controlled by a stabilizing term, *J. Optim. Theory Appl.*, 120 (2004), 275–303.
26. A. CABOT, Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchical minimization, *SIAM J. Optim.*, 15 (2) (2005), 555–572.
27. A. CABOT, H. ENGLER, S. GADAT, On the long time behavior of second order differential equations with asymptotically small dissipation, *Trans. Amer. Math. Soc.*, 361 (2009), 5983–6017.
28. A. CHAMBOLLE, CH. DOSSAL, On the convergence of the iterates of Fista, *J. Opt. Theory Appl.*, 166 (2015), 968–982.
29. R. COMINETTI, J. PEYPOUQUET, S. SORIN, Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization, *J. Differential Equations*, 245 (2008), 3753–3763.
30. S.A. HIRSTOAGA, *Approximation et résolution de problèmes d'équilibre, de point fixe et d'inclusion monotone*. PhD thesis, Université Pierre et Marie Curie - Paris VI, 2006, HAL Id: tel-00137228.
31. M.A. JENDOUBI, R. MAY, On an asymptotically autonomous system with Tikhonov type regularizing term, *Archiv der Mathematik*, 95 (4) (2010), 389–399.
32. R. MAY, C. MNASRI, M. ELLOUMI, Asymptotic for a second order evolution equation with damping and regularizing terms, *AIMS Mathematics*, 6 (5) (2021), 4901–4914.
33. Y. NESTEROV, A method of solving a convex programming problem with convergence rate $O(1/k^2)$, *Soviet Math. Dokl.*, 27 (1983), 372–376.
34. Y. NESTEROV, *Introductory lectures on convex optimization: A basic course*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004.
35. B. POLYAK, *Introduction to Optimization*, New York, NY: Optimization Software-Inc, 1987.
36. W. SIEGEL, Accelerated first-order methods: Differential equations and Lyapunov functions, [arXiv:1903.05671v1 \[math.OC\]](https://arxiv.org/abs/1903.05671v1), 2019.
37. B. SHI, S.S. DU, M. I. JORDAN, W. J. SU, Understanding the acceleration phenomenon via high-resolution differential equations, *Math. Program.* (2021). <https://doi.org/10.1007/s10107-021-01681-8>.
38. W. SU, S. BOYD, E. J. CANDÈS, A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights. *J. Mach. Learn. Res.*, 17(153), (2016), 1–43.
39. A. N. TIKHONOV, *Doklady Akademii Nauk SSSR* 151 (1963) 501–504, (Translated in "Solution of incorrectly formulated problems and the regularization method". *Soviet Mathematics* 4 (1963) 1035–1038).
40. A. N. TIKHONOV, V. Y. ARSENIN, *Solutions of Ill-Posed Problems*, Winston, New York, 1977.
41. D. TORRALBA, *Convergence epigraphique et changements d'échelle en analyse variationnelle et optimisation*, PhD thesis, Université Montpellier, 1996.
42. A. C. WILSON, B. RECHT, M. I. JORDAN, A Lyapunov analysis of momentum methods in optimization, *Journal of Machine Learning Research*, 22 (2021), 1–34.