Accelerated Gradient Methods Combining Tikhonov Regularization with Geometric Damping Driven by the Hessian

Hedy Attouch · Aïcha Balhag · Zaki Chbani · Hassan Riahi

Accepted: 30 March 2023 / Published online: 31 May 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
In a Hilbert framework, for general convex differentiable optimization, we consider accelerated gradient dynamics combining Tikhonov regularization with Hessian-driven damping. The temporal discretization of these dynamics leads to a new class of first-order optimization algorithms with favorable properties. The Tikhonov regularization parameter is assumed to tend to zero as time tends to infinity, which preserves equilibria. The presence of the Tikhonov regularization term induces a strong property of convexity which vanishes asymptotically. To take advantage of the fast convergence rates attached to the heavy ball method in the strongly convex case, we consider inertial dynamics where the viscous damping coefficient is proportional to the square root of the Tikhonov regularization parameter, and hence converges to zero. The geometric damping, controlled by the Hessian of the function to be minimized, induces attenuation of the oscillations. Under an appropriate setting of the parameters, based on Lyapunov’s analysis, we show that the trajectories provide at the same time several remarkable properties: fast convergence of values, fast convergence of gradients towards zero, and strong convergence to the minimum norm minimizer. We show that the corresponding proximal algorithms share the same properties as continuous...
dynamics. The numerical illustrations confirm the results obtained. This study extends a previous paper by the authors regarding similar problems without the presence of Hessian driven damping.

**Keywords** Accelerated gradient methods · Convex optimization · Damped inertial dynamics · Hessian-driven damping · Hierarchical minimization · Nesterov accelerated gradient method · Tikhonov approximation

**Mathematics Subject Classification** 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 \( f : \mathcal{H} \to \mathbb{R} \) a general convex function, which is continuously differentiable, we will develop fast gradient methods for solving the minimization problem

\[
\min \{ f(x) : x \in \mathcal{H} \}.
\]  

Our approach is based on the convergence properties as \( t \to +\infty \) of the trajectories generated by the damped inertial dynamic

\[
(\text{TRISH}) \quad \ddot{x}(t) + \delta \sqrt{\varepsilon(t)} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + \varepsilon(t) x(t) = 0,
\]

and on the link between dynamical systems and the algorithms that result from their temporal discretization. We use (TRISH) as shorthand for Tikhonov regularized inertial system with Hessian-driven damping. As a basic ingredient, this system involves a nonnegative function \( \varepsilon(\cdot) \) which enters both in the viscous damping and the Tikhonov regularization terms. We assume that \( \lim_{t \to +\infty} \varepsilon(t) = 0 \), which preserves the equilibria. According to the structure of (TRISH) this makes the damping coefficient asymptotically vanish, in coordination with the Tikhonov regularization coefficient. The other basic ingredient is the Hessian driven damping term which induces several favorable properties, notably a significant reduction of the oscillations. As a general remark, let us note the originality of the system (TRISH) which is similar but with distinct characteristics of the heavy ball with friction method of Polyak in the strongly convex case, and of the dynamics of Su-Boyd-Candès associated with the accelerated gradient method of Nesterov.

We will show that a judicious setting of \( \varepsilon(\cdot) \) and of the positive parameter \( \delta \) ensures that the trajectories generated by (TRISH) verify the following three properties at the same time:

- rapid convergence of values (one can approach arbitrarily close to the optimal convergence rate),
- rapid convergence of the gradients towards zero.
• strong convergence towards the minimum norm element of \( S = \text{argmin } f \).

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

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

We take for granted the existence and uniqueness results for the Cauchy problem associated with our dynamics. Indeed, the introduction of the Tikhonov regularization term, which is smooth, induces minor modifications of the proof already developed in the case without the Tikhonov term, as developed in [9]. Then, we will show that the corresponding proximal algorithms, obtained by implicit temporal discretization, benefit from similar convergence properties, and will provide an introduction to the CORRESPONDING gradient algorithms with numerical illustrations.

1.1 The Role of the Tikhonov Regularization

Initially designed for the regularization of ill-posed inverse problems [47, 48], 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 [3, 5, 13, 15, 24, 30, 33, 35]. 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 [43], where the damping coefficient \( \gamma > 0 \) is fixed. In [14] Attouch and Czarnecki considered the system

\[
\ddot{x}(t) + \gamma \dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0. \tag{2}
\]

In the slow parametrization case \( \int_0^{+\infty} \varepsilon(t)dt = +\infty \), they proved that any solution \( x(\cdot) \) of (2) converges strongly to the minimum norm element of \( \text{argmin } f \), see also [16, 29, 31, 36]. 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

\[
(AVD)_{\alpha,\varepsilon} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) + \varepsilon(t)x(t) = 0, \tag{3}
\]

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

\[
(AVD)_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla f(x(t)) = 0, \tag{4}
\]
which was introduced by Su, Boyd and Candès in [46]. (AVD)$_\alpha$ is a low resolution ODE of the accelerated gradient method of Nesterov [39, 40] and of the Ravine method [17, 46]. (AVD)$_\alpha$ has been the subject of many recent studies which have given an in-depth understanding of the Nesterov accelerated gradient method, see [4, 6, 8, 11, 21, 32, 38, 44, 46, 50].

As an original aspect of our approach, we rely on the properties of the heavy ball with friction method of Polyak in the strongly convex case, which provides exponential convergence rates. To take advantage of this remarkable property, and adapt it to our situation, we consider the nonautonomous dynamic version of the heavy ball method which at time $t$ is governed by the gradient of the regularized function $x \mapsto f(x) + \frac{\epsilon(t)}{2}\|x\|^2$, where the Tikhonov regularization parameter satisfies $\epsilon(t) \to 0$ as $t \to +\infty$. This idea was first developed in [6, 18]. Let us make this precise.

Recall that a function $f : \mathcal{H} \to \mathbb{R}$ is $\mu$-strongly convex for some $\mu > 0$ if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex. In this setting, we have the following exponential convergence result for the heavy ball with friction dynamic where the viscous damping coefficient is twice the square root of the modulus of strong convexity of $f$, see [42]:

**Theorem 1** Suppose that $f : \mathcal{H} \to \mathbb{R}$ is a function of class $C^1$ which is $\mu$-strongly convex for some $\mu > 0$. Let $x(\cdot) : [t_0, +\infty[ \to \mathcal{H}$ be a solution trajectory of

$$\ddot{x}(t) + 2\sqrt{\mu}\dot{x}(t) + \nabla f(x(t)) = 0. \tag{5}$$

Then, the following property holds: $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(e^{-\sqrt{\mu}t}\right)$ as $t \to +\infty$. To adapt this result to the case of a general convex differentiable function $f : \mathcal{H} \to \mathbb{R}$, a natural idea is to use Tikhonov’s method of regularization. This leads to consider the non-autonomous dynamic which at time $t$ is governed by the gradient of the strongly convex function

$$\varphi_t : \mathcal{H} \to \mathbb{R}, \quad \varphi_t(x) := f(x) + \frac{\epsilon(t)}{2}\|x\|^2.$$ 

The viscosity curve $\epsilon \mapsto x_\epsilon := \arg\min_{\mathcal{H}} \left\{ f(\cdot) + \frac{\epsilon}{2}\|\cdot\|^2 \right\}$ will play a key role in our analysis. By definition of $\varphi_t$, we have $x_{\epsilon(t)} = \arg\min_{\mathcal{H}} \varphi_t$. The first-order optimality condition gives

$$\nabla f(x_{\epsilon(t)}) + \epsilon(t)x_{\epsilon(t)} = 0. \tag{6}$$

We call $t \mapsto x_{\epsilon(t)}$ the parametrized viscosity curve. Then, replacing $f$ by $\varphi_t$ in (5), and noticing that $\varphi_t$ is $\epsilon(t)$-strongly convex, this gives the following dynamic which was introduced in [18] and [6] ($\delta$ is a positive parameter)

$$\dot{x}(t) + \delta\sqrt{\epsilon(t)}\dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t) = 0. \tag{TRIGS}$$

(TRIGS) stands shortly for Tikhonov regularization of inertial gradient systems. In order not to asymptotically modify the equilibria, it is supposed that $\epsilon(t) \to 0$ as $t \to +\infty$.

1 This condition implies that (TRIGS) falls within the framework of the asymptotic version ($t \to +\infty$) of the Browder-Tikhonov regularization method.
inertial gradient systems with asymptotically vanishing damping. It has been shown in [6, 18] that a judicious tuning of $\epsilon(t)$ in (TRIGS) ensures both rapid convergence of values, and strong convergence of the trajectories towards the minimum norm element of $S = \arg\min_{\mathcal{H}} f$ (which is reminiscent of the Tikhonov method).

1.2 The Role of the Hessian-Driven Damping

As is the case with inertial dynamics which are only damped by viscous damping, the system (TRIGS) may exhibit oscillations which are undesirable from an optimization point of view. To remedy this situation, we introduce into the dynamic a geometric damping which is driven by the Hessian of the function $f$ to be minimized. So doing, we obtain the system (TRISH). The presence of the Hessian does not entail numerical difficulties, since the Hessian intervenes in the above ODE in the form $\nabla^2 f(x(t))\dot{x}(t)$, which is nothing but the derivative wrt time of $\nabla f(x(t))$. This explains why the time discretization of this dynamic provides first-order algorithms. The importance of the Hessian driven damping has been demonstrated in several areas. We list some of them below.

- In the field of PDEs for mechanics and physics, it is called strong damping, or geometric damping because it takes into account the geometry of the function to be minimized. In the PDE’s framework, when $f$ is quadratic, and $\nabla f = A$ is a linear elliptic operator, the strong damping involves the action of a fractional power $A^\theta$ of $A$ on the velocity vector. When $\theta \geq \frac{1}{2}$, this induces notably reduced oscillations. The Hessian-driven damping corresponds to $\theta = 1$. It can be combined with various other types of damping, such as the dry friction [2]. It also makes it possible to model shocks which are completely damped in unilateral mechanics [19].

- It has been shown in [17] and [45] that the high resolution ODE of the Ravine and Nesterov methods exhibits the Hessian driven damping. This explains the rapid convergence of the gradients towards zero which is verified by these dynamics and algorithms [9, 10, 22, 26, 45]. Our approach is in accordance with Nesterov [41], where it is conjectured that the introduction of an adapted Tikhonov regularization term helps to make the gradients small.

- The Hessian driven damping comes into the study of Newton’s method in optimization. Given a general maximally monotone operator $A : \mathcal{H} \to 2^\mathcal{H}$, to overcome the ill-posedness of Newton’s continuous method for solving $0 \in A(x)$, the following first-order evolution system was considered by Attouch and Svaiter [23] and studied further in [1, 20]. Formally, this system is written as

$$\gamma(t)\dot{x}(t) + \beta \frac{d}{dt} (A(x(t))) + A(x(t)) = 0.$$ 

It can be considered as a continuous version of the Levenberg-Marquardt method, which acts as a regularization of the Newton method. Under a fairly general assumption on the regularization parameter $\gamma(\cdot)$, this system is well posed and generates trajectories that converge weakly to equilibria. Thus, (TRISH) and its nonsmooth extension can be considered as an inertial and regularized version of this system when $A$ is the subdifferential of a convex lower semicontinuous proper function.
1.3 A Model Result

In Sect. 3, we will prove the following result in the case $\varepsilon(t) = \frac{1}{t^r}$. It is expressed with the help of the parametrized viscosity curve which converges strongly to the minimum norm solution.

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

$$\ddot{x}(t) + \frac{\delta}{t^2} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r} x(t) = 0.$$  \hspace{1cm} (7)

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

This is the first time that these three properties have been obtained within the same dynamic. Let us note that by taking $r$ close to 2, one obtains convergence rates comparable to the most recent results concerning the introduction of the Hessian driven damping in the dynamic associated with the accelerated gradient of Nesterov. Precisely, letting $r \rightarrow 2$ in the formulas above gives $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O} (1/t^2)$, and $\int_{t_0}^{+\infty} \| \nabla f(x(t))\|^2 dt < +\infty$. Thus, by taking $r$ sufficiently close to 2, we can obtain convergence rates arbitrarily close to these rates. The case $r = 2$, which corresponds to the Nesterov accelerated gradient method, is critical: in this case, the strong convergence towards the minimum norm solution is an open question. This choice of $r < 2$, and close to 2, produces a viscosity coefficient close to but slightly higher than that corresponding to the accelerated gradient method of Nesterov. Our results therefore present a certain analogy with the result of Chambolle and Dossal [32] who obtained the convergence of iterates in the accelerated gradient method of Nesterov by slightly increasing the coefficient of viscosity. Our results show also the balance between fast convergence of values and strong convergence to the minimum norm solution.

1.4 Contents

The article is organized as follows. In Sect. 2, for a general Tikhonov regularization parameter $\varepsilon(\cdot)$, we study the asymptotic convergence properties of the solution trajectories of (TRISH). In Sect. 3, we apply these results to the special case $\varepsilon(t) = \frac{1}{t^r}$, $0 < r < 2$, and obtain fast convergence results. Section 4 is devoted to the study of the corresponding first-order algorithms, obtained by temporal discretization. Section 5
contains numerical illustrations. In Sect. 6 we conclude with perspective and open questions.

2 Convergence Results via Lyapunov Analysis

Given a general regularization parameter \( \epsilon(\cdot) \), we successively present the idea guiding the Lyapunov analysis, then some preparatory lemmas, and finally the detailed proof. In the next section, we will particularize our results to the case \( \epsilon(t) = \frac{1}{r}, 0 < r < 2 \), and obtain fast convergence results.

2.1 General Idea of the Proof

As we already mentioned, the function

\[
\varphi_t : H \rightarrow \mathbb{R}, \quad \varphi_t(x) := f(x) + \frac{\epsilon(t)}{2} \|x\|^2 \tag{8}
\]

plays a central role in the Lyapunov analysis, via its strong convexity property. Thus, it is convenient to reformulate (TRISH) with the help of the function \( \varphi_t \), which gives

\[
\begin{aligned}
\ddot{x}(t) + \delta \sqrt{\epsilon(t)} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla \varphi_t(x(t)) &= 0, \\
\end{aligned} \tag{9}
\]

where \( \delta, \beta \) are positive parameters. We recall that \( \epsilon : [t_0, +\infty[ \rightarrow \mathbb{R}^+ \) is a nonincreasing function of class \( C^1 \), such that \( \lim_{t \rightarrow +\infty} \epsilon(t) = 0 \). In the mathematical analysis of inertial gradient dynamics and algorithms with Hessian-driven damping, the basic equality

\[
\frac{d}{dt} \nabla f(x(t)) = \nabla^2 f(x(t)) \dot{x}(t) \tag{10}
\]

makes these systems relevant to first-order methods, a crucial property for numerical purposes. In the presence of the Tikhonov term, to keep the structural property attached to (10), let us introduce the following variant of (TRISH) where the above relation comes with \( \varphi_t \) instead of \( f \):

\[
\begin{aligned}
\ddot{x}(t) + \delta \sqrt{\epsilon(t)} \dot{x}(t) + \beta \frac{d}{dt} (\nabla \varphi_t(x(t))) + \nabla \varphi_t(x(t)) &= 0. \\
\end{aligned} \tag{11}
\]

Adding the suffix E after TRISH recalls that the dynamic has been adapted to take advantage of the Equality in (10), with \( \varphi_t \) instead of \( f \). To encompass these two dynamic systems, we consider

\[
\ddot{x}(t) + \delta \sqrt{\epsilon(t)} \dot{x}(t) + \beta \frac{d}{dt} [\nabla \varphi_t(x(t)) + (p - 1)\epsilon(t)x(t)] + \nabla \varphi_t(x(t)) = 0, \tag{12}
\]

where the parameter \( p \in [0, 1] \). When \( p = 0 \) we get (TRISH), and for \( p = 1 \) we get (TRISHE).
Given $p \in [0, 1]$, let us introduce the real-valued function $t \in [t_0, +\infty[ \mapsto E_p(t) \in \mathbb{R}^+$ that plays a key role in our Lyapunov analysis. It is defined by

$$E_p(t) := \left( \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \right) + \frac{1}{2} \|v_p(t)\|^2$$

(13)

where $\varphi_t$ has been defined in (8), $x_{\varepsilon(t)}$ in (6), and

$$v_p(t) := \lambda \sqrt{\varepsilon(t)} \left( x(t) - x_{\varepsilon(t)} \right) + \dot{x}(t) + \beta \left[ \nabla \varphi_t(x(t)) + (p - 1) \varepsilon(t) x(t) \right],$$

(14)

with $0 \leq \lambda < \delta$. We will show that under a judicious setting of parameters, $E_p(\cdot)$ satisfies the first-order differential inequality

$$\dot{E}_p(t) + \mu(t) E_p(t) + \frac{\beta}{2\delta} (\delta - \lambda) \| \nabla \varphi_t(x(t)) \|^2 \leq \frac{\|x^*\|^2}{2} G(t),$$

(15)

where

$$\mu(t) := -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + (\delta - \lambda) \sqrt{\varepsilon(t)},$$

(16)

and

$$G(t) := (2a + b) \lambda \frac{\dot{\varepsilon}^2(t)}{\varepsilon(t)^3} - \dot{\varepsilon}(t) + (1 - p) \beta \lambda (\delta - \lambda) \varepsilon^2(t),$$

where $a > 1$ and $b > 0$ are parameters involved in the tuning of the function $t \mapsto \varepsilon(t)$. Since $\mu(t) > 0$, this will allow us to estimate the rate of convergence of $E_p(t)$ towards zero. In turn, this provides convergence rates of values and trajectories, as the following lemma shows.

**Lemma 1** Let $x(\cdot) : [t_0, +\infty[ \mapsto \mathcal{H}$ be a solution trajectory of the damped inertial dynamic (12), and $t \in [t_0, +\infty[ \mapsto E_p(t) \in \mathbb{R}^+$ be the energy function defined in (13). Then, the following estimates are satisfied: for any $t \geq t_0$,

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

(17)

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

(18)

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

**Proof** (i) According to the definition of $\varphi_t$, we have

$$f(x(t)) - \min_{\mathcal{H}} f = \varphi_t(x(t)) - \varphi_t(x^*) + \frac{\varepsilon(t)}{2} \left( \|x^*\|^2 - \|x(t)\|^2 \right)$$

$$= \left[ \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \right] + \left[ \varphi_t(x_{\varepsilon(t)}) - \varphi_t(x^*) \right] + \frac{\varepsilon(t)}{2} \left( \|x^*\|^2 - \|x(t)\|^2 \right)$$

$$\leq \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) + \frac{\varepsilon(t)}{2} \|x^*\|^2.$$
By definition of $E_p(t)$ we have

$$\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \leq E_p(t)$$  \hspace{1cm} (19)

which, combined with the above inequality, gives (17).

(ii) By the strong convexity of $\varphi_t$, and $x_{\varepsilon(t)} := \arg\min_{x \in \mathcal{H}} \varphi_t$, we have

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

By combining the inequality above with (19), we get

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

which gives (18).

\(\square\)

### 2.2 Preparatory Results for Lyapunov Analysis

The parametrized viscosity curve $t \mapsto x_{\varepsilon(t)}$ plays a central role in the definition of $E_p(\cdot)$, and therefore in the Lyapunov analysis. We review below some of its topological and differential properties.

#### 2.2.1 Topological Properties

The following properties are immediate consequences of the classical properties of the Tikhonov regularization (see [5] for a general overview of viscosity methods), and of $\lim_{t \to +\infty} \varepsilon(t) = 0$:

- $\forall t \geq t_0, \|x_{\varepsilon(t)}\| \leq \|x^*\|$ \hspace{1cm} (20)
- $\lim_{t \to +\infty} \|x_{\varepsilon(t)} - x^*\| = 0$ where $x^* = \text{proj}_{\arg\min} f$.

#### 2.2.2 Differential Properties

To evaluate the terms $\frac{d}{dt} (\varphi_t(x_{\varepsilon(t)}))$ and $\frac{d}{dt} (x_{\varepsilon(t)})$ which occur in $\frac{d}{dt} E_p(t)$, we use the differentiability properties of the viscosity curve $\varepsilon \mapsto x_\varepsilon = \arg\min_{\xi} \{ f(\xi) + \frac{\varepsilon}{2} \|\xi\|^2 \}$. According to [13, 35, 49], the viscosity curve is Lipschitz continuous on the compact intervals of $]0, +\infty[$. So it is absolutely continuous, and almost everywhere differentiable.

As a technical tool, we will use the Moreau envelope $f_\theta : \mathcal{H} \to \mathbb{R}$ defined for any $\theta > 0$ by

$$f_\theta(x) = \min_{\xi \in \mathcal{H}} \left\{ f(\xi) + \frac{1}{2\theta} \|x - \xi\|^2 \right\}. \hspace{1cm} (22)$$
The infimum in the above expression is achieved at the unique proximal point \( \text{prox}_{\theta f}(x) \), i.e.

\[
f_{\theta}(x) = f(\text{prox}_{\theta f}(x)) + \frac{1}{2\theta} \|x - \text{prox}_{\theta f}(x)\|^2. \tag{23}
\]

One can consult [25, section 12.4] for more details on the Moreau envelope. Since \( \frac{d}{d\theta} f_{\theta}(x) = \frac{1}{2} \|\nabla f_{\theta}(x)\|^2 \), (see [6, Lemma 3]), we have, for a differentiable real function \( \theta : (t_0, +\infty) \to (0, +\infty) \)

\[
\frac{d}{dt} f_{\theta(t)}(x) = -\frac{\dot{\theta}(t)}{2} \|\nabla f_{\theta(t)}(x)\|^2. \tag{24}
\]

Based on these properties we have the following lemma, which was established in [6], and which we reproduce here for ease of reading.

**Lemma 2** 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 \( t \geq t_0 \)

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

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, as stated above. We have

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

According to (24)

\[
\frac{d}{dt} \varphi_t(x_{\varepsilon(t)}) = \frac{d}{dt} \left( f_{\frac{1}{\varepsilon(t)}}(0) \right) = \frac{1}{2} \frac{\dot{\varepsilon}(t)}{\varepsilon^2(t)} \|\nabla f_{\frac{1}{\varepsilon(t)}}(0)\|^2. \tag{25}
\]

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 $\nabla f_{\frac{1}{\epsilon_0}}(0) = \epsilon(t) \left(0 - \operatorname{prox}_{\frac{1}{\epsilon_0}} f(0)\right)$, we get $\nabla f_{\frac{1}{\epsilon_0}}(0) = -\epsilon(t)x_{\epsilon(t)}$. This combined with (25) gives
\[
\frac{d}{dt} \varphi_t(x_{\epsilon(t)}) = \frac{1}{2} \dot{\epsilon}(t) \|x_{\epsilon(t)}\|^2. 
\]

(ii) We have
\[
-\epsilon(t)x_{\epsilon(t)} = \nabla f(x_{\epsilon(t)}) \quad \text{and} \quad -\epsilon(t + h)x_{\epsilon(t+h)} = \nabla f(x_{\epsilon(t+h)}).
\]

According to the monotonicity of $\nabla f$, we have
\[
\left(\epsilon(t)x_{\epsilon(t)} - \epsilon(t + h)x_{\epsilon(t+h)}, x_{\epsilon(t+h)} - x_{\epsilon(t)}\right) \geq 0,
\]
which implies
\[
-\epsilon(t)\|x_{\epsilon(t+h)} - x_{\epsilon(t)}\|^2 + (\epsilon(t) - \epsilon(t + h)) \left\langle x_{\epsilon(t+h)}, x_{\epsilon(t+h)} - x_{\epsilon(t)} \right\rangle \geq 0.
\]

After division by $h^2$, we obtain
\[
\left(\frac{\epsilon(t) - \epsilon(t + h)}{h}\right) \left\langle x_{\epsilon(t+h)}, \frac{x_{\epsilon(t+h)} - x_{\epsilon(t)}}{h} \right\rangle \geq \epsilon(t) \left\| \frac{x_{\epsilon(t+h)} - x_{\epsilon(t)}}{h} \right\|^2.
\]

We now rely on the differentiability properties of the viscosity curve $\epsilon \mapsto x_\epsilon$, which have been recalled above. It is Lipschitz continuous on the compact intervals of $]0, +\infty[$, so almost everywhere differentiable. Therefore, the mapping $t \mapsto x_{\epsilon(t)}$ satisfies the same differentiability properties. By letting $h \to 0$, we obtain that, for almost every $t \geq t_0$
\[
-\dot{\epsilon}(t) \left\langle x_{\epsilon(t)}, \frac{d}{dt} x_{\epsilon(t)} \right\rangle \geq \epsilon(t) \left\| \frac{d}{dt} x_{\epsilon(t)} \right\|^2,
\]
which gives the claim. The last statement follows from Cauchy-Schwarz inequality. \(\square\)

### 2.3 Lyapunov Analysis of (TRISH) and (TRISHE): Main Theorem and Its Proof

Let $p \in [0, 1]$, $\delta > 0$, $\beta > 0$ be given parameters. In order to develop our Lyapunov analysis of
\[
\ddot{x}(t) + \delta \sqrt{\epsilon(t)} \dot{x}(t) + \beta \left[ \nabla \varphi_t(x(t)) + (p - 1)\epsilon(t)x(t) \right] + \nabla \varphi_t(x(t)) = 0, \quad (26)
\]
we assume that the Tikhonov regularization parameter $\epsilon(\cdot)$ satisfies the following growth condition.
Definition 1 The Tikhonov regularization parameter \( t \mapsto \varepsilon(t) \) satisfies the condition \( (\mathcal{H}_p) \) if there exists \( a > 1, b > 0, \lambda \in \left[ \frac{1}{2} \delta, \frac{a}{a + 1} \delta \right] \), and \( t_1 \geq t_0 \) such that for all \( t \geq t_1 \), the following inequalities are satisfied

\[
(\mathcal{H}_p) \begin{cases}
(i) \quad \frac{d}{dt} \left( \frac{1}{\varepsilon(t)} \right) \leq \min \left( 2\lambda - \delta, \frac{1}{2} \left( \frac{a + 1}{a} \lambda \right) \right) \quad \text{and} \quad \delta \beta \leq \frac{1}{\sqrt{\varepsilon(t)}} \; ; \\
(ii) \quad \frac{d}{dt} (\ln(\varepsilon(t))) \leq \frac{4}{\beta} \left( \left( 1 - \frac{1}{2b} \right) \lambda^2 - \delta \lambda + \frac{1}{2} \right) \; ; \\
(iii) \quad \frac{d}{dt} \left( \frac{1}{\varepsilon(t)} \right) \geq \beta (1 - p) \left( \lambda^2 - \delta \lambda + (1 - p) \right) .
\end{cases}
\]

Remark 1 In Sect. 3, we will show that the condition \( (\mathcal{H}_p) \) holds for specific examples of \( \varepsilon(\cdot) \).

Remark 2 The assumption \( a > 1 \) guarantees that the interval \( \left[ \frac{1}{2} \delta, \frac{a}{a + 1} \delta \right] \) is not empty.

Remark 3 Integrating the differential inequality \( (\mathcal{H}_p)(i) \) shows that the damping coefficient in (26) (which is proportional to \( \sqrt{\varepsilon(t)} \)) must be greater than or equal to \( C/t \) for some positive constant \( C \). This is consistent with the theory of inertial gradient systems with time-dependent viscosity coefficient, which shows that the asymptotic optimization property is valid provided that the integral of the viscous damping coefficient over \([t_0, +\infty[\) be infinite, [8, 31]. See also [6, 18] where a similar growth condition on the Tikhonov parameter is considered.

For ease of reading, let us recall the functions that enter the Lyapunov analysis:

\[
E_p(t) := \left( \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \right) + \frac{1}{2} \| v_p(t) \|^2 \\
v_p(t) := \lambda \sqrt{\varepsilon(t)} \left( x(t) - x_{\varepsilon(t)} \right) + \dot{x}(t) + \beta \left[ \nabla \varphi_t(x(t)) + (p - 1) \varepsilon(t) x(t) \right] \\
\mu(t) := -\frac{\dot{\varepsilon}(t)}{2 \varepsilon(t)} + (\delta - \lambda) \sqrt{\varepsilon(t)} \\
\gamma(t) := \exp \left( \int_{t_1}^{t} \mu(s) ds \right) ,
\]

where \( \lambda \) is taken according to condition \( (\mathcal{H}_p) \). We can now state our main convergence result.

Theorem 3 Let \( x(\cdot) : [t_0, +\infty[ \rightarrow \mathcal{H} \) be a solution trajectory of the system (26). Suppose that \( \varepsilon(\cdot) \) satisfies the condition \( (\mathcal{H}_p) \). Then, the following properties are satisfied: for all \( t \geq t_1 \)

\[
E_p(t) \leq \frac{\| x^* \|^2}{2 \gamma(t)} \int_{t_1}^{t} G(s) \gamma(s) ds + \frac{\gamma(t_1) E_p(t_1)}{\gamma(t)} ,
\]

\[
\int_{t_1}^{t} \| \nabla \varphi_s(x(s)) \|^2 ds \leq \frac{2 \delta}{\beta (\delta - \lambda)} E_p(t_1) + \frac{\delta \| x^* \|^2}{\beta (\delta - \lambda)} \int_{t_1}^{t} G(s) ds ,
\]

\( \square \) Springer
where
\[ G(t) = (2a + b)\lambda \frac{\dot{\varepsilon}^2(t)}{\varepsilon^2(t)} - \dot{\varepsilon}(t) + (1 - p)\beta \lambda (\delta - \lambda)\varepsilon^2(t). \] (33)

**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 \( E_p(\cdot) \), see [28, section VIII.2]. According to Lemma 2 (i), for almost all \( t \geq t_0 \) the derivative of \( E_p(\cdot) \) is given by:
\[ \dot{E}_p(t) = \langle \nabla \varphi_p(x(t)), \dot{x}(t) \rangle + \frac{1}{2} \dot{\varepsilon}(t) \| x(t) \|^2 - \frac{1}{2} \dot{\varepsilon}(t) \| x_{\varepsilon(t)} \|^2 + \langle v_p(t), v_p(t) \rangle. \] (34)

According to the definition of \( v_p \) (see (28)), and the Eq. (26), the derivation of \( v_p \) gives
\[
\dot{v}_p(t) = \frac{\lambda}{2} \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \left( x(t) - x_{\varepsilon(t)} \right) + \lambda \sqrt{\varepsilon(t)} \dot{x}(t) - \lambda \sqrt{\varepsilon(t)} \frac{d}{dt} x_{\varepsilon(t)} + \ddot{x}(t) + \beta \frac{d}{dt} \langle \nabla \varphi_p(x(t)) + (p - 1)\varepsilon(t)x(t) \rangle
\]
\[
= \frac{\lambda}{2} \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \left( x(t) - x_{\varepsilon(t)} \right) + (\lambda - \delta) \sqrt{\varepsilon(t)} \dot{x}(t) - \lambda \sqrt{\varepsilon(t)} \frac{d}{dt} x_{\varepsilon(t)} - \nabla \varphi_p(x(t)).
\]

Let us write shortly \( A_p(t) := \nabla \varphi_p(x(t)) + (p - 1)\varepsilon(t)x(t) \). We get
\[
\langle \dot{v}_p(t), v_p(t) \rangle = \left[ \frac{\lambda}{2} \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \left( x(t) - x_{\varepsilon(t)} \right) + (\lambda - \delta) \sqrt{\varepsilon(t)} \dot{x}(t) - \lambda \sqrt{\varepsilon(t)} \frac{d}{dt} x_{\varepsilon(t)} \right] - \langle \nabla \varphi_p(x(t)), \dot{x}(t) \rangle
\]
\[
+ \frac{\lambda}{2} \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \| x(t) - x_{\varepsilon(t)} \|^2 + \lambda \left( \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} + (\lambda - \delta) \varepsilon(t) \right) (x(t) - x_{\varepsilon(t)}, \dot{x}(t))
\]
\[
+ \frac{\lambda}{2} \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \| x(t) - x_{\varepsilon(t)} \|^2 + \beta (\lambda - \delta) \sqrt{\varepsilon(t)} \left( A_p(t), \dot{x}(t) \right)
\]
\[
= \left[ \frac{\beta \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right] \langle \nabla \varphi_p(x(t)), x(t) - x_{\varepsilon(t)} \rangle - \langle \nabla \varphi_p(x(t)), \dot{x}(t) \rangle
\]
\[
- \beta \langle \nabla \varphi_p(x(t)), A_p(t) \rangle - \beta \lambda \sqrt{\varepsilon(t)} \left( A_p(t), \frac{d}{dt} x_{\varepsilon(t)} \right)
\]
\[
- \lambda^2 \varepsilon(t) \left( \frac{d}{dt} x_{\varepsilon(t)}, x(t) - x_{\varepsilon(t)} \right)
\]
\[
- \lambda \sqrt{\varepsilon(t)} \left( \frac{d}{dt} x_{\varepsilon(t)}, \dot{x}(t) \right) + \frac{(p - 1)\lambda \beta}{2} \dot{\varepsilon}(t) \sqrt{\varepsilon(t)}(x(t) - x_{\varepsilon(t)}, x(t)).
\] (35)
Since $\varphi_t$ is $\varepsilon(t)$-strongly convex, we have
\[
\varphi_t(x_{\varepsilon(t)}) - \varphi_t(x(t)) \geq \langle \nabla \varphi_t(x(t)), x_{\varepsilon(t)} - x(t) \rangle + \frac{\varepsilon(t)}{2} \| x(t) - x_{\varepsilon(t)} \|^2.
\]

Recall that $\varepsilon(t)$ is nonincreasing, i.e. $\dot{\varepsilon}(t) \leq 0$. Therefore, by using the above estimation, we get
\[
D_0 \leq \left( \frac{\beta \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \frac{1}{2} \left( \frac{\beta \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right) \varepsilon(t) \| x(t) - x_{\varepsilon(t)} \|^2.
\tag{36}
\]

For all $a > 0$ we have the following elementary inequalities
\[
-\lambda \sqrt{\varepsilon(t)} \left( \frac{d}{dt} x_{\varepsilon(t)}, \dot{x}(t) \right) \leq \frac{\lambda \sqrt{\varepsilon(t)}}{2a} \| \dot{x}(t) \|^2 + \frac{a \lambda \sqrt{\varepsilon(t)}}{2} \left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2.
\tag{37}
\]
\[
-\beta \lambda \sqrt{\varepsilon(t)} \langle A_p(t), \frac{d}{dt} x_{\varepsilon(t)} \rangle \leq \frac{\lambda \beta^2 \sqrt{\varepsilon(t)}}{2a} \| A_p(t) \|^2 + \frac{a \lambda \sqrt{\varepsilon(t)}}{2} \left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2.
\tag{38}
\]

Similarly, for all $b > 0$,
\[
-\lambda^2 \varepsilon(t) \left( \frac{d}{dt} x_{\varepsilon(t)}, x(t) - x_{\varepsilon(t)} \right) \leq \frac{b \lambda \sqrt{\varepsilon(t)}}{2} \left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2 + \frac{\lambda^3 \varepsilon^2(t)}{2b} \| x(t) - x_{\varepsilon(t)} \|^2.
\tag{39}
\]

Note that
\[
C_0 = (\lambda - \delta) \sqrt{\varepsilon(t)} \left( \| \dot{x}(t) \|^2 + \beta \langle A_p(t), \dot{x}(t) \rangle \right)
= \frac{(\lambda - \delta) \sqrt{\varepsilon(t)}}{2} \left( \| \dot{x}(t) \|^2 + \| \dot{x}(t) + \beta A_p(t) \|^2 - \beta^2 \| A_p(t) \|^2 \right). \tag{40}
\]
\[
-\beta \langle \nabla \varphi_t(x(t)), A_p(t) \rangle = -\frac{\beta}{2} \left[ \| \nabla \varphi_t(x(t)) \|^2 + \| A_p(t) \|^2 - \| \nabla \varphi_t(x(t)) - A_p(t) \|^2 \right]
= -\frac{\beta}{2} \left[ \| \nabla \varphi_t(x(t)) \|^2 + \| A_p(t) \|^2 - (p - 1)^2 \varepsilon^2(t) \| x(t) \|^2 \right]. \tag{41}
\]

By combining the relations (36)-(37)-(38)-(39)-(40)-(41) with (35), we obtain
\[
\langle \dot{v}_p(t), v_p(t) \rangle \leq \lambda \left( \frac{\beta \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \lambda \left( \frac{\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + (\lambda - \delta) \varepsilon(t) \right) \langle x(t) - x_{\varepsilon(t)}, \dot{x}(t) \rangle
+ \lambda^2 \left( \frac{\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + (\lambda - \delta) \varepsilon(t) \right) \langle x(t) - x_{\varepsilon(t)}, \dot{x}(t) \rangle
+ \left[ \frac{\lambda^2}{2} \dot{\varepsilon}(t) + \frac{\lambda \varepsilon(t)}{2} \left( \frac{\beta \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right) + \frac{\lambda^3 \varepsilon^2(t)}{2b} \right] \| x(t) - x_{\varepsilon(t)} \|^2
+ \frac{1}{2} \left( 1 + \frac{1}{a} \right) (\lambda - \delta) \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 + \frac{1}{2} (\lambda - \delta) \sqrt{\varepsilon(t)} \| \dot{x}(t) + \beta A_p(t) \|^2
+ \frac{1}{2} (2a + b) \lambda \sqrt{\varepsilon(t)} \left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2 + \frac{(p - 1)\lambda \beta}{2} \frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \langle x(t) - x_{\varepsilon(t)}, x(t) \rangle
\]
\[\begin{align*}
+ \frac{\beta}{2} \left[ \frac{\beta \lambda}{\sqrt{\varepsilon(t)}} - 1 - \beta(\lambda - \delta)\sqrt{\varepsilon(t)} \right] \| A_p(t) \|^2 - \left< \nabla \varphi_t(x(t)), \dot{x}(t) \right>
+ \frac{\beta}{2} (p - 1)^2 \varepsilon^2(t) \| x(t) \|^2 - \frac{\beta}{2} \| \nabla \varphi_t(x(t)) \|^2.
\end{align*}\] (42)

Combining (42) with (34), the terms \(\left< \nabla \varphi_t(x(t)), \dot{x}(t) \right>\) cancel each other out, which gives

\[\begin{align*}
\dot{E}_p(t) & \leq \lambda \left( \frac{\beta \dot{\varepsilon}(t)}{2 \sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) \\
& + \frac{1}{2 \varepsilon(t)} \left[ \dot{\varepsilon}(t) + \beta(p - 1)^2 \varepsilon^2(t) \right] \| x(t) \|^2 \\
& - \frac{1}{2} \dot{x}(t) \| x_{\varepsilon(t)} \|^2 + \lambda \left( \frac{\dot{\varepsilon}(t)}{2 \sqrt{\varepsilon(t)}} + (\lambda - \delta) \varepsilon(t) \right) (x(t) - x_{\varepsilon(t)}, \dot{x}(t)) \\
& + \left[ \frac{\lambda^2}{2} \dot{\varepsilon}(t) + \frac{\lambda \varepsilon(t)}{2} \left( \frac{\beta \dot{\varepsilon}(t)}{2 \sqrt{\varepsilon(t)}} - \sqrt{\varepsilon(t)} \right) + \frac{\lambda^3 \varepsilon^2(t)}{2b} \right] \| x(t) - x_{\varepsilon(t)} \|^2 \\
& + \frac{1}{2} \left( 1 + \frac{1}{a} \right) \lambda - \delta \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 + \frac{1}{2} \left( \lambda - \delta \right) \sqrt{\varepsilon(t)} \| \dot{x}(t) + \beta A_p(t) \|^2 \\
& + \frac{1}{2} (2a + b) \lambda \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 + \frac{\beta}{2} \| x(t) - x_{\varepsilon(t)} \|^2 \\
& + \frac{\beta}{2} \left[ \frac{\beta \lambda}{\sqrt{\varepsilon(t)}} - 1 - \beta(\lambda - \delta)\sqrt{\varepsilon(t)} \right] \| A_p(t) \|^2 - \frac{\beta}{2} \| \nabla \varphi_t(x(t)) \|^2. \tag{43}
\end{align*}\]

To build the differential inequality satisfied by \(E_p(\cdot)\), let us majorize

\[\begin{align*}
\mu(t) E_p(t) &= \mu(t) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \frac{\mu(t)}{2} \| v_p(t) \|^2 \\
& = \mu(t) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \frac{\mu(t) \lambda^2 \varepsilon(t)}{2} \| x(t) - x_{\varepsilon(t)} \|^2 + \frac{\mu(t)}{2} \| \dot{x}(t) + \beta A_p(t) \|^2 \\
& + \mu(t) \lambda \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, \dot{x}(t) + \beta A_p(t) \|^2 \\
& \leq \mu(t) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \frac{\mu(t) \lambda^2 \varepsilon(t)}{2} \| x(t) - x_{\varepsilon(t)} \|^2 + \frac{\mu(t)}{2} \| \dot{x}(t) + \beta A_p(t) \|^2 \\
& + \mu(t) \lambda \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, \dot{x}(t) + \beta \mu(t) \lambda \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, \nabla \varphi_t(x(t)) \|^2 \\
& + (p - 1) \beta \mu(t) \lambda \varepsilon(t) \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, x(t) \|^2 \\
& \leq \mu(t) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)})) + \mu(t) \lambda^2 \varepsilon(t) \| x(t) - x_{\varepsilon(t)} \|^2 + \frac{\mu(t)}{2} \| \dot{x}(t) + \beta A_p(t) \|^2 \\
& + \mu(t) \lambda \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, \dot{x}(t) + \beta \mu(t) \lambda \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, \nabla \varphi_t(x(t)) \|^2 \\
& + (p - 1) \beta \mu(t) \lambda \varepsilon(t) \sqrt{\varepsilon(t)} \| x(t) - x_{\varepsilon(t)}, x(t) \|^2. \tag{44}
\end{align*}\]

By adding (43) and (44), using \(\mu(t) = -\frac{\dot{\varepsilon}(t)}{2 \varepsilon(t)} + (\delta - \lambda) \sqrt{\varepsilon(t)}\), and after simplification, we get
According to Lemma 2, we have

\[
\dot{E}_p(t) + \mu(t)E_p(t) \leq \sqrt{\varepsilon(t)} \left( -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + (\delta - 2\lambda) + \lambda \beta \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} \right) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}))
\]

\[
+ \frac{1}{2} \left[ \dot{\varepsilon}(t) + \beta(p - 1)^2 \varepsilon^2(t) \right] \|x(t)\|^2 - \frac{1}{2} \dot{\varepsilon}(t) \|x_{\varepsilon(t)}\|^2
\]

\[
+ \frac{\lambda}{4} \left[ \beta \sqrt{\varepsilon(t)} \dot{\varepsilon}(t) + 2 \left( 2\delta \lambda - 2\lambda^2 - 1 \right) \varepsilon^3(t) + 2 \frac{\lambda^2}{b} \varepsilon^2(t) \right] \|x(t) - x_{\varepsilon(t)}\|^2
\]

\[
+ \frac{1}{2} \left( 1 + \frac{1}{a} \right) \lambda - \delta \right) \sqrt{\varepsilon(t)} \|x(t)\|^2 - \frac{\dot{\varepsilon}(t)}{4\varepsilon(t)} \|\dot{x}(t)\|^2
\]

\[
+ \frac{\beta}{2} \left[ \frac{\beta \lambda \sqrt{\varepsilon(t)} \dot{\varepsilon}(t)}{a} - 1 - \beta(\lambda - \delta) \sqrt{\varepsilon(t)} \right] \|A_p(t)\|^2
\]

\[
+ \frac{\beta}{2} \left( -1 - \beta \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + \beta (\delta - \lambda) \sqrt{\varepsilon(t)} \right) \|\nabla \varphi_t(x(t))\|^2.
\]

Since \( \varepsilon(\cdot) \) is nonincreasing, \( -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} \|\dot{x}(t)\|^2 \leq \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} \|\dot{x}(t)\|^2 - \frac{\beta^2 \dot{\varepsilon}(t)}{2\varepsilon(t)} \|A_p(t)\|^2. \)

According to Lemma 2, we have

\[
\left\| \frac{d}{dt} x_{\varepsilon(t)} \right\|^2 \leq \frac{\dot{\varepsilon}^2(t)}{\varepsilon^2(t)} \|x_{\varepsilon(t)}\|^2 \leq \frac{\dot{\varepsilon}^2(t)}{\varepsilon^2(t)} \|x^*\|^2.
\]

Combining these relations with \( \langle x(t) - x_{\varepsilon(t)}, x(t) \rangle = \frac{1}{2} \left[ \|x(t) - x_{\varepsilon(t)}\|^2 + \|x(t)\|^2 - \|x_{\varepsilon(t)}\|^2 \right] \), we get

\[
\dot{E}_p(t) + \mu(t)E_p(t) \leq \sqrt{\varepsilon(t)} \left( -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + (\delta - 2\lambda) + \lambda \beta \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} \right) (\varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}))
\]

\[
+ \frac{1}{2} \left[ \dot{\varepsilon}(t) + \beta(p - 1)^2 \varepsilon^2(t) \right] \|x(t)\|^2 - \frac{1}{2} \dot{\varepsilon}(t) \|x_{\varepsilon(t)}\|^2
\]

\[
+ \frac{1}{2} \left( 1 + \frac{1}{a} \right) \lambda - \delta \right) \sqrt{\varepsilon(t)} \|x(t)\|^2 - \frac{\dot{\varepsilon}(t)}{4\varepsilon(t)} \|\dot{x}(t)\|^2
\]

\[
+ \frac{1}{2} \left[ \frac{\beta \lambda \sqrt{\varepsilon(t)} \dot{\varepsilon}(t)}{a} - 1 - \beta(\lambda - \delta) \sqrt{\varepsilon(t)} \right] \|A_p(t)\|^2
\]

\[
+ \frac{\beta}{2} \left( -1 - \beta \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + \beta (\delta - \lambda) \sqrt{\varepsilon(t)} \right) \|\nabla \varphi_t(x(t))\|^2.
\]

(45)
Let us analyze the sign of the coefficients involved in (46). Since $\varepsilon(t)$ is nonincreasing and $p \leq 1$,

$$\dot{E}_p(t) + \mu(t)E_p(t) \leq \sqrt{\varepsilon(t)} \left( -\frac{\dot{\varepsilon}(t)}{2\varepsilon^{\frac{1}{2}}(t)} + (\delta - 2\lambda) + \lambda \beta \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} \right) \left( \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \right)$$

$$+ \frac{1}{2} \left[ \dot{\varepsilon}(t) + \beta(p - 1)(p - 1 + \lambda(\delta - \lambda)) \varepsilon^2(t) \right] \|x(t)\|^2$$

$$+ \frac{\lambda}{4} \left[ \beta \sqrt{\varepsilon(t)} \dot{\varepsilon}(t) - 4 \left( 1 - \frac{1}{2}b \right) \lambda^2 - \delta \lambda + \frac{1}{2} \right] \varepsilon^3(t) \leq 0$$

$$\|x(t) - x_{\varepsilon(t)}\|^2$$

$$+ \frac{1}{2} \left( 1 + \frac{1}{a} \lambda - \delta \right) \sqrt{\varepsilon(t)} - \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} \|\dot{x}(t)\|^2$$

$$+ \frac{\beta}{2} \left[ \beta \left( \frac{\lambda}{a} + \delta - \lambda \right) \sqrt{\varepsilon(t)} - 1 - \frac{\beta \dot{\varepsilon}(t)}{\varepsilon(t)} \right] \|A_p(t)\|^2$$

$$+ \frac{1}{2} \left( 2a + b \right) \lambda \frac{\dot{\varepsilon}^2(t)}{\varepsilon^2(t)} - \dot{\varepsilon}(t) + (1 - p) \beta \lambda (\delta - \lambda) \varepsilon^2(t) \right] \|x_{\varepsilon(t)}\|^2$$

$$+ \frac{\beta}{2} \left[ -1 - \beta \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + \beta (\delta - \lambda) \sqrt{\varepsilon(t)} \right] \|\nabla \varphi_t(x(t))\|^2. \quad (47)$$

• Condition ($\mathcal{H}_p$)(i) implies that

$$\frac{d}{dt} \left( \frac{1}{\sqrt{\varepsilon(t)}} \right) \leq (2\lambda - \delta), \quad \frac{d}{dt} \left( \frac{1}{\sqrt{\varepsilon(t)}} \right) \leq \frac{1}{2} \left( \delta - (1 + \frac{1}{a}) \lambda \right) \quad \text{and} \quad \delta \beta \leq \frac{1}{\sqrt{\varepsilon(t)}}.$$ 

So, we have

$$A = -\frac{\dot{\varepsilon}(t)}{2\varepsilon^{\frac{1}{2}}(t)} + (\delta - 2\lambda) = \frac{d}{dt} \left( \frac{1}{\sqrt{\varepsilon(t)}} \right) + (\delta - 2\lambda) \leq 0.$$ 

• According to ($\mathcal{H}_p$)(ii), we have

$$C = \beta \sqrt{\varepsilon(t)} \dot{\varepsilon}(t) - 4 \left( 1 - \frac{1}{2b} \right) \lambda^2 - \delta \lambda + \frac{1}{2} \varepsilon^3(t).$$
\[ \begin{align*}
&= \varepsilon^2(t) \left[ \beta \frac{d}{dt} (\ln(\varepsilon(t))) - 4 \left( \left(1 - \frac{1}{2b}\right) \lambda^2 - \delta \lambda + \frac{1}{2} \right) \right] \leq 0.
\end{align*} \]

- Similarly, using \((\mathcal{H}_p)(iii)\), we get
\[ \begin{align*}
B &= \dot{\varepsilon}(t) + \beta (p-1) (p-1+\lambda(\delta - \lambda)) \varepsilon^2(t) \\
&= \varepsilon^2(t) \left[ \beta (1-p) (\lambda^2 - \delta \lambda + 1 - p) - \frac{d}{dt} \left( \frac{1}{\varepsilon(t)} \right) \right] \leq 0.
\end{align*} \]

- Condition \((\mathcal{H}_p)(i)\), successively implies that
\[ \begin{align*}
D &= \sqrt{\varepsilon(t)} \left[ \frac{1}{2} \left(1 + \frac{1}{a} \right) \lambda - \delta \right] - \frac{\dot{\varepsilon}(t)}{2 \varepsilon^2(t)} \\
&= \sqrt{\varepsilon(t)} \left[ \frac{d}{dt} \left( \frac{1}{\sqrt{\varepsilon(t)}} \right) + \frac{1}{2} \left(1 + \frac{1}{a} \right) \lambda - \delta \right] \leq 0;
\end{align*} \]
\[ \begin{align*}
\Delta_1 &= \beta \left( \frac{\lambda}{a} + \delta - \lambda \right) \sqrt{\varepsilon(t)} - 1 - \frac{\beta \dot{\varepsilon}(t)}{\varepsilon(t)} \\
&= 2\beta (\delta - \lambda) \sqrt{\varepsilon(t)} - 1 + 2\beta \sqrt{\varepsilon(t)} \left( \frac{\dot{\varepsilon}(t)}{2 \varepsilon^2(t)} - \frac{1}{2} \left( \delta - \left(1 + \frac{1}{a} \right) \lambda \right) \right) \\
&\leq 2\beta (\delta - \lambda) \sqrt{\varepsilon(t)} - 1 = \beta \left( \delta - 2\lambda \right) \sqrt{\varepsilon(t)} + \beta \delta \sqrt{\varepsilon(t)} - 1 \leq 0;
\end{align*} \]
\[ \begin{align*}
\Delta_2 &= \lambda \beta \sqrt{\varepsilon(t)} - 1 + \beta \sqrt{\varepsilon(t)} \left( \delta - 2\lambda \right) + \frac{d}{dt} \left( \frac{1}{\sqrt{\varepsilon(t)}} \right) \\
&\leq \lambda \left( \beta \sqrt{\varepsilon(t)} - \frac{1}{\delta} \right) + \frac{\lambda}{\delta} - 1 \leq \frac{\lambda}{\delta} - 1.
\end{align*} \]

Collecting the above results, it follows from the estimate (47) that
\[ \begin{align*}
\dot{E}_p(t) + \mu(t) E_p(t) &\leq \frac{1}{2} \left[ (2a + b) \lambda \frac{\dot{\varepsilon}^2(t)}{\varepsilon^2(t)} - \dot{\varepsilon}(t) + (1-p) \beta \lambda (\delta - \lambda) \varepsilon^2(t) \right] \| x_{\varepsilon(t)} \|^2 \\
&+ \frac{\mu}{2} \left( \frac{\lambda}{\delta} - 1 \right) \| \nabla \phi_t(x(t)) \|^2.
\end{align*} \]

Since \( \| x_{\varepsilon(t)} \| \leq \| x^* \| \), we get
\[ \begin{align*}
\dot{E}_p(t) + \mu(t) E_p(t) &\leq \frac{\| x^* \|^2}{2} G(t) + \frac{\mu}{2} \left( \frac{\lambda}{\delta} - 1 \right) \| \nabla \phi_t(x(t)) \|^2, \quad (48)
\end{align*} \]

where \( G(t) = (2a + b) \lambda \frac{\dot{\varepsilon}^2(t)}{\varepsilon^2(t)} - \dot{\varepsilon}(t) + (1-p) \beta \lambda (\delta - \lambda) \varepsilon^2(t) \).
We have $\lambda - 1 \leq 0$. By taking $\gamma(t) = \exp \left( \int_{t_1}^{t} \mu(s) ds \right)$, and setting

$$W_p(t) := e^{\int_{t_1}^{t} \mu(s) ds} E_p(t)$$

we conclude that

$$\dot{W}_p(t) = \gamma(t) \left( \dot{E}_p(t) + \mu(t) E_p(t) \right) \leq \frac{\|x^*\|^2}{2} G(t) \gamma(t).$$

By integrating (50) on $[t_1, t]$, and dividing by $\gamma(t)$, we obtain our claim (31)

$$E_p(t) \leq \frac{\|x^*\|^2}{2} \int_{t_1}^{t} G(s) \gamma(s) ds + \frac{\gamma(t_1) E_p(t_1)}{\gamma(t)}.$$  (51)

Coming back to (48), we get by integration

$$E_p(t) - E_p(t_1) + \int_{t_1}^{t} \mu(s) E_p(s) ds + \frac{\beta}{2} \left(1 - \frac{\lambda}{\delta}\right) \int_{t_1}^{t} \|\nabla \varphi(x(s))\|^2 ds$$

$$\leq \frac{\|x^*\|^2}{2} \int_{t_1}^{t} G(s) ds.$$  (52)

Our assertion (32) is then reached by neglecting the positive term $E_p(t) + \int_{t_1}^{t} \mu(s) E_p(s) ds$.  \hfill \Box

**Corollary 1** Let $x(\cdot) : [t_0, +\infty[ \to \mathcal{H}$ be a solution trajectory of the system (TRISHE)

$$\ddot{x}(t) + \delta \sqrt{\varepsilon(t)} \dot{x}(t) + \beta \frac{d}{dt} \left( \nabla \varphi_t(x(t)) \right) + \nabla \varphi_t(x(t)) = 0.$$  (53)

Let us assume that there exists $a, b, \lambda$ and $t_1 \geq t_0$ such that for all $t \geq t_1$, $\mathcal{(H_1)}$ holds. Then,

$$E_1(t) \leq \frac{\|x^*\|^2}{2} \int_{t_1}^{t} \left[ \left( 2a + b \frac{\dot{\varepsilon}(s)}{\varepsilon(s)} - \frac{\dot{\varepsilon}(s)}{\varepsilon(s)} \right) \frac{\gamma(s)}{\gamma(t)} \right] ds + \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)}$$

where $\gamma(t) = \exp \left( \int_{t_1}^{t} \mu(s) ds \right)$ and $\mu(t) = -\frac{\dot{\varepsilon}(t)}{2 \varepsilon(t)} + (\delta - \lambda) \sqrt{\varepsilon(t)}$.

**Corollary 2** Let $x(\cdot) : [t_0, +\infty[ \to \mathcal{H}$ be a solution trajectory of the system (TRISH)

$$\ddot{x}(t) + \delta \sqrt{\varepsilon(t)} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla \varphi_t(x(t)) = 0.$$  (54)
Let us assume that there exists \( a, b, \lambda \) and \( t_1 \geq t_0 \) such that for all \( t \geq t_1 \), condition \((\mathcal{H}_0)\) holds. Then,

\[
E_0(t) \leq \frac{\|x^*\|^2}{2} \int_{t_1}^{t} \left[ \left( 2a + b \right) \frac{\dot{\varepsilon}(s)}{\varepsilon(s)} - \dot{\varepsilon}(s) + \beta \lambda (\delta - \lambda) \varepsilon^2(s) \right] \gamma(s) \, ds + \frac{\gamma(t_1)E_0(t_1)}{\gamma(t)}
\]

where \( \gamma(t) = \exp \left( \int_{t_1}^{t} \mu(s) \, ds \right) \) and \( \mu(t) = -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + (\delta - \lambda) \sqrt{\varepsilon(t)} \).

### 3 Particular Cases

Take \( \varepsilon(t) = \frac{1}{t^r} \), \( 0 < r < 2 \), \( t_0 > 0 \), and consider the systems (TRISHE) and (TRISH). The convergence rate of the values and the strong convergence to the minimum norm solution will be obtained by particularizing Theorem 3 to these situations. In these cases, the integrals that enter into the formulation of Theorem 3 can be calculated explicitly. In fact, obtaining sharp convergence rate of the gradients requires another Lyapunov analysis based on the function \( \mathcal{E}_p \) defined by

\[
\mathcal{E}_p(t) := \left( \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \right) + \frac{1}{2} \| \dot{x}(t) + \beta \left[ \nabla \varphi_t(x(t)) + (p - 1)\varepsilon(t) x(t) \right] \|^2.
\]

We can notice that when \( \lambda = 0 \), we have \( \mathcal{E}_p(t) = E_p(t) \). So, with \( \lambda = 0 \), the estimation (43) becomes

\[
\dot{\mathcal{E}}_p(t) = \dot{E}_p(t) \leq \frac{1}{2} \left[ \dot{\varepsilon}(t) + \beta (p - 1)^2 \varepsilon^2(t) \right] \| x(t) \|^2 - \frac{1}{2} \dot{\varepsilon}(t) \| x_{\varepsilon(t)} \|^2 - \frac{\delta}{2} \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 - \frac{\delta}{2} \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 - \frac{\beta}{2} \| \nabla \varphi_t(x(t)) \|^2.
\]

By supposing \( \delta \beta \leq \frac{1}{\sqrt{\varepsilon(t)}} \), we conclude that

\[
\dot{\mathcal{E}}_p(t) \leq \frac{1}{2} \left[ \dot{\varepsilon}(t) + \beta (p - 1)^2 \varepsilon^2(t) \right] \| x(t) \|^2 - \frac{1}{2} \dot{\varepsilon}(t) \| x^* \|^2 - \frac{\delta}{2} \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 - \frac{\beta}{2} \| \nabla \varphi_t(x(t)) \|^2.
\]

### 3.1 \( p = 1 \): System (TRISHE)

**Theorem 4** Take \( \delta < \sqrt{2} \), \( \varepsilon(t) = \frac{1}{t^r} \) and \( 0 < r < 2 \). Let \( x : [t_0, +\infty[ \rightarrow \mathcal{H} \) be a solution trajectory of
\( \ddot{x}(t) + \left( \frac{\delta}{t^2} + \frac{\beta}{t^r} \right) \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + \left( \frac{1}{t^r} - \frac{r\beta}{t^{r+1}} \right) x(t) = 0. \) (58)

(a) Then, we have convergence of values, strong convergence to the minimum norm solution, and

\[
\begin{align*}
    f(x(t)) - \min_{\mathcal{H}} f &= \mathcal{O}\left( \frac{1}{t^r} \right) \text{ as } t \to +\infty; \quad (59) \\
    \|x(t) - x_{\varepsilon(t)}\|^2 &= \mathcal{O}\left( \frac{1}{t^{2r}} \right) \text{ as } t \to +\infty. \quad (60) \\
    \|\dot{x}(t) + \beta \nabla f(x(t))\| &= \mathcal{O}\left( \frac{1}{t^{\min\left( \frac{2r}{2+r}, r \right)}} \right) \text{ as } t \to +\infty. \quad (61)
\end{align*}
\]

(b) In addition, we have the following integral estimates

\[
\int_{t_1}^{+\infty} t^{r-1} \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_1}^{+\infty} t^{(3r-1)} \|
abla f(x(t))\|^2 dt < +\infty.
\]

**Proof** (a) By taking \( \varepsilon(t) = \frac{1}{t^r} \) in Corollary 1, we get (58). Let us verify that \((\mathcal{H}_1)\) is satisfied by taking \( a, b, \lambda \) satisfying \( a > 1, \ b > \frac{1}{2} \) and \( \lambda \in \left[ \frac{1}{2} \delta, \frac{a}{a+1} \delta \right]. \)

\( \ast \) (i) Since \( \frac{\delta}{2} < \lambda < \frac{a}{a+1} \delta \), we have \( \min(2\lambda - \delta, \frac{1}{2}(\delta - \frac{a+1}{a} \lambda)) > 0 \). On the other hand, since \( r < 2 \) we have \( \lim_{t \to +\infty} t^{\frac{r-2}{2}} = 0 \), which implies that for \( t \geq t_1 \) large enough,

\[
\frac{d}{dt} \left( \frac{1}{\varepsilon(t)} \right) = \frac{r}{2} t^{\frac{r-2}{2}} \leq \min \left( 2\lambda - \delta, \frac{1}{2} \left( \delta - \frac{a+1}{a} \lambda \right) \right) \quad \text{and} \quad \beta \delta \leq t^{\frac{r-2}{2}} = \frac{1}{\sqrt{\varepsilon(t)}}.
\]

\( \ast \) (ii) We have for each \( t > t_0 \), \( \frac{d}{dt} \left( \ln(\varepsilon(t)) \right) = -\frac{r}{t} < 0. \) So to ensure \((\mathcal{H}_1)(ii)\), it suffices to have

\[
(1 - \frac{1}{2b}) \lambda^2 - \delta \lambda + \frac{1}{2} \geq 0. \quad (62)
\]

Since \( \delta < \sqrt{2} \), one can take \( b \geq \frac{1}{2 - \delta \sqrt{2}} > 0 \) such that \( \delta^2 - 2 + \frac{1}{b} \leq 0 \), hence, (62) is satisfied for each \( \lambda > 0 \), and especially for \( \lambda \in \left[ \frac{1}{2} \delta, \frac{a}{a+1} \delta \right]. \)

---

2 One of the Referees offered to justify (62) by taking \( 1 - \frac{1}{2b} < 0 \), and then

\[
\frac{\delta}{2} < \lambda < \min \left\{ \frac{a}{a+1} \delta, \frac{1}{2(\frac{1}{2b} - 1)} \left( \delta + \sqrt{\delta^2 + \frac{1}{b} - 2} \right) \right\},
\]

which is valid, when starting with \( 0 < \delta < 1 \), we can choose \( b > 0 \) such that \( \delta^2 < \frac{4b}{1+2b} \). This remains true since the function \( b \mapsto \frac{4b}{1+2b} \) is increasing and it attains the value 1 when \( b = \frac{1}{2} \).
\( (iii) \) is immediate, since for \( p = 1 \), we have \( \frac{d}{dt} \left( \frac{1}{\varepsilon(t)} \right) = rt^{r-1} \geq 0. \)

This expresses that the condition (\( \mathcal{H}_1 \)) holds. With the notations of Theorem 3, we have

\[
\mu(t) = -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + (\delta - \lambda)\sqrt{\varepsilon(t)} = \frac{r}{2t} + \frac{\delta - \lambda}{t^2} \tag{63}
\]

\[
\gamma(t) = \exp \left( \int_{t_1}^{t} \mu(s) ds \right) = \left( \frac{t}{t_1} \right)^{\frac{r}{2}} \exp \left[ \frac{2(\delta - \lambda)}{2 - r} \left( t^{\frac{2-r}{2}} - t_1^{\frac{2-r}{2}} \right) \right] = C_1 t^{\frac{r}{2}} \exp \left[ \frac{2(\delta - \lambda)}{2 - r} \left( t^{\frac{2-r}{2}} - t_1^{\frac{2-r}{2}} \right) \right]^{-1} \tag{64}
\]

Setting \( \lambda_0 := (2a + b)\lambda, \delta_0 := \frac{2(\delta - \lambda)}{2 - r} \), and replacing \( \varepsilon(t) \) and \( \gamma(t) \) by their values in (53), we get

\[
E_1(t) \leq \frac{r \| x^* \|^2}{2t^{\frac{r}{2}} \exp \left( \delta_0 t^{\frac{2-r}{2}} \right)} \int_{t_1}^{t} \left( \frac{\lambda_0 r}{s^2} + \frac{1}{s^\frac{r+2}{2}} \right) \exp \left( \delta_0 s^{\frac{2-r}{2}} \right) ds + \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)}. \tag{65}
\]

Notice that

\[
\frac{d}{ds} \left( \frac{1}{\rho s} \exp \left( \delta_0 s^{\frac{2-r}{2}} \right) \right) = \left( -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s^{\frac{r+2}{2}}} \right) \exp \left( \delta_0 s^{\frac{2-r}{2}} \right).
\]

For \( s \) large enough, by taking \( 0 < \rho < \frac{1}{a+1} \delta \), we have \( \frac{\lambda_0 r}{s^2} + \frac{1}{s^{\frac{r+2}{2}}} \leq -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s^{\frac{r+2}{2}}} \), which gives

\[
E_1(t) \leq \frac{r}{2t^{\frac{r}{2}} \exp \left( \delta_0 t^{\frac{2-r}{2}} \right)} \int_{t_1}^{t} \left( -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s^{\frac{r+2}{2}}} \right) \exp \left( \delta_0 s^{\frac{2-r}{2}} \right) ds + \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)} = \frac{1}{2t^{\frac{r+2}{2}} \exp \left( \delta_0 t^{\frac{2-r}{2}} \right)} \int_{t_1}^{t} \frac{d}{ds} \left( \frac{1}{\rho s} \exp \left( \delta_0 s^{\frac{2-r}{2}} \right) \right) ds + \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)} = \frac{r}{2\rho t^{\frac{r+2}{2}}} + \frac{r}{t^{\frac{2-r}{2}} \exp \left( \delta_0 t^{\frac{2-r}{2}} \right)} \frac{1}{2\rho t_1} \exp \left( \delta_0 t_1^{\frac{2-r}{2}} \right) + \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)} \leq \frac{r}{2\rho t^{\frac{r+2}{2}}} + \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)}.
\]

We have \( \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)} \leq C t^{-\frac{r}{2}} \exp \left[ -\delta_0 t^{\frac{2-r}{2}} \right] \). Since \( 0 < r < 2 \) and \( \delta_0 > 0 \), we deduce that \( \frac{\gamma(t_1) E_1(t_1)}{\gamma(t)} \) tends to zero at an exponential rate, as \( t \to +\infty \). Therefore,
there exists a positive constant $C$ such that for $t$ large enough

$$E_1(t) \leq \frac{C}{t^\frac{2r}{2}}. \quad (66)$$

By Lemma 1, we deduce that there exists positive constants $C$ and $M$ such that, for $t$ large enough,

$$f(x(t)) - \min_{\mathcal{H}} f \leq C \left( \frac{1}{t^\frac{2r}{2}} + \frac{1}{t^r} \right), \quad \|x(t) - x_{\epsilon(t)}\|^2 \leq \frac{2E_1(t)}{\epsilon(t)} \leq \frac{2C}{t^\frac{2r}{2}}, \quad \text{and}$$

$$\|\dot{x}(t) + \beta \nabla \varphi_t(x(t))\|^2 \leq 2ME_1(t) \leq \frac{MC}{t^{2r}}. \quad (67)$$

Since $0 < r < 2$, we conclude that

$$f(x(t)) - \min_{\mathcal{H}} f = O\left( \frac{1}{t^r} \right), \quad \|x(t) - x_{\epsilon(t)}\|^2 = O\left( \frac{1}{t^\frac{2r}{2}} \right), \quad \text{as} \ t \to +\infty.$$

By (67), we have

$$\|\dot{x}(t) + \beta \nabla f(x(t))\| \leq \|\dot{x}(t) + \beta \nabla \varphi_t(x(t))\| + \epsilon(t)\|x(t)\| \leq \sqrt{\frac{MC}{t^{2r}}} + \frac{1}{t^r}\|x(t)\|. \quad (68)$$

Since $x$ is bounded, we conclude that, as $t \to +\infty$,

$$\|\dot{x}(t) + \beta \nabla f(x(t))\| = \begin{cases} O\left( \frac{1}{t^\frac{2r}{2}} \right) & \text{if} \ r \in [\frac{2}{3}, 2[ \\ O\left( \frac{1}{t^r} \right) & \text{if} \ r \in ]0, \frac{2}{3}[. \end{cases}$$

(b) We now come to the integral estimates of the velocities and gradient terms. For this, we use the pointwise estimates already established, and proceed with the Lyapunov function $E_p$ defined in (56). The system (TRISHE) corresponds to $p = 1$, so we consider

$$E_1(t) := \left( \varphi_t(x(t)) - \varphi_t(x_{\epsilon(t)}) \right) + \frac{1}{2} \|\dot{x}(t) + \beta \nabla \varphi_t(x(t))\|^2. \quad (69)$$

Since for $t > t_1$, $\delta \beta \leq \frac{1}{\sqrt{\epsilon(t)}}$, then according to (57), we have

$$\dot{E}_1(t) \leq \frac{1}{2} \dot{\epsilon}(t)\|x(t)\|^2 - \frac{1}{2} \dot{\epsilon}(t)\|x^\ast\|^2 - \frac{\delta}{2} \sqrt{\epsilon(t)}\|\dot{x}(t)\|^2$$

$$- \frac{\delta}{2} \sqrt{\epsilon(t)}\|\dot{x}(t) + \beta \nabla \varphi_t(x(t))\|^2 - \frac{\beta}{2} \|\nabla \varphi_t(x(t))\|^2 \leq -\frac{1}{2} \dot{\epsilon}(t)\|x^\ast\|^2 - \frac{\delta}{2} \sqrt{\epsilon(t)}\|\dot{x}(t)\|^2 - \frac{\beta}{2} \|\nabla \varphi_t(x(t))\|^2. \quad (68)$$
Equivalently,
\[
\frac{\delta}{2} \sqrt{\varepsilon(t)} \| \dot{x}(t) \|^2 + \frac{\beta}{2} \| \nabla \varphi_t(x(t)) \|^2 \leq -\dot{\mathcal{E}}_1(t) - \frac{1}{2} \ddot{x}(t) \| x^* \|^2.
\]

By multiplying this last equality by \( t^{\frac{3r}{2} - 1} \) and integrating on \([t_1, T]\), we get
\[
\frac{\delta}{2} \int_{t_1}^{T} t^{r-1} \| \dot{x}(t) \|^2 dt + \frac{\beta}{2} \int_{t_1}^{T} t^{\frac{3r}{2} - 1} \| \nabla \varphi_t(x(t)) \|^2 dt \leq -\int_{t_1}^{T} t^{\frac{3r}{2} - 1} \dot{\mathcal{E}}_1(t) dt \\
+ \frac{\| x^* \|^2}{2} \int_{t_1}^{T} t^{\frac{r}{2} - 2} dt.
\]

We have
\[
-\int_{t_1}^{T} t^{\frac{3r}{2} - 1} \dot{\mathcal{E}}_1(t) dt = \left( t^{\frac{3r}{2} - 1} \mathcal{E}_1(t_1) - T t^{\frac{3r}{2} - 1} \mathcal{E}_1(T) \right) + \frac{3r - 2}{2} \int_{t_1}^{T} t^{\frac{3r}{2} - 2} \dot{\mathcal{E}}_1(t) dt \\
\leq t^{\frac{3r}{2} - 1} \mathcal{E}_1(t_1) + \frac{3r - 2}{2} \int_{t_1}^{T} t^{\frac{3r}{2} - 2} (\varphi_t(x(t)) - \varphi_t(x_\varepsilon(t_1))) dt \\
+ \frac{3r - 2}{4} \int_{t_1}^{T} t^{\frac{3r}{2} - 2} \| \dot{x}(t) + \beta \nabla \varphi_t(x(t)) \|^2 dt \\
\leq t^{\frac{3r}{2} - 1} \mathcal{E}_1(t_1) + \frac{3r - 2}{2} \int_{t_1}^{T} t^{\frac{3r}{2} - 2} \mathcal{E}_1(t) dt + \frac{3r - 2}{4} \int_{t_1}^{T} t^{\frac{3r}{2} - 2} \| \dot{x}(t) \|^2 dt + \beta \nabla \varphi_t(x(t)) \|^2 dt.
\]

According to (66) and (67), we deduce that there exists \( C > 0 \) such that
\[
-\int_{t_1}^{T} t^{\frac{3r}{2} - 1} \dot{\mathcal{E}}_1(t) dt \leq t^{\frac{3r}{2} - 1} \mathcal{E}_1(t_1) + C \int_{t_1}^{T} t^{r-3} dt.
\]

From this, we deduce that \(-\int_{t_1}^{+\infty} t^{\frac{3r}{2} - 1} \dot{\mathcal{E}}_1(t) dt < +\infty\). By (69), we conclude that
\[
\frac{\delta}{2} \int_{t_1}^{+\infty} t^{r-1} \| \dot{x}(t) \|^2 dt + \frac{\beta}{2} \int_{t_1}^{+\infty} t^{\frac{3r}{2} - 1} \| \nabla \varphi_t(x(t)) \|^2 dt < +\infty.
\]

Therefore
\[
\int_{t_1}^{+\infty} t^{r-1} \| \dot{x}(t) \|^2 dt < +\infty \quad \text{and} \quad \int_{t_1}^{+\infty} t^{\frac{3r}{2} - 1} \| \nabla \varphi_t(x(t)) \|^2 dt < +\infty.
\]

We also have
\[
\| \nabla f(x(t)) \|^2 \leq \left( \| \nabla \varphi_t(x(t)) \| + \frac{1}{t^r} \| x(t) \| \right)^2 \leq 2 \| \nabla \varphi_t(x(t)) \|^2 + \frac{2}{t^{2r}} \| x(t) \|^2.
\]
Since \( x \) is bounded, we obtain
\[
\int_{t_1}^{+\infty} r^{\frac{3}{2} - 1} \| \nabla f(x(t)) \|^2 dt < +\infty.
\]
This completes the proof. \( \square \)

### 3.2 \( p = 0 \): System (TRISH)

We now come to the corresponding result for (TRISH), stated as a model result in the introduction.

**Theorem 5** Take \( \delta > 2 \), \( \varepsilon(t) = \frac{1}{t^r} \), with \( 1 \leq r < 2 \). Let \( x : \left[ t_0, +\infty \right) \to \mathcal{H} \) be a solution trajectory of

\[
\ddot{x}(t) + \frac{\delta}{t^2} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + \frac{1}{t^r} x(t) = 0. \tag{72}
\]

Then, we have the following estimates

\[
f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left( \frac{1}{t^{r/2}} \right) \text{ as } t \to +\infty; \tag{73}
\]

\[
\|x(t) - x_{\varepsilon(t)}\|^2 = \mathcal{O}\left( \frac{1}{t^{2-r}} \right) \text{ as } t \to +\infty. \tag{74}
\]

\[
\|\dot{x}(t) + \beta \nabla f(x(t))\| = \mathcal{O}\left( \frac{1}{t^{r/2}} \right) \text{ as } t \to +\infty. \tag{75}
\]

\[
\int_{t_1}^{+\infty} t^{r-1} \|\dot{x}(t)\|^2 dt < +\infty, \quad \int_{t_1}^{+\infty} t^{3r-2} \|\nabla f(x(t))\|^2 dt < +\infty. \tag{76}
\]

**Proof** (a) Taking \( \varepsilon(t) = \frac{1}{t^r} \) in Corollary 2 gives (72). So if the condition \((\mathcal{H}_0)\) is satisfied, we get

\[
E_0(t) \leq \frac{\|x^\ast\|^2}{2\gamma(t)} \int_{t_1}^{t} \left[ \left( 2a + b \right) \lambda \left( \frac{2}{\varepsilon(s)} - \frac{1}{\varepsilon(s)} \right) \gamma(s) \right] ds + \frac{\gamma(t_1)E_0(t_1)}{\gamma(t)}. \tag{77}
\]

As in the proof of Theorem 4, conditions (i) and (ii) of \((\mathcal{H}_0)\) are satisfied by taking \( a, b, \lambda \) satisfying \( a > 1, \frac{1}{4} < b < \frac{1}{2} \) and \( \lambda \in \left[ \frac{1}{2}, \frac{a + 1}{a + 1} \right] \).

So, it suffices to check (iii) for \( p = 0 \), i.e., for \( t \) large enough

\[
\frac{d}{dt} \left( \frac{1}{\varepsilon(t)} \right) = rt^{r-1} \geq \beta \left( \lambda^2 - \delta \lambda + 1 \right). \tag{78}
\]

For all \( r > 1 \) this condition is obviously verified since \( \lim_{t \to +\infty} t^{r-1} = +\infty \). If \( 0 < r \leq 1 \), we obtain from \( \delta > 2 \) that \( \lambda^2 - \delta \lambda + 1 \leq 0 \) is satisfied whenever

\[
\frac{1}{2} \left( \delta - \sqrt{\delta^2 - 4} \right) \leq \lambda \leq \frac{1}{2} \left( \delta + \sqrt{\delta^2 - 4} \right). \tag{79}
\]
Combining this condition with $\frac{\delta}{2} < \lambda < \frac{a}{a+1} \delta$, we conclude that (iii) is satisfied for

$$\frac{\delta}{2} < \lambda < \min \left( \frac{a}{a+1} \delta, \frac{1}{2} \left( \delta + \sqrt{\delta^2 - 4} \right) \right).$$

So condition $(H_0)$ is satisfied by taking

$$a > 1, \quad \frac{1}{4} < b < \frac{1}{2}, \quad \frac{1}{2} \delta \leq \lambda \leq \min \left( \frac{a}{a+1} \delta, \frac{1}{2} \left( \delta + \sqrt{\delta^2 - 4} \right) \right).$$

By setting $\lambda_0 := (2a + b) \lambda$, $\delta_0 := \frac{2(\delta - \lambda)}{2 - r}$ and $\beta_0 = \beta \lambda (\delta - \lambda)$ and combining the Eqs. (63) and (64) with (77), we get

$$E_0(t) \leq \frac{r}{2t^2 \exp \left( \frac{\delta_0 t}{rs^2} \right)} \int_{t_1}^{t} \left( \frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} + \frac{\beta_0}{rs \frac{r^2}{2}} \right) \exp \left( \delta_0 s \frac{2-r}{s^2} \right) ds + \frac{\gamma(t_1) E_0(t_1)}{\gamma(t)}. \tag{78}$$

Let us estimate the integral $\int_{t_1}^{t} \left( \frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} + \frac{\beta_0}{rs \frac{r^2}{2}} \right) \exp \left( \delta_0 s \frac{2-r}{s^2} \right) ds$. For $\rho > 0$

$$\frac{d}{ds} \left( \frac{\rho s^2}{\delta_0 s \frac{2-r}{s^2}} \right) = \left( -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s \frac{r^2}{2}} \right) \exp \left( \delta_0 s \frac{2-r}{s^2} \right).$$

So, we need to show that

$$\frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} + \frac{\beta_0}{rs \frac{r^2}{2}} \leq -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s \frac{r^2}{2}}. \tag{79}$$

Since $r \geq 1$, we have for $s$ large enough, $\frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} + \frac{\beta_0}{rs \frac{r^2}{2}} \leq \frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} \leq \frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}}$. By taking $\rho < \frac{r}{(a+1)(r+\beta_0) \delta}$, we have

$$\frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} \leq -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s \frac{r^2}{2}} \iff \frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} \leq \left( \frac{\delta_0 (2-r)}{2 \rho} - 1 - \frac{\beta_0}{r} \right) \frac{1}{s \frac{r^2}{2}} = \frac{\delta - \lambda}{\rho} - 1 - \frac{\beta_0}{r} \iff \frac{\lambda_0 r s^2}{s^2} + \frac{1}{s \frac{r^2}{2}} \leq \frac{\delta - \lambda}{\rho} - 1 - \frac{\beta_0}{r}.$$
We have $0 < r < 2$, which means that \( \lim_{s \to +\infty} \frac{1}{s^{\frac{2-r}{2}}} = 0 \). Combining the fact that \( \lambda < \frac{a}{a+1} \delta \), with the choice of \( \rho \), one can check that

\[
\delta - \lambda - \rho \left( \frac{1 + \frac{\beta_0}{r}}{} \right) > 0
\]

Therefore, for \( s \) large enough, the last above inequalities are satisfied, which implies that, for \( 1 \leq r < 2 \), and \( t_1 \) large enough, we have

\[
E_0(t) \leq \frac{r}{2t^2} \exp\left( \delta_0 t^{\frac{2-r}{2}} \right) \int_{t_1}^{t} \left( -\frac{1}{\rho s^2} + \frac{\delta_0 (2-r)}{2 \rho s^{\frac{r+2}{2}}} + \frac{\beta_0}{r \sigma} \right) \exp\left( \delta_0 s^{\frac{2-r}{2}} \right) ds
\]

Using successively the derivation chain rule, the equation (TRISH), and \( \dot{\varepsilon}(t) \leq 0 \), we get

\[
\frac{d}{dt} E_0(t) = (\nabla \varphi_t(x(t)), \dot{x}(t)) + \frac{\dot{\varepsilon}(t)}{2} \|x(t)\|^2 + \|\dot{x}(t)\|^2 + \beta \nabla f(x(t)) \dot{x}(t)
\]

We now proceed in the same way as in the proof of Theorem 4. Since \( \gamma(t) E_0(t) \) has also an exponential decay to zero, we deduce that for \( t \) large enough, (73), (74) and (75) are satisfied.

(b) We now come to the precise integral estimates of the velocities and gradient terms. Parallel to the study for (TRISHE) we proceed with the Lyapunov function \( E_p \) defined in (56). The system (TRISH) corresponds to \( p = 0 \), so we consider

\[
\dot{E}_0(t) := \left( \varphi_t(x(t)) - \varphi_t(x_{\varepsilon(t)}) \right) + \frac{1}{2} \|\dot{x}(t) + \beta \nabla f(x(t))\|^2.
\]

Using successively the derivation chain rule, the equation (TRISH), and \( \dot{\varepsilon}(t) \leq 0 \), we get

\[
\frac{d}{dt} E_0(t) = (\nabla \varphi_t(x(t)), \dot{x}(t)) + \frac{\dot{\varepsilon}(t)}{2} \|x(t)\|^2 + \|\dot{x}(t)\|^2 + \beta \nabla f(x(t)) \dot{x}(t)
\]
\[- \beta (\nabla \varphi_t(x(t)), \nabla f(x(t))) - \delta \sqrt{\epsilon(t)} \| \dot{x}(t) \|^2 + \frac{\dot{\epsilon}(t)}{2} x(t) \|^2 \]
\[- \delta \beta \sqrt{\epsilon(t)} (\nabla f(x(t)), \dot{x}(t)) \]
\[\leq - \beta \| \nabla f(x(t)) \|^2 - \beta \epsilon(t) \langle x(t), \nabla f(x(t)) \rangle - \delta \sqrt{\epsilon(t)} \| \dot{x}(t) \|^2 \]
\[- \delta \beta \sqrt{\epsilon(t)} (\nabla f(x(t)), \dot{x}(t)) \]
\[\leq - \frac{\beta}{2} \| \nabla f(x(t)) \|^2 + \frac{\beta \epsilon^2(t)}{2} \| x(t) \|^2 - \delta \sqrt{\epsilon(t)} \| \dot{x}(t) \|^2 \]
\[- \delta \beta \sqrt{\epsilon(t)} (\nabla f(x(t)), \dot{x}(t)) \]

Equivalently,
\[\frac{\beta}{2} \| \nabla f(x(t)) \|^2 + \delta \sqrt{\epsilon(t)} \| \dot{x}(t) \|^2 \leq - \frac{d}{dt} \mathcal{E}_0(t) + \frac{\beta \epsilon^2(t)}{2} \| x(t) \|^2 \]
\[- \delta \beta \sqrt{\epsilon(t)} (\nabla f(x(t)), \dot{x}(t)) \].

By multiplying this last equality by $t^{\frac{3r}{2}-1}$ and integrating on $[t_1, T]$, we get
\[\frac{\beta}{2} \int_{t_1}^{T} t^{\frac{3r}{2}-1} \| \nabla f(x(t)) \|^2 dt + \delta \int_{t_1}^{T} t^{r-1} \| \dot{x}(t) \|^2 dt \]
\[\leq - \int_{t_1}^{T} t^{\frac{3r}{2}-1} \frac{d}{dt} \mathcal{E}_0(t) dt + \beta \int_{t_1}^{T} t^{-\frac{r}{2}-1} \| x(t) \|^2 dt \]
\[- \delta \beta \int_{t_1}^{T} t^{r-1} (\nabla f(x(t)), \dot{x}(t)) dt. \quad (80)\]

Let us show that the right-hand side of inequality (80) converges when $T$ goes to infinity.

- For the first term on the right-hand side of inequality (80), we have
\[- \int_{t_1}^{T} t^{\frac{3r}{2}-1} \frac{d}{dt} \mathcal{E}_0(t) dt = - \left[ t^{\frac{3r}{2}-1} \mathcal{E}_0(t) \right]_{t_1}^{T} + \left( \frac{3r}{2} - 1 \right) \int_{t_1}^{T} t^{\frac{3r}{2}-2} \mathcal{E}_0(t) dt \]
\[\leq t^{\frac{3r}{2}-1} \mathcal{E}_0(t_1) + \left( \frac{3r}{2} - 1 \right) \int_{t_1}^{T} t^{\frac{3r}{2}-2} \left( \varphi_t(x(t)) - f(x^*) \right) dt \]
\[+ \left( \frac{3r-2}{4} \right) \int_{t_1}^{T} t^{\frac{3r}{2}-2} \| \dot{x}(t) + \beta \nabla f(x(t)) \|^2 dt \]
\[\leq C_1 + \left( \frac{3r}{2} - 1 \right) \int_{t_1}^{T} t^{\frac{3r}{2}-2} \left( f(x(t)) - f(x^*) \right) dt \]
\[+ \left( \frac{3r-2}{4} \right) \int_{t_1}^{T} t^{\frac{3r}{2}-2} \| x(t) \|^2 dt \]
\[ + \left( \frac{3r - 2}{4} \right) \int_{t_1}^{T} t^{\frac{3}{2} - 2} \| \dot{x}(t) \| \beta \nabla f(x(t)) \|^2 dt. \] (81)

Using (75) and (73), and the fact that \( x(\cdot) \) is bounded, we obtain

\[-\int_{t_1}^{+\infty} t^{\frac{3}{2} - 1} \frac{d}{dt} E_0(t) dt \leq C_1 + C_2 \int_{t_1}^{+\infty} t^{\frac{3}{2} - 2} dt + C_3 \int_{t_1}^{+\infty} t^{r - 3} dt < +\infty, \]

because \( r < 2 \).

(82)

\( \bullet \) Consider now the third term on the right-hand side of inequality (80). According to the equality

\[-\delta \beta \int_{t_1}^{T} t^{r - 1} \langle \nabla f(x(t)), \dot{x}(t) \rangle dt = -\delta \beta \left[ t^{r - 1} \left( f(x(t)) - f(x^*) \right) \right]_{t_1}^{T} \]

\[+\delta \beta (r - 1) \beta \int_{t_1}^{T} t^{r - 2} \left( f(x(t)) - f(x^*) \right) dt \]

\[ \leq \delta \beta t_1^{r - 1} \left( f(x(t_1)) - f(x^*) \right) + \delta \beta (r - 1) \beta \int_{t_1}^{T} t^{r - 2} \left( f(x(t)) - f(x^*) \right) dt. \]

Using (73), we get

\[-\delta \beta \int_{t_1}^{+\infty} t^{r - 1} \langle \nabla f(x(t)), \dot{x}(t) \rangle dt \leq C_4 + C_5 \int_{t_1}^{+\infty} t^{r - 2} dt < +\infty. \] (83)

Collecting (80), (82) and (83), we conclude that

\[ \frac{\beta}{2} \int_{t_1}^{+\infty} t^{\frac{3r - 1}{2}} \| \nabla f(x(t)) \|^2 dt + \delta \int_{t_1}^{+\infty} t^{r - 1} \| \dot{x}(t) \|^2 dt \leq C + \beta \int_{t_1}^{+\infty} t^{\frac{r}{2} - 1} \| x(t) \|^2 dt. \]

Using again that \( x \) is bounded, we deduce that

\[ \frac{\beta}{2} \int_{t_1}^{+\infty} t^{\frac{3r - 1}{2}} \| \nabla f(x(t)) \|^2 dt + \delta \int_{t_1}^{+\infty} t^{r - 1} \| \dot{x}(t) \|^2 dt < +\infty. \]

This completes the proof. \( \square \)

4 Related First-Order Algorithms

In this section, we consider the associated first-order algorithms obtained by temporal discretization of the continuous dynamics (TRISH). This provides a new

---

3 One could as well consider (TRISHE). Not to make the paper too long, we just consider (TRISH).
family of first-order methods to minimize $f$ with favorable convergence properties. At first glance, the presence of the Hessian may seem to lead to numerical difficulties. However, this is not the case because the Hessian comes in the above ODE as $\nabla^2 f(x(t)) \dot{x}(t)$, which is nothing other than the derivative with respect to time of $\nabla f(x(t))$. This explains why the time discretization of these dynamics provides first-order algorithms (see [9, algorithm (IGAHD)] in the case without the Tikhonov regularization). We show that the corresponding proximal algorithms enjoy convergence properties similar to those of continuous dynamics. Following [9], using a renorming technique, we will show in Sect. 4.1.4 that proximal algorithms can efficiently solve structured smooth + non-smooth minimization problems of Lasso type. The general study of gradient algorithms is briefly introduced. This is a subject that requires more in-depth study, and which requires longer developments.

### 4.1 Proximal Algorithms

To get as close as possible to the continuous dynamics (TRISH), we consider the following implicit discretization in time with step $h > 0$. Setting $s = h^2$, for all $k \geq 1$,

\begin{align}
\frac{x_{k+1} - 2x_k + x_{k-1}}{s} + \frac{\delta \varepsilon_k}{\sqrt{s}} (x_{k+1} - x_k) + \frac{\beta}{\sqrt{s}} (\nabla f(x_{k+1}) - \nabla f(x_k)) \\
+ \nabla f(x_{k+1}) + \varepsilon_k x_{k+1} = 0,
\end{align}

(84)

where, according to the above comments, we take $\frac{1}{\sqrt{s}} (\nabla f(x_{k+1}) - \nabla f(x_k))$ as a discretization of $\nabla^2 f(x(t)) \dot{x}(t)$. After multiplication by $s$, (84) writes equivalently as

\begin{align}
(x_{k+1} - x_k) - (x_k - x_{k-1}) + \delta \sqrt{s} \varepsilon_k (x_{k+1} - x_k) + \beta \sqrt{s} (\nabla f(x_{k+1}) - \nabla f(x_k)) \\
+ s \nabla f(x_{k+1}) + s \varepsilon_k x_{k+1} = 0,
\end{align}

which gives

\begin{align}
(1 + \delta \sqrt{s} \varepsilon_k + s \varepsilon_k) (x_{k+1} - x_k) + (s + \beta \sqrt{s}) \nabla f(x_{k+1}) = x_k - x_{k-1} + \beta \sqrt{s} \nabla f(x_k) - s \varepsilon_k x_k.
\end{align}

(85)

Solving (85) with respect to $x_{k+1}$ gives the following inertial proximal algorithm:

| (IPATH): Inertial Proximal Algorithm with Tikhonov regularization and Hessian damping |
| --- |
| **Step** $k$ : Set $\delta_k := \frac{1}{1 + \delta \sqrt{s} \varepsilon_k + s \varepsilon_k}$ and $\lambda_k := \frac{s + \beta \sqrt{s}}{1 + \delta \sqrt{s} \varepsilon_k + s \varepsilon_k}$. |
| (IPATH) $\left\{ \begin{array}{l}
y_k = x_k + \delta_k (x_k - x_{k-1} + \sqrt{s} \beta \nabla f(x_k)) \\
x_{k+1} = \text{prox}_{\lambda_k f} (y_k - s \varepsilon_k \delta_k x_k). \\
\end{array} \right.$ |
(IPA TH) involves both the gradient and the proximal operator attached to $f$. We will then extend our study to the case of a nonsmooth convex function $f$, and thus design a "pure" proximal inertial algorithm. Note that when $\varepsilon_k = 0$, i.e. in the case without Tikhonov regularization, (IPA TH) has a structure similar to that of the inertial proximal algorithm with Hessian driven damping (IPAHD) studied in [9]. In a parallel way to the continuous case, we make the following standing assumption:

\begin{align*}
(A)_{\text{algo}} \quad & f : \mathcal{H} \to \mathbb{R} \text{ is convex and differentiable, } \nabla f \text{ is Lipschitz continuous on bounded sets; } \\
& S := \text{argmin}_f \neq \emptyset. \text{ We denote by } x^* \text{ the element of minimum norm of } S; \\
& (\varepsilon_k) \text{ is a nonincreasing positive sequence such that } \lim_{k \to \infty} \varepsilon_k = 0.
\end{align*}

### 4.1.1 Preparatory Results for Lyapunov Analysis

Lyapunov’s analysis developed below, closely follows the study conducted in the continuous case, which turns out to be a valuable guide. For each $k \geq k_0$, let us define the function $\varphi_k : \mathcal{H} \to \mathbb{R}$ by

$$
\varphi_k(x) := f(x) + \frac{\varepsilon_k}{2} \| x \|^2, \quad (86)
$$

and set

$$
x_{\varepsilon_k} := \text{argmin}_{\mathcal{H}} \varphi_k,
$$

which is the unique minimizer of the strongly convex function $\varphi_k$. The first order optimality condition gives

$$
\nabla f(x_{\varepsilon_k}) + \varepsilon_k x_{\varepsilon_k} = 0. \quad (87)
$$

Equivalently, according to the definition of the proximal mapping

$$
x_{\varepsilon_k} = \text{prox}_{(1/\varepsilon_k)f}(0). \quad (88)
$$

The sequence of iterates $(x_{\varepsilon_k})_k$ is called the (discrete) viscosity curve. It satisfies the following properties which are direct consequences of the classical properties of the Tikhonov regularization

$$
\| x_{\varepsilon_k} \| \leq \| x^* \| \text{ for all } k \geq k_0 \quad (89)
$$

$$
\lim_{k \to +\infty} \| x_{\varepsilon_k} - x^* \| = 0 \text{ where } x^* = \text{proj}_{\text{argmin}_f} 0. \quad (90)
$$

Let us introduce the sequence $(\mathcal{E}_k)$ of nonnegative real numbers that plays a key role in our Lyapunov analysis. It is defined by

$$
\mathcal{E}_k := s \left( \varphi_k(x_k) - \varphi_k(x_{\varepsilon_k}) \right) + \frac{1}{2} \| v_k \|^2, \quad (91)
$$
where \( \varphi_k \) has been defined in (86), and

\[
v_k := \tau_k \left( x_k - x_{\varepsilon_k-1} \right) + \left( x_k - x_{k-1} \right) + \sqrt{s} \beta \nabla f(x_k).
\]  

(92)

The parameter \( \tau_k \) will be adjusted later. The following lemma shows the direct link between the asymptotic behavior of the sequence (\( E_k \)) and the convergence rate of the values and of the iterates.

**Lemma 3** Let \( (x_k) \) be the sequence generated by the algorithm (IPATH), and \( (E_k) \) be the energy sequence defined in (91). Then, the following inequalities are satisfied: for any \( k \ge k_0 \),

\[ f(x_k) - \min_{\mathcal{H}} f \le \frac{1}{s} E_k + \frac{\varepsilon_k}{2} \| x^* \|^2; \]  

(93)

\[ \| x_k - x_{\varepsilon_k} \|^2 \le \frac{2}{s \varepsilon_k} E_k. \]  

(94)

Therefore, \( x_k \) converges strongly to \( x^* \) as soon as \( \lim_{t \to +\infty} \frac{1}{\varepsilon_k} E_k = 0 \).

**Proof** (i) According to the definition of \( \varphi_k \), we have

\[
f(x_k) - \min_{\mathcal{H}} f = \varphi_k(x_k) - \varphi_k(x^*) + \frac{\varepsilon_k}{2} \left( \| x^* \|^2 - \| x_k \|^2 \right)
\]

\[= \left[ \varphi_k(x_k) - \varphi_k(x_{\varepsilon_k}) \right] + \left[ \varphi_k(x_{\varepsilon_k}) - \varphi_k(x^*) \right] + \frac{\varepsilon_k}{2} \left( \| x^* \|^2 - \| x_k \|^2 \right) \]

\[\le \varphi_k(x_k) - \varphi_k(x_{\varepsilon_k}) + \frac{\varepsilon_k}{2} \| x^* \|^2. \]

By definition of \( E_k \) we have

\[
\varphi_k(x_k) - \varphi_k(x_{\varepsilon_k}) \le \frac{1}{s} E_k
\]

(95)

which, combined with the above inequality, gives (93).

(ii) According to the strong convexity of \( \varphi_k \), and \( x_{\varepsilon_k} := \arg\min_{\mathcal{H}} \varphi_k \), we have

\[ \varphi_k(x_k) - \varphi_k(x_{\varepsilon_k}) \ge \frac{\varepsilon_k}{2} \| x_k - x_{\varepsilon_k} \|^2. \]

By combining the inequality above with (91), we get

\[ E_k \ge \frac{s \varepsilon_k}{2} \| x_k - x_{\varepsilon_k} \|^2, \]

which gives (94).
By assumption \((\mathcal{A})_{algo}\), we have \(\lim_{k \to \infty} \epsilon_k = 0\). According to (90), we have that the sequence \((x_{\epsilon_k})_k\) converges strongly to \(x^*\). The conclusion is a direct consequence of inequality (94).

\section*{Lemma 4}
For every \(k \geq k_0\), the following relations are satisfied

\begin{enumerate}[(i)]
\item \(\varphi_k(x_{\epsilon_k}) - \varphi_{k+1}(x_{\epsilon_{k+1}}) \leq \frac{1}{2} (\epsilon_k - \epsilon_{k+1}) \|x_{\epsilon_{k+1}}\|^2\);
\item \(\|x_{\epsilon_{k+1}} - x_{\epsilon_k}\|^2 \leq -\frac{\epsilon_k - \epsilon_{k+1}}{\epsilon_k} \langle x_{\epsilon_{k+1}}, x_{\epsilon_k} - x_{\epsilon_k} \rangle\), and therefore,
\end{enumerate}

\[\|x_{\epsilon_{k+1}} - x_{\epsilon_k}\| \leq -\frac{\epsilon_{k+1} - \epsilon_k}{\epsilon_k} \|x_{\epsilon_{k+1}}\|\].

\section*{Proof}

\textbf{i}) By definition of \(\varphi_k\) and \(x_{\epsilon_k}\) we have

\[\varphi_k(x_{\epsilon_k}) = \inf_{\xi \in \mathcal{H}} \{ f(\xi) + \frac{\epsilon_k}{2} \|\xi\|^2 \} \leq f(x_{\epsilon_{k+1}}) + \frac{\epsilon_k}{2} \|x_{\epsilon_{k+1}}\|^2\]

and

\[\varphi_{k+1}(x_{\epsilon_{k+1}}) = f(x_{\epsilon_{k+1}}) + \frac{\epsilon_{k+1}}{2} \|x_{\epsilon_{k+1}}\|^2\].

Therefore

\[\varphi_k(x_{\epsilon_k}) - \varphi_{k+1}(x_{\epsilon_{k+1}}) \leq f(x_{\epsilon_{k+1}}) + \frac{\epsilon_k}{2} \|x_{\epsilon_{k+1}}\|^2 - f(x_{\epsilon_{k+1}}) - \frac{\epsilon_{k+1}}{2} \|x_{\epsilon_{k+1}}\|^2\]

\[= \frac{1}{2} (\epsilon_k - \epsilon_{k+1}) \|x_{\epsilon_{k+1}}\|^2\].

\textbf{ii}) We have

\[-\epsilon_k x_{\epsilon_k} = \nabla f(x_{\epsilon_k})\quad \text{and}\quad -\epsilon_{k+1} x_{\epsilon_{k+1}} = \nabla f(x_{\epsilon_{k+1}})\].

According to the monotonicity of \(\nabla f\), we have

\[\langle \epsilon_k x_{\epsilon_k} - \epsilon_{k+1} x_{\epsilon_{k+1}}, x_{\epsilon_{k+1}} - x_{\epsilon_k} \rangle = \langle \nabla f(x_{\epsilon_{k+1}}) - \nabla f(x_{\epsilon_k}), x_{\epsilon_{k+1}} - x_{\epsilon_k} \rangle \geq 0\],

which implies \(\epsilon_k \|x_{\epsilon_{k+1}} - x_{\epsilon_k}\|^2 \leq - (\epsilon_{k+1} - \epsilon_k) \langle x_{\epsilon_{k+1}}, x_{\epsilon_{k+1}} - x_{\epsilon_k} \rangle\). This completes the proof.

\section*{4.1.2 Main Result}

Let \(s > 0\), \(\beta > 0\), and \(\delta > 2\) be given parameters. To develop our Lyapunov analysis, we assume that the sequence of Tikhonov regularization parameters \((\epsilon_k)\) satisfies the following growth condition. It is naturally obtained by temporal discretization of the condition \((\mathcal{H}_0)\).
The sequence of Tikhonov regularization parameters \( \varepsilon_k \) satisfies the condition \((H)_{\text{algo}}\) if there exists \( a > 1, b > 0, \lambda \in \left[ \frac{1}{2} \delta, \frac{a}{a+1} \delta \right] \), and \( k_1 \geq k_0 \) such that for all \( k \geq k_1 \),

\[
(H)_{\text{algo}}\begin{cases}
(i) & \frac{1}{\sqrt{s \varepsilon_{k+1}}} - \frac{1}{\sqrt{s \varepsilon_k}} \leq \min \left( 2 \lambda - \delta, \frac{1}{2} \left( \delta - \frac{a+1}{a} \lambda \right) \right) \quad \text{and} \quad \delta \beta \leq \frac{1}{\sqrt{s \varepsilon_k}}; \\
(ii) & \beta \frac{\varepsilon_k}{s \varepsilon_{k+1}} \left( 1 - \sqrt{\frac{s \varepsilon_k}{s \varepsilon_{k+1}}} \right) \leq 2(1 - \frac{1}{2\lambda}) \lambda^2 - 2\delta \lambda + 1; \\
(iii) & \frac{\varepsilon_k - \varepsilon_{k+1}}{\varepsilon_k^2} \geq \beta \left( 1 - \lambda \delta + \lambda^2 \right). 
\end{cases}
\]

Notation: For shortening formulas, given any sequence \((u_k)\) in \( \mathcal{H} \), we write for \( k \geq 1 \)

\[
\dot{u}_k := u_{k+1} - u_k, \quad \ddot{u}_{k-1} := \dot{u}_k - \dot{u}_{k-1} = u_{k+1} - 2u_k + u_{k-1}.
\]

Theorem 6 Suppose \( f : \mathcal{H} \rightarrow \mathbb{R} \) is a convex \( C^1 \) function. Let \((x_k) \subset \mathcal{H}\) be a sequence generated by the algorithm (IPATH). Suppose that \( \sqrt{s} \leq \beta \) and the sequence \((\varepsilon_k)_k\) satisfies the condition \((H)_{\text{algo}}\).

Take \( \tau_k = \lambda \sqrt{s \varepsilon_k} \), and define \( \mathcal{E}_k \) by (91) and (92). Then, the following linear recurrence inequality holds:

\[
\mathcal{E}_{k+1} \leq (1 - \mu_k) \mathcal{E}_k + \frac{\theta_k}{2} \| x^* \|^2
\]

where \( \mu_k \) and \( \theta_k \) are nonnegative numbers defined by

\[
\mu_k = \sqrt{\frac{\varepsilon_k}{\varepsilon_{k+1}}} - 1 + (\delta - \lambda) \sqrt{s \varepsilon_k},
\]

\[
\theta_k = (b + 2a) \lambda \sqrt{s \varepsilon_k} \varepsilon_k \mathcal{E}_{k-1}^2 + s \left( \sqrt{\frac{\varepsilon_k}{\varepsilon_{k+1}}} + (\delta - \lambda) \sqrt{s \varepsilon_k} \right) \mathcal{E}_k
\]

\[
+ \beta s \sqrt{s \lambda} (\delta - \lambda) \epsilon_k \sqrt{\varepsilon_k} \sqrt{\varepsilon_{k+1}}.
\]

Set \( \gamma_k = \exp \left( \sum_{i=k_0}^k \mu_i \right) \). For all \( k \geq k_1 \) we have

\[
\mathcal{E}_{k+1} \leq \frac{\gamma_{k_1}}{\gamma_k} \mathcal{E}_{k_1+1} + \left( \frac{1}{\gamma_k} \sum_{j=k_1+1}^k \gamma_j \theta_j \right) \| x^* \|^2.
\]

Proof According to Lemma 4 i), we obtain for all \( k \geq k_0 \)

\[
\begin{align*}
\mathcal{E}_{k+1} - \mathcal{E}_k &= s (\varphi_{k+1} - \varphi_k) + \frac{1}{2} \| v_{k+1} \|^2 - s (\varphi_k - \varphi_k) - \frac{1}{2} \| v_k \|^2 \\
&= s (\varphi_{k+1} - \varphi_k) - s (\varphi_{k+1} - \varphi_k) + \frac{1}{2} \| v_{k+1} \|^2 - \frac{1}{2} \| v_k \|^2 \\
&\leq s (\varphi_{k+1} - \varphi_k) + \frac{\sqrt{s}}{2} (\varphi_k - \varphi_k) \| x_{k+1} \|^2 + \frac{1}{2} \| v_{k+1} \|^2 - \frac{1}{2} \| v_k \|^2 \\
&= s (\varphi_{k+1} - \varphi_k) + \frac{\sqrt{s}}{2} (\| x_{k+1} \|^2 - \| x_k \|^2) + \frac{1}{2} \| v_{k+1} \|^2 - \| v_k \|^2.
\end{align*}
\]
By convexity of $\frac{1}{2} \| \cdot \|^2$, we have

$$
\frac{1}{2} (\| v_{k+1} \|^2 - \| v_k \|^2) \leq \langle \dot{v}_k, v_{k+1} \rangle.
$$

Let us estimate the last above quantity $\langle \dot{v}_k, v_{k+1} \rangle$. According to the constitutive equation (84) of (TRISH) and the definition (92) of $v_k$, we have for every $k \geq k_0$

$$
\dot{v}_k = (\tau_{k+1} (x_{k+1} - x_{\varepsilon_k}) + \dot{x}_k + \sqrt{s} \beta \nabla f(x_{k+1})) - (\tau_k (x_k - x_{\varepsilon_{k-1}}) + \dot{x}_{k-1} + \sqrt{s} \beta \nabla f(x_k))
$$

$$
= \dot{\tau} (x_{k+1} - x_{\varepsilon_k}) + \tau_k (\dot{x}_k - \dot{x}_{\varepsilon_{k-1}}) + \dot{x}_{k-1} + \sqrt{s} \beta (\nabla f(x_{k+1}) - \nabla f(x_k))
$$

$$
= \dot{\tau} (x_{k+1} - x_{\varepsilon_k}) + (\tau_k - \sqrt{s} \varepsilon_k) \dot{x}_k - \tau_k \dot{x}_{\varepsilon_{k-1}} - s \nabla \varphi_k(x_{k+1}).
$$

Therefore,

$$
\langle \dot{v}_k, v_{k+1} \rangle = \langle \dot{\tau} (x_{k+1} - x_{\varepsilon_k}) + \tau_k (\dot{x}_k - \dot{x}_{\varepsilon_{k-1}}) - \nabla \varphi_k(x_{k+1}),
$$

$$
\tau_{k+1} (x_{k+1} - x_{\varepsilon_k}) + \dot{x}_k + \sqrt{s} \nabla f(x_{k+1}) \rangle
$$

$$
= \tau_{k+1} \dot{\tau} \| x_{k+1} - x_{\varepsilon_k} \|^2 + \langle \tau_k - \sqrt{s} \varepsilon_k, \| \dot{x}_k \|^2
$$

$$
+ \langle \dot{\tau} \tau_{k+1} (\dot{x}_k - \dot{x}_{\varepsilon_{k-1}}) (\dot{x}_k, x_{k+1} - x_{\varepsilon_k}) - \tau_k \varphi_k(x_{k+1}) (x_{k+1} - x_{\varepsilon_k})
$$

$$
- \tau \varepsilon_k \langle \dot{x}_{\varepsilon_{k-1}}, \dot{x}_k \rangle - s \varphi_k (x_{k+1} + x_{k+1} - x_{\varepsilon_k}) + \sqrt{s} \beta (\nabla \varphi_k(x_{k+1})) \rangle
$$

$$
- \beta \sqrt{s} \tau_k \langle \nabla f(x_{k+1}), \dot{x}_{\varepsilon_{k-1}} \rangle - \beta s \sqrt{s} \langle \nabla f(x_{k+1}), \nabla \varphi_k(x_{k+1}) \rangle.
$$

We have $\nabla f(x_{k+1}) = \nabla \varphi_k(x_{k+1}) - \varepsilon_k x_{k+1}$, which gives

$$
\beta \sqrt{s} \dot{\tau}_k \langle \nabla f(x_{k+1}), x_{k+1} - x_{\varepsilon_k} \rangle = \beta \sqrt{s} \dot{\tau}_k \langle \nabla \varphi_k(x_{k+1}), x_{k+1} - x_{\varepsilon_k} \rangle
$$

$$
- \beta \sqrt{s} \dot{\tau}_k \varepsilon_k \langle x_{k+1} - x_{\varepsilon_k} \rangle.
$$

Since $x_{\varepsilon_k}$ is the minimizer of $\varphi_k$, and $\varphi_k$ is strongly convex, we have

$$
0 \geq \varphi_k(x_{\varepsilon_k}) - \varphi_k(x_{k+1}) \geq -\langle \nabla \varphi_k(x_{k+1}), x_{k+1} - x_{\varepsilon_k} \rangle + \frac{\varepsilon_k}{2} \| x_{k+1} - x_{\varepsilon_k} \|^2
$$

(103)

and

$$
\varphi_k(x_k) - \varphi_k(x_{k+1}) \geq -\langle \nabla \varphi_k(x_{k+1}), x_k - x_{k+1} \rangle + \frac{\varepsilon_k}{2} \| x_{k+1} - x_k \|^2.
$$

(104)

Moreover, according to Lemma 4 ii) and inequality (89), we have for any $a > 1$,
we obtain
\[\begin{align*}
-\tau_k \langle \dot{x}_{\epsilon_{k-1}}, \dot{x}_k \rangle & \leq \frac{\alpha \tau_k}{2} \| \dot{x}_{\epsilon_{k-1}} \|^2 + \frac{\tau_k}{2a} \| \dot{x}_k \|^2 \\
-\beta \sqrt{s} \tau_k \langle \nabla f(x_{k+1}), \dot{x}_{\epsilon_{k-1}} \rangle & \leq \frac{\alpha \tau_k}{2} \| \dot{x}_{\epsilon_{k-1}} \|^2 + \frac{s \beta^2 \tau_k}{2a} \| \nabla f(x_{k+1}) \|^2 \\
& + \frac{s \beta^2 \tau_k}{2a} \| \nabla f(x_{k+1}) \|^2
\end{align*}\]

and similarly for any \( b > 0 \)
\[\begin{align*}
-\tau_k \tau_{k+1} \langle \dot{x}_{\epsilon_{k-1}}, x_{k+1} - x_{\epsilon k} \rangle & \leq \frac{b \tau_k \epsilon_{k-1}^2}{2 \epsilon_{k-1}^2} \| x^* \|^2 + \frac{\tau_k \epsilon_{k+1}^2}{2b} \| x_{k+1} - x_{\epsilon k} \|^2.
\end{align*}\]

Let us assume that \((\tau_k)\) is a nonincreasing sequence. By combining the above inequalities with (101), (102), (103), (104) and after reduction we obtain
\[\begin{align*}
\langle \dot{u}_k, v_{k+1} \rangle & \leq \left( \tau_{k+1} \dot{\epsilon}_{k+1} + \epsilon_{k+1} \right) \left( \beta \sqrt{s} \tau_k - s \tau_{k+1} \right) \| x_{k+1} - x_{\epsilon k} \|^2 \\
& + \left( \frac{\tau_{k+1}}{2a} + \tau_k - \delta \sqrt{s} \epsilon_k - \frac{s \epsilon_k}{2} \right) \| \dot{x}_k \|^2 \\
& + \left( \frac{b \tau_k \epsilon_{k-1}^2}{2 \epsilon_{k-1}^2} + \frac{a \tau_k \epsilon_{k-1}^2}{2 \epsilon_{k-1}^2} \right) \| x^* \|^2 + \frac{s \beta^2 \tau_k}{2a} \| \nabla f(x_{k+1}) \|^2 \\
& + \left( \beta \sqrt{s} \tau_k - \delta \sqrt{s} \epsilon_k \right) \langle \nabla f(x_{k+1}), \dot{x}_k \rangle - \beta s \sqrt{s} \langle \nabla f(x_{k+1}), \nabla \varphi_k(x_{k+1}) \rangle \\
& - \beta \sqrt{s} \tau_k \epsilon_k \langle x_{k+1}, x_{k+1} - x_{\epsilon k} \rangle + \langle \dot{\epsilon}_k, \tau_{k+1} (\tau_k - \delta \sqrt{s} \epsilon_k) \rangle \langle \dot{x}_k, x_{k+1} - x_{\epsilon k} \rangle.
\end{align*}\]  

(105)

By combining (105) with
\[\begin{align*}
\beta \sqrt{s} \left( \tau_k - \delta \sqrt{s} \epsilon_k \right) \langle \nabla f(x_{k+1}), \dot{x}_k \rangle & = \frac{1}{2} \left( \tau_k - \delta \sqrt{s} \epsilon_k \right) \left( \| \dot{x}_k + \sqrt{s} \beta \nabla f(x_{k+1}) \|^2 - \| \dot{x}_k \|^2 \right)
\end{align*}\]

and
\[\begin{align*}
-\beta s \sqrt{s} \langle \nabla f(x_{k+1}), \nabla \varphi_k(x_{k+1}) \rangle & = \frac{-\beta s}{2} \sqrt{s} \left( \| \nabla f(x_{k+1}) \|^2 + \| \nabla \varphi_k(x_{k+1}) \|^2 - \epsilon_k^2 \| x_{k+1} \|^2 \right)
\end{align*}\]

we obtain
\[\begin{align*}
\langle \dot{u}_k, v_{k+1} \rangle & \leq \left( \tau_{k+1} \dot{\epsilon}_{k+1} + \epsilon_{k+1} \right) \left( \beta \sqrt{s} \tau_k - s \tau_{k+1} \right) \| x_{k+1} - x_{\epsilon k} \|^2
\end{align*}\]
simplification the last majorization we use Lemma 4

On the other hand, given a positive sequence \((\mu_k)\) we have (in the last majorization we use Lemma 4 i))

\[
\mu_k E_{k+1} = \mu_k \left( s \langle \varphi_{k+1}(x_{k+1}) - \varphi_{k+1}(x_{\epsilon_k}) \rangle + \frac{\tau_{k+1}^2}{2} \| x_{k+1} - x_{\epsilon_k} \| ^2 + \frac{1}{2} \| \dot{x}_k \| \right) \\
+ \sqrt{s} \beta \nabla f(x_{k+1}) \| ^2 \\
+ \mu_k \tau_{k+1} (x_{k+1} - x_{\epsilon_k}) \cdot \dot{x}_k + \sqrt{s} \beta \nabla f(x_{k+1})
\]

Combined with (100), and using that \(-\frac{\epsilon}{2} \| x_{\epsilon_k} \| ^2 \leq -\frac{\epsilon}{2} \| x^* \| ^2\), this gives after simplification

\[
E_{k+1} - E_k \leq \left( \frac{\beta s \sqrt{s} \epsilon_k^2}{2} + \frac{s \dot{\epsilon}_k}{2} \right) \| x_{k+1} \| ^2 + \left( \tau_{k+1} \dot{x}_k + \frac{\epsilon_k}{2} (\beta \sqrt{s} \dot{\epsilon}_k - s \tau_{k+1}) + \frac{\tau_{k+1}^2}{2b} \right) \| x^* \| ^2 \\
+ \frac{1}{2} \left( \frac{\tau_k}{a} + \tau_k - \delta \sqrt{s} \epsilon_k - s \epsilon_k \right) \| \dot{x}_k \| ^2 + \left( \frac{b \tau_k \epsilon_k^2}{2} + \frac{a \tau_k \epsilon_k^2}{2} - \frac{s \dot{\epsilon}_k}{2} \right) \| x^* \| ^2 \\
+ \left( \frac{s \beta^2 \tau_k}{2a} - \frac{s \beta^2}{2} \left( \tau_k - \delta \sqrt{s} \epsilon_k - \frac{\beta s \sqrt{s}}{2} \right) \right) \| \nabla f(x_{k+1}) \| ^2 \\
+ \left( \beta \sqrt{s} \dot{\epsilon}_k - s \tau_{k+1} \right) \| \dot{x}_k \| ^2 + \sqrt{s} \beta \nabla f(x_{k+1}) \| ^2 - \frac{\beta s \sqrt{s}}{2} \| \nabla \varphi_k(x_{k+1}) \| ^2 \\
- \beta \sqrt{s} \epsilon_k \langle x_{k+1}, x_{k+1} - x_{\epsilon_k} \rangle + \left( \dot{x}_k + \tau_{k+1} \left( \tau_k - \delta \sqrt{s} \epsilon_k \right) \langle \dot{x}_k, x_{k+1} - x_{\epsilon_k} \rangle \right)
\]

(106)
\[ s \mu_k (\varphi_k(x_{k+1}) - \varphi_k(x_{s_k})) + \frac{s \mu_k \dot{\varepsilon}_k}{2} (\|x_{k+1}\|^2 - \|x_{s_k+1}\|^2) + \mu_k \tau_k^2 \|x_{k+1} - x_{s_k}\|^2 \\
+ \frac{\mu_k}{2} \|\dot{x}_k + \sqrt{s} \beta \nabla f(x_{k+1})\|^2 + \mu_k \tau_k (x_{k+1} - x_{s_k}, \dot{x}_k) + \mu_k \frac{s \beta^2}{2} \|\nabla \varphi_k(x_{k+1})\|^2 \\
- \sqrt{s} \beta \epsilon_k \mu_k \tau_k (x_{k+1} - x_{s_k}, x_{k+1}). \quad (107) \]

By adding (106) and (107), and by rearranging the right hand terms, we get

\[
(\varepsilon_{k+1} - \varepsilon_k) + \mu_k \varepsilon_{k+1} \leq -s(\tau_{k+1} - \mu_k - \frac{\beta}{\sqrt{s}} \hat{\varepsilon}_k) (\varphi_k(x_{k+1}) - \varphi_k(x_{s_k})) \\
+ \left( \frac{s \mu_k \dot{\varepsilon}_k}{2} + \frac{\beta s \sqrt{\varepsilon_k^2}}{2} + \frac{s \dot{\varepsilon}_k}{2} \right) \|x_{k+1}\|^2 \\
+ \left( \mu_k \tau_k^2 + \tau_k + \frac{\varepsilon_k}{2} (\beta \sqrt{s} \hat{\varepsilon}_k - s \tau_k) + \frac{s \dot{\varepsilon}_k}{2} \right) \|x_{k+1} - x_{s_k}\|^2 \\
+ \frac{1}{2} \left( \tau_k + \frac{\tau_k}{a} + \tau_k - \frac{\sqrt{s} \varepsilon_k}{s} \dot{x}_k \right) \|\dot{x}_k\|^2 + \left( \frac{b \tau_k \dot{\varepsilon}_k}{2} + \frac{\beta s \sqrt{\varepsilon_k^2}}{2} \right) \|\nabla f(x_{k+1})\|^2 \\
+ \frac{1}{2} \left( \tau_k - \frac{\sqrt{s} \varepsilon_k}{s} \mu_k \right) \|\dot{x}_k\|^2 + \sqrt{s} \beta \nabla f(x_{k+1})\|^2 + \frac{1}{2} \left( \frac{s \beta^2}{2} \mu_k - \beta s \sqrt{\varepsilon_k} \right) \|\nabla \varphi_k(x_{k+1})\|^2 \\
- \beta \sqrt{s} \varepsilon_k (\mu_k \tau_k + \frac{\varepsilon_k}{2} (x_{k+1}, x_{k+1} - x_{s_k}) \\
+ (\mu_k \tau_{k+1} + \frac{\dot{x}_k + \tau_k (x_{k+1} - x_{s_k}, x_{k+1} - x_{s_k})) \dot{x}_k, x_{k+1} - x_{s_k} \right). \quad (108) \]

Since we do not know a priori the sign of \( \langle \dot{x}_k, x_{k+1} - x_{s_k} \rangle \), we take the coefficient in front of this term equal to zero, which gives

\[ \dot{\tau}_k + \tau_{k+1} (\tau_k - \frac{\sqrt{s} \varepsilon_k}{s} \mu_k) + \mu_k \tau_{k+1} = 0. \quad (109) \]

Take \( \tau_k = \lambda \sqrt{s \varepsilon_k} \), where \( \lambda \) is a fixed positive parameter. Then, (109) can be equivalently written

\[ \mu_k = -\frac{\sqrt{\varepsilon_k}}{\sqrt{\varepsilon_{k+1}}} + (\delta - \lambda) \sqrt{s \varepsilon_k}. \]

Since \( (\varepsilon_k)_k \) is a nonincreasing sequence, and \( \lambda < \delta \), we have that \( \mu_k > 0 \). Accordingly

\[
\frac{1}{2} \left( \tau_k - \frac{\sqrt{s} \varepsilon_k}{s} \mu_k \right) \|\dot{x}_k + \sqrt{s} \beta \nabla f(x_{k+1})\|^2 \\
= -\frac{\sqrt{\varepsilon_k}}{\sqrt{\varepsilon_{k+1}}} \|\dot{x}_k + \sqrt{s} \beta \nabla f(x_{k+1})\|^2 \\
\leq -\frac{\sqrt{\varepsilon_k}}{\sqrt{\varepsilon_{k+1}}} \|\dot{x}_k\|^2 - \frac{s \beta^2}{\sqrt{s} \varepsilon_k} \|\nabla f(x_{k+1})\|^2 \quad (110) \]
and

$$-\beta \sqrt{s} e_k (\mu_k e_{k+1} + \hat{t}_k) \langle x_{k+1}, x_{k+1} - x_{e_k} \rangle$$
$$= -\beta s \sqrt{\lambda} (\delta - \lambda) e_k \sqrt{e_k} e_{k+1} \langle x_{k+1}, x_{k+1} - x_{e_k} \rangle$$
$$= -\beta s \sqrt{\lambda} (\delta - \lambda) \frac{e_k \sqrt{e_k} e_{k+1}}{2} \left[ \|x_{k+1}\|^2 + \|x_{k+1} - x_{e_k}\|^2 - \|x_{e_k}\|^2 \right].$$

(111)

Combining (108) with (110) and (111), we obtain

$$(e_{k+1} - e_k) + \mu_k e_{k+1} \leq -s \left( \lambda \sqrt{e_{k+1}} + \frac{\sqrt{e_k}}{\sqrt{e_{k+1}}} - (\delta - \lambda) \sqrt{e_k} - \beta \lambda \sqrt{e_k} \right) (\varphi_k(x_{k+1}) - \varphi_k(x_e))$$
$$+ \left( s \mu_k e_k \right) + \frac{\beta s \sqrt{e_k} (\delta - \lambda) e_k \sqrt{e_k} e_{k+1}}{2} \|x_{k+1}\|^2$$
$$+ \left[ s \sqrt{\lambda} e_k e_{k+1} \sqrt{e_k} \left( \lambda (\delta - \lambda) + \frac{\lambda^2}{2b} - s \sqrt{\lambda} \epsilon k \sqrt{e_{k+1}} + \frac{\lambda \beta e_k}{2} - \frac{\beta s \sqrt{\lambda} (\delta - \lambda)}{2} \epsilon_k \sqrt{e_{k+1}} \right) \right] \|x_{k+1} - x_{e_k}\|^2$$
$$+ \frac{1}{2} \left( -2 \frac{\sqrt{e_k}}{\sqrt{e_{k+1}}} + \frac{\sqrt{e_k}}{\sqrt{e_{k+1}}} - (\delta - \lambda) \sqrt{e_k} - \beta \lambda \sqrt{e_k} \right) \|\varphi_k(x_{k+1})\|^2$$
$$+ \frac{b \epsilon k}{2} \left( 1 + \frac{1}{a} \right) \lambda - \delta) \sqrt{e_k} - \beta \lambda \sqrt{e_k} \right) \|\nabla \varphi_k(x_{k+1})\|^2$$
$$+ \frac{\beta s \sqrt{\lambda} (\delta - \lambda) e_k \sqrt{e_k} e_{k+1}}{2} \left( 1 + \frac{1}{a} \right) \lambda - \delta) \sqrt{e_k} - \beta \lambda \sqrt{e_k} \right) \|\nabla f(x_{k+1})\|^2.$$}

According to (89), by putting together the corresponding terms in (112), we obtain the following inequality whose coefficients $A_k$, $B_k$, $C_k$, $D_k$, $E_k$ and $F_k$ are defined and estimated below.

$$(e_{k+1} - e_k) + \mu_k e_{k+1} \leq A_k (\varphi_k(x_{k+1}) - \varphi_k(x_{e_k})) + B_k \|\dot{x}_k\|^2 + C_k \|x_{k+1} - x_{e_k}\|^2$$
$$+ D_k \|x_{k+1}\|^2 + E_k \|\nabla f(x_{k+1})\|^2 + F_k \|\nabla \varphi_k(x_{k+1})\|^2$$
$$+ \beta s \sqrt{\lambda} (\delta - \lambda) \frac{e_k \sqrt{e_k} e_{k+1}}{2} \left( 1 + \frac{1}{a} \right) \lambda - \delta) \sqrt{e_k} - \beta \lambda \sqrt{e_k} \right) \|\nabla f(x_{k+1})\|^2.$$}

$$A_k = -s \left( \lambda \sqrt{e_{k+1}} + \frac{\sqrt{e_k}}{\sqrt{e_{k+1}}} - (\delta - \lambda) \sqrt{e_k} - \beta \lambda \sqrt{e_k} \right)$$
$$= s \sqrt{e_k} \left( \frac{1}{\sqrt{e_{k+1}}} - \frac{1}{\sqrt{e_k}} + \delta - 2\lambda \right)$$
$$+ s \left( \lambda \sqrt{\lambda} (\sqrt{e_k} - \sqrt{e_{k+1}}) - \beta \lambda (\sqrt{e_k} - \sqrt{e_{k+1}}) \right)$$
$$= s \sqrt{e_k} \left( \frac{1}{\sqrt{e_{k+1}}} - \frac{1}{\sqrt{e_k}} + \delta - 2\lambda \right) + \lambda s \left( \sqrt{\lambda} - \beta \right) \left( \sqrt{e_k} - \sqrt{e_{k+1}} \right)$$
$$\leq 0 \quad \geq 0$$
\[ \begin{aligned}
&\leq s \sqrt{s \epsilon_k} \left( \frac{1}{\sqrt{s \epsilon_k+1}} - \frac{1}{\sqrt{s \epsilon_k}} + \delta - 2\lambda \right); \\
B_k &= \frac{1}{2} \left( - \frac{2\sqrt{s \epsilon_k}}{\sqrt{s \epsilon_k+1}} + \left( 1 + \frac{1}{a} \right) \lambda - \delta \right) \sqrt{s \epsilon_k} \leq s \sqrt{s \epsilon_k} \left( \frac{1}{\sqrt{s \epsilon_k+1}} - \frac{1}{\sqrt{s \epsilon_k}} \right) + \frac{1}{2} \left( (1 + \frac{1}{a}) \lambda - \delta \right); \\
C_k &= \left[ s \sqrt{s \lambda \epsilon_k+1} \sqrt{s \epsilon_k} \left( \lambda (\delta - \lambda) + \frac{\lambda^2}{2b} \right) - s \sqrt{s \lambda \epsilon_k+1} \sqrt{\epsilon_k+1} + \frac{\lambda \beta s \epsilon_k}{\sqrt{s \epsilon_k+1}} \right] \\
&\leq s \sqrt{s \lambda \epsilon_k+1} \sqrt{s \epsilon_k} \left( \lambda (\delta - \lambda) + \frac{\lambda^2}{2b} \right) - s \sqrt{s \lambda \epsilon_k+1} \sqrt{\epsilon_k+1} + \frac{\lambda \beta s \epsilon_k}{\sqrt{s \epsilon_k+1}} \lambda \epsilon_k (\text{because } \epsilon_k \text{ is nonincreasing}) \\
D_k &= \left( \frac{\lambda \beta s \epsilon_k}{\sqrt{s \epsilon_k+1}} \sqrt{s \epsilon_k} \right) \left( \lambda (\delta - \lambda) \right) + \frac{\lambda \beta s \epsilon_k}{\sqrt{s \epsilon_k+1}} \sqrt{s \epsilon_k+1} + \frac{s \epsilon_k}{2} - \beta s \epsilon_k \lambda (\delta - \lambda) \sqrt{s \epsilon_k} \sqrt{s \epsilon_k+1} \\
&\leq \frac{\lambda \beta s \epsilon_k}{\sqrt{s \epsilon_k+1}} \sqrt{s \epsilon_k} \left( \lambda (\delta - \lambda) \right) + \frac{\lambda \beta s \epsilon_k}{\sqrt{s \epsilon_k+1}} \sqrt{s \epsilon_k+1} + \frac{s \epsilon_k}{2} - \beta s \epsilon_k \lambda (\delta - \lambda) \sqrt{s \epsilon_k} \sqrt{s \epsilon_k+1} \\
E_k &= s \beta^2 \left( - \frac{s \epsilon_k}{\sqrt{s \epsilon_k+1}} + \frac{1}{2} \left( (1 + \frac{1}{a}) \lambda - \delta \right) \sqrt{s \epsilon_k} - (\lambda - \delta) \sqrt{s \epsilon_k} - \frac{\sqrt{s}}{2\beta} \right)
\end{aligned}\]
\[
\begin{align*}
&= s\beta^2 \sqrt{s\varepsilon_k} \left( \frac{1}{\sqrt{s\varepsilon_{k+1}}} - \frac{1}{\sqrt{s\varepsilon_k}} \right) + \frac{1}{2} \left( 1 + \frac{a}{\lambda} \delta - \frac{1}{\lambda} \right) + s\sqrt{s\beta} \left( \beta (\delta - 2\lambda) \sqrt{s\varepsilon_k} \right) + \frac{s\sqrt{s\beta}}{2} (\beta \delta \sqrt{s\varepsilon_k} - 1) \\
&= s\beta^2 \sqrt{s\varepsilon_k} \left( \frac{1}{\sqrt{s\varepsilon_{k+1}}} - \frac{1}{\sqrt{s\varepsilon_k}} + \frac{1}{2} \left( 1 + \frac{a}{\lambda} \delta - \frac{1}{\lambda} \right) \right) + \frac{s\sqrt{s\beta}}{2} (\beta \delta \sqrt{s\varepsilon_k} - 1); \\
F_k &= \frac{s\beta^2}{2} \left( -\sqrt{s\varepsilon_k} + (\delta - \lambda) \sqrt{s\varepsilon_k} - \sqrt{s\varepsilon_{k+1}} \right) \\
&\leq \frac{s\beta^2}{2} \left( \frac{1}{\sqrt{s\varepsilon_{k+1}}} - \frac{1}{\sqrt{s\varepsilon_k}} + (\delta - 2\lambda) \right) + \frac{s\beta^2}{2} (\delta \sqrt{s\varepsilon_k} - \frac{1}{\beta}).
\end{align*}
\]

We now use the condition \((\mathcal{H})_{algo} (i)\), recalled below for the convenience of the reader. There exists \(a > 1, b > 0, \lambda \in \left[ \frac{1}{2} \delta, \frac{a}{a+1} \delta \right] \), and \(k_1 \geq k_0\) such that for all \(k \geq k_1\),

\[
\frac{1}{\sqrt{s\varepsilon_{k+1}}} - \frac{1}{\sqrt{s\varepsilon_k}} \leq \min \left( 2\lambda - \delta, \delta - \frac{a+1}{2a} \lambda \right) \quad \text{and} \quad \delta \beta \leq \frac{1}{\sqrt{s\varepsilon_k}}.
\]

It is immediate to verify that under this condition the quantities \(A_k, B_k, \varepsilon_k\) and \(F_k\) are less than or equal to zero. Let us now examine the sign of \(C_k\) and \(D_k\).

**For** \(C_k\). We have \((2\delta \lambda - 2(1 - \frac{1}{2\beta}) \lambda^2 - 1) + \beta \sqrt{s\varepsilon_k} (1 - \sqrt{s\varepsilon_{k+1}}) \leq 0\), because condition \((\mathcal{H})_{algo} (ii)\). Thus \(C_k \leq 0\).

**For** \(D_k\). We have \((1 - \lambda \delta + \lambda^2) + \frac{\dot{\varepsilon}_k}{\beta \varepsilon_k^2} \leq 0\), because of the condition \((\mathcal{H})_{algo} (iii)\). Thus \(D_k \leq 0\).

We conclude that

\[
(\varepsilon_{k+1} - \varepsilon_k) + \mu_k \varepsilon_k \leq \frac{\theta_k}{2} \|x^*\|^2,
\]

where \(\mu_k\) and \(\theta_k\) are defined in (97) and (98).

Multiplying \(113\) by \(\gamma_k = \exp \left( \sum_{i=k_0}^{k} \mu_i \right)\) and rearranging the terms, we obtain

\[
(\gamma_k \varepsilon_{k+1} - \gamma_k \varepsilon_k) - (\gamma_k - \gamma_{k-1} - \gamma_k \mu_k) \varepsilon_k \leq \frac{\gamma_k \theta_k}{2} \|x^*\|^2.
\]

Let us analyze the sign of the factor of \(\varepsilon_k\) in\(114\). We have

\[
-(\gamma_k - \gamma_{k-1} - \gamma_k \mu_k) = \gamma_k \left( \exp(-\mu_k) - 1 + \mu_k \right).
\]

Elementary analysis gives that for all \(x \geq 0\)

\[
\exp(-x) - 1 + x \geq 0.
\]
Therefore $-(\gamma_k - \gamma_{k-1} - \gamma_k \mu_k)$ is nonnegative. Returning to (114), we obtain that for $k \geq k_1$

$$\gamma_k \mathcal{E}_{k+1} - \gamma_{k-1} \mathcal{E}_k \leq \frac{\gamma_k \theta_k}{2} \|x^*\|^2.$$ 

By summing the above inequalities between $k_1 + 1$ and $k > k_1$, we end up with

$$\gamma_k \mathcal{E}_{k+1} - \gamma_{k_1} \mathcal{E}_{k_1+1} \leq \frac{1}{2} \left( \sum_{j=k_1+1}^{k} \gamma_j \theta_j \right) \|x^*\|^2,$$

and consequently (99) is satisfied, i.e.,

$$\mathcal{E}_{k+1} \leq \frac{\gamma_{k_1}}{\gamma_k} \mathcal{E}_{k_1+1} + \frac{1}{2 \gamma_k} \left( \sum_{j=k_1+1}^{k} \gamma_j \theta_j \right) \|x^*\|^2.$$ 

This completes the proof of the Lyapunov analysis. 

\[ \square \]

**Remark 4** Note the parallelism between the formulas relating to $\mu(t)$ and $G(t)$ in the continuous case, and the corresponding quantities $\mu_k$ and $\theta_k$ in the algorithmic case. This suggests that the study of the particular case $\epsilon_k = 1/k^r$ can be developed in the same way.

### 4.1.3 Non-smooth Case

Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous and convex function. We rely on the basic properties of the Moreau-Yosida regularization $f_\theta$ defined in (23). We recall that $f_\theta$ is a convex function, whose gradient is $\theta^{-1}$-Lipschitz continuous, such that argmin $f_\theta = \argmin f$. The interested reader may refer to [25, 27] for a comprehensive treatment of the Moreau envelope in a Hilbert setting. Since the set of minimizers is preserved by taking the Moreau envelope, the idea is to replace $f$ by $f_\theta$ in the previous algorithm, and take advantage of the fact that $f_\theta$ is continuously differentiable. The (TRISH) dynamic attached to $f_\theta$ becomes

$$\ddot{x}(t) + \delta \sqrt{\epsilon(t)} \dot{x}(t) + \beta \nabla^2 f_\theta(x(t)) \dot{x}(t) + \nabla f_\theta(x(t)) + \epsilon(t) x(t) = 0.$$ 

However, we do not really need to work on this system (which requires $f_\theta$ to be $C^2$), but with the discretized form which only requires the function to be continuously differentiable, as is the case of $f_\theta$. Then, algorithm (IPAHD) now reads

$$\begin{cases} y_k = x_k + \delta_k \left( x_k - x_{k-1} + \sqrt{s} \beta \nabla f_\theta(x_k) \right) \\ x_{k+1} = \text{prox}_{\lambda_k f_\theta}(y_k - s \epsilon_k \delta_k x_k) \end{cases}.$$
Thus, we just need to formulate the convergence rates obtained in the previous section in terms of \( f \) and its proximal mapping. This follows from the following formulae from proximal calculus [25]:

- \( f_\theta(x) = f(\operatorname{prox}_\theta f(x)) + \frac{1}{2\theta} \| x - \operatorname{prox}_\theta f(x) \|^2, \quad \nabla f_\theta(x) = \frac{1}{\theta} (x - \operatorname{prox}_\theta f(x)). \)
- \( \operatorname{prox}_{\lambda_k f_\theta}(x) = x + \frac{1}{\lambda_k} \operatorname{prox}_{\lambda_k f}(x) - x. \)

We obtain the following relaxed inertial proximal algorithm (NS stands for Non-Smooth):

\[
\begin{align*}
(IPATH) \text{-nonsmooth.} \\
\text{Step } k : & \quad \delta_k := \frac{1}{1 + \delta_\theta \sqrt{s \epsilon_k} + s \epsilon_k} \quad \text{and} \quad \lambda_k := \frac{s + \beta \sqrt{s}}{1 + \delta_\theta \sqrt{s \epsilon_k} + s \epsilon_k}.
\end{align*}
\]

\[
(IPATH - NS) \begin{cases} 
z_k = x_k + \delta_k \left( x_k - x_{k-1} + \sqrt{s \beta} \frac{1}{\theta} (x_k - \operatorname{prox}_{\theta f}(x_k)) \right) - s \epsilon_k \delta_k x_k \\
x_{k+1} = \frac{\theta}{\theta + \lambda_k} z_k + \frac{\lambda_k}{\theta + \lambda_k} \operatorname{prox}_{(\theta + \lambda_k)f}(z_k).
\end{cases}
\]

The above algorithm now only involves proximal operations on \( f \), which makes it applicable to non-smooth optimization problems. We just give an illustration below. Its full development is a subject for further study.

### 4.1.4 Application to Lasso Type Structured Minimization

In many situations, the minimization problem has an additive composite structure \( \min_H (h + g) \), with \( h \) smooth and \( g \) non-smooth. Accelerated proximal-gradient algorithms are effective splitting methods to treat these problems. We will show how to adapt the (IPATH-NS) algorithm to such composite setting, in the case of the Lasso-type problems.

Take \( H = \mathbb{R}^n \) equipped with the usual Euclidean structure. Suppose that the function \( f : H \rightarrow \mathbb{R} \cup \{+\infty\} \) to be minimized has the additive structure

\[
f(x) = \frac{1}{2} \| Ax - b \|_2^2 + g(x), \quad (115)
\]

where \( A \in \mathbb{R}^{m \times n} \) (with \( m \leq n \)), \( b \in \mathbb{R}^m \) and \( g \in \mathcal{C}_0(\mathbb{R}^n) \) (set of closed proper and convex functions). Minimizing such function \( f \) occurs in a variety of fields ranging from inverse problems in signal/image processing, to machine learning and statistics. Typical examples of function \( g \) include the \( \ell_1 \) norm (Lasso), the \( \ell_1 - \ell_2 \) norm (group Lasso), the total variation, or the nuclear norm. In all these such situations, \( g \) is nonsmooth which also makes \( f \) nonsmooth. A direct application of the algorithm (IPATH-NS) would require calculating (at least approximately) the proximal operator of \( f \). It’s not easy in general. To work around this difficulty, we use a change of metric. This "renorming technique" was initiated by Lemaréchal and Sagastizábal in [37]...
to introduce efficient preconditioners into the proximal point algorithm for minimizing convex functions. It was then systematically developed by Esser, Zhang, and Chan [34]. It was adapted by various authors, see for example the recent contribution [9]. For a symmetric and positive definite matrix $M \in \mathbb{R}^{n \times n}$, we denote by $\langle \cdot, \cdot \rangle_M = \langle M \cdot, \cdot \rangle$ the scalar product on $\mathbb{R}^n$ induced by $M$ and by $\| \cdot \|_M$ the associated norm. For a given $f \in \mathcal{C}^0_0(\mathbb{R}^n)$, the Moreau’s envelope $f^M_\lambda$ of index $\lambda > 0$ associated with the metric induced by $M$ is defined by: for $x \in \mathbb{R}^n$

$$f^M_\lambda (x) = \min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\lambda} \|x - y\|^2_M \right\}. \quad (116)$$

Let us denote by $\text{prox}^M_\lambda f (x)$ the unique minimizer in (116), which is the proximal point of $x$, of index $\lambda$, for the metric induced by $M$. The first-order optimality condition for this strongly convex minimization problem gives

$$\text{prox}^M_\lambda f (x) = (M + \lambda \nabla f)^{-1}(Mx). \quad (117)$$

With the particular choice of $f$ in (115), we set

$$M := I_n - \lambda A^T A.$$ 

If $\lambda \in [0, \frac{1}{\| A \|^2_2}]$, then $M$ is positive definite. In this case,

$$\text{prox}^M_\lambda f (x) = \text{prox}_\lambda g \left( x - \lambda A^T (Ax - b) \right). \quad (118)$$

By applying (IPATH-NS) to $f$, and working in the Hilbert space $\mathbb{R}^n$ equipped with the metric $M$, we obtain a splitting algorithm applicable to (115).

### 4.2 Gradient Algorithms

Let us introduce the inertial gradient algorithms associated with the dynamical system (TRISH) that we recall below

$$(\text{TRISH}) \quad \ddot{x}(t) + \delta \dot{x}(t) + \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + \varepsilon(t) x(t) = 0.$$ 

In fact, there are several possibilities for the explicit temporal discretization of (TRISH), as well as with (TRISHE). The study of their convergence properties, and to decide which are the most interesting is a subject that remains largely to be explored. Let’s introduce two of the most promising, which follow naturally depending on whether one considers (TRISH) relevant of Polyak’s heavy ball in the strongly convex case, or of Nesterov’s method

- (TRISH) has been inspired by the heavy ball with friction method of Polyak in the strongly convex case. According to this property, and despite the fact that the strong
convexity property vanishes asymptotically, it is natural to consider the following explicit discretization of (TRISH)

\[ \frac{1}{s} (x_{k+1} - 2x_k + x_{k-1}) + \frac{\delta \sqrt{\varepsilon_k}}{\sqrt{s}} (x_{k+1} - x_k) + \frac{\beta}{\sqrt{s}} (\nabla f(x_k) - \nabla f(x_{k-1})) + \nabla f(x_k) + \varepsilon_k x_k = 0. \]

where \( h > 0 \) is a given positive step size. This gives the following algorithm

(IGATH)-1: Inertial Gradient Algorithm with Tikhonov regularization and Hessian damping.

\[
x_{k+1} = x_k + \frac{1}{1 + \delta \sqrt{s \varepsilon_k}} (x_k - x_{k-1}) - \frac{\sqrt{s}}{1 + \delta \sqrt{s \varepsilon_k}} \left[ \beta (\nabla f(x_k) - \nabla f(x_{k-1})) + \sqrt{s} (\nabla f(x_k) + \varepsilon_k x_k) \right].
\]

- (TRISH) is an inertial dynamic which involves an asymptotic vanishing damping. As such, it is naturally related to the Nesterov accelerated gradient method, as shown in [46]. We therefore follow a discretization similar to that performed in [9] in the case without the Tikhonov regularization term. We consider a similar discretization as above but now the gradient of \( f \) is evaluated at a point \( y_k \) which is chosen according to the Nesterov scheme. This can be written as:

\[ \frac{1}{s} (x_{k+1} - 2x_k + x_{k-1}) + \frac{\delta \sqrt{\varepsilon_k}}{\sqrt{s}} (x_{k+1} - x_k) + \frac{\beta}{\sqrt{s}} (\nabla f(x_k) - \nabla f(x_{k-1})) + \nabla f(y_k) + \varepsilon_k y_k = 0. \]

This gives the following algorithm

(IGATH)-2: Inertial Gradient Algorithm with Tikhonov regularization and Hessian damping.

\[
\begin{align*}
y_k &= x_k + (1 - \delta \sqrt{s \varepsilon_k}) (x_k - x_{k-1}) - \beta \sqrt{s} (\nabla f(x_k) - \nabla f(x_{k-1})) \\
x_{k+1} &= y_k - s (\nabla f(y_k) + \varepsilon_k y_k).
\end{align*}
\]

In addition to the many variants of these algorithms, one might as well consider the gradient algorithm which is obtained by replacing in (IPATH)-nonsmooth the proximal steps by corresponding gradient steps. Considering the first-order system in time and space which is equivalent to (TRISH) is also to be explored.
5 Numerical Illustrations

5.1 Continuous Case

Let us illustrate our results with the following examples where the function $f$ is taken successively strictly convex, then convex with a continuum of solutions. In a third example, we compare the two systems (TRISH) and (TRISHE). The following numerical experiences describe in these three situations the behavior of the trajectories generated by the system (TRIGS) (without the Hessian driven damping) and by the systems (TRISH) and (TRISHE) (with the Hessian driven damping). All these systems take into account the effect of Tikhonov regularization. They are differentiated by the presence, or not, of the Hessian driven damping. According to the model situation described in Theorem 4 and Theorem 5, the Tikhonov regularization parameter is taken equal to $\varepsilon(t) = t^{-r}$, with $0 < r \leq 2$. We consider different values of the parameter $r$ which plays a key role in tuning the viscosity and Tikhonov parameters. We pay particular attention to the case $r$ close to the value 2, which provides fast convergence results. The corresponding dynamical systems are given by:

\begin{align*}
(\text{TRIGS}) & \quad \ddot{x}(t) + \delta t^{-\frac{3}{2}} \dot{x}(t) + \nabla f(x(t)) + t^{-r} \dot{x}(t) = 0 \\
(\text{TRISH}) & \quad \ddot{x}(t) + \delta t^{-\frac{3}{2}} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) + t^{-r} \dot{x}(t) = 0 \\
(\text{TRISHE}) & \quad \ddot{x}(t) + \delta t^{-\frac{3}{2}} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) - rt^{-r-1} x(t) + t^{-r} \dot{x}(t) + \nabla f(x(t)) + t^{-r} \dot{x}(t) = 0.
\end{align*}

We choose $\delta = 3$, $\beta = 1$. To facilitate the comparison of the trajectories corresponding to different dynamics, for example (TRIGS) and (TRISH), they are represented respectively by continuous lines and dotted lines. 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

$$f_1(x) = (x_1 + x_2^2) - 2 \ln(x_1 + 1)(x_2 + 1).$$

The function $f_1$ is strictly convex with $\nabla f_1(x) = \begin{pmatrix} 1 - \frac{2}{x_1 + 1} \\ 2x_2 - \frac{x_1 + 1}{x_2 + 1} \end{pmatrix}$ and $\nabla^2 f_1(x) = \begin{pmatrix} \frac{2}{(x_1 + 1)^2} & 0 \\ 0 & 2 + \frac{2}{(x_2 + 1)^2} \end{pmatrix}$. The unique minimum of $f_1$ is $x^* = (1, (\sqrt{5} - 1)/2)$ and also that of $\phi_t$ is

$$x_{\phi(t)} = \frac{1}{2} \left( \sqrt{t^{2r} + 6t^r + 1} - (t^r + 1), \sqrt{1 + \frac{8t^r}{1 + 2t^r} - 1} \right).$$

The corresponding trajectories to the systems (TRIGS), (TRISHE) are depicted in Fig. 1.
Example 2 Consider the convex function $f_2 : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f_2(x) = \frac{1}{2}(x_1 + x_2 - 1)^2.$$  

We have

$$\nabla f_2(x) = \begin{bmatrix} x_1 + x_2 - 1 \\ x_1 + x_2 - 1 \end{bmatrix} \text{ and } \nabla^2 f_2(x) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

We have $S_2 = \arg\min f_2 = \{(x_1, 1 - x_1) : x_1 \in \mathbb{R}\}$ and $x^* = (\frac{1}{2}, \frac{1}{2})$ is the minimum norm solution. Also, we have the unique minimum of $\varphi_t$ is $x_{\varepsilon(t)} = \left(\frac{1}{1/(2 + r^{-t})}, \frac{1}{1/(2 + r^{-t})}\right)$. The corresponding trajectories to the systems (TRIGS), (TRISHE) are depicted in Fig. 2.

Figures 1 and 2 depict the convergence results of the previous sections. These underline the improvement of convergence of the new second order systems with Hessian-driven damping (TRISH) and (TRISHE) by adapting and greatly reducing the vibrations of the associate solution $x(t)$. We also notice that these new systems also improve the speed of convergence and especially increase the rate of the convergence of the values $f_2(x(t)) - f(x^*)$, the velocities $\|\dot{x}(t)\|_2$ and $\|\nabla f_2(x(t))\|_2$. 
5.2 Algorithmic Case

Let us introduce proximal gradient algorithms inspired by our approach. In Sect. 4.1.4 we considered the case of Lasso problems. Here we take the more general form of $f = h + g$ with $h$ differentiable and $g$ not smooth. We just give an introduction and numerical illustrations, the theoretical study is a subject for further research. More precisely, we seek to minimize on $\mathbb{R}^2$ the functions of the form $h + \|x\|_1$, where $\|\cdot\|_1$ is not smooth. We start from the two systems (TRISH) and (TRISHE):

$$0 \in \dot{x} + \delta \sqrt{\varepsilon} \dot{x} + \beta \frac{d}{dt} (\nabla h(x)) + \nabla h(x) + \partial g(x) + \varepsilon x$$

$$0 \in \dot{x} + \delta \sqrt{\varepsilon} \dot{x} + \beta \frac{d}{dt} (\nabla h(x) + \varepsilon x) + \nabla h(x) + \partial g(x) + \varepsilon x.$$

We consider their respective discretizations

$$0 \in (x_{k+1} - 2x_k + x_{k-1}) + \delta \sqrt{s} \varepsilon_k (x_{k+1} - x_k) + \sqrt{s} \beta (\nabla h(x_k) - \nabla h(x_{k-1})$$

$$+ s (\nabla h(y_k) + \partial g(x_{k+1}) + \varepsilon_k x_k),$$

$$0 \in (x_{k+1} - 2x_k + x_{k-1}) + \delta \sqrt{s} \varepsilon_k (x_{k+1} - x_k) + \sqrt{s} \beta (\nabla h(x_k) - \nabla h(x_{k-1})$$

$$+ \sqrt{s} \beta (\varepsilon_k x_k - \varepsilon_k (x_{k-1})$$

$$+ s (\nabla h(y_k) + \partial g(x_{k+1}) + \varepsilon_k x_k).$$

Choosing $y_k$ according to Nesterov’s scheme gives rise to the following forward-backward algorithms:
Fig. 3 Algorithmic case: convergence rates of values and trajectories for (IPGATH)-1 and (IPGATH)-2 applied to the nonsmooth functions: (left) $\frac{1}{2} (x_1 + x_2 - 1)^2 + \|x\|_1$ and (right) $(x_1 + x_2^2) - 2 \ln(x_1 + 1)(x_2 + 1) + \|x\|_1$

(IPGATH)-1: Inertial Proximal Gradient Algorithm with Tikhonov regularization and Hessian damping.

\[
\begin{align*}
    y_k &= x_k + \frac{\delta_k}{s} (x_k - x_{k-1}) - \frac{\beta \delta_k}{\sqrt{s}} (\nabla h(x_k) - \nabla h(x_{k-1})) \\
    x_{k+1} &= \text{prox}_{\delta_k g} (y_k - \delta_k (\nabla h(y_k) + \epsilon_k x_k))
\end{align*}
\]

and

(IPGATH)-2: Inertial Proximal Gradient Algorithm with Tikhonov regularization and Hessian damping.

\[
\begin{align*}
    y_k &= x_k + (1 - \sqrt{s} \beta \epsilon_k) \frac{\delta_k}{s} (x_k - x_{k-1}) - \frac{\beta \delta_k}{\sqrt{s}} (\nabla h(x_k) - \nabla h(x_{k-1})) \\
    x_{k+1} &= \text{prox}_{\delta_k g} (y_k - \delta_k (\nabla h(y_k) + \epsilon_k^0 x_k))
\end{align*}
\]

where $\delta_k = \frac{s}{1 + \sqrt{s} \epsilon_k}$, $\epsilon_k = \frac{1}{\epsilon_k^0}$, $\epsilon_k^0 = \epsilon_k + \frac{\beta}{\sqrt{s}} (\epsilon_k - \epsilon_{k-1})$.

Numerical tests are shown in Fig. 3. They show the good performance of the two algorithms with respect to the three criteria involved in our study, with an advantage for the second algorithm.
6 Conclusion, Perspective

For convex optimization in Hilbert spaces, we have introduced damped inertial dynamics which combine Hessian driven damping with Tikhonov regularization. These terms induce specific favorable geometric properties. The Tikhonov term regulates the objective function. It makes the dynamic relevant of the heavy ball with friction method for a strongly convex function. The Hessian driven damping has a corrective effect by damping the oscillations that arise with ill-conditioned optimization problems. It turns out that the two techniques combine well and provide a substantial improvement to Nesterov’s accelerated gradient method. While preserving fast convergence of values, they ensure fast convergence of gradients to zero, they significantly reduce oscillations, and provide convergence to the minimum norm solution. When the focus is on the solution of the minimization problem, our approach provides a solution that is independent of the initial condition.

Our Lyapunov analysis provides a solid basis for studying the algorithms obtained by temporal discretization. We have developed a detailed convergence analysis for the associated proximal algorithms, and shown that they enjoy similar favorable properties. We have also introduced gradient algorithms, whose theoretical and numerical study is the subject of further work. We have shown that our approach can naturally be extended to the case of nonsmooth convex optimization, and the study of additively structured "smooth + nonsmooth" convex optimization problems.

One of the main challenges related to our study is whether similar convergence results can be obtained using autonomous systems, see [7] for a first systematic study of this question. Indeed the study of autonomous versions of the Tikhonov method, such as the Haugazeau method, in the context of dynamic systems, and rapid optimization, is a field largely to be explored.

Acknowledgements The research of Aïcha BALHAG was supported by the EIPHI Graduate School (contract ANR-17-EURE-0002).

References

1. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. 161(2), 331–360 (2014)
2. Adly, S., Attouch, H., Vo, V.N.: Newton-type inertial algorithms for solving monotone equations governed by sums of potential and nonpotential operators. Appl. Math. Optim. 85, 44 (2022). https://doi.org/10.1007/s00245-022-09846-3
3. Alvarez, F., Cabot, A.: Asymptotic selection of viscosity equilibria of semilinear evolution equations by the introduction of a slowly vanishing term. Discrete Contin. Dyn. Syst. 15, 921–938 (2006)
4. Apidopoulos, V., Aujol, J.-F., Dossal, Ch.: The differential inclusion modeling the FISTA algorithm and optimality of convergence rate in the case \( b \leq 3 \). SIAM J. Optim. 28(1), 551–574 (2018)
5. Attouch, H.: Viscosity solutions of minimization problems. SIAM J. Optim. 6(3), 769–806 (1996)
6. Attouch, H., Balhag, A., Chbani, Z., Riahi, H.: Damped inertial dynamics with vanishing Tikhonov regularization: Strong asymptotic convergence towards the minimum norm solution. J. Differ. Equ. 311, 29–58 (2022)
7. Attouch, H., Boţ, R.I., Csetnek, E.R.: Fast optimization via inertial dynamics with closed-loop damping. J. Eur. Math. Soc. (JEMS) (2022), published online first https://doi.org/10.4171/JEMS/1231. arXiv:hal-0291.0307
8. Attouch, H., Cabot, A.: Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity. J. Differ. Equ. 263(9), 5412–5458 (2017)
9. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First order optimization algorithms via inertial systems with Hessian driven damping. Math. Program. 193, 113–155 (2022). https://doi.org/10.1007/s10107-020-01591-1
10. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping. Optimization (2021). https://doi.org/10.1080/02331934.2021.2009828
11. Attouch, H., Chbani, Z., Peyrouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. 168(1–2), 123–175 (2018)
12. Attouch, H., Chbani, Z., Riahi, H.: Combining fast inertial dynamics for convex optimization with Tikhonov regularization. J. Math. Anal. Appl. 457, 1065–1094 (2018)
13. Attouch, H., Cominetti, R.: A dynamical approach to convex minimization coupling approximation with the steepest descent method. J. Differ. Equ. 128(2), 519–540 (1996)
14. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of coupled dynamical systems with non-isolated equilibria. J. Differ. Equ. 179, 278–310 (2002)
15. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of coupled dynamical systems with multiscale aspects. J. Differ. Equ. 248, 1315–1344 (2010)
16. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects. J. Differ. Equ. 262(3), 2745–2770 (2017)
17. Attouch, H., Fadili, J.: From the Ravine method to the Nesterov method and vice versa: a dynamic system perspective, accepted in SIAM J. Optim. arXiv:2201.11643v1 [math.OC] (2022)
18. Attouch, H., László, S.: Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution, arXiv:2104.11987v1 [math.OC] (2021)
19. Attouch, H., Maingé, P.E., Redont, P.: A second-order differential system with Hessian-driven damping. Application to non-elastic shock laws. Differ. Equ. Appl. 4(1), 27–65 (2012)
20. Attouch, H., Marques Alves, M., Svaiter, B.F.: A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert Spaces. J. Convex Anal. 23(1), 139–180 (2016)
21. Attouch, H., Peyrouquet, J.: The rate of convergence of Nesterov’s accelerated forward-backward method is actually faster than 1/k^2. SIAM J. Optim. 26(3), 1824–1834 (2016)
22. Attouch, H., Peyrouquet, J., Redont, P.: Fast convex minimization via inertial dynamics with Hessian driven damping. J. Differ. Equ. 261(10), 5734–5783 (2016)
23. Attouch, H., Svaier, B.F.: A continuous dynamical Newton-Like approach to solving monotone inclusions. SIAM J. Control Optim. 49(2), 574–598 (2011)
24. Baillon, J.-B., Cominetti, R.: 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, 263–273 (2001)
25. Bauschke, H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert spaces. CMS Books in Mathematics, Springer, New York (2011)
26. Boţ, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian damping. Math. Program. 189, 151–186 (2021)
27. Brézis, H.: Opérateurs maximaux monotones dans les espaces de Hilbert et équations d’évolution. Lecture Notes 5, North Holland (1972)
28. Brézis, H.: Analyse fonctionnelle. Masson (1983)
29. Cabot, A.: Inertial gradient-like dynamical system controlled by a stabilizing term. J. Optim. Theory Appl. 120, 275–303 (2004)
30. Cabot, A.: Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchical minimization. SIAM J. Optim. 15(2), 555–572 (2005)
31. Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation. Trans. Am. Math. Soc. 361, 5983–6017 (2009)
32. Chambolle, A., Dossal, Ch.: On the convergence of the iterates of Fista. J. Opt. Theory Appl. 166, 968–982 (2015)
33. Cominetti, R., Peyrouquet, J., Sorin, S.: Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization. J. Differ. Equ. 245, 3753–3763 (2008)
34. Esser, E., Zhang, X., Chan, T.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. 3(4), 1015–1046 (2010)
35. Hirstoaga, S.A.: Approximation et résolution de problèmes d’équilibre, de point fixe et d’inclusion monotone. PhD thesis, Université Pierre et Marie Curie - Paris VI, HAL Id: tel-00137228 (2006)
36. Jendoubi, M.A., May, R.: On an asymptotically autonomous system with Tikhonov type regularizing term. Arch. Math. 95(4), 389–399 (2010)
37. Lemaréchal, C., Sagastizábal, C.: Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. SIAM J. Optim. 7(2), 367–385 (1997)
38. May, R., Mnasri, C., Elloumi, M.: Asymptotic for a second order evolution equation with damping and regularizing terms. AIMS Math. 6(5), 4901–4914 (2021)
39. Nesterov, Y.: A method of solving a convex programming problem with convergence rate \(O(1/k^2)\). Soviet Math. Dokl. 27, 372–376 (1983)
40. Nesterov, Y.: Introductory Lectures on Convex Optimization: A Basic Course. Applied Optimization, vol. 87. Kluwer Academic Publishers, Boston, MA (2004)
41. Nesterov, Y.: How to Make the Gradients Small, Optima 88, Math. Optimization Soc. Newsletter (2012)
42. Polyak, B.: Some methods of speeding up the iteration of convergence methods. Comput. Math. Math. Phys. 4, 1–17 (1964)
43. Polyak, B.: Introduction to Optimization. Optimization Software-Inc, New York (1987)
44. Siegel, W.: Accelerated first-order methods: Differential equations and Lyapunov functions, arXiv:1903.05671v1 [math.OC] (2019)
45. Shi, B., Du, S.S., Jordan, M.I., Su, W.J.: Understanding the acceleration phenomenon via high-resolution differential equations. Math. Program. 195, 79–148 (2022). https://doi.org/10.1007/s10107-021-01681-8
46. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. J. Mach. Learn. Res. 17(153), 1–43 (2016)
47. Tikhonov, A.N.: Doklady Akademii Nauk SSSR 151 (1963) 501-504, (Translated in “Solution of incorrectly formulated problems and the regularization method”. Soviet Mathematics 4, 1035–1038 (1963)
48. Tikhonov, A.N., Arsenin, V.Y.: Solutions of Ill-Posed Problems. Winston, New York (1977)
49. Torralba, D.: Convergence épigraphique et changements d’échelle en analyse variationnelle et optimisation, PhD thesis, Université Montpellier (1996)
50. Wilson, A.C., Recht, B., Jordan, M.I.: A Lyapunov analysis of momentum methods in optimization. J. Mach. Learn. Res. 22, 1–34 (2021)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.