Second Order Dynamics Featuring Tikhonov Regularization and Time Scaling

Ernö Robert Csetnek[1] · Mikhail A. Karapetyants[1]

Abstract

In a Hilbert setting we aim to study a second order in time differential equation, combining viscous and Hessian-driven damping, containing a time scaling parameter function and a Tikhonov regularization term. The dynamical system is related to the problem of minimization of a nonsmooth convex function. In the formulation of the problem as well as in our analysis we use the Moreau envelope of the objective function and its gradient and heavily rely on their properties. We show that there is a setting where the newly introduced system preserves and even improves the well-known fast convergence properties of the function and Moreau envelope along the trajectories and also of the gradient of Moreau envelope due to the presence of time scaling. Moreover, in a different setting we prove strong convergence of the trajectories to the element of minimal norm from the set of all minimizers of the objective. The manuscript concludes with various numerical results.

✉ Mikhail A. Karapetyants
mikhail.karapetyants@univie.ac.at

Ernö Robert Csetnek
robert.csetnek@univie.ac.at

[1] Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria

1 Introduction

In the Hilbert setting H, where $\langle \cdot, \cdot \rangle$ denotes the inner product and the norm is defined as usual $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$, we will study the convergence properties of the following second order in time differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t)) + b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0,$$
(1)

with initial conditions $x(t_0) = x_0 \in H$, $\dot{x}(t_0) = \dot{x}_0 \in H$, where α, β and $t_0 > 0$, $\lambda : [t_0, +\infty) \mapsto \mathbb{R}_+$ and $b : [t_0, +\infty) \mapsto \mathbb{R}_+$ are non-negative, non-decreasing and differentiable, $\Phi : H \mapsto \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is a proper, convex and lower semicontinuous function and Φ_λ is its Moreau envelope of the index $\lambda > 0$ and the function $\varepsilon : [t_0, +\infty) \mapsto \mathbb{R}_+$ is continuously differentiable and non-increasing with the property $\lim_{t\to+\infty} \varepsilon(t) = 0$. In addition, we assume that argmin Φ, which is the set of global minimizers of Φ, is not empty and denote by Φ^* the optimal objective value of Φ. The system (1) has a connection to the minimization problem

$$\min_{x \in H} \Phi(x)$$

of a proper, convex and lower semicontinuous function Φ. Studying such systems provides better understanding of their discrete counterpart—optimization algorithms, since there is a strong connection between them, and the question of transitioning from one to another attracts a lot of attention in the modern literature.

One of the main goals of this research is to improve (compared to [23]) the fast rates of convergence for the Moreau envelope of the objective function and the objective function itself to Φ^*, as well as for the gradient of the Moreau envelope of the objective function in terms of the Moreau parameter function λ and the time scaling function b. Moreover, we also deduce the strong convergence of the trajectory of the dynamics to the minimal norm element of argmin Φ. We introduce two settings with different assumptions for each result. To conclude we provide multiple numerical results in order to illustrate our theoretical discoveries.

1.1 Nonsmooth Optimization with Time Scaling

In the smooth setting the pioneering research in studying second order dynamical systems was conducted by Su–Boyd–Candes [30] for the sake of obtaining faster asymptotic convergence for convex functions. They managed to deduce the rates of convergence of the function values being of the order $\frac{1}{t^2}$. Later Attouch–Peypouquet–Redont [20] also established the weak (and in some particular cases the strong) convergence of the trajectories to a minimizer of the objective function. In [19] the same authors continued the development in this direction by adding Hessian-driven damping term in order to obtain the rates for the gradient of the objective function and to eliminate any possible oscillations in the dynamical behaviour of the trajectories.

Concerning the nonsmooth setting we must point out that the Moreau envelope of a proper, convex and lower semicontinuous function $\Phi : H \to \overline{\mathbb{R}}$ proved to be of a significant importance in designing continuous-time approaches and numerical algorithms for the minimization of nonsmooth functions. The rigorous definition of this construction is

$$\Phi_\lambda : H \to \mathbb{R}, \quad \Phi_\lambda(x) = \inf_{y \in H} \left\{ \Phi(y) + \frac{1}{2\lambda} \|x - y\|^2 \right\},$$

where $\lambda > 0$ is the parameter of the Moreau envelope (see, for instance, [21]). One of the most important properties of Moreau approximation is that for every $\lambda > 0$, the functions Φ and Φ_λ share the same optimal objective value and also the same set of minimizers. Moreover, Φ_λ is convex and continuously differentiable with

$$\nabla \Phi_\lambda(x) = \frac{1}{\lambda}(x - \mathrm{prox}_{\lambda\Phi}(x)) \quad \forall x \in H, \tag{2}$$

and $\nabla \Phi_\lambda$ is $\frac{1}{\lambda}$-Lipschitz continuous, where

$$\mathrm{prox}_{\lambda\Phi} : H \to H, \quad \mathrm{prox}_{\lambda\Phi}(x) = \underset{y \in H}{\mathrm{argmin}} \left\{ \Phi(y) + \frac{1}{2\lambda} \|x - y\|^2 \right\},$$

denotes the proximal operator of Φ of parameter λ. The last fact we would like to mention is that for every $x \in H$, the function $\lambda \in (0, +\infty) \to \Phi_\lambda(x)$ is nonincreasing and differentiable (see [14], Lemma A1), namely,

$$\frac{d}{d\lambda} \Phi_\lambda(x) = -\frac{1}{2} \|\nabla \Phi_\lambda(x)\|^2 \quad \forall \lambda > 0. \tag{3}$$

Our research is a logical continuation of the one conducted in [24], where authors applied the time rescaling technique to a nonsmooth optimization problem (for more information on time scaling see also [5, 10, 11, 13]). They considered the following system

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta(t) \frac{d}{dt} \nabla \Phi_{\lambda(t)}(x(t)) + b(t) \nabla \Phi_{\lambda(t)}(x(t)) = 0, \tag{4}$$

where $\alpha \geq 1, t_0 > 0$, and $\beta : [t_0, +\infty) \mapsto [0, +\infty)$ and $b, \lambda : [t_0, +\infty) \mapsto (0, +\infty)$ are differentiable functions. On the one hand, the presence of the Hessian damping term is believed to help reducing the oscillations in the dynamical behaviour and provides the rates for the gradient of the objective function Φ. On the other hand, the time-scaling technique (which is considered to be an artificial way to speed up the convergence of values) affects the convergence rates while bringing more restrictions to the analysis. The following properties were established

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = o\left(\frac{1}{t^2 b(t)}\right) \text{ and } \|\dot{x}(t)\| = o\left(\frac{1}{t}\right) \text{ as } t \to +\infty,$$

from where through proximal mapping the convergence rates for the objective function Φ itself along the trajectory were obtained

$$\Phi\big(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\big) - \Phi^* = o\left(\frac{1}{t^2 b(t)}\right) \text{ and } \|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|$$

$$= o\left(\frac{\sqrt{\lambda(t)}}{t\sqrt{b(t)}}\right) \text{ as } t \to +\infty.$$

Note that by taking $b(\cdot) \equiv 1$ we arrive at the well-known convergence rate of the values being of the order $o\left(\frac{1}{t^2}\right)$. In addition, the following rates for the gradient of the Moreau envelope were deduced

$$\|\nabla\Phi_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t\sqrt{b(t)\lambda(t)}}\right), \text{ as } t \to +\infty.$$

Finally, the weak convergence of the trajectories $x(t)$ to a minimizer of Φ as $t \to +\infty$ was obtained.

In our analysis we borrow some ideas of [24] and develop them further in order to fit the new setting, namely, to adapt to a presence of the whole new term—Tikhonov regularization. The analysis becomes more involved and technical, some fundamental properties of Tikhonov regularization had to be proved for a nonsmooth setting. Its presence affects the set of conditions, which we have to impose on the system parameters: even though some of the conditions are formulated in the same spirit as in [24] (for instance, (11) and (14)), the other ones are completely new due to the presence of the Tikhonov term. Moreover, depending on how fast ε decays, two different setting arise providing different fundamental results (Sects. 3 and 4).

1.2 Tikhonov Regularization

It turned out that having additional term with specific properties in a system equation leads to improving the weak convergence of the trajectories to a minimizer of the objective function Φ to a strong one to the element of minimal norm of argmin Φ. Such systems were studied, for instance, in [4, 6, 9, 12, 17, 23, 27]. The main goal of such a research is to show that these systems preserve all the typical properties of the second order in time dynamical system (fast convergence of the values, the rates for the gradient etc.) but moreover there is an improvement to the strong convergence of the trajectories to the minimal norm solution instead of a weak one to an arbitrary minimizer. One of the many examples of such systems is presented below (see [23])

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0,$$

where $\alpha \geq 3$, $t_0 > 0$, $\Phi : H \mapsto \mathbb{R}$ is twice continuously differentiable and convex and for the rest of the section the function $\varepsilon : [t_0, +\infty) \mapsto \mathbb{R}_+$ is continuously differentiable and non-increasing with the property $\lim_{t\to+\infty} \varepsilon(t) = 0$. In that manuscript

they provided two settings: one for the fast convergence of values obtaining

$$\Phi(x(t)) - \Phi^* = o\left(\frac{1}{t^2}\right), \text{ as } t \to +\infty$$

and the weak convergence of the trajectories to a minimizer of Φ and another setting for the strong convergence of x to x^*, as $t \to +\infty$.

Another fine example is given in [4]:

$$\ddot{x}(t) + \alpha\sqrt{\varepsilon(t)}\dot{x}(t) + \nabla\Phi(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0, \qquad (5)$$

where $\alpha, t_0 > 0$ and $\Phi : H \mapsto \mathbb{R}$ is continuously differentiable and convex. In that paper authors obtained the rates for the function values $\Phi(x(t)) - \Phi^*$, as well as for the quantity $\|x(t) - x_{\varepsilon(t)}\|$, as $t \to +\infty$, where $x_{\varepsilon(t)} = \operatorname{argmin}_H\left(\Phi(x) + \frac{\varepsilon(t)\|x\|^2}{2}\right)$. Thus, they assured the strong convergence of the trajectories to the minimal norm solution $x^* = \operatorname{proj}_{\operatorname{argmin}\Phi}(0)$ under the appropriate assumptions and properly chosen energy functional, using the properties of Tikhonov regularization. The most important thing about this approach is that authors were able to establish fast convergence of values and strong convergence of the trajectories in the very same setting.

The next step was done in [6]:

$$\ddot{x}(t) + \alpha\sqrt{\varepsilon(t)}\dot{x}(t) + \beta\frac{d}{dt}\left(\nabla\varphi_t(x(t)) + (p-1)\varepsilon(t)x(t)\right) + \nabla\varphi_t(x(t)) = 0 \text{ for } t \geq t_0,$$

where $\varphi_t(x) = \Phi(x) + \frac{\varepsilon(t)\|x\|^2}{2}$, $\Phi : H \mapsto \mathbb{R}$ is twice continuously differentiable and convex and $p \in [0, 1]$. This system while preserving all the properties of (5), additionally provides the integral estimate for the norm of the gradient of φ_t.

1.3 Our Contribution

In that paper we will develop the ideas presented in [23] to cover the nonsmooth case with time scaling. We will obtain the fast convergence of the function values (as well as for the gradient of the Moreau envelope of the objective fucntion Φ) for the family of dynamical systems (1) governed by the Moreau envelope of the nonsmooth function Φ and having the Tiknonov term in their formulation:

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = o\left(\frac{1}{t^2 b(t)}\right) \text{ as } t \to +\infty;$$

in terms of the function itself:

$$\Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t))) - \Phi^* = o\left(\frac{1}{t^2 b(t)}\right) \text{ as } t \to +\infty,$$

where

$$\| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t) \| = o\left(\frac{\sqrt{\lambda(t)}}{t\sqrt{b(t)}}\right) \text{ as } t \to +\infty$$

and finally

$$\|\nabla\Phi_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t\sqrt{b(t)\lambda(t)}}\right) \text{ as } t \to +\infty.$$

We will also deduce (under some appropriate conditions) the following result

$$\liminf_{t\to+\infty} \|x(t) - x^*\| = 0,$$

which under some restrictions will be improved to the full strong convergence of the trajectories of (1) to the minimal norm solution.

The paper is organized in the following way. Section 2 is devoted to some preliminary results, which we will need later. We will establish the fast rates of convergence of function values and its Moreau envelope, as well as the gradient of Moreau envelope along the trajectories of the dynamical system (Sect. 3). We will show that under some assumptions the strong convergence of the trajectories to the element of minimal norm from the set of all minimizers of the objective function takes place (Sect. 4). We will provide two settings for the polynomial choice of parameter functions to fulfill the assumptions made through the analysis (Sect. 5) and equip this manuscript with various numerical results (Sect. 6).

2 Preparatory Results

We start with the following lemma (see [21], Proposition 12.22, for the first term of the lemma and [18], Appendix, A1, for the second one).

Lemma 1 *Let* $\Phi : H \mapsto \overline{\mathbb{R}}$ *be a proper, convex and lower semicontinuous function,* $\lambda, \mu > 0$. *Then*

1. $(\Phi_\lambda)_\mu = \Phi_{\lambda+\mu}$.
2. $\operatorname{prox}_{\mu\Phi_\lambda} = \frac{\lambda}{\lambda+\mu} \operatorname{Id} + \frac{\mu}{\lambda+\mu} \operatorname{prox}_{(\lambda+\mu)\Phi}$.

Let us mention two key properties of the Tikhonov regularization, which we will use later in the analysis (see, for instance, [2] or [21] Theorem 23.44 for its classic analogue). First let us introduce the strongly convex function $\varphi_{\varepsilon(t),\lambda(t)} : H \mapsto \mathbb{R}$ as $\varphi_{\varepsilon(t),\lambda(t)}(x) = \Phi_{\lambda(t)}(x) + \frac{\varepsilon(t)\|x\|^2}{2}$ and denote the unique minimizer of $\varphi_{\varepsilon(t),\lambda(t)}$ as $x_{\varepsilon(t),\lambda(t)} = \operatorname{argmin}_H \varphi_{\varepsilon(t),\lambda(t)}$. Thus, the first order optimality condition reads as

$$\nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) + \varepsilon(t)x_{\varepsilon(t),\lambda(t)} = 0. \tag{6}$$

Now we are ready to formulate the following result:

Lemma 2 *Suppose that*

$$\lim_{t \to +\infty} \lambda(t)\varepsilon(t) = 0. \tag{7}$$

Then the following properties of the mapping $t \mapsto x_{\varepsilon(t),\lambda(t)}$ *are satisfied:*

$$\text{for } x^* = \operatorname*{proj}_{\operatorname{argmin} \Phi}(0), \ \|x_{\varepsilon(t),\lambda(t)}\| \leq \|x^*\| \text{ for all } t \geq t_0 \tag{8}$$

and

$$\lim_{t \to +\infty} \|x_{\varepsilon(t),\lambda(t)} - x^*\| = 0. \tag{9}$$

Proof By the monotonicity of $\nabla \Phi_\lambda$ we deduce

$$\big\langle \nabla \Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \nabla \Phi_{\lambda(t)}(x^*), \ x_{\varepsilon(t),\lambda(t)} - x^* \big\rangle \geq 0.$$

By (6) we obtain

$$\big\langle -\varepsilon(t)x_{\varepsilon(t),\lambda(t)}, \ x_{\varepsilon(t),\lambda(t)} - x^* \big\rangle = \varepsilon(t)\left(-\|x_{\varepsilon(t),\lambda(t)}\|^2 + \big\langle x_{\varepsilon(t),\lambda(t)}, x^* \big\rangle \right) \geq 0.$$

Using Cauchy–Schwarz inequality we derive

$$\|x_{\varepsilon(t),\lambda(t)}\| \leq \|x^*\|.$$

This proves the first claim. For the second one consider (6) again and note that it is equivalent to

$$x_{\varepsilon(t),\lambda(t)} = \operatorname{prox}_{\frac{1}{\varepsilon(t)}\Phi_{\lambda(t)}}(0) = \frac{\operatorname{prox}_{\left(\lambda(t)+\frac{1}{\varepsilon(t)}\right)\Phi}(0)}{\lambda(t)\varepsilon(t) + 1}$$

by the item 2. of Lemma 1. Note that $\lambda(t) + \frac{1}{\varepsilon(t)} \to +\infty$, as $t \to +\infty$. Thus, the rest of the proof goes in line with Theorem 23.44 of [21]. $\qquad\square$

Our nearest goal is to deduce the existence and uniqueness of the solutions of the dynamical system (1). Suppose $\beta > 0$. Let us integrate (1) from t_0 to t to obtain

$$\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)) + \int_{t_0}^{t} \left(\frac{\alpha}{s}\dot{x}(s) + b(s)\nabla \Phi_{\lambda(s)}(x(s)) + \varepsilon(s)x(s) \right) ds$$
$$- \big(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x(t_0))\big) = 0.$$

Denoting $z(t) := \int_{t_0}^{t} \left(\frac{\alpha}{s}\dot{x}(s) + b(s)\nabla \Phi_{\lambda(s)}(x(s)) + \varepsilon(s)x(s) \right) ds - \big(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x_0)\big)\big)$ for every $t \geq t_0$ and noticing that $\dot{z}(t) = \frac{\alpha}{t}\dot{x}(t) + b(t)\nabla \Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t)$

we deduce, that (1) is equivalent to

$$\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)) + z(t) = 0,$$
$$\dot{z}(t) - \frac{\alpha}{t}\dot{x}(t) - b(t)\nabla \Phi_{\lambda(t)}(x(t)) - \varepsilon(t)x(t) = 0,$$
$$x(t_0) = x_0, \ z(t_0) = -\left(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x_0)\right).$$

Let us multiply the first line by the function b and the second one by the constant β and then sum them up to get rid of the gradient of the Moreau envelope in the second equation

$$\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)) + z(t) = 0,$$
$$\beta \dot{z}(t) + \left(b(t) - \frac{\alpha\beta}{t}\right)\dot{x}(t) - \beta\varepsilon(t)x(t) + b(t)z(t) = 0,$$
$$x(t_0) = x_0, \ z(t_0) = -\left(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x_0)\right).$$

We denote now $y(t) = \beta z(t) + \left(b(t) - \frac{\alpha\beta}{t}\right)x(t)$, and, after simplification, we obtain the following equivalent formulation for the dynamical system

$$\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)) + \left(\frac{\alpha}{t} - \frac{b(t)}{\beta}\right)x(t) + \frac{1}{\beta}y(t) = 0,$$
$$\dot{y}(t) - \left(\dot{b}(t) + \frac{\alpha\beta}{t^2} + \beta\varepsilon(t) + \frac{b^2(t)}{\beta} - \frac{\alpha b(t)}{t}\right)x(t) + \frac{b(t)}{\beta}y(t) = 0,$$
$$x(t_0) = x_0, \ y(t_0) = -\beta\left(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x_0)\right) + \left(b(t_0) - \frac{\alpha\beta}{t_0}\right)x_0.$$

In case $\beta = 0$ for every $t \geq t_0$, (1) can be equivalently written as

$$\dot{x}(t) - y(t) = 0,$$
$$\dot{y}(t) + \frac{\alpha}{t}y(t) + b(t)\nabla \Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t) = 0,$$
$$x(t_0) = x_0, \ y(t_0) = \dot{x}(t_0).$$

Based on the two reformulations of the dynamical system (1) we formulate the following existence and uniqueness result, which is a consequence of Cauchy-Lipschitz theorem for strong global solutions. The result can be proved in the lines of the proofs of Theorem 1 in [16] or of Theorem 1.1 in [19] with some small adjustments.

Theorem 3 *Suppose that there exists $\lambda_0 > 0$ such that $\lambda(t) \geq \lambda_0$ for all $t \geq t_0$. Then for every $(x_0, \dot{x}(t_0)) \in H \cdot H$ there exists a unique strong global solution $x : [t_0, +\infty) \mapsto H$ of the continuous dynamics (1) which satisfies the Cauchy initial conditions $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$.*

3 Fast Convergence Rates of the Function and Moreau Envelope Values

This chapter is devoted to obtaining the rates of convergence for the Moreau envelope values and for the values of function Φ itself. We will heavily rely on the tools and techniques provided by the Lyapunov analysis. We introduce a slightly modified energy function from [23]. For $2 \leq q \leq \alpha - 1$ we define

$$
\begin{aligned}
E_q(t) = {} & (t^2 b(t) - \beta(q+2-\alpha)t)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + \frac{t^2 \varepsilon(t)}{2}\|x(t)\|^2 \\
& + \frac{1}{2}\|q(x(t) - x^*) + t\left(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\right)\|^2 + \frac{q(\alpha - 1 - q)}{2}\|x(t) - x^*\|^2.
\end{aligned}
\tag{10}
$$

The key assumptions which are essential to our analysis are the following: for all $t \geq t_0$

$$
(\alpha - 3)tb(t) - t^2 \dot{b}(t) + \beta(2 - \alpha) \geq 0, \tag{11}
$$

$$
\exists a \geq 1 \text{ such that } 2\dot{\varepsilon}(t) \leq -a\beta\varepsilon^2(t), \tag{12}
$$

$$
\int_{t_0}^{+\infty} t\varepsilon(t)dt < +\infty, \tag{13}
$$

$$
b(t_0) \geq \frac{\beta}{t_0} \text{ and } b(t_0) > \frac{1}{a} \tag{14}
$$

and

$$
\exists \delta : 0 < \delta < \alpha - 3 \text{ such that } (\alpha - 3)tb(t) - t^2 \dot{b}(t) + \beta(2 - \alpha) \geq \delta tb(t). \tag{15}
$$

Theorem 4 *Suppose $\alpha \geq 3$ and assume that (11), (12), (13), (14) hold for all $t \geq t_0$. Then*

$$
\Phi_{\lambda(t)}(x(t)) - \Phi^* = O\left(\frac{1}{t^2 b(t)}\right), \text{ as } t \to +\infty,
$$

$$
\|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\| = O\left(\frac{1}{t}\right), \text{ as } t \to +\infty.
$$

Moreover, one has for all $a \geq 1$

$$t\varepsilon(t)\|x(t) - x^*\|^2, \quad t\varepsilon(t)\|x(t)\|^2, \quad \left((\alpha - 3)tb(t) - t^2\dot{b}(t) + \beta(2 - \alpha)\right)$$
$$\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) \text{ and}$$
$$\left(\left(t^2b(t) - \beta t\right)\frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2\left(b(t) - \frac{1}{a}\right)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \in L^1\left([t_0, +\infty), \mathbb{R}\right).$$

If, in addition, $\alpha > 3$ and (15) holds, then the trajectory x is bounded and

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

and

$$\int_{t_0}^{+\infty} tb(t)\left(\Phi_{\lambda(s)}(x(s)) - \Phi^*\right) < +\infty.$$

Proof Let us compute the time derivative of the energy function. For every $t \geq t_0$ using (3) we derive

$$\begin{aligned}
\dot{E}_q(t) = {} & \left(2tb(t) + t^2\dot{b}(t) - \beta(q + 2 - \alpha)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + t^2\varepsilon(t)\langle x(t), \dot{x}(t)\rangle \\
& + q(\alpha - 1 - q)\langle\dot{x}(t), x(t) - x^*\rangle \\
& + \left(t^2b(t) - \beta(q + 2 - \alpha)t\right)\left(\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle - \frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\right) \\
& + \frac{2t\varepsilon(t) + t^2\dot{\varepsilon}(t)}{2}\|x(t)\|^2 \\
& + \Big\langle q(x(t) - x^*) + t\left(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\right), (q + 1)\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)) \\
& + t\left(\ddot{x}(t) + \beta\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t))\right)\Big\rangle.
\end{aligned}$$

Define $v(t) = q(x(t) - x^*) + t\left(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\right)$. Using (1) to replace $\ddot{x}(t) + \beta\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t))$ we obtain

$$\begin{aligned}
\langle v(t), \dot{v}(t)\rangle = {} & \Big\langle q(x(t) - x^*) + t\left(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\right), (q + 1 - \alpha)\dot{x}(t) \\
& + (\beta - tb(t))\nabla\Phi_{\lambda(t)}(x(t)) - t\varepsilon(t)x(t)\Big\rangle \\
= {} & q(q + 1 - \alpha)\langle x(t) - x^*, \dot{x}(t)\rangle + (q + 1 - \alpha)t\|\dot{x}(t)\|^2 \\
& + \left(\beta(q + 2 - \alpha)t - t^2b(t)\right)\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle \\
& + \left(\beta^2 t - \beta t^2 b(t)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2
\end{aligned}$$

$$- t^2 \varepsilon(t) \langle x(t), \dot{x}(t) \rangle - \beta t^2 \varepsilon(t) \langle x(t), \nabla \Phi_{\lambda(t)}(x(t)) \rangle$$
$$- qt \left\langle \left(b(t) - \frac{\beta}{t} \right) \nabla \Phi_{\lambda(t)}(x(t)) + \varepsilon(t) x(t), x(t) - x^* \right\rangle.$$

By (14) one has $b(t) - \frac{\beta}{t} > 0$ for all $t \geq t_0$, and thus for a strongly convex function $\varphi_t(x) = \left(b(t) - \frac{\beta}{t} \right) \Phi_{\lambda(t)}(x) + \frac{\varepsilon(t)}{2} \|x\|^2$ we have

$$\varphi_t(x^*) - \varphi_t(x) \geq \left\langle \nabla \varphi_t(x), x^* - x \right\rangle + \frac{\varepsilon(t)}{2} \|x^* - x\|^2$$

or

$$- qt \left\langle \left(b(t) - \frac{\beta}{t} \right) \nabla \Phi_{\lambda(t)}(x(t)) + \varepsilon(t) x(t), x(t) - x^* \right\rangle$$
$$\leq - qt \left(b(t) - \frac{\beta}{t} \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right)$$
$$- qt \frac{\varepsilon(t)}{2} \|x(t)\|^2 - qt \frac{\varepsilon(t)}{2} \|x(t) - x^*\|^2 + qt \frac{\varepsilon(t)}{2} \|x^*\|^2.$$

Therefore, for every $t \geq t_0$

$$\dot{E}_q(t) \leq \left((2-q)tb(t) + t^2 \dot{b}(t) - \beta(2-\alpha) \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + (q+1-\alpha)t \|\dot{x}(t)\|^2$$
$$- \left((t^2 b(t) - \beta(q+2-\alpha)t) \frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2 b(t) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2$$
$$+ \frac{(2-q)t\varepsilon(t) + t^2 \dot{\varepsilon}(t)}{2} \|x(t)\|^2 - qt \frac{\varepsilon(t)}{2} \|x(t) - x^*\|^2 + qt \frac{\varepsilon(t)}{2} \|x^*\|^2$$
$$- \beta t^2 \varepsilon(t) \langle x(t), \nabla \Phi_{\lambda(t)}(x(t)) \rangle.$$

Notice that for $a \geq 1$

$$-\beta t^2 \varepsilon(t) \langle x(t), \nabla \Phi_{\lambda(t)}(x(t)) \rangle \leq \frac{\beta t^2}{a} \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 + \frac{a \beta t^2 \varepsilon^2(t)}{4} \|x(t)\|^2,$$

which leads to

$$\dot{E}_q(t) \leq \left((2-q)tb(t) + t^2 \dot{b}(t) - \beta(2-\alpha) \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + (q+1-\alpha)t \|\dot{x}(t)\|^2$$
$$- \left((t^2 b(t) - \beta(q+2-\alpha)t) \frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2 \left(b(t) - \frac{1}{a} \right) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 \quad (16)$$
$$+ \frac{2(2-q)t\varepsilon(t) + 2t^2 \dot{\varepsilon}(t) + a \beta t^2 \varepsilon^2(t)}{4} \|x(t)\|^2 - qt \frac{\varepsilon(t)}{2} \|x(t) - x^*\|^2 + qt \frac{\varepsilon(t)}{2} \|x^*\|^2$$

for every $t \geq t_0$. Note that $b(t) - \frac{1}{a} > 0$ for all $t \geq t_0$. Then, due to the properties of b, there exists $t^* \geq t_0$ such that $t^2 b(t) - \beta(q+2-\alpha)t \geq 0$ for all $t \geq t^*$ and

all $q \in (2, \alpha - 1]$. Therefore, since $\dot{\lambda}(t) \geq 0$ for all $t \geq t_0$, there exists t^{**}, namely,
$t^{**} = \max \left\{ t^*, \frac{\beta}{b(t_0) - \frac{1}{a}} \right\}$, such that

$$\left(\left(t^2 b(t) - \beta(q + 2 - \alpha)t \right) \frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2 \left(b(t) - \frac{1}{a} \right) \right) \geq 0 \text{ for all } t \geq t^{**}.$$

Consider now two cases with $t \geq t^{**}$. First, take $q = \alpha - 1$ to obtain from (16)

$$
\begin{aligned}
\dot{E}_{\alpha-1}(t) \leq\ & \left((3 - \alpha)t b(t) + t^2 \dot{b}(t) - \beta(2 - \alpha) \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) \\
& - \left(\left(t^2 b(t) - \beta t \right) \frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2 \left(b(t) - \frac{1}{a} \right) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 \\
& + \frac{2(3 - \alpha)t\varepsilon(t) + 2t^2 \dot{\varepsilon}(t) + a\beta t^2 \varepsilon^2(t)}{4} \|x(t)\|^2 - (\alpha - 1)t \frac{\varepsilon(t)}{2} \|x(t) - x^*\|^2 \\
& + (\alpha - 1)t \frac{\varepsilon(t)}{2} \|x^*\|^2
\end{aligned}
\tag{17}
$$

for every $t \geq t_0$. Under the assumptions (11) and (12) we conclude starting from t^{**} that

$$\dot{E}_{\alpha-1}(t) \leq \frac{(\alpha - 1)t\varepsilon(t)}{2} \|x^*\|^2. \tag{18}$$

Under the assumption (13) using the fact that $t \mapsto E_{\alpha-1}(t)$ is bounded from below we deduce the existence of the limit $\lim_{t \to +\infty} E_{\alpha-1}(t)$ due to the Lemma A.1 and, therefore, $t \mapsto E_{\alpha-1}(t)$ is bounded, which leads to

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = O\left(\frac{1}{t^2 b(t)} \right), \text{ as } t \to +\infty.$$

From the boundedness of $t \mapsto \|(\alpha - 1)(x(t) - x^*) + t(\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)))\|^2$ we obtain

$$\|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\| = O\left(\frac{1}{t} \right), \text{ as } t \to +\infty,$$

using the following inequality, which is true for every $t \geq t_0$

$$
\begin{aligned}
t^2 \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2 \leq\ & 2\|(\alpha - 1)(x(t) - x^*) + t(\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)))\|^2 \\
& + 2(\alpha - 1)^2 \|x(t) - x^*\|^2.
\end{aligned}
$$

Moreover, integrating (17) one may obtain the integrability of $t\varepsilon(t)\|x(t) - x^*\|^2$ as well as the other terms in (17). Consider now $q = \alpha - 1 - \delta$, where δ is defined by (15). Thus, (16) becomes

$$\dot{E}_{\alpha-1-\delta}(t) \leq \left((3 - \alpha + \delta)t b(t) + t^2 \dot{b}(t) - \beta(2 - \alpha) \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) - \delta t \|\dot{x}(t)\|^2$$

$$- \left(\left(t^2 b(t) - \beta(1-\delta)t \right) \frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2 \left(b(t) - \frac{1}{a} \right) \right) \| \nabla \Phi_{\lambda(t)}(x(t)) \|^2$$

$$+ \frac{2(3-\alpha+\delta)t\varepsilon(t) + 2t^2\dot{\varepsilon}(t) + a\beta t^2\varepsilon^2(t)}{4} \|x(t)\|^2 - (\alpha-1-\delta)t\frac{\varepsilon(t)}{2} \|x(t) - x^*\|^2$$

$$+ (\alpha-1-\delta)t\frac{\varepsilon(t)}{2} \|x^*\|^2. \tag{19}$$

Under the assumptions (12) and (15) we deduce $\dot{E}_{\alpha-1-\delta}(t) \leq \frac{(\alpha-1-\delta)t\varepsilon(t)}{2}\|x^*\|^2$ starting from t^{**}. Repeating the same argument we derive that $t \mapsto E_{\alpha-1-\delta}(t)$ is bounded. The function $t \mapsto \|x(t) - x^*\|$ is also bounded and so is the trajectory x. Integrating (19) one may additionally obtain the integrability of $t\|\dot{x}(t)\|^2$. From the integrability of $\left((\alpha-3)tb(t) - t^2\dot{b}(t) - \beta(\alpha-2) \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right)$ and (15) we deduce

$$\int_{t_0}^{+\infty} tb(t) \left(\Phi_{\lambda(s)}(x(s)) - \Phi^* \right) < +\infty.$$

$\square$

The next theorem shows that we can actually improve the rates of convergence of the function values in case $\alpha > 3$.

Theorem 5 *Assume that $\alpha > 3$ and (12), (13),(14) and (15) hold. Then*

$$t \left\langle \left(b(t) - \frac{\beta}{t} \right) \nabla \Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle \in L^1\left([t_0, +\infty), \mathbb{R} \right). \tag{20}$$

In addition, $\lim_{t\to+\infty} \psi(t) = 0$, where for $2 \leq q \leq \alpha - 1$

$$\psi(t) = (t^2 b(t) - \beta(q+2-\alpha)t) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + \frac{t^2\varepsilon(t)}{2} \|x(t)\|^2$$
$$+ \frac{t^2}{2} \|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\|^2,$$

which in particular means

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = o\left(\frac{1}{t^2 b(t)} \right) \text{ as } t \to +\infty,$$
$$\|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t} \right) \text{ as } t \to +\infty \tag{21}$$

and moreover,

$$\Phi(\text{prox}_{\lambda(t)\Phi}(x(t))) - \Phi^* = o\left(\frac{1}{t^2 b(t)} \right) \text{ as } t \to +\infty,$$
$$\| \text{prox}_{\lambda(t)\Phi}(x(t)) - x(t) \| = o\left(\frac{\sqrt{\lambda(t)}}{t\sqrt{b(t)}} \right) \text{ as } t \to +\infty$$

and

$$\|\nabla\Phi_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t\sqrt{b(t)\lambda(t)}}\right) \ as \ t \to +\infty.$$

Proof (*i*) Let us first prove an auxiliary estimate (20), which will allow us to obtain the rest of the desired results. We return to

$$\dot{E}_q(t) \leq \left(2tb(t) + t^2\dot{b}(t) - \beta(q + 2 - \alpha)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + (q + 1 - \alpha)t\|\dot{x}(t)\|^2$$
$$- \left((t^2b(t) - \beta(q + 2 - \alpha)t)\frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2\left(b(t) - \frac{1}{a}\right)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2$$
$$+ \frac{4t\varepsilon(t) + 2t^2\dot{\varepsilon}(t) + a\beta t^2\varepsilon^2(t)}{4}\|x(t)\|^2$$
$$- qt\left\langle\left(b(t) - \frac{\beta}{t}\right)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t) - x^*\right\rangle.$$

Under condition (12) we deduce starting from t^{**}

$$\dot{E}_q(t) \leq \left(2tb(t) + t^2\dot{b}(t) - \beta(q + 2 - \alpha)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right)$$
$$+ t\varepsilon(t)\|x(t)\|^2 - qt\left\langle\left(b(t) - \frac{\beta}{t}\right)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t) - x^*\right\rangle.$$

Integrating the last inequality on $[t_0, t]$ we obtain

$$\int_{t_0}^t qs\left\langle\left(b(s) - \frac{\beta}{s}\right)\nabla\Phi_{\lambda(s)}(x(s)), x(s) - x^*\right\rangle ds \leq E_q(t_0) - E_q(t) + \int_{t_0}^t s\varepsilon(s)\|x(s)\|^2 ds$$
$$+ \int_{t_0}^t \left(2sb(s) + s^2\dot{b}(s) - \beta(q + 2 - \alpha)\right)\left(\Phi_{\lambda(s)}(x(s)) - \Phi^*\right) ds - \int_{t_0}^t qs\langle\varepsilon(s)x(s), x(s) - x^*\rangle.$$
$$(22)$$

Since the gradient $\nabla\Phi_\lambda$ is monotone, we know that $\langle\nabla\Phi_{\lambda(t)}(x(t)), x(t) - x^*\rangle \geq 0$. Moreover,

$$- qt\langle\varepsilon(t)x(t), x(t) - x^*\rangle \leq \frac{qt\varepsilon(t)}{2}\left(\|x(t)\|^2 + \|x(t) - x^*\|^2\right). \qquad (23)$$

Notice that by (15) we have

$$(\alpha - 3 - \delta)tb(t) - t^2\dot{b}(t) + \beta(2 - \alpha) > 0$$

or

$$(\alpha - 3 - \delta)tb(t) - t^2\dot{b}(t) > \beta(\alpha - 2).$$

Obviuosly,

$$(\alpha - 3 - \delta)tb(t) - t^2\dot{b}(t) \; > \; \beta(\alpha - 2) \; > \; \beta(\alpha - 2 - q) \text{ for every } q \in (2, \alpha - 1).$$

Introducing $\delta_1 = \alpha - 1 - \delta > 0$ (by the choice of δ) we obtain

$$(\delta_1 - 2)tb(t) - t^2\dot{b}(t) \; > \; -\beta(q + 2 - \alpha)$$

or

$$2tb(t) + t^2\dot{b}(t) - \beta(q + 2 - \alpha) \; < \; \delta_1 tb(t).$$

From Theorem 4 we know that $tb(t)\left(\Phi_{\lambda(s)}(x(s)) - \Phi^*\right)$ is integrable and therefore so is $\left(2tb(t) + t^2\dot{b}(t) - \beta(q + 2 - \alpha)\right)\left(\Phi_{\lambda(s)}(x(s)) - \Phi^*\right)$. Since the function $t \mapsto E_q(t)$ is bounded and the rest of the right hand side of (22) belongs to $L^1\left([t_0, +\infty), \mathbb{R}\right)$ by Theorem 4 and (23), we conclude with (20) due to (14).

(*ii*) In order to derive the convergence rates for the quantities of our interest we require some additional results. Our nearest goal is to establish the existence of the limits

$$\lim_{t \to +\infty} \|x(t) - x^*\| \text{ and } \lim_{t \to +\infty} t\left\langle \dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle.$$

Consider (as was done in [23, 24]) for two different $q_1, q_2 \in (2, \alpha - 1)$ and for every $t \geq t_0$ the difference

$$E_{q_1}(t) - E_{q_2}(t) \; = \; (t^2 b(t) - \beta(q_1 + 2 - \alpha)t)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + \frac{t^2\varepsilon(t)}{2}\|x(t)\|^2$$

$$+ \frac{1}{2}\|q_1(x(t) - x^*) + t(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)))\|^2 + \frac{q_1(\alpha - 1 - q_1)}{2}\|x(t) - x^*\|^2$$

$$- (t^2 b(t) - \beta(q_2 + 2 - \alpha)t)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) - \frac{t^2\varepsilon(t)}{2}\|x(t)\|^2$$

$$- \frac{1}{2}\|q_2(x(t) - x^*) + t(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)))\|^2 - \frac{q_2(\alpha - 1 - q_2)}{2}\|x(t) - x^*\|^2$$

$$= \; (q_1 - q_2)\left(-\beta t\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + t\left\langle \dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle \right.$$

$$+ \frac{\alpha - 1}{2}\|x(t) - x^*\|^2 \bigg).$$

As we have established earlier in Theorem 4 the limits of $E_{q_1}(t) - E_{q_2}(t)$ and $t\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right)$ exists (the latter is actually zero). Therefore, the limit

$$\lim_{t \to +\infty}\left(t\left\langle \dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle + \frac{\alpha - 1}{2}\|x(t) - x^*\|^2 \right) \text{ also exists.}$$

Let us introduce for every $t \geq t_0$ two auxiliary functions

$$k(t) = t \left\langle \dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle + \frac{\alpha - 1}{2} \|x(t) - x^*\|^2$$

and

$$r(t) = \frac{1}{2} \|x(t) - x^*\|^2 + \beta \int_{t_0}^t \left\langle \nabla \Phi_{\lambda(s)}(x(s)), x(s) - x^* \right\rangle ds.$$

Noticing that

$$\dot{r}(t) = \langle x(t) - x^*, \dot{x}(t) \rangle + \beta \left\langle \nabla \Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle$$

we may write for every $t \geq t_0$

$$(\alpha - 1)r(t) + t\dot{r}(t) = k(t) + \beta(\alpha - 1) \int_{t_0}^t \left\langle \nabla \Phi_{\lambda(s)}(x(s)), x(s) - x^* \right\rangle ds.$$

From the fact that $\lim_{t \to +\infty} k(t)$ exists using (20) we obtain that $\lim_{t \to +\infty} (\alpha - 1)r(t) + t\dot{r}(t)$ also exists. Applying Lemma A.2 we deduce the existence of the limit $\lim_{t \to +\infty} r(t)$. Using (20) again we obtain the existence of the limits $\lim_{t \to +\infty} \|x(t) - x^*\|$ and $\lim_{t \to +\infty} t \left\langle \dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle$.

(*iii*) Finally, we are in position to prove (21) and the rest of the convergence rates. The key idea is to show that the limit

$$\lim_{t \to +\infty} \left((t^2 b(t) - \beta(q + 2 - \alpha)t) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + \frac{t^2 \varepsilon(t)}{2} \|x(t)\|^2 \right.$$
$$\left. + \frac{t^2}{2} \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2 \right)$$

exists and is actually zero. Let us return to the definition of our energy functional and rewrite it as

$$E_q(t) = (t^2 b(t) - \beta(q + 2 - \alpha)t) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + \frac{t^2 \varepsilon(t)}{2} \|x(t)\|^2$$
$$+ \frac{t^2}{2} \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2 + qt \left\langle \dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle$$
$$+ \frac{q(\alpha - 1)}{2} \|x(t) - x^*\|^2.$$

Since the limits

$$\lim_{t \to +\infty} E_q(t) \text{ and } \lim_{t \to +\infty} \left(qt \left\langle \dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t), x(t) - x^* \right\rangle + \frac{q(\alpha - 1)}{2} \|x(t) - x^*\|^2 \right) \text{ exist,}$$

it follows that

$$
\lim_{t \to +\infty} \left((t^2 b(t) - \beta(q + 2 - \alpha)t) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + \frac{t^2 \varepsilon(t)}{2} \|x(t)\|^2 \right.
$$
$$
\left. + \frac{t^2}{2} \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2 \right)
$$

exists as well. Denote

$$
\psi(t) = (t^2 b(t) - \beta(q + 2 - \alpha)t) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + \frac{t^2 \varepsilon(t)}{2} \|x(t)\|^2
$$
$$
+ \frac{t^2}{2} \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2
$$

and consider

$$
0 \leq \frac{\psi(t)}{t} \leq 2t b(t) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right) + \frac{t \varepsilon(t)}{2} \|x(t)\|^2
$$
$$
+ \frac{t}{2} \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2. \tag{24}
$$

Let us show that the right hand side of (24) is integrable. Indeed, the first term is integrable by Theorem 4. As we have also established in Theorem 4, starting from t^{**}

$$
\left(t b(t) - \beta \right) \frac{\dot{\lambda}(t)}{2} + \beta t \left(b(t) - \frac{1}{a} \right) \geq \beta^2,
$$

where $a \geq 1$. Then, by (14) and $\dot{\lambda}(t) \geq 0$ for all $t \geq t_0$, we deduce that there exists $t_1 \geq t^{**}$ such that for all $t \geq t_1$

$$
\left(t b(t) - \beta \right) \frac{\dot{\lambda}(t)}{2} + \beta t \left(b(t) - \frac{1}{a} \right) \geq \frac{3 \beta^2}{2}
$$

or

$$
t^2 b(t) \left(\frac{\dot{\lambda}(t)}{2} + \beta \left(1 - \frac{1}{ab(t)} \right) \right) \geq (3\beta + \dot{\lambda}(t)) \frac{\beta t}{2}
$$

or

$$
\left(t^2 b(t) - \beta t \right) \frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2 \left(b(t) - \frac{1}{a} \right) \geq \frac{\beta^2 t}{2},
$$

So, by Theorem 4 the right hand side of (24) belongs to $L^1\big([t_1, +\infty), \mathbb{R}\big)$. Therefore, $\frac{\psi(t)}{t}$ also belongs to $L^1\big([t_1, +\infty), \mathbb{R}\big)$ and since the limit $\lim_{t \to +\infty} \psi(t)$ exists we

deduce that it should be actually zero, which gives us (21). To complete the proof notice that by the definition of the proximal mapping, we have

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = \Phi(\text{prox}_{\lambda(t)\Phi}(x(t))) - \Phi^* + \frac{1}{2\lambda(t)} \| \text{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \quad \forall t \geq t_0.$$

The conclusion follows immediately from (2) and (21).   □

4 Strong Convergence of the Trajectories

In this chapter we will establish the strong convergence of the trajectories to the minimal norm element of argmin Φ.

In order to do so, we will need to modify assumption (13) from the previous chapter:

$$\int_{t_0}^{+\infty} \frac{\varepsilon(t)}{tb(t)} dt < +\infty. \tag{25}$$

Before moving to the main point of the section, let us prove an auxiliary result first.

Theorem 6 *Suppose that $\alpha > 3$, the function λ is bounded for all $t \geq t_0$ and (11), (12), (14) and (25) hold. Then*

$$\lim_{t \to +\infty} \| \text{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\| = 0$$

and

$$\lim_{t \to +\infty} \Phi\left(\text{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* = 0.$$

Proof Let us return to (18):

$$\dot{E}_{\alpha-1}(t) \leq (\alpha - 1)t \frac{\varepsilon(t)}{2} \|x^*\|^2.$$

Let us integrate the last inequality on [T, t]

$$E_{\alpha-1}(t) \leq E_{\alpha-1}(T) + \frac{(\alpha - 1)\|x^*\|^2}{2} \int_T^t s\varepsilon(s)ds.$$

On the other hand, for every $t \geq t_0$

$$E_{\alpha-1}(t) \geq (t^2 b(t) - \beta t)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right).$$

Thus,

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* \leq \frac{E_{\alpha-1}(T)}{t^2 b(t) - \beta t} + \frac{(\alpha-1)\|x^*\|^2}{2(t^2 b(t) - \beta t)} \int_T^t s\varepsilon(s)ds.$$

We deduce due to (25) and the Lemma A.3 that

$$\lim_{t \to +\infty} \frac{1}{t^2 b(t)} \int_T^t s^2 b(s) \frac{\varepsilon(s)}{s b(s)} ds = 0.$$

Therefore,

$$\lim_{t \to +\infty} \frac{(\alpha-1)\|x^*\|^2}{2(t^2 b(t) - \beta t)} \int_T^t s\varepsilon(s)ds = 0$$

and clearly

$$\lim_{t \to +\infty} \frac{E_{\alpha-1}(T)}{t^2 b(t) - \beta t} = 0.$$

Thus, we establish

$$\lim_{t \to +\infty} \Phi_{\lambda(t)}(x(t)) - \Phi^* = 0.$$

By the definition of the proximal mapping

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = \Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* + \frac{1}{2\lambda(t)} \|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \quad \forall t \geq t_0.$$

Using the fact that λ is bounded for all $t \geq t_0$ we deduce

$$\lim_{t \to +\infty} \|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\| = 0$$

and

$$\lim_{t \to +\infty} \Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* = 0.$$

□

For the remaining part of this section we will use a different energy functional. Inspired by [23] we introduce the following functional, which we will heavily rely on throughout this section

$$E_{p,q}(t) = t^{p+1}\left(tb(t) + \beta(\alpha - p - q - 2)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right)$$
$$+ \frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) + \frac{t^p}{2}\|v(t)\|^2, \tag{26}$$

where $v(t) = q(x(t) - x^*) + t(\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)))$ and $p,\ q \geq 0$.

The proof of the following theorem draws inspiration from [10, 15, 23].

Theorem 7 *Suppose that λ is bounded for all $t \geq t_0$, $\alpha > 3$, $b(t_0) \geq \frac{1}{2} + \frac{\beta}{t_0}$ and (11), (12) and (25) are fulfilled. Suppose additionally that for all $t \geq t_0$*

$$\left(\frac{\alpha}{3} - 1\right) t b(t) - t^2 \dot{b}(t) + \frac{\alpha \beta}{3} \geq 0 \tag{27}$$

and moreover that for all $t \geq t_0$

$$2\alpha(\alpha - 3) - 9t^2 \varepsilon(t) + 6\alpha\beta \leq 0, \tag{28}$$

$$18\beta t + 9\beta\dot{\lambda}(t) - 9tb(t)\left(\dot{\lambda}(t) + 2\beta\right) + 3(\alpha + 3)\beta^2 + \alpha^2\beta \leq 0 \tag{29}$$

and

$$\lim_{t \to +\infty} \frac{\beta}{t^{\frac{\alpha}{3}+1}\varepsilon(t)} \int_{t_0}^{t} s^{\frac{\alpha}{3}+1}\varepsilon^2(s)\,ds = 0. \tag{30}$$

If $x : [t_0, +\infty) \mapsto H$ is a solution to (1) and the trajectory $x(t)$ stays either inside or outside the ball $B(0, \|x^*\|)$, then $x(t)$ converges to minimal norm solution $x^* = \mathrm{proj}_{\mathrm{argmin}\,\Phi}(0)$, as $t \to +\infty$. Otherwise, $\liminf_{t \to +\infty} \|x(t) - x^*\| = 0$.

Proof As in [23] we will consider several cases with respect to the trajectory x staying either inside or outside the ball $B(0, \|x^*\|)$.

Case I.

Assume that the trajectory x stays in the complement of the ball B for all $t \geq t_0$. This means nothing but $\|x(t)\| \geq \|x^*\|$ for every $t \geq t_0$.

(i) Our nearest goal is to obtain the upper bound for the derivative of $E_{p,q}$. In order to do so, let us evaluate its time derivative for every $t \geq t_0$ first.

$$\begin{aligned}
\frac{d}{dt}E_{p,q}(t) &= t^p \left((p+2)tb(t) + t^2\dot{b}(t) + (p+1)\beta(\alpha - p - q - 2)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) \\
&+ t^{p+1}(tb(t) + \beta(\alpha - p - q - 2))\left(\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle - \frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\right) \\
&+ \frac{(p+2)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) + t^{p+2}\varepsilon(t)\langle\dot{x}(t), x(t)\rangle \\
&+ \frac{pt^{p-1}}{2}\|q(x(t) - x^*) + t(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)))\|^2 + t^p\langle\dot{v}(t), v(t)\rangle.
\end{aligned} \tag{31}$$

Consider for every $t \geq t_0$ the inner product $\langle\dot{v}(t), v(t)\rangle$:

$$\begin{aligned}
\Bigg\langle (q+1)\dot{x}(t) &+ \beta\nabla\Phi_{\lambda(t)}(x(t)) + t\left(\ddot{x}(t) + \beta\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t))\right), q(x(t) - x^*) + t(\dot{x}(t) \\
&+ \beta\nabla\Phi_{\lambda(t)}(x(t)))\Bigg\rangle
\end{aligned}$$

$$= \Big\langle (q+1-\alpha)\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)) - t\Big(b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t)\Big), q(x(t)-x^*) + t(\dot{x}(t)$$

$$+ \beta\nabla\Phi_{\lambda(t)}(x(t))\Big\rangle$$

$$= q(q+1-\alpha)\langle\dot{x}(t), x(t)-x^*\rangle + (q+1-\alpha)t\left(\|\dot{x}(t)\|^2 + \langle\beta\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle\right)$$

$$+ \beta q\langle\nabla\Phi_{\lambda(t)}(x(t)), x(t)-x^*\rangle + \beta t\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle + \beta^2 t\|\nabla\Phi_{\lambda(t)}(x(t))\|^2$$

$$- qt\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t)-x^*\rangle - t^2\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), \dot{x}(t)\rangle$$

$$- \beta t^2\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), \nabla\Phi_{\lambda(t)}(x(t))\rangle,$$

where above we used (1). Consider now for every $t \geq t_0$,

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

$$+ 2q\beta t\langle\nabla\Phi_{\lambda(t)}(x(t)), x(t)-x^*\rangle + t^2\|\dot{x}(t)\|^2 + 2\beta t^2\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle$$

$$+ \beta^2 t^2\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.$$

The two estimates that we made above lead to (31) becoming

$$\frac{d}{dt}E_{p,q}(t) = t^p\left((p+2)tb(t) + t^2\dot{b}(t) + (p+1)\beta(\alpha-p-q-2)\right)\left(\Phi_{\lambda(t)}(x(t))-\Phi^*\right)$$

$$+ \frac{(p+2)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) + \frac{pq^2 t^{p-1}}{2}\|x(t)-x^*\|^2$$

$$+ \frac{(p+2)\beta^2 t^{p+1}}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 + \left(q+1-\alpha+\frac{p}{2}\right)t^{p+1}\|\dot{x}(t)\|^2$$

$$+ q(q+1-\alpha+p)t^p\langle\dot{x}(t), x(t)-x^*\rangle + q\beta(p+1)t^p\langle\nabla\Phi_{\lambda(t)}(x(t)), x(t)-x^*\rangle$$

$$- qt^{p+1}\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t)-x^*\rangle$$

$$- \beta t^{p+2}\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), \nabla\Phi_{\lambda(t)}(x(t))\rangle$$

$$- \frac{\dot{\lambda}(t)t^{p+1}\left(tb(t) + \beta(\alpha-p-q-2)\right)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.$$

Let us apply the gradient inequality to the strongly convex function $x \mapsto b(t)\Phi_{\lambda(t)}(x) + \frac{\varepsilon(t)\|x\|^2}{2}$:

$$- \langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t)-x^*\rangle + \frac{\varepsilon(t)\|x(t)-x^*\|^2}{2}$$

$$\leq \left(b(t)\Phi^* + \frac{\varepsilon(t)\|x^*\|^2}{2}\right) - \left(b(t)\Phi_{\lambda(t)}(x(t)) + \frac{\varepsilon(t)\|x(t)\|^2}{2}\right)$$

and thus

$$- qt^{p+1}\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t)-x^*\rangle \leq -qt^{p+1}b(t)\left(\Phi_{\lambda(t)}(x(t))-\Phi^*\right)$$

$$- qt^{p+1}\frac{\varepsilon(t)}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) - qt^{p+1}\frac{\varepsilon(t)\|x(t)-x^*\|^2}{2}$$

for every $t \geq t_0$. So, noticing that

$$
\begin{aligned}
& -\beta t^{p+2}\big\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle \\
= & -\beta t^{p+2}b(t)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 - \beta t^{p+2}\varepsilon(t)\big\langle x(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle
\end{aligned}
$$

we deduce

$$
\begin{aligned}
\frac{d}{dt}E_{p,q}(t) \leq\ & t^p\left((p+2-q)tb(t) + t^2\dot{b}(t) + (p+1)\beta(\alpha-p-q-2)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) \\
& + \frac{(p+2-q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) \\
& + \left(\frac{pq^2t^{p-1}}{2} - \frac{qt^{p+1}\varepsilon(t)}{2}\right)\|x(t) - x^*\|^2 \\
& + \frac{(p+2)\beta^2t^{p+1} - 2\beta t^{p+2}b(t) - \dot{\lambda}(t)t^{p+1}\left(tb(t)+\beta(\alpha-p-q-2)\right)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
& + \left(q+1-\alpha+\frac{p}{2}\right)t^{p+1}\|\dot{x}(t)\|^2 + q(q+1-\alpha+p)t^p\langle\dot{x}(t), x(t)-x^*\rangle \\
& + q\beta(p+1)t^p\langle\nabla\Phi_{\lambda(t)}(x(t)), x(t)-x^*\rangle - \beta t^{p+2}\varepsilon(t)\big\langle x(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle.
\end{aligned}
$$

In order to proceed further we will need the following estimates:

$$
\begin{aligned}
& q\beta(p+1)t^p\big\langle\nabla\Phi_{\lambda(t)}(x(t)), x(t)-x^*\big\rangle \\
\leq\ & \frac{q\beta(p+1)t^{p+1}}{4c^2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 + q\beta(p+1)c^2t^{p-1}\|x(t)-x^*\|^2
\end{aligned}
$$

and

$$
-\beta t^{p+2}\varepsilon(t)\big\langle x(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle \leq \frac{\beta t^{p+2}}{a}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 + \frac{a\beta t^{p+2}\varepsilon^2(t)}{4}\|x(t)\|^2
$$

for every $t \geq t_0$, some $c \geq 1$ and $a \geq 1$. Thus,

$$
\begin{aligned}
\frac{d}{dt}E_{p,q}(t) \leq\ & t^p\left((p+2-q)tb(t) + t^2\dot{b}(t) + (p+1)\beta(\alpha-p-q-2)\right)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) \\
& + \left(\frac{(p+2-q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2} + \frac{a\beta t^{p+2}\varepsilon^2(t)}{4}\right)\|x(t)\|^2 \\
& + \left(\frac{pq^2t^{p-1}}{2} - \frac{qt^{p+1}\varepsilon(t)}{2} + q\beta(p+1)c^2t^{p-1}\right)\|x(t)-x^*\|^2 \\
& + \left(\frac{(p+2)\beta^2t^{p+1} - 2\beta t^{p+2}b(t) - \dot{\lambda}(t)t^{p+1}\left(tb(t)+\beta(\alpha-p-q-2)\right)}{2} + \frac{q\beta(p+1)t^{p+1}}{4c^2}\right. \\
& + \left.\frac{\beta t^{p+2}}{a}\right)\cdot\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
& + \left(q+1-\alpha+\frac{p}{2}\right)t^{p+1}\|\dot{x}(t)\|^2 + q(q+1-\alpha+p)t^p\langle\dot{x}(t), x(t)-x^*\rangle \\
& - \left(\frac{(p+2-q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2}\right)\|x^*\|^2.
\end{aligned}
$$

Let us fix

$$q = \frac{2\alpha}{3} \text{ and } p = \frac{\alpha - 3}{3}.$$

First of all, due to this choice

$$q + 1 - \alpha + p = 0$$

and thus we get rid of the term $\langle \dot{x}(t), x(t) - x^* \rangle$. Secondly,

$$q + 1 - \alpha + \frac{p}{2} = -\frac{p}{2} \leq 0. \tag{32}$$

Then

$$p + 2 - q = 1 - \frac{\alpha}{3} \leq 0. \tag{33}$$

So,

$$\frac{d}{dt} E_{p,q}(t) \leq t^p \left((p + 2 - q)tb(t) + t^2 \dot{b}(t) + (p + 1)\beta(\alpha - p - q - 2) \right) \left(\Phi_{\lambda(t)}(x(t)) - \Phi^* \right)$$

$$+ \left(\frac{(p + 2 - q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2} + \frac{\alpha\beta t^{p+2}\varepsilon^2(t)}{4} \right) \|x(t)\|^2$$

$$+ \left(\frac{pq^2 t^{p-1}}{2} - \frac{qt^{p+1}\varepsilon(t)}{2} + q\beta(p+1)c^2 t^{p-1} \right) \|x(t) - x^*\|^2$$

$$+ \left(\frac{(p+2)\beta^2 t^{p+1} - 2\beta t^{p+2} b(t) - \dot{\lambda}(t)t^{p+1}\left(tb(t) + \beta(\alpha - p - q - 2)\right)}{2} + \frac{q\beta(p+1)t^{p+1}}{4c^2} \right.$$

$$+ \left. \frac{\beta t^{p+2}}{a} \right) \cdot \|\nabla\Phi_{\lambda(t)}(x(t))\|^2$$

$$+ \left(q + 1 - \alpha + \frac{p}{2} \right) t^{p+1} \|\dot{x}(t)\|^2 - \left(\frac{(p + 2 - q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2} \right) \|x^*\|^2.$$

Obviously, for t large enough, say, $t \geq t_2 \geq t_0$ the following expression is non-positive due to (27) and $p + 1 = \frac{\alpha}{3} > 0$ and $\alpha - p - q - 2 = -1$

$$(p + 2 - q)tb(t) + t^2 \dot{b}(t) + (p + 1)\beta(\alpha - p - q - 2)$$
$$= \left(1 - \frac{\alpha}{3}\right) tb(t) + t^2 \dot{b}(t) - \frac{\alpha\beta}{3} \leq 0.$$

Moreover, from (28) it follows that for $c = 1$

$$\frac{pq^2 t^{p-1}}{2} - \frac{qt^{p+1}\varepsilon(t)}{2} + q\beta(p+1)c^2 t^{p-1} = \frac{\alpha t^{\frac{\alpha - 6}{3}}}{27} \left(2\alpha(\alpha - 3) - 9t^2\varepsilon(t) + 6\alpha\beta\right) \leq 0$$

for all $t \geq t_0$. Furthermore,

$$\left(\frac{(p + 2 - q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2} + \frac{\alpha\beta t^{p+2}\varepsilon^2(t)}{4} \right) \|x(t)\|^2$$

$$-\left(\frac{(p+2-q)t^{p+1}\varepsilon(t)+t^{p+2}\dot{\varepsilon}(t)}{2}\right)\|x^*\|^2$$

$$=\left(\frac{(p+2-q)t^{p+1}\varepsilon(t)+t^{p+2}\dot{\varepsilon}(t)}{2}+\frac{a\beta t^{p+2}\varepsilon^2(t)}{4}\right)\left(\|x(t)\|^2-\|x^*\|^2\right)$$

$$+\frac{a\beta t^{p+2}\varepsilon^2(t)}{4}\|x^*\|^2.$$

So, under the assumption (12) and the fact that $\|x(t)\| \geq \|x^*\|$ for all $t \geq t_0$ we deduce due to (33)

$$\left(\frac{(p+2-q)t^{p+1}\varepsilon(t)+t^{p+2}\dot{\varepsilon}(t)}{2}+\frac{a\beta t^{p+2}\varepsilon^2(t)}{4}\right)\left(\|x(t)\|^2-\|x^*\|^2\right) \leq 0.$$

Thus, under the assumptions (12), (27), (28) and (29) (the latest leads to the non-positivity of the coefficient of $\|\nabla\Phi_\lambda(x)\|^2$) we conclude due to (32) that for every $t \geq t_2$

$$\frac{d}{dt}E_{p,q}(t) \leq \frac{a\beta t^{\frac{\alpha}{3}+1}\varepsilon^2(t)}{4}\|x^*\|^2. \tag{34}$$

(ii) Let us obtain now the lower bound for $E_{p,q}$. Notice that for $p = \frac{\alpha-3}{3}$ and $q = \frac{2\alpha}{3}$ we have $\alpha - p - q = 1$ and

$$E_{p,q}(t) \geq t^{p+1}\left(tb(t)+\beta(\alpha-p-q-2)\right)\left(\Phi_{\lambda(t)}(x(t))-\Phi^*\right)$$

$$+\frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2-\|x^*\|^2\right)$$

$$= t^{p+1}\left(tb(t)-\beta\right)\left(\Phi_{\lambda(t)}(x(t))-\Phi^*\right)+\frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2-\|x^*\|^2\right)$$

$$\geq \frac{t^{p+2}}{2}\left(\Phi_{\lambda(t)}(x(t))-\Phi^*\right)+\frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2-\|x^*\|^2\right), \tag{35}$$

since $tb(t) - \beta \geq \frac{t}{2}$ for every $t \geq t_0$ by $b(t_0) \geq \frac{1}{2} + \frac{\beta}{t_0}$ and b being non-decreasing. On the other hand, applying the gradient inequality to the strongly convex function $\varphi_{\varepsilon(t),\lambda(t)}(x) = \frac{\Phi_{\lambda(t)}(x)}{2} + \frac{\varepsilon(t)}{2}\|x\|^2$ we deduce for $x_{\varepsilon(t),\lambda(t)} = \operatorname{argmin}_H \varphi_{\varepsilon(t),\lambda(t)}(x)$

$$\varphi_{\varepsilon(t),\lambda(t)}(x) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) \geq \frac{\varepsilon(t)}{2}\|x - x_{\varepsilon(t),\lambda(t)}\|^2 \text{ for every } x \in H.$$

By the definition of $\varphi_{\varepsilon(t),\lambda(t)}(x)$ we deduce

$$\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \varphi_{\varepsilon(t),\lambda(t)}(x^*)$$

$$= \frac{1}{2}\left(\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \Phi^*\right) + \frac{\varepsilon(t)}{2}\left(\|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right)$$

$$\geq \frac{\varepsilon(t)}{2}\left(\|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right).$$

We may now add the last two inequalities to obtain

$$\varphi_{\varepsilon(t),\lambda(t)}(x) - \varphi_{\varepsilon(t),\lambda(t)}(x^*)$$
$$\geq \frac{\varepsilon(t)}{2}\left(\|x - x_{\varepsilon(t),\lambda(t)}\|^2 + \|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right) \text{ for every } x \in H. \quad (36)$$

Plugging (36) into (35) we conclude that for every $t \geq t_2$

$$E_{p,q}(t) \geq \frac{t^{p+2}\varepsilon(t)}{2}\left(\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 + \|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right). \quad (37)$$

(iii) Finally, using the lower and upper bounds for $E_{p,q}$ we can prove the strong convergence of the trajectories to a minimal norm solution. Integrating (34) on $[t_2, t]$ we obtain

$$E_{p,q}(t) \leq E_{p,q}(t_2) + \frac{a\beta\|x^*\|^2}{4}\int_{t_2}^t s^{\frac{\alpha}{3}+1}\varepsilon^2(s)ds$$

and using (37) we deduce for every $t \geq t_2$

$$\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \leq \|x^*\|^2 - \|x_{\varepsilon(t),\lambda(t)}\|^2 + \frac{2E_{p,q}(t_2)}{t^{\frac{\alpha}{3}+1}\varepsilon(t)} + \frac{a\beta\|x^*\|^2}{2t^{\frac{\alpha}{3}+1}\varepsilon(t)}\int_{t_2}^t s^{\frac{\alpha}{3}+1}\varepsilon^2(s)ds.$$

Note that due to (28)

$$t^2\varepsilon(t) \geq \frac{2\alpha(\alpha-3)+6\alpha\beta}{9} = \hat{C} \geq 0$$

and

$$t^{\frac{\alpha}{3}+1}\varepsilon(t) = t^2\varepsilon(t)t^{\frac{\alpha}{3}-1} \geq \hat{C}t^{\frac{\alpha}{3}-1}.$$

Since $\alpha > 3$ we deduce

$$\lim_{t\to+\infty} t^{\frac{\alpha}{3}+1}\varepsilon(t) = +\infty$$

and thus

$$\lim_{t\to+\infty} \frac{2E_{p,q}(t_2)}{t^{\frac{\alpha}{3}+1}\varepsilon(t)} = 0.$$

Finally, by (9) and (30) we conclude

$$\lim_{t\to+\infty} x(t) = x^*.$$

Case II.

Assume now the opposite to the first case, namely, $\|x(t)\| < \|x^*\|$ for every $t \geq t_0$. According to Theorem 6

$$\lim_{t \to +\infty} \| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t) \| = 0$$

and

$$\lim_{t \to +\infty} \Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* = 0.$$

Denote $\xi(t) = \operatorname{prox}_{\lambda(t)\Phi}(x(t))$. Considering a sequence $\{t_k\}_{k \in \mathbb{N}}$ such that $\{x(t_k)\}_{k \in \mathbb{N}}$ converges weakly to an element $\hat{x} \in H$ as $k \to \infty$, we notice that $\{\xi(t_k)\}_{k \in \mathbb{N}}$ converges weakly to $\hat{x}$ as $k \to \infty$. Now, the function Φ being convex and lower semicontinuous in the weak topology, allows us to write

$$\Phi(\hat{x}) \leq \liminf_{k \to \infty} \Phi(\xi(t_k)) = \lim_{t \to +\infty} \Phi(\xi(t)) = \Phi^*$$

and hence, $\hat{x} \in \operatorname{argmin} \Phi$. The norm is weakly semicontinuous, so

$$\|\hat{x}\| \leq \liminf_{k \to \infty} \|\xi(t_k)\| \leq \|x^*\|,$$

which means that $\hat{x} = x^*$ by the uniqueness of the element of the minimum norm in $\operatorname{argmin} \Phi_\lambda$. Therefore, the trajectory x converges weakly to x^* and

$$\|x^*\| \leq \liminf_{t \to +\infty} \|x(t)\| \leq \limsup_{t \to +\infty} \|x(t)\| \leq \|x^*\|$$

and thus

$$\lim_{t \to +\infty} \|x(t)\| = \|x^*\|.$$

From this and the weak convergence of the trajectory x follows the strong one: $\lim_{t \to +\infty} x(t) = x^*$.

Case III.

Assume that for $t \geq t_0$ the trajectory x finds itself both inside and outside the ball $B(0, \|x^*\|)$. Since x is continuous, there exists a sequence $\{t_n\}_{n \in \mathbb{N}} \subseteq [t_0, +\infty)$ such that $t_n \to \infty$ as $n \to \infty$ and $\|x(t_n)\| = \|x^*\|$ for every $n \in \mathbb{N}$. Consider again a weak sequential cluster point $\hat{x}$ of the sequence $\{x(t_n)\}_{n \in \mathbb{N}}$. By repeating the same argument as in the previous case we deduce the weak convergence of $\{x(t_n)\}_{n \in \mathbb{N}}$ to x^*, as $n \to \infty$. Since $\|x(t_n)\| \to \|x^*\|$, as $n \to \infty$, we obtain that $\|x(t_n) - x^*\| \to 0$, as $n \to \infty$, which means $\liminf_{t \to +\infty} \|x(t) - x^*\| = 0$. $\qquad\square$

Remark 1 In this section the condition $\dot{b}(t) \geq 0$ for all $t \geq t_0$ is not necessary. Our conjecture is that we can weaken the setting by omitting this condition and thus widen the range for b, including the functions that decay not faster than $\frac{1}{t^2}$ for the polynomial choice of parameters.

Remark 2 There is no setting which guarantees both fast rates for the values and strong convergence of the trajectories. One of the future goal would be to develop a new approach (based on [6]), which would help us deduce these two results simultaneously.

4.1 Strong Convergence of the Tajectories in Cse $\alpha = 3$

Throughout this section we no longer require that b is non-decreasing. In this case the analogue of Theorem 6 looks as follows.

Theorem 8 *Suppose that for all $t \geq t_0$ the function λ is bounded, $b(t) \equiv b > 0$ is a constant function and (12) and (14) hold. Suppose additionally that (25) holds for constant b, namely*

$$\int_{t_0}^{+\infty} \frac{\varepsilon(t)}{t} dt \; < \; +\infty.$$

Then

$$\lim_{t \to +\infty} \| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t) \| \; = \; 0$$

and

$$\lim_{t \to +\infty} \Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* \; = \; 0.$$

Proof In this case the energy functional becomes

$$E_2(t) = (bt^2 - \beta t)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + \frac{t^2 \varepsilon(t)}{2} \|x(t) + \frac{1}{2}\|2(x(t) - x^*)$$
$$+ t\left(\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\right)\|^2.$$

Relation (16) thus becomes for all $t \geq t_0$

$$\dot{E}_2(t) \leq \beta\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) - \left(\left(bt^2 - \beta t\right)\frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2\left(b - \frac{1}{a}\right)\right)\|\nabla \Phi_{\lambda(t)}(x(t))\|^2$$
$$+ \frac{2t^2 \dot{\varepsilon}(t) + a\beta t^2 \varepsilon^2(t)}{4}\|x(t)\|^2 - t\varepsilon(t)\|x(t) - x^*\|^2 + t\varepsilon(t)\|x^*\|^2.$$

Thus, repeating the same arguments as in Theorem 4 we obtain

$$\dot{E}_2(t) \; \leq \; \beta\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + t\varepsilon(t)\|x^*\|^2.$$

Let us multiply this expression with $t(bt - \beta)$ to obtain

$$t(bt - \beta)\dot{E}_2(t) \leq \beta t(bt - \beta)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right) + t^2(bt - \beta)\varepsilon(t)\|x^*\|^2$$
$$\leq \beta E_2(t) + t^2(bt - \beta)\varepsilon(t)\|x^*\|^2.$$

Now, we will divide by $(bt - \beta)^2$ to conclude

$$\frac{t}{(bt - \beta)}\dot{E}_2(t) \leq \frac{\beta}{(bt - \beta)^2}E_2(t) + \frac{t^2}{(bt - \beta)}\varepsilon(t)\|x^*\|^2$$

or

$$\frac{d}{dt}\left(\frac{t}{bt - \beta}E_2(t)\right) \leq \frac{t^2}{(bt - \beta)}\varepsilon(t)\|x^*\|^2.$$

Integrating the last inequality on $[T, t]$, where $T \geq t_0$, we deduce

$$\frac{t}{bt - \beta}E_2(t) \leq \frac{T}{bT - \beta}E_2(T) + \|x^*\|^2\int_T^t \frac{s^2}{(bs - \beta)}\varepsilon(s)ds.$$

By the definition of E_2 we know

$$E_2(t) \geq (bt^2 - \beta t)\left(\Phi_{\lambda(t)}(x(t)) - \Phi^*\right).$$

Combining these two inequalities, we deduce

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* \leq \frac{T}{t^2(bT - \beta)}E_2(T) + \frac{\|x^*\|^2}{t^2}\int_T^t \frac{s^2}{(bs - \beta)}\varepsilon(s)ds.$$

Now,

$$\lim_{t \to +\infty} \frac{T}{t^2(bT - \beta)}E_2(T) = 0.$$

Applying Lemma A.3 we deduce due to (25)

$$\lim_{t \to +\infty} \frac{bt - \beta}{t^3}\int_T^t \frac{s^3}{(bs - \beta)}\frac{\varepsilon(s)}{s}ds = 0$$

and thus

$$\lim_{t \to +\infty} \frac{\|x^*\|^2}{t^2}\int_T^t \frac{s^2}{(bs - \beta)}\varepsilon(s)ds = 0.$$

Therefore, we establish

$$\lim_{t \to +\infty} \Phi_{\lambda(t)}(x(t)) - \Phi^* = 0.$$

Again, by the definition of the proximal mapping

$$\Phi_{\lambda(t)}(x(t)) - \Phi^* = \Phi\left(\mathrm{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* + \frac{1}{2\lambda(t)}\|\mathrm{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \quad \forall t \geq t_0.$$

Using the fact that λ is bounded for all $t \geq t_0$ we deduce

$$\lim_{t\to+\infty} \|\mathrm{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\| = 0$$

and

$$\lim_{t\to+\infty} \Phi\left(\mathrm{prox}_{\lambda(t)\Phi}(x(t))\right) - \Phi^* = 0.$$

□

We are in position now to formulate the analogue of Theorem 7.

Theorem 9 *Suppose that λ is bounded for all $t \geq t_0$, $b(t) \equiv b \geq \frac{1}{2} + \frac{\beta}{t_0}$ and (12) and (25) hold. Assume, in addition, that*

$$\lim_{t\to+\infty} t^2\varepsilon(t) = +\infty, \tag{38}$$

$$2\beta t + \beta\dot{\lambda}(t) - bt\left(\dot{\lambda}(t) + 2\beta\right) + 2\beta^2 + \beta \leq 0 \text{ for all } t \geq t_0 \tag{39}$$

and

$$\lim_{t\to+\infty} \frac{\beta}{t^2\varepsilon(t)} \int_{t_0}^t s^2\varepsilon^2(s)ds = 0. \tag{40}$$

If $x : [t_0, +\infty) \mapsto H$ is a solution to (1) and the trajectory $x(t)$ stays either inside or outside the ball $B(0, \|x^*\|)$, then $x(t)$ converges to minimal norm solution $x^* = \mathrm{proj}_{\mathrm{argmin}\,\Phi}(0)$, as $t \to +\infty$. Otherwise, $\liminf_{t\to+\infty}\|x(t) - x^*\| = 0$.

Proof The proof goes in line with the one of Theorem 7 by taking $\alpha = 3, b(t) \equiv b > 0$, $q = 2, p = 0$ and referring to Theorem 8 instead of Theorem 6 in the second and third cases. □

5 Analysis of the Conditions

Since all the conditions cannot be satisfied simultaneously, let us treat them separately, namely:

1. In order to obtain the fast convergence rates of the function values we require that for all $t \geq t_0$:

    - $\alpha > 3$;
    - the existence of $a \geq 1$ such that $2\dot{\varepsilon}(t) \leq -a\beta\varepsilon^2(t)$,

- $b(t_0) \geq \frac{\beta}{t_0}$ and $b(t_0) > \frac{1}{a}$;
- $\int_{t_0}^{+\infty} t\varepsilon(t)dt < +\infty$ and
- the existence of $0 < \delta < \alpha - 3$ such that $(\alpha - 3)tb(t) - t^2\dot{b}(t) + \beta(2 - \alpha) \geq \delta tb(t)$.

2. For the strong convergence of the trajectories we require the following for all $t \geq t_0$:

- $\alpha > 3$;
- λ is bounded;
- $\frac{\alpha - 3}{3}b(t) - t\dot{b}(t) + \frac{\alpha\beta}{3} \geq 0$;
- $(\alpha - 3)tb(t) - t^2\dot{b}(t) + \beta(2 - \alpha) \geq 0$;
- the existence of $a \geq 1$ such that $2\dot{\varepsilon}(t) \leq -a\beta\varepsilon^2(t)$, $b(t_0) > \frac{1}{a}$ and $b(t_0) \geq \frac{1}{2} + \frac{\beta}{t_0}$;
- $\int_{t_0}^{+\infty} \frac{\varepsilon(t)}{tb(t)}dt < +\infty$;
- $2\alpha(\alpha - 3) - 9t^2\varepsilon(t) + 6\alpha\beta \leq 0$;
- $18\beta t + 9\beta\dot{\lambda}(t) - 9tb(t)\left(\dot{\lambda}(t) + 2\beta\right) + 3(\alpha + 3)\beta^2 + \alpha^2\beta \leq 0$;
- $\lim_{t \to +\infty} \frac{\beta}{t^{\frac{\alpha}{3}+1}\varepsilon(t)} \int_{t_0}^{t} s^{\frac{\alpha}{3}+1}\varepsilon^2(s)ds = 0$.

We will analyse these conditions in details for the polynomial choice of functions b and ε, namely, $b(t) = bt^n$ and $\varepsilon(t) = \frac{\varepsilon}{t^d}$, where b is positive, $n \geq 0$ and $\varepsilon, d > 0$.

5.1 Setting for the Fast Convergence Rates of the Function Values

The set of the conditions becomes for all $t \geq t_0$

1. $\alpha > 3$;
2. there exists $a \geq 1$ such that $-\frac{2d\varepsilon}{t^{d+1}} \leq -\frac{a\beta\varepsilon^2}{t^{2d}}$,
3. $b(t_0) \geq \frac{\beta}{t_0}$ and $b(t_0) > \frac{1}{a}$;
4. $\int_{t_0}^{+\infty} \frac{\varepsilon}{t^{d-1}}dt < +\infty$ and
5. there exists $0 < \delta < \alpha - 3$ such that $(\alpha - 3)bt^{n+1} - bnt^{n+1} + \beta(2 - \alpha) \geq \delta bt^{n+1}$.

After some simple algebraic computations one may discover that in order to satisfy all the conditions at the same time it is enough to assume

$$\alpha - 3 > n \geq 0 \text{ (condition 5)}$$

and

$$d > 2 \text{ and } d \geq \frac{\beta\varepsilon}{2} \text{ (conditions 2 and 4)},$$

since all the other inequalities could be fulfilled by taking the appropriate t_0, namely,

$$t_0 \geq \max\left\{\sqrt[n+1]{\frac{\beta}{b}}, \sqrt[n+1]{\frac{\beta(\alpha - 2)}{b(\alpha - 3 - n)}}\right\} \text{ and } t_0 > \frac{1}{\sqrt[n]{b}}.$$

5.2 Setting for the Strong Convergence of the Trajectories

The set of the conditions becomes for all $t \geq t_0$

1. $\alpha > 3$;
2. λ is bounded;
3. $\frac{\alpha-3}{3} b t^{n+1} - bn t^{n+1} + \frac{\alpha\beta}{3} \geq 0$;
4. $(\alpha - 3) b t^{n+1} - bn t^{n+1} + \beta(2 - \alpha) \geq 0$;
5. there exists $a \geq 1$ such that $-\frac{2d\varepsilon}{t^{d+1}} \leq -\frac{a\beta\varepsilon^2}{t^{2d}}$, $b(t_0) > \frac{1}{a}$ and $b(t_0) \geq \frac{1}{2} + \frac{\beta}{t_0}$;
6. $\int_{t_0}^{+\infty} \frac{\varepsilon}{b t^{n+d+1}} dt < +\infty$;
7. $2\alpha(\alpha - 3) - \frac{9\varepsilon}{t^{d-2}} + 6\alpha\beta \leq 0$;
8. $18\beta t + 9\beta\dot{\lambda}(t) - 9b t^{n+1}\left(\dot{\lambda}(t) + 2\beta\right) + 3(\alpha + 3)\beta^2 + \alpha^2\beta \leq 0$;
9. $\lim_{t \to +\infty} \frac{\beta}{\varepsilon t^{\frac{\alpha}{3}-d+1}} \int_{t_0}^{t} \varepsilon^2 s^{\frac{\alpha}{3}-2d+1} ds = 0$.

Again, analysis of the set of conditions leads to the following conclusion:

- λ is bounded (condition 2) ;
- $0 \leq n \leq \frac{\alpha-3}{3}$ and $\alpha > 3$ (condition 3) ;
- $\max\left\{1, \frac{\beta\varepsilon}{2}\right\} \leq d \leq 2$ (conditions 5, 7, 8, 9) .

As before, t_0 should be chosen appropriately.

5.3 The Case $\alpha = 3$

In this case the following has to be assumed: there exists $a \geq 1$ such that for all $t \geq t_0$

1. $\lambda(t)$ is bounded;
2. $2\dot{\varepsilon}(t) \leq -a\beta\varepsilon^2(t)$, $b > \frac{1}{a}$ and $b \geq \frac{1}{2} + \frac{\beta}{t_0}$;
3. $\int_{t_0}^{+\infty} \frac{\varepsilon(t)}{t} dt < +\infty$;
4. $\lim_{t \to +\infty} t^2 \varepsilon(t) = +\infty$;
5. $2\beta t + \beta\dot{\lambda}(t) - bt\left(\dot{\lambda}(t) + 2\beta\right) + 2\beta^2 + \beta \leq 0$;
6. $\lim_{t \to +\infty} \frac{\beta}{t^2 \varepsilon(t)} \int_{t_0}^{t} s^2 \varepsilon^2(s) ds = 0$.

Essentially, for the polynomial choice of parameters that means $b \geq 1$ and

- $\lambda(t)$ is bounded (condition 2) ;
- $\max\left\{1, \frac{\beta\varepsilon}{2}\right\} \leq d < 2$ (conditions 4, 5, 6) ,

so with the appropriate choice of t_0 the whole set of conditions is fulfilled.

6 Numerical Examples

6.1 The Rates of Convergence of the Moreau Envelope Values

Consider the objective function $\Phi : \mathbb{R} \to \mathbb{R}$, $\Phi(x) = |x| + \frac{x^2}{2}$ and let us plot the values of its Moreau envelope as well as the gradient of its Moreau envelope for
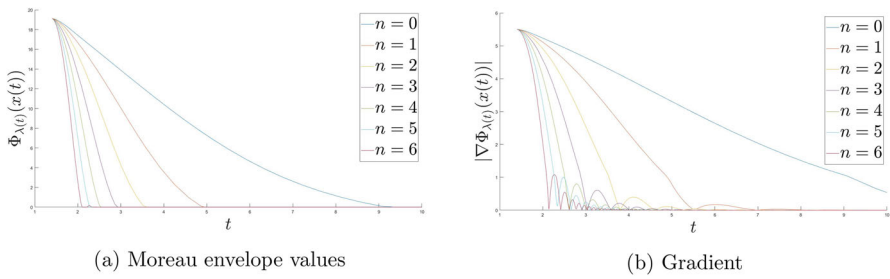
(a) Moreau envelope values   (b) Gradient
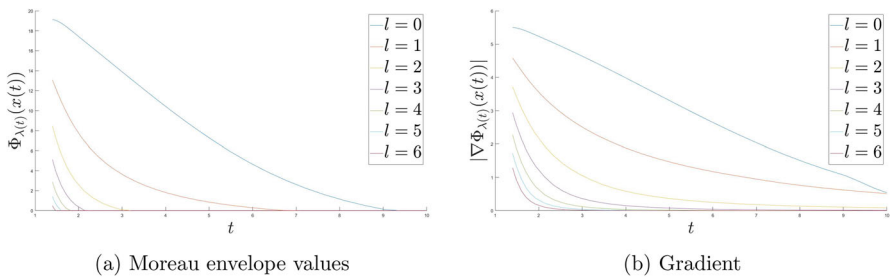
Fig. 1 $l = 0$ and $d = 3$

(a) Moreau envelope values   (b) Gradient

Fig. 2 $d = 3$ and $n = 0$

different polynomial functions λ, ε and b to illustrate the theoretical results with some numerical examples. We take $\lambda(t) = t^l$, $\varepsilon(t) = \frac{1}{t^d}$, $b(t) = t^n$ with $x(t_0) = x_0 = 10$, $\dot{x}(t_0) = 0$, $\alpha = 10$ and $t_0 = 1.4$.

First, let us take different time scaling parameter b with $l = 0$ and $d = 3$ and see how it affects the behaviour of the system (1) (see Fig. 1).

As expected, the faster b grows, the faster the convergence is.

Consider now different Moreau envelope parameter λ with $d = 3$ and $n = 0$ (see Fig. 2).

Note that the difference in the starting point comes from the fact that $t_0 \neq 1$, and for different exponents l the value t_0^l is also different. As predicted by theory, a faster growing function λ leads to faster convergence of not only the gradient of Moreau envelope of the objective function Φ, but also of the values of the Moreau envelope themselves.

Varying the Tikhonov function ε for $n = 0$ and $l = 0$ does not affect the system, which is illustrated by the following plot (see Fig. 3).

(a) Moreau envelope values   (b) Gradient

Fig. 3 $n = 0$ and $l = 0$

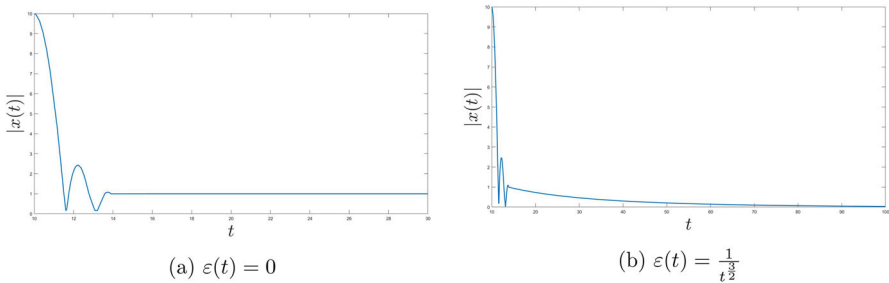(a) $\varepsilon(t) = 0$ (b) $\varepsilon(t) = \frac{1}{t^{\frac{3}{2}}}$
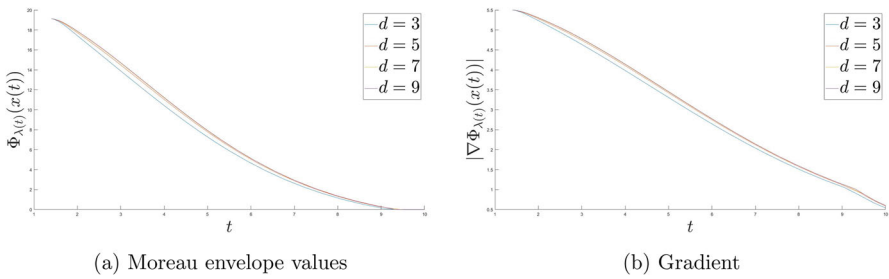
Fig. 4 The role of the Tikhonov term

6.2 Strong Convergence of the Trajectories

For a different objective function let us investigate the strong convergence of the trajectories of (1):

$$\Phi(x) = \begin{cases} |x - 1|, & x > 1 \\ 0, & x \in [-1, 1] \\ |x + 1|, & x < -1. \end{cases}$$

The set argmin Φ is nothing but the segment $[-1, 1]$ and 0 is its element of minimal norm. Let us fix $\alpha = 6$ and $n = 0.7$. First we take constant lambda ($\lambda(t) = 1$ for all $t \geq t_0$) and plot the behaviour of the trajectories of (1) with and without Tikhonov term (see Fig. 4.

As we see in case there is no Tikhonov regularization the trajectories converge to the minimizer 1 of Φ, but the Tikhonov term actually guarantees the convergence towards the minimal norm solution, which is 0.

Another comparison was made for non-constant lambda: $\lambda(t) = 1 - \frac{1}{t^l}$ for $l = 1$ (for different l's the picture is the same), illustrating similar behaviour (see Fig. 5).

Finally, for the same choice of λ let us take different Tikhonov terms to figure out how changing them affects the trajectories of (1) (see Fig. 6).

We see, that the faster ε decays, the slower trajectories converge.

(a) $\varepsilon(t) = 0$ (b) $\varepsilon(t) = \frac{1}{t^{\frac{3}{2}}}$
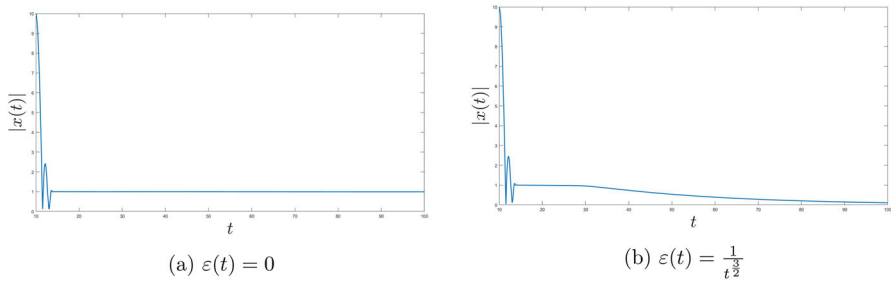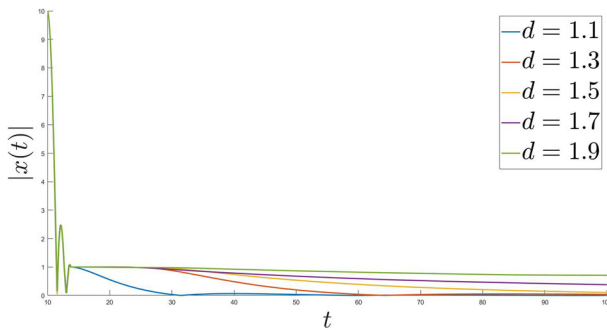
Fig. 5 The role of the Tikhonov term

Fig. 6 $n = 0.7$ and $\lambda(t) = 1 - \frac{1}{t}$

Appendix A

Let us state here some auxiliary lemmas which we used in our analysis. For the proof of the following lemma we refer to [3].

Lemma A.1 *Suppose that $f : [t_0, +\infty) \to \mathbb{R}$ is locally absolutely continuous and bounded from below and there exists $g \in L^1([t_0, +\infty), \mathbb{R})$ such that for almost all $t \geq t_0$*

$$\frac{d}{dt} f(t) \leq g(t).$$

Then there exists $\lim_{t \to +\infty} f(t) \in \mathbb{R}$.

For the proof of the next lemma we refer to [19].

Lemma A.2 *Let H be a real Hilbert space and $x : [t_0, +\infty) \mapsto \mathbb{H}$ be a continuously differentiable function satisfying $x(t) + \frac{t}{\alpha} \dot{x}(t) \to L$ as $t \to +\infty$, with $\alpha > 0$ and $L \in \mathbb{H}$. Then $x(t) \to L$ as $t \to +\infty$.*

For the proof of the final Lemma we refer to [9].

Lemma A.3 *Let $\delta > 0$ and $f \in L^1((\delta, +\infty), \mathbb{R})$ be a non-negative and continuous function. Let $g : [\delta, +\infty) \to [0, +\infty)$ be a non-decreasing function such that $\lim_{t \to +\infty} g(t) = +\infty$. Then it holds*

$$\lim_{t \to +\infty} \frac{1}{g(t)} \int_{\delta}^{t} g(s) f(s) ds = 0.$$

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declaration

Competing interests The authors declare that they have no competing interests subject to the topic of this article.

References

1. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. J. de Mathématiques Pures et Appliquées **81**(8), 747–779 (2002)
2. Attouch, H.: Viscosity solutions of minimization problems. SIAM J. Optim. **6**(3), 769–806 (1996)
3. Attouch, H., Abbas, B., 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)
4. 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)
5. Attouch, H., Balhag, A., Chbani, Z., Riahi, H.: Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling. Evol. Equ. Control Theory **11**(2), 487–514 (2022)
6. Attouch, H., Balhag, A., Chbani, Z., Riahi, H.: Accelerated gradient methods combining Tikhonov regularization with geometric damping driven by the Hessian. Appl. Math. Optim. **88**(29), (2023)
7. Attouch, H., Cabot, A.: Convergence of damped inertial dynamics governed by regularized maximally monotone operators. J. Differ. Equ. **264**, 7138–7182 (2018)
8. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**, 123–175 (2018)
9. Attouch, H., Chbani, Z., Riahi, H.: Combining fast inertial dynamics for convex optimization with Tikhonov regularization. J. Math. Anal. Appl. **457**(2), 1065–1094 (2018)
10. Attouch, H., Chbani, Z., Riahi, H.: Fast proximal methods via time scaling of damped inertial dynamics. SIAM J. Optim. **29**(3), 2227–2256 (2019)

11. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via time scaling of damped inertial gradient dynamics. Pure Appl. Funct. Anal. **6**(6), 1081–1117 (2021)
12. Attouch, H., Chbani, Z., Riahi, H.: Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization (2022). arXiv:2211.10140v1
13. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping. J. Math. Program. Oper. Res. **72**(5), (2023)
14. 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)
15. Attouch, H., Czarnecki, M.-O.: Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria. J. Differ. Equ. **179**, 278–310 (2002)
16. Attouch, H., László, S.C.: Continuous Newton-like inertial dynamics for monotone inclusions. Set-valued Variat. Anal. **29**, 555–581 (2021)
17. Attouch, H., László, S. C.: *Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution* (2021). arXiv:2104.11987
18. Attouch, H., Peypouquet, J.: Convergence of the inertial dynamics and proximal algorithms governed by maximally monotone operators. Math. Program. **174**, 391–432 (2019)
19. Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping damping. J. Differ. Equ. **261**(10), 5734–5783 (2016)
20. Attouch, H., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. Math. Program. **168**, 123–175 (2018)
21. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics, Springer (2016)
22. Boţ, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control. Optim. **54**(3), 1423–1443 (2016)
23. Boţ, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian driven damping. Math. Program. **189**, 151–186 (2021)
24. Boţ, R. I., Karapetyants, M.A.: A fast continuous time approach with time scaling for nonsmooth convex optimization. Adv. Contin. Discrete Models: Theory Appl. **73** (2022)
25. Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation and insights. Trans. Am. Math. Soc. **361**, 5983–6017 (2009)
26. Cabot, A., Engler, H., Gadat, S.: Second order differential equations with asymptotically small dissipation and piecewise flat potentials. Electron. J. Differ. Equ. **17**, 33–38 (2009)
27. László, S.C.: On the strong convergence of the trajectories of a Tikhonov regularized second order dynamical system with asymptotically vanishing damping. J. Differ. Equ. **362**, 355–381 (2023)
28. May, R.: Asymptotic for a second-order evolution equation with convex potential and vanishing damping term. Turk. J. Math. **41**(3), 681–685 (2017)
29. Sell, G.R.: Dynamics of Evolutionary Equations. Springer, New York (2002)
30. 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**, 1–43 (2016)