DISSERTATION | DOCTORAL THESIS

Titel | Title

Fast continuous and discrete time approaches for smooth and nonsmooth optimization featuring time scaling and Tikhonov regularization

verfasst von | submitted by

Mikhail Karapetyants

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doktor der Naturwissenschaften (Dr.rer.nat.)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 796 605 405
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Mathematik
Betreut von | Supervisor:	Univ.-Prof. Dr. Radu Ioan Bot Privatdoz.

Acknowledgements

I would like to take the opportunity to express my gratitude to all the people who contributed, directly and indirectly, to this thesis. First and foremost I would like to thank my supervisor Professor Radu Ioan Boţ for guiding me through my first steps in the field of optimization and being a beacon of knowledge in the infinite sea of mathematics. I would also like to thank Professor Szilárd Csaba László, who was hosting my research stay in Cluj-Napoca, Romania, showing exceptional hospitality and unlimited support, and Dr. habil. Ernö Robert Csetnek, who was always ready to share his exceptional knowledge of mathematics with others. I am very grateful to Professor Jalal Fadili and Professor Zaiwen Wen for agreeing to review this PhD thesis.

This research was financially supported by the Doctoral Programme Vienna Graduate School on Computational Optimization (VGSCO) which is funded by FWF (Austrian Science Fund). I would also like to thank the Vienna School of Mathematics and the VGSCO for all the generous resources and opportunities provided, such as multiple educational courses, summer schools etc.

I should also mention my wonderful colleagues, whom I was lucky to work with. It is my pleasure to thank Alexander, Bo, Buris, Chiara, David, Daniel, Egor, Giancarlo, Khoa, Lev, Matúš, Roman, Rossen, Sorin and acknowledge their positive contribution to my life. Especially, I would like to extend my thankfulness to Axel, Bettina and Michael for helping me during my first days in Austria and showing me how things here work. I very much appreciate our scientific discussions, the evenings after work, which we spent together, and I hope to carry this friendship through the ages.

Last, but not least, I want to thank my family and friends for their constant support – without you all of this would not have been possible. I would like to thank my father, who has always been a role model for me, my mother, whose love and kindness always help people around to get through difficult times, my brother, who I can always rely on, and, of course, my wife, who was the light of my life in the darkest of hours. I also would like to thank my friend Vladimir – our discussions on mathematics and physics inspired many new ideas and expanded my scientific horizons, and my friend Andres, who was always there for me.

Finally, many thanks to our cat, Persik, who keeps bringing a lot of joy in the lives of those around him.

Danke vielmals!

Abstrakt

Die folgende Arbeit widmet sich der Lösung des klassischen Optimierungsproblems der Minimierung einer konvexen Funktion mit verschiedenen Methoden in kontinuierlicher und diskreter Zeit. Es werden sowohl glatte als auch nicht glatte Fälle der Zielfunktionen diskutiert. Der Schwerpunkt dieser Abhandlung liegt auf der Erzielung schneller Konvergenzraten für die Funktionswerte in Kombination mit der Konvergenz der Trajektorien der dynamischen Systeme oder Iterationen der Algorithmen. Besonderes Augenmerk wird auf die Tikhonov-Regularisierungstechnik gelegt, die bekanntermaßen dabei hilft, die schwache Konvergenz der Trajektorien (Iterationen) zur kontrollierten (zum Element der minimalen Norm aus der Menge aller Minimierer der Zielfunktion) starken Konvergenz zu verbessern. In einigen Fällen wird eine Zeitskalierung angewendet, um die Konvergenz der Funktionswerte zu beschleunigen. Es werden verschiedene numerische Experimente vorgestellt, um unser Verständnis der theoretischen Ergebnisse zu verbessern.

Abstract

The following thesis is devoted to solving the classical optimization problem of minimizing a convex function via different methods in continuous and discrete time. Both smooth and nonsmooth cases of the objective functions will be discussed. The main focus throughout this manuscript is to obtain fast convergence rates for the function values combined with the convergence of the trajectories of the dynamical systems or iterates of the algorithms. Major attention is paid to the Tikhonov regularization technique, which is known to help improve the weak convergence of the trajectories (iterates) to the controlled (to the element of the minimal norm from the set of all minimizers of the objective function) strong one. In some cases, time scaling will be applied to accelerate the function values' convergence. Various numerical experiments are presented to improve our understanding of theoretical results.

Key words: Smooth and nonsmooth convex optimization; Damped inertial dynamics; Hessian-driven damping; Time scaling; Moreau envelope; Proximal operator; Tikhonov regularization; Strong convergence; Inertial algorithms.

Contents

1 Introduction

Let H be a real Hilbert space endowed with the scalar product $\langle\cdot,\cdot\rangle$ and norm $\|x\| = \sqrt{\langle x, x\rangle}$ for $x \in H$. Throughout the whole manuscript we will be solving the following optimization problem

$$\min_{x \in H} \Phi(x), \tag{1.1}$$

where $\Phi : H \mapsto \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ is a proper, convex and lower semicontinuous function, and we assume that $\operatorname{argmin} \Phi$, the set of global minimizers of Φ, is not empty. We will denote by $\min \Phi$ the global minimum of Φ and by x^* – a minimizer of Φ, unless other is specified. As it was already mentioned, our main goal is to develop continuous and discrete time approaches in order to obtain fast convergence rates for the function values to its minimum as well as weak convergence of the trajectories (iterates, depending on the approach which we will use) to an element from the set of function's minimizers, or strong convergence of the latter to the minimizer of the smallest norm in case of Tikhonov regularization. Before we turn our attention to the main subject of this thesis, we would like to provide some historical background, state-of-the-art results and preliminary facts, which we will use later in our analysis.

1.1 Summary

In a Hilbert setting we address the classical optimization problem of minimizing a convex function via different methods and techniques, such as binding the gradient of the objective function (or the gradient of its Moreau envelope in case of nonsmooth objective functions) to the second order in time dynamical system and studying its properties for continuous time approaches or via the inertial algorithms for discrete time approaches. In both cases we are interested in showing fast convergence of the function values to its minimum as well as weak convergence of the trajectories (iterates) to a minimizer from the set of all minimizers of the objective function. In case we apply Tikhonov regularization we aim to show the strong convergence of the trajectories (iterates) to the minimizer of the minimal norm. First of all, we will study the convergence properties of a second order in time dynamical system combining viscous and Hessian-driven damping with time scaling in relation with the minimization of a proper, convex and lower semicontinuous function. The system is formulated in terms of the gradient of the Moreau envelope of the objective function with time-dependent parameter. We will show fast convergence rates for the Moreau envelope and its gradient along the trajectory, and also for the velocity of the trajectory of the system. From here we will derive fast convergence rates for the objective function along a path which is the image of the trajectory of the system through the proximal operator of the first. Moreover, we will prove the weak convergence of the trajectory of the system to a global minimizer of the objective function. As a second step, we will study a second order in time differential equation, again combining viscous and Hessian-driven damping, containing a time scaling parameter function and a Tikhonov regularization term. This dynamical system is also related to the problem of minimization of a proper, convex and lower semicontinuous function. In the formulation of the problem as well as in our analysis we will use the Moreau envelope of the objective function and its gradient and will heavily rely on their properties. We will show that 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. Additionally, we will prove strong convergence of the trajectories to the element of minimal norm from the set of all minimizers of the objective. Further, we would like to develop the approach which allows us to obtain both fast convergence of the function values and strong convergence of the trajectories at the same time. So, addressing once more the classical optimization problem of minimizing a proper, convex and lower semicontinuous function via the second order in time dynamics, combining viscous and Hessian-driven damping with a Tikhonov regularization term, we will develop a technique, which at the same time guarantees the fast convergence of the function (and Moreau envelope) values and strong convergence of the trajectories of the system to a minimal norm solution – the element of the minimal norm from all the minimizers of the objective. In our analysis we will once again heavily exploit the Moreau envelope of the objective function and its properties as well as Tikhonov regularization properties, which we previously extended to the nonsmooth case. Moreover, we will deduce the precise rates of convergence of the values for the particular choice of parameters. Here we should mention that the discretization of all the above systems is possible (we will make more precise remarks later on), however they lead to the so-called proximal algorithms due to the presence of Moreau envelope. To avoid this, we would like to switch to smooth objective functions and address the discrete counterpart of the formulated above problems, investigating the strong convergence properties of a Nesterov type algorithm with a Tikhonov regularization term in connection to the problem of minimization of a smooth convex function. We will show that the generated sequences converge strongly to the minimal norm element from the set of all minimizers of the objective function. We will also show that from a practical point of view, Tikhonov regularization does not affect the optimal convergence rate of order $\mathcal{O}(k^{-2})$ for function values. Further, we will obtain fast convergence of the discrete velocity, as well as some estimates concerning the gradient of the objective function. Finally, numerical experiments will be conducted illustrating our theoretical discoveries.

1.2 Preliminary results

Here we would like to gather all the auxiliary facts and lemmas for our future needs.

Definition 1.2.1. A function $\Phi : H \mapsto \overline{\mathbb{R}}$ is called *proper* if $\Phi(x) > -\infty \ \forall x \in H$ and

$\exists x_0 \in H : \Phi(x_0) < +\infty$.

Definition 1.2.2. Let $\Phi : H \mapsto \overline{\mathbb{R}}$. Φ is called *lower semicontinuous* at $x \in H$ if

$$\liminf_{y \to x} \Phi(y) \; = \; \sup_{\delta > 0} \inf_{y \in B(x,\delta)} \Phi(y) \; \geq \; \Phi(x),$$

where $B(x, \delta)$ is the ball centered at x with the radius δ. The function Φ is lower semicontinuous if it is lower semicontinuous at every point in H.

Definition 1.2.3. Let $\Phi : H \mapsto \overline{\mathbb{R}}$. The function Φ is said to be *convex* if $\forall x, y \in H$, $\forall \alpha \in [0, 1]$

$$\Phi\left(\alpha x + (1 - \alpha)y\right) \; \leq \; \alpha\Phi(x) + (1 - \alpha)\Phi(y).$$

Moreover, Φ is said to be *strongly convex* with constant $\mu > 0$ if $\Phi - \frac{\mu}{2}\| \cdot \|^2$ is convex.

Next, we will define the Moreau envelope of a proper, convex and lower semicontinuous function $\Phi : H \mapsto \overline{\mathbb{R}}$ as

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

where $\lambda > 0$ is called the parameter of the Moreau envelope (see, for instance, [31]). For every $\lambda > 0$, the functions Φ and Φ_λ share the same optimal objective value and the same set of minimizers. In addition, Φ_λ is convex and continuously differentiable with

$$\nabla\Phi_\lambda(x) = \frac{1}{\lambda}(x - \operatorname{prox}_{\lambda\Phi}(x)) \quad \forall x \in H, \tag{1.2}$$

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

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

denotes the proximal operator of Φ of parameter λ. For every $x \in H$ and $\lambda, \mu > 0$ we have

$$\| \operatorname{prox}_{\lambda\Phi}(x) - \operatorname{prox}_{\mu\Phi}(x)\| \leq |\lambda - \mu| \|\nabla\Phi_\lambda(x)\|. \tag{1.3}$$

On the other hand, for every $x \in H$, the function $\lambda \in (0, +\infty) \mapsto \Phi_\lambda(x)$ is nonincreasing and differentiable (see, for instance, [6, Lemma A.1]), namely,

$$\frac{d}{d\lambda}\Phi_\lambda(x) = -\frac{1}{2}\|\nabla\Phi_\lambda(x)\|^2 \quad \forall\lambda > 0. \tag{1.4}$$

Next, we will require some definitions related to operators.

Definition 1.2.4. Let $\Phi : H \mapsto \overline{\mathbb{R}}$ be proper and convex. The *subdifferential* of Φ is the set-valued operator, that is, it maps every point in $x \in H$ to a set $\partial\Phi(x) \subseteq H$, defined as

$$\partial\Phi : H \mapsto 2^H : x \mapsto \{u \in H | \; \forall \; y \in H : \langle y - x, u\rangle + \Phi(x) \leq \Phi(y)\}.$$

If $\Phi(x) = \pm\infty$, then $\partial\Phi(x)$ is taken as an empty set. The elements of $\partial\Phi(x)$ are the *subgradients* of Φ at the point x and $\partial\Phi(x)$ is a convex set in the dual space of H.

Definition 1.2.5. Let $A : H \mapsto 2^H$ be a set-valued operator. The operator A is characterized by its *graph*

$$\operatorname{gra} A = \{(x, y) \in H \times H \mid y \in Ax\}.$$

Further, A is monotone if

$$\forall\ (x, u) \in \operatorname{gra} A \text{ and } \forall\ (y, v) \in \operatorname{gra} A : \langle x - y, u - v\rangle \geq 0.$$

Moreover, a monotone operator A is called maximally monotone if there exists no monotone operator $B : H \mapsto 2^H$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

Maximally monotone operators could be characterized in the following way: operator A is maximally monotone if and only if it is monotone and for every pair $(x, u) \in H \times H$,

$$(x, u) \in \operatorname{gra} A \iff \forall\ (y, v) \in \operatorname{gra} A \text{ it holds that } \langle x - y, u - v\rangle \geq 0.$$

The following notions play an important role in the theory of monotone operators.

Definition 1.2.6. Let $A : H \mapsto 2^H$ be maximally monotone and λ be a positive real number. The *resolvent* of A is a single-valued operator defined as

$$J_A = (\operatorname{Id} + A)^{-1}$$

and the *Yosida approximation* of A of index λ is

$$A_\lambda = \frac{1}{\lambda}\left(\operatorname{Id} - J_{\lambda A}\right).$$

Remark 1.2.7. Consider a subdifferential $\partial\Phi$ of a proper, convex and lower semicontinuous function Φ. First, notice that $\partial\Phi$ is maximally monotone. Secondly, for $\lambda > 0$: $J_{\lambda\partial\Phi} = \operatorname{prox}_{\lambda\Phi}$ and $(\partial\Phi)_\lambda = \frac{1}{\lambda}(\operatorname{Id} - \operatorname{prox}_{\lambda\Phi}) = \nabla\Phi_\lambda$ (see, for instance, [31] relation (23.4)).

Next, recall that a differentiable function $\Phi : H \mapsto \mathbb{R}$ is convex if and only if

$$\Phi(x) - \Phi(y) \ \geq\ \langle\nabla\Phi(y), x - y\rangle \ \ \forall (x, y) \in H \times H. \tag{1.5}$$

In the same way, a differentiable function $\Phi : H \mapsto \mathbb{R}$ is strongly convex with constant $\mu > 0$ if and only if

$$\Phi(x) - \Phi(y) \ \geq\ \langle\nabla\Phi(y), x - y\rangle + \frac{\mu}{2}\|x - y\|^2 \quad \forall (x, y) \in H \times H. \tag{1.6}$$

Let us also mention the following lemma (see [31], Proposition 12.22, for the first term of the lemma and [13], Appendix, A1, for the second one).

Lemma 1.2.8. *Let $\Phi : H \mapsto \overline{\mathbb{R}}$ be a proper, convex and lower semicontinuous function, $\lambda, \mu > 0$. Then*

(i) $(\Phi_\lambda)_\mu = \Phi_{\lambda+\mu}$.

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

Proof. By the definition of the Moreau envelope denoting $\alpha = \frac{\lambda}{\lambda+\mu}$ we obtain

$$\begin{aligned}
(\Phi_\lambda)_\mu(x) &= \inf_{z\in H}\left(\inf_{y\in H}\left(\Phi(y) + \frac{1}{2\lambda}\|z-y\|^2\right) + \frac{1}{2\mu}\|z-x\|^2\right)\\
&= \inf_{y\in H}\left(\Phi(y) + \frac{1}{2\alpha\mu}\inf_{z\in H}\left(\alpha\|z-x\|^2 + (1-\alpha)\|z-y\|^2\right)\right)\\
&= \inf_{y\in H}\left(\Phi(y) + \frac{1}{2\alpha\mu}\inf_{z\in H}\left(\|z-(\alpha x+(1-\alpha)y)\|^2 + \alpha(1-\alpha)\|x-y\|^2\right)\right)\\
&= \inf_{y\in H}\left(\Phi(y) + \frac{1}{2(\lambda+\mu)}\|x-y\|^2\right),
\end{aligned}$$

which gives the first claim. Above we used $\|ax+(1-a)y\|^2 + a(1-a)\|x-y\|^2 = a\|x\| + (1-a)\|y\|$ for $x, y \in H$ and $a \in \mathbb{R}$ (see, for instance, [31] Corollary 2.15).

For the second claim first note that according to the semigroup properties of the Moreau envelope (see [31] Proposition 23.7)

$$\frac{1}{\lambda+\mu}\left(\operatorname{Id} - \mathrm{J}_{(\lambda+\mu)\partial\Phi}\right) = \frac{1}{\mu}\left(\operatorname{Id} - \mathrm{J}_{\mu(\partial\Phi)_\lambda}\right),$$

for $\mathrm{J}_{\lambda A} = (\operatorname{Id} + \lambda A)^{-1}$, where A is a maximally monotone operator. Then

$$\frac{\mu}{\lambda+\mu}\left(\operatorname{Id} - \mathrm{J}_{(\lambda+\mu)\partial\Phi}\right) = \operatorname{Id} - \mathrm{J}_{\mu(\partial\Phi)_\lambda},$$

where $(\partial\Phi)_\lambda$ is nothing but $\nabla\Phi_\lambda$. Finally,

$$\mathrm{J}_{\mu\partial(\Phi_\lambda)} = \operatorname{Id} - \frac{\mu}{\lambda+\mu}\operatorname{Id} + \frac{\mu}{\lambda+\mu}\mathrm{J}_{(\lambda+\mu)\partial\Phi},$$

which gives the second claim.

$\square$

In order to obtain strong convergence for the sequence generated by discrete-time schemes one usually needs the following preliminary results. For the proof of the follolwing lemma we refer to [31] Theorem 18.15.

Lemma 1.2.9. *Let $\Phi : H \mapsto \mathbb{R}$ be a convex differentiable function. Then the following are equivalent*

(i) $\nabla\Phi$ *is* $L-$*Lipschitz continuous.*

(ii) $\Phi(x) \le \Phi(y) + \langle\nabla\Phi(y), x-y\rangle + \frac{L}{2}\|y-x\|^2\ \forall x, y \in H$. *(Descent Lemma)*

(iii) $\frac{1}{2L}\|\nabla\Phi(y) - \nabla\Phi(x)\|^2 + \langle\nabla\Phi(y), x-y\rangle + \Phi(y) \le \Phi(x)\ \forall x, y \in H$.

The next result is a modified descent lemma, which in particular contains Lemma 1 from [8], however has a considerable simplified proof.

Lemma 1.2.10. *Let $\Phi : H \mapsto \mathbb{R}$ be a convex differentiable function, with $L-$Lipschitz continuous gradient and let $s > 0$. Then,*

$$\Phi(y - s\nabla\Phi(y)) \le \Phi(x) + \langle \nabla\Phi(y), y - x\rangle + \left(\frac{L}{2}s^2 - s\right)\|\nabla\Phi(y)\|^2 - \frac{1}{2L}\|\nabla\Phi(y) - \nabla\Phi(x)\|^2, \tag{1.7}$$

$\forall x, y \in H$. Assume further that $s \in \left(0, \frac{1}{L}\right]$. Then,

$$\Phi(y - s\nabla\Phi(y)) \le \Phi(x) + \langle \nabla\Phi(y), y - x\rangle - \frac{s}{2}\|\nabla\Phi(y)\|^2 - \frac{s}{2}\|\nabla\Phi(y) - \nabla\Phi(x)\|^2, \ \forall x, y \in H. \tag{1.8}$$

Proof. By taking $x = y - s\nabla\Phi(y)$ in Lemma 1.2.9 item 2, we get

$$\Phi(y - s\nabla\Phi(y)) \le \Phi(y) + \left(\frac{L}{2}s^2 - s\right)\|\nabla\Phi(y)\|^2, \ \forall y \in H. \tag{1.9}$$

From Lemma 1.2.9 item 3 we have

$$\Phi(y) \le \Phi(x) + \langle \nabla\Phi(y), y - x\rangle - \frac{1}{2L}\|\nabla\Phi(y) - \nabla\Phi(x)\|^2, \ \forall x, y \in H. \tag{1.10}$$

Combining (1.9) and (1.10) we get

$$\Phi(y - s\nabla\Phi(y)) \le \Phi(x) + \langle \nabla\Phi(y), y - x\rangle + \left(\frac{L}{2}s^2 - s\right)\|\nabla\Phi(y)\|^2 - \frac{1}{2L}\|\nabla\Phi(y) - \nabla\Phi(x)\|^2,$$

$\forall x, y \in H$, which is nothing else than (1.7).

Assume that $0 < s \le \frac{1}{L}$. Then,

$$\frac{L}{2}s^2 - s \le -\frac{s}{2} \text{ and } -\frac{1}{2L} \le -\frac{s}{2},$$

hence (1.7) leads to (1.8), that is

$$\Phi(y - s\nabla\Phi(y)) \le \Phi(x) + \langle \nabla\Phi(y), y - x\rangle - \frac{s}{2}\|\nabla\Phi(y)\|^2 - \frac{s}{2}\|\nabla\Phi(y) - \nabla\Phi(x)\|^2, \ \forall x, y \in H.$$

□

Finally, we will provide some known in the literature results, which are going to be useful later for our analysis.

Definition 1.2.11. Let I be a finite interval in the real line $\mathbb{R}$. A function $\Phi : I \mapsto \mathbb{R}$ is absolutely continuous on I if for every $\varepsilon > 0$ there exists $\delta > 0$ such that as soon as a finite sequence of pairwise disjoint sub-intervals $\{(x_k, y_k)\}_{k=1}^{N\in\mathbb{N}}$ of I, $y_k > x_k \ \forall k$, satisfies

$$\sum_{k=1}^{N}(y_k - x_k) \ < \ \delta \text{ then } \sum_{k=1}^{N}|\Phi(y_k) - \Phi(x_k)| \ < \ \varepsilon.$$

An absolutely continuous function on I is almost everywhere differentiable on I. We say that a function is locally absolutely continuous if it is absolutely continuous on every compact interval of its domain.

For the proof of the next lemma we refer to [4].

Lemma 1.2.12. *Suppose that $f : [t_0, +\infty) \mapsto \mathbb{R}$ 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 following lemma, which we will use later to obtain the existence of some important limits, we refer to [15].

Lemma 1.2.13. *Let H be a real Hilbert space and $x : [t_0, +\infty) \mapsto \mathbb{H}$ 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 $t \to +\infty$.*

We state a continuous version of Opial's Lemma (see [50]), which is used in the proof of the convergence of the trajectory.

Lemma 1.2.14. *Let S be a non-empty subset of a real Hilbert space H and $x : [0, +\infty) \mapsto H$ a given map. Assume that*

(i) for every $z \in S$, $\lim_{t\to+\infty} \|x(t) - z\|$ exists;

(ii) every weak sequential cluster point of the map x belongs to S.

Then $x(t)$ converges weakly to some element of S as $t \to +\infty$.

Finally, we will need the following lemma later in our analysis to establish the rates of convergence of the function and Moreau envelope values. For its proof we refer to [9].

Lemma 1.2.15. *Let $\delta > 0$ and $f \in L^1((\delta, +\infty), \mathbb{R})$ be a non-negative and continuous function. Let $g : [\delta, +\infty) \mapsto [0, +\infty)$ be a non-decreasing function such that $\lim_{t\to+\infty} g(t) = +\infty$. Then it holds*

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

1.3 Historical remarks

Inertial dynamics were introduced by Polyak in [51] in the form of the so-called heavy ball with friction method

$$\ddot{x}(t) + \alpha\dot{x}(t) + \nabla\Phi(x(t)) = 0,$$

with fixed viscous coefficient $\alpha > 0$, in order to accelerate the gradient method for the minimization of a continuous differentiable function $\Phi : H \mapsto \mathbb{R}$. This system was later studied by Alvarez-Attouch [1, 2] and by Attouch-Goudou-Redont [11]. In these works, for a convex function Φ an asymptotic convergence rate of $\Phi(x(t))$ to $\min \Phi$ of the order $O\left(\frac{1}{t}\right)$, as $t \to +\infty$, as well as an improvement for a strongly convex function Φ to an exponential rate of convergence, were proved. The weak convergence of the trajectories to a minimizer of Φ was also established.

A major step to obtain faster asymptotic convergence in the convex regime was done by Su-Boyd-Candes [52], by considering in the second order dynamical system an asymptotic vanishing damping coefficient

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0, \tag{1.11}$$

for $t \geq t_0$ and $\alpha \geq 3$. Second order dynamical systems with variable and vanishing damping coefficients for optimization were studied, for instance, in [32, 36, 37]. The system (1.11) corresponds to a continuous version of Nesterov's accelerated gradient method [48] (see also [38], which has a connection to the system in question as well; in addition to fast convergence rates of the function values, the authors were able to prove weak convergence of the iterates depending on the choice of the damping parameter α). For the function values, rates of convergence of

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

were obtained. For $\alpha > 3$, in [10] it was shown that the trajectory of (1.11) converges weakly to an element of $\operatorname{argmin} \Phi$, and in [14, 47] the asymptotic convergence rate of the function values was improved to $o\left(\frac{1}{t^2}\right)$, as $t \to +\infty$.

The following system which combines asymptotic vanishing damping with Hessian-driven damping was proposed by Attouch-Peypouquet-Redont in [15]

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0 \tag{1.12}$$

for $t \geq t_0$, where $\Phi : H \mapsto \mathbb{R}$ is twice continuously differentiable and convex, $\alpha \geq 3$ and $\beta \geq 0$. Hessian-driven damping has a natural link with Newton's method and gives rise to dynamical inertial Newton systems [3]. The system (1.12) preserves the convergence properties of (1.11), while having for $\beta > 0$ other important features, namely,

$$\lim_{t \mapsto +\infty} \|\nabla\Phi(x(t))\| = 0 \text{ and } \int_{t_0}^{+\infty} t^2 \|\nabla\Phi(x(t))\|^2 dt < +\infty.$$

In addition, according to [15] possible oscillations exhibited by the solutions of (1.11) are neutralized in (1.12).

1.4 Time scaling

Time scaling of the dynamical system (1.11) was used in order to accelerate the rate of convergence of the values of the function Φ along the trajectory. The system (1.11) becomes through time scaling a dynamical system of the form

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + b(t)\nabla\Phi(x(t)) = 0, \tag{1.13}$$

where $\alpha \geq 3$ and $b : [t_0, +\infty) \mapsto (0, +\infty)$ is a continuous scalar function, as it was introduced and studied by Attouch-Chbani-Riahi in [9]. For (1.13) it was shown that

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

a convergence rate which can be improved to $o\left(\frac{1}{t^2 b(t)}\right)$, as $t \to +\infty$, if $\alpha > 3$.

In [7,8] (see also [5]) the dynamical system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\nabla^2\Phi(x(t))\dot{x}(t) + b(t)\nabla\Phi(x(t)) = 0, \tag{1.14}$$

which combines viscous and Hessian-driven damping with time scaling, where $\alpha \geq 1$ and $\beta, b : [t_0, +\infty) \mapsto (0, +\infty)$ are functions with appropriate differentiability properties, was investigated. A quite general setting formulated in terms of the dynamical system parameter functions was identified in which the properties of (1.14) concerning the convergence of the function values are preserved, while the gradient of Φ strongly converges along the trajectory to zero and the trajectory converges weakly to a minimizer of the objective function. In [7,8] a numerical algorithm obtained via time discretization of (1.14) was studied, exhibiting analogous to the dynamical system convergence properties. The implicit discretization of this system led to proximal algorithms, while some additional assumptions, such as Φ having Lipschitz continuous gradient, allow avoiding having to use the proximal operator in the scheme formulation.

1.5 Nonsmooth optimization

Attouch-Cabot considered in [6] (see also [13] for a more general approach for monotone inclusions) in connection with the minimization of the proper, convex and lower semicontinuous function $\Phi : H \mapsto \overline{\mathbb{R}}$ the following second order differential equation

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi_{\lambda(t)}(x(t)) = 0 \tag{1.15}$$

for $t \geq t_0$, where $\alpha \geq 1$ and $\lambda : [t_0, +\infty) \mapsto (0, +\infty)$ is continuously differentiable and non-decreasing. Convergence rates for the values of the Moreau envelope as well as for the velocity of the system were obtained

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

from where convergence rates for the Φ along the image of the trajectory were deduced

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

as $t \to +\infty$. In addition, the weak convergence of the trajectory x to a minimizer of Φ as $t \to +\infty$ was established.

Attouch-László considered in [12] in the same context the dynamical system

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t)) + \nabla\Phi_{\lambda(t)}(x(t)) = 0 \tag{1.16}$$

where $\alpha > 1$ and $\beta > 0$, and the term $\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t))$ is inspired by the Hessian driven damping and its existence is justified almost everywhere since the mapping $t \mapsto \nabla\Phi_{\lambda(t)}(x(t))$ is locally absolutely continuous (see, for example, [12, Lemma 1]). It was shown that for $\lambda(t) = \lambda t^2$, where $\lambda > 0$, the system (1.16) inherits all major convergence properties of (1.15) and, in addition, the following convergence rates for the gradient of the Moreau envelope of parameter λ and its time derivative along x were established

$$\|\nabla\Phi_{\lambda(t)}(x(t))\| = o\left(\frac{1}{t^2}\right) \text{ and } \left\|\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t))\right\| = o\left(\frac{1}{t^2}\right) \text{ as } t \to +\infty.$$

Once again, the discretization of the above systems is possible, leading to proximal algorithms with similar properties.

1.6 Tikhonov regularization

The presence of the Tikhonov term in the system equation dramatically influences the behaviour of its trajectories, namely, under some appropriate conditions, it improves the convergence of the trajectories from weak to a strong one. Not only that, but it also ensures the convergence not to an arbitrary element from the set of all minimizers of the objective, but to the particular one, which has the smallest norm. Under the presence of the Tikhonov term in the system it is still possible to obtain fast rates of convergence of the function values. Systems with Tikhonov regularization were studied, for instance, in [5, 17, 18, 21, 24, 30, 34, 35, 41, 42, 45].

One of the many examples of such systems is presented below (see [30])

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0, \tag{1.17}$$

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, based on the assumptions made on ε: one for the fast convergence of values, obtaining

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

and the weak convergence of the trajectories to a minimizer of Φ under the condition (among many others)

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

and another setting for the strong convergence of x to the minimal norm solution as $t \to +\infty$, provided the exists a constant C such that for t big enough

$$t^2\varepsilon(t) \ \geq \ C.$$

As we may see, the two assumptions contradict each other for the most natural choice of Tikhonov function, namely, the polynomial one.

The first attempt to obtain both major results in one setting was made in [5]:

$$\ddot{x}(t) + \alpha\sqrt{\varepsilon(t)}\dot{x}(t) + \nabla\Phi(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0, \tag{1.18}$$

where α, $t_0 > 0$ and $\Phi : H \mapsto \mathbb{R}$ is continuously differentiable and convex. In that paper the authors obtained the decaying rates for the function values $\Phi(x(t)) - \min\Phi$, as well as for the quantity $\|x(t) - x_{\varepsilon(t)}\|$, as $t \to +\infty$, where $x_{\varepsilon(t)} = \operatorname{argmin}\left(\Phi(x) + \frac{\varepsilon(t)\|x\|^2}{2}\right)$. Thus, they assured the strong convergence of the trajectories to the minimal norm solution under the appropriate assumptions and properly chosen energy functional, using the properties of Tikhonov regularization, namely, the convergence of $x_{\varepsilon(t)}$ to the minimal norm solution, as $t \to +\infty$. 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.

As a generalization of (1.17) the authors considered the following dynamics (see [18])

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

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 is chosen appropriately. This system inherits the properties of fast convergence rates of the function values, being of the order $\frac{1}{t^2}$, as $t \to +\infty$, and additionally provides the strong convergence results for the trajectories of the system as $t \to +\infty$, in the same setting.

Concerning the nonsmooth case we refer to [34], where it was covered for the more general systems, governed by a maximally monotone operator, but with a different damping. The authors studied the following dynamics

$$\ddot{x}(t) + \frac{\alpha}{t^q}\dot{x}(t) + \beta\frac{d}{dt}\Big(A_{\lambda(t)}(x(t))\Big) + A_{\lambda(t)}(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0, \tag{1.19}$$

where $\alpha > 0$, $\beta \geq 0$, $0 < q \leq 1$ and $\lambda(t) = \lambda t^{2q}$ for $\lambda > 0$, A is a maximally monotone operator and A_λ is its Yosida regularization of the order λ. The system (1.19) is related to the inclusion problem $0 \in Ax$. The authors showed the fast convergence rates for $\|\dot{x}(t)\|$, $\|A_{\lambda(t)}(x(t))\|$ and $\|\frac{d}{dt}A_{\lambda(t)}(x(t))\|$ being of the order $\frac{1}{t^q}$, $\frac{1}{t^{2q}}$ and $\frac{1}{t^{3q}}$, as $t \to +\infty$, correspondingly. Moreover, they established the strong convergence, as $t \to +\infty$, of the trajectories of the system. In the section 4 of [34] the authors also considered a very interesting particular case of $A = \partial\Phi$ using the well-known connection $A_\lambda = (\partial\Phi)_\lambda = \nabla\Phi_\lambda$ for all $\lambda > 0$ (see Remark (1.2.7)).

1.7 Tikhonov regularization in discrete time optimization

Concerning the discrete case, that is, the case of inertial algorithms that converge strongly to the minimal norm solution of a convex optimization problem, there are only a few results in the literature, see [24, 44], which refer to proximal inertial algorithms obtained via implicit discretization of the second order continuous dynamical systems studied in [45, 52]).

Indeed, in [24] the following inertial-proximal algorithm was considered in connection to the optimization problem (1.1): let $x_0, x_1 \in H$ and for every $k \geq 1$ set

$$x_{k+1} = \operatorname{prox}_{\Phi}\left(x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) - \frac{c}{k^2}x_k\right), \tag{1.20}$$

where $\alpha > 3$ and $c > 0$. To our best knowledge this is the first inertial algorithm in the literature for which both the strong convergence result $\liminf_{k\to+\infty}\|x_k - x^*\| = 0$ (throughout this section x^* is the minimal norm element from $\operatorname{argmin}\Phi$) for the generated sequences and fast convergence of the potential energy $\Phi(x_k) - \min\Phi$ and discrete velocity $\|x_k - x_{k-1}\|$, as $k \to +\infty$, were obtained. Note that (1.20) can be obtained via implicit discretization of the dynamical system from [19]:

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

however the rates are not entirely preserved, since in (1.20) one obtains only the rates $\Phi(x_k) - \min\Phi = o(k^{-2s})$ and $\|x_k - x_{k-1}\| = o(k^{-s})$ as $k \to +\infty$, where $s \in \left[\frac{1}{2}, 1\right)$.

In order to show the strong convergence result $\lim_{k\to+\infty}\|x_k - x^*\| = 0$ but also the correlation among the stepsize, inertial parameter and Tikhonov regularization parameter, in [44] the author assumed that the objective function in (1.1) is proper, convex and lower semicontinuous only and associated to this optimization problem the following inertial-proximal algorithm: let $x_0, x_1 \in H$ and for every $k \geq 1$ set

$$x_{k+1} = \operatorname{prox}_{\lambda_k\Phi}\left(x_k + \left(1 - \frac{\alpha}{k^q}\right)(x_k - x_{k-1}) - \frac{c}{k^p}x_k\right), \tag{1.21}$$

where α, q, c, $p > 0$ and (λ_k) is a sequence of positive real numbers. Note that (1.21) can be seen as the discrete counterpart of the continuous dynamical system from [45]:

$$\ddot{x}(t) + \frac{\alpha}{t^q}\dot{x}(t) + \nabla\Phi(x(t)) + \frac{a}{t^p}x(t) = 0 \text{ for } t \geq t_0,$$

where q, p, α and a are positive constants. According to [44], in case the stepsize $\lambda_k \equiv 1$ and $0 < q < 1$, $1 < p < q + 1$ then $\lim_{k\to+\infty}\|x_k - x^*\| = 0$. Further, $\|x_k - x_{k-1}\| = \mathcal{O}(k^{-\frac{q+1}{2}})$ as $k \to +\infty$ and $\Phi(x_k) - \min\Phi = \mathcal{O}(k^{-p})$ as $k \to +\infty$. If $q+1 < p \leq 2$ and for $p = 2$ one has $c > q(1-q)$, then (x_k) converges weakly to a minimizer of Φ. Moreover, $\Phi(x_k) - \min\Phi = \mathcal{O}(k^{-q-1})$ and $\|x_k - x_{k-1}\| = \mathcal{O}(k^{-\frac{q+1}{2}})$ as $k \to +\infty$.

2 Nonsmooth optimization with time scaling

In this chapter we will study the asymptotic behaviour of the second order in time evolution equation in connection with the minimization problem (1.1)

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t)) + b(t)\nabla\Phi_{\lambda(t)}(x(t)) = 0, \tag{2.1}$$

with initial conditions $x(t_0) = x_0 \in H$, $\dot{x}(t_0) = u_0 \in H$, 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. Here and later, we will use the well-known technique of linking the gradient of the Moreau envelope Φ_λ of the objective function Φ to dynamics (2.1) and studying its properties.

Our aim is to derive rates of convergence for the Moreau envelope of the objective function and the objective function itself to $\min \Phi$, as well as for the gradient of the Moreau envelope of the objective function and the velocity of the trajectory to zero in terms of the Moreau parameter function λ and the time scaling function b. In addition, we will provide a setting which also guarantees the weak convergence of the trajectory of the dynamical system to a minimizer of Φ. The theoretical results will be illustrated by numerical experiments.

2.1 Existence and uniqueness of strong global solution

This section is devoted to the topic of existence and uniqueness of a strong global solution of the system of our interest. To this aim we will rewrite (2.1) as a system of the first order in time equations in the product space $H \times H$.

For this section, we additionally require that $\beta : [t_0, +\infty) \mapsto [0, +\infty)$ is twice continuously differentiable with $\beta(t) > 0$ for every $t \geq t_0$. We integrate (2.1) from t_0 to t to obtain

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

Denote $z(t) := \int_{t_0}^{t}\left(\frac{\alpha}{s}\dot{x}(s) + \left(b(s) - \dot{\beta}(s)\right)\nabla\Phi_{\lambda(s)}(x(s))\right)ds - \left(u_0 + \beta(t_0)\nabla\Phi_{\lambda(t_0)}(x_0)\right)$ for every $t \geq t_0$. Since $\dot{z}(t) = \frac{\alpha}{t}\dot{x}(t) + \left(b(t) - \dot{\beta}(t)\right)\nabla\Phi_{\lambda(t)}(x(t))$ we notice, that (2.1) is equivalent to

$$\begin{cases} \dot{x}(t) + \beta(t)\nabla\Phi_{\lambda(t)}(x(t)) + z(t) = 0, \\ \dot{z}(t) - \frac{\alpha}{t}\dot{x}(t) - \left(b(t) - \dot{\beta}(t)\right)\nabla\Phi_{\lambda(t)}(x(t)) = 0, \\ x(t_0) = x_0, \ z(t_0) = -\left(u_0 + \beta(t_0)\nabla\Phi_{\lambda(t_0)}(x_0)\right). \end{cases}$$

After multiplying the first line by $b(t) - \dot{\beta}(t)$ and the second one by $\beta(t)$, by summing them we get rid of the gradient of the Moreau envelope in the second equation

$$\begin{cases} \dot{x}(t) + \beta(t)\nabla\Phi_{\lambda(t)}(x(t)) + z(t) = 0, \\ \beta(t)\dot{z}(t) + \left(b(t) - \dot{\beta}(t) - \frac{\alpha\beta(t)}{t}\right)\dot{x}(t) + \left(b(t) - \dot{\beta}(t)\right)z(t) = 0, \\ x(t_0) = x_0,\ z(t_0) = -\left(u_0 + \beta(t_0)\nabla\Phi_{\lambda(t_0)}(x_0)\right). \end{cases}$$

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

$$\begin{cases} \dot{x}(t) + \beta(t)\nabla\Phi_{\lambda(t)}(x(t)) + \left(\frac{\dot{\beta}(t)-b(t)}{\beta(t)} + \frac{\alpha}{t}\right)x(t) + \frac{1}{\beta(t)}y(t) = 0, \\ \dot{y}(t) + \left(\ddot{\beta}(t) + \frac{3b(t)\dot{\beta}(t)-2\dot{\beta}^2(t)-b^2(t)}{\beta(t)} + \frac{\alpha}{t}\left(b(t) - \dot{\beta}(t) - \frac{\beta(t)}{t}\right) - \dot{b}(t)\right)x(t) \\ \qquad\qquad + \frac{b(t)-2\dot{\beta}(t)}{\beta(t)}y(t) = 0, \\ x(t_0) = x_0,\ y(t_0) = -\beta(t_0)\left(u_0 + \beta(t_0)\nabla\Phi_{\lambda(t_0)}(x_0)\right) + \left(b(t_0) - \dot{\beta}(t_0) - \frac{\alpha\beta(t_0)}{t_0}\right)x_0. \end{cases}$$

In case $\beta(t) = 0$ for every $t \geq t_0$, (2.1) can be equivalently written as

$$\begin{cases} \dot{x}(t) - y(t) = 0, \\ \dot{y}(t) + \frac{\alpha}{t}y(t) + b(t)\nabla\Phi_{\lambda(t)}(x(t)) = 0, \\ x(t_0) = x_0,\ y(t_0) = u_0. \end{cases}$$

Based on the two reformulation of the dynamical system (2.1) we can formulate the following existence and uniqueness result, which is a consequence of the Cauchy-Lipschitz theorem for strong global solutions. The result can be proved in the lines of the proofs of Theorem 1 in [12] or of Theorem 1.1 in [15] with some small adjustments.

Theorem 2.1.1. *Suppose that $\beta : [t_0, +\infty) \mapsto [0, +\infty)$ is twice continuously differentiable such that either $\beta(t) > 0$ for every $t \geq t_0$ or $\beta(t) = 0$ for every $t \geq t_0$, and that there exists $\lambda_0 > 0$ such that $\lambda(t) \geq \lambda_0$ for every $t \geq t_0$. Then for every $(x_0, u_0) \in H \times H$ there exists a unique strong global solution $x : [t_0, +\infty) \mapsto H$, that is, absolutely continuous on any compact subset of $[t_0, +\infty)$, of the continuous dynamics (2.1) which satisfies the Cauchy initial conditions $x(t_0) = x_0$ and $\dot{x}(t_0) = u_0$.*

2.2 Energy function and rates of convergence for function values

In this section we will define for the dynamical system (2.1) an energy function and investigate its dissipativity properties. These will play a crucial role in the derivation of rates of convergence for the Moreau envelope of Φ and the objective function itself.

To shorten the calculations, we introduce the auxiliary function (see also [7, 8])

$$w : [t_0, +\infty) \mapsto \mathbb{R}, \quad w(t) \ = \ b(t) - \dot{\beta}(t) - \frac{\beta(t)}{t}.$$

For $z \in \operatorname{argmin} \Phi$ and

$$0 \leq c \leq \alpha - 1, \tag{2.2}$$

consider the energy function $E_c : [t_0, +\infty) \mapsto [0, +\infty)$,

$$\begin{aligned} E_c(t) &= \big(t^2 w(t) + (\alpha - 1 - c) t \beta(t)\big)\big(\Phi_{\lambda(t)}(x(t)) - \min \Phi\big) \\ &+ \frac{1}{2} \left\| c(x(t) - z) + t\dot{x}(t) + t\beta(t) \nabla \Phi_{\lambda(t)}(x(t)) \right\|^2 \\ &+ \frac{c(\alpha - 1 - c)}{2} \|x(t) - z\|^2. \end{aligned}$$

In the following theorem we formulate sufficient conditions that guarantee the decay of the energy of the the dynamical system (2.1) and discuss some of its consequences.

Theorem 2.2.1. *Suppose that $\alpha \geq 1$, λ is nondecreasing on $[t_0, +\infty)$ and the following conditions*

$$b(t) > \dot{\beta}(t) + \frac{\beta(t)}{t} \text{ for every } t \geq t_0 \tag{2.3}$$

and

$$(\alpha - 3) w(t) - t \dot{w}(t) \geq 0 \text{ for every } t \geq t_0 \tag{2.4}$$

are satisfied. Then, for a solution $x : [t_0, +\infty) \mapsto H$ to (2.1), the following statements are true:

(i) $\dot{E}_c(t) \leq 0$ *for every* $t \geq t_0$*;*

(ii) $\Phi_{\lambda(t)}(x(t)) - \min \Phi \ \leq \ \frac{E_{\alpha-1}(t_0)}{t^2 w(t)}$ *for every* $t \geq t_0$*;*

(iii) $\int_{t_0}^{+\infty} \left(t^2 w(t) \frac{\dot{\lambda}(t)}{2} + t^2 \beta(t) w(t) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 dt \ < \ +\infty$*;*

(iv) $\int_{t_0}^{+\infty} \big((\alpha - 3) t w(t) - t^2 \dot{w}(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min \Phi) dt \ < \ +\infty$*.*
Assuming moreover that $\alpha > 1$ and that

$$\textit{there exists } \varepsilon \in (0, \alpha - 1) \textit{ such that } (\alpha - 3) w(t) - t \dot{w}(t) \ \geq \ \varepsilon b(t) \quad \forall t \geq t_0, \tag{2.5}$$

it holds

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

(vi) the trajectory x is bounded and

(vii) $\int_{t_0}^{+\infty} t b(t) (\Phi_{\lambda(t)}(x(t)) - \min \Phi) dt \ < \ +\infty$*.*

Proof. Denote for simplicity $v(t) = c(x(t) - z) + t\dot{x}(t) + t\beta(t)\nabla\Phi_{\lambda(t)}(x(t))$ $\forall t \geq t_0$. For every $t \geq t_0$ we obtain

$$\begin{aligned}\dot{E}_c(t) &= \big(2tw(t) + t^2\dot{w}(t) + \beta(t)(\alpha - 1 - c) + (\alpha - 1 - c)t\dot{\beta}(t)\big)\big(\Phi_{\lambda(t)}(x(t)) - \min\Phi\big)\\ &+ \big(t^2w(t) + \beta(t)t(\alpha - 1 - c)\big)\Big(\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle - \frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\Big)\\ &+ \Big\langle v(t), \dot{v}(t)\Big\rangle + c(\alpha - 1 - c)\langle x(t) - z, \dot{x}(t)\rangle,\end{aligned}$$

where we used (1.4):

$$\frac{d}{dt}\big(\Phi_{\lambda(t)}(x(t)) - \min\Phi\big) = \big\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\big\rangle - \frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}x(t)\|^2. \tag{2.6}$$

Note that simple calculations yield for every $t \geq t_0$

$$\dot{v}(t) = (c + 1)\dot{x}(t) + t\ddot{x}(t) + t\beta(t)\frac{d}{dt}\big(\nabla\Phi_{\lambda(t)}(x(t))\big) + (\beta(t) + t\dot{\beta}(t))\nabla\Phi_{\lambda(t)}(x(t)).$$

Using (2.1) to replace $\ddot{x}(t)$, we deduce for every $t \geq t_0$

$$\begin{aligned}\Big\langle v(t), \dot{v}(t)\Big\rangle &= c(c + 1 - \alpha)\langle x(t) - z, \dot{x}(t)\rangle\\ &+ c\big(\beta(t) + t\dot{\beta}(t) - tb(t)\big)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle + (c + 1 - \alpha)t\|\dot{x}(t)\|^2\\ &+ \big(\beta(t) + t\dot{\beta}(t) - tb(t)\big)t\big\langle\dot{x}(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle\\ &+ t\beta(t)(c + 1 - \alpha)\big\langle\dot{x}(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle\\ &+ t\beta(t)(\beta(t) + t\dot{\beta}(t) - tb(t))\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.\end{aligned}$$

Overall, since $\beta(t) + t\dot{\beta}(t) - tb(t) = -tw(t)$, we obtain for every $t \geq t_0$

$$\begin{aligned}\dot{E}_c(t) &= \big(2tw(t) + t^2\dot{w}(t) - (\beta(t) + t\dot{\beta}(t))(c + 1 - \alpha)\big)\big(\Phi_{\lambda(t)}(x(t)) - \min\Phi\big)\\ &+ \big(t^2w(t) - t\beta(t)(c + 1 - \alpha)\big)\Big(\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle - \frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\Big)\\ &- ctw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle + (c + 1 - \alpha)t\|\dot{x}(t)\|^2\\ &- t^2w(t)\langle\dot{x}(t), \nabla\Phi_{\lambda(t)}(x(t))\rangle + t\beta(t)(c + 1 - \alpha)\big\langle\dot{x}(t), \nabla\Phi_{\lambda(t)}(x(t))\big\rangle\\ &- t^2\beta(t)w(t)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.\end{aligned}$$

Notice that the terms with $\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\rangle$ cancel each other, thus, after simplification we obtain for every $t \geq t_0$

$$\begin{aligned}\dot{E}_c(t) &= \big(2tw(t) + t^2\dot{w}(t) + (\beta(t) + t\dot{\beta}(t))(\alpha - 1 - c)\big)\big(\Phi_{\lambda(t)}(x(t)) - \min\Phi\big)\\ &- \big(t^2w(t) + t\beta(t)(\alpha - 1 - c)\big)\frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\\ &- ctw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle\\ &- (\alpha - 1 - c)t\|\dot{x}(t)\|^2 - t^2\beta(t)w(t)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.\end{aligned} \tag{2.7}$$

Thanks to (2.3), $w(t)$ is positive for every $t \geq t_0$, thus

$$-ctw(t)\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\rangle \;\leq\; -ctw(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi),$$

which leads to

$$\begin{aligned}\dot{E}_c(t) \;\leq&\; \big((2-c)tw(t) + t^2\dot{w}(t) + (\beta(t) + t\dot{\beta}(t))(\alpha - 1 - c)\big)\big(\Phi_{\lambda(t)}(x(t)) - \min\Phi\big)\\ &- \left(\big(t^2w(t) + t\beta(t)(\alpha - 1 - c)\big)\frac{\dot{\lambda}(t)}{2} + t^2\beta(t)w(t)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\ &- (\alpha - c - 1)t\|\dot{x}(t)\|^2.\end{aligned} \tag{2.8}$$

By (2.2) and the fact that λ is nondecreasing, we deduce that

$$\big(t^2w(t) + t\beta(t)(\alpha - 1 - c)\big)\frac{\dot{\lambda}(t)}{2} + t^2\beta(t)w(t) \;\geq\; 0,$$

so, we obtain for every $t \geq t_0$

$$\begin{aligned}\dot{E}_c(t) \;\leq&\; \big((2-c)tw(t) + t^2\dot{w}(t) + (\beta(t) + t\dot{\beta}(t))(\alpha - 1 - c)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi)\\ &- \left(\big(t^2w(t) + t\beta(t)(\alpha - c - 1)\big)\frac{\dot{\lambda}(t)}{2} + t^2\beta(t)w(t)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.\end{aligned} \tag{2.9}$$

Let us choose $c := \alpha - 1$. According to (2.4) we obtain for the coefficient of $\Phi_{\lambda(t)}(x(t)) - \min\Phi$ in (2.9)

$$(2-c)tw(t) + t^2\dot{w}(t) + (\beta(t) + t\dot{\beta}(t))(\alpha - 1 - c) \;= -t\big((\alpha - 3)w(t) - t\dot{w}(t)\big) \;\leq\; 0.$$

Therefore, (2.9) allows us to deduce for every $t \geq t_0$

$$\begin{aligned}\dot{E}_{\alpha-1}(t) \;\leq&\; -\; \big((\alpha - 3)tw(t) - t^2\dot{w}(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi)\\ &-\; \left(t^2w(t)\frac{\dot{\lambda}(t)}{2} + t^2\beta(t)w(t)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\\ \leq&\; 0.\end{aligned}$$

We have just established that $E_{\alpha-1}$ is nonincreasing, which leads for every $t \geq t_0$ to

$$\begin{aligned}E_{\alpha-1}(t) \;=&\; t^2w(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi)\\ &+\; \frac{1}{2}\left\|(\alpha - 1)(x(t) - z) + t\dot{x}(t) + t\beta(t)\nabla\Phi_{\lambda(t)}(x(t))\right\|^2\\ \leq&\; E_{\alpha-1}(t_0).\end{aligned}$$

From here we obtain for every $t \geq t_0$

$$\Phi_{\lambda(t)}(x(t)) - \min\Phi \;\leq\; \frac{E_{\alpha-1}(t_0)}{t^2w(t)}, \tag{2.10}$$

which proves (ii). Moreover, by integration, we obtain

$$\int_{t_0}^{+\infty} \left(t^2 w(t) \frac{\dot{\lambda}(t)}{2} + t^2 \beta(t) w(t) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 dt \ \leq \ E_{\alpha-1}(t_0) \ < \ +\infty$$

and

$$\int_{t_0}^{+\infty} \big((\alpha-3)tw(t) - t^2\dot{w}(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min \Phi) dt \ \leq \ E_{\alpha-1}(t_0) \ < \ +\infty, \tag{2.11}$$

which are the claims (iii) and (iv).

From now on we assume that $\alpha > 1$ and choose $c := \alpha - 1 - \varepsilon$, where ε is given by (2.5). In this setting, (2.8) reads for every $t \geq t_0$,

$$\begin{aligned} \dot{E}_{\alpha-1-\varepsilon}(t) \ &\leq \big((3-\alpha+\varepsilon)tw(t) + t^2\dot{w}(t) + \varepsilon(\beta(t) + t\dot{\beta}(t))\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\ &\quad - \left(\big(t^2 w(t) + \varepsilon t \beta(t)\big) \frac{\dot{\lambda}(t)}{2} + t^2 \beta(t) w(t) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 - \varepsilon t \|\dot{x}(t)\|^2 \\ &= -t\big((\alpha-3)w(t) - t\dot{w}(t) - \varepsilon b(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\ &\quad - \left(\big(t^2 w(t) + \varepsilon t \beta(t)\big) \frac{\dot{\lambda}(t)}{2} + t^2 \beta(t) w(t) \right) \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 - \varepsilon t \|\dot{x}(t)\|^2. \end{aligned} \tag{2.12}$$

So, under the condition (2.5), $\dot{E}_{\alpha-1-\varepsilon}(t) \ \leq \ 0$ for every $t \geq t_0$. Integrating (2.12) we obtain

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt \ < \ +\infty, \tag{2.13}$$

which gives the claim (v). From the fact that the energy function

$$\begin{aligned} E_{\alpha-1-\varepsilon}(t) \ &= \big(t^2 w(t) + \varepsilon t \beta(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\ &\quad + \ \frac{1}{2} \left\| (\alpha - 1 - \varepsilon)(x(t) - z) + t\dot{x}(t) + t\beta(t)\nabla\Phi_{\lambda(t)}(x(t)) \right\|^2 \\ &\quad + \ \frac{(\alpha-1-\varepsilon)\varepsilon}{2} \|x(t) - z\|^2 \end{aligned}$$

is bounded from above and it is nonnegative on $[t_0, +\infty)$, it follows that the trajectory x is bounded, which is item (vi). Finally, from (2.5) and (2.11) we deduce the claim (vii)

$$\begin{aligned} &\int_{t_0}^{+\infty} \varepsilon t b(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) dt \\ \leq \ &\int_{t_0}^{+\infty} \big((\alpha-3)tw(t) - t^2\dot{w}(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) dt \\ < \ &+\infty, \end{aligned} \tag{2.14}$$

which finishes the proof. □

The following auxiliary result will be needed later.

Lemma 2.2.2. *Suppose that $\alpha > 1$ and (2.5) holds, that λ and β are nondecreasing on $[t_0, +\infty)$, and that (2.3) holds. Then, for a solution $x : [t_0, +\infty) \mapsto H$ to (2.1), it holds*

$$\int_{t_0}^{+\infty} tw(t)\langle \nabla\Phi_{\lambda(t)}(x(t)), x(t) - z\rangle dt \ < \ +\infty. \tag{2.15}$$

Proof. Recall that according to (2.7) we have for every $t \geq t_0$

$$\begin{aligned}
\dot{E}_c(t) \ &= \ \big(2tw(t) + t^2\dot{w}(t) + (\beta(t) + t\dot{\beta}(t))(\alpha - 1 - c)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\
&- \ \big(t^2 w(t) + t\beta(t)(\alpha - 1 - c)\big)\frac{\dot{\lambda}(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
&- \ ctw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle - (\alpha - 1 - c)t\|\dot{x}(t)\|^2 \\
&- \ t^2\beta(t)w(t)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.
\end{aligned}$$

We choose again $c := \alpha - 1$ and split the term $(\alpha - 1)tw(t)\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\rangle$ into the sum of two expressed in terms of ε given by (2.5). For every $t \geq t_0$, we have

$$\begin{aligned}
\dot{E}_{\alpha-1}(t) \ &\leq \ \big(2tw(t) + t^2\dot{w}(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\
&\quad - (\alpha - 1 - \varepsilon)tw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle - \varepsilon tw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle.
\end{aligned}$$

By applying the convex gradient inequality (1.5) we obtain for every $t \geq t_0$

$$\begin{aligned}
\dot{E}_{\alpha-1}(t) \ &\leq \ \big(2tw(t) + t^2\dot{w}(t) - (\alpha - 1 - \varepsilon)tw(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\
&\quad - \ \varepsilon tw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle \\
&= \ \big(t^2\dot{w}(t) - (\alpha - 3 - \varepsilon)\,tw(t)\big)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \\
&\quad - \ \varepsilon tw(t)\big\langle x(t) - z, \nabla\Phi_{\lambda(t)}(x(t))\big\rangle.
\end{aligned} \tag{2.16}$$

Since β is nondecreasing, for every $t \geq t_0$ it holds

$$b(t) = w(t) + \dot{\beta}(t) + \frac{\beta(t)}{t} \geq w(t),$$

thus, (2.5) leads to $t^2\dot{w}(t) - (\alpha - 3 - \varepsilon)tw(t) \leq 0$. Consequently, we obtain from (2.16) by integration

$$\int_{t_0}^{+\infty} tw(t)\langle \nabla\Phi_{\lambda(t)}(x(t)), x(t) - z\rangle dt \ \leq \ \frac{E_{\alpha-1}(t_0)}{\varepsilon} \ < \ +\infty.$$

□

Now we are in position to improve the convergence rates which we obtained previously in (2.10) and to derive from here convergence rates for Φ.

Theorem 2.2.3. *Suppose that $\alpha > 1$ and (2.5) holds, that λ and β are nondecreasing on $[t_0, +\infty)$, and that (2.3) holds. Assume in addition*

$$\int_{t_0}^{+\infty} \left[\frac{\left(\dot{\lambda}(t)\right)^2 t^3 \beta^2(t)}{\lambda^4(t)} - \frac{\dot{\lambda}(t) t^2 b(t)}{2\lambda^2(t)} \right]_+ dt < +\infty, \tag{2.17}$$

where $[\cdot]_+$ denotes the positive part of the expression inside the brackets, and that there exists $C > 0$ such that

$$\frac{d}{dt}\left(t^2 b(t)\right) \;\leq\; Ctb(t) \text{ for every } t \geq t_0. \tag{2.18}$$

Then, for a solution $x : [t_0, +\infty) \mapsto H$ to (2.1), it holds

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi = o\left(\frac{1}{t^2 b(t)}\right) \text{ and } \|\dot{x}(t)\| = o\left(\frac{1}{t}\right) \text{ as } t \to +\infty. \tag{2.19}$$

Moreover,

$$\|\nabla \Phi_{\lambda(t)}(x(t))\| \;=\; o\left(\frac{1}{t\sqrt{b(t)\lambda(t)}}\right) \text{ as } t \to +\infty, \tag{2.20}$$

and

$$\Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t))) - \min \Phi = o\left(\frac{1}{t^2 b(t)}\right) \text{ and } \|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\| \;=\; o\left(\frac{\sqrt{\lambda(t)}}{t\sqrt{b(t)}}\right) \text{ as } t \to +\infty. \tag{2.21}$$

Proof. First we notice that for every $t \geq t_0$ it holds

$$\begin{aligned}
\left\langle \frac{d}{dt}\left(\nabla \Phi_{\lambda(t)}(x(t))\right), \dot{x}(t) \right\rangle &= \left\langle \lim_{h \to 0} \frac{\nabla \Phi_{\lambda(t+h)}(x(t+h)) - \nabla \Phi_{\lambda(t)}(x(t))}{h}, \dot{x}(t) \right\rangle \\
&= \left\langle \lim_{h \to 0} \frac{\nabla \Phi_{\lambda(t+h)}(x(t+h)) - \nabla \Phi_{\lambda(t+h)}(x(t))}{h}, \dot{x}(t) \right\rangle \\
&\quad + \left\langle \lim_{h \to 0} \frac{\nabla \Phi_{\lambda(t+h)}(x(t)) - \nabla \Phi_{\lambda(t)}(x(t))}{h}, \dot{x}(t) \right\rangle.
\end{aligned}$$

For every $h > 0$, by the monotonicity of the gradient of a convex function, we have

$$\left\langle \frac{\nabla \Phi_{\lambda(t+h)}(x(t+h)) - \nabla \Phi_{\lambda(t+h)}(x(t))}{h}, \frac{x(t+h) - x(t)}{h} \right\rangle \;\geq\; 0,$$

so letting h tend to zero we obtain

$$\left\langle \lim_{h \to 0} \frac{\nabla \Phi_{\lambda(t+h)}(x(t+h)) - \nabla \Phi_{\lambda(t+h)}(x(t))}{h}, \dot{x}(t) \right\rangle \;\geq\; 0.$$

Consequently, for every $t \geq t_0$ it holds

$$
\begin{aligned}
&\left\langle \frac{d}{dt}\left(\nabla \Phi_{\lambda(t)}(x(t))\right), \dot{x}(t) \right\rangle \\
\geq\ & \left\langle \lim_{h\to 0} \frac{\nabla \Phi_{\lambda(t+h)}(x(t)) - \nabla \Phi_{\lambda(t)}(x(t))}{h}, \dot{x}(t) \right\rangle \\
=\ & \lim_{h\to 0} \left\langle \frac{(\lambda(t+h)) \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - \lambda(t) \operatorname{prox}_{(\lambda(t+h))\Phi}(x(t)) - \big(\lambda(t+h) - \lambda(t)\big) x(t)}{\lambda(t)\lambda(t+h)h}, \dot{x}(t) \right\rangle \\
=\ & \lim_{h\to 0} \left\langle \frac{\big(\lambda(t+h) - \lambda(t)\big)\big(\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\big)}{\lambda(t)\lambda(t+h)h}, \dot{x}(t) \right\rangle \\
&- \lim_{h\to 0} \left\langle \frac{\operatorname{prox}_{(\lambda(t+h))\Phi}(x(t)) - \operatorname{prox}_{\lambda(t)\Phi}(x(t))}{\lambda(t+h)h}, \dot{x}(t) \right\rangle \\
\geq\ & \frac{\dot{\lambda}(t)}{\lambda^2(t)} \left\langle \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t), \dot{x}(t) \right\rangle - \lim_{h\to 0} \frac{(\lambda(t+h) - \lambda(t)) \|\nabla \Phi_{\lambda(t)}(x(t))\| \|\dot{x}(t)\|}{\lambda(t+h)h} \\
=\ & \frac{\dot{\lambda}(t)}{\lambda^2(t)} \left\langle \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t), \dot{x}(t) \right\rangle - \frac{\dot{\lambda}(t) \|\nabla \Phi_{\lambda(t)}(x(t))\| \|\dot{x}(t)\|}{\lambda(t)} \\
=\ & -\frac{\dot{\lambda}(t)}{\lambda(t)} \left\langle \nabla \Phi_{\lambda(t)}(x(t)), \dot{x}(t) \right\rangle - \frac{\dot{\lambda}(t) \|\nabla \Phi_{\lambda(t)}(x(t))\| \|\dot{x}(t)\|}{\lambda(t)} \\
\geq\ & -\frac{2\dot{\lambda}(t) \|\nabla \Phi_{\lambda(t)}(x(t))\| \|\dot{x}(t)\|}{\lambda(t)},
\end{aligned}
$$

where we used (1.2), (1.3) and the Cauchy-Schwarz inequality. Now we multiply (2.1) by $t^2\dot{x}(t)$ to deduce, by using the inequality above and (2.6), for every $t \geq t_0$

$$
\begin{aligned}
0 &= t^2 \langle \ddot{x}(t), \dot{x}(t) \rangle + \alpha t \|\dot{x}(t)\|^2 + t^2 \beta(t) \left\langle \frac{d}{dt}\left(\nabla \Phi_{\lambda(t)}(x(t))\right), \dot{x}(t) \right\rangle \\
&\quad + t^2 b(t) \big\langle \nabla \Phi_{\lambda(t)}(x(t)), \dot{x}(t) \big\rangle \\
&\geq t^2 \frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t)\|^2\right) + \alpha t \|\dot{x}(t)\|^2 + t^2 b(t) \frac{d}{dt}\big(\Phi_{\lambda(t)}(x(t)) - \min \Phi\big) \\
&\quad + \frac{\dot{\lambda}(t) t^2 b(t)}{2} \|\nabla \Phi_{\lambda(t)} x(t)\|^2 - \frac{2t^2 \beta(t) \dot{\lambda}(t)}{\lambda(t)} \|\nabla \Phi_{\lambda(t)}(x(t))\| \|\dot{x}(t)\| \\
&\geq \frac{d}{dt}\left(\frac{t^2}{2}\|\dot{x}(t)\|^2 + t^2 b(t) (\Phi_{\lambda(t)}(x(t)) - \min \Phi)\right) + (\alpha - 1) t \|\dot{x}(t)\|^2 \\
&\quad - \big(\Phi_{\lambda(t)}(x(t)) - \min \Phi\big) \frac{d}{dt}\left(t^2 b(t)\right) \\
&\quad - \left\{ \left[\left(\frac{\dot{\lambda}(t)}{\lambda(t)}\right)^2 t^3 \beta^2(t) - \frac{\dot{\lambda}(t) t^2 b(t)}{2} \right] \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 + t \|\dot{x}(t)\|^2 \right\}.
\end{aligned}
$$

Using (2.18) we obtain for every $t \geq t_0$

$$\frac{d}{dt}\left(\frac{t^2}{2}\|\dot{x}(t)\|^2 + t^2 b(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi)\right) \leq \; [2-\alpha]_+ t\|\dot{x}(t)\|^2$$
$$+ \; \left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right)Ctb(t) + \left[\left(\frac{\dot{\lambda}(t)}{\lambda(t)}\right)^2 t^3\beta^2(t) - \frac{\dot{\lambda}(t)t^2 b(t)}{2}\right]_+ \|\nabla\Phi_{\lambda(t)}(x(t))\|^2.$$

Next we show the integrability of the right-hand side of the expression above. The first term is integrable according to Theorem 2.2.1 (v) and the second one is integrable according to Theorem 2.2.1 (vii). Further, since

$$\|\nabla\Phi_{\lambda(t)}(x(t)) - \nabla\Phi_{\lambda(t)}(z)\| \; \leq \; \frac{1}{\lambda(t)}\|x(t) - z\| \;\forall t \geq t_0,$$

and taking into the account the boundedness of the trajectory x established in Theorem 2.2.1 (vi) and that $z \in \operatorname{argmin}\Phi$, we deduce

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

So, under the assumption (2.17), we obtain that there exists $\widetilde{C} > 0$ such that for every $t \geq t_0$

$$\int_{t_0}^{t}\left[\left(\frac{\dot{\lambda}(s)}{\lambda(s)}\right)^2 s^3\beta^2(s) - \frac{\dot{\lambda}(s)s^2 b(s)}{2}\right]_+ \|\nabla\Phi_{\lambda(s)}(x(s))\|^2 ds$$
$$\leq \widetilde{C}\int_{t_0}^{t}\left[\frac{\left(\dot{\lambda}(s)\right)^2 s^3\beta^2(s)}{\lambda^4(s)} - \frac{\dot{\lambda}(s)s^2 b(s)}{2\lambda^2(s)}\right]_+ ds$$
$$< \; +\infty.$$

Applying Lemma 1.2.12, we conclude that the following limit

$$L := \lim_{t\to+\infty}\left(\frac{t^2}{2}\|\dot{x}(t)\|^2 + t^2 b(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi)\right) \geq 0$$

exists. We will show that $L = 0$. Supposing that $L > 0$, we deduce that there exists $t^* \geq t_0$ such that for every $t \geq t^*$

$$\frac{t}{2}\|\dot{x}(t)\|^2 + tb(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \geq \frac{L}{2t}.$$

Integrating the last inequality on $[t^*, +\infty)$, we arrive at the contradiction with the integrability of the left-hand side as proved in Theorem 2.2.1 (v) and (vii). Therefore, $L = 0$ and we obtain

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

Using the definition of the proximal mapping, we derive

$$\Phi_{\lambda(t)}(x(t))-\min\Phi = \Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t)))-\min\Phi+\frac{1}{2\lambda(t)}\|\operatorname{prox}_{\lambda(t)\Phi}(x(t))-x(t)\|^2 \quad (2.22)$$

$\forall t \geq t_0$, which yields

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

as $t \to +\infty$. According to (1.2) we obtain from here

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

□

2.3 Convergence of the trajectories

In this section we will investigate the weak convergence of the trajectory x to a minimizer of Φ.

Theorem 2.3.1. *Suppose that $\alpha > 1$, (2.3) and (2.5) hold and that λ and β are nondecreasing on $[t_0, +\infty)$. Assume in addition that*

$$\lim_{t\to+\infty} \frac{\beta(t)}{tw(t)} = 0 \quad (2.23)$$

and

$$\sup_{t\geq t_0} \frac{\lambda(t)}{t} < +\infty. \quad (2.24)$$

If $x : [t_0, +\infty) \mapsto H$ is a solution to (2.1), then $x(t)$ converges weakly to a minimizer of Φ as $t \to +\infty$.

Proof. Let $z \in \operatorname{argmin}\Phi$. Previously, in Theorem 2.2.1, we established the existence of the limit of $E_c(t)$ as $t \to +\infty$ for $c = \alpha - 1$ and $c = \alpha - 1 - \varepsilon$, where $\varepsilon \in (0, \alpha - 1)$ is given by (2.5). Thus, computing the difference

$$\begin{aligned} E_{\alpha-1-\varepsilon}(t) - E_{\alpha-1}(t) &= \varepsilon t\beta(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) + \frac{\varepsilon(\alpha-1)}{2}\|x(t)-z\|^2 \\ &\quad - \varepsilon\langle(\alpha-1)(x(t)-z) + t(\dot{x}(t) + \beta(t)\nabla\Phi_{\lambda(t)}(x(t)), x(t)-z\rangle \\ &= \varepsilon t\beta(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) - \frac{\varepsilon(\alpha-1)}{2}\|x(t)-z\|^2 \end{aligned}$$

$$- \quad \varepsilon t\langle \dot{x}(t) + \beta(t)\nabla\Phi_{\lambda(t)}(x(t)), x(t) - z\rangle,$$

we deduce that the limit of the right-hand side exists. Thanks to (2.10), we derive for every $t \geq t_0$

$$t\beta(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) \;\leq\; t\beta(t)\frac{E_{\alpha-1}(t_0)}{t^2 w(t)} \;=\; E_{\alpha-1}(t_0)\frac{\beta(t)}{tw(t)}$$

and from here, based on the assumption (2.23), we obtain

$$\lim_{t\to+\infty} t\beta(t)(\Phi_{\lambda(t)}(x(t)) - \min\Phi) = 0. \tag{2.25}$$

Hereby, we derived that the limit of the quantity

$$p(t) := \frac{\alpha-1}{2}\|x(t) - z\|^2 + t\langle \dot{x}(t), x(t) - z\rangle + t\beta(t)\langle \nabla\Phi_{\lambda(t)}(x(t)), x(t) - z\rangle$$

exists as $t \to +\infty$. Now we are ready to prove the existence of the limit of $\|x(t) - z\|$ as $t \to +\infty$. Denote

$$q(t) := \frac{\alpha-1}{2}\|x(t) - z\|^2 + (\alpha - 1)\int_{t_0}^{t} \beta(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle ds \;\forall t \geq t_0.$$

For every $t \geq t_0$, it holds that

$$p(t) = q(t) + \frac{t}{\alpha-1}\dot{q}(t) - (\alpha - 1)\int_{t_0}^{t} \beta(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle ds,$$

since

$$\dot{q}(t) = (\alpha - 1)\langle x(t) - z, \dot{x}(t)\rangle + \big(\alpha - 1\big)\big(\beta(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle\big)$$

and

$$\begin{aligned} q(t) + \frac{t}{\alpha-1}\dot{q}(t) \;&=\; \frac{\alpha-1}{2}\|x(t) - z\|^2 + (\alpha - 1)\int_{t_0}^{t} \beta(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle ds \\ &\quad+\; t\langle x(t) - z, \dot{x}(t)\rangle + t\big(\beta(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle\big). \end{aligned}$$

By Lemma 2.2.2 we established that $\int_{t_0}^{+\infty} sw(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle ds < +\infty$. In turn, (2.23) yields that

$$\lim_{t\to+\infty}\int_{t_0}^{t} \beta(s)\langle \nabla\Phi_{\lambda(s)}(x(s)), x(s) - z\rangle ds \text{ exists .} \tag{2.26}$$

Finally,

$$\lim_{t\to+\infty}\left(q(t) + \frac{t}{\alpha-1}\dot{q}(t)\right) \text{ also exists.}$$

Applying now Lemma 1.2.13, we immediately get the existence of the limit of $q(t)$ as $t \to +\infty$. By the definition of q and (2.26) we establish the first statement of the Opial's Lemma (see Lemma 1.2.14), namely, that, for any $z \in \operatorname{argmin} \Phi$

$$\lim_{t\to+\infty} \|x(t) - z\| \ \text{ exists.}$$

To establish the second claim of the Opial's Lemma, first note that from (2.22) and (2.25) we have, $\lim_{t\mapsto+\infty} t\beta(t)(\Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t))) - \min \Phi) = 0$ and $\lim_{t\mapsto+\infty} \frac{t\beta(t)}{\lambda(t)} \| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 = 0$. Using that β is nondecreasing and assumption (2.24), we deduce

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

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

$$\Phi(z) \ \leq \ \liminf_{k\to+\infty} \Phi(\operatorname{prox}_{\lambda(t_k)\Phi}(x(t_k))) \ = \ \lim_{t\to+\infty} \Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t))) \ = \ \min \Phi.$$

Hence, $z \in \operatorname{argmin} \Phi$, and the second statement of the Opial's Lemma is shown. This gives the weak convergence of the trajectory $x(t)$ to a minimizer of Φ as $t \mapsto +\infty$. □

Remark 2.3.2. In the hypotheses of Theorem 2.2.3, in order to obtain the convergence of the trajectories, besides (2.24) it is enough to assume that

$$\sup_{t\geq t_0} \frac{\beta(t)}{tw(t)} < +\infty$$

in order to guarantee (2.26). Indeed, in this case (2.25) follows from the conclusion of Theorem 2.2.3

$$\lim_{t\mapsto+\infty} t\beta(t)(\Phi_{\lambda(t)}(x(t)) - \min \Phi) \leq \lim_{t\mapsto+\infty} t^2 b(t)(\Phi_{\lambda(t)}(x(t)) - \min \Phi)\frac{\beta(t)}{tw(t)} = 0.$$

Remark 2.3.3. **(implicit discretization)** Implicit discretization of the dynamical system (2.1) with fixed step size $h > 0$ leads to the numerical scheme that reads for every $k \geq 1$ (see also [7,8])

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{h^2} + \frac{\alpha(x_{k+1} - x_k)}{kh^2} + \frac{\beta_k \left(\nabla\Phi_{\lambda_{k+1}}(x_{k+1}) - \nabla\Phi_{\lambda_k}(x_k)\right)}{h} + b_k\nabla\Phi_{\lambda_{k+1}}(x_{k+1}) \ = \ 0,$$

where x_k, λ_k, β_k and b_k denote $x(kh)$, $\lambda(kh)$, $\beta(kh)$ and $b(kh)$, respectively. Rearranging the terms one obtains for every $k \geq 1$

$$x_{k+1} + \frac{kh(\beta_k + hb_k)\nabla\Phi_{\lambda_{k+1}}(x_{k+1})}{\alpha + k} \ = \ x_k + \frac{k(x_k - x_{k-1})}{\alpha + k} + \frac{kh\beta_k\nabla\Phi_{\lambda_k}(x_k)}{\alpha + k}$$

or, equivalently,

$$(\forall k \geq 1) \quad \begin{cases} y_k & = \; x_k + \frac{k}{\alpha+k}(x_k - x_{k-1}) + \frac{kh\beta_k}{\alpha+k}\nabla\Phi_{\lambda_k}(x_k), \\ x_{k+1} & = \; \operatorname{prox}_{\frac{kh(\beta_k+hb_k)}{\alpha+k}\Phi_{\lambda_k}}(y_k). \end{cases}$$

Relation (1.2), namely,

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

and the second property of the proximal mapping from Lemma 1.2.8, namely,

$$\operatorname{prox}_{\mu\Phi_\lambda}(x) = \frac{\lambda}{\lambda+\mu}x + \frac{\mu}{\lambda+\mu}\operatorname{prox}_{(\lambda+\mu)\Phi}(x) \quad \forall x \in H, \forall \mu, \lambda > 0,$$

lead to the following formulation of the implicit numerical algorithm

$$(\forall k \geq 1) \quad \begin{cases} y_k & = x_k + \frac{k}{\alpha+k}(x_k - x_{k-1}) + \frac{kh\beta_k}{\lambda_k(\alpha+k)}(x_k - \operatorname{prox}_{\lambda_k\Phi}(x_k)), \\ x_{k+1} & = \frac{\lambda_k(\alpha+k)}{\lambda_k(\alpha+k)+kh(\beta_k+hb_k)}y_k \\ & \qquad + \frac{kh(\beta_k+hb_k)}{\lambda_k(\alpha+k)+kh(\beta_k+hb_k)}\operatorname{prox}_{\frac{\lambda_k(\alpha+k)+kh(\beta_k+hb_k)}{\alpha+k}\Phi}(y_k), \end{cases}$$

where $x_0,\ x_1 \in H$ are given starting points.

2.4 Polynomial choices for the system parameter functions

According to the previous two sections, in order to guarantee both the fast convergence rates in Theorem 2.2.3 and the convergence of the trajectory to a minimizer of Φ in Theorem 2.3.1, by taking also into account Remark 2.3.2, it is enough to make the following assumptions on the system parameter functions

(I) $\alpha > 1$ and there exists $\varepsilon \in (0, \alpha - 1)$ such that $(\alpha-3)w(t) - t\dot{w}(t) \geq \varepsilon b(t)$ for every $t \geq t_0$;

(II) β and λ are nondecreasing on $[t_0, +\infty)$;

(III) $b(t) > \dot{\beta}(t) + \frac{\beta(t)}{t}$ for every $t \geq t_0$;

(IV) $\int_{t_0}^{+\infty}\left[\frac{\beta^2(t)(\dot{\lambda}(t))^2t^3}{\lambda^4(t)} - \frac{\dot{\lambda}(t)t^2b(t)}{2\lambda^2(t)}\right]_+ dt < +\infty$;

(V) there exists $C > 0$ such that $\frac{d}{dt}\left(t^2b(t)\right) \;\leq\; Ctb(t)$ for every $t \geq t_0$;

(VI) $\sup_{t\geq t_0}\frac{\beta(t)}{tw(t)} < +\infty$;

(VII) $\sup_{t\geq t_0}\frac{\lambda(t)}{t} \;<\; +\infty$.

In this section we will investigate the fulfillment of these conditions for

$$b(t) = bt^n, \quad \beta(t) = \beta t^m \quad \text{and} \quad \lambda(t) = \lambda t^l,$$

where $n, m, l \in \mathbb{R}$, $b, \lambda > 0$ and $\beta \geq 0$.

For this choice of b, condition (V) is fulfilled.

We assume first that $\beta = 0$. Then the conditions (III), (IV) and (VI) are fulfilled, while the conditions (II) and (VI) are nothing else than $0 \leq l \leq 1$. Condition (I) asks for $\alpha > 1$ and for the existence of $\varepsilon \in (0, \alpha - 1)$ such that for every $t \geq t_0$

$$(\alpha - 3 - n - \varepsilon)bt^n \geq 0$$

or, equivalently, $\alpha - 3 - n \geq \varepsilon$. To this end it is enough to have that $\alpha - 3 > n$.

In case $\beta > 0$, conditions (II) and (VII) are nothing else than $m \geq 0$ and $0 \leq l \leq 1$. Condition (III) reads for every $t \geq t_0$

$$bt^n > m\beta t^{m-1} + \beta t^{m-1} = (m+1)\beta t^{m-1},$$

or, equivalently,

$$t^{n-m+1} > \frac{(m+1)\beta}{b}.$$

From here we get

$$0 \leq m \leq n + 1.$$

and $b > (m+1)\beta t_0^{m-1-n}$.

Condition (VI) requires that

$$\sup_{t \geq t_0} \frac{\beta t^m}{t(bt^n - \beta m t^{m-1} - \beta t^{m-1})} = \sup_{t \geq t_0} \frac{\beta t^m}{bt^{n+1} - \beta t^m (m+1)} < +\infty$$

and it is obviously fulfilled.

Condition (I) asks for $\alpha > 1$ and for the existence of $\varepsilon \in (0, \alpha - 1)$ such that for every $t \geq t_0$

$$(\alpha - 3)(bt^n - \beta m t^{m-1} - \beta t^{m-1}) - t(bnt^{n-1} - \beta m(m-1)t^{m-2} - \beta(m-1)t^{m-2}) \geq \varepsilon bt^n.$$

After simplification we obtain that for every $t \geq t_0$

$$(\alpha - 3 - n - \varepsilon)bt^n + \beta(m+1)(m+2-\alpha)t^{m-1} \geq 0$$

or, equivalently,

$$(\alpha - 3 - n - \varepsilon)bt^{n-m+1} \geq \beta(m+1)(\alpha - m - 2).$$

On the one hand we have $m = n + 1$ and $(\alpha - 3 - n)\left(1 - \frac{\beta(n+2)}{b}\right) > \varepsilon$, which requires that $\alpha - 3 - n > 0$. On the other hand, we have $m < n + 1$, which also requires that $\alpha - 3 - n > 0$.

Consequently, we have to assume that

$$\alpha - 3 > n \quad \text{and} \quad b > \frac{\beta(m+1)(\alpha - m - 2)}{(\alpha - 3 - n)t_0^{n-m+1}}.$$

In this case, there will be always an $\varepsilon \in (0, \alpha - 1)$ such that $\alpha - 3 - n - \varepsilon > 0$ and

$$b > \frac{\beta(m+1)(\alpha - m - 2)}{(\alpha - 3 - n - \varepsilon)t_0^{n-m+1}} > \frac{\beta(m+1)(\alpha - m - 2)}{(\alpha - 3 - n)t_0^{n-m+1}},$$

in other words, which satisfies condition (I).

Finally, let us have a closer look at condition (V). This reads as

$$\int_{t_0}^{+\infty} \left[\frac{\beta^2 t^{2m} (l\lambda)^2 t^{2l-2} t^3}{\lambda^4 t^{4l}} - \frac{l\lambda t^{l-1} t^2 b t^n}{2\lambda^2 t^{2l}} \right]_+ dt < +\infty$$

or, equivalently,

$$\begin{aligned}
&\int_{t_0}^{+\infty} \left[\left(\frac{\beta l}{\lambda} \right)^2 t^{2m-2l+1} - \frac{lb}{2\lambda} t^{n-l+1} \right]_+ dt \\
= &\int_{t_0}^{+\infty} \left[\left(\frac{\beta l}{\lambda} \right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} \right]_+ t^{2m-2l+1} dt \\
< &\ +\infty.
\end{aligned}$$

1. In case

$$n + l > 2m,$$

there exists $t_1 \geq t_0$ such that for every $t \geq t_1$

$$\left(\frac{\beta l}{\lambda} \right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} \leq 0.$$

Therefore, we obtain

$$\begin{aligned}
&\int_{t_0}^{+\infty} \left[\left(\frac{\beta l}{\lambda} \right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} \right]_+ t^{2m-2l+1} dt \\
= &\int_{t_0}^{t_1} \left[\left(\frac{\beta l}{\lambda} \right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} \right]_+ t^{2m-2l+1} dt + \int_{t_1}^{+\infty} \left[\left(\frac{\beta l}{\lambda} \right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} \right]_+ t^{2m-2l+1} dt \\
= &\int_{t_0}^{t_1} \left[\left(\frac{\beta l}{\lambda} \right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} \right]_+ t^{2m-2l+1} dt < +\infty,
\end{aligned}$$

thus (V) is fulfilled.

2. In case

$$n + l < 2m,$$

there exist $\delta > 0$ and $t_2 \geq t_0$ such that for every $t \geq t_2$

$$\left(\frac{\beta l}{\lambda}\right)^2 - \frac{lb}{2\lambda} t^{n+l-2m} > \delta.$$

Taking into account that $2m - 2l + 1 > n + 1 - l \geq -1$, we have

$$\begin{aligned} &\int_{t_0}^{+\infty} \left[\left(\frac{\beta l}{\lambda}\right)^2 - \frac{lb}{2\lambda} t^{n+l-2m}\right]_+ t^{2m-2l+1} dt \\ \geq &\int_{t_0}^{t_2} \left[\left(\frac{\beta l}{\lambda}\right)^2 - \frac{lb}{2\lambda} t^{n+l-2m}\right]_+ t^{2m-2l+1} dt + \delta \int_{t_2}^{+\infty} t^{2m-2l+1} dt \ = +\infty, \end{aligned}$$

thus (V) is not fulfilled.

3. It is only left to consider the case

$$n + l = 2m.$$

Condition (V) becomes

$$\int_{t_0}^{+\infty} \left[\left(\frac{\beta l}{\lambda}\right)^2 - \frac{lb}{2\lambda}\right]_+ t^{2m-2l+1} dt < +\infty.$$

If $b \geq \frac{2\beta^2 l}{\lambda}$, then it is fulfilled. Otherwise, since $2m - 2l + 1 = n - l + 1 \geq -1$, it is not fulfilled.

Summarising, all convergence statements in Theorem 2.2.3 and Theorem 2.3.1 hold in the two settings

1. $\alpha > 1$, $\beta = 0$, $\alpha - 3 > n$, $0 \leq l \leq 1$, and $b, \lambda > 0$;
2. $\alpha > 1$, $\beta > 0$, $\alpha - 3 > n$, $0 \leq l \leq 1$, $0 \leq m \leq n+1$, $b > \frac{(m+1)(\alpha-m-2)\beta}{(\alpha-3-n)t_0^{n-m+1}}$, $\lambda > 0$, and either $2m < n + l$, or $2m = n + l$ and $b \geq \frac{2l\beta^2}{\lambda}$.

Remark 2.4.1. Theorem 2.2.3 is providing for the choices $b(t) = bt^n$ and $\lambda(t) = \lambda t^l$ the following convergence rates

$$\Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t))) - \min \Phi = o\left(\frac{1}{t^{n+2}}\right), \quad \|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\| \ = \ o\left(\frac{1}{t^{\frac{n}{2}+1-\frac{l}{2}}}\right)$$

and

$$\|\nabla \Phi_{\lambda(t)}(x(t))\| \ = \ o\left(\frac{1}{t^{\frac{n}{2}+1+\frac{l}{2}}}\right),$$

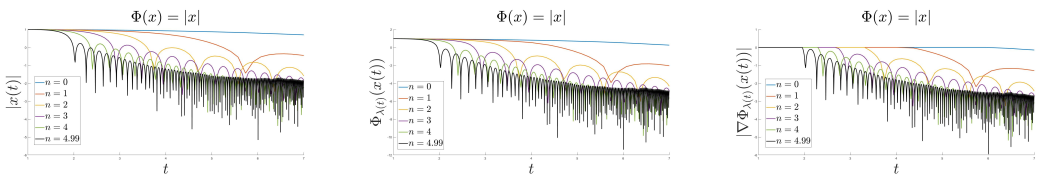
as $t \to +\infty$. Clearly, the bigger the n is the faster the convergence is. On the other hand, concerning the exponent l things are a bit more complicated: we may gain in one case, but inevitably lose in the other. Interesting case is when $l = 0$, which corresponds to λ being a constant function. In this case, one can notice a balance between accelerating the convergence of $\|\nabla \Phi_{\lambda(t)}(x(t))\|$ and slowing the latter for $\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|$, since none of them are affected by l anymore.

2.5 Numerical experiments

In this section we will conduct series of experiments to investigate the influence of the system parameters λ, β and b on the convergence behaviour of dynamical system. We will successively fix two of them and vary the last one in order to do so. For the numerical experiments we will restrict ourselves to the polynomial choices addressed in the previous section $\lambda(t) = t^l$, $\beta(t) = t^m$, $b(t) = bt^n$ with $b = \frac{(m+1)(\alpha-m-2)\beta}{(\alpha-3-n)t_0^{n-m+1}} + 1$, as well as $x(t_0) = x_0 = 10$, $\dot{x}(t_0) = 0$, and $t_0 = 1$.

2.5.1 The influence of b on the dynamical behaviour

First let us choose as objective function $\Phi : \mathbb{R} \mapsto \mathbb{R}_+, \Phi(x) = |x|$, fix $m = 0$, $\alpha = 9$ and $l = 1$, and vary n.



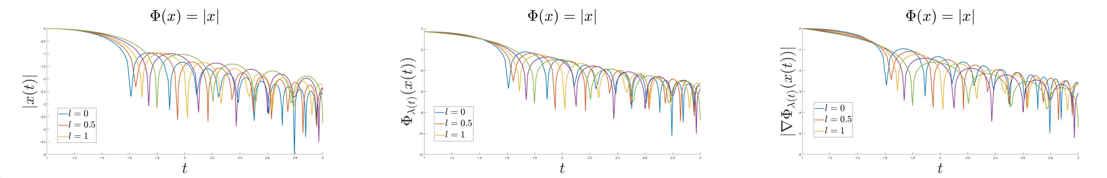
(a) Trajectories (b) Moreau envelope values (c) Moreau envelope gradient

Figure 2.1: $m = 0$, $\alpha = 9$ and $l = 1$

In Figure 2.1 we clearly see that the faster the exponent of the function b grows the faster the convergence of the function values of the Moreau envelope and its gradient are, starting with the slowest pace for $n = 0$ and accelerating until $n = 4.99$, confirming the theoretical convergence rates. In addition, the increase in the exponent of b seems to improve the convergence behaviour of the trajectory, too. Fast growing exponents for b will improve the convergence greatly, however, as seen in the previous section, they are limited by the upper bound value $\alpha - 3$.

2.5.2 The influence of λ on the dynamical behaviour

For the same objective function as in the previous subsection, we study the behaviour of the dynamics when varying the exponent l to investigate the influence of the function λ. To this end we fix $m = 0$, $\alpha = 9$ and $n = 5 < \alpha - 3$, and take for l three different values from 0 to 1. We also choose the starting point $x_0 = 1$ in order to provide a better illustration.



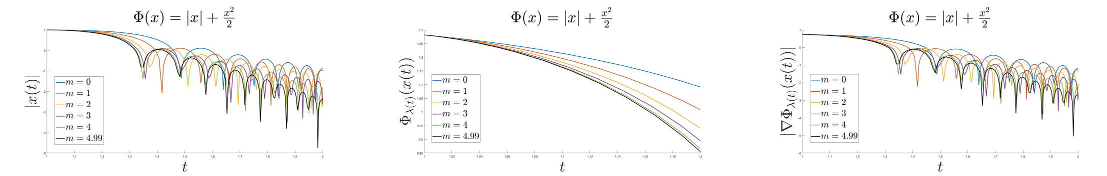
(a) Trajectories (b) Moreau envelope values (c) Moreau envelope gradient

Figure 2.2: $m = 0$, $\alpha = 9$ and $n = 5$

One can notice in Figure 2.2 that the convergence behaviour of the functions values of the Moreau envelope and its gradient is better the higher l is, whereas, interestingly enough, for the convergence of the trajectories an opposite phenomenon takes place.

2.5.3 The influence of β on the dynamical behaviour

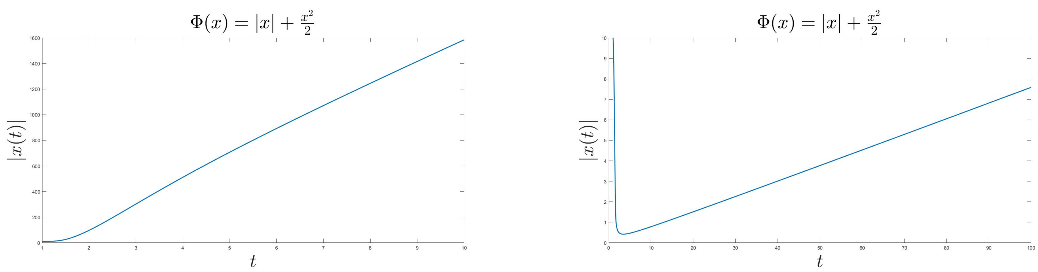
Let $\Phi : \mathbb{R} \mapsto \mathbb{R}_+, \Phi(x) = |x| + \frac{x^2}{2}$, $\alpha = 13$, $n = 9 < \alpha - 3$ and $l = 1$. We vary the exponent m such that $2m < n + l$ to study the influence of the function β on the convergence behaviour of the system.



(a) Trajectories (b) Moreau envelope values (c) Moreau envelope gradient

Figure 2.3: $n = 9$, $\alpha = 13$ and $l = 1$

In Figure 2.3 we see that, even though m does not explicitly appear in the theoretical convergence rates for the gradient of the Moreau envelope and the trajectory of the system, it influences the convergence behaviour of both of them as well as of the function values of the Moreau envelope, in the sense that these are faster the higher the values of m are.



(a) $n = 9$, $\alpha = 13$, $l = 1$ and $m = 12$ (b) $n = 4$, $\alpha = 2$, $l = 4$ and $m = 6$

Figure 2.4: Divergence of the trajectories

Finally, we consider two parameter choices which lie outside the convergence setting derived in the previous section and notice that these fundamentally affects the convergence of the trajectory. In Figure 2.4(a) we choose m such that that condition $2m < n+l$ is violated, and in Figure 2.4(b) we choose α and n such that the condition $\alpha - 3 > n$ is also violated. One can see that in both settings the trajectories diverge.

3 Nonsmooth optimization with time scaling and Tikhonov regularization

This chapter is devoted to studying the convergence properties of the following second order in time differential equation in connection with the minimization problem (1.1)

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t)) + b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0, \qquad (3.1)$$

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 and the function $\varepsilon : [t_0, +\infty) \mapsto \mathbb{R}_+$ is continuously differentiable and non-increasing with the property $\lim_{t\to+\infty} \varepsilon(t) = 0$. 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 [30]) the fast rates of convergence for the Moreau envelope of the objective function and the objective function itself to $\min\Phi$, 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 $\operatorname{argmin}\Phi$. To conclude, we provide multiple numerical results in order to illustrate our theoretical discoveries.

3.1 Preparatory results

Let us mention two key properties of the Tikhonov regularization, which we will use later in the analysis (see, for instance, [16] or [32] 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 $x_{\varepsilon(t),\lambda(t)} = \operatorname{argmin}\varphi_{\varepsilon(t),\lambda(t)}$. Thus, the first order optimality condition reads as

$$\nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) + \varepsilon(t)x_{\varepsilon(t),\lambda(t)} = 0. \qquad (3.2)$$

Now we are ready to formulate the following result:

Lemma 3.1.1. *Suppose that*

$$\lim_{t\to+\infty} \lambda(t)\varepsilon(t) \ = \ 0. \qquad (3.3)$$

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

$$\textit{for } x^* = \operatorname{proj}_{\operatorname{argmin}\Phi}(0), \ \|x_{\varepsilon(t),\lambda(t)}\| \leq \|x^*\| \textit{ for all } t \geq t_0 \tag{3.4}$$

and

$$\lim_{t\to+\infty} \|x_{\varepsilon(t),\lambda(t)} - x^*\| = 0. \tag{3.5}$$

Proof. Suppose that $t \geq t_0$. By the monotonicity of $\nabla\Phi_{\lambda(t)}$ we deduce

$$\left\langle \nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \nabla\Phi_{\lambda(t)}(x^*), x_{\varepsilon(t),\lambda(t)} - x^* \right\rangle \ \geq \ 0 \text{ for all } t \geq t_0.$$

By (3.2) we obtain

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

Using Cauchy-Schwarz inequality we derive

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

This proves the first claim. For the second one, consider (3.2) again and note that it is equivalent to

$$x_{\varepsilon(t),\lambda(t)} = \operatorname{prox}_{\frac{1}{\varepsilon(t)}\Phi_{\lambda(t)}}(0) = \frac{\operatorname{prox}_{\left(\lambda(t)+\frac{1}{\varepsilon(t)}\right)\Phi}(0)}{\lambda(t)\varepsilon(t)+1} \tag{3.6}$$

by the second claim of Lemma 1.2.8. Note that by (3.3) we have $\lambda(t) + \frac{1}{\varepsilon(t)} \to +\infty$ and $\lambda(t)\varepsilon(t) + 1 \to 1$, as $t \to +\infty$. From now on, the proof is inspired by Theorem 23.44 of [32]. Take $z \in \operatorname{argmin}\Phi = \operatorname{argmin}\Phi_\lambda$ for each $\lambda > 0$. From (3.6) and from the fact that the resolvent of maximally monotone operator is maximally monotone and firmly nonexpansive (see, for instance, Corollary 23.11(i) of [32]) and Cauchy-Schwarz inequality it follows that for every $t \geq t_0$ (note that z could be represented as $z = \operatorname{prox}_{\frac{1}{\varepsilon(t)}\Phi_{\lambda(t)}}(z)$)

$$\left\|z - x_{\varepsilon(t),\lambda(t)}\right\| \|z\| \ \geq \ \left\langle z - x_{\varepsilon(t),\lambda(t)}, z\right\rangle \ \geq \ \left\|z - x_{\varepsilon(t),\lambda(t)}\right\|^2, \tag{3.7}$$

which gives the boundedness of $x_{\varepsilon(t),\lambda(t)}$ for every $t \geq t_0$. Now, let y be a weak sequential cluster point of $\{x_{\varepsilon(t_k),\lambda(t_k)}\}_{k\in\mathbb{N}}$, namely, $x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})} \rightharpoonup y$, as $n \to +\infty$. From (3.2) we deduce

$$\nabla\Phi_{\lambda(t_{k_n})}(x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}) + \varepsilon(t_{k_n})x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})} = 0.$$

Using (see Remark (1.2.7))

$$\nabla\Phi_\lambda \ = \ (\partial\Phi)_\lambda \ = \ \frac{1}{\lambda}\left(\operatorname{Id} - (\operatorname{Id} + \lambda\partial\Phi)^{-1}\right) \quad \forall\lambda > 0$$

we further obtain

$$(\operatorname{Id} + \lambda(t_{k_n})\partial\Phi)^{-1}\left(x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\right) \ = \ \left(\lambda(t_{k_n})\varepsilon(t_{k_n}) + 1\right)x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})},$$

which is equivalent to

$$\begin{aligned} x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})} &\in \Big(\lambda(t_{k_n})\varepsilon(t_{k_n})+1\Big)x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})} \\ &\quad + \lambda(t_{k_n})\partial\Phi\left(\left(\lambda(t_{k_n})\varepsilon(t_{k_n})+1\right)x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\right) \end{aligned}$$

or

$$\frac{-\varepsilon(t_{k_n})}{\lambda(t_{k_n})\varepsilon(t_{k_n})+1}x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})} \in \partial\Phi\left(x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\right). \tag{3.8}$$

The sequence

$$\left\{x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}, \frac{-\varepsilon(t_{k_n})}{\lambda(t_{k_n})\varepsilon(t_{k_n})+1}x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\right\}$$

lies in $\operatorname{gra}\partial\Phi$ by (3.8) and converges to $(y,0)$ in $H^{weak}\times H^{strong}$ topology due to the sequence $\{x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\}_{n\in\mathbb{N}}$ being also bounded and (3.3). Therefore, since $\operatorname{gra}\partial\Phi$ is sequentially closed (see Proposition 20.38(ii) of [32]) it follows that $y\in\operatorname{argmin}\Phi$. From (3.7) we derive

$$\|y-x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\|^2 \leq \left\langle y-0, y-x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\right\rangle \to 0, \text{ as } n\to+\infty,$$

by the definition of weak convergence, thus, $x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\to y$, as $n\to+\infty$. On the other hand, (3.7) leads to

$$\begin{aligned} 0 &\geq \|z-x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\|^2 - \left\langle z-x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}, z-0\right\rangle \\ &= \left\langle z-x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}, 0-x_{\varepsilon(t_{k_n}),\lambda(t_{k_n})}\right\rangle \to \langle z-y, 0-y\rangle \end{aligned}$$

and thus $y=x^*$ by the characterization of x^*, namely,

$$\text{for } x^*\in\operatorname{argmin}\Phi \text{ and } \forall z\in\operatorname{argmin}\Phi \text{ it holds that } \langle z-x^*, 0-x^*\rangle\leq 0.$$

So, x^* being the only weak sequential cluster point of the bounded sequence $\left\{x_{\varepsilon(t_k),\lambda(t_k)}\right\}_{k\in\mathbb{N}}$ means that $x_{\varepsilon(t_k),\lambda(t_k)}\rightharpoonup x^*$, as $k\to+\infty$, by Lemma 2.46 of [32]. By (3.7) again we deduce

$$\|x_{\varepsilon(t_k),\lambda(t_k)}-x^*\|^2 \leq \left\langle 0-x^*, x_{\varepsilon(t_k),\lambda(t_k)}-x^*\right\rangle \to 0, \text{ as } k\to+\infty$$

and so the second claim follows.

□

Our nearest goal is to deduce the existence and uniqueness of the solutions of the dynamical system (3.1). Suppose $\beta>0$. Let us integrate (3.1) from t_0 to t to obtain

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

Denoting $z(t) := \int_{t_0}^{t} \left(\frac{\alpha}{s}\dot{x}(s) + b(s)\nabla\Phi_{\lambda(s)}(x(s)) + \varepsilon(s)x(s)\right) ds - \left(\dot{x}(t_0) + \beta\nabla\Phi_{\lambda(t_0)}(x_0)\right)$ 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 (3.1) is equivalent to

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

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

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

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

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

In case $\beta = 0$ for every $t \geq t_0$, (3.1) can be equivalently written as

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

Based on the two reformulations of the dynamical system (3.1) 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 [12] or of Theorem 1.1 in [15] with some small adjustments.

Theorem 3.1.2. *Suppose that there exists $\lambda_0 > 0$ such that $\lambda(t) \geq \lambda_0$ for every $t \geq t_0$. Then for every $(x_0, \dot{x}(t_0)) \in H \times H$ there exists a unique strong global solution $x : [t_0, +\infty) \mapsto H$ of the continuous dynamics (3.1) which satisfies the Cauchy initial conditions $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$.*

3.2 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. The construction of the energy function is inspired by [30]. For $2 \leq q \leq \alpha - 1$ we define

$$\begin{aligned} E_q(t) &= (t^2 b(t) - \beta(q+2-\alpha)t)\left(\Phi_{\lambda(t)}(x(t)) - \min \Phi\right) + \frac{t^2\varepsilon(t)}{2}\|x(t)\|^2 \\ &+ \frac{1}{2}\|q(x(t)-x^*) + t(\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\|^2 + \frac{q(\alpha-1-q)}{2}\|x(t)-x^*\|^2. \end{aligned} \tag{3.9}$$

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

$$(\alpha-3)tb(t) - t^2\dot{b}(t) + \beta(2-\alpha) \;\geq\; 0, \tag{3.10}$$

$$\exists a \geq 1 \text{ such that } \; 2\dot{\varepsilon}(t) \;\leq\; -a\beta\varepsilon^2(t), \tag{3.11}$$

$$\int_{t_0}^{+\infty} t\varepsilon(t)dt \;<\; +\infty, \tag{3.12}$$

$$b(t_0) \;\geq\; \frac{\beta}{t_0} \text{ and } b(t_0) > \frac{1}{a} \tag{3.13}$$

and

$$\exists \delta : 0 < \delta < \alpha - 3 \text{ such that } (\alpha-3)tb(t) - t^2\dot{b}(t) + \beta(2-\alpha) \;\geq\; \delta t b(t). \tag{3.14}$$

Theorem 3.2.1. *Suppose $\alpha \geq 3$ and assume that (3.10) – (3.13) hold for every $t \geq t_0$. Then*

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

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

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

$$t\varepsilon(t)\|x(t)-x^*\|^2,\; t\varepsilon(t)\|x(t)\|^2, \left((\alpha-3)tb(t) - t^2\dot{b}(t) + \beta(2-\alpha)\right)\left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right)$$

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

If, in addition, $\alpha > 3$ and (3.14) holds, then the trajectory x is bounded and

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

and

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

Proof. Let us compute the time derivative of the energy function. For every $t \geq t_0$ using (1.4) we derive

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

Define $v(t) = q(x(t) - x^*) + t(\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))$ $\forall t \geq t_0$. Using (3.1) to replace $\ddot{x}(t) + \beta\frac{d}{dt}\nabla\Phi_{\lambda(t)}(x(t))$ we obtain

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

By (3.13) one has $b(t) - \frac{\beta}{t} > 0$ for every $t \geq t_0$, 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

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

or

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

Therefore, for every $t \geq t_0$

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

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

$$\begin{aligned}
\dot{E}_q(t) \ \leq \ & \Big((2-q)tb(t) + t^2\dot{b}(t) - \beta(2-\alpha)\Big)\left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right) + (q+1-\alpha)t\|\dot{x}(t)\|^2 \\
- \ & \left(\Big(t^2b(t) - \beta(q+2-\alpha)t\Big)\frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2\left(b(t) - \frac{1}{a}\right)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
+ \ & \frac{2(2-q)t\varepsilon(t) + 2t^2\dot{\varepsilon}(t) + a\beta t^2\varepsilon^2(t)}{4}\|x(t)\|^2 - qt\frac{\varepsilon(t)}{2}\|x(t) - x^*\|^2 \\
+ \ & qt\frac{\varepsilon(t)}{2}\|x^*\|^2
\end{aligned} \tag{3.15}$$

for every $t \geq t_0$. Note that $b(t) - \frac{1}{a} > 0$ for every $t \geq t_0$. Then, due to the properties of b, there exists $t^* \geq t_0$ such that $t^2b(t) - \beta(q+2-\alpha)t \ \geq \ 0$ for every $t \geq t^*$ and all $q \in (2, \alpha-1]$. Therefore, since $\dot{\lambda}(t) \geq 0$ for every $t \geq t_0$, there exists t^{**}, namely, $t^{**} = \max\left\{t^*, \frac{\beta}{b(t_0) - \frac{1}{a}}\right\}$, such that

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

Consider now two cases with $t \geq t^{**}$. First, take $q = \alpha - 1$ to obtain from (3.15)

$$\begin{aligned}
\dot{E}_{\alpha-1}(t) \ \leq \ & \Big((3-\alpha)tb(t) + t^2\dot{b}(t) - \beta(2-\alpha)\Big)\left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right) \\
- \ & \left(\Big(t^2b(t) - \beta t\Big)\frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2\left(b(t) - \frac{1}{a}\right)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
+ \ & \frac{2(3-\alpha)t\varepsilon(t) + 2t^2\dot{\varepsilon}(t) + a\beta t^2\varepsilon^2(t)}{4}\|x(t)\|^2 - (\alpha-1)t\frac{\varepsilon(t)}{2}\|x(t) - x^*\|^2 \\
+ \ & (\alpha-1)t\frac{\varepsilon(t)}{2}\|x^*\|^2
\end{aligned} \tag{3.16}$$

for every $t \geq t_0$. Under the assumptions (3.10) and (3.11) we conclude starting from t^{**} that

$$\dot{E}_{\alpha-1}(t) \ \leq \ \frac{(\alpha-1)t\varepsilon(t)}{2}\|x^*\|^2. \tag{3.17}$$

Under the assumption (3.12) 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 1.2.12 and, therefore, $t \mapsto E_{\alpha-1}(t)$ is bounded, which leads to

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

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 (3.16) one may obtain the integrability of $t \mapsto t\varepsilon(t)\|x(t)-x^*\|^2$ as well as the other terms in (3.16). Consider now $q = \alpha - 1 - \delta$, where δ is defined by (3.14). Thus, (3.15) becomes

$$\begin{aligned}\dot{E}_{\alpha-1-\delta}(t) \ \leq\ & \Big((3-\alpha+\delta)tb(t) + t^2\dot{b}(t) - \beta(2-\alpha)\Big)\left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right) - \delta t\|\dot{x}(t)\|^2 \\ &- \left(\left(t^2 b(t) - \beta(1-\delta)t\right)\frac{\dot{\lambda}(t)}{2} - \beta^2 t + \beta t^2\left(b(t) - \frac{1}{a}\right)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\ &+ \frac{2(3-\alpha+\delta)t\varepsilon(t) + 2t^2\dot{\varepsilon}(t) + a\beta t^2\varepsilon^2(t)}{4}\|x(t)\|^2 \\ &- (\alpha-1-\delta)t\frac{\varepsilon(t)}{2}\|x(t)-x^*\|^2 + (\alpha-1-\delta)t\frac{\varepsilon(t)}{2}\|x^*\|^2.\end{aligned} \tag{3.18}$$

Under the assumptions (3.11) and (3.14) 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 (3.18) one may additionally obtain the integrability of $t \mapsto t\|\dot{x}(t)\|^2$. From the integrability of $t \mapsto \Big((\alpha-3)tb(t) - t^2\dot{b}(t) - \beta(\alpha-2)\Big)\left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right)$ and (3.14) we deduce

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

□

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

Theorem 3.2.2. *Assume that* $\alpha > 3$ *and* (3.11) – (3.14) *hold. Then*

$$t \mapsto t \left\langle \left(b(t) - \frac{\beta}{t} \right) \nabla \Phi_{\lambda(t)}(x(t)), x(t) - x^* \right\rangle \in L^1\big([t_0, +\infty), \mathbb{R}\big). \tag{3.19}$$

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

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

which in particular means

$$\begin{aligned} \Phi_{\lambda(t)}(x(t)) - \min \Phi &= o\left(\frac{1}{t^2 b(t)} \right) \text{ as } t \to +\infty, \\ \|\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\| &= o\left(\frac{1}{t} \right) \text{ as } t \to +\infty \end{aligned} \tag{3.20}$$

and moreover,

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

and

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

Proof. (i) Let us first prove an auxiliary estimate (3.19), which will allow us to obtain the rest of the desired results. We return to

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

Under condition (3.11) we deduce starting from t^{**}

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

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

$$\begin{aligned}&\int_{t_0}^t qs\left\langle\left(b(s) - \frac{\beta}{s}\right)\nabla\Phi_{\lambda(s)}(x(s)), x(s) - x^*\right\rangle ds\\ &\leq E_q(t_0) - E_q(t) + \int_{t_0}^t s\varepsilon(s)\|x(s)\|^2 ds\\ &\quad + \int_{t_0}^t \Big(2sb(s) + s^2\dot{b}(s) - \beta(q+2-\alpha)\Big)\left(\Phi_{\lambda(s)}(x(s)) - \min\Phi\right) ds\\ &\quad - \int_{t_0}^t qs\langle\varepsilon(s)x(s), x(s) - x^*\rangle.\end{aligned} \tag{3.21}$$

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

$$-\,qt\langle\varepsilon(t)x(t), x(t) - x^*\rangle \ \leq \ \frac{qt\varepsilon(t)}{2}\left(\|x(t)\|^2 + \|x(t) - x^*\|^2\right). \tag{3.22}$$

Notice that by (3.14) 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 3.2.1 we know that $t \mapsto tb(t)\left(\Phi_{\lambda(s)}(x(s)) - \min\Phi\right)$ is integrable and therefore so is $t \mapsto \Big(2tb(t) + t^2\dot{b}(t) - \beta(q+2-\alpha)\Big)\left(\Phi_{\lambda(s)}(x(s)) - \min\Phi\right)$. Since the function $t \mapsto E_q(t)$ is bounded and the rest of the right-hand side of (3.21) belongs to $L^1\left([t_0, +\infty), \mathbb{R}\right)$ by Theorem 3.2.1 and (3.22), we conclude with (3.19) due to (3.13).

(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 [30,33]) for two different $q_1, q_2 \in (2, \alpha-1)$ and for every $t \geq t_0$ the difference

$$\begin{aligned}
E_{q_1}(t)-E_{q_2}(t) \;&=\; (t^2b(t)-\beta(q_1+2-\alpha)t)\left(\Phi_{\lambda(t)}(x(t))-\min\Phi\right)+\frac{t^2\varepsilon(t)}{2}\|x(t)\|^2\\
&+\;\frac{1}{2}\|q_1(x(t)-x^*)+t(\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\|^2+\frac{q_1(\alpha-1-q_1)}{2}\|x(t)-x^*\|^2\\
&-(t^2b(t)-\beta(q_2+2-\alpha)t)\left(\Phi_{\lambda(t)}(x(t))-\min\Phi\right)-\frac{t^2\varepsilon(t)}{2}\|x(t)\|^2\\
&-\;\frac{1}{2}\|q_2(x(t)-x^*)+t(\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\|^2-\frac{q_2(\alpha-1-q_2)}{2}\|x(t)-x^*\|^2\\
=(q_1-q_2)&\Bigg(-\beta t\left(\Phi_{\lambda(t)}(x(t))-\min\Phi\right)+t\left\langle\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t)),x(t)-x^*\right\rangle\\
&+\;\frac{\alpha-1}{2}\|x(t)-x^*\|^2\Bigg).
\end{aligned}$$

As we have established earlier in Theorem 3.2.1 the limits of $E_{q_1}(t)-E_{q_2}(t)$ and $t\left(\Phi_{\lambda(t)}(x(t))-\min\Phi\right)$, as $t\to+\infty$, exist (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 (3.19) we obtain that $\lim_{t\to+\infty}(\alpha-1)r(t)+t\dot{r}(t)$ also exists. Applying Lemma 1.2.13 we deduce the existence of the limit $\lim_{t\to+\infty} r(t)$. Using (3.19) 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 (3.20) and the rest of the convergence rates. The key idea is to show that the limit

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

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

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

Since the limits

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

exist, it follows that

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

exists as well. Denote

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

and consider

$$0 \ \leq \ \frac{\psi(t)}{t} \ \leq \ 2tb(t)\left(\Phi_{\lambda(t)}(x(t))-\min\Phi\right)+\frac{t\varepsilon(t)}{2}\|x(t)\|^2+\frac{t}{2}\|\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\|^2. \tag{3.23}$$

Let us show that the right-hand side of (3.23) is integrable. Indeed, the first term is integrable by Theorem 3.2.1. As we have also established in Theorem 3.2.1, starting from t^{**}

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

where $a \geq 1$. Then, by (3.13) and $\dot{\lambda}(t) \geq 0$ for every $t \geq t_0$, we deduce that there exists $t_1 \geq t^{**}$ such that for every $t \geq t_1$

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

or

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

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 3.2.1 the right-hand side of (3.23) belongs to $L^1\big([t_1, +\infty), \mathbb{R}\big)$. Therefore, $t \mapsto \frac{\psi(t)}{t}$ 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 (3.20). To complete the proof notice that by the definition of the proximal mapping, we have

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

The conclusion follows immediately from (1.2) and (3.20).

□

3.3 Strong convergence of the trajectories

In this section we will establish the strong convergence of the trajectories to the minimal norm element of $\operatorname{argmin}\Phi$.

In order to do so, we will need to modify assumption (3.12):

$$\boxed{\int_{t_0}^{+\infty}\frac{\varepsilon(t)}{tb(t)}dt \;<\; +\infty. \qquad (3.24)}$$

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

Theorem 3.3.1. *Suppose that $\alpha > 3$, the function λ is bounded for every $t \geq t_0$ and (3.10), (3.11), (3.13) and (3.24) hold. Then*

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

and

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

Proof. Let us return to (3.17):

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

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 $t \geq t_0$

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

Thus,

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

We deduce due to (3.24) and the Lemma 1.2.15 that

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

Therefore,

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

and clearly

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

Thus, we establish

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

By the definition of the proximal mapping

$$\Phi_{\lambda(t)}(x(t)) - \min\Phi = \;\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min\Phi + \frac{1}{2\lambda(t)}\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2$$

$\forall t \geq t_0$. Using the fact that λ is bounded for every $t \geq t_0$ we deduce

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

and

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

□

For the remaining part of this section we will use a different energy functional. Inspired by [30] we introduce the following functional, which we will heavily rely on throughout this section

$$\begin{aligned} E_{p,q}(t) = & t^{p+1}\left(tb(t)+\beta(\alpha-p-q-2)\right)\left(\Phi_{\lambda(t)}(x(t))-\min\Phi\right) \\ & + \frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2-\|x^*\|^2\right)+\frac{t^p}{2}\left\|v(t)\right\|^2, \end{aligned} \tag{3.25}$$

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

The proof of the following theorem draws inspiration from [20, 23, 30].

Theorem 3.3.2. *Suppose that λ is bounded for every $t\geq t_0$, $\alpha>3$, $b(t_0)\geq\frac{1}{2}+\frac{\beta}{t_0}$ and (3.10), (3.11) and (3.24) are fulfilled. Suppose additionally that for every $t\geq t_0$*

$$\left(\frac{\alpha}{3}-1\right)tb(t)-t^2\dot{b}(t)+\frac{\alpha\beta}{3}\ \geq\ 0, \tag{3.26}$$

$$2\alpha(\alpha-3)-9t^2\varepsilon(t)+6\alpha\beta\ \leq\ 0, \tag{3.27}$$

$$18\beta t+9\beta\dot{\lambda}(t)-9tb(t)\left(\dot{\lambda}(t)+2\beta\right)+3(\alpha+3)\beta^2+\alpha^2\beta\ \leq\ 0 \tag{3.28}$$

and

$$\lim_{t\to+\infty}\frac{\beta}{t^{\frac{\alpha}{3}+1}\varepsilon(t)}\int_{t_0}^{t}s^{\frac{\alpha}{3}+1}\varepsilon^2(s)ds\ =\ 0. \tag{3.29}$$

If $x:[t_0,+\infty)\mapsto H$ is a solution to (3.1) and the trajectory $x(t)$ stays either inside or outside the ball $B(0,\|x^*\|)$, then $x(t)$ converges to minimal norm solution $x^*=\mathrm{proj}_{\operatorname{argmin}\Phi}(0)$, as $t\to+\infty$. Otherwise, $\liminf_{t\to+\infty}\|x(t)-x^*\|=0$.

Proof. The proof of this theorem is inspired by the analogous one in [30], namely, we will consider several cases with respect to the trajectory x staying either inside or outside the ball $B\left(0,\|x^*\|\right)$.

Case I.

Assume that the trajectory x stays in the complement of the ball $B\left(0,\|x^*\|\right)$ for every $t\geq t_0$. 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:

$$\begin{aligned}
&\frac{d}{dt}E_{p,q}(t) \\
&= t^p \Big((p+2)tb(t) + t^2\dot{b}(t) + (p+1)\beta(\alpha - p - q - 2)\Big) \left(\Phi_{\lambda(t)}(x(t)) - \min \Phi\right) \\
&\quad + \ t^{p+1}\left(tb(t) + \beta(\alpha - p - q - 2)\right) \left(\left\langle \nabla \Phi_{\lambda(t)}(x(t)), \dot{x}(t) \right\rangle - \frac{\dot{\lambda}(t)}{2} \|\nabla \Phi_{\lambda(t)}(x(t))\|^2 \right) \\
&\quad + \ \frac{(p+2)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2} \left(\|x(t)\|^2 - \|x^*\|^2\right) + t^{p+2}\varepsilon(t)\langle \dot{x}(t), x(t)\rangle \\
&\quad + \ \frac{pt^{p-1}}{2} \|q(x(t) - x^*) + t\left(\dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t))\right)\|^2 + t^p \langle \dot{v}(t), v(t)\rangle.
\end{aligned} \tag{3.30}$$

Consider for every $t \geq t_0$ the inner product

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

where above we used (3.1). Consider now for every $t \geq t_0$,

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

The two estimates that we made above lead to (3.30) becoming

$$\begin{aligned}
&\frac{d}{dt}E_{p,q}(t) \\
&= t^p \Big((p+2)tb(t) + t^2\dot{b}(t) + (p+1)\beta(\alpha - p - q - 2)\Big)\left(\Phi_{\lambda(t)}(x(t)) - \min \Phi\right) \\
&\quad + \ \frac{(p+2)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) + \frac{pq^2t^{p-1}}{2}\|x(t) - x^*\|^2 \\
&\quad + \ \frac{(p+2)\beta^2 t^{p+1}}{2}\|\nabla \Phi_{\lambda(t)}(x(t))\|^2 + \left(q + 1 - \alpha + \frac{p}{2}\right)t^{p+1}\|\dot{x}(t)\|^2
\end{aligned}$$

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

Let us apply the gradient inequality (1.6) to the strongly convex function $x \mapsto b(t)\Phi_{\lambda(t)}(x) + \frac{\varepsilon(t)\|x\|^2}{2}$:

$$
\begin{aligned}
&-\Big\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t)-x^*\Big\rangle + \frac{\varepsilon(t)\|x(t)-x^*\|^2}{2} \\
\leq\ &\left(b(t)\min\Phi + \frac{\varepsilon(t)\|x^*\|^2}{2}\right) - \left(b(t)\Phi_{\lambda(t)}(x(t)) + \frac{\varepsilon(t)\|x(t)\|^2}{2}\right)
\end{aligned}
$$

and thus

$$
\begin{aligned}
&-qt^{p+1}\Big\langle b(t)\nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t), x(t)-x^*\Big\rangle \ \leq\ -qt^{p+1}b(t)\left(\Phi_{\lambda(t)}(x(t)) - \min\Phi\right) \\
&-qt^{p+1}\frac{\varepsilon(t)}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) - qt^{p+1}\frac{\varepsilon(t)\|x(t)-x^*\|^2}{2}
\end{aligned}
$$

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

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

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

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

$$\leq \frac{q\beta(p+1)t^{p+1}}{4c^2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 + q\beta(p+1)c^2t^{p-1}\|x(t)-x^*\|^2$$

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,

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

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 $\big\langle\dot{x}(t), x(t)-x^*\big\rangle$. Secondly,

$$q+1-\alpha+\frac{p}{2} \;=\; -\frac{p}{2} \;\leq\; 0. \tag{3.31}$$

Then

$$p+2-q \;=\; 1-\frac{\alpha}{3} \;\leq\; 0. \tag{3.32}$$

So, for every $t \geq t_0$

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

$$+ \left(\frac{(p+2)\beta^2 t^{p+1} - 2\beta t^{p+2} b(t) - \dot{\lambda}(t) t^{p+1} \left(tb(t) + \beta(\alpha - p - q - 2) \right)}{2} \right.$$
$$\left. + \frac{q\beta(p+1)t^{p+1}}{4c^2} + \frac{\beta t^{p+2}}{a} \right) \cdot \|\nabla \Phi_{\lambda(t)}(x(t))\|^2$$
$$+ \left(q + 1 - \alpha + \frac{p}{2} \right) t^{p+1} \|\dot{x}(t)\|^2 - \left(\frac{(p+2-q)t^{p+1}\varepsilon(t) + t^{p+2}\dot{\varepsilon}(t)}{2} \right) \|x^*\|^2.$$

Obviously, for t large enough, say, $t \geq t_2 \geq t_0$ the following expression is non-positive due to (3.26) 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 (3.27) 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 every $t \geq t_0$. Furthermore,

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

So, under the assumption (3.11) and the fact that $\|x(t)\| \geq \|x^*\|$ for every $t \geq t_0$ we deduce due to (3.32)

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

Thus, under the assumptions (3.11), (3.26), (3.27) and (3.28) (the latest leads to the non-positivity of the coefficient of $\|\nabla \Phi_\lambda(x)\|^2$) we conclude due to (3.31) that for every $t \geq t_2$

$$\frac{d}{dt} E_{p,q}(t) \;\leq\; \frac{a\beta t^{\frac{\alpha}{3}+1}\varepsilon^2(t)}{4} \|x^*\|^2. \tag{3.33}$$

(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

$$\begin{aligned}
&E_{p,q}(t) \\
&\geq t^{p+1}\left(tb(t) + \beta(\alpha - p - q - 2)\right)\left(\Phi_{\lambda(t)}(x(t)) - \min \Phi\right) + \frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) \\
&= t^{p+1}\left(tb(t) - \beta\right)\left(\Phi_{\lambda(t)}(x(t)) - \min \Phi\right) + \frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right) \\
&\geq \frac{t^{p+2}}{2}\left(\Phi_{\lambda(t)}(x(t)) - \min \Phi\right) + \frac{\varepsilon(t)t^{p+2}}{2}\left(\|x(t)\|^2 - \|x^*\|^2\right),
\end{aligned} \tag{3.34}$$

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 (1.6) 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 $x_{\varepsilon(t),\lambda(t)} = \operatorname{argmin} \varphi_{\varepsilon(t),\lambda(t)}(x)$

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

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

$$\begin{aligned}
\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \varphi_{\varepsilon(t),\lambda(t)}(x^*) &= \frac{1}{2}\left(\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \min \Phi\right) \\
&\quad + \frac{\varepsilon(t)}{2}\left(\|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right) \\
&\geq \frac{\varepsilon(t)}{2}\left(\|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right).
\end{aligned}$$

We may now add the last two inequalities to obtain

$$\varphi_{\varepsilon(t),\lambda(t)}(x) - \varphi_{\varepsilon(t),\lambda(t)}(x^*) \geq \frac{\varepsilon(t)}{2}\left(\|x - x_{\varepsilon(t),\lambda(t)}\|^2 + \|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right) \tag{3.35}$$

for every $x \in H$. Plugging (3.35) into (3.34) we conclude that for every $t \geq t_2$

$$E_{p,q}(t) \geq \frac{t^{p+2}\varepsilon(t)}{2}\left(\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 + \|x_{\varepsilon(t),\lambda(t)}\|^2 - \|x^*\|^2\right). \tag{3.36}$$

(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 (3.33) 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 (3.36) 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 (3.27)

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

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 (3.5) and (3.29) 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 3.3.1

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

and

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

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 $\{\operatorname{prox}_{\lambda(t_k)\Phi}(x(t_k))\}_{k\in\mathbb{N}}$ 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

$$\Phi(\hat{x}) \;\leq\; \liminf_{k\to\infty} \Phi(\operatorname{prox}_{\lambda(t_k)\Phi}(x(t_k))) \;=\; \lim_{t\to+\infty} \Phi(\operatorname{prox}_{\lambda(t)\Phi}(x(t))) \;=\; \min\Phi$$

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

$$\|\hat{x}\| \;\leq\; \liminf_{k\to\infty} \|\operatorname{prox}_{\lambda(t_k)\Phi}(x(t_k))\| \;\leq\; \|x^*\|,$$

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, namely

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

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 $\{t_k\}_{k\in\mathbb{N}} \subseteq [t_0, +\infty)$ such that $t_k \to \infty$ as $k \to \infty$ and $\|x(t_k)\| = \|x^*\|$ for every $k \in \mathbb{N}$. Consider again a weak sequential cluster point $\hat{x}$ of the sequence $\{x(t_k)\}_{k\in\mathbb{N}}$. By repeating the same argument as in the previous case we deduce the weak convergence of $\{x(t_k)\}_{k\in\mathbb{N}}$ to x^*, as $k \to \infty$. Since $\|x(t_k)\| \to \|x^*\|$, as $k \to \infty$, we obtain that $\|x(t_k) - x^*\| \to 0$, as $k \to \infty$, which means $\liminf_{t\to+\infty} \|x(t) - x^*\| = 0$.

□

Remark 3.3.3. In this section the condition $\dot{b}(t) \geq 0$ for every $t \geq t_0$ 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.

3.3.1 Strong convergence of the trajectories in case $\alpha = 3$.

Throughout this section we no longer require that b is non-decreasing. In this case the analogue of Theorem 3.3.1 looks as follows

Theorem 3.3.4. *Suppose that for every $t \geq t_0$ the function λ is bounded, $b(t) \equiv b > 0$ is a constant function and (3.11) and (3.13) hold. Suppose additionally that (3.24) holds for constant b, namely*

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

Then

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

and

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

Proof. In this case the energy functional becomes

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

Relation (3.15) thus becomes for every $t \geq t_0$

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

Thus, repeating the same arguments as in Theorem 3.2.1 we obtain

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

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

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

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

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

Combining these two inequalities, we deduce

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

Now,

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

Applying Lemma 1.2.15 we deduce due to (3.24)

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

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

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

Again, by the definition of the proximal mapping

$$\Phi_{\lambda(t)}(x(t)) - \min\Phi = \;\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min\Phi + \frac{1}{2\lambda(t)}\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2$$

$\forall t \geq t_0$. Using the fact that λ is bounded for every $t \geq t_0$ we deduce

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

and

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

□

We are in position now to formulate the analogue of Theorem 3.3.2.

Theorem 3.3.5. *Suppose that λ is bounded for every $t \geq t_0$, $b(t) \equiv b \geq \frac{1}{2} + \frac{\beta}{t_0}$ and (3.11) and (3.24) hold. Assume, in addition, that*

$$\lim_{t\to+\infty} t^2\varepsilon(t) \;=\; +\infty, \tag{3.37}$$

$$2\beta t + \beta\dot{\lambda}(t) - bt\left(\dot{\lambda}(t) + 2\beta\right) + 2\beta^2 + \beta \;\leq\; 0 \text{ for all } t \geq t_0 \tag{3.38}$$

and

$$\lim_{t\to+\infty} \frac{\beta}{t^2\varepsilon(t)} \int_{t_0}^t s^2\varepsilon^2(s)ds \;=\; 0. \tag{3.39}$$

If $x : [t_0, +\infty) \mapsto H$ is a solution to (3.1) and the trajectory $x(t)$ stays either inside or outside the ball $B(0, \|x^*\|)$, then $x(t)$ converges to minimal norm solution $x^* = \operatorname{proj}_{\operatorname{argmin}\Phi}(0)$, as $t \to +\infty$. Otherwise, $\liminf_{t\to+\infty} \|x(t) - x^*\| = 0$.

Proof. The proof goes in line with the one of Theorem 3.3.2 by taking $\alpha = 3$, $b(t) \equiv b > 0$, $q = 2$, $p = 0$ and referring to Theorem 3.3.4 instead of Theorem 3.3.1 in the second and third cases.

□

3.4 Analysis of the conditions

We would like to conduct a detailed analysis of the conditions, which we imposed on our parameters in order to obtain the desired results.

1. In order to obtain the fast convergence rates of the function values we require that for every $t \geq t_0$:
   - $\alpha \ > \ 3$;
   - 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 every $t \geq t_0$:
   - $\alpha \ > \ 3$;
   - 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 $\varepsilon, d > 0$.

3.4.1 Setting for the fast convergence rates of the function values

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

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,

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

3.4.2 Setting for the strong convergence of the trajectories

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

1. $\alpha \ > \ 3$;
2. λ is bounded;
3. $\frac{\alpha-3}{3}bt^{n+1} - bnt^{n+1} + \frac{\alpha\beta}{3} \ \geq \ 0$;
4. $(\alpha-3)bt^{n+1} - bnt^{n+1} + \beta(2-\alpha) \ \geq \ 0$;
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. $\int_{t_0}^{+\infty} \frac{\varepsilon}{bt^{n+d+1}} dt \ < \ +\infty$;
7. $2\alpha(\alpha-3) - \frac{9\varepsilon}{t^{d-2}} + 6\alpha\beta \ \leq \ 0$;
8. $18\beta t + 9\beta\dot{\lambda}(t) - 9bt^{n+1}\left(\dot{\lambda}(t) + 2\beta\right) + 3(\alpha+3)\beta^2 + \alpha^2\beta \ \leq \ 0$;
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.

3.4.3 The case $\alpha = 3$

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

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) ,

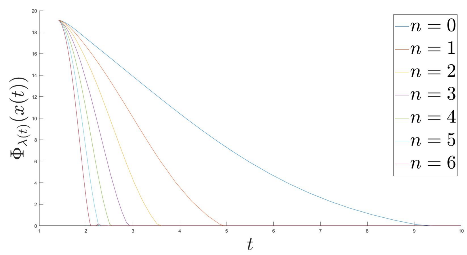
so with the appropriate choice of t_0 the whole set of conditions is fulfilled.

3.5 Numerical examples

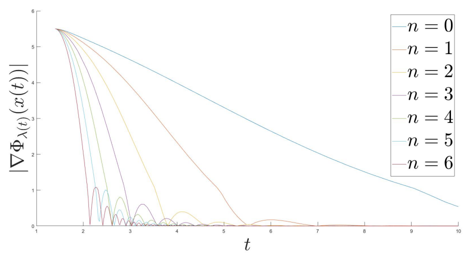
3.5.1 The rates of convergence of the Moreau envelope values

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 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 (3.1) (see Figure 3.1):



(a) Moreau envelope values

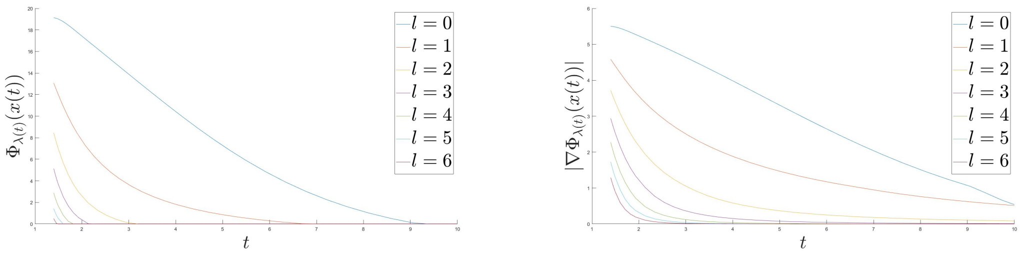


(b) Gradient

Figure 3.1: $l = 0$ and $d = 3$.

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 Figure 3.2):

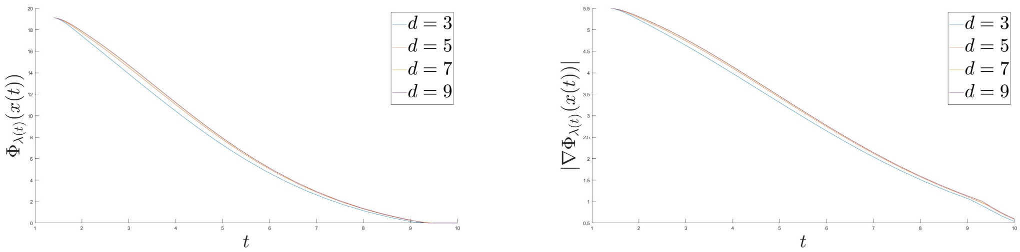


(a) Moreau envelope values (b) Gradient

Figure 3.2: $d = 3$ and $n = 0$.

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 Figure 3.3):



(a) Moreau envelope values (b) Gradient

Figure 3.3: $n = 0$ and $l = 0$.

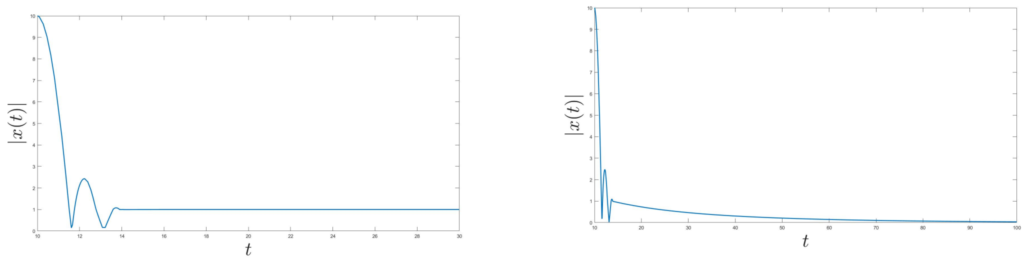
3.5.2 Strong convergence of the trajectories

For a different objective function let us investigate the strong convergence of the trajectories of (3.1):

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

The set $\operatorname{argmin} \Phi$ 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 every $t \geq t_0$)

and plot the behaviour of the trajectories of (3.1) with and without Tikhonov term (see Figure 3.4).

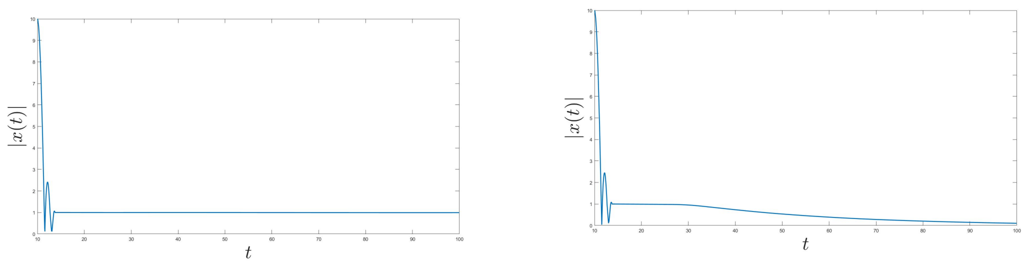


(a) $\varepsilon(t) = 0$ (b) $\varepsilon(t) = \frac{1}{t^{\frac{3}{2}}}$

Figure 3.4: The role of the Tikhonov term.

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 Figure 3.5).



(a) $\varepsilon(t) = 0$ (b) $\varepsilon(t) = \frac{1}{t^{\frac{3}{2}}}$

Figure 3.5: The role of the Tikhonov term.

Finally, for the same choice of λ let us take different Tikhonov terms to figure out how changing them affects the trajectories of (3.1) (see Figure 3.6):

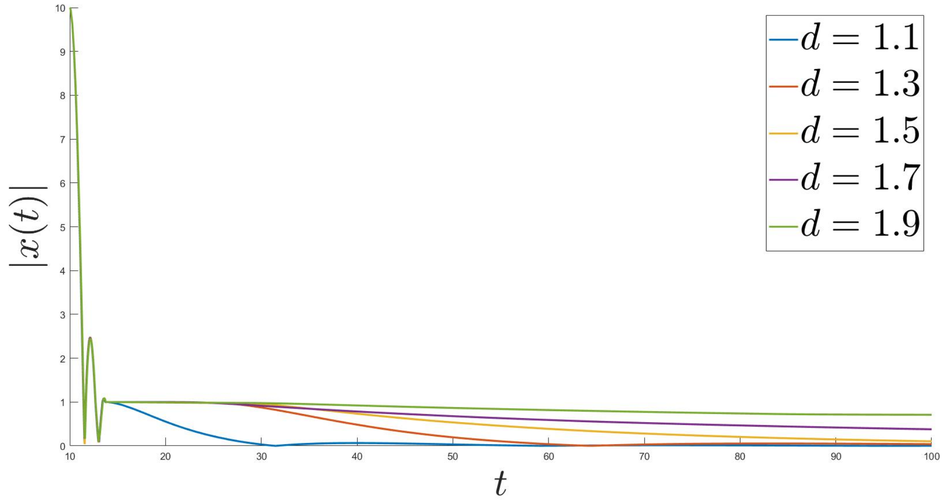


Figure 3.6: $n = 0.7$ and $\lambda(t) = 1 - \frac{1}{t}$.

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

4 Nonsmooth optimization with Tikhonov regularization featuring Tikhonov-dependent damping

In this chapter we would like to approach the minimization problem (1.1) via the following second order in time differential equation

$$\ddot{x}(t) + \alpha\sqrt{\varepsilon(t)}\dot{x}(t) + \beta\frac{d}{dt}\left(\nabla\Phi_{\lambda(t)}(x(t))\right) + \nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t) = 0 \text{ for } t \geq t_0, \quad (4.1)$$

and study its convergence properties, showing that alongside the trajectory – the solution of (4.1) – the function Φ converges to its minimum. The initial conditions are $x(t_0) = x_0 \in H$ and $\dot{x}(t_0) = x_1 \in H$ and α, β and $t_0 > 0$. The function $\lambda : [t_0, +\infty) \mapsto \mathbb{R}_+$ is assumed to be continuously differentiable and nondecreasing while the function $\varepsilon : [t_0, +\infty) \mapsto \mathbb{R}_+$ is continuously differentiable and nonincreasing with the property $\lim_{t\to+\infty} \varepsilon(t) = 0$.

The main goal of the research is to provide the setting in a nonsmooth case, where we would have fast convergence of the function values combined with strong convergence of the trajectories to the element of the minimal norm from the set of all minimizers of the objective function. This analysis is an extrapolation of the one conducted in [18] to the case of a nonsmooth objective function. We also aim to provide the exact rates of convergence of the values for the polynomial choice of the smoothing parameter λ and Tikhonov function ε. As a conclusion, multiple numerical experiments were conducted allowing better understanding of the theoretical results.

4.1 Preliminaries

Let us begin with the following estimates will be used later to evaluate the derivative of our energy function.

Lemma 4.1.1. *The following properties are satisfied:*

(i) for every $t \geq t_0$, $\frac{d}{dt}\left(\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) = \frac{1}{2}\left(\dot{\varepsilon}(t) - \dot{\lambda}(t)\varepsilon^2(t)\right)\|x_{\varepsilon(t),\lambda(t)}\|^2$;

(ii) the function $t \mapsto x_{\varepsilon(t),\lambda(t)}$ is Lipschitz continuous on the compact intervals of $(t_0, +\infty)$, thus, is almost everywhere differentiable. Moreover, for almost every $t \geq t_0$

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

Proof. By the definition of $\varphi_{\varepsilon(t),\lambda(t)}$

$$\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) = \inf_{y\in H}\left(\Phi_{\lambda(t)}(y) + \frac{\varepsilon(t)}{2}\|y-0\|^2\right) = \left(\Phi_{\lambda(t)}\right)_{\frac{1}{\varepsilon(t)}}(0) = \Phi_{\lambda(t)+\frac{1}{\varepsilon(t)}}(0)$$

by (i) of Lemma 1.2.8. Thus,

$$\frac{d}{dt}\left(\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) \;=\; \frac{d}{dt}\left(\Phi_{\lambda(t)+\frac{1}{\varepsilon(t)}}(0)\right) = \frac{1}{2}\left(\frac{\dot\varepsilon(t)}{\varepsilon^2(t)} - \dot\lambda(t)\right)\|\nabla\Phi_{\lambda(t)+\frac{1}{\varepsilon(t)}}(0)\|^2,$$

by (1.4). From (3.2) we obtain

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

where the second equality comes from (ii) of Lemma 1.2.8. Combining the last two equalities with (1.3) we obtain

$$\begin{aligned}
\frac{d}{dt}\left(\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) &= \frac{1}{2}\left(\frac{\dot\varepsilon(t)}{\varepsilon^2(t)} - \dot\lambda(t)\right)\left\|\frac{0-\operatorname{prox}_{\left(\lambda(t)+\frac{1}{\varepsilon(t)}\right)\Phi}(0)}{\lambda(t)+\frac{1}{\varepsilon(t)}}\right\|^2 \\
&= \frac{1}{2}\left(\frac{\dot\varepsilon(t)}{\varepsilon^2(t)} - \dot\lambda(t)\right)\left\|\frac{-\left(\lambda(t)\varepsilon(t)+1\right)x_{\varepsilon(t),\lambda(t)}}{\lambda(t)+\frac{1}{\varepsilon(t)}}\right\|^2 \\
&= \frac{1}{2}\left(\dot\varepsilon(t) - \dot\lambda(t)\varepsilon^2(t)\right)\|x_{\varepsilon(t),\lambda(t)}\|^2,
\end{aligned}$$

which is the first claim.

To obtain the second claim we start with (3.2) noticing that for $h>0$

$$\nabla\Phi_{\lambda(t)}\left(x_{\varepsilon(t),\lambda(t)}\right) = -\varepsilon(t)x_{\varepsilon(t),\lambda(t)} \text{ and } \nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t+h),\lambda(t+h)}) = -\varepsilon(t+h)x_{\varepsilon(t+h),\lambda(t+h)}.$$

Consider

$$\begin{aligned}
&\nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t+h),\lambda(t+h)}) - \nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) \\
&= \nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t+h),\lambda(t+h)}) - \nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t),\lambda(t)}) \\
&\qquad + \nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t),\lambda(t)}) - \nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}).
\end{aligned}$$

Taking the inner product of each part of this equality with $x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}$, we notice that

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

by the monotonicity of $\nabla\Phi_{\lambda(t+h)}$. So,

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

$$\begin{aligned}
&= -\varepsilon(t)\|x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}\|^2 \\
&\quad + (\varepsilon(t) - \varepsilon(t+h))\,\langle x_{\varepsilon(t+h),\lambda(t+h)}, x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}\rangle \\
&\geq \langle \nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t),\lambda(t)}) - \nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)}), x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}\rangle.
\end{aligned}$$

Let us divide the last inequality by h^2 to obtain

$$\begin{aligned}
&\frac{\varepsilon(t) - \varepsilon(t+h)}{h}\left\langle x_{\varepsilon(t+h),\lambda(t+h)}, \frac{x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}}{h}\right\rangle \\
&\geq \varepsilon(t)\left\|\frac{x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}}{h}\right\|^2 \\
&\quad + \left\langle \frac{\nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t),\lambda(t)}) - \nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)})}{h}, \frac{x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}}{h}\right\rangle.
\end{aligned}$$

Now notice that, since the mapping $t \mapsto x_{\varepsilon(t),\lambda(t)}$ is Lipschitz continuous on the compact intervals of $\mathbb{R}_+ \setminus \{0\}$ (according to [6]), therefore, almost everywhere differentiable. Tending h to zero we deduce for almost every $t \geq t_0$

$$\begin{aligned}
&-\dot{\varepsilon}(t)\left\langle x_{\varepsilon(t),\lambda(t)}, \frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\rangle \\
&\geq \varepsilon(t)\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|^2 - \frac{2\dot{\lambda}(t)\|\nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\|\,\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|}{\lambda(t)},
\end{aligned}$$

where we used the following estimate from [33]

$$\begin{aligned}
&\lim_{h\to 0}\left\langle \frac{\nabla\Phi_{\lambda(t+h)}(x_{\varepsilon(t),\lambda(t)}) - \nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)})}{h}, \frac{x_{\varepsilon(t+h),\lambda(t+h)} - x_{\varepsilon(t),\lambda(t)}}{h}\right\rangle \\
&\geq -\frac{2\dot{\lambda}(t)\|\nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\|\,\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|}{\lambda(t)}.
\end{aligned}$$

On the other hand, Cauchy-Schwartz inequality yields

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

Combining the last two inequalities we arrive at

$$-\dot{\varepsilon}(t)\|x_{\varepsilon(t),\lambda(t)}\| + \frac{2\dot{\lambda}(t)}{\lambda(t)}\|\nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\| \geq \varepsilon(t)\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|.$$

Replacing $\nabla\Phi_{\lambda(t)}(x_{\varepsilon(t),\lambda(t)})$ using (3.2) gives us the second claim.

□

Our nearest goal is to deduce the existence and uniqueness of the solution of the dynamical system (4.1). Suppose $\beta > 0$. Let us integrate (4.1) from t_0 to t to obtain

$$\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t)) + \int_{t_0}^{t}\left(\alpha\sqrt{\varepsilon(s)}\dot{x}(s) + \nabla\Phi_{\lambda(s)}(x(s)) + \varepsilon(s)x(s)\right)ds - \dot{x}(t_0)$$

$$- \quad \beta \nabla \Phi_{\lambda(t_0)}(x(t_0)) \quad = \quad 0.$$

Denoting

$$z(t) := \int_{t_0}^{t} \left(\alpha \sqrt{\varepsilon(s)} \dot{x}(s) + \nabla \Phi_{\lambda(s)}(x(s)) + \varepsilon(s) x(s) \right) ds - \left(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x_0) \right)$$

for every $t \geq t_0$ and noticing that $\dot{z}(t) = \alpha \sqrt{\varepsilon(t)} \dot{x}(t) + \nabla \Phi_{\lambda(t)}(x(t)) + \varepsilon(t) x(t)$ we deduce, that (4.1) is equivalent to

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

Let us multiply the second one by β and then by summing it with the first line we get rid of the gradient of the Moreau envelope in the second equation

$$\begin{cases} \dot{x}(t) + \beta \nabla \Phi_{\lambda(t)}(x(t)) + z(t) = 0, \\ \beta \dot{z}(t) + \left(1 - \alpha \beta \sqrt{\varepsilon(t)} \right) \dot{x}(t) - \beta \varepsilon(t) x(t) + z(t) = 0, \\ x(t_0) = x_0, \ z(t_0) = -\left(\dot{x}(t_0) + \beta \nabla \Phi_{\lambda(t_0)}(x_0) \right). \end{cases}$$

We denote now $y(t) = \beta z(t) + \left(1 - \alpha \beta \sqrt{\varepsilon(t)} \right) x(t)$, and, after simplification, we obtain the following equivalent formulation for the dynamical system

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

In case $\beta = 0$ for every $t \geq t_0$, (4.1) can be equivalently written as

$$\begin{cases} \dot{x}(t) - y(t) = 0, \\ \dot{y}(t) + \alpha \sqrt{\varepsilon(t)} y(t) + \nabla \Phi_{\lambda(t)}(x(t)) + \varepsilon(t) x(t) = 0, \\ x(t_0) = x_0, \ y(t_0) = \dot{x}(t_0). \end{cases}$$

Therefore, based on the two reformulations of the dynamical system (4.1) above we provide the following existence and uniqueness result, which is a consequence of the Cauchy-Lipschitz theorem for strong global solutions. The proof follows the lines of the proofs of Theorem 1 in [12] or of Theorem 1.1 in [15] with some small adjustments.

Theorem 4.1.2. *Suppose that there exists $\lambda_0 > 0$ such that $\lambda(t) \geq \lambda_0$ for every $t \geq t_0$. Then for every $(x_0, \dot{x}(t_0)) \in H \times H$ there exists a unique strong global solution $x : [t_0, +\infty) \mapsto H$ of the continuous dynamics (4.1) which satisfies the Cauchy initial conditions $x(t_0) = x_0$ and $\dot{x}(t_0) = \dot{x}_0$.*

Remark 4.1.3. We would like to stress that Theorem 11 of [34] does not cover the case presented in this chapter.

1. First of all, the systems (4.1) and (1.19) have different damping coefficients. The damping in (4.1) depends on the Tikhonov function ε, while the damping in (1.19) is taken in a polynomial form $\frac{1}{t^q}$. Thus, if we take $\varepsilon(t) = \frac{1}{t^{2q}}$ in (1.19) to mimic the relation between the damping parameter and the Tikhonov function as in (4.1), then one of the conditions of Theorem 11 becomes
$$\int_{t_0}^{+\infty} t^{3q}\varepsilon^2(t)dt \;=\; \int_{t_0}^{+\infty} \frac{1}{t^q}dt \;<\; +\infty,$$
where $0 < q < 1$, which is obviously not fulfilled.

2. Secondly, the smoothing parameter λ in [34] is fixed, while our analysis holds for more general choice of λ. However, if we want to consider the polynomial case of parameters (Section 4), then we indeed arrive at a similar restriction for λ: in Section 4 we will discover that for strong convergence of the trajectories and polynomial choice of parameters, $\lambda(t) = t^l$, we have to take $0 \le l < 2$, which is a wider range than $0 < q < 1$ for $\lambda(t) = t^{2q}$.

3. Finally, the sets of conditions $(C_0) - (C_4)$ and our assumptions (4.3) – (4.6) lead to different settings in case of polynomial choice of the function $\varepsilon(t) = \frac{1}{t^d}$, $d > 0$ (Section 4). Namely, according to Corollary 9 of [34] the setting to satisfy all the conditions in the analysis in this case for $\beta > 0$ is $\max\left\{1-q, \frac{3q+1}{2}\right\} < d \le 1+q$ with $0 < q < 1$, whereas as we will see later (Section 4) our set of conditions allows $1 \le d \le 2$, which is more flexible in terms of the upper bound while being almost the same in terms of the lower bound. Thus, $d = 2$ is not an option in [34], but the lower limitation for the choice of d could be wider depending on the choice of q. Since the rates of convergence are better for the bigger values of d (Theorem 4.3.1), this additional flexibility of the upper bound justifies, in our opinion, the restrictions for the lower one.

4.2 Abstract convergence results of the function values and strong convergence of the trajectories

This section is devoted to establishing some crucial estimates for the following quantities $\Phi_{\lambda(t)}(x(t)) - \min \Phi$ and $\|x(t) - x_{\varepsilon(t),\lambda(t)}\|$ for every $t \ge t_0$. In order to do so we will use the ideas and methods of Lyapunov analysis. We introduce the energy function

$$\begin{aligned} E(t) &= \varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) \\ &+ \frac{1}{2}\left\|\gamma\sqrt{\varepsilon(t)}\left(x(t) - x_{\varepsilon(t),\lambda(t)}\right) + \dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\right\|^2, \end{aligned} \tag{4.2}$$

where $\frac{\alpha}{2} \le \gamma < \alpha$. The next theorem provides the main result of this section.

Theorem 4.2.1. *Let $x : [t_0, +\infty) \mapsto H$ be a solution of (4.1). Then for every $t \geq t_0$*

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \ \leq \ E(t) + \frac{\varepsilon(t)}{2}\|x^*\|^2,$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \ \leq \ E(t) + \frac{\varepsilon(t)}{2}\|x^*\|^2,$$

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \ \leq \ 2\lambda(t)E(t) + \lambda(t)\varepsilon(t)\|x^*\|^2$$

and

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

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

Proof. Consider for every $t \geq t_0$

$$\begin{aligned}
&\Phi_{\lambda(t)}(x(t)) - \min \Phi \\
&= \varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x^*) + \frac{\varepsilon(t)}{2}\left(\|x^*\|^2 - \|x(t)\|^2\right) \\
&= \varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) + \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \varphi_{\varepsilon(t),\lambda(t)}(x^*) \\
&\quad + \frac{\varepsilon(t)}{2}\left(\|x^*\|^2 - \|x(t)\|^2\right) \\
&\leq \varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) + \frac{\varepsilon(t)}{2}\left(\|x^*\|^2 - \|x(t)\|^2\right)
\end{aligned}$$

Using the definition of E we obtain for every $t \geq t_0$

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \ \leq \ E(t) + \frac{\varepsilon(t)}{2}\|x^*\|^2.$$

By the definition of the proximal mapping

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \ = \ \Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi + \frac{1}{2\lambda(t)}\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2$$

$\forall t \geq t_0$. Thus,

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \ \leq \ E(t) + \frac{\varepsilon(t)}{2}\|x^*\|^2$$

and

$$\frac{1}{2\lambda(t)}\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \ \leq \ E(t) + \frac{\varepsilon(t)}{2}\|x^*\|^2.$$

The second result immediately follows from the $\varepsilon(t)$-strong convexity of $\varphi_{\varepsilon(t),\lambda(t)}$:

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

and thus

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

Finally, by $\lim_{t\to+\infty}\varepsilon(t)=0$ and (3.5) we deduce the strong convergence of the trajectories to x^* as soon as $\lim_{t\to+\infty}\frac{E(t)}{\varepsilon(t)}=0$.

□

Theorem 4.2.1 provided some abstract estimates for the important quantities. In order to show that these estimates are actually meaningful, we will have to first estimate the energy functional E. The idea is to show that this energy function satisfies the following differential inequality, as it was done in [18],

$$\dot{E}(t)+\mu(t)E(t)+\frac{\beta}{2}\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2 \;\leq\; \frac{g(t)\|x^*\|^2}{2} \text{ for all } t\geq t_0,$$

where $\mu(t)=(\alpha-\gamma)\sqrt{\varepsilon(t)}-\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}$ and g are positive functions. The next theorem provides the analysis needed to obtain the desired inequality.

Theorem 4.2.2. *Let $x:[t_0,+\infty)\mapsto H$ be a solution of (4.1). Assume that (3.3) holds and suppose that there exist $a,c>0$ such that for t large enough it holds that*

$$\frac{d}{dt}\left(\frac{1}{\sqrt{\varepsilon(t)}}\right) \;\leq\; \min\left\{2\gamma-\alpha-\frac{\gamma\beta\dot{\varepsilon}(t)}{2\varepsilon(t)},\; \alpha-\gamma\frac{a+1}{a}\right\} \tag{4.3}$$

$$\left(2\gamma(\alpha-\gamma)+\frac{\gamma}{c}-1\right)\varepsilon(t)-\beta\dot{\varepsilon}(t) \;\leq\; 0, \tag{4.4}$$

$$2\beta\varepsilon^2(t)+\left(2-\gamma\beta\sqrt{\varepsilon(t)}\right)\dot{\varepsilon}(t) \;\leq\; 0 \tag{4.5}$$

and

$$\left(\frac{\gamma}{a}+2(\alpha-\gamma)\right)\beta^2\sqrt{\varepsilon(t)}-\frac{3\beta^2\dot{\varepsilon}(t)}{2\varepsilon(t)}-\dot{\lambda}(t) \;\leq\; \beta. \tag{4.6}$$

Then there exists $t_1\geq t_0$ such that for every $t\geq t_1$

$$\beta\int_{t_1}^{t}\|\nabla\varphi_{\varepsilon(s)}(x(s))\|^2ds \;\leq\; 2E(t_1)+\|x^*\|^2\int_{t_1}^{t}g(s)ds$$

and

$$E(t) \;\leq\; \frac{\|x^*\|^2}{2\Gamma(t)}\int_{t_1}^{t}\Gamma(s)g(s)ds+\frac{\Gamma(t_1)E(t_1)}{\Gamma(t)},$$

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

Proof. We start with computing the derivative of the energy function (4.2). Let us denote $v(t)=\gamma\sqrt{\varepsilon(t)}\left(x(t)-x_{\varepsilon(t),\lambda(t)}\right)+\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))$. Once again, by the classical derivation chain rule using (i) from Lemma 4.1.1 and (1.4) we obtain for every $t\geq t_0$

$$\dot{E}(t) \;=\langle\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t)),\dot{x}(t)\rangle+\frac{\dot{\varepsilon}(t)}{2}\|x(t)\|^2+\frac{1}{2}\left(\dot{\lambda}(t)\varepsilon^2(t)-\dot{\varepsilon}(t)\right)\|x_{\varepsilon(t),\lambda(t)}\|^2$$

$$+ \ \langle \dot{v}(t), v(t) \rangle - \frac{\dot{\lambda}(t)}{2} \|\nabla \Phi_{\lambda(t)}(x(t))\|^2.$$

Our nearest goal is to obtain the upper bound for $\dot{E}$. Let us calculate for every $t \geq t_0$

$$\begin{aligned}
\dot{v}(t) &= \frac{\gamma \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} \left(x(t) - x_{\varepsilon(t),\lambda(t)} \right) + \gamma \sqrt{\varepsilon(t)} \dot{x}(t) - \gamma \sqrt{\varepsilon(t)} \frac{d}{dt} x_{\varepsilon(t),\lambda(t)} + \ddot{x}(t) \\
&+ \ \beta \frac{d}{dt} \left(\nabla \Phi_{\lambda(t)}(x(t)) \right) \\
&= \frac{\gamma \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} \left(x(t) - x_{\varepsilon(t),\lambda(t)} \right) + (\gamma - \alpha) \sqrt{\varepsilon(t)} \dot{x}(t) - \gamma \sqrt{\varepsilon(t)} \frac{d}{dt} x_{\varepsilon(t),\lambda(t)} \\
&- \ \nabla \Phi_{\lambda(t)}(x(t)) - \varepsilon(t) x(t) \\
&= \frac{\gamma \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} \left(x(t) - x_{\varepsilon(t),\lambda(t)} \right) + (\gamma - \alpha) \sqrt{\varepsilon(t)} \dot{x}(t) - \gamma \sqrt{\varepsilon(t)} \frac{d}{dt} x_{\varepsilon(t),\lambda(t)} \\
&- \ \nabla \varphi_{\varepsilon(t),\lambda(t)}(x(t)),
\end{aligned}$$

where above we used (4.1). Thus, for every $t \geq t_0$

$$\begin{aligned}
&\langle \dot{v}(t), v(t) \rangle \\
= \ &\frac{\gamma^2 \dot{\varepsilon}(t)}{2} \|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 + \left(\frac{\gamma \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + \gamma(\gamma - \alpha)\varepsilon(t) \right) \langle x(t) - x_{\varepsilon(t),\lambda(t)}, \dot{x}(t) \rangle \\
- \ &\gamma^2 \varepsilon(t) \left\langle \frac{d}{dt} x_{\varepsilon(t),\lambda(t)}, x(t) - x_{\varepsilon(t),\lambda(t)} \right\rangle - \gamma \sqrt{\varepsilon(t)} \left\langle x(t) - x_{\varepsilon(t),\lambda(t)}, \nabla \varphi_{\varepsilon(t),\lambda(t)}(x(t)) \right\rangle \\
+ \ &(\gamma - \alpha) \sqrt{\varepsilon(t)} \|\dot{x}(t)\|^2 - \gamma \sqrt{\varepsilon(t)} \left\langle \frac{d}{dt} x_{\varepsilon(t),\lambda(t)}, \dot{x}(t) \right\rangle - \left\langle \nabla \varphi_{\varepsilon(t),\lambda(t)}(x(t)), \dot{x}(t) \right\rangle \\
+ \ &\frac{\gamma \beta \dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} \left\langle x(t) - x_{\varepsilon(t),\lambda(t)}, \nabla \Phi_{\lambda(t)}(x(t)) \right\rangle + \beta (\gamma - \alpha) \sqrt{\varepsilon(t)} \left\langle \nabla \Phi_{\lambda(t)}(x(t)), \dot{x}(t) \right\rangle \\
- \ &\gamma \beta \sqrt{\varepsilon(t)} \left\langle \frac{d}{dt} x_{\varepsilon(t),\lambda(t)}, \nabla \Phi_{\lambda(t)}(x(t)) \right\rangle - \beta \left\langle \nabla \varphi_{\varepsilon(t),\lambda(t)}(x(t)), \nabla \Phi_{\lambda(t)}(x(t)) \right\rangle.
\end{aligned}$$

Let us use the previous estimates to evaluate the quantity $\langle \dot{v}(t), v(t) \rangle$. Namely, by the $\varepsilon(t)$-strong convexity of $\varphi_{\varepsilon(t),\lambda(t)}$ for every $t \geq t_0$

$$\begin{aligned}
&\varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \varphi_{\varepsilon(t),\lambda(t)}(x(t)) \\
\geq \ &\left\langle \nabla \varphi_{\varepsilon(t),\lambda(t)}(x(t)), x_{\varepsilon(t),\lambda(t)} - x(t) \right\rangle + \frac{\varepsilon(t)}{2} \|x_{\varepsilon(t),\lambda(t)} - x(t)\|^2
\end{aligned}$$

and then for every $t \geq t_0$

$$\begin{aligned}
&- \gamma \sqrt{\varepsilon(t)} \left\langle x(t) - x_{\varepsilon(t),\lambda(t)}, \nabla \varphi_{\varepsilon(t),\lambda(t)}(x(t)) \right\rangle \\
\leq \ &- \gamma \sqrt{\varepsilon(t)} \left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) \right) - \frac{\gamma \varepsilon^{\frac{3}{2}}(t)}{2} \|x_{\varepsilon(t),\lambda(t)} - x(t)\|^2.
\end{aligned}$$

Again, by the $\varepsilon(t)$-strong convexity of $\varphi_{\varepsilon(t),\lambda(t)}$ since $\dot{\varepsilon}(t) \leq 0$ for every $t \geq t_0$

$$\begin{aligned}
&\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}}\left\langle x(t) - x_{\varepsilon(t),\lambda(t)}, \nabla\Phi_{\lambda(t)}(x(t)) + \varepsilon(t)x(t) - \varepsilon(t)x(t)\right\rangle \\
\leq\ &\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}}\Bigg(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) - \frac{\varepsilon(t)}{2}\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \\
&- \ \varepsilon(t)\left\langle x(t) - x_{\varepsilon(t),\lambda(t)}, x(t)\right\rangle\Bigg).
\end{aligned}$$

Furthermore,

$$-\varepsilon(t)\left\langle x(t) - x_{\varepsilon(t),\lambda(t)}, x(t)\right\rangle \ = \ -\frac{\varepsilon(t)}{2}\Big(\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 + \|x(t)\|^2 - \|x_{\varepsilon(t),\lambda(t)}\|^2\Big).$$

Notice that for $a > 0$

$$-\gamma\sqrt{\varepsilon(t)}\left\langle \frac{d}{dt}x_{\varepsilon(t),\lambda(t)}, \dot{x}(t)\right\rangle \ \leq \ \frac{\gamma\sqrt{\varepsilon(t)}}{2a}\|\dot{x}(t)\|^2 + \frac{a\gamma\sqrt{\varepsilon(t)}}{2}\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|^2$$

as well as

$$\begin{aligned}
&- \gamma\beta\sqrt{\varepsilon(t)}\left\langle \frac{d}{dt}x_{\varepsilon(t),\lambda(t)}, \nabla\Phi_{\lambda(t)}(x(t))\right\rangle \\
\leq\ &\frac{\gamma\beta^2\sqrt{\varepsilon(t)}}{2a}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 + \frac{a\gamma\sqrt{\varepsilon(t)}}{2}\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|^2.
\end{aligned}$$

In the same spirit for $b > 0$

$$\begin{aligned}
&- \gamma^2\varepsilon(t)\left\langle \frac{d}{dt}x_{\varepsilon(t),\lambda(t)}, x(t) - x_{\varepsilon(t),\lambda(t)}\right\rangle \\
\leq\ &\frac{b\gamma\sqrt{\varepsilon(t)}}{2}\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|^2 + \frac{\gamma^3\varepsilon^{\frac{3}{2}}(t)}{2b}\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&(\gamma - \alpha)\sqrt{\varepsilon(t)}\left(\|\dot{x}(t)\|^2 + \beta\left\langle\nabla\Phi_{\lambda(t)}(x(t)), \dot{x}(t)\right\rangle\right) \\
=\ &\frac{(\gamma - \alpha)\sqrt{\varepsilon(t)}}{2}\Big(\|\dot{x}(t)\|^2 + \|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 - \beta^2\|\nabla\Phi_{\lambda(t)}(x(t))\|^2\Big)
\end{aligned}$$

and

$$\begin{aligned}
&- \beta\left\langle\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t)), \nabla\Phi_{\lambda(t)}(x(t))\right\rangle \\
= &- \frac{\beta}{2}\Big(\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2 + \|\nabla\Phi_{\lambda(t)}(x(t))\|^2 - \|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \nabla\Phi_{\lambda(t)}(x(t))\|^2\Big) \\
= &- \frac{\beta}{2}\Big(\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2 + \|\nabla\Phi_{\lambda(t)}(x(t))\|^2 - \varepsilon^2(t)\|x(t)\|^2\Big).
\end{aligned} \tag{4.7}$$

Combining all the estimates above we arrive for every $t \geq t_0$ at

$$\begin{aligned}
\langle \dot{v}(t), v(t) \rangle \ \leq\ & \left(\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \gamma\sqrt{\varepsilon(t)} \right) \left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) \right) \\
+ & \left(\frac{\gamma\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + \gamma(\gamma-\alpha)\varepsilon(t) \right) \langle x(t) - x_{\varepsilon(t),\lambda(t)}, \dot{x}(t) \rangle \\
+ & \left(\frac{\gamma^2\dot{\varepsilon}(t)}{2} + \frac{\gamma^3\varepsilon^{\frac{3}{2}}(t)}{2b} - \frac{\gamma\varepsilon^{\frac{3}{2}}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2} \right) \|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \\
+ & \left(\frac{\gamma}{a} + \gamma - \alpha \right) \frac{\sqrt{\varepsilon(t)}}{2} \|\dot{x}(t)\|^2 + \frac{(\gamma-\alpha)\sqrt{\varepsilon(t)}}{2} \|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
+ & \frac{\gamma(2a+b)\sqrt{\varepsilon(t)}}{2} \left\| \frac{d}{dt} x_{\varepsilon(t),\lambda(t)} \right\|^2 \\
+ & \frac{1}{2} \left(\frac{\gamma\beta^2\sqrt{\varepsilon(t)}}{a} - \beta - \beta^2(\gamma-\alpha)\sqrt{\varepsilon(t)} \right) \|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
- & \left\langle \nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t)), \dot{x}(t) \right\rangle + \left(\frac{\beta\varepsilon^2(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{4} \right) \|x(t)\|^2 \\
- & \frac{\beta}{2} \|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2 + \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{4} \|x_{\varepsilon(t),\lambda(t)}\|^2.
\end{aligned}$$

Returning to the expression for $\dot{E}(t)$ we notice that the terms $\left\langle \nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t)), \dot{x}(t) \right\rangle$ cancel each other out:

$$\begin{aligned}
\dot{E}(t) \ \leq\ & \left(\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} - \gamma\sqrt{\varepsilon(t)} \right) \left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)}) \right) \\
+ & \left(\frac{\gamma\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + \gamma(\gamma-\alpha)\varepsilon(t) \right) \langle x(t) - x_{\varepsilon(t),\lambda(t)}, \dot{x}(t) \rangle \\
+ & \left(\frac{\gamma^2\dot{\varepsilon}(t)}{2} + \frac{\gamma^3\varepsilon^{\frac{3}{2}}(t)}{2b} - \frac{\gamma\varepsilon^{\frac{3}{2}}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2} \right) \|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \\
+ & \left(\frac{\gamma}{a} + \gamma - \alpha \right) \frac{\sqrt{\varepsilon(t)}}{2} \|\dot{x}(t)\|^2 + \frac{(\gamma-\alpha)\sqrt{\varepsilon(t)}}{2} \|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
+ & \frac{\gamma(2a+b)\sqrt{\varepsilon(t)}}{2} \left\| \frac{d}{dt} x_{\varepsilon(t),\lambda(t)} \right\|^2 \\
+ & \frac{1}{2} \left(\frac{\gamma\beta^2\sqrt{\varepsilon(t)}}{a} - \beta - \beta^2(\gamma-\alpha)\sqrt{\varepsilon(t)} - \dot{\lambda}(t) \right) \|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
+ & \frac{1}{2} \left(\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2} \right) \|x_{\varepsilon(t),\lambda(t)}\|^2
\end{aligned}$$

$$+ \left(\frac{\beta\varepsilon^2(t)+\dot{\varepsilon}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{4}\right)\|x(t)\|^2$$
$$- \frac{\beta}{2}\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2 \text{ for all } t \geq t_0.$$

Let us now consider

$$\begin{aligned}
\mu(t)E(t) &= \mu(t)\left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) \\
&+ \frac{\mu(t)}{2}\left\|\gamma\sqrt{\varepsilon(t)}\left(x(t)-x_{\varepsilon(t),\lambda(t)}\right) + \dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\right\|^2 \\
&= \mu(t)\left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) + \frac{\gamma^2\mu(t)\varepsilon(t)}{2}\|x(t)-x_{\varepsilon(t),\lambda(t)}\|^2 \\
&+ \frac{\mu(t)}{2}\|\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 + \gamma\mu(t)\sqrt{\varepsilon(t)}\left\langle x(t)-x_{\varepsilon(t),\lambda(t)}, \dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\right\rangle \\
&\leq \mu(t)\left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) + \gamma^2\mu(t)\varepsilon(t)\|x(t)-x_{\varepsilon(t),\lambda(t)}\|^2 \\
&+ \frac{\mu(t)}{2}\|\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 + \gamma\mu(t)\sqrt{\varepsilon(t)}\left\langle x(t)-x_{\varepsilon(t),\lambda(t)}, \dot{x}(t)\right\rangle \\
&+ \frac{\beta^2\mu(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2,
\end{aligned}$$

since

$$\begin{aligned}
&\gamma\beta\mu(t)\sqrt{\varepsilon(t)}\left\langle x(t)-x_{\varepsilon(t),\lambda(t)}, \nabla\Phi_{\lambda(t)}(x(t))\right\rangle \\
&\leq \gamma\beta\mu(t)\sqrt{\varepsilon(t)}\|x(t)-x_{\varepsilon(t),\lambda(t)}\|\|\nabla\Phi_{\lambda(t)}(x(t))\| \\
&\leq \frac{\gamma^2\mu(t)\varepsilon(t)}{2}\|x(t)-x_{\varepsilon(t),\lambda(t)}\|^2 + \frac{\beta^2\mu(t)}{2}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.
\end{aligned}$$

Therefore, using $\mu(t) = (\alpha-\gamma)\sqrt{\varepsilon(t)} - \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}$ (the terms with $\langle x(t)-x_{\varepsilon(t),\lambda(t)}, \dot{x}(t)\rangle$ disappear), we obtain for every $t \geq t_0$

$$\begin{aligned}
&\dot{E}(t) + \mu(t)E(t) \\
&\leq \left(\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + (\alpha-2\gamma)\sqrt{\varepsilon(t)} - \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}\right)\left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) \\
&+ \left(\gamma^2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t) + \frac{\gamma^3\varepsilon^{\frac{3}{2}}(t)}{2b} - \frac{\gamma\varepsilon^{\frac{3}{2}}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\right)\|x(t)-x_{\varepsilon(t),\lambda(t)}\|^2 \\
&+ \left(\frac{\gamma}{a}+\gamma-\alpha\right)\frac{\sqrt{\varepsilon(t)}}{2}\|\dot{x}(t)\|^2 - \frac{\dot{\varepsilon}(t)}{4\varepsilon(t)}\|\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
&+ \frac{\gamma(2a+b)\sqrt{\varepsilon(t)}}{2}\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|^2 \\
&+ \frac{1}{2}\left(\frac{\gamma\beta^2\sqrt{\varepsilon(t)}}{a} - \beta + 2\beta^2(\alpha-\gamma)\sqrt{\varepsilon(t)} - \frac{\beta^2\dot{\varepsilon}(t)}{2\varepsilon(t)} - \dot{\lambda}(t)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{2}\left(\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\right)\|x_{\varepsilon(t),\lambda(t)}\|^2 \\
&+ \left(\frac{\beta\varepsilon^2(t) + \dot{\varepsilon}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{4}\right)\|x(t)\|^2 - \frac{\beta}{2}\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2.
\end{aligned}
$$

Further we have ($\dot{\varepsilon}(t) \leq 0$ for every $t \geq t_0$)

$$
-\frac{\dot{\varepsilon}(t)}{4\varepsilon(t)}\|\dot{x}(t) + \beta\nabla\Phi_{\lambda(t)}(x(t))\|^2 \;\leq\; -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}\|\dot{x}(t)\|^2 - \frac{\beta^2\dot{\varepsilon}(t)}{2\varepsilon(t)}\|\nabla\Phi_{\lambda(t)}(x(t))\|^2.
$$

As we have established earlier by the second claim of Lemma 4.1.1 and (3.4)

$$
\left\|\frac{d}{dt}x_{\varepsilon(t),\lambda(t)}\right\|^2 \;\leq\; \left(\frac{2\dot{\lambda}(t)}{\lambda(t)} - \frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\right)^2\|x_{\varepsilon(t),\lambda(t)}\|^2 \;\leq\; \left(\frac{2\dot{\lambda}(t)}{\lambda(t)} - \frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\right)^2\|x^*\|^2,
$$

and since there exists $t_1 \geq t_0$ such that $\left(\sqrt{\varepsilon(t)} \to 0, \text{ as } t \to +\infty\right)$

$$
\dot{\lambda}(t)\varepsilon^2(t) + \left(\frac{\gamma\beta\sqrt{\varepsilon(t)}}{2} - 1\right)\dot{\varepsilon}(t) \;\geq\; 0 \text{ for all } t \geq t_1,
$$

we deduce for every $t \geq t_1$

$$
\begin{aligned}
&\frac{1}{2}\left(\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\right)\|x_{\varepsilon(t),\lambda(t)}\|^2 \\
\leq\; &\frac{1}{2}\left(\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\right)\|x^*\|^2.
\end{aligned}
$$

Choosing $b = c\gamma$ with $c > 0$ we obtain for every $t \geq t_1$

$$
\begin{aligned}
&\dot{E}(t) + \mu(t)E(t) \\
\leq\; &\left(\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} + (\alpha - 2\gamma)\sqrt{\varepsilon(t)} - \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}\right)\left(\varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\right) \\
&+ \left(\gamma^2(\alpha - \gamma)\varepsilon^{\frac{3}{2}}(t) + \frac{\gamma^2\varepsilon^{\frac{3}{2}}(t)}{2c} - \frac{\gamma\varepsilon^{\frac{3}{2}}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\right)\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \\
&+ \left(\frac{\gamma}{a} + \gamma - \alpha - \frac{\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)}\right)\frac{\sqrt{\varepsilon(t)}}{2}\|\dot{x}(t)\|^2 \\
&+ \frac{1}{2}\left(\frac{\gamma\beta^2\sqrt{\varepsilon(t)}}{a} - \beta + 2\beta^2(\alpha - \gamma)\sqrt{\varepsilon(t)} - \frac{3\beta^2\dot{\varepsilon}(t)}{2\varepsilon(t)} - \dot{\lambda}(t)\right)\|\nabla\Phi_{\lambda(t)}(x(t))\|^2 \\
&+ \left(\frac{\beta\varepsilon^2(t) + \dot{\varepsilon}(t)}{2} - \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{4}\right)\|x(t)\|^2
\end{aligned}
$$

$$+\left(\frac{1}{2}\left(\dot{\lambda}(t)\varepsilon^2(t)-\dot{\varepsilon}(t)+\frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\right)+\frac{\gamma(2a+c\gamma)\sqrt{\varepsilon(t)}}{2}\left(\frac{2\dot{\lambda}(t)}{\lambda(t)}-\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\right)^2\right)$$
$$\cdot\|x^*\|^2-\frac{\beta}{2}\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2.$$

Let us investigate the signs of the terms in the inequality above when t is large enough to satisfy what we assumed before (4.3) – (4.6). First of all,

$$(\alpha-2\gamma)\sqrt{\varepsilon(t)}-\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}+\frac{\gamma\beta\dot{\varepsilon}(t)}{2\sqrt{\varepsilon(t)}} = \left(\frac{d}{dt}\left(\frac{1}{\sqrt{\varepsilon(t)}}\right)+\alpha-2\gamma+\frac{\gamma\beta\dot{\varepsilon}(t)}{2\varepsilon(t)}\right)\sqrt{\varepsilon(t)} \leq 0$$

due to (4.3). Secondly,

$$\begin{aligned}
&\gamma^2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)+\frac{\gamma^2\varepsilon^{\frac{3}{2}}(t)}{2c}-\frac{\gamma\varepsilon^{\frac{3}{2}}(t)}{2}-\frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\\
=\ &\frac{\gamma\varepsilon^{\frac{3}{2}}(t)}{2}\left(2\gamma(\alpha-\gamma)+\frac{\gamma}{c}-1\right)-\frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}\\
=\ &\frac{\gamma\sqrt{\varepsilon(t)}}{2}\left(\left(2\gamma(\alpha-\gamma)+\frac{\gamma}{c}-1\right)\varepsilon(t)-\beta\dot{\varepsilon}(t)\right) \leq 0
\end{aligned}$$

due to (4.4). Next we have

$$\frac{\gamma}{a}+\gamma-\alpha-\frac{\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)} = \frac{d}{dt}\left(\frac{1}{\sqrt{\varepsilon(t)}}\right)+\gamma\frac{a+1}{a}-\alpha \leq 0$$

due to (4.3). Then,

$$\frac{\gamma\beta^2\sqrt{\varepsilon(t)}}{a}-\beta+2\beta^2(\alpha-\gamma)\sqrt{\varepsilon(t)}-\frac{3\beta^2\dot{\varepsilon}(t)}{2\varepsilon(t)}-\dot{\lambda}(t) \leq 0$$

due to (4.6). Finally,

$$\frac{\beta\varepsilon^2(t)+\dot{\varepsilon}(t)}{2}-\frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{4} \leq 0$$

due to (4.5), since

$$2\beta\varepsilon^2(t)+\left(2-\gamma\beta\sqrt{\varepsilon(t)}\right)\dot{\varepsilon}(t) \leq 0.$$

So, at the end we deduce for every $t\geq t_1$

$$\begin{aligned}
&\dot{E}(t)+\mu(t)E(t)\\
\leq\ &\frac{1}{2}\left(\dot{\lambda}(t)\varepsilon^2(t)-\dot{\varepsilon}(t)+\frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2}+\gamma(2a+c\gamma)\sqrt{\varepsilon(t)}\left(\frac{2\dot{\lambda}(t)}{\lambda(t)}-\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\right)^2\right)\|x^*\|^2\\
&-\frac{\beta}{2}\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2 = \frac{g(t)\|x^*\|^2}{2}-\frac{\beta}{2}\|\nabla\varphi_{\varepsilon(t),\lambda(t)}(x(t))\|^2.
\end{aligned} \tag{4.8}$$

Integrating (4.8) from t_1 to t we obtain

$$E(t) - E(t_1) + \int_{t_1}^{t} \mu(s)E(s)ds + \frac{\beta}{2}\int_{t_1}^{t} \|\nabla\varphi_{\varepsilon(s)}(x(s))\|^2 ds \ \leq \ \frac{\|x^*\|^2}{2}\int_{t_1}^{t} g(s)ds$$

or, neglecting the positive terms,

$$\frac{\beta}{2}\int_{t_1}^{t} \|\nabla\varphi_{\varepsilon(s)}(x(s))\|^2 ds \ \leq \ E(t_1) + \frac{\|x^*\|^2}{2}\int_{t_1}^{t} g(s)ds.$$

From (4.8) we also obtain for every $t \geq t_1$

$$\dot{E}(t) + \mu(t)E(t) \ \leq \ \frac{g(t)\|x^*\|^2}{2}.$$

Multiplying this with $\Gamma(t) = \exp\left(\int_{t_1}^{t} \mu(s)ds\right)$ and integrating again on $[t_1, t]$ we deduce

$$E(t) \ \leq \ \frac{\|x^*\|^2}{2\Gamma(t)}\int_{t_1}^{t} \Gamma(s)g(s)ds + \frac{\Gamma(t_1)E(t_1)}{\Gamma(t)}.$$

□

Now that we have an estimate for the energy function as well, we are able to derive, under which conditions the quantities of the right-hand side of the estimates of the Theorem 4.2.1 do converge to zero, as $t \to +\infty$. This result is given by the next theorem.

Theorem 4.2.3. *Let $x : [t_0, +\infty) \mapsto H$ be a solution of (4.1). Assume that (3.3) holds and suppose that there exist $a, c > 0$ such that for t large enough assumptions (4.3) - (4.6) hold. Suppose additionally that*

$$\begin{aligned}
\lim_{t\to+\infty} \sqrt{\varepsilon(t)}\exp\left((\alpha-\gamma)\int_{t_1}^{t}\sqrt{\varepsilon(s)}ds\right) &= +\infty,\\
\lim_{t\to+\infty} \frac{\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)} &= 0,\\
\lim_{t\to+\infty} \dot{\lambda}(t)\sqrt{\varepsilon(t)} &= 0 \ \textit{and}\\
\lim_{t\to+\infty} \frac{\dot{\lambda}(t)}{\lambda(t)\sqrt{\varepsilon(t)}} &= 0.
\end{aligned} \tag{4.9}$$

Then

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

Proof. Let us notice that we can simplify the proof a bit due to

$$\begin{aligned}
&\int_{t_1}^{t}\left(\dot{\lambda}(s)\varepsilon^2(s)-\dot{\varepsilon}(s)+\frac{\gamma\beta\dot{\varepsilon}(s)\sqrt{\varepsilon(s)}}{2}+\gamma(2a+c\gamma)\sqrt{\varepsilon(s)}\left(\frac{2\dot{\lambda}(s)}{\lambda(s)}-\frac{\dot{\varepsilon}(s)}{\varepsilon(s)}\right)^2\right)\Gamma(s)ds\\
\leq\ &\int_{t_1}^{t}\left(\dot{\lambda}(s)\varepsilon^2(s)-\dot{\varepsilon}(s)+\gamma(2a+c\gamma)\sqrt{\varepsilon(s)}\left(\frac{2\dot{\lambda}(s)}{\lambda(s)}-\frac{\dot{\varepsilon}(s)}{\varepsilon(s)}\right)^2\right)\Gamma(s)ds.
\end{aligned}$$

The proof of the theorem will be divided into several steps.

The asymptotic behaviour of the function Γ

Let us start with the function $\Gamma(t) = \exp\left(\int_{t_1}^{t}\mu(s)ds\right)$ for $\mu(t) = -\frac{\dot{\varepsilon}(t)}{2\varepsilon(t)} + (\alpha-\gamma)\sqrt{\varepsilon(t)}$.

$$\begin{aligned}
\Gamma(t) &= \exp\left(-\frac{1}{2}\int_{t_1}^{t}\frac{\dot{\varepsilon}(s)}{\varepsilon(s)}ds + (\alpha-\gamma)\int_{t_1}^{t}\sqrt{\varepsilon(s)}ds\right)\\
&= \exp\left(\frac{1}{2}\ln\frac{\varepsilon(t_1)}{\varepsilon(t)} + (\alpha-\gamma)\int_{t_1}^{t}\sqrt{\varepsilon(s)}ds\right)\\
&= \sqrt{\frac{\varepsilon(t_1)}{\varepsilon(t)}}\exp\left((\alpha-\gamma)\int_{t_1}^{t}\sqrt{\varepsilon(s)}ds\right).
\end{aligned}$$

Since $\varepsilon(t)$ is positive for every $t \geq t_1 \geq t_0$, the integral is nonnegative and the whole exponent is lower bounded by 1. Using the property of Tikhonov function, namely, $\lim_{t\to+\infty}\varepsilon(t) = 0$, we deduce that

$$\Gamma(t) \geq \sqrt{\frac{\varepsilon(t_1)}{\varepsilon(t)}} \to +\infty \text{ as } t \to +\infty.$$

The asymptotic behaviour of the function E

For now it is enough to assume the following, but later we will have to strengthen these assumptions:

$$\begin{aligned}
\lim_{t\to+\infty}\dot{\lambda}(t)\varepsilon^{\frac{3}{2}}(t) &= 0 \text{ and}\\
\lim_{t\to+\infty}\frac{\dot{\lambda}(t)}{\lambda(t)} &= 0.
\end{aligned} \tag{4.10}$$

Let us recall the form of the energy function

$$\begin{aligned}
E(t) &= \varphi_{\varepsilon(t),\lambda(t)}(x(t)) - \varphi_{\varepsilon(t),\lambda(t)}(x_{\varepsilon(t),\lambda(t)})\\
&+ \frac{1}{2}\left\|\gamma\sqrt{\varepsilon(t)}\left(x(t)-x_{\varepsilon(t),\lambda(t)}\right)+\dot{x}(t)+\beta\nabla\Phi_{\lambda(t)}(x(t))\right\|^2,
\end{aligned}$$

where $\frac{\alpha}{2} \leq \gamma < \alpha$. Let us study the behaviour of the function

$$\frac{\int_{t_1}^{t}\left[\left(\dot{\lambda}(s)\varepsilon^2(s)-\dot{\varepsilon}(s)+\frac{\left(2\dot{\lambda}(s)\varepsilon(s)-\lambda(s)\dot{\varepsilon}(s)\right)^2(2a+c\gamma)\gamma}{\lambda^2(s)\varepsilon^{\frac{3}{2}}(s)}\right)\exp\left(\int_{t_1}^{s}\mu(u)du\right)\right]ds}{\exp\left(\int_{t_1}^{t}\mu(u)du\right)}$$

$$= \frac{\int_{t_1}^{t}\tilde{g}(s)\Gamma(s)ds}{\Gamma(t)},$$

as $t \to +\infty$. Since $\tilde{g}(t)\Gamma(t) \geq 0$ for every $t \geq t_1$ so is the integral $\int_{t_1}^{t}\tilde{g}(s)\Gamma(s)ds$. If $0 \leq \int_{t_1}^{t}\tilde{g}(s)\Gamma(s)ds \leq +\infty$, then $E(t)$ goes to zero as $t \to +\infty$ due to the properties of h and Theorem 4.2.2. Otherwise, we may apply L'Hospital's rule to obtain

$$\lim_{t\to+\infty}\frac{\int_{t_1}^{t}\tilde{g}(s)\Gamma(s)ds}{\Gamma(t)} = \lim_{t\to+\infty}\frac{\tilde{g}(t)}{\mu(t)}$$

$$= \lim_{t\to+\infty}\left(\frac{2\dot{\lambda}(t)\varepsilon^3(t)-2\varepsilon(t)\dot{\varepsilon}(t)}{2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)-\dot{\varepsilon}(t)}+\frac{2\varepsilon(t)\left(2\dot{\lambda}(t)\varepsilon(t)-\lambda(t)\dot{\varepsilon}(t)\right)^2(a+c\gamma)\gamma}{\lambda^2(t)\varepsilon^{\frac{3}{2}}(t)\left(2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)-\dot{\varepsilon}(t)\right)}\right),$$

if the latest exists, which we are going to show now. Consider

$$\frac{2\dot{\lambda}(t)\varepsilon^3(t)-2\varepsilon(t)\dot{\varepsilon}(t)}{2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)-\dot{\varepsilon}(t)} = 2\frac{\dot{\lambda}(t)\varepsilon^{\frac{3}{2}}(t)-\frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}}}{2(\alpha-\gamma)-\frac{\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)}} \leq \frac{\dot{\lambda}(t)\varepsilon^{\frac{3}{2}}(t)-\frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}}}{\alpha-\gamma},$$

since $-\frac{\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)} \geq 0$. Notice that

$$\lim_{t\to+\infty}\frac{-\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} = 0 \text{ by (4.3), since } 0 \leq -\frac{\dot{\varepsilon}(t)}{\sqrt{\varepsilon(t)}} \leq 2\left(\alpha-\gamma\frac{a+1}{a}\right)\varepsilon(t)\to 0,$$

as $t \to +\infty$. So, by (4.10) we deduce

$$\lim_{t\to+\infty}\frac{2\dot{\lambda}(t)\varepsilon^3(t)-2\varepsilon(t)\dot{\varepsilon}(t)}{2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)-\dot{\varepsilon}(t)} = 0.$$

Consider now

$$\frac{2\left(2\dot{\lambda}(t)\varepsilon(t)-\lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{\lambda^2(t)\sqrt{\varepsilon(t)}\left(2(\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)-\dot{\varepsilon}(t)\right)} \leq \frac{2\left(2\dot{\lambda}(t)\varepsilon(t)-\lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{2(\alpha-\gamma)\lambda^2(t)\varepsilon^2(t)}$$

$$= \frac{(2a+c\gamma)\gamma}{\alpha-\gamma}\left(\frac{2\dot{\lambda}(t)}{\lambda(t)}-\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\right)^2.$$

Again, by (4.3) we know that

$$0 \leq -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)} \leq 2\left(\alpha-\gamma\frac{a+1}{a}\right)\sqrt{\varepsilon(t)}\to 0, \text{ as } t\to+\infty.$$

So, again using (4.10) we deduce that $\lim_{t\to+\infty}E(t) = 0$.

The asymptotic behaviour of the function $\frac{E}{\varepsilon}$

For this part we assume the full set of conditions (4.9). In the same spirit let us analyse the asymptotic behaviour of $\frac{E(t)}{\varepsilon(t)}$ as $t \to +\infty$. From Theorem 4.2.2 we know that

$$\frac{E(t)}{\varepsilon(t)} \leq \frac{\int_{t_1}^{t} \tilde{g}(s)\Gamma(s)ds}{\varepsilon(t)\Gamma(t)}\|x^*\| + \frac{E(t_1)}{\varepsilon(t)\Gamma(t)}.$$

By (4.9) we immediately deduce that $\lim_{t\to+\infty} \frac{E(t_1)}{\varepsilon(t)\Gamma(t)} = 0$. For the first term let us use the same technique as in the previous chapter and apply L'Hospital's rule to obtain

$$\begin{aligned}
\lim_{t\to+\infty} \frac{\int_{t_1}^{t} \tilde{g}(s)\Gamma(s)ds}{\varepsilon(t)\Gamma(t)} &= \lim_{t\to+\infty} \frac{\tilde{g}(t)\Gamma(t)}{\dot{\varepsilon}(t)\Gamma(t) + \mu(t)\varepsilon(t)\Gamma(t)} = \lim_{t\to+\infty} \frac{\tilde{g}(t)}{\dot{\varepsilon}(t) + \mu(t)\varepsilon(t)} \\
&= \lim_{t\to+\infty} \left(\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\left(2\dot{\lambda}(t)\varepsilon(t) - \lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{\lambda^2(t)\varepsilon^{\frac{3}{2}}(t)} \right) \frac{1}{\dot{\varepsilon}(t) + \mu(t)\varepsilon(t)} \\
&= \lim_{t\to+\infty} \left(\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\left(2\dot{\lambda}(t)\varepsilon(t) - \lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{\lambda^2(t)\varepsilon^{\frac{3}{2}}(t)} \right) \frac{1}{\frac{\dot{\varepsilon}(t)}{2} + (\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)} \\
&= \lim_{t\to+\infty} \left(\frac{\dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t)}{\frac{\dot{\varepsilon}(t)}{2} + (\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)} + \frac{\left(2\dot{\lambda}(t)\varepsilon(t) - \lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{\lambda^2(t)\varepsilon^{\frac{3}{2}}(t)\left(\frac{\dot{\varepsilon}(t)}{2} + (\alpha-\gamma)\varepsilon^{\frac{3}{2}}(t)\right)} \right) \\
&= \lim_{t\to+\infty} \frac{\dot{\lambda}(t)\lambda^2(t)\varepsilon^{\frac{7}{2}}(t) - \dot{\varepsilon}(t)\lambda^2(t)\varepsilon^{\frac{3}{2}}(t) + \left(2\dot{\lambda}(t)\varepsilon(t) - \lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{\frac{\lambda^2(t)\varepsilon^{\frac{3}{2}}(t)\dot{\varepsilon}(t)}{2} + (\alpha-\gamma)\lambda^2(t)\varepsilon^3(t)} \\
&= \lim_{t\to+\infty} \frac{\dot{\lambda}(t)\sqrt{\varepsilon(t)} - \frac{\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)} + \left(\frac{2\dot{\lambda}(t)\varepsilon(t) - \lambda(t)\dot{\varepsilon}(t)}{\lambda(t)\varepsilon^{\frac{3}{2}}(t)}\right)^2(2a+c\gamma)\gamma}{\frac{\dot{\varepsilon}(t)}{2\varepsilon^{\frac{3}{2}}(t)} + \alpha - \gamma} = 0 \text{ by (4.9).}
\end{aligned}$$

Thus, we have established that $\lim_{t\to+\infty} \frac{E(t)}{\varepsilon(t)} = 0$.

The asymptotic behaviour of the function λE

In this section we need to assume the full set of conditions (4.9) again. We will study the behaviour of $\lambda(t)E(t)$, as $t \to +\infty$. Again, from Theorem 4.2.2 we know that

$$\lambda(t)E(t) \leq \frac{\lambda(t)\int_{t_1}^{t} \tilde{g}(s)\Gamma(s)ds}{\Gamma(t)}\|x^*\| + \frac{E(t_1)\lambda(t)}{\Gamma(t)}.$$

We immediately obtain that $\frac{E(t_1)\lambda(t)}{\Gamma(t)} \to 0$ as $t \to +\infty$, since

$$\lim_{t\to+\infty} \frac{\exp\left((\alpha-\gamma)\int_{t_1}^{t}\sqrt{\varepsilon(s)}ds\right)}{\lambda(t)\sqrt{\varepsilon(t)}} = \lim_{t\to+\infty} \frac{\sqrt{\varepsilon(t)}\exp\left((\alpha-\gamma)\int_{t_1}^{t}\sqrt{\varepsilon(s)}ds\right)}{\lambda(t)\varepsilon(t)} = +\infty$$

by (3.3) and (4.9). Arguing in the same way we deduce for the first term

$$\lim_{t\to+\infty}\frac{\int_{t_1}^{t}\tilde{g}(s)\Gamma(s)ds}{\frac{\Gamma(t)}{\lambda(t)}}=\lim_{t\to+\infty}\frac{\tilde{g}(t)\Gamma(t)}{\frac{\mu(t)\Gamma(t)\lambda(t)-\Gamma(t)\dot{\lambda}(t)}{\lambda^2(t)}}=\lim_{t\to+\infty}\frac{\tilde{g}(t)\lambda^2(t)}{\mu(t)\lambda(t)-\dot{\lambda}(t)}.$$

Consider

$$\begin{aligned}\frac{\tilde{g}(t)\lambda^2(t)}{\mu(t)\lambda(t)-\dot{\lambda}(t)}&=\frac{\left(\dot{\lambda}(t)\varepsilon^2(t)-\dot{\varepsilon}(t)+\frac{\left(2\dot{\lambda}(t)\varepsilon(t)-\lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{\lambda^2(t)\varepsilon^{\frac{3}{2}}(t)}\right)\lambda^2(t)}{-\frac{\dot{\varepsilon}(t)\lambda(t)}{2\varepsilon(t)}+(\alpha-\gamma)\sqrt{\varepsilon(t)}\lambda(t)-\dot{\lambda}(t)}\\
&=2\frac{\dot{\lambda}(t)\varepsilon^3(t)\sqrt{\varepsilon(t)}\lambda^2(t)-\dot{\varepsilon}(t)\varepsilon(t)\sqrt{\varepsilon(t)}\lambda^2(t)+\left(2\dot{\lambda}(t)\varepsilon(t)-\lambda(t)\dot{\varepsilon}(t)\right)^2(2a+c\gamma)\gamma}{-\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}\lambda(t)+2\left(\alpha-\gamma\right)\varepsilon^2(t)\lambda(t)-2\dot{\lambda}(t)\varepsilon(t)\sqrt{\varepsilon(t)}}\\
&=2\frac{\dot{\lambda}(t)\varepsilon^{\frac{3}{2}}(t)\lambda(t)-\frac{\dot{\varepsilon}(t)\lambda(t)}{\sqrt{\varepsilon(t)}}+\left(\frac{2\dot{\lambda}(t)\varepsilon(t)-\lambda(t)\dot{\varepsilon}(t)}{\varepsilon(t)\sqrt{\lambda(t)}}\right)^2(2a+c\gamma)\gamma}{\frac{-\dot{\varepsilon}(t)}{\varepsilon^{\frac{3}{2}}(t)}+2\left(\alpha-\gamma\right)-\frac{2\dot{\lambda}(t)}{\lambda(t)\sqrt{\varepsilon(t)}}}\to 0 \text{ as } t\to+\infty\end{aligned}$$

by (3.3), (4.3) and (4.9), since

$$\begin{aligned}0&\leq\dot{\lambda}(t)\varepsilon^{\frac{3}{2}}(t)\lambda(t)=\dot{\lambda}(t)\sqrt{\varepsilon(t)}\varepsilon(t)\lambda(t)\to 0,\ \text{as } t\to+\infty,\\
0&\leq\frac{-\dot{\varepsilon}(t)\lambda(t)}{\sqrt{\varepsilon(t)}}\leq 2\left(\alpha-\gamma\frac{a+1}{a}\right)\lambda(t)\varepsilon(t)\to 0,\ \text{as } t\to+\infty,\\
0&\leq\frac{-\dot{\varepsilon}(t)\sqrt{\lambda(t)}}{\varepsilon(t)}\leq 2\left(\alpha-\gamma\frac{a+1}{a}\right)\sqrt{\varepsilon(t)\lambda(t)}\to 0,\ \text{as } t\to+\infty\end{aligned}$$

and

$$\frac{\dot{\lambda}(t)}{\sqrt{\lambda(t)}}\to 0,\ \text{as } t\to+\infty.$$

Thus, we have established that $\lim_{t\to+\infty}\frac{E(t)}{\lambda(t)}=0$.

□

As we can see, the results of Theorem 4.2.3 together with (3.3) and $\lim_{t\to+\infty}\varepsilon(t)=0$ guarantee the convergence to zero, as $t\to+\infty$, of the quantities in Theorem 4.2.1.

4.3 Polynomial choice of parameters

In this section we would like to specify the form of the functions λ and ε, namely, taking $\lambda(t)=t^l$ and $\varepsilon(t)=\frac{1}{t^d}$, $l\geq 0$ and $d>0$, and show that the main results still hold in this case. First of all, equation (4.1) becomes

$$\ddot{x}(t)+\frac{\alpha}{t^{\frac{d}{2}}}\dot{x}(t)+\beta\frac{d}{dt}\left(\nabla\Phi_{t^l}(x(t))\right)+\nabla\Phi_{t^l}(x(t))+\frac{1}{t^d}x(t)=0 \text{ for } t\geq t_0. \tag{4.11}$$

The second step would be to show that the main result of Theorem 4.2.2 is valid and obtain the precise rates of convergence for the function values and trajectories using Theorem 4.2.1. In order to do so let us formulate the next theorem.

Theorem 4.3.1. *Let $x : [t_0, +\infty) \mapsto H$ be a solution of (4.11). Assume that $1 \le d < 2$ and $0 \le l < d$. Then for t large enough*

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \ \le \ \frac{1}{t^d},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \ \le \ \frac{1}{t^d},$$

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \ \le \ \frac{1}{t^{\frac{d}{2}+1-l}}$$

and

$$\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \ \le \ \frac{1}{t^{1-\frac{d}{2}}}.$$

If $d = 2$ and $0 \le l < 2$ we deduce the following estimates for t large enough.

If $0 < \alpha < 2$, then

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \ \le \ \frac{1}{t^{\frac{\alpha}{2}+1}},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \ \le \ \frac{1}{t^{\frac{\alpha}{2}+1}}$$

and

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \ \le \ \frac{1}{t^{\frac{\alpha}{2}-l+1}}.$$

If $\alpha \ge 2$, then

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \ \le \ \frac{1}{t^2},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \ \le \ \frac{1}{t^2}$$

and

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \ \le \ \frac{1}{t^{2-l}}.$$

Proof. It is easy to check that the choice above of d and l satisfies conditions (3.3) and (4.3) – (4.6). Indeed, the set of assumptions for $\lambda(t) = t^l$ and $\varepsilon(t) = \frac{1}{t^d}$, l, $d > 0$ becomes

1. $\lim_{t\to+\infty} t^{l-d} \ = \ 0;$

   There exist $\frac{\alpha}{2} \le \gamma < \alpha$, $a > 0$ and $c > 0$ such that for all t large enough

2. $\frac{d}{2}t^{\frac{d}{2}-1} \ \le \ \min\left\{2\gamma - \alpha + \frac{\gamma\beta d}{2t},\ \alpha - \gamma\frac{a+1}{a}\right\};$

3. $\left(2\gamma(\alpha-\gamma) + \frac{\gamma}{c} - 1\right)\frac{1}{t^d} + \frac{d\beta}{t^{d+1}} \ \le \ 0;$

4. $\frac{2\beta}{t^{2d}} - d\left(2 - \frac{\gamma\beta}{t^{\frac{d}{2}}}\right)\frac{1}{t^{d+1}} \ \leq \ 0$

   and

5. $\left(\frac{\gamma}{a} + 2(\alpha-\gamma)\right)\beta^2\frac{1}{t^{\frac{d}{2}}} + \frac{3d\beta^2}{2t} - lt^{l-1} \ \leq \ \beta.$

The conditions above are, in turn, equivalent to

1. $l < d$;
2. $d \leq 2$;
3. $2\gamma(\alpha-\gamma) \ < \ 1$;
4. $d \geq 1$

   and

5. is always satisfied starting from t large enough.

Finally, we deduce for l and d

$$\begin{aligned}1 &\leq d \leq 2,\\ 0 &\leq l < d,\end{aligned}$$

which is our desired setting to satisfy all the conditions and to guarantee that the main results of this paper are valid.

Therefore, the results of Theorem 4.2.2 are valid in this case. Namely, in Theorem 4.2.2, we have obtained

$$E(t) \ \leq \ \frac{\|x^*\|^2}{2\Gamma(t)}\int_{t_1}^{t}\Gamma(s)g(s)ds + \frac{\Gamma(t_1)E(t_1)}{\Gamma(t)},$$

where $g(t) = \dot{\lambda}(t)\varepsilon^2(t) - \dot{\varepsilon}(t) + \frac{\gamma\beta\dot{\varepsilon}(t)\sqrt{\varepsilon(t)}}{2} + \gamma(2a+c\gamma)\sqrt{\varepsilon(t)}\left(\frac{2\dot{\lambda}(t)}{\lambda(t)} - \frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\right)^2$, $\Gamma(t) = \exp\left(\int_{t_1}^{t}\mu(s)ds\right)$ and $\mu(t) = (\alpha-\gamma)\sqrt{\varepsilon(t)} - \frac{\dot{\varepsilon}(t)}{2\varepsilon(t)}$. Now, our goal is to deduce the actual rates of convergence of the function values and trajectories. The proof will be divided into several sections for the convenience of the reader.

The functions μ and Γ

Let us consider the case when $1 \leq d < 2$. The case when $d = 2$ will be treated separately. The function μ thus writes as follows $\mu(t) = \frac{\alpha-\gamma}{t^{\frac{d}{2}}} + \frac{d}{2t}$. Then,

$$\Gamma(t) \ = \ \exp\left(\int_{t_1}^{t}\left[\frac{\alpha-\gamma}{s^{\frac{d}{2}}} + \frac{d}{2s}\right]ds\right) \ = \ \left(\frac{t}{t_1}\right)^{\frac{d}{2}}\exp\left(\int_{t_1}^{t}\frac{\alpha-\gamma}{s^{\frac{d}{2}}}ds\right)$$

$$= \left(\frac{t}{t_1}\right)^{\frac{d}{2}} \exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}}\left[t^{1-\frac{d}{2}} - t_1^{1-\frac{d}{2}}\right]\right) = Ct^{\frac{d}{2}} \exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}}\right),$$

where $C = \left(t_1^{\frac{d}{2}} \exp\left[\frac{\alpha-\gamma}{1-\frac{d}{2}} t_1^{1-\frac{d}{2}}\right]\right)^{-1}$. So, $\frac{\Gamma(t_1)E(t_1)}{\Gamma(t)}$ goes to zero exponentially, as time goes to infinity due to $1 \le d < 2$.

The function g

First notice that

$$\begin{aligned} g(t) &= \frac{lt^{l-1}}{t^{2d}} + \frac{d}{t^{d+1}} - \frac{\gamma\beta d}{2t^{d+1+\frac{d}{2}}} + \frac{\gamma(2a+c\gamma)}{t^{\frac{d}{2}}}\left(\frac{2l}{t} + \frac{d}{t}\right)^2 \\ &= lt^{l-1-2d} + \frac{d}{t^{d+1}} - \frac{\gamma\beta d}{2t^{d+1+\frac{d}{2}}} + \frac{\gamma(2a+c\gamma)(2l+d)^2}{t^{\frac{d}{2}+2}} \\ &= \frac{l}{t^{2d-l+1}} + \frac{d}{t^{d+1}} - \frac{\gamma\beta d}{2t^{d+1+\frac{d}{2}}} + \frac{C_1}{t^{\frac{d}{2}+2}}, \end{aligned}$$

where $C_1 = \gamma(2a+c\gamma)(2l+d)^2$. Then,

$$\Gamma(t)g(t) = C\left(\frac{l}{t^{\frac{3d}{2}+1-l}} + \frac{d}{t^{\frac{d}{2}+1}} - \frac{\gamma\beta d}{2t^{d+1}} + \frac{C_1}{t^2}\right)\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}}\right).$$

Let us notice that the behaviour of $\frac{l}{t^{\frac{3d}{2}+1-l}} + \frac{d}{t^{\frac{d}{2}+1}} - \frac{\gamma\beta d}{2t^{d+1}} + \frac{C_1}{t^2}$ is dictated by the term $\frac{1}{t^{\frac{d}{2}+1}}$, as $t \to +\infty$, since $1 \le d < 2$ and $0 \le l < d$.

Integrating the product Γg

The technique, which will be used in this section, is inspired by [18]. First of all, notice that for some $\delta > 0$

$$\frac{d}{dt}\left(\frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}}\right)}{\delta t}\right) = \left(-\frac{1}{\delta t^2} + \frac{\alpha-\gamma}{\delta t^{\frac{d}{2}+1}}\right)\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}}\right).$$

Secondly, there exists such δ that starting from some $t_2 \ge t_1$ it holds that

$$\frac{l}{t^{\frac{3d}{2}+1-l}} + \frac{d}{t^{\frac{d}{2}+1}} - \frac{\gamma\beta d}{2t^{d+1}} + \frac{C_1}{t^2} \le -\frac{1}{\delta t^2} + \frac{\alpha-\gamma}{\delta t^{\frac{d}{2}+1}}.$$

Thus,

$$C\int_{t_2}^{t}\left(\frac{l}{s^{\frac{3d}{2}+1-l}} + \frac{d}{s^{\frac{d}{2}+1}} - \frac{\gamma\beta d}{2s^{d+1}} + \frac{C_1}{s^2}\right)\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} s^{1-\frac{d}{2}}\right) ds$$

$$
\begin{aligned}
&\leq C \int_{t_2}^{t} \left(-\frac{1}{\delta s^2} + \frac{\alpha-\gamma}{\delta s^{\frac{d}{2}+1}} \right) \exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} s^{1-\frac{d}{2}} \right) ds \\
&= C \int_{t_2}^{t} \frac{d}{ds} \left(\frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} s^{1-\frac{d}{2}} \right)}{\delta s} \right) ds \;=\; C \left(\frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}} \right)}{\delta t} - \frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t_2^{1-\frac{d}{2}} \right)}{\delta t_2} \right) \\
&= C \frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}} \right)}{\delta t} - C_2,
\end{aligned}
$$

where $C_2 = C \frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t_2^{1-\frac{d}{2}} \right)}{\delta t_2}$.

Finalizing the estimates

Let us return to

$$
\begin{aligned}
E(t) \;&\leq\; \frac{\|x^*\|^2}{2\Gamma(t)} \int_{t_1}^{t} \Gamma(s) g(s) ds + \frac{\Gamma(t_1) E(t_1)}{\Gamma(t)} \\
&= \frac{\|x^*\|^2}{2\Gamma(t)} \int_{t_1}^{t_2} \Gamma(s) g(s) ds + \frac{\|x^*\|^2}{2\Gamma(t)} \int_{t_2}^{t} \Gamma(s) g(s) ds + \frac{\Gamma(t_1) E(t_1)}{\Gamma(t)} \\
&\leq \frac{\|x^*\|^2}{2} \left(\frac{1}{\Gamma(t)} \int_{t_1}^{t_2} \Gamma(s) g(s) ds - \frac{C_2}{\Gamma(t)} + C \frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}} \right)}{\delta t \Gamma(t)} \right) + \frac{\Gamma(t_1) E(t_1)}{\Gamma(t)}.
\end{aligned}
$$

This expression converges to zero at a speed of the slowest decaying term (all the other decay exponentially):

$$
C \frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}} \right)}{\delta t \Gamma(t)} \;=\; C \frac{\exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}} \right)}{\delta C t^{\frac{d}{2}+1} \exp\left(\frac{\alpha-\gamma}{1-\frac{d}{2}} t^{1-\frac{d}{2}} \right)} \;=\; \frac{1}{\delta t^{\frac{d}{2}+1}}.
$$

Thus, there exists a constant $C_3 > 0$ such that for every $t \geq t_2$

$$
E(t) \;\leq\; \frac{C_3}{t^{\frac{d}{2}+1}}.
$$

The rates themselves

Now we can deduce the actual rates for the quantities in Theorem 4.2.1. for every $t \geq t_2$

$$
\Phi_{\lambda(t)}(x(t)) - \min \Phi \;\leq\; \frac{C_3}{t^{\frac{d}{2}+1}} + \frac{1}{2t^d} \|x^*\|^2,
$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min\Phi \;\leq\; \frac{C_3}{t^{\frac{d}{2}+1}} + \frac{1}{2t^d}\|x^*\|^2,$$

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;\leq\; 2C_3 t^{l-\frac{d}{2}-1} + t^{l-d}\|x^*\|^2$$

and

$$\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \;\leq\; \frac{C_3}{t^{1-\frac{d}{2}}}.$$

Finally, there exist constants $C_4, C_5 > 0$ such that for every $t \geq t_2$

$$\Phi_{\lambda(t)}(x(t)) - \min\Phi \;\leq\; \frac{C_4}{t^d},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min\Phi \;\leq\; \frac{C_4}{t^d},$$

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;\leq\; \frac{C_5}{t^{\frac{d}{2}+1-l}}$$

and

$$\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \;\leq\; \frac{C_3}{t^{1-\frac{d}{2}}}.$$

The rates of convergence of the function values in case $d = 2$

This particular case is of a great interest, as it is in a way a bordering case, when one cannot show the strong convergence of the trajectories, but still can show the fast convergence of the values. In this case the functions μ and Γ are

$$\mu(t) \;=\; \frac{\alpha - \gamma + 1}{t}$$

and

$$\Gamma(t) \;=\; \exp\left(\int_{t_1}^{t} \frac{\alpha-\gamma+1}{s} ds\right) \;=\; \left(\frac{t}{t_1}\right)^{\alpha-\gamma+1} \;=\; Ct^{\alpha-\gamma+1},$$

where $C = \frac{1}{t_1^{\alpha-\gamma+1}}$. The function g is

$$g(t) \;=\; \frac{l}{t^{5-l}} + \frac{2}{t^3} - \frac{\gamma\beta}{t^4} + \frac{C_1}{t^3} \;=\; \frac{l}{t^{5-l}} + \frac{2+C_1}{t^3} - \frac{\gamma\beta}{t^4},$$

where $C_1 = 4\gamma(2a + c\gamma)(l+1)^2$. Thus,

$$\Gamma(t)g(t) \;=\; C\left(lt^{\alpha-\gamma+l-4} + (2+C_1)\,t^{\alpha-\gamma-2} - \gamma\beta t^{\alpha-\gamma-3}\right).$$

So,

$$C\int_{t_1}^{t}\left(ls^{\alpha-\gamma+l-4} + (2+C_1)\,s^{\alpha-\gamma-2} - \gamma\beta s^{\alpha-\gamma-3}\right) ds$$

$$
\begin{aligned}
&= C\left(\frac{ls^{\alpha-\gamma+l-3}}{\alpha-\gamma+l-3}+\frac{(2+C_1)\,s^{\alpha-\gamma-1}}{\alpha-\gamma-1}-\frac{\gamma\beta s^{\alpha-\gamma-2}}{\alpha-\gamma-2}\right)\Bigg|_{t_1}^{t} \\
&= C\left(\frac{lt^{\alpha-\gamma+l-3}}{\alpha-\gamma+l-3}+\frac{(2+C_1)\,t^{\alpha-\gamma-1}}{\alpha-\gamma-1}-\frac{\gamma\beta t^{\alpha-\gamma-2}}{\alpha-\gamma-2}\right)-C_2,
\end{aligned}
$$

where $C_2 = C\left(\frac{lt_1^{\alpha-\gamma+l-3}}{\alpha-\gamma+l-3}+\frac{(2+C_1)t_1^{\alpha-\gamma-1}}{\alpha-\gamma-1}-\frac{\gamma\beta t_1^{\alpha-\gamma-2}}{\alpha-\gamma-2}\right)$. By Theorem 5 we have

$$
\begin{aligned}
E(t) \;&\leq\; \frac{\|x^*\|^2}{2}\,\frac{C\left(\frac{lt^{\alpha-\gamma+l-3}}{\alpha-\gamma+l-3}+\frac{(2+C_1)t^{\alpha-\gamma-1}}{\alpha-\gamma-1}-\frac{\gamma\beta t^{\alpha-\gamma-2}}{\alpha-\gamma-2}\right)-C_2}{Ct^{\alpha-\gamma+1}}+\frac{Ct_1^{\alpha-\gamma+1}E(t_1)}{Ct^{\alpha-\gamma+1}} \\
&=\; \frac{\|x^*\|^2}{2}\,\frac{\frac{lt^{\alpha-\gamma+l-3}}{\alpha-\gamma+l-3}+\frac{(2+C_1)t^{\alpha-\gamma-1}}{\alpha-\gamma-1}-\frac{\gamma\beta t^{\alpha-\gamma-2}}{\alpha-\gamma-2}}{t^{\alpha-\gamma+1}}+\frac{C_3}{t^{\alpha-\gamma+1}} \\
&=\; \frac{\|x^*\|^2}{2}\left(\frac{lt^{l-4}}{\alpha-\gamma+l-3}+\frac{(2+C_1)\,t^{-2}}{\alpha-\gamma-1}-\frac{\gamma\beta t^{-3}}{\alpha-\gamma-2}\right)+\frac{C_3}{t^{\alpha-\gamma+1}}
\end{aligned}
$$

where $C_3 = \frac{2Ct_1^{\alpha-\gamma+1}E(t_1)-C_2\|x^*\|^2}{2C}$. We know that $\frac{\alpha}{2} \leq \gamma < \alpha$ and $0 \leq l < 2$. Thus, in the brackets the term with t^{-2} is dominating, as $t \to +\infty$. Moreover, $\alpha-\gamma+1 > 1$. So, the behaviour of the entire expression depends on the value of α. There exists a constant C_4 such that for every $t \geq t_1$

$$E(t) \;\leq\; \frac{C_4}{t^2}+\frac{C_3}{t^{\alpha-\gamma+1}}.$$

That leads us to the following rates for every $t \geq t_1$

$$
\begin{aligned}
\Phi_{\lambda(t)}(x(t)) - \min\Phi \;&\leq\; \frac{C_4}{t^2}+\frac{C_3}{t^{\alpha-\gamma+1}}+\frac{\|x^*\|^2}{2t^2}, \\
\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min\Phi \;&\leq\; \frac{C_4}{t^2}+\frac{C_3}{t^{\alpha-\gamma+1}}+\frac{\|x^*\|^2}{2t^2}, \\
\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;&\leq\; \frac{2C_4}{t^{2-l}}+\frac{2C_3}{t^{\alpha-\gamma-l+1}}+\frac{\|x^*\|^2}{t^{2-l}}
\end{aligned}
$$

and

$$\|x(t) - x_{\varepsilon(t),\lambda(t)}\|^2 \;\leq\; 2C_4+\frac{2C_3}{t^{\alpha-\gamma-1}}.$$

As we can see, the strong convergence of the trajectories can no longer be shown. Nevertheless, for $C_5 = \frac{2C_4+\|x^*\|^2}{2}$ we deduce for every $t \geq t_1$

$$
\begin{aligned}
\Phi_{\lambda(t)}(x(t)) - \min\Phi \;&\leq\; \frac{C_5}{t^2}+\frac{C_3}{t^{\alpha-\gamma+1}}, \\
\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min\Phi \;&\leq\; \frac{C_5}{t^2}+\frac{C_3}{t^{\alpha-\gamma+1}}
\end{aligned}
$$

and

$$\|\operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;\leq\; \frac{2C_5}{t^{2-l}}+\frac{2C_3}{t^{\alpha-\gamma-l+1}}.$$

Since we are free to choose γ such that $\frac{\alpha}{2} \leq \gamma < \alpha$, and since we want to have as fast rates as possible, we should take $\gamma = \frac{\alpha}{2}$.

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \;\leq\; \frac{C_5}{t^2} + \frac{C_3}{t^{\frac{\alpha}{2}+1}},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \;\leq\; \frac{C_5}{t^2} + \frac{C_3}{t^{\frac{\alpha}{2}+1}}$$

and

$$\| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;\leq\; \frac{2C_5}{t^{2-l}} + \frac{2C_3}{t^{\frac{\alpha}{2}-l+1}}.$$

Here we have to consider several cases.

1. If $0 < \alpha < 2$, then there exists C_6 such that for every $t \geq t_1$

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \;\leq\; \frac{C_6}{t^{\frac{\alpha}{2}+1}},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \;\leq\; \frac{C_6}{t^{\frac{\alpha}{2}+1}}$$

and

$$\| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;\leq\; \frac{2C_6}{t^{\frac{\alpha}{2}-l+1}}.$$

2. If $\alpha \geq 2$, then there exists C_6 such that for every $t \geq t_1$

$$\Phi_{\lambda(t)}(x(t)) - \min \Phi \;\leq\; \frac{C_6}{t^2},$$

$$\Phi\left(\operatorname{prox}_{\lambda(t)\Phi}(x(t))\right) - \min \Phi \;\leq\; \frac{C_6}{t^2}$$

and

$$\| \operatorname{prox}_{\lambda(t)\Phi}(x(t)) - x(t)\|^2 \;\leq\; \frac{2C_6}{t^{2-l}}.$$

□

Remark 4.3.2. Probably, it is possible to show the weak convergence of the trajectories to a minimizer of the objective function in case $d = 2$.

Remark 4.3.3. Condition 3 does not contradict with the choice of γ, namely, γ could be chosen to satisfy both of them at the same time:

$$2\gamma(\alpha - \gamma) \;<\; 1 \text{ and } \frac{\alpha}{2} \leq \gamma < \alpha.$$

Indeed, 3 implies

$$\gamma^2 - \alpha\gamma + \frac{1}{2} > 0.$$

If $\alpha < \sqrt{2}$, then $\gamma^2 - \alpha\gamma + \frac{1}{2}$ is always positive and we are free to choose γ such that $\frac{\alpha}{2} \leq \gamma < \alpha$. Otherwise, $\alpha \geq \sqrt{2}$ means that

$$\gamma \; < \; \frac{\alpha - \sqrt{\alpha^2 - 2}}{2} \text{ or } \gamma \; > \; \frac{\alpha + \sqrt{\alpha^2 - 2}}{2}$$

and thus we take

$$\frac{\alpha}{2} \; \leq \; \frac{\alpha + \sqrt{\alpha^2 - 2}}{2} \; < \; \gamma \; < \; \alpha.$$

4.4 Numerical examples

4.4.1 The rates of convergence of the Moreau envelope values

Let us consider the following objective function $\Phi : \mathbb{R} \to \mathbb{R}$, $\Phi(x) = |x| + \frac{x^2}{2}$ and plot the values of its Moreau envelope for different polynomial functions λ and ε in order to illustrate the theoretical results with some numerical examples. We set $\lambda(t) = t^l$ and $\varepsilon(t) = \frac{1}{t^d}$ with $x(t_0) = x_0 = 10$, $\dot{x}(t_0) = 0$, $\alpha = 10$, $\beta = 1$ and $t_0 = 1$.

Consider different Moreau envelope parameters λ with $d = 1.9$:

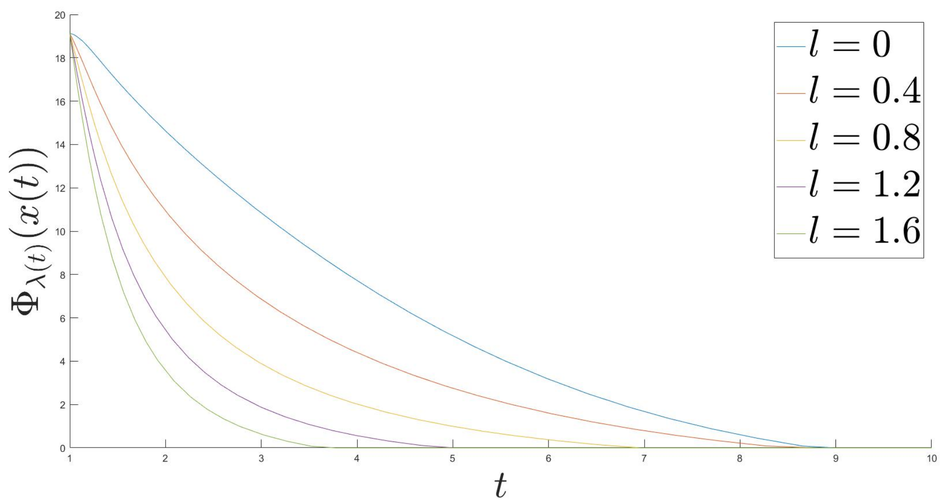


Figure 4.1: Moreau envelope values for $d = 1.9$

We notice that a faster growing function λ implies faster convergence of the Moreau envelope of the objective function Φ.

Increasing the speed of decay of the Tikhonov function ε for a fixed $l = 1$ accelerates the convergence of the Moreau envelope values, which was predicted by the theory:

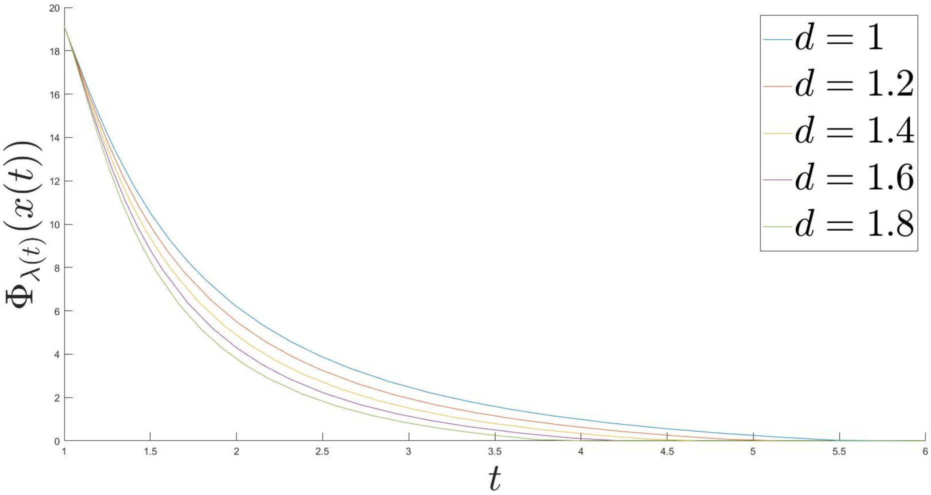


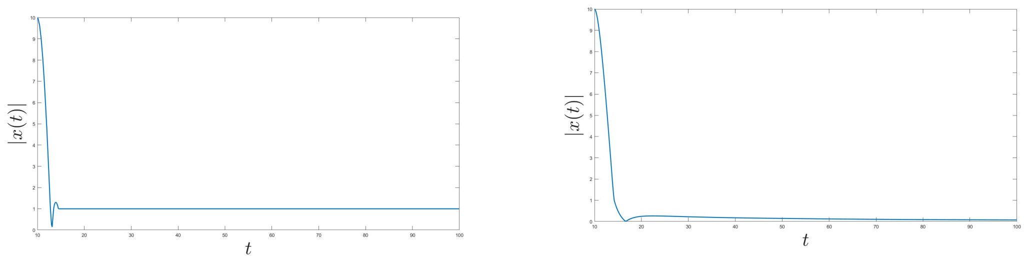
Figure 4.2: Moreau envelope values for $l = 1$

4.4.2 Strong convergence of the trajectories

For the different objective function let us investigate the strong convergence of the trajectories of (4.1) and show some examples when the trajectories actually diverge due to one of the key assumptions of the analysis not being fulfilled. We define

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

The set $\operatorname{argmin}\Phi$ is the segment $[-1,1]$ and clearly 0 is its element of the minimal norm. Let us investigate the influence of the Tikhonov term on the behaviour of the trajectories of the system for $\lambda(t) = t$.



(a) $\varepsilon(t) = 0$

(b) $\varepsilon(t) = \frac{1}{t^{\frac{3}{2}}}$

Figure 4.3: The role of Tikhonov term.

As we can see in case Tikhonov function is missing the trajectories converge to a minimizer 1 of Φ, however, Tikhonov term ensures the convergence to the minimal norm solution 0.

Finally, for the same choice of λ and Φ let us take different Tikhonov functions to study their effect on the trajectories of (4.1). For this purpose we increase the starting point to $x(t_0) = 100$.

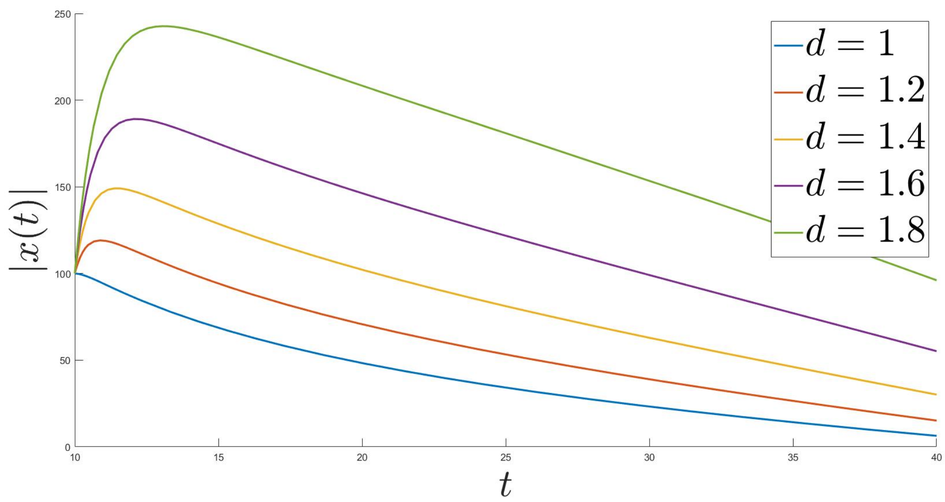


Figure 4.4: $l = 1$.

As we see, the faster ε decays, the slower trajectories converge, which totally corresponds to the theoretical results.

To end this section let us break some of the fundamental conditions of our analysis and show that there is no convergence of the trajectories in this case.

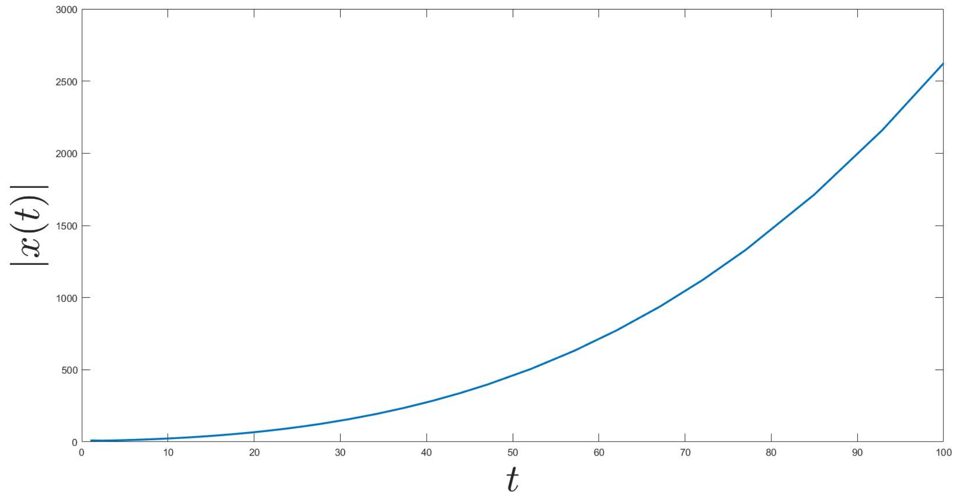


Figure 4.5: l and d do not meet the requirements.

5 Discrete time optimization: inertial algorithm with double Tikhonov regularization

In this chapter we will address the minimization problem (1.1) in a smooth setting using the tools of discrete time optimization, namely, we will associate to (1.1) the following inertial-gradient type algorithm. Let $x_0, x_1 \in H$ and for every $k \geq 1$ set

$$\begin{cases} y_k = x_k + b_{k-1}(x_k - x_{k-1}) - c_k x_k \\ x_{k+1} = y_k - s\nabla\Phi(y_k) - s\varepsilon_k y_k. \end{cases}, \tag{5.1}$$

where $\Phi : H \mapsto \mathbb{R}$ is a convex, continuously Fréchet differentiable function, with L-Lipschitz continuous gradient, whose set of minimizers $\operatorname{argmin} \Phi$ is nonempty. Here $(b_k)_{k\geq 0}$ and $(c_k)_{k\geq 1}$ are given sequences and we assume that the stepsize s satisfies $0 < s < \frac{1}{L}$. Further, we assume that $(\varepsilon_k)_{k\geq 1}$ is a non-increasing positive sequence that satisfies $\lim_{k\to+\infty} \varepsilon_k = 0$.

5.1 Motivation and preliminary results

Observe that in case the inertial parameter $(b_k)_{k\geq 0}$ satisfies $\lim_{k\to+\infty} b_k = 1$ and we take $c_k \equiv 0$, $\varepsilon_k \equiv 0$, then scheme (5.1) has the form of the famous Nesterov algorithm, (see [38,48] and also [14,45]).

Indeed, Nesterov's algorithm in its original form can be written as: $x_0 = x_1 \in H$ and for every $k \geq 1$

$$\begin{cases} y_k = x_k + b_{k-1}(x_k - x_{k-1}) \\ x_{k+1} = y_k - s\nabla\Phi(y_k), \end{cases} \tag{5.2}$$

with the stepsize $0 < s \leq \frac{1}{L}$ and the inertial parameter $b_{k-1} = \frac{t_k - 1}{t_{k+1}}$, where the sequence (t_k) satisfies the recursion $t_1 = 1$, $t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2}$, for every $k \geq 1$. It is easy to verify that indeed $\frac{t_k-1}{t_{k+1}} \to 1$ as $k \to +\infty$ and it is obvious that $t_k \geq \frac{k+1}{2}$ for every $k \geq 1$. Further, if one consider $t_k = r(k-1)$, $r \leq \frac{1}{2}$ then one has $t_{k+1} \leq \frac{1+\sqrt{4t_k^2+1}}{2}$, hence in the literature t_k is usually taken in the form $t_k = \frac{k-1}{\alpha-1}$, $\alpha \geq 3$. In that case algorithm (5.2) becomes: $x_0 = x_1 \in H$ and for $k \geq 1$

$$x_{k+1} = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) - s\nabla\Phi\left(x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1})\right). \tag{5.3}$$

According to [48], the sequences generated by (5.2) satisfy $\Phi(x_k) - \min \Phi = \mathcal{O}(k^{-2})$ as $k \to +\infty$. However, whether the sequences generated by (5.2) converge, is still a widely open problem. In contrast (see, [14]), the sequences generated by (5.3) converge in the weak topology of H to a minimizer of Φ and $\Phi(x_k) - \min \Phi = o(k^{-2})$ as $k \to +\infty$, provided $\alpha > 3$. Nevertheless, also here it is not known whether the generated sequences converge for $\alpha = 3$.

In this section our goal is to provide conditions such that the sequences generated by (5.1) converge, in the strong topology of H, to a minimizer of the objective function Φ. Even more, we propose to have an a priori control to which element from $\operatorname{argmin} \Phi$ our algorithm converges, that is, the minimum norm minimizer of the objective function Φ. At the same time we aim to preserve (as much as possible) the rate $\mathcal{O}(k^{-2})$ for the potential energy $\Phi(x_k) - \min \Phi$, as $k \to +\infty$.

These are the reasons why we introduced the two Tikhonov regularization terms in scheme (5.1). Indeed, the terms $c_k x_k$ and $\varepsilon_k y_k$ in scheme (5.1) play the role of Tikhonov regularization terms, and may assure the strong convergence of the generated sequences to the element of minimal norm from $\operatorname{argmin} \Phi$, (see [6, 9, 22, 24, 26, 27, 29, 30, 39, 40, 44, 45, 55, 56]). Our analysis reveals that the inertial parameter and the Tikhonov regularization parameters are strongly correlated. This fact is in concordance with some recent results from the literature concerning the strong convergence of the trajectories of some continuous second order dynamical systems to a minimal norm minimizer of a convex function or to the minimal norm zero of a maximally monotone operator [5, 9, 18, 21, 24, 26, 29, 30, 34, 41, 45, 46]. In this context we emphasize, that until recently it was an open question whether one can obtain fast convergence of the function values in a trajectory to the minimum of the objective function and at the same time strong convergence results of the generated trajectories.

In concordance to the results obtained in [24] and [44] for inertial-proximal algorithms, the main goal of this paper is to obtain similar results for Nesterov type inertial algorithms. Unfortunately our parameters in scheme (5.1) will not have such simple forms as the parameters in the algorithm (1.20) or algorithm (1.21) and this is due to the fact that we cannot use a discrete Lyapunov function of similar form as the ones considered in [24, 44], instead we have to construct a new discrete Lyapunov function suitable for our analysis. Therefore, the forms of the given sequences $(b_k)_{k\geq 0}$ and $(c_k)_{k\geq 1}$ in scheme (5.1) are crucial in order to obtain our results. More precisely, the sequences $(b_k)_{k\geq 0}$ and $(c_k)_{k\geq 1}$ must satisfy some quite technical conditions, (noted by (B) and (C)), as it will be emphasized later on. A comprehensive analysis on these conditions will also be carried out, here we just underline that in case we specify the Tikhonov regularization parameter ε_k as $\varepsilon_k = \frac{c}{k^p}$, $c, p > 0$ and the sequences $(b_k)_{k\geq 0}$ and $(c_k)_{k\geq 1}$ satisfy the above-mentioned conditions, then the main result of the paper can be summarized in the following theorem.

Theorem 5.1.1. *Assume that $p < 2$ and let $(x_k)_{k\geq 0}$ and $(y_k)_{k\geq 1}$ be the sequences generated by scheme (5.1). Then, (x_k) and (y_k) converge strongly to x^*, where $\{x^*\} = \operatorname{proj}_{\operatorname{argmin} \Phi}(0)$ is the minimum norm minimizer of our objective function Φ. Further, $\Phi(x_k) - \min \Phi = \mathcal{O}\left(k^{-p}\right)$, as $k \to +\infty$, and $\Phi(y_k) - \min \Phi = \mathcal{O}\left(k^{-p}\right)$, as $k \to +\infty$. Additionally, $\|\nabla \Phi(x_k)\| = o\left(k^{-\frac{p}{2}}\right)$ as $k \to +\infty$, $\|\nabla \Phi(y_k)\| = o\left(k^{-\frac{p}{2}}\right)$ as $k \to +\infty$ and*

$\|x_k - x_{k-1}\| \; = \; o\left(k^{-\frac{p}{2}}\right)$ *as* $k \to +\infty$.

Remark 5.1.2. As it was expected in the view of the results concerning the previously presented continuous dynamical systems and algorithms, neither scheme (5.1) will not provide the rate $\Phi\left(x_k\right) - \min \Phi = O\left(k^{-2}\right)$ as $k \to +\infty$. However, when p is very close to 2, then the convergence rate of the values $\Phi(x_k) - \min \Phi = \mathcal{O}\left(k^{-p}\right),$ as $k \to +\infty$, is practically as good as the rate $\Phi\left(x_k\right) - \min \Phi = O\left(k^{-2}\right),$ as $k \to +\infty$, obtained for Nesterov's algorithm.

We continue the present section by emphasizing the main idea behind the Tikhonov regularization, which will assure strong convergence results for the sequence generated by our scheme (5.1) to a minimizer of minimal norm of the objective function. By $\overline{x}_k$ we denote the unique solution of the strongly convex minimization problem

$$\min_{x \in H}\left(\Phi(x) + \frac{\varepsilon_k}{2}\|x\|^2\right).$$

We know, (see, for instance, [22]), that $\lim\limits_{k \to +\infty} \overline{x}_k = x^*$, where $x^* = \operatorname*{argmin}\limits_{x \in \operatorname{argmin} \Phi} \|x\|$ is the minimal norm element from the set $\operatorname{argmin} \Phi$. Obviously, $\{x^*\} = \operatorname{proj}_{\operatorname{argmin} \Phi} 0$ and we have the inequality $\|\overline{x}_k\| \le \|x^*\|$ (see [30]).

Since $\overline{x}_k$ is the unique minimizer of the strongly convex function $\Phi_k(x) = \Phi(x) + \frac{\varepsilon_k}{2}\|x\|^2$, obviously one has

$$\nabla \Phi_k(\overline{x}_k) = \nabla \Phi(\overline{x}_k) + \varepsilon_k \overline{x}_k = 0. \tag{5.4}$$

Further, from Lemma A.1 c) from [44] we have

$$\|\overline{x}_{k+1} - \overline{x}_k\| \le \min\left(\frac{\varepsilon_k - \varepsilon_{k+1}}{\varepsilon_{k+1}}\|\overline{x}_k\|, \frac{\varepsilon_k - \varepsilon_{k+1}}{\varepsilon_k}\|\overline{x}_{k+1}\|\right). \tag{5.5}$$

Note that since Φ_k is strongly convex, from the gradient inequality (1.6) we have

$$\Phi_k(y) - \Phi_k(x) \ge \langle \nabla \Phi_k(x), y - x\rangle + \frac{\varepsilon_k}{2}\|x - y\|^2, \text{ for all } x, y \in H. \tag{5.6}$$

In particular

$$\Phi_k(x) - \Phi_k(\overline{x}_k) \ge \frac{\varepsilon_k}{2}\|x - \overline{x}_k\|^2, \text{ for all } x \in H. \tag{5.7}$$

Moreover, observe that for all $x, y \in H$, one has

$$\begin{aligned} \Phi(x) - \Phi(y) &= (\Phi_k(x) - \Phi_k(\overline{x}_k)) + (\Phi_k(\overline{x}_k) - \Phi_k(y)) + \frac{\varepsilon_k}{2}(\|y\|^2 - \|x\|^2) \\ &\le \Phi_k(x) - \Phi_k(\overline{x}_k) + \frac{\varepsilon_k}{2}\|y\|^2. \end{aligned} \tag{5.8}$$

Note that $\nabla \Phi_k(x) = \nabla \Phi(x) + \varepsilon_k x$, consequently if $\nabla \Phi$ is L-Lipschitz continuous then the Lipschitz constant of the gradient of Φ_k is $L + \varepsilon_k$.

Hence, if we apply Lemma 1.2.10 to Φ_k we get that for all $s \in \left(0, \frac{1}{L+\varepsilon_k}\right]$ one has

$$\begin{aligned} &\Phi_k(y - s\nabla \Phi_k(y)) \\ &\le \Phi_k(x) + \langle \nabla \Phi_k(y), y - x\rangle - \frac{s}{2}\|\nabla \Phi_k(y)\|^2 - \frac{s}{2}\|\nabla \Phi_k(y) - \nabla \Phi_k(x)\|^2, \forall x, y \in H. \end{aligned} \tag{5.9}$$

Now, we can rewrite scheme (5.1) in a more convenable equivalent form, by using the strongly convex function Φ_k. Indeed, since $\nabla\Phi_k(x) = \nabla\Phi(x) + \varepsilon_k x$, scheme (5.1) can equivalently be written as: $x_0, x_1 \in H$ and for every $k \geq 1$

$$\begin{cases} y_k = x_k + b_{k-1}(x_k - x_{k-1}) - c_k x_k \\ x_{k+1} = y_k - s\nabla\Phi_k(y_k). \end{cases} \tag{5.10}$$

5.2 Convergence results

In this section we provide sufficient conditions such that the sequences generated by (5.10) converge strongly to the minimum norm minimizer of Φ and at the same time fast convergence of the function values in the generated sequences and also fast convergence of the discrete velocity to zero are obtained. Moreover, we also show some pointwise estimates for the gradient of the objective function.

In order to obtain our general result concerning the strong convergence of the sequences generated by the scheme (5.10) we need to use (5.9), hence we adjust the indexes in scheme (5.10) as follows.

Assume that $0 < s < \frac{1}{L}$ and let $k_0 \in \mathbb{N}$ such that the following assumption holds.

$$(S) \quad s \in \left(0, \frac{1}{L + \varepsilon_{k_0}}\right] \subseteq \left(0, \frac{1}{L + \varepsilon_k}\right], \text{ for all } k \geq k_0.$$

Note that such index k_0 exists, since ε_k is non-increasing and $\varepsilon_k \to 0$ as $k \to +\infty$. Observe that, since ε_k is non-increasing there exists $k_1 \geq k_0$ such that $1 - s\varepsilon_k > 0$ for every $k \geq k_1$.

Consider further the sequence $(q_k)_{k\geq 0}$ which after an index $k_2 \geq k_1$, satisfies the condition (Q), that is

$$(Q) \quad (1 - s\varepsilon_{k+1})^2 q_{k+1}^2 - (1 - s\varepsilon_k)^2 q_k^2 - 2sq_{k+1} + s(1 - s\varepsilon_k)^2 q_k \leq 0,\ q_k \geq \frac{2s}{(1 - s\varepsilon_k)^2}$$

for every $k \geq k_2$. Note that $q_k \geq 2s > 0$ for every $k \geq k_2$.

Now, having a sequence (q_k) that satisfies condition (Q), the inertial parameter $(b_k)_{k\geq 0}$ and the regularization parameter $(c_k)_{k\geq 1}$ in scheme (5.1) are defined via the conditions

$$(B) \quad \begin{cases} b_{k-1} = 0, \text{ if } k = 1 \text{ or } (1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)q_{k-1}q_k = 0 \\[2ex] b_{k-1} = \dfrac{(q_{k-1} - s)((1 - s\varepsilon_{k-1})^2 q_{k-1} - 2s)}{(1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)q_{k-1}q_k}, \text{ otherwise} \end{cases}$$

and

$$(C) \quad \begin{cases} c_k = 0, \text{ if } k = 1 \text{ or } (1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)q_{k-1}q_k = 0 \\[2ex] c_k = \dfrac{2s}{(1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)^2 q_k}\left(\dfrac{s}{q_{k-1}} - \dfrac{s^2\varepsilon_k}{q_{k-1}} - s(\varepsilon_{k-1} - \varepsilon_k)\right), \text{ otherwise.} \end{cases}$$

Note that despite of the complex form of these parameters, from a practical perspective, scheme (5.1) can easily be implemented. Further, one can show that after an index big enough, the form of the sequences b_k and c_k will be given by the second conditions in (B) and (C).

Indeed, let $\overline{k} = k_2 + 1$ and observe that $(1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)q_{k-1}q_k > 0$ for every $k \geq \overline{k}$, consequently the sequences b_k and c_k defined at (B) and (C) have the following forms:

$$b_{k-1} = \frac{(q_{k-1} - s)((1 - s\varepsilon_{k-1})^2 q_{k-1} - 2s)}{(1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)q_{k-1}q_k}, \text{ for all } k \geq \overline{k}$$

and

$$c_k = \frac{2s}{(1 - s\varepsilon_{k-1})(1 - s\varepsilon_k)^2 q_k} \left(\frac{s}{q_{k-1}} - \frac{s^2 \varepsilon_k}{q_{k-1}} - s(\varepsilon_{k-1} - \varepsilon_k) \right), \text{ for all } k \geq \overline{k}.$$

Remark 5.2.1. Let us show that our condition (Q) is similar to the original Nesterov condition. Indeed, setting $\varepsilon_k \equiv 0$ and $q_k = 2st_k$ with $t_1 \geq 1$ we deduce for k big enough (Q) is equivalent to

$$4s^2 t_{k+1}^2 - 4s^2 t_k^2 - 4s^2 t_{k+1} + 2s^2 t_k \leq 0 \text{ and } t_k \geq 1.$$

The first condition becomes after simplification

$$2t_{k+1}^2 - 2t_{k+1} + \left(t_k - 2t_k^2\right) \leq 0,$$

hence

$$t_{k+1} \in \left[\frac{1 - \sqrt{4t_k^2 - 2t_k + 1}}{2}, \frac{1 + \sqrt{4t_k^2 - 2t_k + 1}}{2} \right],$$

So, one can take $t_{k+1} = \frac{1+\sqrt{4t_k^2 - 2t_k + 1}}{2}$ which is very similar to Nesterov's initial choice of t_k, namely, $t_1 = 1$, $t_{k+1} = \frac{1+\sqrt{4t_k^2+1}}{2}$, for all $k \geq 1$. Even more, for the choice $t_{k+1} = \frac{1+\sqrt{4t_k^2 - 2t_k + 1}}{2}$ one has $t_k \geq 1$ for every $k \geq 1$.

Thus being said, now we are able to prove the main result of the paper.

Theorem 5.2.2. *For a sequence (q_k) satisfying (Q) and the stepsize s satisfying (S), consider the sequences $(b_k)_{k\geq 0}$ and $(c_k)_{k\geq 1}$ defined at (B) and (C) and let $(x_k)_{k\geq 0}$, $(y_k)_{k\geq 1}$ be the sequences generated by scheme (5.1). Assume that the sequence $\left(\frac{q_k \varepsilon_k}{q_{k-1}\varepsilon_{k-1}}\right)_{k\geq \overline{k}}$ is bounded, the sequence $(q_k^2 \varepsilon_k)_{k\geq \overline{k}-1}$ is increasing, further $\lim_{k\to+\infty} q_k^2 \varepsilon_k = +\infty$ and $\lim_{k\to+\infty} \frac{q_k(\varepsilon_k - \varepsilon_{k+1})}{\varepsilon_k} = 0$.*

Then, (x_k) converges strongly to x^*, where $\{x^*\} = \operatorname{proj}_{\operatorname{argmin} \Phi}(0)$ is the minimum norm minimizer of our objective function Φ. Moreover $\|x_k - y_k\| = o\left(\sqrt{\varepsilon_k}\right)$ as $k \to +\infty$, hence (y_k) also converges strongly to x^*.

Further, the following estimates hold:

$$\Phi_k(x_k) - \Phi_k(\overline{x}_k) = o(\varepsilon_k) \text{ as } k \to +\infty,$$

$$\Phi(x_k) - \min \Phi = \mathcal{O}\left(\varepsilon_k\right), \ \textit{as } k \to +\infty \textit{ and } \Phi(y_k) - \min \Phi = \mathcal{O}\left(\varepsilon_k\right), \ \textit{as } k \to +\infty,$$

$$\|x_k - x_{k-1}\| \ = \ o\left(\sqrt{\varepsilon_k}\right) \ \textit{as } k \to +\infty,$$

and

$$\|\nabla\Phi(x_k)\| = \ o\left(\sqrt{\varepsilon_k}\right) \ \textit{as } k \to +\infty \textit{ and } \|\nabla\Phi(y_k)\| = \ o\left(\sqrt{\varepsilon_k}\right) \ \textit{as } k \to +\infty.$$

Proof. For the convenience of the reader we divide our proof in three steps and we explain our techniques.

In step 1 we construct a discrete Lyapunov function in the form

$$E_k = (p_k + q_k)(\Phi_{k+1}(x_{k+1}) - \Phi_{k+1}(\overline{x}_{k+1})) + \|\eta_{k+1} - x^*\|^2$$

and we show the key inequality

$$E_k - E_{k-1} + \frac{s}{q_{k-1}}E_{k-1} \leq \frac{s}{q_{k-1}}\|x^*\|^2 + q_k\varepsilon_k\langle x^*, x^* - \overline{x}_{k+1}\rangle + (p_k + q_k)\frac{\varepsilon_k - \varepsilon_{k+1}}{2}\|\overline{x}_{k+1}\|^2$$

holds for every $k \geq \overline{k}$, where $\overline{k} = k_2 + 1$ and k_2 was defined at (Q).

The forms of the sequences (p_k) and (η_k) will follow from our computations, that is

$$p_k = \frac{(1 - s\varepsilon_k)^2 q_k^2}{2s} - q_k$$

and

$$\eta_k \ = \ \frac{(1 - s\varepsilon_k)q_k}{2s\left(1 - \frac{s}{q_{k-1}}\right)}y_k - \left(\frac{(1 - s\varepsilon_k)q_k}{2s\left(1 - \frac{s}{q_{k-1}}\right)} - \frac{1}{\left(1 - \frac{s}{q_{k-1}}\right)(1 - s\varepsilon_k)}\right)x_k.$$

The form of the sequences (p_k) and (η_k) are actually the reason why we need the conditions (Q), (B) and (C).

In step 2 our aim is to use telescopic sums. To this purpose we multiply the key inequality with $\pi_k = \frac{1}{\prod_{i=\overline{k}}^{k}\left(1 - \frac{s}{q_{i-1}}\right)}$ and we obtain

$$\begin{aligned} &\pi_k E_k - \pi_{k-1}E_{k-1} \\ &\leq (\pi_k - \pi_{k-1})\|x^*\|^2 + q_k\varepsilon_k\pi_k\langle x^*, x^* - \overline{x}_{k+1}\rangle + (p_k + q_k)\frac{\varepsilon_k - \varepsilon_{k+1}}{2}\pi_k\|\overline{x}_{k+1}\|^2, \end{aligned}$$

for all $k > \overline{k}$.

By summing up the above inequality, after some easy computations one gets that

$$E_k = o(q_k^2\varepsilon_k) \text{ as } k \to +\infty.$$

In step 3 by using the above rate we deduce the strong convergence of the generated sequences and the rates, concerning the potential energy, discrete velocity and the value of the gradient in a generated sequence, from the conclusion of the theorem.

Step 1. Let us construct the energy functional E_k. To this purpose assume that $k \geq \overline{k}$ and take $y = y_k$, $x = x_k$ in (5.9) in order to deduce

$$\Phi_k(x_{k+1}) - \Phi_k(x_k) \leq \langle \nabla\Phi_k(y_k), y_k - x_k\rangle - \frac{s}{2}\|\nabla\Phi_k(y_k)\|^2 - \frac{s}{2}\|\nabla\Phi_k(y_k) - \nabla\Phi_k(x_k)\|^2 \tag{5.11}$$

$\forall k \geq \overline{k}$. Now we take $y = y_k$, $x = x^*$ in (5.9) and taking into account that $\nabla\Phi_k(x^*) = \nabla\Phi(x^*) + \varepsilon_k x^*$ and $\nabla\Phi(x^*) = 0$ we obtain

$$\Phi_k(x_{k+1}) - \Phi_k(x^*) \leq \langle \nabla\Phi_k(y_k), y_k - x^*\rangle - \frac{s}{2}\|\nabla\Phi_k(y_k)\|^2 - \frac{s}{2}\|\nabla\Phi_k(y_k) - \varepsilon_k x^*\|^2,\ \forall k \geq \overline{k}. \tag{5.12}$$

Consider the sequence $(p_k)_{k\geq\overline{k}}$ defined by

$$p_k = \frac{(1 - s\varepsilon_k)^2 q_k^2}{2s} - q_k, \tag{5.13}$$

for every $k \geq \overline{k}$. Note that due to assumption (Q) one has $p_k \geq 0$ for every $k \geq \overline{k}$.

We multiply (5.11) with p_k and (5.12) with q_k and sum them up. In the left-hand side of the inequality obtained we get

$$(p_k + q_k)\Phi_k(x_{k+1}) - p_k\Phi_k(x_k) - q_k\Phi_k(x^*).$$

The right-hand side of the inequality obtained becomes

$$\begin{aligned}
&\langle \nabla\Phi_k(y_k), p_k(y_k - x_k)\rangle - \frac{s}{2}p_k\|\nabla\Phi_k(y_k)\|^2 - \frac{s}{2}p_k\|\nabla\Phi_k(y_k) - \nabla\Phi_k(x_k)\|^2\\
&+\ \langle \nabla\Phi_k(y_k), q_k(y_k - x^*)\rangle - \frac{s}{2}q_k\|\nabla\Phi_k(y_k)\|^2 - \frac{s}{2}q_k\|\nabla\Phi_k(y_k) - \varepsilon_k x^*\|^2\\
=&\ \langle \nabla\Phi_k(y_k), (p_k + q_k)y_k - p_k x_k - q_k x^*\rangle - \frac{s}{2}(p_k + q_k)\langle \nabla\Phi_k(y_k), \nabla\Phi_k(y_k)\rangle\\
&-\ \frac{s}{2}p_k\|\nabla\Phi_k(y_k) - \nabla\Phi_k(x_k)\|^2 - \frac{s}{2}q_k\left(\|\nabla\Phi_k(y_k)\|^2 - 2\varepsilon_k\langle \nabla\Phi_k(y_k), x^*\rangle + \varepsilon_k^2\|x^*\|^2\right)\\
=&\ \left\langle \nabla\Phi_k(y_k), (p_k + q_k)y_k - p_k x_k + (s\varepsilon_k - 1)q_k x^* - \frac{s}{2}(p_k + q_k)\nabla\Phi_k(y_k)\right\rangle\\
&-\ \frac{s}{2}p_k\|\nabla\Phi_k(y_k) - \nabla\Phi_k(x_k)\|^2 - \frac{s}{2}q_k\|\nabla\Phi_k(y_k)\|^2 - \frac{s}{2}q_k\varepsilon_k^2\|x^*\|^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
&(p_k + q_k)\Phi_k(x_{k+1}) - p_k\Phi_k(x_k) - q_k\Phi_k(x^*)\\
\leq&\ -\frac{s}{2}p_k\|\nabla\Phi_k(y_k) - \nabla\Phi_k(x_k)\|^2 - \frac{s}{2}q_k\|\nabla\Phi_k(y_k)\|^2 - \frac{s}{2}q_k\varepsilon_k^2\|x^*\|^2\\
&+\left\langle \nabla\Phi_k(y_k), (p_k + q_k)y_k - p_k x_k + (s\varepsilon_k - 1)q_k x^* - \frac{s}{2}(p_k + q_k)\nabla\Phi_k(y_k)\right\rangle,
\end{aligned} \tag{5.14}$$

for every $k \geq \overline{k}$.

Now by neglecting the non-positive terms from the right-hand side of (5.14) we obtain

$$(p_k + q_k)\Phi_k(x_{k+1}) - p_k\Phi_k(x_k) - q_k\Phi_k(x^*) \tag{5.15}$$

$$\leq \left\langle \nabla\Phi_k(y_k), (p_k+q_k)y_k - p_k x_k + (s\varepsilon_k - 1)q_k x^* - \frac{s}{2}(p_k+q_k)\nabla\Phi_k(y_k) \right\rangle,$$

for every $k \geq \overline{k}$.

In what follows we deal with the left-hand side of (5.15). In order to use telescopic sums and taking into account that the minimizer of the strongly convex function Φ_k is $\overline{x}_k$, we manage to have on the left-hand side (5.15) the difference

$$(p_k+q_k)(\Phi_{k+1}(x_{k+1}) - \Phi_{k+1}(\overline{x}_{k+1})) - (p_{k-1}+q_{k-1})(\Phi_k(x_k) - \Phi_k(\overline{x}_k)).$$

First, by using the fact that (ε_k) is non-increasing we have

$$\Phi_k(x_{k+1}) = \Phi_{k+1}(x_{k+1}) + \frac{\varepsilon_k - \varepsilon_{k+1}}{2}\|x_{k+1}\|^2 \geq \Phi_{k+1}(x_{k+1}),$$

consequently it holds

$$\begin{aligned}
&(p_k+q_k)\Phi_k(x_{k+1}) - p_k\Phi_k(x_k) - q_k\Phi_k(x^*) \\
&= (p_k+q_k)(\Phi_k(x_{k+1}) - \Phi_{k+1}(\overline{x}_{k+1})) + (p_k+q_k)\Phi_{k+1}(\overline{x}_{k+1}) \qquad (5.16)\\
&\quad - p_k(\Phi_k(x_k) - \Phi_k(\overline{x}_k)) - p_k\Phi_k(\overline{x}_k) - q_k\Phi_k(x^*) \\
&\geq (p_k+q_k)(\Phi_{k+1}(x_{k+1}) - \Phi_{k+1}(\overline{x}_{k+1})) - (p_{k-1}+q_{k-1})(\Phi_k(x_k) - \Phi_k(\overline{x}_k)) \\
&\quad + (p_{k-1}+q_{k-1}-p_k)(\Phi_k(x_k) - \Phi_k(\overline{x}_k)) + p_k(\Phi_{k+1}(\overline{x}_{k+1}) - \Phi_k(\overline{x}_k)) \\
&\quad + \ q_k(\Phi_{k+1}(\overline{x}_{k+1}) - \Phi_k(x^*)).
\end{aligned}$$

Let us further estimate the terms $p_k(\Phi_{k+1}(\overline{x}_{k+1}) - \Phi_k(\overline{x}_k))$ and $q_k(\Phi_{k+1}(\overline{x}_{k+1}) - \Phi_k(x^*))$ in (5.16).

On one hand, according to (5.7) one has $\Phi_k(\overline{x}_{k+1}) - \Phi_k(\overline{x}_k) \geq \frac{\varepsilon_k}{2}\|\overline{x}_{k+1} - \overline{x}_k\|^2$ hence

$$\begin{aligned}
p_k(\Phi_{k+1}(\overline{x}_{k+1}) - \Phi_k(\overline{x}_k)) &= p_k\left(\Phi_k(\overline{x}_{k+1}) - \Phi_k(\overline{x}_k) + \frac{\varepsilon_{k+1}-\varepsilon_k}{2}\|\overline{x}_{k+1}\|^2\right) \qquad (5.17)\\
&\geq p_k\frac{\varepsilon_k}{2}\|\overline{x}_{k+1} - \overline{x}_k\|^2 + p_k\frac{\varepsilon_{k+1}-\varepsilon_k}{2}\|\overline{x}_{k+1}\|^2 \\
&\geq p_k\frac{\varepsilon_{k+1}-\varepsilon_k}{2}\|\overline{x}_{k+1}\|^2. \qquad (5.18)
\end{aligned}$$

On the other hand, by using the gradient inequality (1.6), we get

$$\begin{aligned}
&q_k(\Phi_{k+1}(\overline{x}_{k+1}) - \Phi_k(x^*)) = q_k\left(\Phi_k(\overline{x}_{k+1}) - \Phi_k(x^*) + \frac{\varepsilon_{k+1}-\varepsilon_k}{2}\|\overline{x}_{k+1}\|^2\right) \geq \qquad (5.19)\\
&q_k\varepsilon_k\langle x^*, \overline{x}_{k+1} - x^*\rangle + q_k\frac{\varepsilon_k}{2}\|\overline{x}_{k+1} - x^*\|^2 + q_k\frac{\varepsilon_{k+1}-\varepsilon_k}{2}\|\overline{x}_{k+1}\|^2 \geq \\
&q_k\varepsilon_k\langle x^*, \overline{x}_{k+1} - x^*\rangle + q_k\frac{\varepsilon_{k+1}-\varepsilon_k}{2}\|\overline{x}_{k+1}\|^2.
\end{aligned}$$

Hence, combining (5.15), (5.16), (5.17) and (5.19) we obtain

$$(p_k+q_k)(\Phi_{k+1}(x_{k+1}) - \Phi_{k+1}(\overline{x}_{k+1})) - (p_{k-1}+q_{k-1})(\Phi_k(x_k) - \Phi_k(\overline{x}_k))$$

$$
\begin{aligned}
&+ (p_{k-1}+q_{k-1}-p_k)(\Phi_k(x_k)-\Phi_k(\overline{x}_k))\\
&\leq q_k\varepsilon_k\langle x^*, x^*-\overline{x}_{k+1}\rangle + (p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\|\overline{x}_{k+1}\|^2 \qquad (5.20)\\
&+ \left\langle \nabla\Phi_k(y_k), (p_k+q_k)y_k - p_kx_k + (s\varepsilon_k-1)q_kx^* - \frac{s}{2}(p_k+q_k)\nabla\Phi_k(y_k)\right\rangle,
\end{aligned}
$$

for every $k \geq \overline{k}$. Now, according to the form of p_k and condition (Q) one has

$$
\begin{aligned}
p_{k-1}+q_{k-1}-p_k &= \frac{(1-s\varepsilon_{k-1})^2q_{k-1}^2}{2s} - \frac{(1-s\varepsilon_k)^2q_k^2}{2s} + q_k \;\geq\; s\frac{(1-s\varepsilon_{k-1})^2}{2s}q_{k-1}\\
&= (p_{k-1}+q_{k-1})\frac{s}{q_{k-1}},
\end{aligned}
$$

for every $k \geq \overline{k}$. Consequently, (5.20) leads to

$$
\begin{aligned}
&(p_k+q_k)(\Phi_{k+1}(x_{k+1})-\Phi_{k+1}(\overline{x}_{k+1})) - (p_{k-1}+q_{k-1})(\Phi_k(x_k)-\Phi_k(\overline{x}_k))\\
&+ (p_{k-1}+q_{k-1})\frac{s}{q_{k-1}}(\Phi_k(x_k)-\Phi_k(\overline{x}_k))\\
&\leq q_k\varepsilon_k\langle x^*, x^*-\overline{x}_{k+1}\rangle + (p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\|\overline{x}_{k+1}\|^2 \qquad (5.21)\\
&+ \left\langle \nabla\Phi_k(y_k), (p_k+q_k)y_k - p_kx_k + (s\varepsilon_k-1)q_kx^* - \frac{s}{2}(p_k+q_k)\nabla\Phi_k(y_k)\right\rangle,
\end{aligned}
$$

for every $k \geq \overline{k}$. In what follows, in order to obtain the key inequality

$$
E_k - E_{k-1} + \frac{s}{q_{k-1}}E_{k-1} \leq \frac{s}{q_{k-1}}\|x^*\|^2 + q_k\varepsilon_k\langle x^*, x^*-\overline{x}_{k+1}\rangle + (p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\|\overline{x}_{k+1}\|^2,
$$

we show that there exists a sequence (η_k) such that

$$
\begin{aligned}
&\left\langle \nabla\Phi_k(y_k), (p_k+q_k)y_k - p_kx_k + (s\varepsilon_k-1)q_kx^* - \frac{s}{2}(p_k+q_k)\nabla\Phi_k(y_k)\right\rangle \leq \qquad (5.22)\\
&\left(1-\frac{s}{q_{k-1}}\right)\|\eta_k-x^*\|^2 - \|\eta_{k+1}-x^*\|^2 + \frac{s}{q_{k-1}}\|x^*\|^2, \text{ for all } k \geq \overline{k}.
\end{aligned}
$$

Indeed, for every $k \geq \overline{k}$, consider the sequence

$$
\begin{aligned}
\eta_k &= \frac{(p_k+q_k)y_k - p_kx_k}{\left(1-\frac{s}{q_{k-1}}\right)(1-s\varepsilon_k)q_k}\\
&= \frac{(1-s\varepsilon_k)q_k}{2s\left(1-\frac{s}{q_{k-1}}\right)}y_k - \left(\frac{(1-s\varepsilon_k)q_k}{2s\left(1-\frac{s}{q_{k-1}}\right)} - \frac{1}{\left(1-\frac{s}{q_{k-1}}\right)(1-s\varepsilon_k)}\right)x_k.
\end{aligned}
$$

We show next, that η_{k+1} can be expressed as a linear combination of η_k and $\nabla\Phi_k(y_k)$. Indeed,

$$
\eta_{k+1} = \frac{(1-s\varepsilon_{k+1})q_{k+1}}{2s\left(1-\frac{s}{q_k}\right)}y_{k+1} - \left(\frac{(1-s\varepsilon_{k+1})q_{k+1}}{2s\left(1-\frac{s}{q_k}\right)} - \frac{1}{\left(1-\frac{s}{q_k}\right)(1-s\varepsilon_{k+1})}\right)x_{k+1},
$$

and according to scheme (5.10) we have $y_{k+1} = (1 + b_k - c_{k+1})x_{k+1} - b_k x_k$ and $x_{k+1} = y_k - s\nabla\Phi_k(y_k)$. Hence,

$$\begin{aligned}
\eta_{k+1} &= \frac{(1 - s\varepsilon_{k+1})q_{k+1}((1 + b_k - c_{k+1})x_{k+1} - b_k x_k)}{2s\left(1 - \frac{s}{q_k}\right)} \\
&\quad - \left(\frac{(1 - s\varepsilon_{k+1})q_{k+1}}{2s\left(1 - \frac{s}{q_k}\right)} - \frac{1}{\left(1 - \frac{s}{q_k}\right)(1 - s\varepsilon_{k+1})}\right)x_{k+1} \\
&= \left(\frac{(1 - s\varepsilon_{k+1})q_{k+1}(b_k - c_{k+1})}{2s\left(1 - \frac{s}{q_k}\right)} + \frac{1}{\left(1 - \frac{s}{q_k}\right)(1 - s\varepsilon_{k+1})}\right)x_{k+1} \\
&\quad - \frac{(1 - s\varepsilon_{k+1})q_{k+1}b_k}{2s\left(1 - \frac{s}{q_k}\right)}x_k \qquad (5.23) \\
&= \left(\frac{(1 - s\varepsilon_{k+1})q_{k+1}(b_k - c_{k+1})}{2s\left(1 - \frac{s}{q_k}\right)} + \frac{1}{\left(1 - \frac{s}{q_k}\right)(1 - s\varepsilon_{k+1})}\right)y_k \\
&\quad - \frac{(1 - s\varepsilon_{k+1})q_{k+1}b_k}{2s\left(1 - \frac{s}{q_k}\right)}x_k \qquad (5.24) \\
&\quad - \left(\frac{(1 - s\varepsilon_{k+1})q_{k+1}(b_k - c_{k+1})}{2s\left(1 - \frac{s}{q_k}\right)} + \frac{1}{\left(1 - \frac{s}{q_k}\right)(1 - s\varepsilon_{k+1})}\right)s\nabla\Phi_k(y_k) \\
&= \frac{(1 - s\varepsilon_k)q_k}{2s}y_k - \left(\frac{(1 - s\varepsilon_k)q_k}{2s} - \frac{1}{1 - s\varepsilon_k}\right)x_k - \frac{(1 - s\varepsilon_k)q_k}{2}\nabla\Phi_k(y_k) \\
&= \left(1 - \frac{s}{q_{k-1}}\right)\eta_k - \frac{(1 - s\varepsilon_k)q_k}{2}\nabla\Phi_k(y_k), \text{ for all } k \geq \overline{k}. \qquad (5.25)
\end{aligned}$$

The last two equalities come from the fact that

$$b_k = \frac{(q_k - s)((1 - s\varepsilon_k)^2 q_k - 2s)}{(1 - s\varepsilon_k)(1 - s\varepsilon_{k+1})q_k q_{k+1}}, \text{ for all } k \geq \overline{k}$$

and

$$c_{k+1} = \frac{2s}{(1 - s\varepsilon_k)(1 - s\varepsilon_{k+1})^2 q_{k+1}}\left(\frac{s}{q_k} - \frac{s^2\varepsilon_{k+1}}{q_k} - s(\varepsilon_k - \varepsilon_{k+1})\right),$$

hence

$$\begin{aligned}
\frac{(1 - s\varepsilon_{k+1})q_{k+1}b_k}{2s\left(1 - \frac{s}{q_k}\right)} &= \frac{(1 - s\varepsilon_{k+1})q_k q_{k+1}}{2s\,(q_k - s)} \cdot \frac{(q_k - s)((1 - s\varepsilon_k)^2 q_k - 2s)}{(1 - s\varepsilon_k)(1 - s\varepsilon_{k+1})q_k q_{k+1}} \\
&= \frac{(1 - s\varepsilon_k)q_k}{2s} - \frac{1}{1 - s\varepsilon_k}
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{(1-s\varepsilon_{k+1})q_{k+1}(b_k-c_{k+1})}{2s\left(1-\frac{s}{q_k}\right)}+\frac{1}{\left(1-\frac{s}{q_k}\right)(1-s\varepsilon_{k+1})}\\
=\ &\frac{(1-s\varepsilon_k)q_k}{2s}-\frac{1}{1-s\varepsilon_k}+\frac{1}{\left(1-\frac{s}{q_k}\right)(1-s\varepsilon_{k+1})}\\
&-\frac{(1-s\varepsilon_{k+1})q_kq_{k+1}}{2s\,(q_k-s)}\frac{2s}{(1-s\varepsilon_k)(1-s\varepsilon_{k+1})^2q_{k+1}}\\
&\cdot\left(\frac{s(1-s\varepsilon_{k+1})}{q_k}+((1-s\varepsilon_k)-(1-s\varepsilon_{k+1}))\right)=\\
&\frac{(1-s\varepsilon_k)q_k}{2s}-\frac{1}{1-s\varepsilon_k}+\frac{q_k}{(q_k-s)\,(1-s\varepsilon_{k+1})}\\
&-\frac{q_k}{(q_k-s)\,(1-s\varepsilon_k)(1-s\varepsilon_{k+1})}\left(\frac{(s-q_k)(1-s\varepsilon_{k+1})}{q_k}+(1-s\varepsilon_k)\right)\\
=\ &\frac{(1-s\varepsilon_k)q_k}{2s}.
\end{aligned}
$$

Further,

$$
\left(1-\frac{s}{q_{k-1}}\right)\eta_k=\frac{(1-s\varepsilon_k)q_k}{2s}y_k-\left(\frac{(1-s\varepsilon_k)q_k}{2s}-\frac{1}{(1-s\varepsilon_k)}\right)x_k
$$

Now, by using (5.23) and the facts that $\frac{(1-s\varepsilon_k)^2q_k^2}{4}=\frac{s}{2}(p_k+q_k)$ and $(1-s\varepsilon_k)q_k\left(1-\frac{s}{q_{k-1}}\right)\eta_k=(p_k+q_k)y_k-p_kx_k$ we get

$$
\begin{aligned}
&\left(1-\frac{s}{q_{k-1}}\right)\|\eta_k-x^*\|^2-\|\eta_{k+1}-x^*\|^2+\frac{s}{q_{k-1}}\|x^*\|^2\\
=\ &\left(1-\frac{s}{q_{k-1}}\right)\|\eta_k\|^2-\|\eta_{k+1}\|^2-2\left\langle\left(1-\frac{s}{q_{k-1}}\right)\eta_k-\eta_{k+1},x^*\right\rangle\\
=\ &\left(1-\frac{s}{q_{k-1}}\right)\|\eta_k\|^2-\left\|\left(1-\frac{s}{q_{k-1}}\right)\eta_k-\frac{(1-s\varepsilon_k)q_k}{2}\nabla\Phi_k(y_k)\right\|^2\\
&-\ 2\left\langle\frac{(1-s\varepsilon_k)q_k}{2}\nabla\Phi_k(y_k),x^*\right\rangle\\
=\ &\left(1-\frac{s}{q_{k-1}}\right)\frac{s}{q_{k-1}}\|\eta_k\|^2+\left\langle\left(1-\frac{s}{q_{k-1}}\right)\eta_k,(1-s\varepsilon_k)q_k\nabla\Phi_k(y_k)\right\rangle\\
&-\ \frac{(1-s\varepsilon_k)^2q_k^2}{4}\|\nabla\Phi_k(y_k)\|^2-\langle(1-s\varepsilon_k)q_k\nabla\Phi_k(y_k),x^*\rangle\\
=\ &\left\langle\nabla\Phi_k(y_k),(p_k+q_k)y_k-p_kx_k+(s\varepsilon_k-1)q_kx^*-\frac{s}{2}(p_k+q_k)\nabla\Phi_k(y_k)\right\rangle\\
&+\ \left(1-\frac{s}{q_{k-1}}\right)\frac{s}{q_{k-1}}\|\eta_k\|^2.
\end{aligned}
$$

Since, $q_k \geq 2s$ for every $k \geq \overline{k}-1$ we get that $1-\frac{s}{q_{k-1}} \geq \frac{1}{2}$ for every $k \geq \overline{k}$ hence the previous equality implies (5.22).

Consequently, for the discrete Lyapunov function

$$E_k = (p_k + q_k)(\Phi_{k+1}(x_{k+1}) - \Phi_{k+1}(\overline{x}_{k+1})) + \|\eta_{k+1} - x^*\|^2,$$

(5.21) and (5.22) lead to the key inequality

$$E_k - E_{k-1} + \frac{s}{q_{k-1}} E_{k-1} \tag{5.26}$$

$$\leq \frac{s}{q_{k-1}}\|x^*\|^2 + q_k\varepsilon_k\langle x^*, x^* - \overline{x}_{k+1}\rangle + (p_k+q_k)\frac{\varepsilon_k - \varepsilon_{k+1}}{2}\|\overline{x}_{k+1}\|^2, \tag{5.27}$$

for every $k \geq \overline{k}$.

Step 2. Our goal is to write the left hand side of (5.26) as the difference of two consecutive iterates in order to use telescopic sums. To this purpose consider the sequence $\pi_k = \frac{1}{\prod_{i=\overline{k}}^{k}\left(1-\frac{s}{q_{i-1}}\right)}$. Note that $(\pi_k)_{k\geq\overline{k}}$ is well defined, positive and increasing since by the hypotheses we have $q_{k-1} \geq 2s$ for every $k \geq \overline{k}$. Further, since $\frac{1}{1-\frac{s}{q_{i-1}}} \leq 2$ we have that $\pi_k \leq 2^{k-\overline{k}+1}$.

Now, by multiplying (5.26) with π_k we obtain

$$\begin{aligned} &\pi_k E_k - \pi_{k-1}E_{k-1} \\ &\leq \frac{s}{q_{k-1}}\pi_k\|x^*\|^2 + q_k\varepsilon_k\pi_k\langle x^*, x^* - \overline{x}_{k+1}\rangle + (p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\pi_k\|\overline{x}_{k+1}\|^2 \\ &= (\pi_k - \pi_{k-1})\|x^*\|^2 + q_k\varepsilon_k\pi_k\langle x^*, x^* - \overline{x}_{k+1}\rangle + (p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\pi_k\|\overline{x}_{k+1}\|^2, \end{aligned} \tag{5.28}$$

for all $k > \overline{k}$.

By summing up (5.28) from $k = \overline{k}+1$ to $k = n > \overline{k}+1$ we obtain

$$\pi_n E_n \leq \pi_n\|x^*\|^2 + \sum_{k=\overline{k}+1}^{n} q_k\varepsilon_k\pi_k\langle x^*, x^* - \overline{x}_{k+1}\rangle \tag{5.29}$$

$$+ \sum_{k=\overline{k}+1}^{n}(p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\pi_k\|\overline{x}_{k+1}\|^2 + C, \tag{5.30}$$

where C is a nonnegative constant that satisfies $C \geq \pi_{\overline{k}}E_{\overline{k}} - \pi_{\overline{k}}\|x^*\|^2$.

Next we show that

$$\frac{\pi_n\|x^*\|^2 + \sum_{k=\overline{k}+1}^{n} q_k\varepsilon_k\pi_k\langle x^*, x^* - \overline{x}_{k+1}\rangle + \sum_{k=\overline{k}+1}^{n}(p_k+q_k)\frac{\varepsilon_k-\varepsilon_{k+1}}{2}\pi_k\|\overline{x}_{k+1}\|^2 + C}{\pi_n}$$

$$= o(q_n^2\varepsilon_n),$$

as $n \to +\infty$. Indeed, according to the hypotheses $q_n^2\varepsilon_n \to +\infty$ as $n \to +\infty$ and we know that (π_n) is increasing, hence $\frac{\pi_n\|x^*\|+C}{\pi_n} = o(q_n^2\varepsilon_n)$ as $n \to +\infty$.

Further, since according to the hypotheses the sequence $\left(\frac{q_n\varepsilon_n}{q_{n-1}\varepsilon_{n-1}}\right)$ is bounded, the sequence $(q_n^2\varepsilon_n\pi_n)$ is increasing and $\lim_{n\to+\infty} q_n^2\varepsilon_n\pi_n = +\infty$, by using the fact that $\lim_{n\to+\infty}\langle x^*, x^* - \overline{x}_{n+1}\rangle = 0$, via the Cesàro-Stolz theorem we get

$$\begin{aligned}
\lim_{n\to+\infty} \frac{\sum_{k=\overline{k}+1}^n q_k\varepsilon_k\pi_k\langle x^*, x^* - \overline{x}_{k+1}\rangle}{q_n^2\varepsilon_n\pi_n} &= \lim_{n\to+\infty} \frac{q_n\varepsilon_n\pi_n\langle x^*, x^* - \overline{x}_{n+1}\rangle}{q_n^2\varepsilon_n\pi_n - q_{n-1}^2\varepsilon_{n-1}\pi_{n-1}} \\
&= \lim_{n\to+\infty} \frac{\langle x^*, x^* - \overline{x}_{n+1}\rangle}{q_n - \frac{q_{n-1}^2\varepsilon_{n-1}\pi_{n-1}}{q_n\varepsilon_n\pi_n}} \\
&= \lim_{n\to+\infty} \frac{\langle x^*, x^* - \overline{x}_{n+1}\rangle}{\frac{q_n^2\varepsilon_n - q_{n-1}^2\varepsilon_{n-1}}{q_n\varepsilon_n} + s\frac{q_{n-1}\varepsilon_{n-1}}{q_n\varepsilon_n}} \\
&\leq \lim_{n\to+\infty} \frac{\langle x^*, x^* - \overline{x}_{n+1}\rangle}{s\frac{q_{n-1}\varepsilon_{n-1}}{q_n\varepsilon_n}} = 0.
\end{aligned}$$

Finally, according to the hypotheses $\lim_{n\to+\infty} \frac{q_n(\varepsilon_n - \varepsilon_{n+1})}{\varepsilon_n} = 0$, hence by using the fact that $p_n + q_n = \frac{(1-s\varepsilon_n)^2 q_n^2}{2s}$, for some $M > 0$ one has

$$\begin{aligned}
&\lim_{n\to+\infty} \frac{\sum_{k=\overline{k}+1}^n (p_k + q_k)\frac{\varepsilon_k - \varepsilon_{k+1}}{2}\pi_k\|\overline{x}_{k+1}\|^2}{q_n^2\varepsilon_n\pi_n} \\
&= \frac{1}{4s}\lim_{n\to+\infty} \frac{(1-s\varepsilon_n)^2 q_n^2(\varepsilon_n - \varepsilon_{n+1})\pi_n\|\overline{x}_{n+1}\|^2}{q_n^2\varepsilon_n\pi_n - q_{n-1}^2\varepsilon_{n-1}\pi_{n-1}} \\
&= \frac{1}{4s}\lim_{n\to+\infty} \frac{(1-s\varepsilon_n)^2(\varepsilon_n - \varepsilon_{n+1})\|\overline{x}_{n+1}\|^2}{\varepsilon_n - \frac{q_{n-1}^2\varepsilon_{n-1}\pi_{n-1}}{q_n^2\pi_n}} \\
&= \frac{1}{4s}\lim_{n\to+\infty} \frac{(1-s\varepsilon_n)^2(\varepsilon_n - \varepsilon_{n+1})\|\overline{x}_{n+1}\|^2}{\frac{q_n^2\varepsilon_n - q_{n-1}^2\varepsilon_{n-1}}{q_n^2} + s\frac{q_{n-1}\varepsilon_{n-1}}{q_n^2}} \\
&= \frac{1}{4s}\lim_{n\to+\infty} \frac{(1-s\varepsilon_n)^2\|\overline{x}_{n+1}\|^2}{\frac{q_n^2\varepsilon_n - q_{n-1}^2\varepsilon_{n-1}}{q_n\varepsilon_n} + s\frac{q_{n-1}\varepsilon_{n-1}}{q_n\varepsilon_n}} \cdot \frac{q_n(\varepsilon_n - \varepsilon_{n+1})}{\varepsilon_n} \\
&\leq M \lim_{n\to+\infty} \frac{q_n(\varepsilon_n - \varepsilon_{n+1})}{\varepsilon_n} = 0.
\end{aligned}$$

Consequently, from (5.29) we get

$$E_n = o(q_n^2\varepsilon_n) \text{ as } n \to +\infty. \tag{5.31}$$

Step 3. Now using (5.31) we derive the strong convergence of (x_n) and (y_n) to x^* and the convergence rates concerning the potential energies, discrete velocity and values of the gradient stated in the conclusion of the theorem.

Indeed, taking into account the form of E_n and the fact that $p_n + q_n = \frac{(1-s\varepsilon_n)^2 q_n^2}{2s}$, (5.31) leads to $(p_n + q_n)(\Phi_{n+1}(x_{n+1}) - \Phi_{n+1}(\overline{x}_{n+1})) = o(q_n^2\varepsilon_n)$ as $n \to +\infty$. In other words, $\Phi_n(x_n) - \Phi_n(\overline{x}_n) = o(\varepsilon_n)$ as $n \to +\infty$ and by using (5.8) we get

$$\Phi(x_n) - \min \Phi = \mathcal{O}(\varepsilon_n) \text{ as } n \to +\infty.$$

In order to show strong convergence, we use (5.7) and we get

$$\begin{aligned}\lim_{n\to+\infty} \|x_n - x^*\|^2 &\leq 2 \lim_{n\to+\infty} \left(\|x_n - \overline{x}_n\|^2 + \|\overline{x}_n - x^*\|^2\right)\\ &\leq 4 \lim_{n\to+\infty} \frac{\Phi_n(x_n) - \Phi_n(\overline{x}_n)}{\varepsilon_n} + 2 \lim_{n\to+\infty} \|\overline{x}_n - x^*\|^2 = 0.\end{aligned}$$

Concerning the rates of convergence for the discrete velocity $\|x_n - x_{n-1}\|$ we conclude the following. From the definition of E_n and the fact that $E_n = o\left(q_n^2\varepsilon_n\right)$ as $n \to +\infty$ we have that

$$\|\eta_n - x^*\| = o\left(q_n\sqrt{\varepsilon_n}\right) \text{ as } n \to +\infty.$$

Now, using the definition of η_n and the fact that $y_n = x_n + b_{n-1}(x_n - x_{n-1}) - c_n x_n$ we derive

$$\begin{aligned}&\eta_n - x^*\\ &= \frac{(1-s\varepsilon_n)q_n}{2s(1-\frac{s}{q_{n-1}})} y_n - \left(\frac{(1-s\varepsilon_n)q_n}{2s(1-\frac{s}{q_{n-1}})} - \frac{1}{(1-\frac{s}{q_{n-1}})(1-s\varepsilon_n)}\right) x_n - x^*\\ &= \frac{(1-s\varepsilon_n)q_n}{2s(1-\frac{s}{q_{n-1}})} b_{n-1}(x_n - x_{n-1}) + \left(\frac{1}{(1-\frac{s}{q_{n-1}})(1-s\varepsilon_n)} - \frac{(1-s\varepsilon_n)q_n}{2s(1-\frac{s}{q_{n-1}})} c_n\right) x_n - x^*\end{aligned}$$

Now, since $(q_n^2\varepsilon_n)$ is increasing, and (ε_n) is non-increasing we deduce that (q_n) is increasing. Further, since $\lim_{n\to+\infty} q_n^2\varepsilon_n = +\infty$ and $\lim_{n\to+\infty} \varepsilon_n = 0$ we obtain that $\lim_{n\to+\infty} q_n = +\infty$. Consequently, for n big enough one has

$$b_{n-1} = \frac{(q_{n-1} - s)((1-s\varepsilon_{n-1})^2 q_{n-1} - 2s)}{(1-s\varepsilon_{n-1})(1-s\varepsilon_n)q_{n-1}q_n} \leq \frac{(1-s\varepsilon_{n-1})q_{n-1}}{(1-s\varepsilon_n)q_n} < 1$$

and

$$\lim_{n\to+\infty} \frac{c_n}{\sqrt{\varepsilon_n}} = \lim_{n\to+\infty} \frac{2s}{q_n\sqrt{\varepsilon_n}} \frac{1}{(1-s\varepsilon_{n-1})(1-s\varepsilon_n)^2} \left(\frac{s}{q_{n-1}} - \frac{s^2\varepsilon_n}{q_{n-1}} - s(\varepsilon_{n-1} - \varepsilon_n)\right) = 0.$$

Let us show that $\liminf_{n\to+\infty} b_{n-1} > 0$. To this purpose, observe that it is enough to show that $\liminf_{n\to+\infty} \frac{q_{n-1}}{q_n} > 0$. So, assume the contrary, that is $\liminf_{n\to+\infty} \frac{q_{n-1}}{q_n} = 0$. Then, $\limsup_{n\to+\infty} \frac{q_n}{q_{n-1}} = +\infty$. Now, from the assumption $\lim_{n\to+\infty} \frac{q_n(\varepsilon_n - \varepsilon_{n+1})}{\varepsilon_n} = 0$ and the fact that $\lim_{n\to+\infty} q_n = +\infty$ we get that $\lim_{n\to+\infty} \frac{\varepsilon_n - \varepsilon_{n+1}}{\varepsilon_n} = 0$, hence $\lim_{n\to+\infty} \frac{\varepsilon_n}{\varepsilon_{n-1}} = 1$. Consequently, $\limsup_{n\to+\infty} \frac{\varepsilon_n q_n}{\varepsilon_{n-1}q_{n-1}} = +\infty$, which contradicts the assumption in the hypotheses that the sequence $\left(\frac{q_n\varepsilon_n}{q_{n-1}\varepsilon_{n-1}}\right)$ is bounded.

Thus, by using the fact that (x_n) is bounded we have

$$\lim_{n\to+\infty}\frac{\left(\frac{1}{(1-\frac{s}{q_{n-1}})(1-s\varepsilon_n)}-\frac{(1-s\varepsilon_n)q_n}{2s(1-\frac{s}{q_{n-1}})}c_n\right)x_n-x^*}{q_n\sqrt{\varepsilon_n}}=0,$$

which combined with the fact that $\lim_{n\to+\infty}\frac{\eta_n-x^*}{q_n\sqrt{\varepsilon_n}}=0$ yields

$$\lim_{n\to+\infty}\frac{(1-s\varepsilon_n)}{2s(1-\frac{s}{q_{n-1}})}b_{n-1}\frac{x_n-x_{n-1}}{\sqrt{\varepsilon_n}}=0.$$

But $\liminf_{n\to+\infty}b_{n-1}>0$ and $\lim_{n\to+\infty}\frac{(1-s\varepsilon_n)}{2s(1-\frac{s}{q_{n-1}})}=\frac{1}{2s}$, hence $\|x_n-x_{n-1}\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$.

From here and the fact that $c_n=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$, we deduce at once that $\|y_n-x_n\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$, hence in particular

$$\lim_{n\to+\infty}y_n=x^*.$$

From (5.23) we have $\left(1-\frac{s}{q_{n-1}}\right)\eta_n-\eta_{n+1}=\frac{(1-s\varepsilon_n)q_n}{2}\nabla\Phi_n(y_n)$ and using the fact that $\|\eta_n\|=o\left(q_n\sqrt{\varepsilon_n}\right)$ as $n\to+\infty$ we obtain that $\|\nabla\Phi_n(y_n)\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$, and further by using the fact that $\sqrt{\varepsilon_n}\to 0$, $n\to+\infty$ that $\|\nabla\Phi_n(y_n)-\varepsilon_n y_n\|=|\nabla\Phi(y_n)\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$.

By using the $L-$Lipschitz continuity of $\nabla\Phi$ we get $\|\nabla\Phi(x_n)-\nabla\Phi(y_n)\|\leq L\|y_n-x_n\|$ which combined with the facts that $\|y_n-x_n\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$ and $\|\nabla\Phi(y_n)\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$ lead to $\|\nabla\Phi(x_n)\|=o(\sqrt{\varepsilon_n})$ as $n\to+\infty$.

Finally, by using Lemma 1.2.9 item 2 we get

$$\Phi(y_n)-\min\Phi\leq\Phi(x_n)-\min\Phi+\|\nabla\Phi(x_n)\|\|y_n-x_n\|+\frac{L}{2}\|y_n-x_n\|^2$$

hence, from the facts that $\Phi(x_n)-\min\Phi=\mathcal{O}(\varepsilon_n)$, $\|y_n-x_n\|=o(\sqrt{\varepsilon_n})$ and $\|\nabla\Phi(x_n)\|=o(\sqrt{\varepsilon_n})$ s $n\to+\infty$ we obtain that

$$\Phi(y_n)-\min\Phi=\mathcal{O}(\varepsilon_n)\text{ as }n\to+\infty.$$

□

Remark 5.2.3. Note that under the hypotheses of Theorem 5.2.2 one has $\lim_{k\to+\infty}c_k=0$ and c_k is positive for k big enough. Indeed, in the proof of the Theorem 5.2.2 we have shown even more, that is, $\lim_{k\to+\infty}\frac{c_k}{\sqrt{\varepsilon_k}}=0$.

Taking into account that

$$c_k=\frac{2s}{(1-s\varepsilon_{k-1})(1-s\varepsilon_k)^2q_k}\left(\frac{s}{q_{k-1}}-\frac{s^2\varepsilon_k}{q_{k-1}}-s(\varepsilon_{k-1}-\varepsilon_k)\right),$$

in order to show that c_k is positive for k big enough it is enough to show that $\frac{s}{q_{k-1}} - \frac{s^2\varepsilon_k}{q_{k-1}} - s(\varepsilon_{k-1} - \varepsilon_k) > 0$ for k big enough. That is,

$$1 - s\varepsilon_k > q_{k-1}(\varepsilon_{k-1} - \varepsilon_k) \text{ for } k \text{ big enough.}$$

The latter follows directly from the facts that $\lim_{k\to+\infty} \frac{q_{k-1}(\varepsilon_{k-1}-\varepsilon_k)}{\varepsilon_{k-1}} = 0$, $\lim_{k\to+\infty} \frac{\varepsilon_k}{\varepsilon_{k-1}} = 1$ and $\lim_{k\to+\infty} \varepsilon_{k-1} = 0$.

5.3 Particular choice of the parameter sequences (q_k) and (ε_k)

Let us consider a specific choice of the sequences (q_k) and (ε_k) being of the polynomial type, namely, $q_k = ak^q$, $\varepsilon_k = \frac{c}{k^p}$, where $1 \geq q > 0$, $p > 0$ and a and c are positive real numbers. Let us fix $0 < s < \frac{1}{L}$. Then from condition (S) we have $s \leq \frac{1}{L+\varepsilon_{k_0}}$ for some $k_0 \in \mathbb{N}$, and it is an easy computation that

$$k_0 = \operatorname{int}\left(\frac{cs}{1-Ls}\right)^{\frac{1}{p}} + 1,$$

where $\operatorname{int}(x)$ denotes the integer part of x.

Now we compute the index $k_1 \geq k_0$ such that $1 - s\varepsilon_k > 0$ for every $k \geq k_1$. Note that $1 - s\varepsilon_k > 0$ whenever $k \geq \operatorname{int}(cs)^{\frac{1}{p}} + 1$, consequently one can take $k_1 = k_0$.

Condition (Q) in this case becomes: after an index $k_2 \geq k_1$ it holds that

$$\left(1 - \frac{sc}{(k+1)^p}\right)^2 a^2(k+1)^{2q} - \left(1 - \frac{sc}{k^p}\right)^2 a^2k^{2q} - 2sa(k+1)^q + s\left(1 - \frac{sc}{k^p}\right)^2 ak^q \ \leq \ 0$$

and

$$ak^q \ \geq \ \frac{2s}{(1-\frac{sc}{k^p})^2}$$

for all $k \geq k_2$. Note that the second condition is always fulfilled starting from k large enough due to q and p being positive. Let us show that also the first condition in (Q) holds for every $0 < q \leq 1$. First consider the case $q < 1$. Since

$$\left(1 - \frac{sc}{(k+1)^p}\right)^2 a^2(k+1)^{2q} - \left(1 - \frac{sc}{k^p}\right)^2 a^2k^{2q} = \mathcal{O}(k^{2q-1}) \text{ as } k \to +\infty$$

and

$$-2sa(k+1)^q + s\left(1 - \frac{sc}{k^p}\right)^2 ak^q = -sak^q + \mathcal{O}(k^{q-1}) + \mathcal{O}(k^{q-p}) \text{ as } k \to +\infty$$

we obtain that

$$\begin{aligned}&\left(1 - \frac{sc}{(k+1)^p}\right)^2 a^2(k+1)^{2q} - \left(1 - \frac{sc}{k^p}\right)^2 a^2k^{2q} - 2sa(k+1)^q + s\left(1 - \frac{sc}{k^p}\right)^2 ak^q \\ = \ & -sak^q + \mathcal{O}(k^{2q-1}) + \mathcal{O}(k^{q-1}) + \mathcal{O}(k^{q-p}) \text{ as } k \to +\infty.\end{aligned}$$

Obviously, the right-hand side of the previous inequality goes to $-\infty$ as $k \to +\infty$, hence there exists an index $k_2 \geq k_1$ such that

$$\left(1 - \frac{sc}{(k+1)^p}\right)^2 a^2(k+1)^{2q} - \left(1 - \frac{sc}{k^p}\right)^2 a^2k^{2q} - 2sa(k+1)^q + s\left(1 - \frac{sc}{k^p}\right)^2 ak^q \leq 0$$

for every $k \geq k_2$, that is, (Q) holds.

Now, if $q = 1$ we obtain

$$\begin{aligned} &\left(1 - \frac{sc}{(k+1)^p}\right)^2 a^2(k+1)^2 - \left(1 - \frac{sc}{k^p}\right)^2 a^2k^2 - 2sa(k+1) + s\left(1 - \frac{sc}{k^p}\right)^2 ak \\ &= (2a^2 - as)k + \mathcal{O}(k^{1-p}) + \mathcal{O}(1) \text{ as } k \to +\infty, \end{aligned}$$

hence, arguing as before, we deduce that (Q) holds provided $2a^2 - as < 0$, that is $a < \frac{s}{2}$.

Concerning the sequence $(b_k)_{k\geq 0}$ in this particular case condition (B) becomes:

$$(Bp) \quad \begin{cases} b_{k-1} = 0, \text{ if } k \in \left\{1, (cs)^{\frac{1}{p}}, (cs)^{\frac{1}{p}} + 1\right\} \\ \\ b_{k-1} = \dfrac{k^p(a(k-1)^q - s)(a((k-1)^p - cs)^2(k-1)^q - 2s(k-1)^{2p})}{a^2(k-1)^{q+p}k^q((k-1)^p - cs)(k^p - cs)}, \text{otherwise.} \end{cases}$$

Note that $b_k \to 1$ as $k \to +\infty$.

Further, condition (C) becomes:

$$(Cp) \quad \begin{cases} c_k = 0, \text{ if } k \in \left\{1, (cs)^{\frac{1}{p}}, (cs)^{\frac{1}{p}} + 1\right\} \\ \\ c_k = \dfrac{2s^2k^p((k-1)^pk^p - c(k-1)^p - ac(k-1)^qk^p + ac(k-1)^{q+p})}{a^2(k-1)^qk^q((k-1)^p - cs)(k^p - cs)^2}, \text{otherwise.} \end{cases}$$

Note that $c_k > 0$ for k big enough, further $c_k \to 0$ as $k \to +\infty$. Indeed, $-ac(k-1)^qk^p + ac(k-1)^{q+p} = \mathcal{O}(k^{q+p-1})$ and $(k-1)^pk^p - c(k-1)^p = \mathcal{O}(k^{2p})$ as $k \to +\infty$, and $2p > q + p - 1$, hence $c_k > 0$ for k big enough. Further, the order of numerator in c_k is $\mathcal{O}(k^{3p})$ as $k \to +\infty$ and the order of denominator is $\mathcal{O}(k^{3p+2q})$ as $k \to +\infty$, therefore $c_k \to 0$ as $k \to +\infty$.

Hence, indeed in this case the term c_kx_k in scheme (5.1) plays the role of a Tikhonov regularization term. More precisely, scheme (5.1) reads as: $x_0, x_1 \in H$ and for every $k \geq 1$

$$\begin{cases} y_k = x_k, \text{ if } k \in \left\{1, (cs)^{\frac{1}{p}}, (cs)^{\frac{1}{p}} + 1\right\} \\ y_k = x_k + \dfrac{k^p(a(k-1)^q - s)(a((k-1)^p - cs)^2(k-1)^q - 2s(k-1)^{2p})}{a^2(k-1)^{q+p}k^q((k-1)^p - cs)(k^p - cs)}(x_k - x_{k-1}) \\ \quad - \dfrac{2s^2k^p((k-1)^pk^p - c(k-1)^p - ac(k-1)^qk^p + ac(k-1)^{q+p})}{a^2(k-1)^qk^q((k-1)^p - cs)(k^p - cs)^2}x_k, \text{ otherwise} \\ x_{k+1} = y_k - s\nabla\Phi(y_k) - \frac{cs}{k^p}y_k. \end{cases} \tag{5.32}$$

We underline that from a numerical point of view scheme (5.32) can easily be implemented. In this particular case, we have the following result.

Theorem 5.3.1. *Let $0 < q < 1$ and $0 < p < 2q$ and for a fixed the stepsize $s \in \left(0, \frac{1}{L}\right)$ consider the sequences $(x_k)_{k\in\mathbb{N}}$, $(y_k)_{k\in\mathbb{N}}$ generated by scheme (5.32). Then, (x_k) and (y_k) converge strongly to x^*, where $\{x^*\} = \operatorname{proj}_{\operatorname{argmin}\Phi}(0)$ is the minimum norm minimizer of our objective function Φ.*

Further,

$$\Phi_k(x_k) - \Phi_k(\overline{x}_k) = o(k^{-p}) \text{ as } k \to +\infty,$$

$$\Phi(x_k) - \min\Phi = \mathcal{O}\left(k^{-p}\right), \text{ as } k \to +\infty \text{ and } \Phi(y_k) - \min\Phi = \mathcal{O}\left(k^{-p}\right), \text{ as } k \to +\infty,$$

$$\|x_k - x_{k-1}\| = o\left(k^{-\frac{p}{2}}\right) \text{ as } k \to +\infty$$

and

$$\|\nabla\Phi(x_k)\| = o\left(k^{-\frac{p}{2}}\right) \text{ as } k \to +\infty \text{ and } \|\nabla\Phi(y_k)\| = o\left(k^{-\frac{p}{2}}\right) \text{ as } k \to +\infty$$

Proof. We only need to show that the following conditions from the hypotheses of Theorem 5.2.2 hold:

- the sequence $\frac{q_k\varepsilon_k}{q_{k-1}\varepsilon_{k-1}} = \frac{k^{q-p}}{(k-1)^{q-p}}$ is bounded if we take the starting index big enough;
- the sequence $q_k^2\varepsilon_k = a^2ck^{2q-p}$ is increasing after a starting index big enough;
- $\lim_{k\to+\infty} q_k^2\varepsilon_k = \lim_{k\to+\infty} a^2ck^{2q-p} = +\infty$;
- $\lim_{k\to+\infty} \frac{q_k(\varepsilon_k-\varepsilon_{k+1})}{\varepsilon_k} = \lim_{k\to+\infty} \frac{ak^q\left(\frac{1}{k^p}-\frac{1}{(k+1)^p}\right)}{\frac{1}{k^p}} = 0.$

First of all, the sequence $\frac{k^{q-p}}{(k-1)^{q-p}}$ is indeed bounded if we take the starting index big enough since $\lim_{k\to+\infty} \frac{k^{q-p}}{(k-1)^{q-p}} = 1$. Secondly, the sequence a^2ck^{2q-p} is increasing when $2q > p$ and $\lim_{k\to+\infty} a^2ck^{2q-p} = +\infty$ when $2q > p$. Finally,

$$\lim_{k\to+\infty} \frac{ak^q\left(\frac{1}{k^p}-\frac{1}{(k+1)^p}\right)}{\frac{1}{k^p}} = \lim_{k\to+\infty} \frac{ak^q}{k+1}\,\frac{(k+1)^p - k^p}{(k+1)^{p-1}} = 0,$$

since $q < 1$. $\square$

Remark 5.3.2. Note that condition (Q) also holds in the case $q_k = ak$, $a < \frac{s}{2}$. Unfortunately, in this case Theorem 5.2.2 cannot be applied, since for $p < 2q = 2$ one has

$$\lim_{k\to+\infty} \frac{q_k(\varepsilon_k - \varepsilon_{k+1})}{\varepsilon_k} = \lim_{k\to+\infty} \frac{ak\left(\frac{1}{k^p}-\frac{1}{(k+1)^p}\right)}{\frac{1}{k^p}} = p \neq 0.$$

Remark 5.3.3. We emphasize that scheme (5.32) can be seen as a Nesterov type algorithm with two Tikhonov regularization terms. Indeed, the extrapolation parameter (b_k) goes to 1 as $k \to +\infty$ such as in the Nesterov algorithm. Further, the terms c_kx_k and ε_ky_k

can be thought as Tikhonov regularization terms since both c_k and ε_k are nonnegative sequences (after k big enough), and go to 0 as $k \to +\infty$. Unfortunately, we could not allow the case $q = 1$ and $p = 2$ in our algorithm. Nevertheless, if p is close to 2, (and q is close to 1), then from a numerical perspective the convergence rates obtained for the potential energy $\Phi(x_k) - \min \Phi$ and discrete velocity $\|x_k - x_{k-1}\|$ are as good as the rates obtained for the famous Nesterov algorithm, see [48]. Moreover, our algorithm assures the strong convergence of the generated sequences to the minimum norm minimizer - a feature that makes it unique in the literature.

5.4 Numerical experiments

In this section we consider some numerical experiments in order to sustain the theoretical results obtained in Theorem 5.3.1. To this purpose, let us consider the objective function

$$\Phi : \mathbb{R}^2 \mapsto \mathbb{R},\ \ \Phi(x, y) = (ax + by)^2,$$

where $a, b \in \mathbb{R} \setminus \{0\}$. Then obviously Φ is smooth and convex and its gradient is Lipschitz continuous, having Lipschitz constant $L = 2\sqrt{2}\sqrt{(a^2 + b^2)\max(a^2, b^2)}$. Observe that the minimal value of Φ is 0 and the set $\operatorname{argmin} \Phi$ is $\left\{\left(x, -\frac{a}{b}x\right) : x \in \mathbb{R}\right\}$, further clearly $(0, 0)$ is the minimizer of minimal norm. For simplicity in the following experiments concerning scheme (5.32) we take everywhere $s = 0.1$, (which will always satisfy $s < \frac{1}{L}$), and fix the starting points $x_0 = (1, -1)$ and $x_1 = (-1, 1)$.

In our first experiment we fix $a = 0.1$ and $b = 100$ and $q_k = k^{0.99}$. Further, in scheme (5.32) we set $\varepsilon_k = \frac{1}{k^p}$, where $p \in \{0.3, 0.6, 0.9, 1.2, 1.5, 1.95\}$. Note that for these values one always has $p < 2q$. Further, we consider the case when there is no Tikhonov regularization, that is the case of a Nesterov type algorithm by taking $\varepsilon_k \equiv 0$ and $c_k \equiv 0$. In order to show the theoretical rates obtained in Theorem 5.3.1 for the discrete velocity $\|x_k - x_{k-1}\|$ and the potential energy $\Phi(x_k) - \min \Phi$ we also represent the values $\frac{1}{k}$ and $\frac{\Phi(x_1)}{k^2}$, respectively.

So, we run scheme (5.32) for 20 iterations, the results are shown in Figure 5.1.

Observe that indeed our algorithm has a similar (even better) behavior as the Nesterov type algorithm, the convergence rates for the discrete velocity and potential energy are of order $o(k^{-1})$ and $O(k^{-2})$, respectively. Further, the presence of the Tikhonov regularization terms do not affect the optimal rates, even more, while we increase p these rates become better.

In our second experiment for $a = 1$, $b = 5$ and $q_k = k^{0.8}$ we show the influence of the Tikhonov regularization terms $\varepsilon_k y_k$ and $c_k x_k$ on the behavior of the iterates of the algorithm. In the next figures we represent the first component of the iterates x_k with red, whereas the second component will be represented with blue.

First, we take $\varepsilon_k = k^{-\frac{3}{2}}$, (and the corresponding c_k), in order to show convergence to the minimum norm minimizer and also we analyze what happens if we renounce to both Tikhonov regularization terms. According to Figure 5.2(a) in the presence of the Tikhonov regularization terms $c_k x_k$ and $\varepsilon_k y_k$ the generated sequence x_k converges to

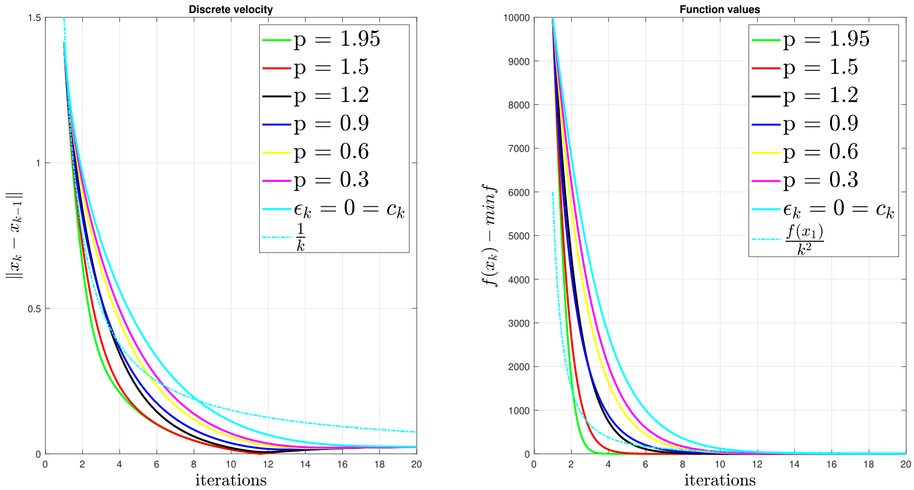
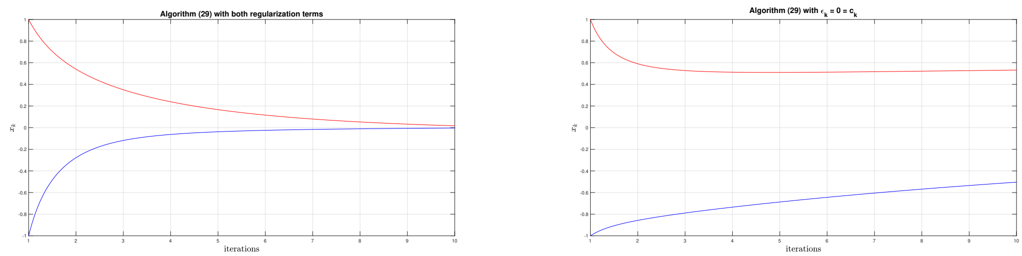


Figure 5.1: Different choices of ε_k for $q_k = k^{0.99}$

$(0,0)$, the minimum norm minimizer of Φ. In contrast, if we take $c_k \equiv 0$, $\varepsilon_k \equiv 0$, according to Figure 5.2(b), x_k does not converge to the minimum norm minimizer.



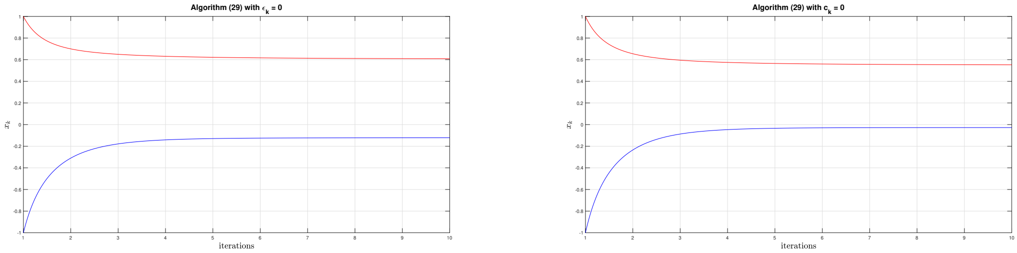
(a) With Tikhonov regularization

(b) Without Tikhonov regularization

Figure 5.2: If we drop the Tikhonov regularization terms in scheme (5.32) we do not have convergence to the minimum norm solution anymore.

Next we show that in order to have convergence to the minimum norm minimizer the presence of both Tikhonov regularization terms are essential. Note that in case $\varepsilon_k \equiv 0$ the parameter c_k in scheme (5.32) becomes $c_k = \frac{2s^2}{(k-1)^q k^q}$, hence the term $c_k x_k$ in the formulation of y_k still has the role of a Tikhonov regularization term. We represent this case at Figure 5.3(a). Further, we consider the case $c_k \equiv 0$, but $\varepsilon_k = k^{-\frac{3}{2}}$, see Figure 5.3(b).

As we can see, in the absence of one of the Tikhonov regularization terms we do not have convergence to the minimum norm minimizer of the objective function Φ. Hence, according to the last two figures the presence of two Tikhonov regularization terms in



(a) The absence of the term $\varepsilon_k y_k$ (b) The absence of the term $c_k x_k$

Figure 5.3: Dropping one of the Tikhonov regularization terms in scheme (5.32) we do not have convergence to the minimum norm solution anymore.

our algorithm is fully justified.

6 Conclusion

In this thesis, several topics related to convex optimization for both smooth and nonsmooth objective functions, via discrete and continuous time methods, featuring time scaling and Tikhonov regularization, were discussed. In the second chapter we delved into the topic of nonsmooth optimization featuring the time scaling. The main benefit lies in accelerating the rates of convergence of the function values while also providing the weak convergence of the generated trajectories. We approached the classical problem of minimizing a proper, convex and lower semicontinuous function via Moreau approximation with time-dependent parameter. Naturally, we linked the gradient of the Moreau envelope of the objective to the second order in time differential equation and used Lyapunov analysis to derive the desired results. This chapter is based on the paper [33] together with my supervisor, Professor R. I. Boţ.

One of the many possible continuations of this research was to consider applying Tikhonov regularization technique, which was done in Chapter 3. The idea was not only to improve the weak convergence of the trajectories to the strong one, but also to control this convergence, meaning that we will know exactly to which element of the Hilbert space the convergence takes place – to the element of the minimal norm from the set of all minimizers of the objective function. Another important goal was to preserve the accelerated (thanks to the presence of time scaling coefficient) rates of convergence of the function values. The technique follows the same ideas of Lyapunov analysis from the previous chapter. In the end, we were able to obtain two different settings: one for the strong controlled convergence and the other one – for the fast minimization rates. The results of this chapter were presented in the paper [35] together with Dr. habil. E. R. Csetnek.

The crucial disadvantage of the approach which was developed in Chapter 3 is the necessity to split the results into two settings: one cannot have both major results at the same time. This issue was addressed in the fourth chapter, where a new approach providing one common setting for both strong convergence of the trajectories and fast convergence of function values was developed. It should be mentioned that the time scaling coefficient no longer enters the system (4.1), but it is definitely possible to extend the analysis to cover this case. The paper [41] served as an inspiration for writing this chapter.

Last, but not least, we addressed the discrete counterpart of the discussed problems in Chapter 5, where we considered the inertial algorithm featuring double Tikhonov regularization. The goals were the same: to obtain strong convergence of the iterates to the minimal norm solution, as well as fast rates of convergence of the function to its global minimum. Unlike other similar results in the literature, the presented algorithm is not a proximal one, being thus the explicit discretization of the appropriate dynamical system. Moreover, for the original Nesterov algorithm the convergence of the generated sequences is still an open problem, so the strong convergence result for the Nesterov type algorithm could be seen as a big step forward, being, to our best knowledge, the first result of such type in the literature. The chapter was based on the paper [42] together with Professor S. C. László.

All in all, in this thesis several open questions were addressed, and some interesting particular cases were discussed, but what is more important, it gave me many ideas for the further research and helped establish many connections with prominent researchers for future fruitful collaborations.

Bibliography

[1] H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, Lecture Notes in Economics and Mathematical Systems 481, 25–35, 2000.

[2] H. Attouch, F. Alvarez, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9, 3–11, 2001.

[3] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping*, Application to optimization and mechanics, Journal de Mathématiques Pures et Appliquées 81(8), 747–779, 2002.

[4] H. Attouch, B. Abbas, B. F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and Applications 161(2), 331-360, 2014.

[5] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, *Fast convex optimization via inertial dynamics combining viscous and Hessian-driven damping with time rescaling*, Evolution Equations & Control Theory 11(2), 487-514, 2022.

[6] H. Attouch, A. Cabot, *Convergence of damped inertial dynamics governed by regularized maximally monotone operators*, Journal of Differential Equations 264, 7138–7182, 2018.

[7] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *First-order optimization algorithms via inertial systems with Hessian driven damping*, Mathematical Programming 193, 113-155, 2022.

[8] H. Attouch, Z. Chbani, J. Fadili, H. Riahi, *Convergence of iterates for first-order optimization algorithms with inertia and Hessian driven damping*, Optimization, DOI: 10.1080/02331934.2021.2009828, 2022.

[9] H. Attouch, Z. Chbani, H. Riahi, *Fast convex optimization via time scaling of damped inertial gradient dynamics*, SIAM Journal on Optimization 29(3), 2227-2256, 2019.

[10] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Mathematical Programming 168, 123-175, 2018.

[11] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function by asymptotical analysis of a dissipative dynamical system*, Communications in Contemporary Mathematics 2(1), 1-34, 2000.

[12] H. Attouch, S. C. László, *Continuous Newton-like inertial dynamics for monotone inclusions*, Set-Valued and Variational Analysis 29, 555–581, 2021.

[13] H. Attouch, J. Peypouquet, *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Mathematical Programming 174(1-2), 391–432, 2019.

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

[15] H. Attouch, J. Peypouquet, P. Redont, *Fast convex optimization via inertial dynamics with Hessian driven damping damping*, Journal of Differential Equations 261(10), 5734-5783, 2016.

[16] H. Attouch, *Viscosity solutions of minimization problems*, SIAM Journal of Optimization 6(3), 769-806, 1996.

[17] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, *Damped inertial dynamics with vanishing Tikhonov regularization: strong asymptotic convergence towards the minimum norm solution*, Journal of Differential Equations 311, 29-58, 2022.

[18] H. Attouch, A. Balhag, Z. Chbani, H. Riahi, *Accelerated gradient methods combining Tikhonov regularization with geometric damping driven by the Hessian*, Applied Mathematics and Optimization 88(29), 2023.

[19] H. Attouch, Z. Chbani, H. Riahi, *Combining fast inertial dynamics for convex optimization with Tikhonov regularization*, Journal of Mathematical Analysis and Applications 457(2), 1065-1094, 2018.

[20] H. Attouch, Z. Chbani, H. Riahi, *Fast Proximal Methods via Time Scaling of Damped Inertial Dynamics*, SIAM Journal on Optimization 29(3), 2227-2256, 2019.

[21] H. Attouch, Z. Chbani, H. Riahi, *Accelerated gradient methods with strong convergence to the minimum norm minimizer: a dynamic approach combining time scaling, averaging, and Tikhonov regularization*, arXiv:2211.10140v1, 2022.

[22] H. Attouch, R. Cominetti, *A dynamical approach to convex minimization coupling approximation with the steepest descent method*, Journal of Differential Equations 128(2), 519-540, 1996.

[23] H. Attouch, M-O. Czarnecki, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria*, Journal of Differential Equations 179, 278-310, 2002.

[24] H. Attouch, S. C. László, *Convex optimization via inertial algorithms with vanishing Tikhonov regularization: fast convergence to the minimum norm solution*, arXiv:2104.11987, 2021.

[25] H. Attouch, J. Peypouquet, *Convergence of the inertial dynamics and proximal algorithms governed by maximally monotone operators*, Mathematical Programming 174, 391–432, 2019.

[26] C.D. Alecsa, S.C. László, *Tikhonov regularization of a perturbed heavy ball system with vanishing damping*, SIAM Journal on Optimization 31(4), 2921-2954, 2021.

[27] H. Attouch, L.M. Briceño-Arias, P.L. Combettes, *A strongly convergent primal-dual method for nonoverlapping domain decomposition*, Numerische Mathematik 133(3), 443-470, 2016.

[28] H. Attouch, A. Cabot, Z. Chbani, H. Riahi, *Inertial Forward–Backward Algorithms with Perturbations: Application to Tikhonov Regularization*, Journal of Optimization: Theory and Applications 179, 1–36, 2018.

[29] R.I. Boţ, S.M. Grad, D. Meier, M. Staudigl, *Inducing strong convergence of trajectories in dynamical systems associated to monotone inclusions with composite structure*, Advanced Nonlinear Analysis 10, 450–476, 2021.

[30] R. I. Boţ, E. R. Csetnek, S. C. László, *Tikhonov regularization of a second order dynamical system with Hessian driven damping*, Mathematical Programming 189, 151–186, 2021.

[31] H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, 2016.

[32] R. I. Boţ, E.R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM Journal on Control and Optimization 54(3), 1423-1443, 2016.

[33] R. I. Boţ, M.A. Karapetyants, *A fast continuous time approach with time scaling for nonsmooth convex optimization*, Advances in Continuous and Discrete Models: Theory and Applications 73, 2022.

[34] R. I. Boţ, E. R. Csetnek, S. C. László, *On the strong convergence of continuous Newton-like inertial dynamics with Tikhonov regularization for monotone inclusions*, Journal of Mathematical Analysis and Applications 530(2), 2023.

[35] R.E. Csetnek, M.A. Karapetyants, *A fast continuous time approach for nonsmooth convex optimization with time scaling and Tikhonov regularization*, Accepted for publication in Journal of Optimization Theory and Applications, 2024.

[36] A. Cabot, H. Engler, S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation and insights*, Transactions of the American Mathematical Society 361, 5983–6017, 2009.

[37] A. Cabot, H. Engler, S. Gadat, *Second order differential equations with asymptotically small dissipation and piecewise flat potentials*, Electronic Journal of Differential Equations 17, 33–38, 2009.

[38] A. Chambolle, C. Dossal, *On the convergence of the iterates of the "fast iterative shrinkage/ thresholding algorithm"*, Journal of Optimization: Theory and Applications 166(3), 968–982, 2015.

[39] R. Cominetti, J. Peypouquet, S. Sorin, *Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization*, Journal of Differential Equations 245, 3753-3763, 2008.

[40] M.A. Jendoubi, R. May, *On an asymptotically autonomous system with Tikhonov type regularizing term*, Archiv der Mathematik 95(4), 389-399, 2010.

[41] M. Karapetyants, *A fast continuous time approach for nonsmooth convex optimization using Tikhonov regularization technique*, Computational Optimization and Applications 87, 531-569, 2024.

[42] M. Karapetyants, S.C. László, *A Nesterov type algorithm with double Tikhonov regularization: fast convergence of the function values and strong convergence to the minimal norm solution*, Applied Mathematics and Optimization 90(17), 2024.

[43] S.C. László, *Convergence rates for an inertial algorithm of gradient type associated to a smooth nonconvex minimization*, Mathematical Programming 190, 285–329, 2021.

[44] S.C. László, *On the convergence of an inertial proximal algorithm with a Tikhonov regularization term*, preprint, 2023.

[45] S. C. László, *On the strong convergence of the trajectories of a Tikhonov regularized second order dynamical system with asymptotically vanishing damping*, Journal of Differential Equations 362, 355-381, 2023.

[46] S.C. László, *Solving convex optimization problems via a second order dynamical system with implicit Hessian damping and Tikhonov regularization*, preprint, 2024.

[47] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Mathematics 41(3), 681-685, 2017.

[48] Y. Nesterov, *A method for solving the convex programming problem with convergence rate* $O\left(\frac{1}{k^2}\right)$, Doklady Akademii Nauk SSSR 269(3), 543-547, 1983.

[49] Y. Nesterov, *Introductory lectures on convex optimization: a basic course*, Kluwer Academic Publishers, Dordrecht, 2004.

[50] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bulletin of the American Mathematical Society 73(4), 591-597, 1967.

[51] B. T. Polyak, *Some methods of speeding up the convergence of iterative methods*, USSR Computational Mathematics and Mathematical Physics 4(5), 1-17, 1964.

[52] W. Su, S. Boyd, E.J. Candès, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, Journal of Machine Learning Research 17, 1-43, 2016.

[53] G. R. Sell, Dynamics of Evolutionary Equations, Springer, New York, 2002.

[54] B. Shi, S.S. Du, M.I. Jordan, Weijie J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, preprint, 2018.

[55] A.N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Doklady Akademii Nauk SSSR 151, 501-504, 1963.

[56] A.N. Tikhonov, V.Y. Arsenin, *Solutions of ill-posed problems*, Winston, New York, 1977.