Pure and Applied Functional Analysis
Volume 6, Number 6, 2021, 1081–1117

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

FAST CONVEX OPTIMIZATION VIA TIME SCALING OF DAMPED INERTIAL GRADIENT DYNAMICS

HEDY ATTOUCH, ZAKI CHBANI, AND HASSAN RIAHI

Abstract. In a Hilbert space setting, in order to develop fast first-order methods for convex optimization, we study the asymptotic convergence properties ($t \to +\infty$) of the trajectories of the inertial dynamics

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

The function Φ to minimize is assumed to be convex, continuously differentiable, $\gamma(t)$ is a positive damping coefficient, and $\beta(t)$ is a time scale coefficient. Convergence rates for the values $\Phi(x(t)) - \min_{\mathcal{H}} \Phi$ and the velocities are obtained under conditions involving only $\beta(t)$ and $\gamma(t)$. In this general framework (Φ is only assumed to be convex with a non-empty solution set), the fast convergence property is closely related to the asymptotic vanishing property $\gamma(t) \to 0$, and to the temporal scaling $\beta(t) \to +\infty$. We show the optimality of the convergence rates thus obtained, and study their stability under external perturbation of the system. The discrete time versions of the results provide convergence rates for a large class of inertial proximal algorithms. In doing so, we encompass the classical results on the subject, including Güler's inertial proximal algorithm.

1. Introduction

Unless specified, throughout the paper we make the following standing assumptions

$$\begin{cases} \mathcal{H} \text{ is a real Hilbert space;} \\ \Phi : \mathcal{H} \to \mathbb{R} \text{ is a convex function of class } \mathcal{C}^1,\ S = \operatorname{argmin} \Phi \neq \emptyset; \\ t_0 > 0 \text{ is the origin of time;} \\ \gamma \text{ and } \beta : [t_0, +\infty[\to \mathbb{R}_+ \text{ are non-negative continuous functions.} \end{cases}$$

We are interested in solving by (fast) first-order methods the convex optimization problem

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

To deal with large-scale problems, first-order methods have become very popular in recent decades. Their direct link with gradient dynamics allows them to be approached with powerful tools from various fields such as mechanics, control theory,

---

2020 *Mathematics Subject Classification.* 37N40, 46N10, 49M30, 65K05, 65K10, 90B50, 90C25; Secondary: 37N40, 46N10, 49M30, 65K05, 65K10, 90B50, 90C25.

Key words and phrases. Convex optimization, inertial gradient system, inertial proximal algorithms, Lyapunov analysis, Nesterov accelerated gradient method, time rescaling.

differential geometry and algebraic geometry. With this respect, the object of our study is the damped Inertial Gradient System

$$(\text{IGS})_{\gamma,\beta} \qquad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla\Phi(x(t)) = 0,$$

where $\nabla\Phi : \mathcal{H} \to \mathcal{H}$ is the gradient Φ. $(\text{IGS})_{\gamma,\beta}$ involves two time-dependent parameters: $\gamma(t)$ is a positive viscous damping coefficient, and $\beta(t)$ is a time scale coefficient. The optimization property of this system comes from the viscous friction term $\gamma(t)\dot{x}(t)$ which dissipates the global energy (potential + kinetic) of the system. Linking the tuning of the parameters $\gamma(t)$ and $\beta(t)$ to the corresponding convergence rate of the values $\Phi(x(t)) - \min_{\mathcal{H}} \Phi$ is a subtle question. We will answer this question in a general context, and obtain at the same time parallel results for the associated proximal algorithms. We will take advantage of the fact that proximal algorithms (obtained by implicit time discretization) usually inherit the properties of the continuous dynamics from which they come. Let us successively examine the role of the parameters $\gamma(t)$ and $\beta(t)$.

1.1. **The damping parameter $\gamma(t)$.** Let's review some historical facts about the choice of $\gamma(t)$ in

$$(\text{IGS})_{\gamma,1} \qquad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla\Phi(x(t)) = 0,$$

in relation to convex minimization. B. Polyak initiated the use of inertial dynamics to accelerate the gradient method. In [34] (1964), he introduced the Heavy Ball with Friction system

$$(\text{HBF}) \qquad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla\Phi(x(t)) = 0,$$

which has a fixed viscous damping coefficient $\gamma > 0$ (the mass has been normalized equal to 1). In a seminal paper [1] (2000), for a general convex function Φ, Alvarez proved that each trajectory of (HBF) converges weakly to a minimizer of Φ. For a strongly convex function Φ, (HBF) provides the convergence of $\Phi(x(t))$ to $\min_{\mathcal{H}} \Phi$ at an exponential rate by taking γ equal to $2\sqrt{\mu}$, where μ is the modulus of strong convexity of Φ. For general convex functions, the asymptotic convergence rate of (HBF) is $\mathcal{O}(\frac{1}{t})$ (in the worst case). It's not better than the steepest descent. A decisive step was taken in 2014 by Su-Boyd-Candès [39] who introduced the inertial system

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

For general convex functions, it provides a continuous version of the accelerated gradient method of Nesterov [28, 29] (described a little further). For $\alpha \geq 3$, each trajectory $x(\cdot)$ of $(\text{AVD})_{\alpha}$ satisfies the asymptotic convergence rate of the values $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^2}\right)$. As a specific feature, the viscous damping coefficient $\frac{\alpha}{t}$ vanishes (tends to zero) as time t goes to infinity, hence the terminology. Its close relationship with the Nesterov accelerated gradient method makes $(\text{AVD})_{\alpha}$ an interesting dynamic for convex optimization. As such, it has been the subject of many recent studies, see [3], [5, 6], [7], [8], [9], [13], [15], [16], [19], [27], [39]. They allow to better understand Nesterov's method. Let's explain why $\frac{\alpha}{t}$ is a clever tuning of the damping coefficient.

i) In [20], Cabot-Engler-Gaddat showed that the damping coefficient $\gamma(\cdot)$ in $(\text{IGS})_{\gamma,1}$ dissipates the energy, and hence makes the dynamic interesting for optimization, as long as $\int_{t_0}^{+\infty} \gamma(t)dt = +\infty$. The damping coefficient can go to zero asymptotically but not too fast. The smallest which is admissible is of order $\frac{1}{t}$. It is the one which enforces the most the inertial effect with respect to the friction effect.

ii) The tuning of the parameter α in front of $\frac{1}{t}$ comes naturally from the Lyapunov analysis and the optimality of the convergence rates obtained. Indeed, the case $\alpha = 3$, which corresponds to Nesterov's historical algorithm, turns out to be critical. In the case $\alpha = 3$, the question of the convergence of the trajectories remains an open problem (except in one dimension where convergence holds [9]). As a remarkable property, for $\alpha > 3$, it has been shown by Attouch-Chbani-Peypouquet-Redont [8] and May [27] that each trajectory converges weakly to a minimizer. Corresponding results for the algorithmic case have been obtained in [21] and [13]. Moreover for $\alpha > 3$, it is shown in [13] and [27] that the asymptotic convergence rate of the values is $o(\frac{1}{t^2})$. The subcritical case $\alpha < 3$ has been examined by Apidopoulos-Aujol-Dossal [3] and Attouch-Chbani-Riahi [9], with the convergence rate of the values $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$.

A unifying view on the subject, dealing with the case of a general damping coefficient $\gamma(t)$ in $(\text{IGS})_{\gamma,1}$, has been developed by Attouch-Cabot [5] and Attouch-Cabot-Chbani-Riahi [7].

1.2. **The time scaling parameter $\beta(t)$.** Let's illustrate the role of $\beta(t)$ in the following model situation. Start from the $(\text{AVD})_\alpha$ system with $\gamma(t) = \frac{\alpha}{t}$ and $\alpha \geq 3$. Given a trajectory $x(\cdot)$ of $(\text{AVD})_\alpha$, as explained above

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

Let's make the change of time variable in $(\text{AVD})_\alpha$: $t = s^p$, where p is a positive parameter. Set $y(s) := x(s^p)$. By the derivation chain rule, we have

$$\ddot{y}(s) + \frac{\alpha_p}{s}\dot{y}(s) + p^2 s^{2(p-1)} \nabla\Phi(y(s)) = 0, \tag{1.1}$$

where $\alpha_p = 1 + (\alpha - 1)p$. The convergence rate of values becomes

$$\Phi(y(s)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{s^{2p}}\right) \quad \text{as } s \to +\infty. \tag{1.2}$$

For $p > 1$, we have $\alpha_p > \alpha$, so the damping parameters for (1.1) are similar to those of $(\text{AVD})_\alpha$. The only major difference is the coefficient $s^{2(p-1)}$ in front of $\nabla\Phi(y(s))$, which explodes when $s \to +\infty$. From (1.2) we observe that the convergence rate of values can be made arbitrarily fast (in the scale of powers of $\frac{1}{s}$) with p large. The physical intuition is clear. Fast optimization is associated with the fast parameterization of the trajectories of the $(\text{AVD})_\alpha$ system. One of our objective is to transpose these results to the proximal algorithms [36], taking advantage of the fact that implicit discretization usually preserves the properties of the continuous dynamics. The case $\gamma(t) = \frac{\alpha}{t}$, and $\beta(t)$ general, i.e.,

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

has been analyzed by Attouch-Chbani-Riahi [10]. As explained above, the varying parameter $t \mapsto \beta(t)$ comes naturally with the time reparametrization of the above dynamics, and plays a key role in the acceleration of its asymptotic convergence properties (the key idea is to take $\beta(t) \to +\infty$ as $t \to +\infty$ in a controlled way). The importance of the time scaling to accelerate algorithms was also stressed in [38], [41].

1.3. **Linking inertial dynamics with first-order algorithms.** Let's first recall some classic results concerning the algorithms associated with the continuous steepest descent. Given $\lambda > 0$, the gradient descent method

$$x_{k+1} = x_k - \lambda\nabla\Phi(x_k)$$

and the proximal point method

$$x_{k+1} = \text{prox}_{\lambda\Phi}(x_k) = (I + \lambda\partial\Phi)^{-1}(x_k) = \text{argmin}_{\xi\in\mathcal{H}}\left\{\lambda\Phi(\xi) + \frac{1}{2}\|x - \xi\|^2\right\}$$

are the basic blocks of the first-order methods for convex optimization. By interpreting λ as a fixed time step, they can be respectively obtained as the forward (explicit) discretization of the continuous steepest descent

$$\dot{x}(t) + \lambda\nabla\Phi(x(t)) = 0, \tag{1.3}$$

and the backward (implicit) discretization of the differential inclusion

$$\dot{x}(t) + \lambda\partial\Phi(x(t)) \ni 0. \tag{1.4}$$

The gradient method goes back to Cauchy (1847). The proximal algorithm was first introduced by Martinet [26] (1970), and then developed by Rockafellar [36] who extended it to solve monotone inclusions. One can consult [17], [31], [32], [33], [35], for a recent account on the proximal methods, that play a central role in nonsmooth optimization as a basic block of many splitting algorithms.

Let's now come with second-order evolution equations, and illustrate their link with algorithms. Time discretization with a fixed time step $h > 0$ of $(\text{IGS})_{\gamma,\beta}$

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

gives

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \gamma(kh)\frac{1}{h}(x_k - x_{k-1}) + \beta(kh)\nabla\Phi(\xi_k) = 0. \tag{1.5}$$

Set $\alpha_k = 1 - h^2\gamma(kh)$ and $\beta_k = h^2\beta(kh)$. Following the choice of ξ_k, we obtain one of the following algorithms:

- **Implicit:** $\xi_k = x_{k+1}$ gives the Inertial Proximal algorithm (Beck-Teboulle [18], Güler [24]):

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{\beta_k\Phi}(y_k), \end{cases}$$

- **Nesterov choice:** $\xi_k = x_k + \alpha_k(x_k - x_{k-1})$ gives the Inertial Gradient algorithm:

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = y_k - \beta_k \nabla\Phi(y_k). \end{cases}$$

Nesterov's scheme corresponds to replacing the proximal step by a gradient step. This is illustrated below.

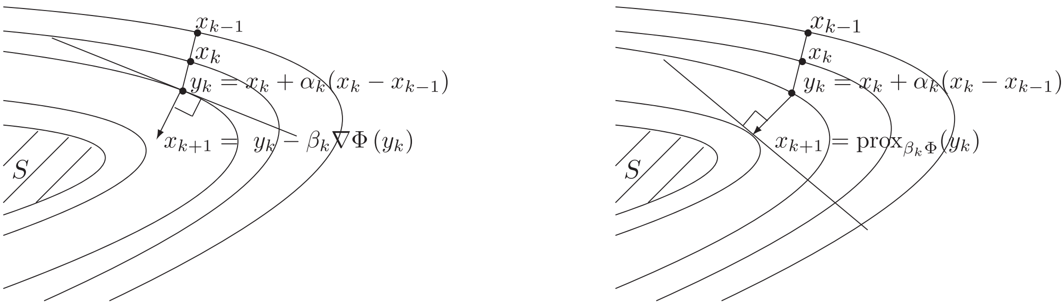

Figure 1. Inertial Gradient algorithm   Inertial Proximal algorithm.

For additive structured optimization, the combination of the two algorithms yields FISTA type proximal-gradient algorithms. Their study was initiated by Beck-Teboulle [18], see Attouch-Cabot [6] for recent developments.

1.4. **Presentation of the results.** We are interested in the joint setting of the parameters $\gamma(t)$ and $\beta(t)$ providing a fast convergence of the values $\Phi(x(t)) - \min\Phi$ as $t \to +\infty$, and its algorithmic counterpart. In addition to Güler's accelerated proximal algorithm, our dynamic approach with general damping and scaling coefficients gives rise to a whole family of proximal algorithms with fast convergence properties. We start with the study of the continuous dynamic (sections 2 to 4), then consider the algorithmic aspects. In section 2, based on Lyapunov analysis, we study the asymptotic behaviour of the trajectories for $(\mathrm{IGS})_{\gamma,\beta}$. According to [5], the description of our result uses the auxiliary function $\Gamma_\gamma : [t_0, +\infty[\to \mathbb{R}_+$ defined by

$$\Gamma_\gamma(t) = p(t)\int_t^{+\infty} \frac{du}{p(u)} \quad \text{where} \quad p(t) = e^{\int_{t_0}^t \gamma(u)\,du}.$$

In our main result, Theorem 2.1, we show that, under the following growth condition on $\beta(\cdot)$

$$(\mathrm{H})_{\gamma,\beta} \qquad \Gamma_\gamma(t)\dot\beta(t) \le \beta(t)\left(3 - 2\gamma(t)\Gamma_\gamma(t)\right),$$

we have the convergence rate of values

$$\Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \text{as } t \to +\infty.$$

In Theorems 2.5 and 4.2, we complete this result by showing that, under a slightly stronger growth condition on $\beta(\cdot)$, we can pass from the capital $\mathcal{O}$ to the small o estimate both for the function values and the velocities:

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \text{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{\Gamma_\gamma(t)}\right) \quad \text{as } t \to +\infty.$$

Some special cases for $\gamma(\cdot)$ and $\beta(\cdot)$ of particular interest are examined in section 3. In section 4, we consider a perturbed version of the initial evolution system $(\mathrm{IGS})_{\gamma,\beta}$. Assuming that the perturbation term g satisfies $\int_{t_0}^{+\infty} \Gamma_\gamma(t)\|g(t)\|dt < +\infty$ (which means that it vanishes fast enough), we obtain similar convergence rates of the values (Theorem 4.1) and weak convergence of the trajectories (Theorem 4.2). In section 5, we make the link with the Güler inertial proximal algorithm, and provide its continuous interpretation. In section 6 (Theorem 6.1), we study the fast convergence of values for a class of inexact inertial proximal algorithms, which takes into account a unified discretization of the damping term. We conclude with some Perspectives. In the Appendix section, we prove the existence and uniqueness of the global solution of the Cauchy problem associated with $(\mathrm{IGS})_{\gamma,\beta,e}$, and complete with some technical lemmas.

2. Convergence rates for the (IGS) system.

Based on the right tuning of the damping coefficient $\gamma(t)$ and of the scaling coefficient $\beta(t)$, we will analyze the convergence rate of the trajectories of the inertial dynamic $(\mathrm{IGS})_{\gamma,\beta}$

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

We take for granted the existence of solutions to this system. The existence and uniqueness of a classical global solution for the corresponding Cauchy problem is detailed in the Appendix, Theorem 8.2. In the elementary case $\Phi \equiv 0$, the direct integration of $(\mathrm{IGS})_{\gamma,\beta}$ makes appear the function

$$p(t) = e^{\int_{t_0}^t \gamma(u)\,du}. \tag{2.1}$$

In the sequel, we will systematically assume that the following condition $(\mathrm{H})_0$ is satisfied

$$(\mathrm{H})_0 \qquad \int_{t_0}^{+\infty} \frac{du}{p(u)} < +\infty.$$

Under this assumption, we can define the function $\Gamma_\gamma : [t_0, +\infty[\to \mathbb{R}_+$

$$\Gamma_\gamma(t) = p(t) \int_t^{+\infty} \frac{du}{p(u)}. \tag{2.2}$$

Note that the definition of Γ_γ does not depend on the choice of the origin of time t_0. The subscript γ highlights the close link between Γ_γ and the damping function γ.

The asymptotic properties of the dynamic system $(\mathrm{IGS})_{\gamma,\beta}$ are based on the behavior of the functions $t \mapsto \Gamma_\gamma(t)$ and $t \mapsto \beta(t)$ as $t \to +\infty$. By differentiating (2.2), we immediately obtain the differential relation

$$\dot{\Gamma}_\gamma(t) = \gamma(t)\Gamma_\gamma(t) - 1, \tag{2.3}$$

which plays a central role in Lyapunov's analysis. Let us define the rescaled global energy function

$$W(t) := \frac{1}{2}\|\dot{x}(t)\|^2 + \beta(t)\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right), \tag{2.4}$$

and the anchor function

$$h(t) := \frac{1}{2}\|x(t) - z\|^2,$$

where $z \in \operatorname{argmin} \Phi$ is given. They are the basic constitutive blocks of the function $\mathcal{E} : [t_0, +\infty[\to \mathbb{R}^+$

$$\mathcal{E}(t) := \Gamma_\gamma(t)^2 W(t) + h(t) + \Gamma_\gamma(t)\dot{h}(t) \tag{2.5}$$

that will serve for the Lyapunov analysis. We have

$$\begin{aligned}
\mathcal{E}(t) &= \Gamma_\gamma(t)^2 W(t) + h(t) + \Gamma_\gamma(t)\dot{h}(t) \\
&= \Gamma_\gamma(t)^2 \left(\frac{1}{2}\|\dot{x}(t)\|^2 + \beta(t)\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right)\right) \\
&+ \frac{1}{2}\|x(t) - z\|^2 + \Gamma_\gamma(t)\langle \dot{x}(t), x(t) - z\rangle,
\end{aligned}$$

which gives

$$\mathcal{E}(t) = \Gamma_\gamma(t)^2 \beta(t)\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) + \frac{1}{2}\|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\|^2. \tag{2.6}$$

Hence, $\mathcal{E}(\cdot)$ is a non-negative function.

2.1. **$\mathcal{O}$-rate of convergence for the values.** Based on the decreasing property of $\mathcal{E}(\cdot)$, we are going to prove the following theorem.

Theorem 2.1. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex function which is continuously differentiable and such that* $\operatorname{argmin} \Phi \neq \emptyset$. *Suppose that $\gamma : [t_0, +\infty[\to \mathbb{R}_+$ is a continuous function that satisfies* $(\mathrm{H})_0$. *Suppose that $\beta : [t_0, +\infty[\to \mathbb{R}_+$ is a positive continuous function. Assume that the following growth condition* $(\mathrm{H})_{\gamma,\beta}$ *linking $\gamma(t)$ with $\beta(t)$ is satisfied:*

$$\Gamma_\gamma(t)\dot{\beta}(t) \leq \beta(t)\left(3 - 2\gamma(t)\Gamma_\gamma(t)\right). \tag{$\mathrm{H}_{\gamma,\beta}$}$$

Then, for every solution trajectory $x : [t_0, +\infty[\to \mathcal{H}$ of

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

we have the following properties:

(i) The following convergence rate of the values is satisfied:

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \text{as } t \to +\infty. \tag{2.7}$$

Precisely, for all $t \geq t_0$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{\beta(t)\Gamma_\gamma(t)^2}, \tag{2.8}$$

with

$$C = \Gamma_\gamma(t_0)^2\beta(t_0)\left(\Phi(x(t_0)) - \min_{\mathcal{H}} \Phi\right) + d(x(t_0), \operatorname{argmin}\Phi)^2 + \Gamma_\gamma(t_0)^2\|\dot{x}(t_0)\|^2.$$

(ii) *The solution trajectory* $x(\cdot)$ *is bounded on* $[t_0, +\infty[$.

Proof. Set briefly $m := \min_{\mathcal{H}} \Phi$.
(i) Let's compute the time derivative of $\mathcal{E}(\cdot)$, as formulated in (2.5). We first compute the derivative of its main ingredients, namely, W and h. The classical derivation chain rule and $(\mathrm{IGS})_{\gamma,\beta}$ give

$$\begin{aligned}\dot{W}(t) &= \langle \dot{x}(t), \ddot{x}(t)\rangle + \beta(t)\langle \dot{x}(t), \nabla\Phi(x(t))\rangle + \dot{\beta}(t)\left(\Phi(x(t)) - m\right)\\ &= \langle \dot{x}(t), \ddot{x}(t) + \beta(t)\nabla\Phi(x(t))\rangle + \dot{\beta}(t)\left(\Phi(x(t)) - m\right)\\ &= -\gamma(t)\|\dot{x}(t)\|^2 + \dot{\beta}(t)\left(\Phi(x(t)) - m\right).\end{aligned} \tag{2.9}$$

On the other hand, $\dot{h}(t) = \langle \dot{x}(t), x(t) - z\rangle$ and $\ddot{h}(t) = \|\dot{x}(t)\|^2 + \langle \ddot{x}(t), x(t) - z\rangle$. It ensues that

$$\begin{aligned}\ddot{h}(t) + \gamma(t)\dot{h}(t) &= \|\dot{x}(t)\|^2 + \langle \ddot{x}(t) + \gamma(t)\dot{x}(t), x(t) - z\rangle\\ &= \|\dot{x}(t)\|^2 - \beta(t)\langle \nabla\Phi(x(t)), x(t) - z\rangle\\ &\leq \|\dot{x}(t)\|^2 - \beta(t)(\Phi(x(t)) - m) \leq \|\dot{x}(t)\|^2,\end{aligned} \tag{2.10}$$

where the above inequality follows from the convexity of Φ.
We have now all the ingredients to derivate $\mathcal{E}(\cdot)$, as defined in (2.5). Collecting the above results we obtain

$$\begin{aligned}\dot{\mathcal{E}}(t) &= \Gamma_\gamma(t)^2\dot{W}(t) + 2\Gamma_\gamma(t)\dot{\Gamma}_\gamma(t)W(t) + \dot{h}(t) + \dot{\Gamma}_\gamma(t)\dot{h}(t) + \Gamma_\gamma(t)\ddot{h}(t)\\ &= \Gamma_\gamma(t)^2[-\gamma(t)\|\dot{x}(t)\|^2 + \dot{\beta}(t)\left(\Phi(x(t)) - m\right)]\\ &+ 2\Gamma_\gamma(t)\dot{\Gamma}_\gamma(t)\left[\frac{1}{2}\|\dot{x}(t)\|^2 + \beta(t)\left(\Phi(x(t)) - m\right)\right]\\ &+ \dot{h}(t) + \dot{\Gamma}_\gamma(t)\dot{h}(t) + \Gamma_\gamma(t)\ddot{h}(t).\end{aligned}$$

According to $1 + \dot{\Gamma}_\gamma(t) = \gamma(t)\Gamma_\gamma(t)$ and (2.10), the above last line writes

$$\begin{aligned}\dot{h}(t) + \dot{\Gamma}_\gamma(t)\dot{h}(t) + \Gamma_\gamma(t)\ddot{h}(t) &= \Gamma_\gamma(t)\left(\ddot{h}(t) + \gamma(t)\dot{h}(t)\right)\\ &\leq \Gamma_\gamma(t)\left(\|\dot{x}(t)\|^2 - \beta(t)(\Phi(x(t)) - m)\right).\end{aligned}$$

Combining the above results, we obtain

$$\begin{aligned}\dot{\mathcal{E}}(t) &\leq \Gamma_\gamma(t)^2\left[-\gamma(t)\|\dot{x}(t)\|^2 + \dot{\beta}(t)\left(\Phi(x(t)) - m\right)\right]\\ &+ 2\Gamma_\gamma(t)\dot{\Gamma}_\gamma(t)\left[\frac{1}{2}\|\dot{x}(t)\|^2 + \beta(t)\left(\Phi(x(t)) - m\right)\right]\\ &\quad +\Gamma_\gamma(t)\left(\|\dot{x}(t)\|^2 - \beta(t)(\Phi(x(t)) - m)\right)\\ &\leq \Gamma_\gamma(t)\|\dot{x}(t)\|^2\left[1 + \dot{\Gamma}_\gamma(t) - \gamma(t)\Gamma_\gamma(t)\right]\\ &+ \left(\Phi(x(t)) - m\right)\Gamma_\gamma(t)\left(\Gamma_\gamma(t)\dot{\beta}(t) + 2\dot{\Gamma}_\gamma(t)\beta(t) - \beta(t)\right).\end{aligned}$$

Using again the relation $1+\dot{\Gamma}_\gamma(t) = \gamma(t)\Gamma_\gamma(t)$, we finally get

$$\dot{\mathcal{E}}(t) \le (\Phi(x(t)) - m)\,\Gamma_\gamma(t)\left(\Gamma_\gamma(t)\dot{\beta}(t) + 2\dot{\Gamma}_\gamma(t)\beta(t) - \beta(t)\right). \tag{2.11}$$

According to $1+\dot{\Gamma}_\gamma(t) = \gamma(t)\Gamma_\gamma(t)$, this is equivalent to

$$\dot{\mathcal{E}}(t) \le (\Phi(x(t)) - m)\,\Gamma_\gamma(t)\left(\Gamma_\gamma(t)\dot{\beta}(t) + \beta(t)(2\gamma(t)\Gamma_\gamma(t) - 3)\right). \tag{2.12}$$

Therefore, by assumption $(\mathrm{H})_{\gamma,\beta}$, we deduce that $\dot{\mathcal{E}}(t) \le 0$. Hence $\mathcal{E}(t) \le \mathcal{E}(t_0)$ on $[t_0, +\infty[$. According to the formulation (2.6) of $\mathcal{E}(t)$, we deduce that, for all $t \ge t_0$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \le \frac{\mathcal{E}(t_0)}{\beta(t)\Gamma_\gamma(t)^2}. \tag{2.13}$$

According to the definition of $\mathcal{E}(t_0)$, we have

$$\begin{aligned} \mathcal{E}(t_0) &= \Gamma_\gamma(t_0)^2\beta(t_0)\left(\Phi(x(t_0)) - \min_{\mathcal{H}}\Phi\right) + \frac{1}{2}\|x(t_0) - z + \Gamma_\gamma(t_0)\dot{x}(t_0)\|^2 \\ &\le \Gamma_\gamma(t_0)^2\beta(t_0)\left(\Phi(x(t_0)) - \min_{\mathcal{H}}\Phi\right) + \|x(t_0) - z\|^2 + \Gamma_\gamma(t_0)^2\|\dot{x}(t_0)\|^2. \end{aligned}$$

The estimate (2.13) being valid for any $z \in \operatorname{argmin}\Phi$, we finally obtain

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \le \frac{C}{\beta(t)\Gamma_\gamma(t)^2}, \tag{2.14}$$

with

$$C = \Gamma_\gamma(t_0)^2\beta(t_0)\left(\Phi(x(t_0)) - \min_{\mathcal{H}}\Phi\right) + d(x(t_0), \operatorname{argmin}\Phi)^2 + \Gamma_\gamma(t_0)^2\|\dot{x}(t_0)\|^2.$$

(ii) Let us now prove that the solution $x(\cdot)$ of $(\mathrm{IGS})_{\gamma,\beta}$ is bounded on $[t_0, +\infty[$. According to (2.6), and $\mathcal{E}(\cdot)$ decreasing, we have

$$\|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\|^2 \le 2\mathcal{E}(t) \le 2\mathcal{E}(t_0).$$

After developing the above inequality, we obtain

$$\|x(t) - z\|^2 + 2\Gamma_\gamma(t)\,\langle x(t) - z, \dot{x}(t)\rangle \le 2\mathcal{E}(t_0). \tag{2.15}$$

Set $h(t) := \frac{1}{2}\|x(t) - z\|^2$, and $q(t) := \int_t^{+\infty} \dfrac{ds}{p(s)}$. We have $\{q(t) : t \ge t_0\}$ is bounded since $\int_{t_0}^{+\infty} \frac{ds}{p(s)} < +\infty$. After dividing (2.15) by $p(t)$, and noticing that $\frac{\Gamma_\gamma(t)}{p(t)} = q(t)$, we obtain, with $C = \mathcal{E}(t_0)$

$$\frac{1}{p(t)}h(t) + q(t)\dot{h}(t) \le \frac{C}{p(t)}, \quad \forall t \in [t_0, +\infty[.$$

Since $\dot{q}(t) = -\dfrac{1}{p(t)}$, we equivalently have $q(t)\dot{h}(t) - \dot{q}(t)(h(t) - C) \le 0$. After dividing by $q(t)^2$, we obtain

$$\frac{d}{dt}\left(\frac{h(t) - C}{q(t)}\right) = \frac{1}{q(t)^2}\left(q(t)\dot{h}(t) - \dot{q}(t)(h(t) - C)\right) \le 0.$$

Integration of this inequality gives $h(t) \le C_1(1 + q(t))$ for some $C_1 > 0$. Therefore, $x(\cdot)$ is bounded. □

Remark 2.2. Let us analyze the condition $(\mathrm{H})_{\gamma,\beta}$. We return to its formulation (2.16)

$$\Gamma_\gamma(t)\dot{\beta}(t) + 2\dot{\Gamma}_\gamma(t)\beta(t) - \beta(t) \leq 0. \tag{2.16}$$

After multiplication by $\Gamma_\gamma(t)$, and by setting $\xi(t) = \beta(t)\Gamma_\gamma^2(t)$ it writes equivalently as

$$\dot{\xi}(t) - \frac{1}{\Gamma_\gamma(t)}\xi(t) \leq 0. \tag{2.17}$$

Integration of this first-order differential equation immediately gives

$$\xi(t) \leq \xi(t_0)\exp\left(\int_{t_0}^t \frac{1}{\Gamma_\gamma(s)}ds\right) \text{ on } [t_0, +\infty[. \tag{2.18}$$

Therefore

$$\beta(t) \leq \beta(t_0)\frac{\Gamma_\gamma^2(t_0)}{\Gamma_\gamma^2(t)}\exp\left(\int_{t_0}^t \frac{1}{\Gamma_\gamma(u)}du\right) \text{ on } [t_0, +\infty[. \tag{2.19}$$

Thus, the condition $(\mathrm{H})_{\gamma,\beta}$ imposes a growth limitation for $\beta(\cdot)$, which depends on $\gamma(\cdot)$.

Remark 2.3. The assumption $\operatorname{argmin}\Phi \neq \emptyset$ is crucial to guarantee that the trajectory remains bounded. Otherwise when $\beta \equiv 1$, Φ is minorized and does not attain its infimum, i.e., $\operatorname{argmin}\Phi = \emptyset$, we may have $\lim_{t\to+\infty}\Phi(x(t)) = \inf_{\mathcal{H}}\Phi$ (see [8]). From this, we easily deduce that $\lim_{t\to+\infty}\|x(t)\| = +\infty$.

2.2. **Rate of decay of the global energy.** To obtain fast convergence of velocities to zero, we need to introduce the following slightly strengthened condition $(\mathrm{H})^+_{\gamma,\beta}$: there exist $t_1 \geq t_0$ and $\rho > 0$ such that for all $t \geq t_1$

$$\Gamma_\gamma(t)\dot{\beta}(t) \leq \beta(t)\Big(3 - \rho - 2\gamma(t)\Gamma_\gamma(t)\Big). \tag{$\mathrm{H})^+_{\gamma,\beta}$$

Note that $(\mathrm{H})_{\gamma,\beta}$ corresponds to the case $\rho = 0$ in $(\mathrm{H})^+_{\gamma,\beta}$.
Let's first establish an integral estimate for the rescaled global energy function W.

Proposition 2.4. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex function of class $\mathcal{C}^1$ such that* $\operatorname{argmin}\Phi \neq \emptyset$. *Suppose that the property* $(\mathrm{H})^+_{\gamma,\beta}$ *is satisfied. Then, for any solution* $x(\cdot)$ *of* $(\mathrm{IGS})_{\gamma,\beta}$, *the following integral energy estimate holds*

$$\int_{t_0}^{+\infty} \Gamma_\gamma(t)W(t)dt < +\infty, \tag{2.20}$$

where

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

Equivalently

$$\int_{t_0}^{+\infty} \Gamma_\gamma(t)\|\dot{x}(t)\|^2 dt < +\infty \quad \textit{and} \quad \int_{t_0}^{+\infty} \Gamma_\gamma(t)\beta(t)\left(\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right)dt < +\infty.$$

Proof. Set $m := \min_{\mathcal{H}} \Phi$. Start from the energy estimate (2.9). After multiplication by the positive scalar $\Gamma_\gamma(t)^2$, we get

$$\Gamma_\gamma(t)^2 \dot{W}(t) + \gamma(t)\Gamma_\gamma(t)^2\|\dot{x}(t)\|^2 = \dot{\beta}(t)\Gamma_\gamma(t)^2 \left(\Phi(x(t)) - m\right).$$

After integration by parts on (t_0, t), we obtain

$$\Gamma_\gamma(t)^2W(t) - \Gamma_\gamma(t_0)^2W(t_0) - 2\int_{t_0}^t \Gamma_\gamma(s)\dot{\Gamma}_\gamma(s)W(s)ds + \int_{t_0}^t \gamma(s)\Gamma_\gamma(s)^2\|\dot{x}(s)\|^2 ds$$
$$= \int_{t_0}^t \dot{\beta}(s)\Gamma_\gamma(s)^2 \left(\Phi(x(s)) - m\right) ds.$$

Replace $W(s)$ by its formulation $W(s) := \frac{1}{2}\|\dot{x}(s)\|^2 + \beta(s)\left(\Phi(x(s)) - \min_{\mathcal{H}} \Phi\right)$ in the third term of the left member of the above formula. After simplification, we get

$$\Gamma_\gamma(t)^2W(t) + \int_{t_0}^t \Gamma_\gamma(s)[\Gamma_\gamma(s)\gamma(s) - \dot{\Gamma}_\gamma(s)]\|\dot{x}(s)\|^2 ds$$
$$= \Gamma_\gamma(t_0)^2W(t_0) + \int_{t_0}^t [2\Gamma_\gamma(s)\dot{\Gamma}_\gamma(s)\beta(s) + \dot{\beta}(s)\Gamma_\gamma(s)^2]\left(\Phi(x(s)) - m\right) ds.$$

By (2.3), we have $\Gamma_\gamma(s)\gamma(s) - \dot{\Gamma}_\gamma(s) = 1$.
By $(\mathrm{H})_{\gamma,\beta}$ and (2.3), we have $\frac{d}{ds}\left(\Gamma_\gamma(s)^2\beta(s)\right) = 2\Gamma_\gamma(s)\dot{\Gamma}_\gamma(s)\beta(s) + \dot{\beta}(s)\Gamma_\gamma(s)^2 \leq \Gamma_\gamma(s)\beta(s)$.
Collecting the above results, we obtain
(2.21)

$$\Gamma_\gamma(t)^2W(t) + \int_{t_0}^t \Gamma_\gamma(s)\|\dot{x}(s)\|^2 ds \quad \leq \quad \int_{t_0}^t \Gamma_\gamma(s)\beta(s)\left(\Phi(x(s)) - m\right) ds + \Gamma_\gamma(t_0)^2W(t_0).$$

In order to estimate the second member of (2.21), we return to (2.12)

$$\dot{\mathcal{E}}(t) \leq \Gamma_\gamma(t)\left(\Gamma_\gamma(t)\dot{\beta}(t) + \beta(t)(2\gamma(t)\Gamma_\gamma(t) - 3)\right)\left(\Phi(x(t)) - m\right).$$

According to the property $(\mathrm{H})^+_{\gamma,\beta}$, that we rewrite as follows

$$\Gamma_\gamma(t)\dot{\beta}(t) + \beta(t)\Big(2\gamma(t)\Gamma_\gamma(t) - 3\Big) \leq -\rho\beta(t), \tag{2.22}$$

we get

$$\dot{\mathcal{E}}(t) + \rho\Gamma_\gamma(t)\beta(t)\left(\Phi(x(t)) - m\right) \leq 0.$$

Integrating this inequality, and using that $\mathcal{E}(t)$ is non-negative, we obtain

$$\int_{t_0}^{+\infty} \Gamma_\gamma(s)\beta(s)\left(\Phi(x(s)) - m\right) ds \leq \frac{\mathcal{E}(t_0)}{\rho} < +\infty. \tag{2.23}$$

Combining this inequality with (2.21), we obtain

$$\Gamma_\gamma(t)^2W(t) + \int_{t_0}^t \Gamma_\gamma(s)\|\dot{x}(s)\|^2 ds \quad \leq \quad \frac{\mathcal{E}(t_0)}{\rho} + \Gamma_\gamma(t_0)^2W(t_0). \tag{2.24}$$

Hence

$$\int_{t_0}^{+\infty} \Gamma_\gamma(s)\|\dot{x}(s)\|^2 ds \quad \leq \quad \frac{\mathcal{E}(t_0)}{\rho} + \Gamma_\gamma(t_0)^2W(t_0). \tag{2.25}$$

By adding the inequalities (2.23) and (2.25), and using the definition of W, we finally get

$$\int_{t_0}^{+\infty} \Gamma_\gamma(s) W(s) ds \leq \frac{3}{2}\frac{\mathcal{E}(t_0)}{\rho} + \frac{1}{2}\Gamma_\gamma(t_0)^2 W(t_0),$$

which gives (2.20). $\square$

2.3. **From $\mathcal{O}$ to o convergence rate of the values.** According to Attouch-Peypouquet [13] and Attouch-Cabot [5], let us improve the convergence rates obtained in the previous sections by passing from capital $\mathcal{O}$ to small o estimates.

Theorem 2.5. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex function of class $\mathcal{C}^1$ such that* $\operatorname{argmin} \Phi \neq \emptyset$. *Suppose that the property* $(\mathrm{H})^+_{\gamma,\beta}$ *is satisfied.*

(i) *If* $\displaystyle\int_{t_0}^{+\infty} \frac{1}{\Gamma_\gamma(t)} dt = +\infty$, *then, as* $t \to +\infty$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \textit{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{\Gamma_\gamma(t)}\right).$$

(ii) *If* $\displaystyle\int_{t_0}^{+\infty} \beta(t)\Gamma_\gamma(t) dt = +\infty$ *and* $2\gamma(t)\beta(t) + \dot{\beta}(t)$ *is non-negative on* $[t_0, +\infty[$, *then, as* $t \to +\infty$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds}\right) \quad \textit{and} \quad \|\dot{x}(t)\|^2 = o\left(\frac{\beta(t)}{\int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds}\right).$$

Proof. We examine successively the two cases:

$$\int_{t_0}^{+\infty} \frac{1}{\Gamma_\gamma(t)} dt = +\infty \quad \text{and} \quad \int_{t_0}^{+\infty} \beta(t)\Gamma_\gamma(t) dt = +\infty.$$

(i) Using the classical derivation chain rule, the definition of W, and the formula (2.9) for $\dot{W}$, we obtain

$$\begin{aligned} \frac{d}{dt}\left(\Gamma_\gamma(t)^2 W(t)\right) &= \Gamma_\gamma(t)\left(2\dot{\Gamma}_\gamma(t) W(t) + \Gamma_\gamma(t)\dot{W}(t)\right) \\ &= \Gamma_\gamma(t)\left(2\dot{\Gamma}_\gamma(t)\beta(t) + \Gamma_\gamma(t)\dot{\beta}(t)\right)(\Phi(x(t)) - m) \\ &+ \Gamma_\gamma(t)\left(\dot{\Gamma}_\gamma(t) - \gamma(t)\Gamma_\gamma(t)\right)\|\dot{x}(t)\|^2. \end{aligned}$$

According to $\dot{\Gamma}_\gamma(t) - \gamma(t)\Gamma_\gamma(t) = -1$, and to the property $(\mathrm{H})_{\gamma,\beta}$, i.e., $\dfrac{d}{dt}\left(\Gamma_\gamma(t)^2\beta(t)\right) \leq \Gamma_\gamma(t)\beta(t)$, we obtain

$$\frac{d}{dt}\left(\Gamma_\gamma(t)^2 W(t)\right) \leq \Gamma_\gamma(t)\beta(t)\left(\Phi(x(t)) - m\right).$$

According to Proposition 2.4, we have $\int_{t_0}^{+\infty} \Gamma_\gamma(t)\beta(t)\left(\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right) dt < +\infty$. The non-negative function $k(t) := \Gamma_\gamma(t)^2 W(t)$ verifies $\left(\dfrac{dk}{dt}\right)^+ \in L^1(t_0; +\infty)$. This implies that the limit of k exists, as $t \to +\infty$, that is,

$$\lim_{t\to+\infty} \Gamma_\gamma(t)^2 W(t) \quad \text{exists.}$$

Let's show that, under the assumption $\int_{t_0}^{+\infty} \frac{dt}{\Gamma_\gamma(t)} = +\infty$, this implies

$$\lim_{t\to+\infty} \Gamma_\gamma(t)^2 W(t) = 0.$$

Otherwise, there would exist some $c > 0$ such that $\Gamma_\gamma(t)W(t) \geq \frac{c}{\Gamma_\gamma(t)}$ for t sufficiently large. According to (2.20) $\int_{t_0}^{+\infty} \Gamma_\gamma(t)W(t) < +\infty$, this would imply $\int_{t_0}^{+\infty} \frac{dt}{\Gamma_\gamma(t)} < +\infty$, a contradiction. Thus we have obtained

$$\Phi(x(t)) - m = o\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \text{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{\Gamma_\gamma(t)}\right) \text{ as } t \to +\infty.$$

(ii) Let's define the functions $u, v : [t_0, +\infty[\to \mathbb{R}_+$ by

$$u(t) := \int_{t_0}^{t} \beta(s)\Gamma_\gamma(s)ds \quad \text{and} \quad v(t) := \frac{W(t)}{\beta(t)}.$$

Using the classical derivation chain rule, the definition of W, and the formula (2.9) for $\dot{W}$, we obtain

$$\begin{aligned}
\frac{dv}{dt}(t) &= -\frac{\dot{\beta}(t)}{\beta(t)^2}W(t) + \frac{1}{\beta(t)}\dot{W}(t) \\
&= -\frac{\dot{\beta}(t)}{\beta(t)^2}\left(\frac{1}{2}\|\dot{x}(t)\|^2 + \beta(t)\left(\Phi(x(t)) - m\right)\right) \\
&+ \frac{1}{\beta(t)}\left(-\gamma(t)\|\dot{x}(t)\|^2 + \dot{\beta}(t)\left(\Phi(x(t)) - m\right)\right) \\
&= -\frac{2\gamma(t)\beta(t) + \dot{\beta}(t)}{2\beta^2(t)}\|\dot{x}(t)\|^2.
\end{aligned}$$

By assumption, $2\gamma(t)\beta(t) + \dot{\beta}(t)$ is non-negative on $[t_0, +\infty[$. Therefore $\frac{dv}{dt}(t) \leq 0$, which gives that the function $v(\cdot)$ is non-increasing on $[t_0, +\infty[$.
By assumption, $\int_{t_0}^{+\infty} \beta(s)\Gamma_\gamma(s)ds = +\infty$. Therefore, the function u is an increasing bijection from $[t_0, +\infty[$ onto $[0, +\infty[$. Set $r(t) := u^{-1}(\delta u(t))$ where $\delta \in]0, 1[$ is fixed and $t \in [t_0, +\infty[$. In view of the increasing property of u, we have $r(t) \leq t$ for all $t \in [t_0, +\infty[$. By definition of u and of $r(t)$, we have

$$\int_{t_0}^{r(t)} \beta(s)\Gamma_\gamma(s)ds = u(r(t)) = \delta u(t) = \delta \int_{t_0}^{t} \beta(s)\Gamma_\gamma(s)ds.$$

Therefore,

$$\int_{r(t)}^{t} \beta(s)\Gamma_\gamma(s)ds = \int_{t_0}^{t} \beta(s)\Gamma_\gamma(s)ds - \int_{t_0}^{r(t)} \beta(s)\Gamma_\gamma(s)ds = (1-\delta)\int_{t_0}^{t} \beta(s)\Gamma_\gamma(s)ds.$$

Recalling that v is non-increasing, we deduce that

$$\begin{aligned}\int_{r(t)}^t \Gamma_\gamma(s) W(s) ds &= \int_{r(t)}^t \beta(s)\Gamma_\gamma(s) v(s) ds \\ &\geq v(t) \int_{r(t)}^t \beta(s)\Gamma_\gamma(s) ds = (1-\delta)\frac{W(t)}{\beta(t)} \int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds.\end{aligned}$$

Since $\displaystyle\int_{t_0}^{+\infty} \Gamma_\gamma(t) W(t) dt < +\infty$, we deduce that

$$0 \leq \lim_{t\to+\infty} \frac{W(t)}{\beta(t)} \int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds \leq \frac{1}{1-\delta} \lim_{t\to+\infty} \int_{r(t)}^t \Gamma_\gamma(s) W(s) ds = 0.$$

We conclude that $W(t) = o\left(\dfrac{\beta(t)}{\int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds}\right)$. Equivalently

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds}\right) \quad \text{and} \quad \|\dot{x}(t)\|^2 = o\left(\frac{\beta(t)}{\int_{t_0}^t \beta(s)\Gamma_\gamma(s) ds}\right).$$

□

Remark 2.6. Note that the non-negativity condition on $2\gamma(t)\beta(t) + \dot{\beta}(t)$ implies, after integration,

$$\beta(t) \geq \beta(t_0) e^{-2\int_{t_0}^t \gamma(s) ds} = \frac{\beta(t_0)}{p(t)^2}. \tag{2.26}$$

Remark 2.7. The convergence of the trajectories generated by the dynamic system $(\mathrm{IGS})_{\gamma,\beta}$ will be analyzed in section 4, in the more general case where additional perturbations are taken into consideration.

3. Particular cases

3.1. Case $\beta(t) = 1$, $\gamma(\cdot)$ general.

The asymptotic convergence properties of the dynamical system $(\mathrm{IGS})_{\gamma,1}$ are based on the behavior of $\Gamma_\gamma(t)$ as $t \to +\infty$. In this case, the condition $(\mathrm{H})_{\gamma,\beta}$ reduces to

$$\gamma(t)\Gamma_\gamma(t) \leq \frac{3}{2}, \tag{H_1}$$

which is the condition introduced by Attouch-Cabot in [5]. Thus, we recover Corollary 3.4. in [5], that is,

Corollary 3.1. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex function of class $\mathcal{C}^1$ such that* $\operatorname{argmin} \Phi \neq \emptyset$. *Let us assume that $\gamma : [t_0, +\infty[\to \mathbb{R}_+$ is a continuous function satisfying* $(\mathrm{H})_0$ *and* $(\mathrm{H})_1$. *Then, every solution trajectory $x : [t_0, +\infty[\to \mathcal{H}$ of* $(\mathrm{IGS})_{\gamma,\beta}$ *satisfies the following convergence rate of the values:*

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{\Gamma_\gamma(t)^2}\right) \quad \textit{as } t \to +\infty. \tag{3.1}$$

The condition $(\mathrm{H})_{\gamma,\beta}^{+}$, which means $\exists t_1 \geq t_0, \exists \rho > 0$ such that $2\gamma(t)\Gamma_\gamma(t) - 3 + \rho \leq 0$ for every $t \geq t_1$, becomes

$$(\mathrm{H})_1^+ \qquad \limsup_{t\to+\infty} \gamma(t)\Gamma_\gamma(t) < \frac{3}{2}.$$

This condition is equivalent to existence of $m < \frac{3}{2}$ such that $\gamma(t)\Gamma_\gamma(t) \leq m$ on $[t_1, +\infty[$. Therefore, the assertion (ii) of Theorem 2.5 gives

Corollary 3.2. ([5, Theorem 3.6]) *Suppose that the conditions of Corollary 3.1 are satisfied, together with* $\int_{t_0}^{+\infty} \Gamma_\gamma(t)dt = +\infty$ *and* $(\mathrm{H})_1^+$. *Then, as* $t \to +\infty$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\int_{t_0}^t \Gamma_\gamma(s)ds}\right), \quad \textit{and} \quad \|\dot{x}(t)\|^2 = o\left(\frac{1}{\int_{t_0}^t \Gamma_\gamma(s)ds}\right). \tag{3.2}$$

The assertion (i) of Theorem 2.5, in the case $\beta(t) = 1$, is new and can be formulated as follows.

Corollary 3.3. *Suppose that the conditions of Corollary 3.1 are satisfied together with* $\int_{t_0}^{+\infty} \frac{1}{\Gamma_\gamma(t)}dt = +\infty$ *and* $(\mathrm{H})_1^+$. *Then, as* $t \to +\infty$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\Gamma_\gamma(t)^2}\right) \quad \textit{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{\Gamma_\gamma(t)}\right). \tag{3.3}$$

3.2. **Case $\beta(\cdot)$ general, $\gamma(t) = \frac{\alpha}{t}$, $\alpha > 1$.** In this case, the inertial dynamic $(\mathrm{IGS})_{\gamma,\beta}$ writes

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta(t)\nabla\Phi(x(t)) = 0. \tag{3.4}$$

Elementary computation gives $p(t) = \left(\frac{t}{t_0}\right)^\alpha$. Hence $\int_{t_0}^{+\infty} \frac{du}{p(u)} < +\infty$ for $\alpha > 1$, and condition $(\mathrm{H})_0$ is satisfied. From this, we readily obtain $\Gamma_\gamma(t) = \frac{t}{\alpha-1}$. Condition $(\mathrm{H})_{\gamma,\beta}$ reduces to

$$(\mathrm{H})_2 \qquad t\dot{\beta}(t) \leq (\alpha - 3)\beta(t) \text{ for } t \geq t_0.$$

As a consequence of Theorem 2.1 we recover the convergence rate of values of [10, Theorem 8.1], which can be formulated as follows:

Corollary 3.4. *Let* $\Phi : \mathcal{H} \to \mathbb{R}$ *be a convex function of class* $\mathcal{C}^1$ *such that* $\operatorname{argmin}\Phi \neq \emptyset$. *Let us assume that* $\beta : [t_0, +\infty[\to \mathbb{R}_+$ *is a continuous function, and that condition* $(\mathrm{H})_2$ *is satisfied. Then every solution trajectory* $x : [t_0, +\infty[\to \mathcal{H}$ *of* (3.4) *satisfies the following convergence rate of the values:*

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^2\beta(t)}\right) \quad \textit{as } t \to +\infty. \tag{3.5}$$

Similar computation gives the following formulation of condition $(\mathrm{H})_{\gamma,\beta}^+$:

$$(\mathrm{H})_2^+ \qquad t\dot{\beta}(t) \leq \beta(t)\left(\alpha - 3 - \rho(\alpha-1)\right) \iff \limsup_{t\to+\infty} \frac{t\dot{\beta}(t)}{\beta(t)} < \alpha - 3.$$

Noticing that $\displaystyle\int_{t_0}^{+\infty} \frac{1}{\Gamma_\gamma(t)}dt = +\infty$, a direct application of Theorem 2.5 gives the following result:

Corollary 3.5. *(i) Under the condition* $(\mathrm{H})_2^+$*, every solution trajectory* $x : [t_0, +\infty[\to \mathcal{H}$ *of* (3.4) *satisfies the following convergence rate of the values:*

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^2\beta(t)}\right) \quad \text{as } t \to +\infty. \tag{3.6}$$

(ii) Suppose moreover that

$$\int_{t_0}^{+\infty} t\beta(t)dt = +\infty \quad \text{and} \quad 2\alpha\beta(t) + t\dot{\beta}(t) \geq 0, \ \forall t \in [t_0, +\infty[, \tag{3.7}$$

then

(3.8)

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\int_{t_0}^t s\beta(s)ds}\right) \quad \text{and} \quad \|\dot{x}(t)\|^2 = o\left(\frac{\beta(t)}{\int_{t_0}^t s\beta(s)ds}\right) \quad \text{as } t \to +\infty.$$

Remark 3.6. a) When $\beta(t) \equiv 1$ we get the classical results for the evolution equation

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

which can be seen as a continuous version of the Nesterov method. In this case, the condition $(\mathrm{H})_2$ simply writes $\alpha \geq 3$. As a result, for $\alpha \geq 3$, we get $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^2}\right)$, and for $\alpha > 3$ $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^2}\right)$ as $t \to +\infty$, see [5], [8], [27], [39].

b) When $\beta(t) = t^p$, we readily obtain that the condition $(\mathrm{H})_2$ is equivalent to $\alpha \geq 3 + p$.

3.3. **Case** $\beta(t) = \beta_0 t^a \ln(t)^b$**,** $\gamma(t) = \frac{\alpha}{t}$**.** In this case, we readily obtain

$$t^2\beta(t) = \mathcal{O}\left(t^{a+2}\ln(t)^b\right) \text{ and } \int_{t_0}^t s\beta(s)ds = \mathcal{O}\left(t^{a+2}\ln(t)^b\right).$$

By specializing Corollary 3.4 and Corollary 3.5 to this situation, we obtain the following statement.

Corollary 3.7. *Let* $x : [t_0, +\infty[\to \mathcal{H}$ *be a solution trajectory of* (3.4) *with* $\beta(t) = \beta_0 t^a \ln(t)^b$*, and* $\alpha > 1$*.*

(i) *If* $a \leq \alpha - 3$ *and* $b \leq 0$*, then* $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{a+2}\ln(t)^b}\right)$ *as* $t \to +\infty$*.*

(ii) *If* $a < \alpha - 3$*, then* $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^{a+2}\ln(t)^b}\right)$ *as* $t \to +\infty$*.*

Proof. To show assertions (i) and (ii), we need to examine conditions $(\mathrm{H})_2$ and $(\mathrm{H})_2^+$, respectively. For $t \geq t_0 > 0$,

$$\frac{t\dot{\beta}(t)}{\beta(t)} - \alpha + 3 = a - \alpha + 3 + \frac{b}{\ln(t)}.$$

(i) If $a \leq \alpha - 3$ and $b \leq 0$, then $\dfrac{t\dot{\beta}(t)}{\beta(t)} - \alpha + 3 \leq 0$ for every $t \geq t_0$ and $(\mathrm{H})_2$ is satisfied.

(ii) If $a < \alpha - 3$, then have $\limsup_{t\to+\infty} \left(\dfrac{t\dot{\beta}(t)}{\beta(t)} - \alpha + 3 \right) < 0$, and $(\mathrm{H})_2^+$ is satisfied. $\square$

3.4. **Optimality of the results.** Let us show the optimality of the convergence rate obtained in Theorem 2.1. This will result from the following example showing that when $\gamma(t) = \frac{\alpha}{t}$, and $\beta(t) = t^\delta$, $\mathcal{O}(1/t^{2+\delta})$ is the worst possible case for the convergence rate of values. The following example was obtained by rescaling the example of [8]. Take $\mathcal{H} = \mathbb{R}$ and $\Phi(x) = c|x|^r$, where c and r are positive parameters. We consider the evolution equation

$$(\mathrm{IGS})_{\frac{\alpha}{t},t^\delta} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + t^\delta \nabla\Phi(x(t)) = 0,$$

and look for non-negative solutions of the form $x(t) = \frac{1}{t^\theta}$, with $\theta > 0$. This corresponds to trajectories that are completely damped. We begin by determining the values of c, r,θ providing such solutions. On the one hand,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) = \theta(\theta+1-\alpha)\frac{1}{t^{\theta+2}}.$$

On the other hand, $\nabla\Phi(x) = cr|x|^{r-2}x$, which gives

$$t^\delta \nabla\Phi(x(t)) = cr\frac{1}{t^{\theta(r-1)-\delta}}.$$

Thus, $x(t) = \frac{1}{t^\theta}$ is solution of $(\mathrm{IGS})_{\frac{\alpha}{t},t^\delta}$ if and only if,

i) $\theta + 2 = \theta(r-1) - \delta$, which is equivalent to $r > 2$ and $\theta = \frac{2+\delta}{r-2}$; and
ii) $cr = \theta(\alpha - \theta - 1)$, which is equivalent to $\alpha > \frac{r+\delta}{r-2}$ and $c = \frac{2+\delta}{r(r-2)}(\alpha - \frac{r+\delta}{r-2})$.

We have $\min\Phi = 0$ and

$$\Phi(x(t)) = \frac{2+\delta}{r(r-2)}\left(\alpha - \frac{r+\delta}{r-2}\right)\frac{1}{t^{\frac{r(2+\delta)}{r-2}}}.$$

The convergence rate of $\Phi(x(t))$ to 0 depends on the parameter r. When r goes to infinity, the exponent $\frac{r(2+\delta)}{r-2} > 2+\delta$ tends to $2+\delta$. This limit situation is obtained by taking a function Φ which becomes very flat around the set of its minimizers. Therefore, without other geometrical hypotheses on Φ, we cannot expect a convergence rate better than $\mathcal{O}(1/t^{2+\delta})$. It is precisely the convergence rate provided by Theorem 2.1 and Corollary 3.5, since $t^2\beta(t) = t^{2+\delta}$ in this situation.
Without rescaling, the optimality of the convergence rate of the values for the damped inertial gradient systems has been analyzed in a recent work of Apidopoulos-Aujol-Dossal-Rondepierre [4].

3.5. **Numerical examples.** The following examples have been implemented with the Scilab opensource software version 5.5.2. In the context of convex optimization, we consider successively the strongly convex case with different conditionings, then the strictly convex case without strong convexity, and finally the case of a convex function with a continuum of solutions. The initial time t_0 has been taken equal to one.

3.5.1. **Example 1. (Strongly convex).** We consider the strongly convex functions

$\Phi_1(x_1,x_2) := 5\cdot 10^{-3}x_1^2 + x_2^2$ (ill-conditioned)

$\Phi_2(x_1,x_2) := 5x_1^2 + x_2^2$ (well-conditioned),

which respectively illustrate the ill-conditioned case, and the well-conditioned case.

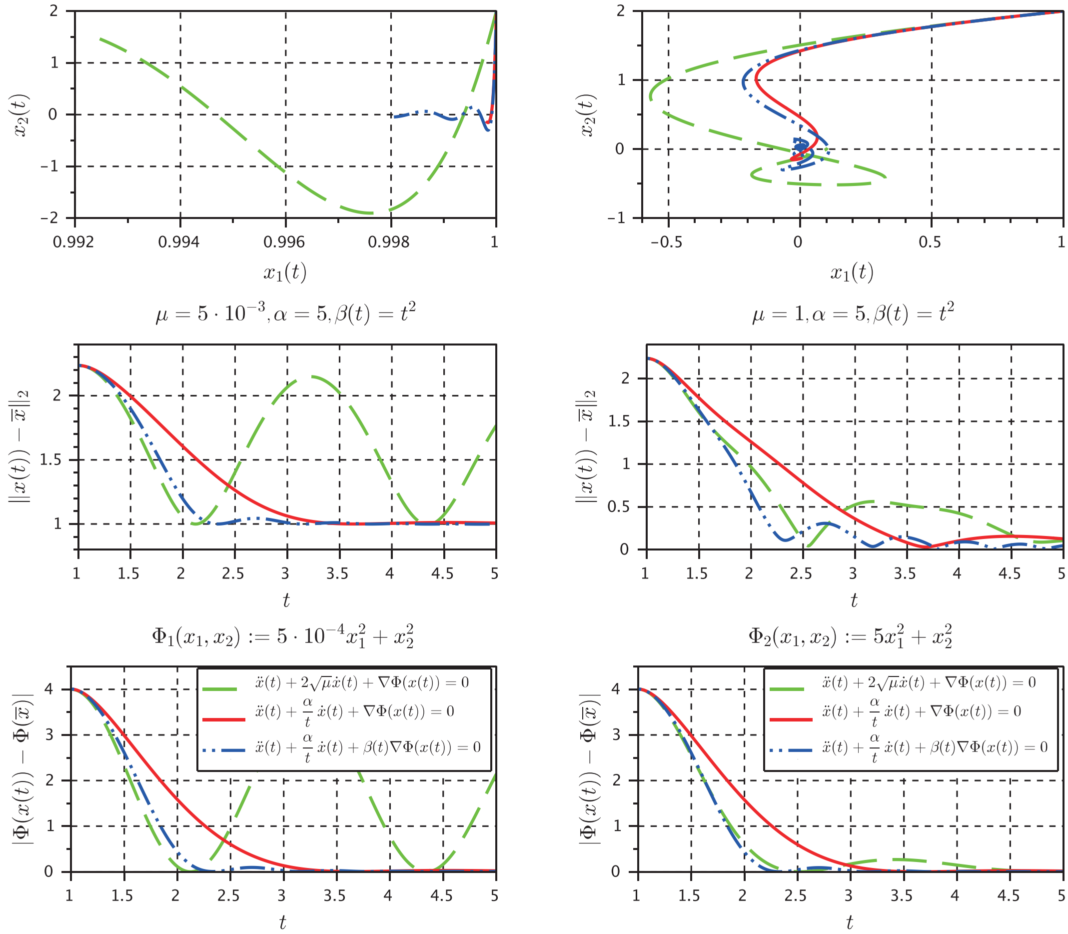

FIGURE 2. Graphical view of of $\|x(t)-\bar{x}\|_2$ and $|\Phi_i(x(t))-\Phi_i(\bar{x})|$ for different values of $\gamma(t)$ and $\beta(t)$.

We investigate numerically the convergence behaviour of $\|x(t)-\bar{x}\|_2$ and $|\Phi_i(x(t))-\Phi_i(\bar{x})|$ $(i=1,2)$ where $x(\cdot)$ is the solution of the dynamical system $(\text{IGS})_{\gamma,\beta}$, with the initial conditions $(x_1(1),x_2(1)) = (1,2)$ and $(\dot{x}_1(1),\dot{x}_2(1)) = (0,0)$. Three choices of γ and β are considered:

- $\gamma(t) \equiv 2\sqrt{\mu_i}$ and $\beta(t)\equiv 1$, where μ_i is the smallest eigenvalue for $\nabla^2\Phi_i$. This choice is in accordance with the linear convergence results for the heavy ball method in the strongly convex case,
- $\gamma(t) = \frac{5}{t}$ and $\beta(t)\equiv 1$,
- $\gamma(t) = \frac{5}{t}$ and $\beta(t) = t^2$.

The functions Φ_i reach their infimum which is equal to zero uniquely at $(0,0)$. According to our theoretical results, we have the asymptotic convergence of values to zero. Compared results are shown in Figure 2. For the ill-conditioned function Φ_1, the convergence of the values is faster for $\gamma(t) = \frac{5}{t}$ than for $\gamma(t)$ constant.

3.5.2. **Example 2. (Convex and non strongly convex).**

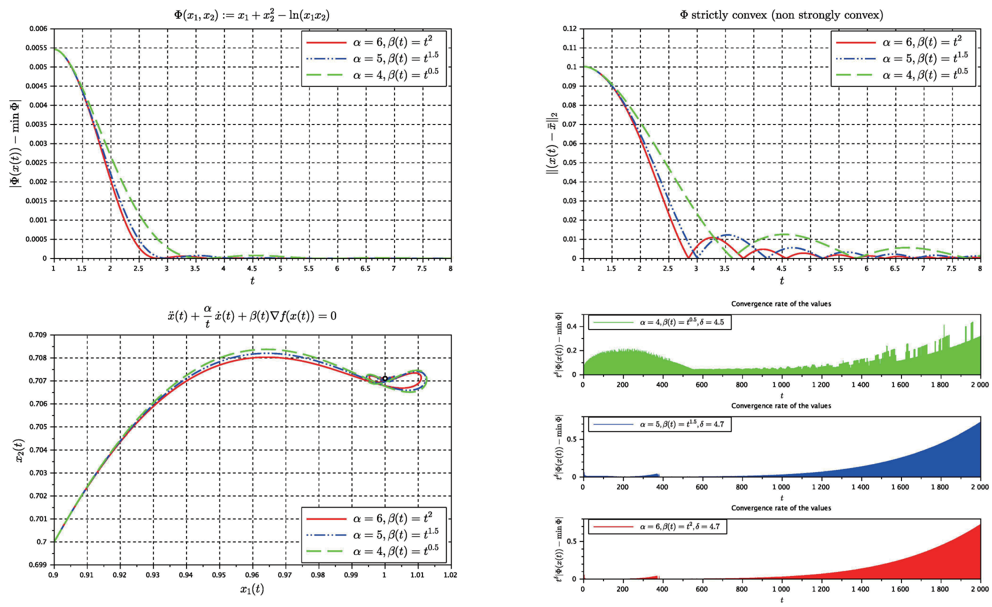

FIGURE 3.

Let's analyze the convergence properties, as $t \to +\infty$, of the trajectories of the dynamical system $(\text{IGS})_{\gamma,\beta}$ in the case where Φ is defined on $]0,+\infty[^2$ by $\Phi(x_1,x_2) = x_1 + x_2^2 - \ln(x_1x_2)$.
We can easily verify that Φ is a strictly convex (non-strongly convex) function, and that $\bar{x} = (1, \sqrt{2}/2)$ is the unique global minimum point of Φ. We take $\gamma(t) = \frac{\alpha}{t}$ and $\beta(t) = t^p$ so as to satisfy condition $(\text{H})_2$, which in this case is equivalent to $\alpha \geq 3 + p$. Specifically we compare the convergence rates for the solutions u_1, u_2, u_3 of the system $(\text{IGS})_{\gamma,\beta}$ for $(\alpha_1 = 4, \beta_1(t) = t^{0.5}), (\alpha_2 = 5, \beta_2(t) = t^{1.5})$ and $(\alpha_3 = 6, \beta_3(t) = t^2)$. This is illustrated in Figure 3, where the corresponding trajectories are displayed on the same screen with their optimal end point $\bar{x}$. We observe that the numerical examples illustrated in Figure 3 are in agreement with the convergence rates predicted in Corollary 3.4. Precisely, in accordance with $\Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}\left(\frac{1}{t^2\beta(t)}\right) = \mathcal{O}\left(\frac{1}{t^{2+p}}\right)$ as $t \to +\infty$, we have, with $\delta_i \geq 2 + p_i$,

$$\max_{i=1,2,3}\ \max_{1\leq t\leq 2000}\left[t^{\delta_i}(|\Phi(u_i(t)) - \min\Phi|)\right] \leq 0.5.$$

3.5.3. **Example 3. (Non unique global solution).** Take $\Phi : \mathbb{R}^2 \to \mathbb{R}+$ defined by $\Phi(x) = (x_1 + x_2)^2$. The function Φ is convex but non strongly convex. Its solution set is the whole line $\operatorname{argmin}\Phi = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = -x_1\}$. Depending on the initial data, we observe the convergence of the trajectories to different equilibria belonging to $\operatorname{argmin}\Phi$. In Figure 4 (left) are represented the trajectories of $(\text{IGS})_{\gamma,\beta}$ corresponding to the initial position $(x_1(1), x_2(1)) = (1, 2)$ and various initial velocity vectors $(\dot{x}_1(1), \dot{x}_2(1))$. In Figure 4 (right) we observe a similar phenomenon when we change the initial point $(x_1(1), x_2(1))$ while keeping the same initial velocity vector $(\dot{x}_1(1), \dot{x}_2(1)) = (15, 0)$.

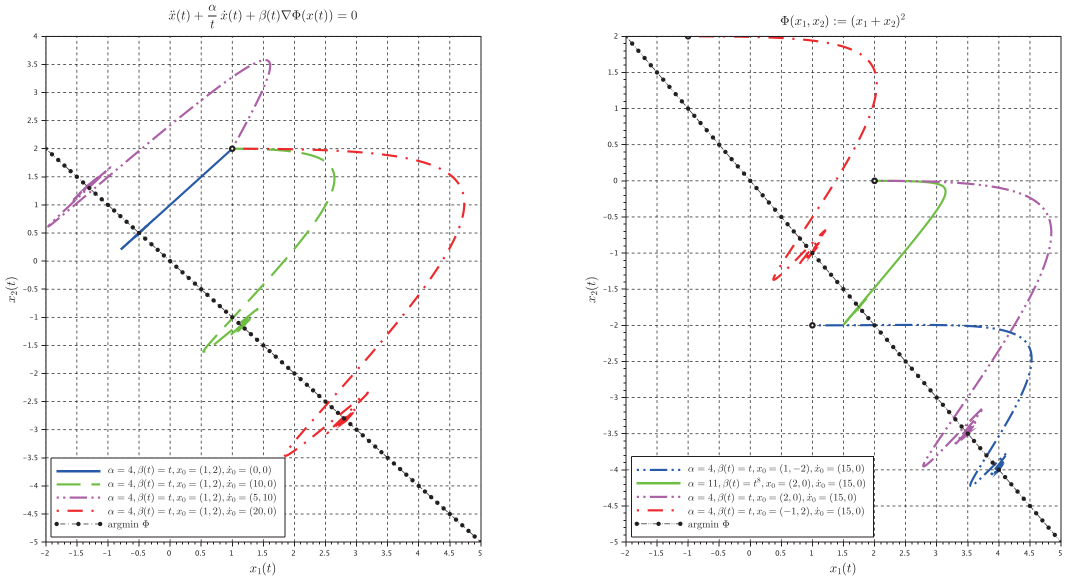

FIGURE 4.

4. EXISTENCE AND STABILITY WITH RESPECT TO PERTURBATIONS

Consider the perturbed version of the initial evolution system $(\text{IGS})_{\gamma,\beta}$

$$(\text{IGS})_{\gamma,\beta,e} \qquad \ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla\Phi(x(t)) = e(t),$$

where the second member $e(\cdot)$ can be interpreted as an external action on the system, an error, a perturbation or a control term. For the existence and uniqueness of classical global solution of the Cauchy problem associated with the evolution system $(\text{IGS})_{\gamma,\beta,e}$, we refer to Theorem 8.2.
In this section, we show that the results of the previous sections remain satisfied if the perturbation g is sufficiently small asymptotically.

Theorem 4.1. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a $\mathcal{C}^1$ convex function with* $\operatorname{argmin}\Phi \neq \emptyset$. *Take $\beta : [t_0, +\infty[\to \mathbb{R}_+^*$ a continuous function. Suppose that $\gamma : [t_0, +\infty[\to \mathbb{R}_+$ is a continuous function that satisfies* $(\text{H})_0$. *Suppose that the function $e : [t_0, +\infty[\to \mathcal{H}$ is locally integrable and verifies*

$$(4.1) \qquad (\text{H})_{\text{e}} \quad \int_{t_0}^{+\infty} \Gamma_\gamma(t)\|e(t)\|dt < +\infty.$$

Then, for any solution $x : [t_0, +\infty[\to \mathcal{H}$ *of* $(\mathrm{IGS})_{\gamma,\beta,e}$, *the following statements are satisfied:*

(i) *Under the condition* $(\mathrm{H})_{\gamma,\beta}$, $x(t)$ *and* $\Gamma_\gamma(t)\dot{x}(t)$ *are bounded on* $[t_0, +\infty[$, *and we have the convergence rate of the values:*

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \text{as } t \to +\infty. \tag{4.2}$$

(ii) *Moreover, under condition* $(\mathrm{H})^+_{\gamma,\beta}$, *we have :*

$$\int_{t_0}^{+\infty} \beta(t)\Gamma_\gamma(t)\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) dt < +\infty. \tag{4.3}$$

Proof. The guiding idea of the proof is the same as in the Theorem 2.1. As a Lyapunov function, we use the same energy function $\mathcal{E}(\cdot)$, which is defined for $t \geq t_0$ by

$$\mathcal{E}(t) \;=\; \beta(t)\Gamma_\gamma(t)^2\left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\|^2.$$

The time derivative of $\mathcal{E}(\cdot)$ is given by

$$\begin{aligned}\dot{\mathcal{E}}(t) = \;& \frac{d}{dt}\left(\beta(t)\Gamma_\gamma(t)^2\right)\left(\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right) + \beta(t)\Gamma_\gamma(t)^2\langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\ & + \left\langle \frac{d}{dt}\left(x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\right), x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\right\rangle.\end{aligned} \tag{4.4}$$

Then, according to (2.3), note that the dynamical system $(\mathrm{IGS})_{\gamma,\beta,e}$ can be formulated as

$$\frac{d}{dt}\left(\Gamma_\gamma(t)\dot{x}(t) + x(t) - z\right) = \Gamma_\gamma(t)e(t) - \beta(t)\Gamma_\gamma(t)\nabla\Phi(x(t)). \tag{4.5}$$

From (4.4) and (4.5) we deduce that

$$\begin{aligned}\dot{\mathcal{E}}(t) \;=\;& \frac{d}{dt}\left(\beta(t)\Gamma_\gamma(t)^2\right)\left(\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right) + \beta(t)\Gamma_\gamma(t)^2\langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\ & + \langle \Gamma_\gamma(t)e(t) - \beta(t)\Gamma_\gamma(t)\nabla\Phi(x(t)), x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\rangle \\ =\;& \frac{d}{dt}\left(\beta(t)\Gamma_\gamma(t)^2\right)\left(\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right) + \langle \Gamma_\gamma(t)e(t), x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\rangle \\ & - \langle \beta(t)\Gamma_\gamma(t)\nabla\Phi(x(t)), x(t) - z\rangle.\end{aligned}$$

Using the Cauchy-Schwarz inequality and the convexity of Φ, that is,

$$\langle \nabla\Phi(x(t)), x(t) - z\rangle \geq \Phi(x(t)) - \Phi(z),$$

we obtain

$$\begin{aligned}\dot{\mathcal{E}}(t) \leq & \left[\frac{d}{dt}\left(\beta\Gamma_\gamma^2\right)(t) - (\beta\Gamma_\gamma)(t)\right]\left(\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right) \\ & + \Gamma_\gamma(t)\|e(t)\| \cdot \|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\|.\end{aligned} \tag{4.6}$$

Using the condition $(\mathrm{H})_{\gamma,\beta}$ and the definition of $\mathcal{E}$, we deduce that

$$\dot{\mathcal{E}}(t) \leq \Gamma_\gamma(t)\|e(t)\| \cdot \|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\| \leq \sqrt{2}\Gamma_\gamma(t)\|e(t)\|\sqrt{\mathcal{E}(t)}, \tag{4.7}$$

By integrating the differential inequality (4.7), and using the assumption 4.1, we obtain

$$\sqrt{\mathcal{E}(t)} \leq \sqrt{\mathcal{E}(t_0)} + \frac{1}{\sqrt{2}} \int_{t_0}^{+\infty} \Gamma_\gamma(s)\|e(s)\|ds = Cte < +\infty,$$

which gives the claim (4.2). In addition, we obtain that $\|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\|^2$ is bounded, which gives

$$\|x(t) - z\|^2 + 2\Gamma_\gamma(t) \langle x(t) - z, \dot{x}(t)\rangle \leq C.$$

Set $h(t) := \frac{1}{2} \|x(t) - z\|^2$. The above inequality gives

$$h(t) + \Gamma_\gamma(t)\dot{h}(t) \leq \frac{1}{2}C.$$

By an argument similar to that of the unperturbed case, the integration of the above differential inequality gives that the trajectory $x(\cdot)$ is bounded. Returning to the boundedness of $\|x(t) - z + \Gamma_\gamma(t)\dot{x}(t)\|$, we also conclude that $\Gamma_\gamma(t)\dot{x}(t)$ is bounded on $[t_0, +\infty[$. To prove the second affirmation of the theorem, let us return to the relation (4.6). Since $x(t) - z + \Gamma_\gamma(t)\dot{x}(t)$ is bounded on $[t_0, \infty[$, there exists some $C > 0$ such that

$$\dot{\mathcal{E}}(t) \leq \left[\frac{d}{dt}\left(\beta\Gamma_\gamma^2\right)(t) - (\beta\Gamma_\gamma)(t)\right]\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) + C\Gamma_\gamma(t)\|e(t)\|.$$

Integrating on $[t_0, +\infty[$, and using condition $(\mathrm{H})_{\gamma,\beta}^{+}$ and (2.3), we finally get

$$\int_{t_0}^{+\infty} \beta(t)\Gamma_\gamma(t)\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) dt \leq \frac{1}{\rho}\left(\mathcal{E}(t_0) + C\int_{t_0}^{+\infty} \Gamma_\gamma(t)\|e(t)\|dt\right) < +\infty,$$

which completes the proof. □

As in the unperturbed case, we can now pass from capital $\mathcal{O}$ estimates to small o estimates under the slightly stronger hypothesis $(\mathrm{H})_{\gamma,\beta}^{+}$. In addition, we obtain the convergence of trajectories.

Theorem 4.2. *Let $\Phi : \mathcal{H} \to \mathbb{R}$ be a convex continuously differentiable function such that* $\operatorname{argmin} \Phi$ *is nonempty. Suppose that the properties* $(\mathrm{H})_{\gamma,\beta}^{+}$ *and* (4.1) *are satisfied. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a solution trajectory of* $(\mathrm{IGS})_{\gamma,\beta,e}$. *Then, as $t \to +\infty$, $x(t)$ converges weakly to a point in* $\operatorname{argmin} \Phi$.
If moreover $\displaystyle\int_{t_0}^{+\infty} \frac{1}{\Gamma_\gamma(t)}dt = +\infty$, we obtain

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{\beta(t)\Gamma_\gamma(t)^2}\right) \quad \textit{and} \quad \|\dot{x}(t)\| = o\left(\frac{1}{\Gamma_\gamma(t)}\right) \quad \textit{as } t \to +\infty.$$

Proof. For the weak convergence, the proof is based on Opial's Lemma 8.3. By elementary calculus, convexity of Φ, and equation $(\mathrm{IGS})_{\gamma,\beta,e}$, one can first establish, as in (2.10), that for any $z \in \operatorname{argmin} \Phi$, the function $h_z(t) := \frac{1}{2}\|x(t) - z\|^2$ satisfies

$$\ddot{h}_z(t) + \gamma(t)\dot{h}_z(t) \leq \|\dot{x}(t)\|^2 + \langle e(t), x(t) - z\rangle.$$

Taking the norm of each member and using the boundedness of $x(t)$, see Theorem 4.1, we deduce that

$$\ddot{h}_z(t) + \gamma(t)\dot{h}_z(t) \leq \|\dot{x}(t)\|^2 + M\|e(t)\|. \tag{4.8}$$

By multiplying this differential inequality by $p(t)$, using $p(t)\gamma(t) = \dot{p}(t)$ and integrating, we obtain

$$\dot{h}_z(t) \leq \frac{1}{p(t)}\int_{t_0}^{t}\left(\|\dot{x}(s)\|^2 + M\|e(s)\|\right)ds + \frac{p(t_0)}{p(t)}\dot{h}_z(t_0). \tag{4.9}$$

Set $L := \int_{t_0}^{+\infty}\frac{dt}{p(t)}$. By integrating (4.9) and applying Fubini theorem, we obtain

$$\begin{aligned}\int_{t_0}^{+\infty}\left[\dot{h}_z\right]^+(t)dt &\leq \int_{t_0}^{+\infty}\left(\frac{1}{p(t)}\int_{t_0}^{t}\left(\|\dot{x}(s)\|^2 + M\|e(s)\|\right)ds\right)dt + p(t_0)\dot{h}_z(t_0)L\\ &= \int_{t_0}^{+\infty}\left(\|\dot{x}(s)\|^2 + M\|e(s)\|\right)\left(\int_{s}^{+\infty}\frac{dt}{p(t)}\right)ds + p(t_0)\dot{h}_z(t_0)L\\ &= \int_{t_0}^{+\infty}\Gamma_\gamma(s)\left(\|\dot{x}(s)\|^2 + M\|g(s)\|\right)ds + p(t_0)\dot{h}_z(t_0)L.\end{aligned}$$

It remains to prove the estimate $\int_{t_0}^{+\infty}\Gamma_\gamma(t)\|\dot{x}(t)\|^2dt < +\infty$. As in the proof of Proposition 2.4, multiplying $(\mathrm{IGS})_{\gamma,\beta,e}$ by the vector $\dot{x}$ and by $\Gamma_\gamma(t)^2$, integrating on (t_0,t) and using $(\mathrm{H})_{\gamma,\beta}$, we obtain

$$\begin{aligned}\int_{t_0}^{t}\Gamma_\gamma(s)\|\dot{x}(s)\|^2ds &\leq \int_{t_0}^{t}\Gamma_\gamma(s)\beta(s)\left(\Phi(x(t)) - m\right)ds + W(t_0) - W(t)\\ &+ \int_{t_0}^{t}\Gamma_\gamma(s)^2\langle g(s),\dot{x}(s)\rangle ds\\ &\leq \int_{t_0}^{+\infty}\Gamma_\gamma(s)\beta(s)\left(\Phi(x(t)) - m\right)ds + W(t_0)\\ &+ \sup_{s\in[t_0,+\infty]}\Gamma_\gamma(s)\|\dot{x}(s)\|\int_{t_0}^{+\infty}\Gamma_\gamma(s)\|e(s)\|ds.\end{aligned}$$

Using 4.1, the boundedness of $\Gamma_\gamma(s)\|\dot{x}(s)\|$ on $[t_0,+\infty]$, and (4.3), we obtain

$$\int_{t_0}^{+\infty}\Gamma_\gamma(t)\|\dot{x}(t)\|^2dt < +\infty, \tag{4.10}$$

and consequently $\left[\dot{h}_z\right]^+ \in L^1(t_0,+\infty)$. Since h_z is non-negative, this implies the convergence of $h_z(t)$ as $t \to +\infty$. The second item of Lemma 8.3 is a direct consequence of the minimizing property (2.7) of the trajectory, and of the lower semicontinuity for the weak topology of the convex continuous function Φ.

Following the lines of the proof of the first statement of Theorem 2.1, we can show that

$$\begin{aligned}
&\int_{t_0}^{+\infty} \frac{d}{dt}\left(\Gamma_\gamma(t)^2 W(t)\right) dt \\
&\leq \int_{t_0}^{+\infty} \Gamma_\gamma(t)\beta(t)\left(\Phi(x(t)) - m\right) dt + \int_{t_0}^{+\infty} \Gamma_\gamma(t)^2 \|e(t)\| \|\dot{x}(t)\| dt \\
&\leq \int_{t_0}^{+\infty} \Gamma_\gamma(t)\beta(t)\left(\Phi(x(t)) - m\right) dt + \sup_{t\geq t_0} \Gamma_\gamma(t)\|\dot{x}(t)\| \int_{t_0}^{+\infty} \Gamma_\gamma(t)\|e(t)\| dt < +\infty.
\end{aligned}$$

From (4.3) and (4.10) we have

$$\int_{t_0}^{+\infty} \Gamma_\gamma(t) W(t) dt < +\infty.$$

In addition, by using $\int_{t_0}^{+\infty} \frac{dt}{\Gamma_\gamma(t)} = +\infty$, we obtain that $\lim_{t\to+\infty} \Gamma_\gamma(t)^2 W(t) = 0$, which ends the proof. $\square$

5. Continuous modeling of Güler's Inertial Proximal Point Algorithm

In [23], Güler first studied the convergence rate of the proximal point algorithm without inertia. Then, in [24], he introduced the so-called Inertial Proximal Point Algorithm, which combines the ideas of Nesterov and Martinet as follows:

(5.1)

- Initialization of ν_0 and $A_0 > 0$.
- Step k :
  - Choose $\beta_k > 0$, and calculate $g_k > 0$ by solving
  - $g_k^2 + g_k A_k \beta_k - A_k \beta_k = 0$.
  - $y_k = (1 - g_k)x_k + g_k \nu_k$;
  - $x_{k+1} = \text{prox}_{\beta_k \Phi}(y_k)$;
  - $\nu_{k+1} = \nu_k + \frac{1}{g_k}(x_{k+1} - y_k)$;
  - $A_{k+1} = (1 - g_k)A_k$.

Let us show that Güler's proximal algorithm (5.1) can be written as:

$$(\text{IP})_{\alpha_k,\beta_k} \qquad \begin{cases} y_k = x_k + (1-\alpha_k)(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{\beta_k \Phi}(y_k). \end{cases}$$

This result was first obtained by the authors in [10]. We reproduce it here briefly for the convenience of the reader. First verify that, for all $k \geq 1$

$$\nu_k = x_{k-1} + \frac{1}{g_{k-1}}\left(x_k - x_{k-1}\right). \tag{5.2}$$

For this, we use an induction argument. Suppose (5.2) is satisfied at step k, then show that it will be at step $k+1$. Using successively (5.1) and (5.2), we obtain

$$\begin{aligned} \nu_{k+1} &= \nu_k + \frac{1}{g_k}(x_{k+1} - y_k) = x_{k-1} + \frac{1}{g_{k-1}}(x_k - x_{k-1}) + \frac{1}{g_k}(x_{k+1} - y_k) \\ &= \frac{1}{g_k}x_{k+1} + x_{k-1} + \frac{1}{g_{k-1}}(x_k - x_{k-1}) - \frac{1}{g_k}((1-g_k)x_k + g_k\nu_k) \\ &= \frac{1}{g_k}x_{k+1} - \frac{1-g_k}{g_k}x_k = x_k + \frac{1}{g_k}(x_{k+1} - x_k), \end{aligned}$$

which shows that (5.2) is satisfied at step $k+1$. Then, combining once again (5.1) and (5.2), we obtain

$$\begin{aligned} y_k &= (1-g_k)x_k + g_k\nu_k = (1-g_k)x_k + g_k\left(x_{k-1} + \frac{1}{g_{k-1}}(x_k - x_{k-1})\right) \\ &= x_k + \left(\frac{g_k}{g_{k-1}} - g_k\right)(x_k - x_{k-1}). \end{aligned}$$

Hence, Güler's proximal algorithm can be written as $(\mathrm{IP})_{\alpha_k,\beta_k}$ with

$$\alpha_k = g_k\left(\frac{1}{g_{k-1}} - 1\right). \tag{5.3}$$

By construction of g_k, we have $g_k = \frac{1}{2}\left(-A_k\beta_k + \sqrt{(A_k\beta_k)^2 + 4A_k\beta_k}\right)$, which, by elementary calculation, gives $0 \leq g_k < 1$. According to (5.3), we deduce that $\alpha_k > 0$. As a result, this makes Güler's algorithm (5.1) as an inertial proximal algorithm $(\mathrm{IP})_{\alpha_k,\beta_k}$. From (5.1), we also get:

$$A_k = A_0\prod_{j=0}^{k-1}(1-g_j) \quad \text{and} \quad g_k^2 = A_k\beta_k(1-g_k) = \beta_k A_{k+1}; \tag{5.4}$$

and then, we obtain the following relation between β_k and g_k:

$$\beta_k = \frac{g_k^2}{A_0\prod_{j=0}^{k}(1-g_j)}. \tag{5.5}$$

According to (5.3)–(5.5), we have obtained that all the parameters entering into Güler's algorithm can be expressed according to the single parameter g_k.
Let's come with the dynamic interpretation of Güler's algorithm, as formulated in $(\mathrm{IP})_{\alpha_k,\beta_k}$. According to the formulation (5.3) of α_k we get

$$x_{k+1} + \beta_k\partial\Phi(x_{k+1}) \ni y_k = x_k + g_k\left(\frac{1}{g_{k-1}} - 1\right)(x_k - x_{k-1}).$$

Equivalently,

$$x_{k+1} - 2x_k + x_{k-1} + \left(g_k - \frac{g_k - g_{k-1}}{g_{k-1}}\right)(x_k - x_{k-1}) + \beta_k\partial\Phi(x_{k+1}) = 0. \tag{5.6}$$

This can be interpreted as a backward time discretization of the second-order evolution equation (when Φ is smooth)

$$\ddot{x}(t) + \left(g(t) - \frac{\dot{g}(t)}{g(t)}\right)\dot{x}(t) + \beta(t)\nabla\Phi(x(t)) = 0. \tag{5.7}$$

So, modeling the Güler's accelerated Backward algorithm $(\mathrm{IP})_{\alpha_k,\beta_k}$, we derive a second-order ordinary differential equation

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

where the damping coefficient is expressed as

$$\gamma(t) = g(t) - \frac{\dot{g}(t)}{g(t)} = g(t)\left(1 + \frac{d}{dt}\left(\frac{1}{g}\right)(t)\right). \tag{5.8}$$

The algorithm $(\mathrm{IP})_{\alpha_k,\beta_k}$ features a new optimal convergence rate than Nesterov's method and also can be applied for nonsmooth convex function Φ. The terminology $(\mathrm{IGS})_{\gamma,\beta}$ refers to the Inertial Gradient System with damping coefficient $\gamma(t)$ and time scale coefficient $\beta(t)$, see [5] for an extended study in the case $\beta \equiv 1$. Indeed, the parameter $\beta(\cdot)$ comes naturally with the time scaling of these dynamics.
Taking $w := \frac{1}{g}$, then (5.8) is equivalent to solve the non-autonomous linear differential equation

$$\dot{w}(t) - \gamma(t)w(t) = -1.$$

Set $p(t) := \exp\left(\int_{t_0}^{t}\gamma(\tau)d\tau\right)$ for $t \geq t_0$, then, following [5, Proposition 2.1] and assuming that $\int_{t_0}^{+\infty}\frac{ds}{p(s)} < +\infty$, we obtain $\Gamma_\gamma(t) = p(t)\int_t^{\infty}\frac{1}{p(s)}ds$ as the unique solution satisfying the limit condition $\lim_{t\to+\infty}\frac{\Gamma_\gamma(t)}{p(t)} = 0$. Hence the general solution of (5.8) is $g(t) = \frac{1}{Cp(t)+\Gamma_\gamma(t)}$ with $\frac{1}{C} = \lim\limits_{t\to+\infty} p(t)g(t)$.
When $\beta(t) \equiv \beta > 0$ is fixed, fast convergence of the values is obtained in [5, Corollary 3.4] under the condition $\gamma(t)\Gamma_\gamma(t) \leq \frac{3}{2}$. According to (5.8), in terms of $g(t)$, this condition takes the equivalent form

$$g(t)\left(1 + \frac{d}{dt}\left(\frac{1}{g}\right)(t)\right)\frac{1}{g(t)} = 1 + \frac{d}{dt}\left(\frac{1}{g}\right)(t) \leq \frac{3}{2}.$$

So, the condition $\gamma(t)\Gamma_\gamma(t) \leq \frac{3}{2}$ becomes $\frac{d}{dt}\left(\frac{1}{g}\right)(t) \leq \frac{1}{2}$, with the corresponding convergence rate of the values

$$\Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}\left(g(t)^2\right). \tag{5.9}$$

This shows the obvious interest in formulating the damping coefficient in the form $g(t)\left(1 + \frac{d}{dt}\left(\frac{1}{g}\right)(t)\right)$. There is no loss of generality, and the conditions for obtaining rapid convergence results can be formulated directly on the data g. For example, let's start with $e(t) = \frac{\alpha-1}{t}$. From (5.8), we immediately obtain $\gamma(t) = \frac{\alpha}{t}$. Then, (5.9) shows that the well-known condition $\alpha \geq 3$ provides the $\mathcal{O}\left(\frac{1}{t^2}\right)$ convergence rate of values (see [5, 39]).
Following [6, Theorem 1], when $\beta_k \equiv \beta > 0$ is fixed, we obtain a fast convergence of the values under the condition $\forall k \geq 1$, $t_{k+1}^2 - t_k^2 \leq t_{k+1}$ where

$t_k := 1 + \sum_{i=k}^{+\infty} \prod_{j=k}^{i} \alpha_j$. This condition can be equivalently formulated in terms of g_k as

$$g_k \geq 1 - \left(\frac{g_k}{g_{k-1}}\right)^2, \tag{5.10}$$

with the corresponding convergence rate of the values

$$\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(g_k^2\right). \tag{5.11}$$

Starting from $g_k = \frac{\alpha-1}{k}$, (5.8) immediately gives $\gamma_k = \frac{\alpha}{k}$. Then (5.10)-(5.11) shows that the well-known condition $\alpha \geq 3$ provides the $\mathcal{O}\left(\frac{1}{k^2}\right)$ convergence rate of values, a classical result (see [8, 13, 21, 39]).

6. A class of Inexact inertial proximal algorithms

Motivated by the results above, we consider the proximal algorithms that can be obtained (when Φ is smooth) by various temporal discretizations of the second-order evolution equation

$$\ddot{x}(t) + g(t)\left(1 + \frac{d}{dt}\left(\frac{1}{g}\right)(t)\right)\dot{x}(t) + \lambda(t)\nabla\Phi(x(t)) = e(t), \tag{6.1}$$

where, as in section 4, $e(t)$ stands for a perturbation or error term. We'll see that the convergence analysis for the inertial proximal algorithm can be developed within this setting. As a major advantage, the convergence results can be expressed directly on the parameters describing the algorithm.

6.1. **A parametrized family of proximal inertial algorithms.** Let us start from the second-order evolution equation (6.1) and introduce various temporal discretizations. When considering the implicit discretization for the potential term, which gives proximal algorithms, we can take a general convex lower semicontinuous proper function Φ. As novelty, we introduce a parameter $\theta \in [0,1]$ which takes into account different discretizations of the damping term: for $k \geq 1$,

$$\begin{aligned}(x_{k+1} - 2x_k + x_{k-1}) + g_k(1-\theta)\left(1 + \frac{1}{g_{k+1}} - \frac{1}{g_k}\right)(x_{k+1} - x_k)\\ + g_k\theta\left(1 + \frac{1}{g_k} - \frac{1}{g_{k-1}}\right)(x_k - x_{k-1}) + \lambda_k \partial\Phi(x_{k+1}) \ni e_k.\end{aligned}$$

After dividing by g_k, we obtain

$$\begin{aligned}&\left(\frac{1}{g_k} + (1-\theta)\left(1 + \frac{1}{g_{k+1}} - \frac{1}{g_k}\right)\right)(x_{k+1} - x_k)\\ &\quad - \left(\frac{1}{g_k} + \theta\left(\frac{1}{g_{k-1}} - \frac{1}{g_k} - 1\right)\right)(x_k - x_{k-1}) + \frac{\lambda_k}{g_k}\partial\Phi(x_{k+1}) \ni \frac{1}{g_k}e_k.\end{aligned}$$

Set, for $k \geq 1$, $\theta_k := \frac{1}{g_k} + \theta\left(\frac{1}{g_{k-1}} - \frac{1}{g_k} - 1\right)$, then

$$\frac{1}{g_k} + (1-\theta)\left(1 + \frac{1}{g_{k+1}} - \frac{1}{g_k}\right) = 1 + \theta_{k+1}.$$

So, we can reformulate the above algorithm in the condensed form

$$(6.2)\quad x_{k+1} + \frac{\lambda_k}{g_k(1+\theta_{k+1})}\partial\Phi(x_{k+1}) \ni x_k + \frac{\theta_k}{1+\theta_{k+1}}(x_k - x_{k-1}) + \frac{1}{g_k(1+\theta_{k+1})}e_k.$$

The following formulation of the algorithm (6.2) combines additive errors with the use of ϵ-subgradients. It extends the framework of the Inertial Proximal algorithm studied in [11]. One can consult [37] and [40] for related results concerning the introduction of errors in proximal based algorithms. Setting

$$r_k := \tfrac{1}{g_k(1+\theta_{k+1})}e_k,$$

we obtain

Inexact Inertial Proximal algorithm.

$$(\mathrm{IP})_{\alpha_k,\beta_k,\epsilon_k,r_k} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} \approx \operatorname{prox}^{\epsilon_k}_{\beta_k\Phi}(y_k - r_k), \end{cases}$$

$$\alpha_k := \tfrac{\theta_k}{1+\theta_{k+1}};\quad \beta_k := \tfrac{\lambda_k}{g_k(1+\theta_{k+1})};$$

$$\theta_k := \tfrac{1}{g_k} + \theta\left(\tfrac{1}{g_{k-1}} - \tfrac{1}{g_k} - 1\right).$$

By definition of the inexact proximal operator, the iteration at step k of $(\mathrm{IP})_{\alpha_k,\beta_k,\epsilon_k}$ can be written as

$$\frac{1}{\beta_k}(y_k - x_{k+1} - r_k) \in \partial_{\epsilon_k}\Phi(x_{k+1}),$$

where $\partial_{\epsilon_k}\Phi(x) := \{u \in \mathcal{H} : \Phi(x) \le \Phi(y) - \langle u, y-x\rangle + \epsilon_k,\ \forall y \in \mathcal{H}\}$.

6.2. **Convergence rates.** The objective of this section is to study the rapid convergence of values for sequences generated by the algorithm $(\mathrm{IP})_{\alpha_k,\beta_k,\epsilon_k}$.

Theorem 6.1. *Consider the algorithm* $(\mathrm{IP})_{\alpha_k,\beta_k,\epsilon_k,r_k}$ *and suppose that* $0 < g_k \le 1$, $0 \le \theta \le 1$, *and the parameters* (g_k), (λ_k) *and* θ *satisfy the growth condition: there exists* $k_1 \in \mathbb{N}$ *such that for all* $k \ge k_1$

$$(K_{g_k,\lambda_k,\theta})\qquad \lambda_{k+1} \le \frac{g_{k+1}}{g_k}\frac{\theta_{k+1}+1}{\theta_{k+2}}\lambda_k.$$

Suppose that the sequences $(r_k) \subset \mathcal{H}$, $(\epsilon_k) \subset \mathbb{R}_+$ *satisfy the summability properties*

$$(6.3)\qquad \sum_k (1+\theta_{k+1})\|r_k\| < +\infty \text{ and } \sum_k (1+\theta_{k+1})\frac{\lambda_k\epsilon_k}{g_k} < +\infty.$$

Then, for any sequence (x_k) *generated by the algorithm* $(\mathrm{IP})_{\alpha_k,\beta_k,\epsilon_k,r_k}$, *we have*

$$\begin{cases} (i)\ \Phi(x_k) - \min_{\mathcal{H}}\Phi = \mathcal{O}\left(\dfrac{g_{k-1}}{\lambda_{k-1}(1+\theta_k)}\right), \quad \textit{as } k \to +\infty; \\ (ii)\ \sum_{k\ge1}\beta_{k,\theta}\left(\Phi(x_k) - \min_{\mathcal{H}}\Phi\right) < +\infty \\ \qquad \textit{where } \beta_{k,\theta} := \frac{\lambda_{k-1}}{g_{k-1}}(1+\theta_k) - \frac{\lambda_k}{g_k}\theta_{k+1} \ \textit{ is non-negative by } (K_{g_k,\lambda_k,\theta}). \end{cases}$$

Proof. To make the presentation simpler, without loss of generality, we take $k_1 = 1$. By definition of the proximal operator, the iteration at step k of the algorithm $(\mathrm{IP})_{\alpha_k,\beta_k,\epsilon_k,r_k}$ writes

$$\frac{1}{\beta_k}(y_k - x_{k+1} - r_k) \in \partial_{\epsilon_k}\Phi(x_{k+1}).$$

Equivalently, we have the following subdifferential inequalities: for any $x \in \mathcal{H}$

$$\Phi(x) + \epsilon_k \geq \Phi(x_{k+1}) + \frac{1}{\beta_k}\left(\langle x - x_{k+1}, y_k - x_{k+1}\rangle - \langle x - x_{k+1}, r_k\rangle\right). \tag{6.4}$$

Let us write successively inequality (6.4) at $x = x_k$ and $x = x^* \in \operatorname{argmin}\Phi$. We obtain the two inequalities

$$\Phi(x_k) + \epsilon_k \geq \Phi(x_{k+1}) + \frac{1}{\beta_k}(\langle x_k - x_{k+1}, y_k - x_{k+1}\rangle - \langle x_k - x_{k+1}, r_k\rangle), \tag{6.5}$$

$$\Phi(x^*) + \epsilon_k \geq \Phi(x_{k+1}) + \frac{1}{\beta_k}(\langle x^* - x_{k+1}, y_k - x_{k+1}\rangle - \langle x^* - x_{k+1}, r_k\rangle). \tag{6.6}$$

Using $x_k - x_{k+1} = x_k - y_k + y_k - x_{k+1}$ in (6.5) and $x^* - x_{k+1} = x^* - y_k + y_k - x_{k+1}$ in (6.6) we obtain

$$\begin{aligned} \Phi(x_k) + \epsilon_k &\geq \Phi(x_{k+1}) \\ &+ \frac{1}{\beta_k}(\langle x_k - y_k, y_k - x_{k+1}\rangle - \langle x_k - x_{k+1}, r_k\rangle + \|y_k - x_{k+1}\|^2), \end{aligned} \tag{6.7}$$

$$\begin{aligned} \Phi(x^*) + \epsilon_k &\geq \Phi(x_{k+1}) \\ &+ \frac{1}{\beta_k}(\langle x^* - y_k, y_k - x_{k+1}\rangle - \langle x^* - x_{k+1}, r_k\rangle + \|y_k - x_{k+1}\|^2). \end{aligned} \tag{6.8}$$

Multiplying (6.7) by $\dfrac{\theta_k}{\alpha_k} - 1 \geq 0$, then adding (6.8), we derive that

$$\begin{aligned} \frac{\epsilon_k\theta_k}{\alpha_k} + \left(\frac{\theta_k}{\alpha_k} - 1\right)(\Phi(x_k) - \Phi(x^*)) &\geq \frac{\theta_k}{\alpha_k}(\Phi(x_{k+1}) - \Phi(x^*)) + \frac{\theta_k}{\alpha_k\beta_k}\|y_k - x_{k+1}\|^2 \\ &+ \frac{1}{\beta_k}\left\langle x_{k+1} - y_k, \left(\frac{\theta_k}{\alpha_k} - 1\right)(y_k - x_k) + y_k - x^*\right\rangle \\ &+ \frac{1}{\beta_k}\left\langle \left(\frac{\theta_k}{\alpha_k} - 1\right)(x_{k+1} - x_k) + x_{k+1} - x^*, r_k\right\rangle. \end{aligned} \tag{6.9}$$

By definition of y_k we have

$$\begin{aligned} \left(\frac{\theta_k}{\alpha_k} - 1\right)(y_k - x_k) + y_k &= \left(\frac{\theta_k}{\alpha_k} - 1\right)\alpha_k(x_k - x_{k-1}) + x_k + \alpha_k(x_k - x_{k-1}) \\ &= x_k + \theta_k(x_k - x_{k-1}) = z_k \end{aligned}$$

where $z_k := x_k + \theta_k(x_k - x_{k-1})$. Moreover

$$\left(\frac{\theta_k}{\alpha_k} - 1\right)(x_{k+1} - x_k) + x_{k+1} = \theta_{k+1}(x_{k+1} - x_k) + x_{k+1} = z_{k+1}.$$

We then deduce from (6.9) that

$$\begin{aligned}\frac{\epsilon_k\theta_k}{\alpha_k} &+ \left(\frac{\theta_k}{\alpha_k}-1\right)(\Phi(x_k)-\Phi(x^*)) \geq \frac{\theta_k}{\alpha_k}(\Phi(x_{k+1})-\Phi(x^*))\\ &+ \frac{1}{\beta_k}\langle x_{k+1}-y_k, z_k-x^*\rangle + \frac{1}{\beta_k}\langle r_k, z_{k+1}-x^*\rangle + \frac{\theta_k}{\alpha_k\beta_k}\|y_k-x_{k+1}\|^2.\end{aligned}$$

Equivalently, after multiplication by β_k

$$\begin{aligned}\frac{\epsilon_k\theta_k\beta_k}{\alpha_k} &+ \beta_k\left(\frac{\theta_k}{\alpha_k}-1\right)(\Phi(x_k)-\Phi(x^*)) \geq \frac{\beta_k\theta_k}{\alpha_k}(\Phi(x_{k+1})-\Phi(x^*))\\ &+ \langle x_{k+1}-y_k, z_k-x^*\rangle + \langle r_k, z_{k+1}-x^*\rangle + \frac{\theta_k}{\alpha_k}\|y_k-x_{k+1}\|^2.\end{aligned} \tag{6.10}$$

To write (6.10) in a recursive form, we use $z_{k+1}-z_k = (1+\theta_{k+1})(x_{k+1}-y_k)$. It ensues that

$$\|z_{k+1}-x^*\|^2 = \|z_k-x^*\|^2 + 2(1+\theta_{k+1})\langle x_{k+1}-y_k, z_k-x^*\rangle + (1+\theta_{k+1})^2\|x_{k+1}-y_k\|^2,$$

which gives

$$\begin{aligned}\langle x_{k+1}-y_k, z_k-x^*\rangle = &\frac{1}{2(1+\theta_{k+1})}\left(\|z_{k+1}-x^*\|^2-\|z_k-x^*\|^2\right)\\ &-\frac{(1+\theta_{k+1})}{2}\|x_{k+1}-y_k\|^2.\end{aligned}$$

Using this equality in (6.10), we obtain

$$\begin{aligned}\frac{\epsilon_k\theta_k\beta_k}{\alpha_k} + \beta_k\left(\frac{\theta_k}{\alpha_k}-1\right)(\Phi(x_k)-\Phi(x^*)) &\geq \frac{\beta_k\theta_k}{\alpha_k}(\Phi(x_{k+1})-\Phi(x^*)) + \langle r_k, z_{k+1}-x^*\rangle\\ + \frac{1}{2(1+\theta_{k+1})}&\left(\|z_{k+1}-x^*\|^2-\|z_k-x^*\|^2\right) + \frac{(1+\theta_{k+1})}{2}\|x_{k+1}-y_k\|^2,\end{aligned}$$

where we have used $\frac{\theta_k}{\alpha_k}-\frac{(1+\theta_{k+1})}{2}=\frac{(1+\theta_{k+1})}{2}$ (a consequence of the definition of α_k). After multiplication by $(1+\theta_{k+1})$, and neglecting the non-negative term $\frac{(1+\theta_{k+1})}{2}\|x_{k+1}-y_k\|^2$, we obtain

$$\begin{aligned}&\frac{\epsilon_k\theta_k\beta_k}{\alpha_k}(1+\theta_{k+1}) + \beta_k(1+\theta_{k+1})\left(\frac{\theta_k}{\alpha_k}-1\right)(\Phi(x_k)-\Phi(x^*)) + \frac{1}{2}\|z_k-x^*\|^2\\ &\geq \frac{\beta_k(1+\theta_{k+1})\theta_k}{\alpha_k}(\Phi(x_{k+1})-\Phi(x^*))\\ &+\frac{1}{2}\|z_{k+1}-x^*\|^2 + (1+\theta_{k+1})\langle r_k, z_{k+1}-x^*\rangle.\end{aligned} \tag{6.11}$$

According to $\beta_k(1+\theta_{k+1}) = \frac{\lambda_k}{g_k}$, and $\frac{\theta_k}{\alpha_k}-1=\theta_{k+1}$ we have

$$\beta_k(1+\theta_{k+1})\left(\frac{\theta_k}{\alpha_k}-1\right) = \frac{\lambda_k}{g_k}\theta_{k+1}.$$

Hence, (6.11) can be equivalently written as

$$\frac{\lambda_k}{g_k}(1+\theta_{k+1})\epsilon_k + \frac{\lambda_k}{g_k}\theta_{k+1}(\Phi(x_k)-\Phi(x^*)) + \frac{1}{2}\|z_k - x^*\|^2$$
$$\geq \frac{\lambda_k}{g_k}(1+\theta_{k+1})(\Phi(x_{k+1})-\Phi(x^*)) + \frac{1}{2}\|z_{k+1}-x^*\|^2 + (1+\theta_{k+1})\langle r_k, z_{k+1}-x^*\rangle.$$

This naturally leads us to introduce the sequence $(\mathcal{E}_k)$

$$\mathcal{E}_k = \frac{\lambda_{k-1}}{g_{k-1}}(1+\theta_k)(\Phi(x_k)-\Phi(x^*)) + \frac{1}{2}\|z_k - x^*\|^2. \tag{6.12}$$

Thus, we have obtained the following inequality

$$\begin{aligned} \mathcal{E}_k + \frac{\lambda_k}{g_k}(1+\theta_{k+1})\epsilon_k \quad \geq \quad & \mathcal{E}_{k+1} + \left(\frac{\lambda_{k-1}}{g_{k-1}}(1+\theta_k) - \frac{\lambda_k}{g_k}\theta_{k+1}\right)(\Phi(x_k) - \min_{\mathcal{H}}\Phi) \\ & +(1+\theta_{k+1})\langle r_k, z_{k+1}-x^*\rangle. \end{aligned} \tag{6.13}$$

Under condition $(K_{g_k,\lambda_k,\theta})$ we have $\frac{\lambda_{k-1}}{g_{k-1}}(1+\theta_k) - \frac{\lambda_k}{g_k}\theta_{k+1} \geq 0$. Hence,

$$\mathcal{E}_{k+1} \leq \mathcal{E}_k + (1+\theta_{k+1})\|r_k\| \cdot \|z_{k+1}-x^*\| + \frac{\lambda_k}{g_k}(1+\theta_{k+1})\epsilon_k. \tag{6.14}$$

By summing inequalities (6.14) from $j=1$ to $k-1$, we obtain

$$\mathcal{E}_k \leq \mathcal{E}_1 + \sum_{j=1}^{k-1}(1+\theta_{j+1})\|r_j\| \cdot \|z_{j+1}-x^*\| + \sum_{j=1}^{k-1}(1+\theta_{j+1})\frac{\lambda_j\epsilon_j}{g_j}. \tag{6.15}$$

Since $\mathcal{E}_k \geq \frac{1}{2}\|z_k - x^*\|^2$ and $A := \sum(1+\theta_{j+1})\frac{\lambda_j\epsilon_j}{g_j} < +\infty$, we deduce that

$$\|z_k - x^*\|^2 \leq 2\mathcal{E}_1 + 2A + \sum_{j=1}^{k} 2(1+\theta_j)\|r_{j-1}\| \cdot \|z_j - x^*\|. \tag{6.16}$$

Let us apply the Gronwall Lemma 8.4 with $a_j = \|z_j - x^*\|$, $b_j = 2(1+\theta_j)\|r_{j-1}\|$, and $c = \sqrt{2(\mathcal{E}_1 + A)}$. According to assumption (6.3), we obtain

$$\|z_k - x^*\| \leq c + \sum_{j=1}^{\infty} 2(1+\theta_j)\|r_{j-1}\| < +\infty.$$

Returning to (6.15), we deduce from the convergence of the series $B := \sum(1 + \theta_{j+1})\|r_j\|$ that

$$\mathcal{E}_k \leq C := \mathcal{E}_1 + B\,(c+B) + A < +\infty. \tag{6.17}$$

By definition of $\mathcal{E}_k$, we obtain, for all $k \geq k_1$

$$\frac{\lambda_{k-1}}{g_{k-1}}(1+\theta_k)(\Phi(x_k)-\Phi(x^*)) \leq \mathcal{E}_k \leq C,$$

which gives the claim. The last item follows directly by summing (6.13). $\square$

Depending on the choice of the parameter θ, we obtain a specific algorithm, with its convergence rate. Let's consider the following cases of particular interest:

a) Case $\theta = 1$ corresponds to the explicit discretization of the damping term:
(6.18)

$$(x_{k+1}-2x_k+x_{k-1})+g_k\left(1+\frac{1}{g_k}-\frac{1}{g_{k-1}}\right)(x_k-x_{k-1})+\beta_k\partial_{\epsilon_k}\Phi(x_{k+1})\ni -\frac{g_kg_{k-1}}{1-g_{k-1}}e_k,$$

since, for $\theta = 1$ in the formula giving the parameters, we have

$$\theta_k := \frac{1}{g_k}+\left(\frac{1}{g_{k-1}}-\frac{1}{g_k}-1\right)=\frac{1}{g_{k-1}}-1,\ \theta_{k+1}+1=\frac{1}{g_k} \text{ and } \beta_k=\lambda_k.$$

So, the formula $(K_{g_k,\lambda_k,\theta})$ in Theorem 6.1 becomes

$$\beta_{k+1}\leq \frac{g_{k+1}}{g_k}\frac{\frac{1}{g_k}}{\frac{1}{g_{k+1}}-1}\beta_k=\frac{g_{k+1}}{g_k^2(\frac{1}{g_{k+1}}-1)}\beta_k. \tag{6.19}$$

So we recover the same growth condition as in [11, Theorem 4]. Let us now compare the convergence rates. Theorem 6.1 gives

$$\Phi(x_k)-\min_{\mathcal{H}}\Phi=\mathcal{O}\left(\frac{g_{k-1}}{\lambda_{k-1}(1+\theta_k)}\right).$$

From $\theta_k+1=\frac{1}{g_{k-1}}$ we get $\frac{g_{k-1}}{\lambda_{k-1}(1+\theta_k)}=\frac{g_{k-1}^2}{\lambda_{k-1}}=\frac{1}{t_k^2\beta_{k-1}}$, and then we recover the same convergence rate as in [11, Theorem 4].

b) Case $\theta = 0$ corresponds to the implicit discretization of the damping term
(6.20)

$$(x_{k+1}-2x_k+x_{k-1})+g_k\left(1+\frac{1}{g_{k+1}}-\frac{1}{\alpha_k}\right)(x_{k+1}-x_k)+\beta_k\partial\Phi(x_{k+1})\ni -\frac{g_kg_{k-1}}{1-g_{k-1}}e_k.$$

We have $\theta_k=\frac{1}{g_k}$, $\alpha_k=\frac{g_{k+1}}{g_k(1+g_{k+1})}$ and $\beta_k=\frac{\lambda_kg_{k+1}}{g_k(1+g_{k+1})}$, which by Theorem 6.1 give that, under the condition $(K_{g_k,\lambda_k,\theta})$, i.e., $\lambda_{k+1}\leq\frac{(1+g_{k+1})g_{k+2}}{g_k}\lambda_k$, we have

$$\Phi(x_k)-\min_{\mathcal{H}}\Phi=\mathcal{O}\left(\frac{g_kg_{k-1}}{\lambda_{k-1}(1+g_k)}\right),\ \text{ as } k\to+\infty.$$

Consider the case $g_k=\frac{\alpha-1}{k-1}$. This gives $\alpha_k=\frac{k-1}{k+\alpha-1}$, which corresponds to a variant of the Nesterov acceleration scheme considered by several authors (see [8], [21], [39]). An elementary calculation shows that the growth condition above and the corresponding convergence rate give results comparable to those of the explicit case.

7. Perspectives

In general, the presence of oscillations is not a desirable property for optimization problems. In this respect, in order to improve the inertial methods, various strategies have recently been developed. It would be interesting to combine them with time scaling. Let us list some of them.

- To overcome the presence of wild oscillations (which may occur for ill-conditioned minimization problems), one has to consider a damping which takes into account the geometry of Φ. In this direction, the Hessian-driven damping

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla^2\Phi(x(t))\dot{x}(t) + \nabla\Phi(x(t)) = 0, \tag{7.1}$$

  combines the Nesterov acceleration with the Newton method, see [2], [14]. At first glance, the presence of the Hessian may seem like a numerical difficulty. The crucial point is that the Hessian comes in the form $\nabla^2\Phi(x(t))\dot{x}(t)$, which is equal to the derivative of $\nabla\Phi(x(t))$. As a consequence, the temporal discretization of this dynamics provides first-order algorithms. Time scaling in this context would lead to consider the Hessian-driven damping dynamics

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2\Phi(x(t))\dot{x}(t) + b(t)\nabla\Phi(x(t)) = 0.$$

- To avoid oscillations, while retaining the advantage of the inertial effect, the restarting method considers the damping coefficient as a control variable. The strategy is to maintain a high speed along the orbit by stopping the dynamic when the speed begins to decrease. After stopping, restart with zero speed, see Su-Boyd-Candès (2016). As well, time scaling in this framework is an interesting subject to study.
- In [22] Ghisi-Gobbino-Haraux consider $t \mapsto \gamma(t)$ as a pulsating function that alternates big and small values in a suitable way. They prove the effectiveness of the method for quadratic minimization, and apply it to ordinary differential equations and partial differential equations of hyperbolic type.
- In the above approaches, the damping is considered as a control variable. In this respect, it would be interesting to consider the damping as a closed-loop control, as opposed to the open-loop approach developed in most of the papers devoted to inertial methods in optimization.
- In our study of the continuous dynamic, the potential function Φ to minimize has been assumed to be continuously differentiable. Indeed, the entire study can be performed to deal with nonsmooth functions, simply replacing Φ with its Moreau envelopes. This operation preserves the optimal value and optimal set, and leads to relaxed proximal algorithms whose numerical complexity is the same. Without time scaling, this approach has been recently developed by Attouch-Peypouquet in [12].

8. Appendix

In what follows, we prove the existence and the uniqueness of a global solution to the Cauchy problem associated with the evolution system $(\mathrm{IGS})_{\gamma,\beta,e}$ that we recall below

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla\Phi(x(t)) = e(t). \tag{$(\mathrm{IGS})_{\gamma,\beta,e}$}$$

We will use the following lemma.

Lemma 8.1 ([25, Prop. 6.2.1])**.** *Let $F : I \times \mathcal{X} \to \mathcal{X}$ where $I = [t_0, +\infty[$ and $\mathcal{X}$ is a Banach space. Assume that*

(i) *for every $x \in \mathcal{X}$, $F(\cdot, x) \in L^1_{loc}(I, \mathcal{X})$;*

(ii) *for a.e. $t \in I$, for every $x, y \in \mathcal{X}$,*

$$\|F(t,x) - F(t,y)\| \le K(t, \|x\| + \|y\|)\|x - y\|, \text{ where } K(\cdot, r) \in L^1_{loc}(I), \forall r \in \mathbb{R}_+;$$

(iii) *for a.e. $t \in I$, for every $x \in \mathcal{X}$,*

$$\|F(t,x)\| \le P(t)(1 + \|x\|), \text{ where } P \in L^1_{loc}(I).$$

Then, for every $s \in I, x \in \mathcal{X}$, there exists a unique solution $u_{s,x} \in W^{1,1}_{loc}(I, \mathcal{X})$ of the Cauchy problem:

$$\dot{u}_{s,x}(t) = F(t, u_{s,x}(t)) \text{ for a.e. } t \in I, \text{ and } u_{s,x}(s) = x.$$

For simplicity, we give a short proof in the case where the gradient of Φ is Lipschitz continuous.

Theorem 8.2. *Suppose that $\Phi : \mathcal{H} \to \mathbb{R}$ is convex, $\mathcal{C}^1$, with Lipschitz continuous gradient $\nabla\Phi$. Assume that $\beta, \gamma : [t_0, +\infty[\to \mathbb{R}^*_+$ and $e : [t_0, +\infty[\to \mathcal{H}$ are locally integrable. Then, the evolution system* $(\text{IGS})_{\gamma,\beta,e}$*, with initial condition $(x(t_0), \dot{x}(t_0)) = (x_0, \dot{x}_0) \in \mathcal{H} \times \mathcal{H}$, admits a unique global solution $x : [t_0, +\infty[\to \mathcal{H}$.*

Proof. To prove the existence and uniqueness for the Cauchy problem$(\text{IGS})_{\gamma,\beta,e}$ with initial condition $(x(t_0), \dot{x}(t_0)) = (x_0, \dot{x}_0)$, we formulate it in the phase space. Set $I = [t_0, +\infty[$, and define $F : I \times \mathcal{H} \times \mathcal{H} \to \mathcal{H}$ by

$$F(t, x, y) = (y, e(t) - \beta(t)\nabla\Phi(x) - \gamma(t)y).$$

Set $u(t) = (x(t), y(t))$. The Cauchy problem for $(\text{IGS})_{\gamma,\beta,e}$ can be equivalently formulated as

$$\begin{cases} \dot{u}(t) &= F(t, u(t)) \quad \text{for a.e. } t \in I, \\ u(t_0) &= (x_0, \dot{x}_0). \end{cases} \tag{8.1}$$

Let us verify the three conditions of Lemma 8.1.

(i) For each $(x, y) \in \mathcal{H} \times \mathcal{H}$, $F(\cdot, x, y) \in L^1_{loc}(I, \mathcal{H})$, since the functions e, β and γ are so.

(ii) Denote by L the Lipschitz constant of $\nabla\Phi$. For every $u = (x, y), u' = (x', y') \in \mathcal{H} \times \mathcal{H}$ and a.e. $t \in I$

$$\begin{aligned} \|F(t,u) - F(t,u')\| &= \|y - y'\| + \|\beta(t)(\nabla\Phi(x) - \nabla\Phi(x')) + \gamma(t)(y - y')\| \\ &\le (1 + L\beta(t) + \gamma(t))\left(\|x - x'\| + \|y - y'\|\right) \\ &= (1 + L\beta(t) + \gamma(t))\|(x, y) - (x', y')\| \end{aligned}$$

and then the second condition is verified, since the real function $t \mapsto 1 + L\beta(t) + \gamma(t)$ belongs to $L^1_{loc}(I, \mathbb{R})$.

(iii) For every $u = (x, y) \in \mathcal{H} \times \mathcal{H}$ and a.e. $t \in I$

$$\begin{aligned} \|F(t,u)\| &= \|y\| + \|\beta(t)(\nabla\Phi(x) - \nabla\Phi(x_0)) + \beta(t)\nabla\Phi(x_0) + \gamma(t)y + e(t)\| \\ &\le (1 + \gamma(t))\|y\| + L\beta(t)\|x - x_0\| + \beta(t)\|\nabla\Phi(x_0)\| + \|e(t)\| \\ &\le \max(1 + \gamma(t), L\beta(t), \beta(L\|x_0\| + \|\nabla\Phi(x_0)\| + \|e(t)\|)\,(1 + \|x\| + \|y\|) \\ &= r(t)(1 + \|u\|), \end{aligned}$$

where $r(t) = \max(1+\gamma(t), L\beta(t), \beta(L\|x_0\| + \|\nabla\Phi(x_0)\| + \|e(t)\|)$. Since $r(\cdot) \in L^1_{loc}(I,\mathbb{R})$, we conclude that all the conditions of Lemma 8.1 are satisfied. So, there exists a unique global solution of $(\mathrm{IGS})_{\gamma,\beta,e}$ satisfying the initial condition $(x(t_0),\dot{x}(t_0)) = (x_0,\dot{x}_0)$. $\square$

Lemma 8.3. ([30]) *Let S be a nonempty subset of $\mathcal{H}$ and let $x : [t_0,+\infty[\to \mathcal{H}$. Assume that*
(i) for every $z \in S$, $\lim_{t\to\infty} \|x(t) - z\|$ exists;
(ii) every weak sequential cluster point of $x(t)$, as $t \to \infty$, belongs to S.
Then $x(t)$ converges weakly as $t \to \infty$ to a point in S.

Lemma 8.4 ([8, Lemma 5.14])**.** *Let (a_k) be a sequence of non-negative numbers such that, for all $k \in \mathbb{N}$, $a_k^2 \le c^2 + \sum_{j=1}^k b_j a_j$, where (b_j) is a summable sequence of non-negative numbers, and $c \ge 0$. Then, for all $k \in \mathbb{N}$, $a_k \le c + \sum_{j=1}^{\infty} b_j$.*

References

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

[2] F. Álvarez, H. Attouch, J. Bolte and P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl. **81** (2002), 747–779.

[3] V. Apidopoulos, J.-F. Aujol and Ch. Dossal, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule*, Mathematical Programming **180** (2020), 137–156

[4] V. Apidopoulos, J.-F. Aujol, Ch. Dossal and A. Rondepierre, *Convergence rates of an inertial gradient descent algorithm under growth and flatness condition*, Mathematical Programming **187** (2021), 151–193

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

[6] H. Attouch and A. Cabot, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., **28** (2018), 849–874.

[7] H. Attouch, A. Cabot, Z. Chbani and H. Riahi, *Rate of convergence of inertial gradient dynamics with time-dependent viscous damping coefficient*, Evolution Equations and Control Theory **7** (2018), 353–371.

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

[9] H. Attouch, Z. Chbani and H. Riahi, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case* $\alpha \le 3$, ESAIM: COCV, **25** (2019).

[10] H. Attouch, Z. Chbani and H. Riahi, *Fast proximal methods via time scaling of damped inertial dynamics*, SIAM J. Optim. **29** (2019), 2227–2256.

[11] H. Attouch, Z. Chbani and H. Riahi, *Convergence rates of inertial proximal algorithms with general extrapolation and proximal coefficients*, Vietnam Journal of Mathematics **48** (2020), 247–276.

[12] H. Attouch and J. Peypouquet, *Convergence rate of proximal inertial algorithms associated with Moreau envelopes of convex functions*, in: Splitting Algorithms, Modern Operator Theory, and Applications, H. Bauschke, R. Burachik, R. Luke (eds), Springer, 2019.

[13] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$, SIAM J. Optim. **26** (2016), 1824–1834.

[14] H. Attouch, J. Peypouquet and P. Redont, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations **261** (2016), 5734–5783.

[15] J.-F. Aujol and Ch. Dossal, *Stability of over-relaxations for the Forward-Backward algorithm, application to FISTA*, SIAM J. Optim. **25** (2015), 2408–2433.

[16] J.-F. Aujol and Ch. Dossal, *Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for $b > 0$*, 2017, https://hal.inria.fr/hal-01547251v2.

[17] H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CSM Books in Mathematics, Springer, 2011.

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

[19] R. I. Bot and E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM J. Control Optim. **54** (2016), 1423–1443.

[20] A. Cabot, H. Engler and S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation* Transactions of the American Mathematical Society **361** (2009), 5983–6017.

[21] A. Chambolle and Ch. Dossal, *On the convergence of the iterates of the Fast Iterative Shrinkage Thresholding Algorithm*, Journal of Optimization Theory and Applications **166** (2015), 968–982.

[22] M. Ghisi, M. Gobbino and A. Haraux, *The remarkable effectiveness of time-dependent damping terms for second order evolution equations*, SIAM J. Control Optim. **54** (2016), 1266–1294.

[23] O. Güler, *On the convergence of the proximal point algorithm for convex optimization*, SIAM J. Control Optim. **29** (1991), 403–419.

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

[25] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris, 1991.

[26] B. Martinet, *Brève communication. Régularisation d'inéquations variationnelles par approximations successives*, ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique **4** (1970), 154–158.

[27] R. May, *Asymptotic for a second-order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Mathematics **41** (2017), 681–685.

[28] Y. Nesterov, *A method of solving a convex programming problem with convergence rate O(1/k2)*, Soviet Mathematics Doklady, **27** (1983), 372–376.

[29] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004.

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

[31] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in Optimization **1** (2013), 123–231.

[32] J. Peypouquet, *Convex Optimization in Normed Spaces: Theory, Methods and Examples*, Springer, 2015.

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

[34] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. Math. Phys. **4** (1964), 1–17.

[35] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.

[36] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976), 877–898.

[37] M. Schmidt, N. Le Roux and F. Bach, *Convergence rates of inexact proximal-gradient methods for convex optimization*, NIPS'11 - 25 th Annual Conference on Neural Information Processing Systems, Dec 2011, Grenada, Spain. (2011) HAL inria-00618152v3.

[38] B. Shi, S. S. Du, M. I. Jordan and W. J. Su, *Understanding the acceleration phenomenon via high-resolution differential equations*, Math. Program, (2021), https://doi.org/10.1007/s10107-021-01681-8.

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

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

[41] A. Wibisono, A.C. Wilson and M. I. Jordan, *A variational perspective on accelerated methods in optimization*, Proceedings of the National Academy of Sciences, **113** (2016), E7351–E7358.

Manuscript received May 22 2019
revised September 9 2019

H. Attouch
IMAG, Univ. Montpellier, CNRS, Montpellier, France
E-mail address: hedy.attouch@umontpellier.fr

Z. Chbani
Cadi Ayyad university, Faculty of Sciences Semlalia, Mathematics, 40000 Marrakech, Morroco
E-mail address: chbaniz@uca.ac.ma

H. Riahi
Cadi Ayyad university, Faculty of Sciences Semlalia, Mathematics, 40000 Marrakech, Morroco
E-mail address: h-riahi@uca.ac.ma