



An Extension of the Second Order Dynamical System that Models Nesterov's Convex Gradient Method

Cristian Daniel Alecsa^{1,2} · Szilárd Csaba László³ · Titus Pința⁴

Published online: 8 June 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In this paper we deal with a general second order continuous dynamical system associated to a convex minimization problem with a Fréchet differentiable objective function. We show that inertial algorithms, such as Nesterov's algorithm, can be obtained via the natural explicit discretization from our dynamical system. Our dynamical system can be viewed as a perturbed version of the heavy ball method with vanishing damping, however the perturbation is made in the argument of the gradient of the objective function. This perturbation seems to have a smoothing effect for the energy error and eliminates the oscillations obtained for this error in the case of the heavy ball method with vanishing damping, as some numerical experiments show. We prove that the value of the objective function in a generated trajectory converges in order $\mathcal{O}(1/t^2)$ to the global minimum of the objective function. Moreover, we obtain that a trajectory generated by the dynamical system converges to a minimum point of the objective function.

This work was supported by a Grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2016-0266.

✉ Szilárd Csaba László
laszlosziszi@yahoo.com

Cristian Daniel Alecsa
alecsa@rist.ro

Titus Pința
titus.pinta@maths.ox.ac.uk

¹ Tiberiu Popoviciu Institute of Numerical Analysis, Romanian Academy, Cluj-Napoca, Romania

² Romanian Institute of Science and Technology, Cluj-Napoca, Romania

³ Department of Mathematics, Technical University of Cluj-Napoca, Str. Memorandumului nr. 28, 400114 Cluj-Napoca, Romania

⁴ Mathematical Institute, University of Oxford, Oxford, England

Keywords Convex optimization · Heavy ball method · Continuous second order dynamical system · Convergence rate · Inertial algorithm

Mathematics Subject Classification 34G20 · 47J35 · 90C25 · 90C30

1 Introduction

Since Su et al. in [32] showed that Nesterov's accelerated convex gradient method has the exact limit the second order differential equation that governs the heavy ball system with vanishing damping, that is,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla g(x(t)) = 0, \quad x(t_0) = u_0, \quad \dot{x}(t_0) = v_0, \quad t_0 > 0, \quad u_0, v_0 \in \mathbb{R}^m, \quad (1)$$

with $\alpha = 3$, the latter system has been intensively studied in the literature in connection to the minimization problem $\inf_{x \in \mathbb{R}^m} g(x)$. Here $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is a convex Fréchet differentiable function with Lipschitz continuous gradient.

In [32] the authors proved that

$$g(x(t)) - \min g = \mathcal{O}\left(\frac{1}{t^2}\right)$$

for every $\alpha \geq 3$, however they did not show the convergence of a generated trajectory to a minimum of the objective function g .

In [9], Attouch et al. considered the case $\alpha > 3$ in (1), and showed that the generated trajectory $x(t)$ converges to a minimizer of g as $t \rightarrow +\infty$. Actually in [9] the authors considered the perturbed version of the heavy ball system with vanishing damping, that is,

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla g(x(t)) = h(t), \quad x(t_0) = u_0, \quad \dot{x}(t_0) = v_0, \quad t_0 > 0, \quad u_0, v_0 \in \mathbb{R}^m, \quad (2)$$

where $h : [t_0, +\infty) \rightarrow \mathbb{R}$ is a small perturbation term that satisfies $\int_{t_0}^{+\infty} t \|h(t)\| dt < +\infty$. Beside the convergence of a generated trajectory $x(t)$ to a minimizer of g , they showed that also in this case the convergence rate of the objective function along the trajectory, that is $g(x(t)) - \min g$, is of order $\mathcal{O}\left(\frac{1}{t^2}\right)$.

Another perturbed version of (1) was studied by Attouch et al. in [8], see also [3,7,11,30]. They assumed that the objective g is twice continuously differentiable and the perturbation of their system is made at the damping term. More precisely, they studied the dynamical system with Hessian driven damping

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \nabla^2 g(x(t))\dot{x}(t) + \nabla g(x(t)) = 0, \quad x(t_0) = u_0, \quad \dot{x}(t_0) = v_0, \quad t_0 > 0, \quad u_0, v_0 \in \mathbb{R}^m, \quad (3)$$

where $\alpha > 0$ and $\beta > 0$. In case $\alpha > 3, \beta > 0$ they showed the convergence of a generated trajectory to a minimizer of g . Moreover, they obtained that in this case the convergence rate of the objective function along the trajectory, that is $g(x(t)) - \min g$, is of order $o\left(\frac{1}{t^2}\right)$.

Further, Attouch et al. in [10] studied the subcritical case $\alpha \leq 3$ and they proved that in this case the convergence rates of the objective function g along the trajectory generated by (1), i.e $g(x(t)) - \min g$, is of order $\mathcal{O}\left(\frac{1}{t^{\frac{2}{3}\alpha}}\right)$.

Another approach is due to Aujol et al. [14], who assumed that beside convexity, the objective g in (1) satisfies some geometrical conditions, such as the Łojasiewicz property. The importance of their results obtained in [14] is underlined by the fact that applying the classical Nesterov scheme on a convex objective function without studying its geometrical properties may lead to sub-optimal algorithms.

It is worth mentioning the work of Aujol and Dossal [13], who did not assumed the convexity of g , but the convexity of the function $(g(x(t)) - g(x^*))^\beta$, where β is strongly related to the damping parameter α and x^* is a global minimizer of g . Under these assumptions, they obtained some general convergence rates and also the convergence of the generated trajectories of (1). In case $\beta = 1$ they results reduce to the results obtained in [8–10].

However, the convergence of the trajectories generated by the continuous heavy ball system with vanishing damping (1), in the general case when g is non-convex is still an open question. Some important steps in this direction have been made in [19] (see also [18]), where convergence of the trajectories of a perturbed system, have been obtained in a non-convex setting. More precisely in [19] is considered the system

$$\ddot{x}(t) + \left(\gamma + \frac{\alpha}{t}\right)\dot{x}(t) + \nabla g(x(t)) = 0, \quad x(t_0) = u_0, \quad \dot{x}(t_0) = v_0, \quad (4)$$

where $t_0 > 0, u_0, v_0 \in \mathbb{R}^m, \gamma > 0, \alpha \in \mathbb{R}$. Note that here α can take non-positive values. For $\alpha = 0$ we recover the dynamical system studied in [16]. According to [19], the trajectory generated by the dynamical system (4) converges to a critical point of g , if a regularization of g satisfies the Kurdyka–Łojasiewicz property.

Further results concerning the heavy ball method and its extensions can be found in [4,6,15,21–23].

1.1 An Extension of the Heavy Ball Method and the Nesterov Type Algorithms Obtained via Explicit Discretization

What one can notice concerning the heavy ball system and its variants is, that despite of the result of Su et al. [32], these systems will never give through the natural implicit/explicit discretization the Nesterov algorithm. This is due to the fact that the gradient of g is evaluated in $x(t)$, and this via discretization will become $\nabla g(x_n)$ or $\nabla g(x_{n+1})$ and never of the form $\nabla g(x_n + \alpha_n(x_n - x_{n-1}))$ as Nesterov’s gradient method requires. Another observation is that using the same approach as in [32] one can show, see [25], that (1), (and also (4)), models beside Nesterov’s algorithm other algorithms too. In this paper we overcome the deficiencies emphasized above by intro-

ducing a dynamical system that via explicit discretization leads to inertial algorithms of gradient type. To this end, let us consider the optimization problem

$$(P) \quad \inf_{x \in \mathbb{R}^m} g(x) \quad (5)$$

where $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is a convex Fréchet differentiable, function with L_g -Lipschitz continuous gradient, i.e. there exists $L_g \geq 0$ such that $\|\nabla g(x) - \nabla g(y)\| \leq L_g \|x - y\|$ for all $x, y \in \mathbb{R}^m$.

We associate to (5) the following second order dynamical system:

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0, \end{cases} \quad (6)$$

where $u_0, v_0 \in \mathbb{R}^m$, $t_0 > 0$ and $\alpha > 0$, $\beta \in \mathbb{R}$, $\gamma \geq 0$.

Remark 1 The connection of (6) with the heavy ball system with vanishing damping (1) is obvious, the latter one can be obtained from (6) for $\gamma = \beta = 0$. Further, by using the Taylor expansion for $\nabla g(\cdot)$ we get

$$\nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \approx \nabla g(x(t)) + \left(\gamma + \frac{\beta}{t} \right) \nabla^2 g(x(t)) \dot{x}(t). \quad (7)$$

Hence, (6) is strongly related to the following version of heavy ball method with Hessian driven damping:

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \left(\gamma + \frac{\beta}{t} \right) \nabla^2 g(x(t)) \dot{x}(t) + \nabla g(x(t)) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0. \end{cases} \quad (8)$$

We emphasize that second order dynamical systems with Hessian driven damping similar to (8) have been intensively studied in the literature, see [3,7,8,11,21,30]

The study of the dynamical system (6) in connection to the optimization problem (5) is motivated by the following facts.

1. The dynamical system (6) leads via explicit discretization to inertial algorithms. In particular Nesterov's algorithm can be obtained via this natural discretization.
2. A generated trajectory and the objective function value in this trajectory in general have a better convergence behaviour than a trajectory generated by the heavy ball system with vanishing damping (1), as some numerical examples show.
3. The same numerical experiments reveal that the perturbation term $\left(\gamma + \frac{\beta}{t} \right) \dot{x}(t)$ in the argument of the gradient of the objective function g has a smoothing effect and annihilates the oscillations obtained in case of the dynamical system (1) for the errors $g(x(t)) - \min g$ and $\|x(t) - x^*\|$, where x^* is a minimizer of g . In view of Remark 1 this fact is not surprising, since according to [30] the system (8) has the same smoothing effect.

4. Despite of the similarities between the systems (6) and (8) underlined in Remark 1, the trajectories generated by the dynamical system (6) seems to have a better convergence behaviour than the trajectories generated by (8) as some numerical experiments show.
5. A trajectory $x(t)$ generated by the dynamical system (6) ensures the convergence rate of order $\mathcal{O}\left(\frac{1}{t^2}\right)$ for the decay $g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - \min g$, provided it holds that $\alpha > 3, \gamma > 0, \beta \in \mathbb{R}$ or $\alpha \geq 3, \gamma = 0, \beta \geq 0$.
6. A trajectory $x(t)$ generated by the dynamical system (6) ensures the same convergence rate of order $\mathcal{O}\left(\frac{1}{t^2}\right)$ for the decay $g(x(t)) - \min g$ as the heavy ball method with vanishing damping, for the case $\alpha > 3, \gamma = 0, \beta \geq 0$.
7. The convergence of a generated trajectory $x(t)$ to a minimizer of the objective function g can be obtained in case $\alpha > 3, \gamma > 0$ and $\beta \in \mathbb{R}$ and also in the case $\alpha > 3, \gamma = 0$ and $\beta \geq 0$.

Remark 2 Nevertheless, in case $\gamma = 0$ and $\beta < 0$ the dynamical system (6) can generate periodical solutions, hence the convergence of a generated trajectory to a minimizer of the objective is hopeless. To illustrate this fact, for $\beta < 0, \alpha > 0, \gamma = 0$ consider the strongly convex objective function $g : \mathbb{R} \rightarrow \mathbb{R}, g(x) = \frac{\alpha}{-2\beta}x^2$. Then, taking into account that $\gamma = 0$ the dynamical system (6) becomes

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{-\beta}x(t) = 0, \\ x(0) = 0, \dot{x}(0) = \sqrt{\frac{\alpha}{-\beta}}. \end{cases} \tag{9}$$

Now, the periodical function $x(t) = \sin \sqrt{\frac{\alpha}{-\beta}}t$ is a solution of (9), consequently do not exist the limit $\lim_{t \rightarrow +\infty} x(t)$.

We emphasize that the formulation of the dynamical system (6) is natural since by explicit discretization leads to inertial gradient methods, in particular the famous Polyak and Nesterov numerical schemes can be obtained from (6). For other inertial algorithms we refer to [2,17,25].

Indeed, explicit Euler discretization of (6), with the constant stepsize $h, t_n = nh, x_n = x(t_n)$ leads to

$$\frac{x_{n+1} - 2x_n + x_{n-1}}{h^2} + \frac{\alpha}{nh^2}(x_n - x_{n-1}) + \nabla g\left(x_n + \left(\frac{\gamma}{h} + \frac{\beta}{nh^2}\right)(x_n - x_{n-1})\right) = 0.$$

Equivalently, the latter equation can be written as

$$x_{n+1} = x_n + \left(1 - \frac{\alpha}{n}\right)(x_n - x_{n-1}) - h^2 \nabla g\left(x_n + \left(\frac{\gamma}{h} + \frac{\beta}{nh^2}\right)(x_n - x_{n-1})\right). \tag{10}$$

Now, setting $h^2 = s$ and denoting the constants $\frac{\gamma}{h}$ and $\frac{\beta}{h^2}$ still with γ and β , we get the following general inertial algorithm:

Let $x_0, x_{-1} \in \mathbb{R}^m$ and for all $n \in \mathbb{N}$ consider the sequences

$$\begin{cases} y_n = x_n + \left(1 - \frac{\alpha}{n}\right)(x_n - x_{n-1}) \\ z_n = x_n + \left(\gamma + \frac{\beta}{n}\right)(x_n - x_{n-1}) \\ x_{n+1} = y_n - s \nabla g(z_n). \end{cases} \quad (11)$$

However, from a practical point of view it is more convenient to work with the following equivalent formulation: Let $x_0, x_{-1} \in \mathbb{R}^m$ and for all $n \in \mathbb{N}$ consider the sequences

$$\begin{cases} y_n = x_n + \frac{n}{n+\alpha}(x_n - x_{n-1}) \\ z_n = x_n + \frac{\gamma n + \beta}{n+\alpha}(x_n - x_{n-1}) \\ x_{n+1} = y_n - s \nabla g(z_n), \end{cases} \quad (12)$$

where $\alpha > 0$, $\beta \in \mathbb{R}$ and $\gamma \geq 0$.

Remark 3 Notice that for $\gamma = \beta = 0$ we obtain a variant of Polyak's algorithm [29] and for $\gamma = 1$ and $\beta = 0$ we obtain Nesterov's algorithm [27,28]. An interesting fact is that Algorithm (12) allows different inertial steps and this approach seems to be new in the literature.

Remark 4 Note that according to Theorem 9 below, the dynamical system (6) can equivalently be rewritten as a first order system. Further, since the explicit Euler method applied to a first order system is an explicit Runge–Kutta method of order 1, the interested reader can use the results from [33] in order to obtain numerical schemes from (6) via Runge–Kutta discretization. In the view of the results from [33] the Runge–Kutta discretization of the dynamical system (6) may have an acceleration effect.

Remark 5 Independently to us, very recently, a system similar to (6) was studied by Muehlebach and Jordan in [26] and they show that Nesterov's accelerated gradient method can be obtained from their system via a semi-implicit Euler discretization scheme. They considered a constant damping instead of $\frac{\alpha}{t}$ and also took $\beta = 0$. Further they also treated the ODE

$$\ddot{x}(t) + \frac{3}{t+2}\dot{x}(t) + s \nabla g\left(x(t) + \frac{t-1}{t+2}\dot{x}(t)\right) = 0$$

which for $s = 1$ is obviously equivalent to the particular case of the governing ODE from (6), obtained for $\alpha = 3$, $\gamma = 1$, $\beta = -3$. However, the freedom of controlling the parameters β and γ in (6) is essential as the next numerical experiments show.

1.2 Some Numerical Experiments

In this section we consider two numerical experiments for the trajectories generated by the dynamical system (6) for a convex but not strongly convex objective function.

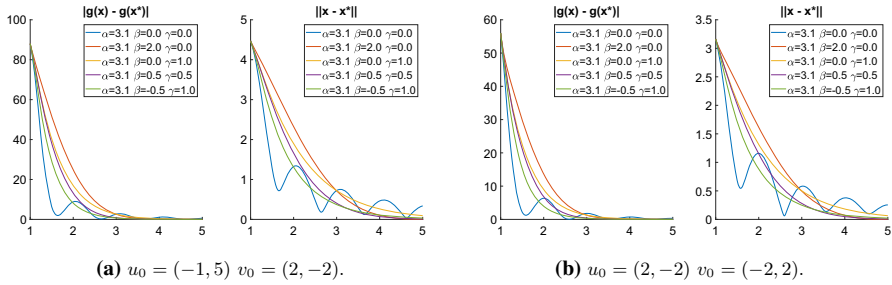


Fig. 1 Error analysis with different parameters in dynamical system (6) for a convex, but not strongly convex, objective function

Everywhere in the following numerical experiments we consider the continuous time dynamical systems (6) and (8), solved numerically with the ode45 adaptive method in MATLAB. We solved these dynamical systems with ode45 on the interval [1, 100] and the plot in Figs. 1, 2 show the energy error $|g(x(t)) - g(x^*)|$ on the left and the iterate error $\|x(t) - x^*\|$ on the right.

We show the evolution of the two errors with respect to different values for α , β and γ , including the case that yields the Heavy Ball with Friction. One can observe, see Fig. 1, that the best choice is not $\gamma = \beta = 0$ which is the case of heavy ball system with vanishing damping (1). Further, in the view of Remark 1, it seems interesting to compare the convergence behaviour of the trajectories generated by the dynamical systems (6) and (8) for different values for α , β and γ , see Fig. 2.

In the next experiments we consider the convex, but not strongly convex function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$g(x, y) = x^4 + 5y^2 - 4x - 10y + 8.$$

Then, $\nabla g(x, y) = (4x^3 - 4, 10y - 10)$, further $x^* = (1, 1)$ is the unique minimizer of g and $g^* = g(1, 1) = 0$.

In our first experiment we compare the convergence behaviour of the generated trajectories of (6) by taking into account the following instances.

α	β	γ	Color
3.1	0	0	Blue
3.1	2	0	Red
3.1	0	1	Yellow
3.1	0.5	0.5	Purple
3.1	-0.5	1	Green

The result are depicted in Fig. 1 for the starting points $u_0 = (-1, 5)$, $v_0 = (2, -2)$ and $u_0 = (2, -2)$, $v_0 = (-2, 2)$, respectively.

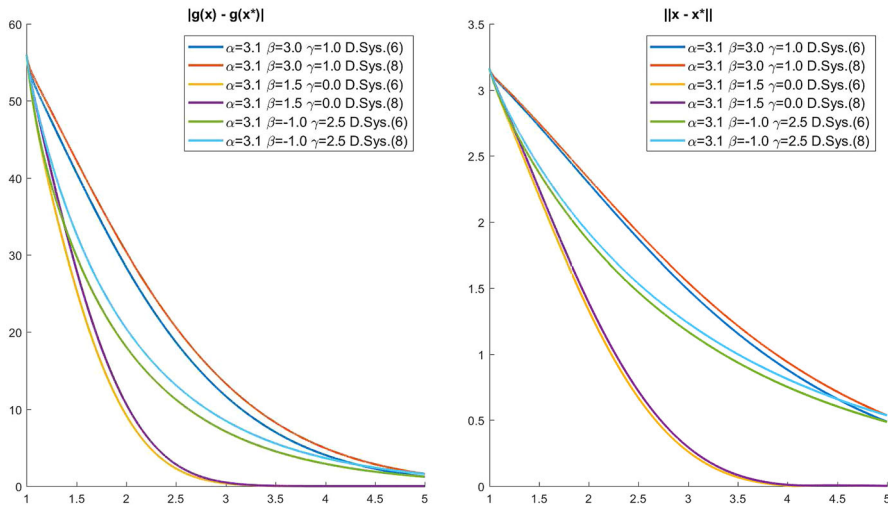


Fig. 2 Convergence properties of the trajectories generated by the dynamical systems (6) and (8)

Remark 6 Observe that in all cases the trajectories generated by dynamical system (6) have a better behaviour than the trajectories generated by the heavy ball system with vanishing damping (1). The choice of the parameters α , γ and β will be validated by the theoretical results from Sect. 3, where we show that in case $\alpha > 3$, $\gamma > 0$, $\beta \in \mathbb{R}$ and also in the case $\alpha > 3$, $\gamma = 0$, $\beta \geq 0$ the energy error $g(x(t)) - \min g$ is of order $\mathcal{O}\left(\frac{1}{t^2}\right)$ just as the case of heavy ball method. Further, for the values $\alpha > 3$, $\gamma > 0$, $\beta \in \mathbb{R}$ and for $\alpha > 3$, $\gamma = 0$, $\beta \geq 0$ we are able to show that a generated trajectory $x(t)$ converges to a minimum of the objective function g .

Remark 7 As we have emphasized before, it seems that the perturbation $\left(\gamma + \frac{\beta}{t}\right) \dot{x}(t)$ in the argument of the gradient of the objective function has a smoothing effect. The explanation of this fact is that, according to Remark 1, the trajectories generated by the dynamical system (6) and the trajectories generated by the dynamical system (8) are similar. Further, the behaviour of the trajectories generated by the dynamical system (8), in particular the explanation for this smoothing effect, are enlightened in [30].

In our second numerical experiment we consider the same convex function g and we compare the convergence behaviour of the trajectories generated by the dynamical system (6) (abbreviated D.Sys. (6)) and the trajectories generated by the dynamical system (8) (abbreviated D.Sys. (8)) for the following instances.

The result are depicted in Fig. 2 for the starting points $u_0 = (2, -2)$ and $v_0 = (-2, 2)$. One can observe that also for these instances the trajectories generated by (6) have slightly better convergence properties than the trajectories generated by (8).

α	β	γ
3.1	3	1
3.1	1.5	0
3.1	- 1	2.5

1.3 The Organization of the Paper

The outline of the paper is the following. After a short section concerning the existence and uniqueness of the trajectories generated by the dynamical system (6) in Sect. 3 we deal with the convergence analysis of the generated trajectories. We introduce a general energy functional, which will play the role of a Lyapunov function associated to the dynamical system (6) (see the full computations given in the Appendix). We obtain a rate of order $\mathcal{O}(1/t^2)$ for the decay $g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - \min g$. Further, we show that also the error $g(x(t)) - \min g$ has a rate of order $\mathcal{O}(1/t^2)$. Finally, we show the convergence of the generated trajectories to a minimum of the objective function g . In Sect. 4 we conclude our paper and we present some possible related future researches.

2 Existence and Uniqueness

The first step toward our existence and uniqueness result obtained in the present section concerns the definition of a strong global solution of the dynamical system (6).

Definition 1 We call the function $x : [t_0, +\infty) \rightarrow \mathbb{R}^m$ a strong global solution of the dynamical system (6) if satisfies the following properties:

- (i) $x, \dot{x} : [t_0, +\infty) \rightarrow \mathbb{R}^m$ are locally absolutely continuous;
- (ii) $\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) = 0$ for almost every $t \geq t_0$;
- (iii) $x(t_0) = u_0$ and $\dot{x}(t_0) = v_0$.

For brevity reasons, we recall that a mapping $x : [t_0, +\infty) \rightarrow \mathbb{R}^m$ is called locally absolutely continuous if it is absolutely continuous on every compact interval $[t_0, T]$, where $T > t_0$. Further, we have the following equivalent characterizations for an absolutely continuous function $x : [t_0, T] \rightarrow \mathbb{R}^m$, (see, for instance, [1,5]):

- (a) there exists $y : [t_0, T] \rightarrow \mathbb{R}^m$ an integrable function, such that

$$x(t) = x(t_0) + \int_{t_0}^t y(s)ds, \forall t \in [t_0, T];$$

- (b) x is a continuous function and its distributional derivative is Lebesgue integrable on the interval $[t_0, T]$;

- (c) for every $\varepsilon > 0$, there exists $\eta > 0$, such that for every finite family $I_k = (a_k, b_k)$ from $[t_0, T]$, the following implication is valid :

$$\left[I_k \cap I_j = \emptyset \text{ and } \sum_k |b_k - a_k| < \eta \right] \Rightarrow \left[\sum_k \|x(b_k) - x(a_k)\| < \varepsilon \right].$$

Remark 8 Let $x : [t_0, +\infty) \rightarrow \mathbb{R}^m$ be a locally absolutely continuous function. Then x is differentiable almost everywhere and its derivative coincides with its distributional derivative almost everywhere. On the other hand, we have the equality $\dot{x}(t) = y(t)$ for almost every $t \in [t_0, +\infty)$, where $y = y(t)$ is defined at the integration formula (a).

The first result of the present section concerns the existence and uniqueness of the trajectories generated by the dynamical system (6). We prove existence and uniqueness of a strong global solution of (6) by making use of the Cauchy-Lipschitz-Picard Theorem for absolutely continues trajectories (see for example [24, Proposition 6.2.1], [31, Theorem 54]). The key argument is that one can rewrite (6) as a particular first order dynamical system in a suitably chosen product space (see also [3,12,19,20]).

Theorem 9 Let $(u_0, v_0) \in \mathbb{R}^m \times \mathbb{R}^m$. Then, the dynamical system (6) admits a unique strong global solution.

It is worthwhile to emphasize that we do not need to assume that the objective function g is convex in order to obtain existence and uniqueness of the trajectories generated by the dynamical system (6). Further, we need the following result.

Proposition 10 For the starting points $(u_0, v_0) \in \mathbb{R}^m \times \mathbb{R}^m$ let x be the unique strong global solution of the dynamical system (6). Then, \ddot{x} is locally absolutely continuous on $[t_0, +\infty)$. In particular the third order derivative $x^{(3)}$ exists almost everywhere on $[t_0, +\infty)$.

The proofs of Theorem 9 and Proposition 10 are postponed to Appendix.

3 Convergence Analysis

3.1 On a General Energy Functional Associated to the Dynamical System (6)

In order to obtain convergence rates for the function values in the trajectories generated by the dynamical system (6), we need to introduce an appropriate energy functional which will play the role of a Lyapunov function. The form of such an energy functional associated to heavy ball system with vanishing damping and its extensions is well known in the literature, see for instance [8,9,13,14]. However, the novelty of the dynamical system (6), compared with the extended/perturbed variants of the heavy ball system studied in the above mentioned papers, consists in the fact that in system (6) the perturbation is carried out in the argument of the gradient of the objective function. This seems to be a new approach in the literature, therefore the previously

mentioned energy functionals are not suitable for a valuable convergence analysis of the dynamical system (6).

Hence, let us denote $\alpha(t) = \frac{\alpha}{t}$ and $\beta(t) = \gamma + \frac{\beta}{t}$, and assume that $\operatorname{argmin} g \neq \emptyset$. Further, let $g^* = \min g = g(x^*)$, $x^* \in \operatorname{argmin} g$. In connection to the dynamical system (6), we introduce the general energy functional

$$\mathcal{E}(t) = a(t)(g(x(t) + \beta(t)\dot{x}(t)) - g^*) + \frac{1}{2}\|b(t)(x(t) - x^*) + c(t)\dot{x}(t)\|^2 + \frac{d(t)}{2}\|x(t) - x^*\|^2, \tag{13}$$

which can be seen as an extension of the energy function studied in [9] in connection to the heavy ball system with vanishing damping.

Our purpose is to define the non-negative functions $a(t)$, $b(t)$, $c(t)$, $d(t)$ such that $\dot{\mathcal{E}}(t) \leq 0$, that is, the function $\mathcal{E}(t)$ is non-increasing after a $t_1 \geq t_0$. Indeed, if $\mathcal{E}(t)$ is non-increasing for $t \geq t_1$, then

$$a(t)(g(x(t) + \beta(t)\dot{x}(t)) - g^*) \leq \mathcal{E}(t) \leq \mathcal{E}(t_1),$$

in other words

$$g(x(t) + \beta(t)\dot{x}(t)) - g^* \leq \frac{\mathcal{E}(t_1)}{a(t)}, \text{ for all } t \geq t_1.$$

Concerning the derivative of the energy functional $\mathcal{E}(t)$ we can prove the following result (see the Sect. A2 at the Appendix).

$$\begin{aligned} \dot{\mathcal{E}}(t) &\leq (a'(t) - b(t)c(t))(g(x(t) + \beta(t)\dot{x}(t)) - g^*) \\ &\quad - a(t)\beta(t)\|\nabla g(x(t) + \beta(t)\dot{x}(t))\|^2 \\ &\quad + \left(-a(t)\alpha(t)\beta(t) + a(t)\beta'(t) + a(t) - c^2(t) + b(t)c(t)\beta(t)\right) \\ &\quad \langle \nabla g(x(t) + \beta(t)\dot{x}(t)), \dot{x}(t) \rangle \\ &\quad + \left(b'(t)b(t) + \frac{d'(t)}{2}\right)\|x(t) - x^*\|^2 \\ &\quad + \left(b^2(t) + b(t)c'(t) + b'(t)c(t) - b(t)c(t)\alpha(t) + d(t)\right)\langle \dot{x}(t), x(t) - x^* \rangle \\ &\quad + c(t)\left(b(t) + c'(t) - c(t)\alpha(t)\right)\|\dot{x}(t)\|^2. \end{aligned} \tag{14}$$

Consequently, in order to have $\dot{\mathcal{E}}(t) \leq 0$ for all $t \geq t_1$, $t_1 \geq t_0$, one must assume that for all $t \geq t_1$ the following inequalities hold:

$$a'(t) - b(t)c(t) \leq 0, \quad (15)$$

$$-a(t)\beta(t) \leq 0, \quad (16)$$

$$-a(t)\alpha(t)\beta(t) + a(t)\beta'(t) + a(t) - c^2(t) + b(t)c(t)\beta(t) = 0, \quad (17)$$

$$b'(t)b(t) + \frac{d'(t)}{2} \leq 0, \quad (18)$$

$$b'(t)c(t) + b(t)(b(t) + c'(t) - c(t)\alpha(t)) + d(t) = 0, \quad (19)$$

$$c(t)(b(t) + c'(t) - c(t)\alpha(t)) \leq 0. \quad (20)$$

Remark 11 Observe that (16) implies that $\beta(t) \geq 0$ for all $t \geq t_1$, $t_1 \geq t_0$ and this shows that in dynamical system (6), one must have $\beta \geq 0$ whenever $\gamma = 0$. Further, (18) is satisfied whenever $b(t)$ and $d(t)$ are constant functions. It is obvious that there exists t_1 such that for all $t \geq t_1$ we have $-\alpha(t)\beta(t) + \beta'(t) + 1 = 1 - \frac{\alpha\gamma}{t} - \frac{\alpha\beta + \beta}{t^2} > 0$, hence from (17) we get

$$a(t) = \frac{c^2(t) - b(t)c(t)\beta(t)}{-\alpha(t)\beta(t) + \beta'(t) + 1}.$$

Since (19) and (20) do not depend by $\beta(t)$, it seem natural to choose $c(t)$, $b(t)$ and $d(t)$ the same as in case of heavy ball system with vanishing damping (see [9]), that is, $c(t) = t$, $b(t) = b \in (0, \alpha - 1]$ and $d(t) = b(\alpha - 1 - b)$, for all $t \geq t_0$, provided $\alpha > 1$. Now, an easy computation shows that in this case

$$a(t) = t^2 + \gamma(\alpha - b)t + \beta + (\beta + \alpha\gamma^2)(\alpha - b) + \frac{(\alpha\beta\gamma + (\gamma\beta + 2\alpha\beta\gamma + \alpha^2\gamma^3)(\alpha - b))t + (\alpha\beta + \beta)(\beta + (\beta + \alpha\gamma^2)(\alpha - b))}{t^2 - \alpha\gamma t - (\alpha\beta + \beta)},$$

hence (15) is satisfied whenever $b > 2$, which implies that $\alpha > 3$.

However, if $\gamma = 0$ then

$$a(t) = t^2 + \beta + \beta(\alpha - b) + \frac{((\alpha\beta + \beta)(\beta + \beta(\alpha - b)))}{t^2 - (\alpha\beta + \beta)},$$

hence (15) holds also for $b = 2$ and $\alpha = 3$.

3.2 Error Estimates for the Values

In this section we obtain convergence rate of order $\mathcal{O}(1/t^2)$, $t \rightarrow +\infty$ for the difference $g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - g^*$ where $g^* = g(x^*) = \min g$, $x^* \in \operatorname{argmin} g \neq \emptyset$. From here we are able to show that $g(x(t)) - g^*$ also has a convergence rate of order $\mathcal{O}(1/t^2)$, $t \rightarrow +\infty$. However, just as in the case of heavy ball system with vanishing damping, in order to obtain these rates, it is necessary to assume $\alpha \geq 3$ in our system (6). We have the following result.

Theorem 12 Let x be the unique strong global solution of (6) and assume that $\operatorname{argmin} g \neq \emptyset$. Assume further that $\alpha > 3$, $\gamma > 0$ and $\beta \in \mathbb{R}$.

Then, there exists $K \geq 0$ and $t_1 \geq t_0$ such that

$$g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - \min g \leq \frac{K}{t^2}, \text{ for all } t \geq t_1. \tag{21}$$

Further,

$$\int_{t_0}^{+\infty} t \left(g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - \min g \right) dt < +\infty, \tag{22}$$

and

$$\int_{t_0}^{+\infty} t^2 \left\| \nabla g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) \right\|^2 dt < +\infty. \tag{23}$$

Proof Let $\min g = g^* = g(x^*)$, $x^* \in \operatorname{argmin} g$. Consider the energy functional (13) with

$$b(t) = \alpha - 1 > 2, \quad c(t) = t, \quad d(t) = 0.$$

According to Remark 11 we have

$$a(t) = t^2 + \gamma t + 2\beta + \alpha\gamma^2 + \frac{(3\alpha\beta\gamma + \alpha^2\gamma^3 + \beta\gamma)t + \beta(\alpha + 1)(2\beta + \alpha\gamma^2)}{t^2 - \alpha\gamma t - \beta(\alpha + 1)}$$

and the conditions (17)–(20) are satisfied with equality for every $t \geq t_0$.

Obviously, if t is big enough on has that $a(t) > 0$ and since $\gamma > 0$ it holds that $\gamma + \frac{\beta}{t} > 0$ for t big enough, even if $\beta < 0$. Hence, there exists $t' \geq t_0$ such (16) is satisfied for every $t \geq t'$.

Now, $a'(t) - b(t)c(t) = (3 - \alpha)t + \gamma + \mathcal{O}\left(\frac{1}{t^2}\right)$ and by taking into account that $\alpha > 3$ we obtain that there exists $t'' \geq t_0$ such that (15) holds for every $t \geq t''$.

Let $t''' = \max(t', t'')$. We conclude that, $\dot{\mathcal{E}}(t) \leq 0$ for all $t \geq t'''$, hence $\mathcal{E}(t)$ is non-increasing, i.e.

$$a(t) \left(g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - \min g \right) \leq \mathcal{E}(t) \leq \mathcal{E}(t'''), \text{ for all } t \geq t''.$$

But, obviously there exists $t_1 \geq t'''$ such that $a(t) \geq t^2$ for all $t \geq t_1$ and by denoting $K = \mathcal{E}(t''')$ we obtain

$$g\left(x(t) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t)\right) - \min g \leq \frac{K}{t^2}, \text{ for all } t \geq t_1.$$

Next we prove (22). Since $a'(t) - b(t)c(t) = (3 - \alpha)t + \gamma + \mathcal{O}\left(\frac{1}{t^2}\right)$ there exist $t_2 \geq t_1$ such that

$$a'(t) - b(t)c(t) \leq \frac{(3 - \alpha)t}{2}, \text{ for all } t \geq t_2.$$

Hence from (14) we have

$$\dot{\mathcal{E}}(t) \leq \frac{(3 - \alpha)t}{2} \left(g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) - \min g \right), \text{ for all } t \geq t_2,$$

and by integrating from t_2 to $T > t_2$ we get

$$\int_{t_2}^T t \left(g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) - \min g \right) dt \leq \frac{2}{3 - \alpha} (\mathcal{E}(T) - \mathcal{E}(t_2)).$$

Now, by letting $T \rightarrow +\infty$ and taking into account that \mathcal{E} is non-increasing, the conclusion follows.

For proving (23) observe that there exists $t_3 \geq t_1$ such that $-a(t) \left(\gamma + \frac{\beta}{t} \right) \leq -\frac{\gamma}{2}t^2$ for all $t \geq t_3$. Consequently

$$\dot{\mathcal{E}}(t) \leq -\frac{\gamma}{2}t^2 \left\| \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\|^2, \text{ for all } t \geq t_3.$$

Integrating from t_3 to $T > t_3$ and letting $T \rightarrow +\infty$ we obtain the desired conclusion. \square

Remark 13 Note that according to Remark 11 the condition (16) holds even if $\gamma = 0$ and $\beta \geq 0$. Consequently, even in the case $\alpha > 3$, $\gamma = 0$, $\beta \geq 0$ the conclusions (21) and (22) in Theorem 12 hold. However, if $\beta = 0$ then we cannot obtain (23). Nevertheless, if $\beta > 0$ then (23) becomes:

$$\int_{t_0}^{+\infty} t \left\| \nabla g \left(x(t) + \frac{\beta}{t} \dot{x}(t) \right) \right\|^2 dt < +\infty.$$

Remark 14 Concerning the case $\alpha = 3$, note that (16) is satisfied if one assumes that $\gamma = 0$ but $\beta \geq 0$. Moreover, in this case (15) is also satisfied since, one has $a'(t) - b(t)c(t) = -\frac{4\beta^2(\alpha+1)}{(t^2 - \beta(\alpha+1))^2} \leq 0$. Hence, also in the case $\alpha = 3$, $\gamma = 0$, $\beta \geq 0$ (21) in the conclusion of Theorem 12 holds.

Moreover, if we assume $\beta > 0$ then (23) becomes:

$$\int_{t_0}^{+\infty} t \left\| \nabla g \left(x(t) + \frac{\beta}{t} \dot{x}(t) \right) \right\|^2 dt < +\infty.$$

Next we show that also the error $g(x(t)) - \min g$ is of order $\mathcal{O}(1/t^2)$. For obtaining this result we need the Descent Lemma, see [28].

Lemma 15 Let $g : \mathbb{R}^m \rightarrow \mathbb{R}$ be a Fréchet differentiable function with L_g Lipschitz continuous gradient. Then one has

$$g(y) \leq g(x) + \langle \nabla g(x), y - x \rangle + \frac{L_g}{2} \|y - x\|^2, \quad \forall x, y \in \mathbb{R}^m.$$

Theorem 16 Let x be the unique strong global solution of (6) and assume that $\operatorname{argmin} g \neq \emptyset$. If $\alpha > 3$, $\gamma > 0$ and $\beta \in \mathbb{R}$ or $\alpha > 3$, $\gamma = 0$ and $\beta \geq 0$, then x is bounded and there exists $K \geq 0$ and $t_1 \geq t_0$ such that

$$\|\dot{x}(t)\| \leq \frac{K}{t}, \quad \text{for all } t \geq t_1. \tag{24}$$

Further,

$$\int_{t_0}^{+\infty} t \|\dot{x}(t)\|^2 dt < +\infty \tag{25}$$

and for $\gamma = 0$ and $\beta \geq 0$ there exists $K_1 \geq 0$ and $t_2 \geq t_1$ such that

$$g(x(t)) - \min g \leq \frac{K_1}{t^2}, \quad \text{for all } t \geq t_2. \tag{26}$$

Proof For $x^* \in \operatorname{argmin} g$ let $g^* = g(x^*) = \min g$ and consider the energy function (13) with $b(t) = b$, where $2 < b < \alpha - 1$, $c(t) = t$, $d(t) = b(\alpha - 1 - b) > 0$ and

$$\begin{aligned} a(t) &= \frac{c^2(t) - b(t)c(t) \left(\gamma + \frac{\beta}{t} \right)}{1 - \frac{\beta}{t^2} - \frac{\alpha}{t} \left(\gamma + \frac{\beta}{t} \right)} = \frac{(t^2 - b\gamma t - b\beta)t^2}{t^2 - \alpha\gamma t - \beta(\alpha + 1)} \\ &= \left(1 + \frac{(\alpha - b)\gamma t - \beta(\alpha + 1 - b)}{t^2 - \alpha\gamma t - \beta(\alpha + 1)} \right) t^2. \end{aligned}$$

According to Remark 11 there exists $t_1 \geq t_0$ such that the conditions (15)–(20) are satisfied.

From the definition of \mathcal{E} one has

$$\|b(t)(x(t) - x^*) + c(t)\dot{x}(t)\| \leq \sqrt{2\mathcal{E}(t)},$$

and

$$\|x(t) - x^*\| \leq \sqrt{\frac{2}{d(t)}\mathcal{E}(t)}$$

that is

$$\|b(x(t) - x^*) + t\dot{x}(t)\| \leq \sqrt{2\mathcal{E}(t)},$$

and

$$\|x(t) - x^*\| \leq \sqrt{\frac{2}{b(\alpha - 1 - b)}\mathcal{E}(t)}.$$

By using the fact that \mathcal{E} non-increasing on an interval $[t_1, +\infty)$ the latter inequality assures that x is bounded.

Now, by using the inequality $\|X - Y\| \geq \|X\| - \|Y\|$ we get

$$\|b(x(t) - x^*) + t\dot{x}(t)\| \geq t\|\dot{x}(t)\| - \|b(x(t) - x^*)\|,$$

hence for all $t \geq t_1$ one has

$$t\|\dot{x}(t)\| \leq \sqrt{2\mathcal{E}(t)} + \|b(x(t) - x^*)\| \leq \left(1 + \sqrt{\frac{1}{\alpha - 1 - b}}\right) \sqrt{2\mathcal{E}(t)} \leq K,$$

where $K = \left(1 + \sqrt{\frac{1}{\alpha - 1 - b}}\right) \sqrt{2\mathcal{E}(t_1)}$.

Further, (20) becomes $(b + 1 - \alpha)t < 0$, hence for all $t \geq t_1$ one has

$$\dot{\mathcal{E}}(t) \leq (b + 1 - \alpha)t\|\dot{x}(t)\|^2.$$

By integrating from t_1 to $T > t_1$ one gets

$$\int_{t_1}^T t\|\dot{x}(t)\|^2 dt \leq \frac{1}{\alpha - 1 - b}(\mathcal{E}(t_1) - \mathcal{E}(T)) \leq \frac{1}{\alpha - 1 - b}\mathcal{E}(t_1).$$

By letting $T \rightarrow +\infty$ we obtain

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

Now, by considering $\gamma = 0$, $\beta \geq 0$ and by using Lemma 15 with $y = x(t)$ and $x = x(t) + \frac{\beta}{t}\dot{x}(t)$ we obtain

$$\begin{aligned} g(x(t)) - g\left(x(t) + \frac{\beta}{t}\dot{x}(t)\right) &\leq \left\langle \nabla g\left(x(t) + \frac{\beta}{t}\dot{x}(t)\right), -\frac{\beta}{t}\dot{x}(t) \right\rangle + \frac{L_g}{2} \left\| \frac{\beta}{t}\dot{x}(t) \right\|^2 \\ &\leq \frac{\beta}{t} \left\| \nabla g\left(x(t) + \frac{\beta}{t}\dot{x}(t)\right) \right\| \|\dot{x}(t)\| + \frac{L_g}{2} \frac{\beta^2}{t^2} \|\dot{x}(t)\|^2. \end{aligned} \tag{27}$$

From (24), the Lipschitz continuity of ∇g and the facts that x and \dot{x} are bounded, we get that there exists $t'_2 \geq t_1$ and $K' > 0$ such that

$$\frac{\beta}{t} \left\| \nabla g\left(x(t) + \frac{\beta}{t}\dot{x}(t)\right) \right\| \|\dot{x}(t)\| \leq \frac{K'}{t^2}, \quad \text{for all } t \geq t'_2.$$

Further (24) assures that

$$\frac{L_g \beta^2}{2 t^2} \|\dot{x}(t)\| \leq \frac{L_g \beta^2 K}{2 t^3}, \quad \text{for all } t \geq t_1.$$

Combining the latter relations with (27) we get that there exists $K'_1 > 0$ such that

$$g(x(t)) - g\left(x(t) + \frac{\beta}{t} \dot{x}(t)\right) \leq \frac{K'_1}{t^2}, \quad \text{for all } t \geq t'_2. \tag{28}$$

Now, by adding (21) and (28) we get that there exists $K_1 > 0$ and $t_2 \geq t'_2$ such that

$$g(x(t)) - \min g \leq \frac{K_1}{t^2}, \quad \text{for all } t \geq t_2.$$

□

Remark 17 Note that (23), which holds whenever $\alpha > 3$ and $\gamma > 0$, assures that

$$t \left\| \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\| \in L^2([t_0, +\infty), \mathbb{R}).$$

Further, (25) assures that

$$\sqrt{t} \|\dot{x}(t)\| \in L^2([t_0, +\infty), \mathbb{R}). \tag{29}$$

Consequently, the system (6) leads to

$$\begin{aligned} \|t\ddot{x}(t)\| &= \left\| \alpha \dot{x}(t) + t \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\| \\ &\leq \alpha \|\dot{x}(t)\| + t \left\| \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\| \end{aligned}$$

hence, by (38) (see Lemma 27 in Appendix), that is $\|x^{(3)}(t)\| \leq K(\|\dot{x}(t)\| + \|\ddot{x}(t)\|)$ for some $K > 0$, we get

$$t \|\ddot{x}(t)\| \in L^2([t_0, +\infty), \mathbb{R}) \text{ and } \sqrt{t} \|x^{(3)}(t)\| \in L^2([t_0, +\infty), \mathbb{R}). \tag{30}$$

Remark 18 Notice that (29) and (30) and (38) assure in particular that

$$\lim_{t \rightarrow +\infty} \|\dot{x}(t)\| = \lim_{t \rightarrow +\infty} \|\ddot{x}(t)\| = 0 \text{ and } \lim_{t \rightarrow +\infty} \|x^{(3)}(t)\| = 0. \tag{31}$$

Remark 19 In the case $\gamma = 0$ and $\beta \geq 0$ according to Remark 13 one has

$$\sqrt{t} \left\| \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\| \in L^2([t_0, +\infty), \mathbb{R}).$$

Hence, as in Remark 17 we derive that

$$\sqrt{t}\|\ddot{x}(t)\|, \sqrt{t}\|x^{(3)}(t)\| \in L^2([t_0, +\infty), \mathbb{R})$$

and consequently (31) holds.

3.3 The Convergence of the Generated Trajectories

In this section we show convergence of the generated trajectories to a minimum point of the objective g . Consider the set of limit points of the trajectory x , that is

$$\begin{aligned} \omega(x) = \{ \bar{x} \in \mathbb{R}^m : \exists (t_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}, t_n \rightarrow +\infty, n \\ \rightarrow +\infty \text{ such that } x(t_n) \rightarrow \bar{x}, n \rightarrow +\infty \}. \end{aligned}$$

We show that $\omega(x) \subseteq \operatorname{argmin} g$. We emphasize that since g is convex one has

$$\operatorname{argmin} g = \operatorname{crit} g := \{ \bar{x} \in \mathbb{R}^m : \nabla g(\bar{x}) = 0 \}.$$

We have the following result.

Lemma 20 *Let x be the unique strong global solution of (6) and assume that $\operatorname{argmin} g \neq \emptyset$. If $\alpha > 3$, $\gamma > 0$ and $\beta \in \mathbb{R}$, then the following assumptions hold.*

- (i) $\omega(x) = \omega\left(x(\cdot) + \left(\gamma + \frac{\beta}{t}\right)\dot{x}(\cdot)\right)$;
- (ii) $\omega(x) \subseteq \operatorname{argmin} g$.

Proof Indeed (31) assures that $\lim_{t \rightarrow +\infty} \left(\gamma + \frac{\beta}{t}\right)\dot{x}(t) = 0$, which immediately proves (i).

For proving (ii) consider $\bar{x} \in \omega(x)$. Then, there exists $(t_n)_{n \in \mathbb{N}} \subseteq \mathbb{R}, t_n \rightarrow +\infty, n \rightarrow +\infty$ such that $\lim_{n \rightarrow +\infty} x(t_n) = \bar{x}$. Now, since ∇g is continuous and $\lim_{n \rightarrow +\infty} \left(x(t_n) + \left(\gamma + \frac{\beta}{t_n}\right)\dot{x}(t_n)\right) = \bar{x}$ one has

$$\lim_{n \rightarrow +\infty} \nabla g\left(x(t_n) + \left(\gamma + \frac{\beta}{t_n}\right)\dot{x}(t_n)\right) = \nabla g(\bar{x}).$$

Further, according to (31)

$$\lim_{n \rightarrow +\infty} \left(\ddot{x}(t_n) + \frac{\alpha}{t_n}\dot{x}(t_n)\right) = 0.$$

Now, the system (6) gives

$$0 = \lim_{n \rightarrow +\infty} \left(\ddot{x}(t_n) + \frac{\alpha}{t_n}\dot{x}(t_n) + \nabla g\left(x(t_n) + \left(\gamma + \frac{\beta}{t_n}\right)\dot{x}(t_n)\right)\right) = \nabla g(\bar{x})$$

that is $\bar{x} \in \operatorname{argmin} g$ and this proves (ii). □

Remark 21 Obviously, according to Remark 19 the conclusion of Lemma 20 remains valid also in the case $\alpha > 3$, $\gamma = 0$ and $\beta \geq 0$.

Our aim is to prove that the limit $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$ exists for every $x^* \in \operatorname{argmin} g$, and for this we need the following result from [9].

Lemma 22 (Lemma A.4. [9]) *Let $t_0 > 0$, and let $w : [t_0, +\infty) \rightarrow \mathbb{R}$ be a continuously differentiable function which is bounded from below. Assume that*

$$t\ddot{w}(t) + \alpha\dot{w}(t) \leq G(t)$$

for some $\alpha > 1$, almost every $t > t_0$, and some non-negative function $G \in L^1(t_0, +\infty)$. Then, the positive part $[\dot{w}]_+$ of \dot{w} belongs to $L^1(t_0, +\infty)$ and limit $\lim_{t \rightarrow +\infty} w(t)$ exists.

Now we can prove the following.

Lemma 23 *Let x be the unique strong global solution of (6) and assume that $\operatorname{argmin} g \neq \emptyset$. If $\alpha > 3$, $\gamma > 0$ and $\beta \in \mathbb{R}$, then for every $x^* \in \operatorname{argmin} g$ there exists the limit*

$$\lim_{t \rightarrow +\infty} \|x(t) - x^*\|.$$

Proof Let $x^* \in \operatorname{argmin} g$ and define the function $h_{x^*} : [t_0, +\infty) \rightarrow \mathbb{R}$, $h_{x^*}(t) = \frac{1}{2} \|x(t) - x^*\|^2$. Using the chain rule with respect to the differentiation of h_{x^*} , we obtain that

$$\dot{h}_{x^*}(t) = \langle \dot{x}(t), x(t) - x^* \rangle$$

and that

$$\ddot{h}_{x^*}(t) = \langle \ddot{x}(t), x(t) - x^* \rangle + \|\dot{x}(t)\|^2.$$

Let us denote $g^* = g(x^*) = \min(g)$. Using the dynamical system (6), one has that

$$\ddot{h}_{x^*}(t) + \frac{\alpha}{t} \dot{h}_{x^*}(t) = - \left\langle \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right), x(t) - x^* \right\rangle + \|\dot{x}(t)\|^2.$$

This means that

$$\begin{aligned} \ddot{h}_{x^*}(t) + \frac{\alpha}{t} \dot{h}_{x^*}(t) &= - \left\langle \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right), x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) - x^* \right\rangle \\ &\quad + \|\dot{x}(t)\|^2 + \left\langle \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right), \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right\rangle. \end{aligned} \tag{32}$$

By the convexity of the mapping g and taking into account that $g^* = g(x^*) = \min g$, we obtain

$$\begin{aligned} & -\left\langle \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right), x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) - x^* \right\rangle \\ & \leq g^* - g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \leq 0, \end{aligned}$$

hence by using (6) the inequality (32) becomes

$$\begin{aligned} \ddot{h}_{x^*}(t) + \frac{\alpha}{t} \dot{h}_{x^*}(t) & \leq g^* - g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) + \left(1 - \frac{\alpha}{t} \left(\gamma + \frac{\beta}{t} \right) \right) \|\dot{x}(t)\|^2 \\ & \quad - \left(\gamma + \frac{\beta}{t} \right) \langle \dot{x}(t), \ddot{x}(t) \rangle. \end{aligned}$$

Consequently, one has

$$t \ddot{h}_{x^*}(t) + \alpha \dot{h}_{x^*}(t) \leq \left(t - \alpha \left(\gamma + \frac{\beta}{t} \right) \right) \|\dot{x}(t)\|^2 - (\gamma t + \beta) \langle \dot{x}(t), \ddot{x}(t) \rangle. \quad (33)$$

Now, according to (29) one has

$$t \|\dot{x}(t)\|^2 \in L^1([t_0, +\infty), \mathbb{R}).$$

Consequently,

$$\left(t - \alpha \left(\gamma + \frac{\beta}{t} \right) \right) \|\dot{x}(t)\|^2 \in L^1([t_0, +\infty), \mathbb{R}). \quad (34)$$

Further,

$$-(\gamma t + \beta) \langle \dot{x}(t), \ddot{x}(t) \rangle \leq \frac{1}{2} \|\dot{x}(t)\|^2 + \frac{1}{2} (\gamma t + \beta)^2 \|\ddot{x}(t)\|^2$$

and (30) assures that

$$t^2 \|\ddot{x}(t)\|^2 \in L^1([t_0, +\infty), \mathbb{R}),$$

hence

$$-(\gamma t + \beta) \langle \dot{x}(t), \ddot{x}(t) \rangle \in L^1([t_0, +\infty), \mathbb{R}). \quad (35)$$

According to (34) and (35) we get that the function

$$\begin{aligned} G(t) & = \left(t - \alpha \left(\gamma + \frac{\beta}{t} \right) \right) \|\dot{x}(t)\|^2 + \frac{1}{2} \|\dot{x}(t)\|^2 \\ & \quad + \frac{1}{2} (\gamma t + \beta)^2 \|\ddot{x}(t)\|^2 \in L^1([t_0, +\infty), \mathbb{R}). \end{aligned}$$

Moreover, it is obvious that there exists $t_1 \geq t_0$ such that G is non negative for every $t \geq t_1$. From (33) one has

$$t\ddot{h}_{x^*}(t) + \alpha\dot{h}_{x^*}(t) \leq G(t), \text{ for every } t \geq t_0, \tag{36}$$

hence Lemma 22 leads to the existence of the limit

$$\lim_{t \rightarrow +\infty} h_{x^*}(t)$$

and consequently the limit

$$\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$$

also exists. □

Now, we present the main result of this subsection regarding the convergence of the solution of the dynamical system (6) as $t \rightarrow +\infty$.

Theorem 24 *Let x be the unique strong global solution of the dynamical system (6) and assume that $\text{argmin } g \neq \emptyset$. If $\alpha > 3$, $\gamma > 0$ and $\beta \in \mathbb{R}$, then $x(t)$ converges to a point in $\text{argmin } g$ as $t \rightarrow +\infty$.*

Proof Let $x^* \in \text{argmin } g$. Then, by Lemma 23, we know that $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$ exists, hence x is bounded. But then from Bolzano–Weierstrass Theorem we get that there exists $\bar{x} \in \omega(x)$. In other words, there exists a sequence $t_n \rightarrow +\infty$, $n \rightarrow +\infty$, such that $\lim_{n \rightarrow +\infty} x(t_n) = \bar{x}$. From Lemma 20, we obtain that $\bar{x} \in \text{argmin } g$. Using Lemma 23 again, we find that $\lim_{t \rightarrow +\infty} \|x(t) - \bar{x}\| = \lim_{n \rightarrow +\infty} \|x(t_n) - \bar{x}\| = 0$, hence

$$x(t) \rightarrow \bar{x} \in \text{argmin } g, t \rightarrow +\infty.$$

□

Remark 25 As it was expected, Theorem 24 remains true if in its hypothesis we assume only that $\alpha > 3$, $\gamma = 0$ and $\beta \geq 0$. Indeed, note that under these assumptions the conclusion of Lemma 23 holds, since G from its proof becomes

$$G(t) = \left(t - \frac{\alpha\beta}{t}\right) \|\dot{x}(t)\|^2 + \frac{1}{2} \|\dot{x}(t)\|^2 + \frac{1}{2} \beta^2 \|\ddot{x}(t)\|^2$$

and according to Remark 19 $G(t) \in L^1([t_0, +\infty), \mathbb{R})$. Moreover, it is obvious that there exists $t_1 \geq t_0$ such that G is non negative for every $t \geq t_1$.

This fact combined with Remark 21 lead to the desired conclusion.

4 Conclusion

In this paper we study a second order dynamical system which can be viewed as an extension of the heavy ball system with vanishing damping. This dynamical system is actually a perturbed version of the heavy ball system with vanishing damping, but the perturbation is made in the argument of the gradient of the objective function. Our system is also strongly related to the heavy ball system with Hessian driven damping. Numerical experiments show that the above mentioned perturbation brings a smoothing effect in the behaviour of the energy error $g(x(t)) - \min g$ and also in the behaviour of the absolute error $\|x(t) - x^*\|$, where $x(t)$ is a trajectory generated by our dynamical system and x^* is a minimum of the objective g . We show that our system via explicit discretization leads to inertial algorithms. A related future research is the convergence analysis of algorithm (12), since this algorithm contains as particular case the famous Nesterov algorithm. However, since Algorithm (12) may allow different inertial steps, its area of applicability can be considerable.

We have shown existence and uniqueness of the trajectories generated by our dynamical system even for the case the objective function g is non-convex. Further, we treated the cases when the energy error $g(x(t)) - \min g$ is of order $\mathcal{O}\left(\frac{1}{t^2}\right)$ and we obtained the convergence of a generated trajectory to a minimum of the objective function g . Another related research is the convergence analysis of the generated trajectories in the case when the objective function g is possible non-convex. This would be a novelty in the literature even for the case $\alpha > 3$, $\gamma = \beta = 0$, that is, for the case of the heavy ball system with vanishing damping.

Finally, we underline that the dynamical system (6) can easily be extended to proximal-gradient dynamical systems, (see [18,20] and the references therein).

Acknowledgements The authors are thankful to two anonymous referees for their valuable remarks and suggestions which improved the quality of the paper.

A. Complements to the Section Existence and uniqueness

In what follows we give a detailed proof for Theorem 9.

Proof (Theorem 9).

By making use of the notation $X(t) = (x(t), \dot{x}(t))$ the system (6) can be rewritten as a first order dynamical system:

$$\begin{cases} \dot{X}(t) = F(t, X(t)) \\ X(t_0) = (u_0, v_0), \end{cases} \quad (37)$$

where $F : [t_0, +\infty) \times \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathbb{R}^m \times \mathbb{R}^m$, $F(t, u, v) = \left(v, -\frac{\alpha}{t}v - \nabla g\left(u + \left(\gamma + \frac{\beta}{t}\right)v\right)\right)$.

The existence and uniqueness of a strong global solution of (6) follows according to the Cauchy–Lipschitz–Picard Theorem applied to the first order dynamical system

(37). In order to prove the existence and uniqueness of the trajectories generated by (37) we show the following:

- (I) For every $t \in [t_0, +\infty)$ the mapping $F(t, \cdot, \cdot)$ is $L(t)$ -Lipschitz continuous and $L(\cdot) \in L^1_{loc}([t_0, +\infty))$.
- (II) For all $u, v \in \mathbb{R}^m$ one has $F(\cdot, u, v) \in L^1_{loc}([t_0, +\infty), \mathbb{R}^m \times \mathbb{R}^m)$.

Let us prove (I). Let $t \in [t_0, +\infty)$ be fixed and consider the pairs (u, v) and (\bar{u}, \bar{v}) from $\mathbb{R}^m \times \mathbb{R}^m$. Using the Lipschitz continuity of ∇g and the obvious inequality $\|A + B\|^2 \leq 2\|A\|^2 + 2\|B\|^2$ for all $A, B \in \mathbb{R}^m$, we make the following estimations :

$$\begin{aligned} & \|F(t, u, v) - F(t, \bar{u}, \bar{v})\| \\ &= \sqrt{\|v - \bar{v}\|^2 + \left\| \frac{\alpha}{t}(v - \bar{v}) + \nabla g\left(\bar{u} + \left(\gamma + \frac{\beta}{t}\right)\bar{v}\right) - \nabla g\left(u + \left(\gamma + \frac{\beta}{t}\right)v\right) \right\|^2} \\ &\leq \sqrt{\left(1 + 2\left(\frac{\alpha}{t}\right)^2\right)\|v - \bar{v}\|^2 + 2L_g^2\left\| \left(u - \bar{u}\right) + \left(\gamma + \frac{\beta}{t}\right)(v - \bar{v}) \right\|^2} \\ &\leq \sqrt{1 + 4L_g^2 + 2\left(\frac{\alpha}{t}\right)^2 + 4L_g^2\left(\gamma + \frac{\beta}{t}\right)^2} \| (u, v) - (\bar{u}, \bar{v}) \|. \end{aligned}$$

By employing the notation $L(t) = \sqrt{1 + 4L_g^2 + 2\left(\frac{\alpha}{t}\right)^2 + 4L_g^2\left(\gamma + \frac{\beta}{t}\right)^2}$, we have that

$$\|F(t, u, v) - F(t, \bar{u}, \bar{v})\| \leq L(t) \cdot \|(u, v) - (\bar{u}, \bar{v})\|.$$

Obviously the function $t \rightarrow L(t)$ is continuous on $[t_0, +\infty)$, hence $L(\cdot)$ is integrable on $[t_0, T]$ for all $t_0 < T < +\infty$.

For proving (II) consider $(u, v) \in \mathbb{R}^m \times \mathbb{R}^m$ a fixed pair of elements and let $T > t_0$. We consider the following estimations:

$$\begin{aligned} \int_{t_0}^T \|F(t, u, v)\| dt &= \int_{t_0}^T \sqrt{\|v\|^2 + \left\| \frac{\alpha}{t}v + \nabla g\left(u + \left(\gamma + \frac{\beta}{t}\right)v\right) \right\|^2} dt \\ &\leq \int_{t_0}^T \sqrt{\left(1 + 2\left(\frac{\alpha}{t}\right)^2 + 4L_g^2\left(\gamma + \frac{\beta}{t}\right)^2\right)\|v\|^2 + 4\|\nabla g(u)\|^2} dt \\ &\leq \sqrt{\|v\|^2 + \|\nabla g(u)\|^2} \int_{t_0}^T \sqrt{5 + 2\left(\frac{\alpha}{t}\right)^2 + 4L_g^2\left(\gamma + \frac{\beta}{t}\right)^2} dt \end{aligned}$$

and the conclusion follows by the continuity of the function $t \rightarrow \sqrt{5 + 2\left(\frac{\alpha}{t}\right)^2 + 4L_g^2\left(\gamma + \frac{\beta}{t}\right)^2}$ on $[t_0, T]$.

The Cauchy–Lipschitz–Picard theorem guarantees existence and uniqueness of the trajectory of the first order dynamical system (37) and thus of the second order dynamical system (6). \square

Remark 26 Note that we did not use the convexity assumption imposed on g in the proof of Theorem 9. However, we emphasize that according to the proof of Theorem 9, the assumption that g has a Lipschitz continuous gradient is essential in order to obtain existence and uniqueness of the trajectories generated by the dynamical system (6).

Next we present the complete proof of Proposition 10. For simplicity, in the proof of the following sequel we employ the 1-norm on $\mathbb{R}^m \times \mathbb{R}^m$, defined as $\|(x_1, x_2)\|_1 = \|x_1\| + \|x_2\|$, for all $x_1, x_2 \in \mathbb{R}^m$. Obviously one has

$$\frac{1}{\sqrt{2}} \|(x_1, x_2)\|_1 \leq \|(x_1, x_2)\| = \sqrt{\|x_1\|^2 + \|x_2\|^2} \leq \|(x_1, x_2)\|_1, \text{ for all } x_1, x_2 \in \mathbb{R}^m.$$

Proof (Proposition 10).

We show that $\dot{X}(t) = (\dot{x}(t), \ddot{x}(t))$ is locally absolutely continuous, hence \ddot{x} is also locally absolutely continuous. This implies by Remark 8 that $x^{(3)}$ exists almost everywhere on $[t_0, +\infty)$.

Let $T > t_0$ and $s, t \in [t_0, T]$. We consider the following chain of inequalities :

$$\begin{aligned} & \|\dot{X}(s) - \dot{X}(t)\|_1 \\ &= \|F(s, X(s)) - F(t, X(t))\|_1 \\ &= \left\| \left(\dot{x}(s) - \dot{x}(t), \frac{\alpha}{t} \dot{x}(t) - \frac{\alpha}{s} \dot{x}(t) + \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right. \right. \\ &\quad \left. \left. - \nabla g \left(x(s) + \left(\gamma + \frac{\beta}{s} \right) \dot{x}(s) \right) \right) \right\|_1 \\ &\leq \|\dot{x}(s) - \dot{x}(t)\| + \left\| \frac{\alpha}{s} \dot{x}(s) - \frac{\alpha}{t} \dot{x}(t) \right\| \\ &\quad + \left\| \nabla g \left(x(s) + \left(\gamma + \frac{\beta}{s} \right) \dot{x}(s) \right) - \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\| \end{aligned}$$

and by using the L_g -Lipschitz continuity of ∇g we obtain

$$\begin{aligned} \|\dot{X}(s) - \dot{X}(t)\|_1 &\leq \|\dot{x}(s) - \dot{x}(t)\| + \left\| \frac{\alpha}{s} \dot{x}(s) - \frac{\alpha}{t} \dot{x}(t) \right\| + L_g \|x(s) - x(t)\| \\ &\quad + L_g \left\| \left(\gamma + \frac{\beta}{s} \right) \dot{x}(s) - \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right\| \\ &\leq \|\dot{x}(s) - \dot{x}(t)\| + \frac{\alpha}{s} \|\dot{x}(s) - \dot{x}(t)\| + \left| \frac{\alpha}{s} \right. \\ &\quad \left. - \frac{\alpha}{t} \right| \|\dot{x}(t)\| + L_g \|x(s) - x(t)\| \\ &\quad + L_g \left(\gamma + \frac{\beta}{s} \right) \|\dot{x}(s) - \dot{x}(t)\| \\ &\quad + L_g \left| \frac{\beta}{s} - \frac{\beta}{t} \right| \cdot \|\dot{x}(t)\| \\ &= \left(1 + \frac{\alpha}{s} + L_g \left(\gamma + \frac{\beta}{s} \right) \right) \|\dot{x}(s) - \dot{x}(t)\| + L_g \|x(s) - x(t)\| \\ &\quad + (\alpha + L_g |\beta|) \cdot \left| \frac{1}{s} - \frac{1}{t} \right| \|\dot{x}(t)\|. \end{aligned}$$

Further, let us introduce the following additional notations:

$$\begin{aligned} L_1 &:= \max_{s \in [t_0, T]} \left(1 + \frac{\alpha}{s} + L_g \left(\gamma + \frac{\beta}{s} \right) \right) = 1 + \frac{\alpha}{t_0} + L_g \left(\gamma + \frac{\beta}{t_0} \right) \\ \text{and } L_2 &:= (\alpha + L_g |\beta|) \max_{t \in [t_0, T]} \|\dot{x}(t)\|. \end{aligned}$$

Then, one has

$$\begin{aligned} \|\dot{X}(s) - \dot{X}(t)\| &\leq \|\dot{X}(s) - \dot{X}(t)\|_1 \leq L_1 \|\dot{x}(s) - \dot{x}(t)\| + L_g \|x(s) - x(t)\| \\ &\quad + L_2 \left| \frac{1}{s} - \frac{1}{t} \right|. \end{aligned}$$

By the fact that x is the strong global solution for the dynamical system (6), it follows that x and \dot{x} are absolutely continuous on the interval $[t_0, T]$. Moreover, the function $t \rightarrow \frac{1}{t}$ belongs to $C^1([t_0, T], \mathbb{R})$, hence it is also absolutely continuous on the interval $[t_0, T]$. Let $\varepsilon > 0$. Then, there exists $\eta > 0$, such that for $I_k = (a_k, b_k) \subseteq [t_0, T]$ satisfying $I_k \cap I_j = \emptyset$ and $\sum_k |b_k - a_k| < \eta$, we have that

$$\sum_k \|\dot{x}(b_k) - \dot{x}(a_k)\| < \frac{\varepsilon}{3L_1}, \quad \sum_k \|x(b_k) - x(a_k)\| < \frac{\varepsilon}{3L_g} \quad \text{and} \quad \sum_k \left| \frac{1}{b_k} - \frac{1}{a_k} \right| < \frac{\varepsilon}{3L_2}.$$

Summing all up, we obtain

$$\begin{aligned} \sum_k \|\dot{X}(b_k) - \dot{X}(a_k)\| &\leq L_1 \sum_k \|\dot{x}(b_k) - \dot{x}(a_k)\| + L_g \sum_k \|x(b_k) - x(a_k)\| \\ &\quad + L_2 \sum_k \left| \frac{1}{b_k} - \frac{1}{a_k} \right| < \varepsilon, \end{aligned}$$

consequently \dot{X} is absolutely continuous on $[t_0, T]$. and the conclusion follows. \square

Concerning an upper bound estimate of the third order derivative $x^{(3)}$ the following result holds.

Lemma 27 *For the initial values $(u_0, v_0) \in \mathbb{R}^m \times \mathbb{R}^m$ consider x the unique strong global solution of the second-order dynamical system (6). Then, there exists $K > 0$ such that for almost every $t \in [t_0, +\infty)$, we have that:*

$$\|x^{(3)}(t)\| \leq K(\|\dot{x}(t)\| + \|\ddot{x}(t)\|). \quad (38)$$

Proof For $h > 0$ we consider the following inequalities :

$$\begin{aligned} &\|\dot{X}(t+h) - \dot{X}(t)\|_1 \\ &= \|F(t+h, X(t+h)) - F(t, X(t))\|_1 \\ &\leq \|(\dot{x}(t+h) - \dot{x}(t))\| + \alpha \left\| \frac{1}{t+h} \dot{x}(t+h) - \frac{1}{t} \dot{x}(t) \right\| \\ &\quad + \left\| \nabla g \left(x(t+h) + \left(\gamma + \frac{\beta}{t+h} \right) \dot{x}(t+h) \right) - \nabla g \left(x(t) + \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right) \right\| \\ &\leq \|(\dot{x}(t+h) - \dot{x}(t))\| + \alpha \left\| \frac{1}{t+h} \dot{x}(t+h) - \frac{1}{t} \dot{x}(t) \right\| \\ &\quad + L_g \|x(t+h) - x(t)\| + L_g \left\| \left(\gamma + \frac{\beta}{t+h} \right) \dot{x}(t+h) - \left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right\|. \end{aligned}$$

Now, dividing by $h > 0$ and taking the limit $h \rightarrow 0$, it follows that

$$\|\ddot{X}(t)\|_1 \leq \|\ddot{x}(t)\| + \alpha \left\| \left(\frac{1}{t} \dot{x}(t) \right)' \right\| + L_g \|\dot{x}(t)\| + L_g \left\| \left(\left(\gamma + \frac{\beta}{t} \right) \dot{x}(t) \right)' \right\|.$$

Consequently,

$$\|x^{(3)}(t)\| \leq \alpha \left\| -\frac{1}{t^2} \dot{x}(t) + \frac{1}{t} \ddot{x}(t) \right\| + L_g \|\dot{x}(t)\| + L_g \left\| -\frac{\beta}{t^2} \dot{x}(t) + \left(\gamma + \frac{\beta}{t} \right) \ddot{x}(t) \right\|.$$

Finally,

$$\begin{aligned} \|x^{(3)}(t)\| &\leq \left(L_g + \frac{\alpha + L_g|\beta|}{t^2}\right) \|\dot{x}(t)\| + \left(\frac{\alpha}{t} + L_g \left|\gamma + \frac{\beta}{t}\right|\right) \|\ddot{x}(t)\| \\ &\leq K(\|\dot{x}(t)\| + \|\ddot{x}(t)\|), \end{aligned}$$

where $K = \max \left\{ \max_{t \geq t_0} \left(L_g + \frac{\alpha + L_g|\beta|}{t^2}\right), \max_{t \geq t_0} \left(\frac{\alpha}{t} + L_g \left|\gamma + \frac{\beta}{t}\right|\right) \right\}$. \square

B. On the Derivative of the Energy Functional (13)

In what follows we derive the conditions which must be imposed on the positive functions $a(t), b(t), c(t), d(t)$ in order to obtain $\dot{\mathcal{E}}(t) \leq 0$ for every $t \geq t_1$. We have,

$$\begin{aligned} \dot{\mathcal{E}}(t) &= a'(t)(g(x(t) + \beta(t)\dot{x}(t)) - g^*) \\ &\quad + a(t)\langle \nabla g(x(t) + \beta(t)\dot{x}(t)), \beta(t)\ddot{x}(t) + (\beta'(t) + 1)\dot{x}(t) \rangle \\ &\quad + \langle b'(t)(x(t) - x^*) + (b(t) + c'(t))\dot{x}(t) \rangle \\ &\quad + c(t)\ddot{x}(t), b(t)(x(t) - x^*) + c(t)\dot{x}(t) \rangle \\ &\quad + \frac{d'(t)}{2} \|x(t) - x^*\|^2 + d(t)\langle \dot{x}(t), x(t) - x^* \rangle. \end{aligned}$$

Now, from (6) we get $\ddot{x}(t) = -\alpha(t)\dot{x}(t) - \nabla g(x(t) + \beta(t)\dot{x}(t))$, hence

$$\begin{aligned} \dot{\mathcal{E}}(t) &= a'(t)(g(x(t) + \beta(t)\dot{x}(t)) - g^*) \\ &\quad + a(t)\langle \nabla g(x(t) + \beta(t)\dot{x}(t)), -\beta(t)\nabla g(x(t) + \beta(t)\dot{x}(t)) \\ &\quad + (-\beta(t)\alpha(t) + \beta'(t) + 1)\dot{x}(t) \rangle \\ &\quad + \langle b'(t)(x(t) - x^*) + (b(t) + c'(t) - c(t)\alpha(t))\dot{x}(t) \rangle \\ &\quad - c(t)\nabla g(x(t) + \beta(t)\dot{x}(t)), b(t)(x(t) - x^*) + c(t)\dot{x}(t) \rangle \\ &\quad + \frac{d'(t)}{2} \|x(t) - x^*\|^2 + d(t)\langle \dot{x}(t), x(t) - x^* \rangle. \end{aligned}$$

Consequently,

$$\begin{aligned}
 \dot{\mathcal{E}}(t) &= a'(t)(g(x(t) + \beta(t)\dot{x}(t)) - g^*) \\
 &\quad - a(t)\beta(t)\|\nabla g(x(t) + \beta(t)\dot{x}(t))\|^2 \\
 &\quad + \left(-a(t)\alpha(t)\beta(t) + a(t)\beta'(t) + a(t) - c^2(t)\right) \langle \nabla g(x(t) + \beta(t)\dot{x}(t)), \dot{x}(t) \rangle \\
 &\quad + \left(b'(t)b(t) + \frac{d'(t)}{2}\right) \|x(t) - x^*\|^2 \\
 &\quad + \left(b^2(t) + b(t)c'(t) + b'(t)c(t) - b(t)c(t)\alpha(t) + d(t)\right) \langle \dot{x}(t), x(t) - x^* \rangle \\
 &\quad + c(t) \left(b(t) + c'(t) - c(t)\alpha(t)\right) \|\dot{x}(t)\|^2 \\
 &\quad - b(t)c(t)\langle \nabla g(x(t) + \beta(t)\dot{x}(t)), x(t) - x^* \rangle. \tag{39}
 \end{aligned}$$

But

$$\begin{aligned}
 \langle \nabla g(x(t) + \beta(t)\dot{x}(t)), x(t) - x^* \rangle &= \langle \nabla g(x(t) + \beta(t)\dot{x}(t)), x(t) + \beta(t)\dot{x}(t) - x^* \rangle \\
 &\quad - \langle \nabla g(x(t) + \beta(t)\dot{x}(t)), \beta(t)\dot{x}(t) \rangle
 \end{aligned}$$

and by the convexity of g we have $\langle \nabla g(x(t) + \beta(t)\dot{x}(t)), x(t) + \beta(t)\dot{x}(t) - x^* \rangle \geq g(x(t) + \beta(t)\dot{x}(t)) - g^*$, hence

$$\begin{aligned}
 &-b(t)c(t)\langle \nabla g(x(t) + \beta(t)\dot{x}(t)), x(t) - x^* \rangle \\
 &\leq -b(t)c(t)(g(x(t) + \beta(t)\dot{x}(t)) - g^*) + b(t)c(t)\beta(t)\langle \nabla g(x(t) \\
 &\quad + \beta(t)\dot{x}(t)), \dot{x}(t) \rangle.
 \end{aligned}$$

Therefore, (39) becomes

$$\begin{aligned}
 \dot{\mathcal{E}}(t) &\leq (a'(t) - b(t)c(t))(g(x(t) + \beta(t)\dot{x}(t)) - g^*) - a(t)\beta(t)\|\nabla g(x(t) + \beta(t)\dot{x}(t))\|^2 \\
 &\quad + \left(-a(t)\alpha(t)\beta(t) + a(t)\beta'(t) + a(t) - c^2(t) + b(t)c(t)\beta(t)\right) \\
 &\quad \langle \nabla g(x(t) + \beta(t)\dot{x}(t)), \dot{x}(t) \rangle \\
 &\quad + \left(b'(t)b(t) + \frac{d'(t)}{2}\right) \|x(t) - x^*\|^2 \\
 &\quad + \left(b^2(t) + b(t)c'(t) + b'(t)c(t) - b(t)c(t)\alpha(t) + d(t)\right) \langle \dot{x}(t), x(t) - x^* \rangle \\
 &\quad + c(t) \left(b(t) + c'(t) - c(t)\alpha(t)\right) \|\dot{x}(t)\|^2. \tag{40}
 \end{aligned}$$

From here one easily can observe that the conditions (15)–(20) assure that

$$\dot{\mathcal{E}}(t) \leq 0, \quad \text{for all } t \geq t_1.$$

Remark 28 Note that we did not use the form of the sequences $\alpha(t)$ and $\beta(t)$ in the above computations. Consequently, the energy functional (13) may be suitable also

for the following general system.

$$\begin{cases} \ddot{x}(t) + \alpha(t)\dot{x}(t) + \nabla g(x(t) + \beta(t)\dot{x}(t)) = 0 \\ x(t_0) = u_0, \dot{x}(t_0) = v_0, \end{cases}$$

where $u_0, v_0 \in \mathbb{R}^m$ and $\alpha, \beta : [t_0 + \infty) \rightarrow (0, +\infty)$, $t_0 \geq 0$ are continuously derivable functions.

References

1. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward–backward methods for structured monotone inclusions in Hilbert spaces. *J. Optim. Theory Appl.* **161**(2), 331–360 (2014)
2. Alecsa, C., László, S.C.: Viorel, A: A gradient type algorithm with backward inertial steps associated to a nonconvex minimization problem. *Num. Algorithms* (2019). <https://doi.org/10.1007/s11075-019-00765-z>
3. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics. *Journal de Mathématiques Pures et Appliquées* **81**(8), 747–779 (2002)
4. Attouch, H., Chbani, Z.: Fast inertial dynamics and fista algorithms in convex optimization. *Perturbation aspects.* (2015). [arXiv:1507.01367](https://arxiv.org/abs/1507.01367)
5. Attouch, H., Svaiter, B.F.: A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM J. Control Optim.* **49**(2), 574–598 (2011)
6. Attouch, H., Goudou, X., Redont, P.: The heavy ball with friction method, I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system. *Commun. Contemp. Math.* **2**, 1–34 (2000)
7. Attouch, H., Peypouquet, J., Redont, P.: A dynamical approach to an inertial forward–backward algorithm for convex minimization. *SIAM J. Optim.* **24**(1), 232–256 (2014)
8. Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping. *J. Differ. Equ.* **261**(10), 5734–5783 (2016)
9. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Math. Program.* **168**(1–2), 123–175 (2018)
10. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: COCV* **25**, 2 (2019)
11. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping. (2019). [arXiv:1907.10536](https://arxiv.org/abs/1907.10536)
12. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via time scaling of damped inertial gradient dynamics. <https://hal.archives-ouvertes.fr/hal-02138954> (2019)
13. Aujol, J.F., Dossal, Ch.: Optimal rate of convergence of an ODE associated to the Fast Gradient Descent schemes for $b > 0$. <https://hal.inria.fr/hal-01547251v2/document> (2017)
14. Aujol, J.F., Dossal, C., Rondepierre, A.: Optimal convergence rates for Nesterov acceleration. *SIAM J. Optim.* **29**(4), 3131–3153 (2019)
15. Balti, M., May, R.: Asymptotic for the perturbed heavy ball system with vanishing damping term. *Evol. Equ. Control Theory* **6**(2), 177–186 (2017)
16. Bégout, P., Bolte, J., Jendoubi, M.A.: On damped second-order gradient systems. *J. Differ. Equ.* **259**, 3115–3143 (2015)
17. Boţ, R.I., Csetnek, E.R., László, S.C.: An inertial forward–backward algorithm for minimizing the sum of two non-convex functions. *Euro J. Comput. Optim.* **4**(1), 3–25 (2016)
18. Boţ, R.I., Csetnek, E.R., László, S.C.: Approaching nonsmooth nonconvex minimization through second-order proximal-gradient dynamical systems. *J. Evol. Equ.* **18**(3), 1291–1318 (2018)
19. Boţ, R.I., Csetnek, E.R., László, S.C.: A second order dynamical approach with variable damping to nonconvex smooth minimization. *Appl. Anal.* **99**(3), 361–378 (2018)
20. Boţ, R.I., Csetnek, E.R., László, S.C.: A primal-dual dynamical approach to structured convex minimization problems. (2019). [arXiv:1905.08290](https://arxiv.org/abs/1905.08290)

21. Boţ, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian driven damping. (2019). [arXiv:1911.12845](https://arxiv.org/abs/1911.12845)
22. Cabot, A., Engler, H., Gadat, S.: On the long time behavior of second order differential equations with asymptotically small dissipation. *Trans. Am. Math. Soc.* **361**, 5983–6017 (2009)
23. Cabot, A., Engler, H., Gadat, S.: Second order differential equations with asymptotically small dissipation and piecewise at potentials. *Electr. J. Differ. Equ.* **17**, 33–38 (2009)
24. Haraux, A.: *Systèmes Dynamiques Dissipatifs et Applications*. Recherches en Mathématiques Appliquées 17, Masson (1991)
25. László, S.C.: Convergence rates for an inertial algorithm of gradient type associated to a smooth nonconvex minimization. (2018). [arXiv:1811.09616](https://arxiv.org/abs/1811.09616)
26. Muehlebach, M., Jordan, M.I.: A Dynamical systems perspective on Nesterov acceleration. (2019). [arXiv:1905.07436](https://arxiv.org/abs/1905.07436)
27. Nesterov, Y.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR* **269**(3), 543–547 (1983). (Russian)
28. Nesterov, Y.: *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Dordrecht (2004)
29. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. *U.S.S.R. Comput. Math. Math. Phys.* **4**(5), 1–17 (1964)
30. Shi, B., Du, S.S., Jordan, M.I., Su, W.J.: Understanding the acceleration phenomenon via high-resolution differential equations. (2018). [arXiv:1810.08907](https://arxiv.org/abs/1810.08907)
31. Sontag, E.D.: *Mathematical Control Theory. Deterministic Finite-Dimensional Systems*. Texts in Applied Mathematics 6, vol. 2. Springer, New York (1998)
32. Su, W., Boyd, S., Candes, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *J. Mach. Learn. Res.* **17**, 1–43 (2016)
33. Zhang, J., Mokhtari, A., Sra, S., Jadbabaie, A.: Direct Runge-Kutta discretization achieves acceleration. In: *Advances in Neural Information Processing Systems*, pp. 3900–3909 (2018)

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