

Second order forward-backward dynamical systems for monotone inclusion problems

Radu Ioan Boţ *

Ernö Robert Csetnek †

August 29, 2018

Abstract. We begin by considering second order dynamical systems of the form $\ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0$, where $B : \mathcal{H} \rightarrow \mathcal{H}$ is a cocoercive operator defined on a real Hilbert space \mathcal{H} , $\lambda : [0, +\infty) \rightarrow [0, +\infty)$ is a relaxation function and $\gamma : [0, +\infty) \rightarrow [0, +\infty)$ a damping function, both depending on time. For the generated trajectories, we show existence and uniqueness of the generated trajectories as well as their weak asymptotic convergence to a zero of the operator B . The framework allows to address from similar perspectives second order dynamical systems associated with the problem of finding zeros of the sum of a maximally monotone operator and a cocoercive one. This captures as particular case the minimization of the sum of a nonsmooth convex function with a smooth convex one. Furthermore, we prove that when B is the gradient of a smooth convex function the value of the latter converges along the ergodic trajectory to its minimal value with a rate of $\mathcal{O}(1/t)$.

Key Words. dynamical systems, monotone inclusions, convex optimization problems, continuous forward-backward method

AMS subject classification. 34G25, 47J25, 47H05, 90C25

1 Introduction and preliminaries

This paper is motivated by the heavy ball with friction dynamical system

$$\ddot{x} + \gamma\dot{x} + \nabla f(x) = 0, \quad (1)$$

which is a nonlinear oscillator with damping $\gamma > 0$ and potential $f : \mathcal{H} \rightarrow \mathbb{R}$, supposed to be a convex and differentiable function defined on the real Hilbert space \mathcal{H} . The system (1) is a simplified version of the differential system describing the motion of a heavy ball that keeps rolling over the graph of the function f under its own inertia until friction stops it at a critical point of f (see [14]). Motivated by different models of friction, in [3, 23] a generalized version of (1) has been investigated in finite-dimensional spaces, by replacing the damping $\gamma\dot{x}$ with $\partial\Phi(\dot{x})$, which is the convex subdifferential of a convex function $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$ at \dot{x} .

The second order dynamical system (1) has been considered by several authors in the context of minimizing the function f , these investigations being either concerned with the asymptotic convergence of the generated trajectories to a critical point of f or with the convergence of the function value along the trajectories to its global minimum value (see [4, 8, 9, 14]). It is also worth to mention that the time discretization of the heavy ball with friction dynamical system leads

*University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at.

†University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: ernoe.robert.csetnek@univie.ac.at. Research supported by FWF (Austrian Science Fund), Lise Meitner Programme, project M 1682-N25.

to so-called inertial type algorithms, which are numerical schemes sharing the feature that the current iterate of the generated sequence is defined by making use of the previous two iterates (see, for instance, [4–6, 22, 28, 31]).

In order to approach the minimization of f over a nonempty, convex and closed set $C \subseteq \mathcal{H}$, the gradient-projection second order dynamical system

$$\ddot{x} + \gamma \dot{x} + x - P_C(x - \eta \nabla f(x)) = 0 \quad (2)$$

has been considered, where $P_C : \mathcal{H} \rightarrow C$ denotes the projection onto the set C and $\eta > 0$. Convergence statements for the trajectories to a global minimizer of f over C have been provided in [8, 9]. Furthermore, in [9], these investigations have been expanded to more general second order dynamical systems of the form

$$\ddot{x} + \gamma \dot{x} + x - Tx = 0, \quad (3)$$

where $T : \mathcal{H} \rightarrow \mathcal{H}$ is a nonexpansive operator. It has been shown that when $\gamma^2 > 2$ the trajectory of (8) converges weakly to an element in the fixed points set of T , provided the latter is nonempty.

In this manuscript, we first treat the second order dynamical system

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0, \quad (4)$$

where $B : \mathcal{H} \rightarrow \mathcal{H}$ is a cocoercive operator, $\lambda : [0, +\infty) \rightarrow [0, +\infty)$ is a relaxation function in time and $\gamma : [0, +\infty) \rightarrow [0, +\infty)$ is a continuous damping parameter. We refer the reader to [10, 15, 24, 25, 35] for other works where second order differential equations with time dependent damping have been considered and investigated in connection with optimization problems. On the other hand, second order dynamical systems governed by cocoercive operators have been recently considered also in [12], however, with constant relaxation and damping functions. The existence and uniqueness of strong global solutions for (4) is obtained by applying the classical Cauchy-Lipschitz-Picard Theorem (see [27]). We also show that under mild assumptions on the relaxation function the trajectory $x(t)$ converges weakly as $t \rightarrow +\infty$ to a zero of the operator B , provided the latter has a nonempty set of zeros. To this end we use the continuous version of the Opial Lemma (see also [4, 8, 9], where similar techniques have been used).

Further, we approach the problem of finding a zero of the sum of a maximally monotone operator and a cocoercive one via a second order dynamical system formulated by making use of the resolvent of the set-valued operator, see (24). Dynamical systems of implicit type have been already considered in the literature in [1, 2, 13, 16, 18, 20, 21]. We specialize these investigations to the minimization of the sum of a nonsmooth convex function with a smooth convex function, which is approached by means of a second order dynamical system of forward-backward type. This fact allows us to recover and improve results given in [8, 9] in the context of studying (2). We also emphasize the fact that the explicit discretization of the second order forward-backward dynamical system gives rise to a relaxed forward-backward algorithm with inertial effects. By approaching minimization problems from continuous perspective we expect to gain more insights into how to properly choose the relaxation and damping parameters in the corresponding iterative schemes in order to improve their convergence behavior. Finally, whenever B is the gradient of a smooth convex function we show that the value of the latter converges along the ergodic trajectories generated by (4) to its minimum value with a rate of convergence of $\mathcal{O}(1/t)$.

Throughout this paper $\mathbb{N} = \{0, 1, 2, \dots\}$ denotes the set of nonnegative integers and \mathcal{H} a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

2 Existence and uniqueness of strong global solutions

This section is devoted to the study of existence and uniqueness of strong global solutions of a second order dynamical system governed by Lipschitz continuous operators.

Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be an L -Lipschitz continuous operator (that is $L \geq 0$ and $\|Bx - By\| \leq L\|x - y\|$ for all $x, y \in \mathcal{H}$), $\lambda, \gamma : [0, +\infty) \rightarrow [0, +\infty)$ be Lebesgue measurable functions, $u_0, v_0 \in \mathcal{H}$ and consider the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \quad (5)$$

As in [2, 16], we consider the following definition of an absolutely continuous function.

Definition 1 (see, for instance, [2, 16]) A function $x : [0, b] \rightarrow \mathcal{H}$ (where $b > 0$) is said to be absolutely continuous if one of the following equivalent properties holds:

(i) there exists an integrable function $y : [0, b] \rightarrow \mathcal{H}$ such that

$$x(t) = x(0) + \int_0^t y(s)ds \quad \forall t \in [0, b];$$

(ii) x is continuous and its distributional derivative \dot{x} is Lebesgue integrable on $[0, b]$;

(iii) for every $\varepsilon > 0$, there exists $\eta > 0$ such that for any finite family of intervals $I_k = (a_k, b_k)$ we have the implication

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

Remark 1 (a) It follows from the definition that an absolutely continuous function is differentiable almost everywhere, its derivative coincides with its distributional derivative almost everywhere and one can recover the function from its derivative $\dot{x} = y$ by the integration formula (i).

(b) If $x : [0, b] \rightarrow \mathcal{H}$ (where $b > 0$) is absolutely continuous, then the function $z = B \circ x$ is absolutely continuous, too. This can be easily seen by using the characterization of absolute continuity in Definition 1(iii). Moreover, z is almost everywhere differentiable and the inequality $\|\dot{z}(\cdot)\| \leq L\|\dot{x}(\cdot)\|$ holds almost everywhere.

Definition 2 We say that $x : [0, +\infty) \rightarrow \mathcal{H}$ is a strong global solution of (5) if the following properties are satisfied:

- (i) $x, \dot{x} : [0, +\infty) \rightarrow \mathcal{H}$ are locally absolutely continuous, in other words, absolutely continuous on each interval $[0, b]$ for $0 < b < +\infty$;
- (ii) $\ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0$ for almost every $t \in [0, +\infty)$;
- (iii) $x(0) = u_0$ and $\dot{x}(0) = v_0$.

For proving the existence and uniqueness of strong global solutions of (5) we use the Cauchy-Lipschitz-Picard Theorem for absolutely continuous trajectories. The key observation here is that one can rewrite (5) as a particular first order dynamical system in a suitably chosen product space (see also [7]).

Theorem 2 Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be an L -Lipschitz continuous operator and $\lambda, \gamma : [0, +\infty) \rightarrow [0, +\infty)$ be Lebesgue measurable functions such that $\lambda, \gamma \in L^1_{\text{loc}}([0, +\infty))$ (that is $\lambda, \gamma \in L^1([0, b])$ for every $0 < b < +\infty$). Then for each $u_0, v_0 \in \mathcal{H}$ there exists a unique strong global solution of the dynamical system (5).

Proof. The system (5) can be equivalently written as a first order dynamical system in the phase space $\mathcal{H} \times \mathcal{H}$

$$\begin{cases} \dot{Y}(t) = F(t, Y(t)) \\ Y(0) = (u_0, v_0), \end{cases} \quad (6)$$

with

$$Y : [0, +\infty) \rightarrow \mathcal{H} \times \mathcal{H}, \quad Y(t) = (x(t), \dot{x}(t))$$

and

$$F : [0, +\infty) \times \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H}, \quad F(t, u, v) = (v, -\gamma(t)v - \lambda(t)Bu).$$

We endow $\mathcal{H} \times \mathcal{H}$ with scalar product $\langle (u, v), (\bar{u}, \bar{v}) \rangle_{\mathcal{H} \times \mathcal{H}} = \langle u, \bar{u} \rangle + \langle v, \bar{v} \rangle$ and corresponding norm $\|(u, v)\|_{\mathcal{H} \times \mathcal{H}} = \sqrt{\|u\|^2 + \|v\|^2}$.

(a) For arbitrary $u, \bar{u}, v, \bar{v} \in \mathcal{H}$, by using the Lipschitz continuity of the involved operators, we obtain

$$\begin{aligned} \|F(t, u, v) - F(t, \bar{u}, \bar{v})\|_{\mathcal{H} \times \mathcal{H}} &= \sqrt{\|v - \bar{v}\|^2 + \|\gamma(t)(\bar{v} - v) + \lambda(t)(B\bar{u} - Bu)\|^2} \\ &\leq \sqrt{(1 + 2\gamma^2(t))\|v - \bar{v}\|^2 + 2L^2\lambda^2(t)\|u - \bar{u}\|^2} \\ &\leq \sqrt{1 + 2\gamma^2(t) + 2L^2\lambda^2(t)} \|(u, \bar{u}) - (v, \bar{v})\|_{\mathcal{H} \times \mathcal{H}} \\ &\leq (1 + \sqrt{2}\gamma(t) + \sqrt{2}L\lambda(t)) \|(u, \bar{u}) - (v, \bar{v})\|_{\mathcal{H} \times \mathcal{H}} \quad \forall t \geq 0. \end{aligned}$$

As $\lambda, \gamma \in L^1_{\text{loc}}([0, +\infty))$, the Lipschitz constant of $F(t, \cdot, \cdot)$ is locally integrable.

(b) Next we show that

$$\forall u, v \in \mathcal{H}, \quad \forall b > 0, \quad F(\cdot, u, v) \in L^1([0, b], \mathcal{H} \times \mathcal{H}). \quad (7)$$

For arbitrary $u, v \in \mathcal{H}$ and $b > 0$ it holds

$$\begin{aligned} \int_0^b \|F(t, u, v)\|_{\mathcal{H} \times \mathcal{H}} dt &= \int_0^b \sqrt{\|v\|^2 + \|\gamma(t)v + \lambda(t)Bu\|^2} dt \\ &\leq \int_0^b \sqrt{(1 + 2\gamma^2(t))\|v\|^2 + 2\lambda^2(t)\|Bu\|^2} dt \\ &\leq \int_0^b \left((1 + \sqrt{2}\gamma(t))\|v\| + \sqrt{2}\lambda(t)\|Bu\| \right) dt \end{aligned}$$

and from here, by using the assumptions made on λ, γ , (7) follows.

In the light of the statements (a) and (b), the existence and uniqueness of a strong global solution for (6) follow from the Cauchy-Lipschitz-Picard Theorem for first order dynamical systems (see, for example, [27, Proposition 6.2.1]). The conclusion is a consequence of the equivalence of (5) and (6). \blacksquare

3 Convergence of the trajectories

In this section we address the convergence properties of the trajectories generated by the dynamical system (5) by assuming that $B : \mathcal{H} \rightarrow \mathcal{H}$ is β -cocoercive for $\beta > 0$, that is $\beta\|Bx - By\|^2 \leq \langle x - y, Bx - By \rangle$ for all $x, y \in \mathcal{H}$. This implies that B is $\frac{1}{\beta}$ -Lipschitz continuous. If $B = \nabla g$, where $g : \mathcal{H} \rightarrow \mathbb{R}$ is a convex and differentiable function such that ∇g is $\frac{1}{\beta}$ -Lipschitz continuous, then the reverse implication holds, too. Indeed, according to the Baillon-Haddad Theorem, ∇g is a β -cocoercive operator (see [17, Corollary 18.16]).

The following results, which can be interpreted as continuous versions of the quasi-Fejér monotonicity for sequences, plays an important role in the forthcoming investigations. For their proofs we refer the reader to [2, Lemma 5.1] and [2, Lemma 5.2], respectively.

Lemma 3 Suppose that $F : [0, +\infty) \rightarrow \mathbb{R}$ is locally absolutely continuous and bounded below and that there exists $G \in L^1([0, +\infty))$ such that for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}F(t) \leq G(t).$$

Then there exists $\lim_{t \rightarrow \infty} F(t) \in \mathbb{R}$.

Lemma 4 If $1 \leq p < \infty$, $1 \leq r \leq \infty$, $F : [0, +\infty) \rightarrow [0, +\infty)$ is locally absolutely continuous, $F \in L^p([0, +\infty))$, $G : [0, +\infty) \rightarrow \mathbb{R}$, $G \in L^r([0, +\infty))$ and for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}F(t) \leq G(t),$$

then $\lim_{t \rightarrow +\infty} F(t) = 0$.

The next result which we recall here is the continuous version of the Opial Lemma (see, for example, [2, Lemma 5.3], [1, Lemma 1.10]).

Lemma 5 Let $S \subseteq \mathcal{H}$ be a nonempty set and $x : [0, +\infty) \rightarrow \mathcal{H}$ a given map. Assume that

- (i) for every $x^* \in S$, $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$ exists;
- (ii) every weak sequential cluster point of the map x belongs to S .

Then there exists $x_\infty \in S$ such that $x(t)$ converges weakly to x_∞ as $t \rightarrow +\infty$.

In order to prove the convergence of the trajectories of (5), we make the following assumptions on the relaxation function λ and the damping parameter γ , respectively:

- (A1) $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq \frac{1 + \theta}{\beta}. \quad (8)$$

Due to Definition 1 and Remark 1(a), $\dot{\lambda}(t), \dot{\gamma}(t)$ exists for almost every $t \geq 0$ and $\dot{\lambda}, \dot{\gamma}$ are Lebesgue integrable on each interval $[0, b]$ for $0 < b < +\infty$. This combined with $\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t)$ and the fact that λ, γ take only positive values yield the existence of a positive lower bound $\underline{\lambda}$ for λ and of a positive upper bound $\bar{\gamma}$ for γ . Furthermore, the second assumption in (8) provides also a positive upper bound $\bar{\lambda}$ for λ and a positive lower bound $\underline{\gamma}$ for γ . Notice that the couple of functions

$$\lambda(t) = \frac{1}{ae^{-\rho t} + b} \text{ and } \gamma(t) = a'e^{-\rho' t} + b',$$

where $a, a', \rho, \rho' \geq 0$ and $b, b' > 0$ fulfill the inequality $b'^2 b > \frac{1}{\beta}$, verify the conditions in assumption (A1).

We would also like to point out that under the conditions considered in (A1) the global version of the Picard-Lindelöf Theorem allows us to conclude that, for $u_0, v_0 \in \mathcal{H}$, there exists a unique trajectory $x : [0, +\infty) \rightarrow \mathcal{H}$ which is a C^2 function and which satisfies the relation (ii) in Definition 2 for every $t \in [0, +\infty)$. The considerations we make in the following take into account this fact.

We state now the convergence result.

Theorem 6 Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a β -cocoercive operator for $\beta > 0$ such that $\text{Zer } B := \{u \in \mathcal{H} : Bu = 0\} \neq \emptyset$, $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A1) and $u_0, v_0 \in \mathcal{H}$. Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (5). Then the following statements are true:

- (i) the trajectory x is bounded and $\dot{x}, \ddot{x}, Bx \in L^2([0, +\infty); \mathcal{H})$;
- (ii) $\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} B(x(t)) = 0$;
- (iii) $x(t)$ converges weakly to an element in $\text{Zer } B$ as $t \rightarrow +\infty$.

Proof. (i) Take an arbitrary $x^* \in \text{Zer } B$ and consider for every $t \in [0, +\infty)$ the function $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$. We have $\dot{h}(t) = \langle x(t) - x^*, \dot{x}(t) \rangle$ and $\ddot{h}(t) = \|\dot{x}(t)\|^2 + \langle x(t) - x^*, \ddot{x}(t) \rangle$ for every $t \in [0, +\infty)$. Taking into account (5), we get for every $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t) \langle x(t) - x^*, B(x(t)) \rangle = \|\dot{x}(t)\|^2. \quad (9)$$

The cocoercivity of B and the fact that $Bx^* = 0$ yields for every $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \beta\lambda(t)\|B(x(t))\|^2 \leq \|\dot{x}(t)\|^2.$$

Taking again into account (5) one obtains for every $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{\beta}{\lambda(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 \leq \|\dot{x}(t)\|^2$$

or, equivalently,

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{\beta\gamma(t)}{\lambda(t)} \frac{d}{dt}(\|\dot{x}(t)\|^2) + \left(\frac{\beta\gamma^2(t)}{\lambda(t)} - 1 \right) \|\dot{x}(t)\|^2 + \frac{\beta}{\lambda(t)} \|\ddot{x}(t)\|^2 \leq 0.$$

Combining this inequality with

$$\frac{\gamma(t)}{\lambda(t)} \frac{d}{dt}(\|\dot{x}(t)\|^2) = \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) - \frac{\dot{\gamma}(t)\lambda(t) - \gamma(t)\dot{\lambda}(t)}{\lambda^2(t)} \|\dot{x}(t)\|^2 \quad (10)$$

and

$$\gamma(t)\dot{h}(t) = \frac{d}{dt}(\gamma h)(t) - \dot{\gamma}(t)h(t) \geq \frac{d}{dt}(\gamma h)(t), \quad (11)$$

it yields for every $t \in [0, +\infty)$

$$\begin{aligned} & \ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \\ & \beta \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) + \left(\frac{\beta\gamma^2(t)}{\lambda(t)} + \beta \frac{-\dot{\gamma}(t)\lambda(t) + \gamma(t)\dot{\lambda}(t)}{\lambda^2(t)} - 1 \right) \|\dot{x}(t)\|^2 + \frac{\beta}{\lambda(t)} \|\ddot{x}(t)\|^2 \leq 0. \end{aligned}$$

Now, assumption (A1) delivers for almost every $t \in [0, +\infty)$ the inequality

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) + \theta \|\dot{x}(t)\|^2 + \beta \bar{\lambda}^{-1} \|\ddot{x}(t)\|^2 \leq 0. \quad (12)$$

This implies that the function $t \mapsto \dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2$, which is locally absolutely continuous, is monotonically decreasing. Hence there exists a real number M such that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \leq M, \quad (13)$$

which yields that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \underline{\gamma}h(t) \leq M.$$

By multiplying this inequality with $\exp(\underline{\gamma}t)$ and then integrating from 0 to T , where $T > 0$, one easily obtains

$$h(T) \leq h(0) \exp(-\underline{\gamma}T) + \frac{M}{\underline{\gamma}}(1 - \exp(-\underline{\gamma}T)),$$

thus

$$h \text{ is bounded} \quad (14)$$

and, consequently,

$$\text{the trajectory } x \text{ is bounded.} \quad (15)$$

On the other hand, from (13), it follows that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \beta \gamma \bar{\lambda}^{-1} \|\dot{x}(t)\|^2 \leq M,$$

hence

$$\langle x(t) - x^*, \dot{x}(t) \rangle + \beta \gamma \bar{\lambda}^{-1} \|\dot{x}(t)\|^2 \leq M.$$

This inequality in combination with (15) yields

$$\dot{x} \text{ is bounded,} \quad (16)$$

which further implies that

$$\dot{h} \text{ is bounded.} \quad (17)$$

Integrating the inequality (12) we obtain that there exists a real number $N \in \mathbb{R}$ such that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 + \theta \int_0^t \|\dot{x}(s)\|^2 ds + \beta \bar{\lambda}^{-1} \int_0^t \|\ddot{x}(s)\|^2 ds \leq N.$$

From here, via (17), we conclude that $\dot{x}(\cdot), \ddot{x}(\cdot) \in L^2([0, +\infty); \mathcal{H})$. Finally, from (5) and (A1) we deduce $Bx \in L^2([0, +\infty); \mathcal{H})$ and the proof of (i) is complete.

(ii) For every $t \in [0, +\infty)$ it holds

$$\frac{d}{dt} \left(\frac{1}{2} \|\dot{x}(t)\|^2 \right) = \langle \dot{x}(t), \ddot{x}(t) \rangle \leq \frac{1}{2} \|\dot{x}(t)\|^2 + \frac{1}{2} \|\ddot{x}(t)\|^2$$

and Lemma 4 together with (i) lead to $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$.

Further, by taking into consideration Remark 1(b), for every $t \in [0, +\infty)$ we have

$$\frac{d}{dt} \left(\frac{1}{2} \|B(x(t))\|^2 \right) = \left\langle B(x(t)), \frac{d}{dt}(Bx(t)) \right\rangle \leq \frac{1}{2} \|B(x(t))\|^2 + \frac{1}{2\beta^2} \|\dot{x}(t)\|^2.$$

By using again Lemma 4 and (i) we get $\lim_{t \rightarrow +\infty} B(x(t)) = 0$, while the fact that $\lim_{t \rightarrow +\infty} \ddot{x}(t) = 0$ follows from (5) and (A2).

(iii) We are going to prove that both assumptions in Opial Lemma are fulfilled. The first one concerns the existence of $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$. As seen in the proof of part (i), the function $t \mapsto \dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2$ is monotonically decreasing, thus from (i), (ii) and (A1) we deduce that $\lim_{t \rightarrow +\infty} \gamma(t)h(t)$ exists and it is a real number. By taking also into account that $\exists \lim_{t \rightarrow +\infty} \gamma(t) \in (0, \infty)$, we obtain the existence of $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$.

We come now to the second assumption of the Opial Lemma. Let \bar{x} be a weak sequential cluster point of x , that is, there exists a sequence $t_n \rightarrow +\infty$ (as $n \rightarrow +\infty$) such that $(x(t_n))_{n \in \mathbb{N}}$ converges weakly to \bar{x} . Since B is a maximally monotone operator (see for instance [17, Example 20.28]), its graph is sequentially closed with respect to the weak-strong topology of the product space $\mathcal{H} \times \mathcal{H}$. By using also that $\lim_{n \rightarrow +\infty} B(x(t_n)) = 0$, we conclude that $B\bar{x} = 0$, hence $\bar{x} \in \text{Zer } B$ and the proof is complete. \blacksquare

A standard choice of a cocoercive operator defined on a real Hilbert spaces is $B = \text{Id} - T$, where $T : \mathcal{H} \rightarrow \mathcal{H}$ is a *nonexpansive operator*, that is, a 1-Lipschitz continuous operator. As it easily follows from the nonexpansiveness of T , B is in this case 1/2-cocoercive. For this particular operator B the dynamical system (5) becomes

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)(x(t) - T(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \quad (18)$$

while assumption (A1) reads

(A2) $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq 2(1 + \theta). \quad (19)$$

Theorem 6 gives rise to the following result.

Corollary 7 *Let $T : \mathcal{H} \rightarrow \mathcal{H}$ be a nonexpansive operator such that $\text{Fix } T := \{u \in \mathcal{H} : Tu = u\} \neq \emptyset$, $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A2) and $u_0, v_0 \in \mathcal{H}$. Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (18). Then the following statements are true:*

- (i) *the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - T)x \in L^2([0, +\infty); \mathcal{H})$;*
- (ii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - T)(x(t)) = 0$;*
- (iii) *$x(t)$ converges weakly to a point in $\text{Fix } T$ as $t \rightarrow +\infty$.*

Remark 8 In the particular case when $\gamma(t) = \gamma > 0$ for all $t \geq 0$ and $\lambda(t) = 1$ for all $t \in [0, +\infty)$ the dynamical system (18) becomes

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) - T(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \quad (20)$$

The convergence of the trajectories generated by (20) has been studied in [9, Theorem 3.2] under the condition $\gamma^2 > 2$. In this case (A2) is obviously fulfilled for an arbitrary $0 < \theta \leq (\gamma^2 - 2)/2$. However, different to [9], we allow in Corollary 7 nonconstant damping and relaxation functions depending on time. We would also like to notice that in [4] an anisotropic damping has been considered in the context of approaching the minimization of a smooth convex function via second order dynamical systems.

We close the section by addressing an immediate consequence of the above corollary applied to second order dynamical systems governed by averaged operators. The operator $R : \mathcal{H} \rightarrow \mathcal{H}$ is said to be α -averaged for $\alpha \in (0, 1)$, if there exists a nonexpansive operator $T : \mathcal{H} \rightarrow \mathcal{H}$ such that $R = (1 - \alpha)\text{Id} + \alpha T$. For $\alpha = \frac{1}{2}$ we obtain as an important representative of this class the *firmly nonexpansive* operators. For properties and insights concerning these families of operators we refer to the monograph [17].

We consider the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)(x(t) - R(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{cases} \quad (21)$$

and formulate the assumption

(A3) $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq 2\alpha(1 + \theta). \quad (22)$$

Corollary 9 *Let $R : \mathcal{H} \rightarrow \mathcal{H}$ be an α -averaged operator for $\alpha \in (0, 1)$ such that $\text{Fix } R \neq \emptyset$, $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A3) and $u_0, v_0 \in \mathcal{H}$. Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (21). Then the following statements are true:*

- (i) *the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - R)x \in L^2([0, +\infty); \mathcal{H})$;*
- (ii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - R)(x(t)) = 0$;*
- (iii) *$x(t)$ converges weakly to a point in $\text{Fix } R$ as $t \rightarrow +\infty$.*

Proof. Since R is α -averaged, there exists a nonexpansive operator $T : \mathcal{H} \rightarrow \mathcal{H}$ such that $R = (1 - \alpha)\text{Id} + \alpha T$. The conclusion is a direct consequence of Corollary 7, by taking into account that (21) is equivalent to

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \alpha\lambda(t)(x(t) - T(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases}$$

and $\text{Fix } R = \text{Fix } T$. ■

4 Forward-backward second order dynamical systems

In this section we approach the monotone inclusion problem

$$\text{find } 0 \in A(x) + B(x),$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator and $B : \mathcal{H} \rightarrow \mathcal{H}$ is a β -cocoercive operator for $\beta > 0$ via a second order forward-backward dynamical system with relaxation and damping functions depending on time.

For readers convenience we recall at the beginning some standard notions and results in monotone operator theory (see also [17, 19, 34]). For an arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\text{Gr } A = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its graph. We use also the notation $\text{Zer } A = \{x \in \mathcal{H} : 0 \in Ax\}$ for the set of zeros of A . We say that A is monotone, if $\langle x - y, u - v \rangle \geq 0$ for all $(x, u), (y, v) \in \text{Gr } A$. A monotone operator A is said to be maximally monotone, if there exists no proper monotone extension of the graph of A on $\mathcal{H} \times \mathcal{H}$. The resolvent of A , $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by $J_A = (\text{Id} + A)^{-1}$. If A is maximally monotone, then $J_A : \mathcal{H} \rightarrow \mathcal{H}$ is single-valued and maximally monotone (see [17, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [17, Proposition 23.2])

$$p \in J_{\gamma A} x \text{ if and only if } (p, \gamma^{-1}(x - p)) \in \text{Gr } A. \quad (23)$$

The operator A is said to be uniformly monotone if there exists an increasing function $\phi_A : [0, +\infty) \rightarrow [0, +\infty]$ that vanishes only at 0 and fulfills $\langle x - y, u - v \rangle \geq \phi_A(\|x - y\|)$ for all $(x, u), (y, v) \in \text{Gr } A$. A popular class of operators having this property is the one of strongly monotone operators. We say that A is γ -strongly monotone for $\gamma > 0$, if $\langle x - y, u - v \rangle \geq \gamma\|x - y\|^2$ for all $(x, u), (y, v) \in \text{Gr } A$.

For $\eta > 0$ we consider the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t) \left[x(t) - J_{\eta A} \left(x(t) - \eta B(x(t)) \right) \right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \quad (24)$$

Further, we consider the following assumption, where $\delta := \frac{4\beta - \eta}{2\beta}$:

(A4) $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq \frac{2(1 + \theta)}{\delta}. \quad (25)$$

Theorem 10 Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $B : \mathcal{H} \rightarrow \mathcal{H}$ be a β -cocoercive operator for $\beta > 0$ such that $\text{Zer}(A + B) \neq \emptyset$. Let $\eta \in (0, 2\beta)$ and set $\delta := \frac{4\beta - \eta}{2\beta}$. Let $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A4), $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (24). Then the following statements are true:

- (i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - J_{\eta A} \circ (\text{Id} - \eta B))x \in L^2([0, +\infty); \mathcal{H})$;
- (ii) $\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - J_{\eta A} \circ (\text{Id} - \eta B))(x(t)) = 0$;
- (iii) $x(t)$ converges weakly to a point in $\text{Zer}(A + B)$ as $t \rightarrow +\infty$;
- (iv) if $x^* \in \text{Zer}(A + B)$, then $B(x(\cdot)) - Bx^* \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t \rightarrow +\infty} B(x(t)) = Bx^*$ and B is constant on $\text{Zer}(A + B)$;
- (v) if A or B is uniformly monotone, then $x(t)$ converges strongly to the unique point in $\text{Zer}(A + B)$ as $t \rightarrow +\infty$.

Proof. (i)-(iii) It is immediate that the dynamical system (24) can be written in the form

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)(x(t) - R(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \quad (26)$$

where $R = J_{\eta A} \circ (\text{Id} - \eta B)$. According to [17, Corollary 23.8 and Remark 4.24(iii)], $J_{\eta A}$ is $1/2$ -cocoercive. Moreover, by [17, Proposition 4.33], $\text{Id} - \eta B$ is $\eta/(2\beta)$ -averaged. Combining this with [30, Theorem 3(b)], we derive that R is $1/\delta$ -averaged. The statements (i)-(iii) follow now from Corollary 9 by noticing that $\text{Fix } R = \text{Zer}(A + B)$ (see [17, Proposition 25.1(iv)]).

(iv) The fact that B is constant on $\text{Zer}(A + B)$ follows from the cocoercivity of B and the monotonicity of A . A proof of this statement when A is the subdifferential of a proper, convex and lower semicontinuous function is given for instance in [1, Lemma 1.7].

Let be an arbitrary $x^* \in \text{Zer}(A + B)$. From the definition of the resolvent we have for every $t \in [0, +\infty)$

$$-B(x(t)) - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \in A\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right), \quad (27)$$

which combined with $-Bx^* \in Ax^*$ and the monotonicity of A leads to

$$0 \leq \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -B(x(t)) + Bx^* - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle. \quad (28)$$

The cocoercivity of B yields for every $t \in [0, +\infty)$

$$\begin{aligned} \beta\|B(x(t)) - Bx^*\|^2 &\leq \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t), -B(x(t)) + Bx^* \right\rangle - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 \\ &\quad + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle \\ &\leq \frac{1}{2\beta} \left\| \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) \right\|^2 + \frac{\beta}{2}\|B(x(t)) - Bx^*\|^2 \\ &\quad - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle \\ &= \frac{\eta - 2\beta}{2\eta\beta\lambda^2(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + \frac{\beta}{2}\|B(x(t)) - Bx^*\|^2 \\ &\quad + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle \\ &\leq \frac{\beta}{2}\|B(x(t)) - Bx^*\|^2 + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle. \end{aligned}$$

For evaluating the last term of the above inequality we use the function $h : [0, +\infty) \rightarrow \mathbb{R}$, $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$, already used in the proof of Theorem 6. From

$$\left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle = -\frac{1}{\eta\lambda(t)} \left(\ddot{h}(t) + \gamma(t)\dot{h}(t) - \|\dot{x}(t)\|^2 \right) \quad (29)$$

we obtain for every $t \in [0, +\infty)$

$$\frac{\beta\lambda(t)}{2}\|B(x(t)) - Bx^*\|^2 + \frac{1}{\eta} \left(\ddot{h}(t) + \gamma(t)\dot{h}(t) \right) \leq \frac{1}{\eta}\|\dot{x}(t)\|^2.$$

Taking into account also the relation (11) and the bounds for λ , we get for every $t \in [0, +\infty)$

$$\frac{\beta\lambda}{2}\|B(x(t)) - Bx^*\|^2 + \frac{1}{\eta} \left(\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) \right) \leq \frac{1}{\eta}\|\dot{x}(t)\|^2.$$

After integration we obtain that for every $T \in [0, +\infty)$

$$\frac{\beta\lambda}{2} \int_0^T \|B(x(t)) - Bx^*\|^2 dt + \frac{1}{\eta} \left(\dot{h}(T) - \dot{h}(0) + \gamma(T)h(T) - \gamma(0)h(0) \right) \leq \frac{1}{\eta} \int_0^T \|\dot{x}(t)\|^2 dt.$$

Since $\dot{x} \in L^2([0, +\infty); \mathcal{H})$, γ has a positive upper bound, $h(T) \geq 0, \gamma(T) \geq 0$ for every $T \in [0, +\infty)$ and $\lim_{T \rightarrow +\infty} \dot{h}(T) = 0$, it follows that $B(x(\cdot)) - Bx^* \in L^2([0, +\infty); \mathcal{H})$.

Further, by taking into consideration Remark 1(b), we have

$$\frac{d}{dt} \left(\frac{1}{2}\|B(x(t)) - Bx^*\|^2 \right) = \left\langle B(x(t)) - Bx^*, \frac{d}{dt}(Bx(t)) \right\rangle \leq \frac{1}{2}\|B(x(t)) - Bx^*\|^2 + \frac{1}{2\beta^2}\|\dot{x}(t)\|^2$$

and from here, in the light of Lemma 4, it follows that $\lim_{t \rightarrow +\infty} B(x(t)) = Bx^*$.

(v) Let x^* be the unique element of $\text{Zer}(A + B)$. For the beginning we suppose that A is uniformly monotone with corresponding function $\phi_A : [0, +\infty) \rightarrow [0, +\infty]$, which is increasing and vanishes only at 0.

By similar arguments as in the proof of statement (iv), for every $t \in [0, +\infty)$ we have

$$\begin{aligned} \phi_A \left(\left\| \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^* \right\| \right) &\leq \\ \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -B(x(t)) + Bx^* - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle, \end{aligned}$$

which combined with the monotonicity of B yields

$$\begin{aligned} \phi_A \left(\left\| \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^* \right\| \right) &\leq \\ \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t), -B(x(t)) + Bx^* \right\rangle - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + \\ \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle &\leq \\ \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t), -B(x(t)) + Bx^* \right\rangle + \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \right\rangle. \end{aligned}$$

As λ and γ are bounded by positive constants, by using (i)-(iv) it follows that the right-hand side of the last inequality converges to 0 as $t \rightarrow +\infty$. Hence

$$\lim_{t \rightarrow +\infty} \phi_A \left(\left\| \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^* \right\| \right) = 0$$

and the properties of the function ϕ_A allow to conclude that $\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*$ converges strongly to 0 as $t \rightarrow +\infty$. By using again the boundedness of λ and γ and assumption (ii) we obtain that $x(t)$ converges strongly to x^* as $t \rightarrow +\infty$.

Finally, suppose that B is uniformly monotone with corresponding function $\phi_B : [0, +\infty) \rightarrow [0, +\infty]$, which is increasing and vanishes only at 0. The conclusion follows by letting t in the inequality

$$\langle x(t) - x^*, B(x(t)) - Bx^* \rangle \geq \phi_B(\|x(t) - x^*\|) \quad \forall t \in [0, +\infty)$$

converge to $+\infty$ and by using that x is bounded and $\lim_{t \rightarrow +\infty} (B(x(t)) - Bx^*) = 0$. \blacksquare

Remark 11 We would like to emphasize the fact that the statements in Theorem 10 remain valid also for $\eta := 2\beta$. Indeed, in this case the cocoercivity of B implies that $\text{Id} - \eta B$ is nonexpansive, hence the operator $R = J_{\eta A} \circ (\text{Id} - \eta B)$ used in the proof is nonexpansive, too, and so the statements in (i)-(iii) follow from Corollary 7. Furthermore, the proof of the statements (iv) and (v) can be repeated also for $\eta = 2\beta$.

In the remaining of this section we turn our attention to optimization problems of the form

$$\min_{x \in \mathcal{H}} f(x) + g(x),$$

where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function and $g : \mathcal{H} \rightarrow \mathbb{R}$ is a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$ -Lipschitz continuous gradient for $\beta > 0$.

We recall some standard notations and facts in convex analysis. For a proper, convex and lower semicontinuous function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, its (convex) subdifferential at $x \in \mathcal{H}$ is defined as

$$\partial f(x) = \{u \in \mathcal{H} : f(y) \geq f(x) + \langle u, y - x \rangle \quad \forall y \in \mathcal{H}\}.$$

When seen as a set-valued mapping, it is a maximally monotone operator (see [32]) and its resolvent is given by $J_{\eta \partial f} = \text{prox}_{\eta f}$ (see [17]), where $\text{prox}_{\eta f} : \mathcal{H} \rightarrow \mathcal{H}$,

$$\text{prox}_{\eta f}(x) = \underset{y \in \mathcal{H}}{\text{argmin}} \left\{ f(y) + \frac{1}{2\eta} \|y - x\|^2 \right\}, \quad (30)$$

denotes the proximal point operator of f and $\eta > 0$. According to [17, Definition 10.5], f is said to be uniformly convex with modulus function $\phi : [0, +\infty) \rightarrow [0, +\infty]$, if ϕ is increasing, vanishes only at 0 and fulfills $f(\alpha x + (1 - \alpha)y) + \alpha(1 - \alpha)\phi(\|x - y\|) \leq \alpha f(x) + (1 - \alpha)f(y)$ for all $\alpha \in (0, 1)$ and $x, y \in \text{dom } f := \{x \in \mathcal{H} : f(x) < +\infty\}$. Notice that if this inequality holds for $\phi = (\nu/2)|\cdot|^2$ for $\nu > 0$, then f is said to be ν -strongly convex.

In the following statement we approach the minimizers of $f + g$ via the second order dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t) \left[x(t) - \text{prox}_{\eta f} \left(x(t) - \eta \nabla g(x(t)) \right) \right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \quad (31)$$

Corollary 12 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function and $g : \mathcal{H} \rightarrow \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$ -Lipschitz continuous gradient for $\beta > 0$ such that $\underset{x \in \mathcal{H}}{\text{argmin}} \{f(x) + g(x)\} \neq \emptyset$. Let $\eta \in (0, 2\beta]$ and set $\delta := \frac{4\beta - \eta}{2\beta}$. Let $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A4), $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (31). Then the following statements are true:*

- (i) *the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))x \in L^2([0, +\infty); \mathcal{H})$;*
- (ii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))(x(t)) = 0$;*
- (iii) *$x(t)$ converges weakly to a minimizer of $f + g$ as $t \rightarrow +\infty$;*

(iv) if x^* is a minimizer of $f + g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t \rightarrow +\infty} \nabla g(x(t)) = \nabla g(x^*)$ and ∇g is constant on $\operatorname{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}$;
(v) if f or g is uniformly convex, then $x(t)$ converges strongly to the unique minimizer of $f + g$ as $t \rightarrow +\infty$.

Proof. The statements are direct consequences of the corresponding ones from Theorem 10 (see also Remark 11), by choosing $A := \partial f$ and $B := \nabla g$, by taking into account that

$$\operatorname{Zer}(\partial f + \nabla g) = \operatorname{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}.$$

For statement (v) we also use the fact that if f is uniformly convex with modulus ϕ , then ∂f is uniformly monotone with modulus 2ϕ (see [17, Example 22.3(iii)]). ■

Remark 13 Consider again the setting in Remark 8, namely, when $\gamma(t) = \gamma > 0$ for every $t \geq 0$ and $\lambda(t) = 1$ for every $t \in [0, +\infty)$. Furthermore, for C a nonempty, convex, closed subset of \mathcal{H} , let $f = \delta_C$ be the indicator function of C , which is defined as being equal to 0 for $x \in C$ and to $+\infty$, else. The dynamical system (31) attached in this setting to the minimization of g over C becomes

$$\begin{cases} \ddot{x}(t) + \gamma \dot{x}(t) + x(t) - P_C(x(t) - \eta \nabla g(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \quad (32)$$

where P_C denotes the projection onto the set C .

The asymptotic convergence of the trajectories of (32) has been studied in [9, Theorem 3.1] under the conditions $\gamma^2 > 2$ and $0 < \eta \leq 2\beta$. In this case assumption (A4) trivially holds by choosing θ such that $0 < \theta \leq (\gamma^2 - 2)/2 \leq (\delta\gamma^2 - 2)/2$. Thus, in order to verify (A4) in case $\lambda(t) = 1$ for every $t \in [0, +\infty)$ one needs to equivalently assume that $\gamma^2 > 2/\delta$. Since $\delta \geq 1$, this provides a slight improvement over [9, Theorem 3.1] in what concerns the choice of γ . We refer the reader also to [8] for an analysis of the convergence rates of trajectories of the dynamical system (32) when g is endowed with supplementary properties.

For the two main convergence statements provided in this section it was essential to choose the step size η in the interval $(0, 2\beta]$ (see Theorem 10, Remark 11 and Corollary 12). This, because of the fact that in this way we were able to guarantee for the generated trajectories the existence of the limit $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|^2$, where x^* denotes a solution of the problem under investigation. It is interesting to observe that, when dealing with convex optimization problems, one can go also beyond this classical restriction concerning the choice of the step size (a similar phenomenon has been reported also in [1, Section 5.2]). This is pointed out in the following result, which is valid under the assumption

(A5) $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq \eta\theta + \frac{\eta}{\beta} + 1. \quad (33)$$

and for the proof of which we use instead of $\|x(\cdot) - x^*\|^2$ a modified energy functional.

Corollary 14 Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \rightarrow \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$ -Lipschitz continuous gradient for $\beta > 0$ such that $\operatorname{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\} \neq \emptyset$. Let be $\eta > 0$, $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A5), $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (31). Then the following statements are true:

- (i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))x \in L^2([0, +\infty); \mathcal{H})$;
- (ii) $\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))(x(t)) = 0$;
- (iii) $x(t)$ converges weakly to a minimizer of $f + g$ as $t \rightarrow +\infty$;
- (iv) if x^* is a minimizer of $f + g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t \rightarrow +\infty} \nabla g(x(t)) = \nabla g(x^*)$ and ∇g is constant on $\text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}$;
- (v) if f or g is uniformly convex, then $x(t)$ converges strongly to the unique minimizer of $f + g$ as $t \rightarrow +\infty$.

Proof. Consider an arbitrary element $x^* \in \text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\} = \text{Zer}(\partial f + \nabla g)$. Similarly to the proof of Theorem 10(iv), we derive for every $t \in [0, +\infty)$ (see the first inequality after (28))

$$\begin{aligned} & \beta \|\nabla g(x(t)) - \nabla g(x^*)\|^2 \leq \\ & \frac{1}{\lambda(t)} \left(\langle \ddot{x}(t), -\nabla g(x(t)) + \nabla g(x^*) \rangle + \gamma(t) \langle \dot{x}(t), -\nabla g(x(t)) + \nabla g(x^*) \rangle \right) - \\ & \frac{1}{\eta \lambda^2(t)} \|\ddot{x}(t) + \gamma(t) \dot{x}(t)\|^2 + \left\langle x(t) - x^*, -\frac{1}{\eta \lambda(t)} \ddot{x}(t) - \frac{\gamma(t)}{\eta \lambda(t)} \dot{x}(t) \right\rangle. \end{aligned} \quad (34)$$

In what follows we evaluate the right-hand side of the above inequality and introduce to this end the function

$$q : [0, +\infty) \rightarrow \mathbb{R}, \quad q(t) = g(x(t)) - g(x^*) - \langle \nabla g(x^*), x(t) - x^* \rangle.$$

Due to the convexity of g one has

$$q(t) \geq 0 \quad \forall t \geq 0.$$

Further, for every $t \in [0, +\infty)$

$$\dot{q}(t) = \langle \dot{x}(t), \nabla g(x(t)) - \nabla g(x^*) \rangle,$$

thus

$$\gamma(t) \langle \dot{x}(t), -\nabla g(x(t)) + \nabla g(x^*) \rangle = -\gamma(t) \dot{q}(t) = -\frac{d}{dt}(\gamma q)(t) + \dot{\gamma}(t) q(t) \leq -\frac{d}{dt}(\gamma q)(t) \quad (35)$$

On the other hand, for every $t \in [0, +\infty)$

$$\ddot{q}(t) = \langle \ddot{x}(t), \nabla g(x(t)) - \nabla g(x^*) \rangle + \left\langle \dot{x}(t), \frac{d}{dt} \nabla g(x(t)) \right\rangle,$$

hence

$$\langle \ddot{x}(t), -\nabla g(x(t)) + \nabla g(x^*) \rangle \leq -\ddot{q}(t) + \frac{1}{\beta} \|\dot{x}(t)\|^2. \quad (36)$$

We have for almost every $t \in [0, +\infty)$ (see also (10))

$$\begin{aligned} \frac{1}{\lambda(t)} \|\ddot{x}(t) + \gamma(t) \dot{x}(t)\|^2 &= \frac{1}{\lambda(t)} \|\ddot{x}(t)\|^2 + \frac{\gamma^2(t)}{\lambda(t)} \|\dot{x}(t)\|^2 + \\ & \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) - \frac{\dot{\gamma}(t) \lambda(t) - \gamma(t) \dot{\lambda}(t)}{\lambda^2(t)} \|\dot{x}(t)\|^2. \end{aligned} \quad (37)$$

Finally, by multiplying (34) with $\lambda(t)$ and by using (35), (36), (37) and (29) we obtain after rearranging the terms for almost every $t \in [0, +\infty)$ that

$$\begin{aligned} & \beta \lambda(t) \|\nabla g(x(t)) - \nabla g(x^*)\|^2 + \frac{d}{dt^2} \left(\frac{1}{\eta} h + q \right) + \frac{d}{dt} \left(\gamma(t) \left(\frac{1}{\eta} h + q \right) \right) + \frac{1}{\eta} \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) \\ & + \left(\frac{\gamma^2(t)}{\eta \lambda(t)} + \frac{-\dot{\gamma}(t) \lambda(t) + \gamma(t) \dot{\lambda}(t)}{\eta \lambda^2(t)} - \frac{1}{\beta} - \frac{1}{\eta} \right) \|\dot{x}(t)\|^2 + \frac{1}{\eta \lambda(t)} \|\ddot{x}(t)\|^2 \leq 0. \end{aligned}$$

This relation gives rise via (A5) to

$$\begin{aligned} & \beta\lambda(t)\|\nabla g(x(t)) - \nabla g(x^*)\|^2 + \frac{d}{dt^2} \left(\frac{1}{\eta}h + q \right) + \frac{d}{dt} \left(\gamma(t) \left(\frac{1}{\eta}h + q \right) \right) \\ & + \frac{1}{\eta} \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) + \theta \|\dot{x}(t)\|^2 + \frac{1}{\eta\lambda(t)} \|\ddot{x}(t)\|^2 \leq 0, \end{aligned} \quad (38)$$

for almost every $t \in [0, +\infty)$. This implies that the function

$$t \mapsto \frac{d}{dt} \left(\frac{1}{\eta}h + q \right) (t) + \gamma(t) \left(\frac{1}{\eta}h + q \right) (t) + \frac{1}{\eta} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) \quad (39)$$

is monotonically decreasing. Arguing as in the proof of Theorem 6, by taking into account that λ, γ have positive upper and lower bounds, it follows that $\frac{1}{\eta}h + q, h, q, x, \dot{x}, \dot{h}, \dot{q}$ are bounded and $\dot{x}, \ddot{x}, (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))x \in L^2([0, +\infty); \mathcal{H})$. Furthermore, $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$. Since $\frac{d}{dt} (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))x \in L^2([0, +\infty); \mathcal{H})$ (see Remark 1(b)), we derive from Lemma 4 that $\lim_{t \rightarrow +\infty} (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))(x(t)) = 0$. As

$$\ddot{x}(t) = -\gamma(t)\dot{x}(t) - \lambda(t)(\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))(x(t))$$

for every $t \in [0, +\infty)$, we obtain that $\lim_{t \rightarrow +\infty} \ddot{x}(t) = 0$. From (38) it also follows that $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0, +\infty); \mathcal{H})$ and, by applying again Lemma 4, it yields $\lim_{t \rightarrow +\infty} \nabla g(x(t)) = \nabla g(x^*)$. In this way the statements (i), (ii) and (iv) are shown.

(iii) Since the function in (39) is monotonically decreasing, from (i), (ii) and (iv) it follows that the limit $\lim_{t \rightarrow +\infty} \left(\gamma(t) \left(\frac{1}{\eta}h + u \right) (t) \right)$ exists and it is a real number. From $\lim_{t \rightarrow +\infty} \gamma(t) \in (0, +\infty)$ we get that $\exists \lim_{t \rightarrow +\infty} \left(\frac{1}{\eta}h + u \right) (t) \in \mathbb{R}$.

Furthermore, since x^* has been chosen as an arbitrary minimizer of $f + g$, we conclude that for all $x^* \in \text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}$ the limit

$$\lim_{t \rightarrow +\infty} E(t, x^*) \in \mathbb{R}$$

exists, where

$$E(t, x^*) = \frac{1}{2\eta} \|x(t) - x^*\|^2 + g(x(t)) - g(x^*) - \langle \nabla g(x^*), x(t) - x^* \rangle.$$

In what follows we use a similar technique as in [18] (see, also, [1, Section 5.2]). Since $x(\cdot)$ is bounded, it has at least one weak sequential cluster point.

We prove first that each weak sequential cluster point of $x(\cdot)$ is a minimizer of $f + g$. Let $x^* \in \text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}$ and $t_n \rightarrow +\infty$ (as $n \rightarrow +\infty$) be such that $(x(t_n))_{n \in \mathbb{N}}$ converges weakly to \bar{x} . Since $(x(t_n), \nabla g(x(t_n))) \in \text{Gr}(\nabla g)$, $\lim_{n \rightarrow +\infty} \nabla g(x(t_n)) = \nabla g(x^*)$ and $\text{Gr}(\nabla g)$ is sequentially closed in the weak-strong topology, we obtain $\nabla g(\bar{x}) = \nabla g(x^*)$.

From (27) written for $t = t_n$, $A = \partial f$ and $B = \nabla g$, by letting n converge to $+\infty$ and by using that $\text{Gr}(\partial f)$ is sequentially closed in the weak-strong topology, we obtain $-\nabla g(x^*) \in \partial f(\bar{x})$. This, combined with $\nabla g(\bar{x}) = \nabla g(x^*)$, delivers $-\nabla g(\bar{x}) \in \partial f(\bar{x})$, hence $\bar{x} \in \text{Zer}(\partial f + \nabla g) = \text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}$.

Next we show that $x(\cdot)$ has at most one weak sequential cluster point, fact which guarantees that it has exactly one weak sequential cluster point. This implies the weak convergence of the trajectory to a minimizer of $f + g$.

Let x_1^*, x_2^* be two weak sequential cluster points of $x(\cdot)$. This means that there exist $t_n \rightarrow +\infty$ (as $n \rightarrow +\infty$) and $t'_n \rightarrow +\infty$ (as $n \rightarrow +\infty$) such that $(x(t_n))_{n \in \mathbb{N}}$ converges weakly to x_1^* (as $n \rightarrow +\infty$) and $(x(t'_n))_{n \in \mathbb{N}}$ converges weakly to x_2^* (as $n \rightarrow +\infty$). Since $x_1^*, x_2^* \in \text{argmin}_{x \in \mathcal{H}} \{f(x) +$

$g(x)\}$, we have $\lim_{t \rightarrow +\infty} E(t, x_1^*) \in \mathbb{R}$ and $\lim_{t \rightarrow +\infty} E(t, x_2^*) \in \mathbb{R}$, hence $\exists \lim_{t \rightarrow +\infty} (E(t, x_1^*) - E(t, x_2^*)) \in \mathbb{R}$. We obtain

$$\exists \lim_{t \rightarrow +\infty} \left(\frac{1}{\eta} \langle x(t), x_2^* - x_1^* \rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x(t) \rangle \right) \in \mathbb{R},$$

which, when expressed by means of the sequences $(t_n)_{n \in \mathbb{N}}$ and $(t'_n)_{n \in \mathbb{N}}$, leads to

$$\frac{1}{\eta} \langle x_1^*, x_2^* - x_1^* \rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_1^* \rangle = \frac{1}{\eta} \langle x_2^*, x_2^* - x_1^* \rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_2^* \rangle.$$

This is the same with

$$\frac{1}{\eta} \|x_1^* - x_2^*\|^2 + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_2^* - x_1^* \rangle = 0$$

and by the monotonicity of ∇g we conclude that $x_1^* = x_2^*$.

(v) The proof of this statement follows in analogy to the one of the corresponding statement of Theorem 10(v) written for $A = \partial f$ and $B = \nabla g$. \blacksquare

Remark 15 When $\gamma(t) = \gamma > 0$ for every $t \geq 0$ and $\lambda(t) = 1$ for every $t \in [0, +\infty)$, the second inequality in (33) is verified if and only if $\gamma^2 > \frac{\eta}{\beta} + 1$. In other words, (A5) allows in this particular setting a more relaxed choice for the parameters γ, η and β , beyond the standard assumptions $0 < \eta \leq 2\beta$ and $\gamma^2 > 2$ considered in [9].

Remark 16 The explicit discretization of (31) with respect to the time variable t , with step size $h_n > 0$, relaxation variable $\lambda_n > 0$, damping variable $\gamma_n > 0$ and initial points $x_0 := u_0$ and $x_1 := v_0$ yields the following iterative scheme

$$\frac{x_{n+1} - 2x_n + x_{n-1}}{h_n^2} + \gamma_n \frac{x_{n+1} - x_n}{h_n} = \lambda_n \left[\text{prox}_{\eta f} \left(x_n - \eta \nabla g(x_n) \right) - x_n \right] \quad \forall n \geq 1.$$

For $h_n = 1$ this becomes

$$x_{n+1} = \left(1 - \frac{\lambda_n}{1 + \gamma_n} \right) x_n + \frac{\lambda_n}{1 + \gamma_n} \text{prox}_{\eta f} \left(x_n - \eta \nabla g(x_n) \right) + \frac{\lambda_n}{1 + \gamma_n} (x_n - x_{n-1}) \quad \forall n \geq 1,$$

which is a relaxed forward-backward algorithm for minimizing $f + g$ with inertial effects. For more on inertial-type forward-backward algorithms we refer the reader to [28].

In the following we provide a rate for the convergence for a convex and (Fréchet) differentiable function $g : \mathcal{H} \rightarrow \mathbb{R}$ with Lipschitz continuous gradient to its minimum value along the ergodic trajectory generated by

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \quad (40)$$

To this end we make the following assumption:

(A6) $\lambda : [0, +\infty) \rightarrow (0, +\infty)$ is locally absolutely continuous, $\gamma : [0, +\infty) \rightarrow (0, +\infty)$ is twice differentiable and there exists $\zeta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$0 < \zeta \leq \gamma(t)\lambda(t) - \dot{\lambda}(t), \quad \dot{\gamma}(t) \leq 0 \text{ and } 2\dot{\gamma}(t)\gamma(t) - \ddot{\gamma}(t) \leq 0. \quad (41)$$

Let us mention that the following result is in the spirit of a convergence rate statement recently given in [26, Theorem 1] for the objective function values on a sequence iteratively generated by an inertial gradient-type algorithm.

Theorem 17 Let $g : \mathcal{H} \rightarrow \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$ -Lipschitz continuous gradient for $\beta > 0$ such that $\operatorname{argmin}_{x \in \mathcal{H}} g(x) \neq \emptyset$. Let $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A6) $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (40).

Then for every minimizer x^* of g and every $T > 0$ it holds

$$0 \leq g\left(\frac{1}{T} \int_0^T x(t) dt\right) - g(x^*) \leq \frac{1}{2\zeta T} \left[\|v_0 + \gamma(0)(u_0 - x^*)\|^2 + \left(\frac{\lambda(0)}{\beta} - \dot{\gamma}(0)\right) \|u_0 - x^*\|^2 \right].$$

Proof. By using (40), the convexity of g and (A6) we get for almost every $t \in [0, +\infty)$

$$\begin{aligned} & \frac{d}{dt} \left(\frac{1}{2} \|\dot{x}(t) + \gamma(t)(x(t) - x^*)\|^2 + \lambda(t)g(x(t)) - \frac{\dot{\gamma}(t)}{2} \|x(t) - x^*\|^2 \right) \\ &= \langle \ddot{x}(t) + \dot{\gamma}(t)(x(t) - x^*) + \gamma(t)\dot{x}(t), \dot{x}(t) + \gamma(t)(x(t) - x^*) \rangle \\ & \quad - \frac{\ddot{\gamma}(t)}{2} \|x(t) - x^*\|^2 - \dot{\gamma}(t) \langle \dot{x}(t), x(t) - x^* \rangle + \dot{\lambda}(t)g(x(t)) + \lambda(t) \langle \dot{x}(t), \nabla g(x(t)) \rangle \\ &= -\gamma(t)\lambda(t) \langle \nabla g(x(t)), x(t) - x^* \rangle + \dot{\lambda}(t)g(x(t)) + \left(\dot{\gamma}(t)\gamma(t) - \frac{\ddot{\gamma}(t)}{2} \right) \|x(t) - x^*\|^2 \\ &\leq -\gamma(t)\lambda(t) \langle \nabla g(x(t)), x(t) - x^* \rangle + \dot{\lambda}(t)g(x(t)) \\ &\leq (\dot{\lambda}(t) - \gamma(t)\lambda(t))(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*) \\ &\leq -\zeta(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*). \end{aligned}$$

We obtain after integration

$$\begin{aligned} & \frac{1}{2} \|\dot{x}(T) + \gamma(T)(x(T) - x^*)\|^2 + \lambda(T)g(x(T)) - \frac{\dot{\gamma}(T)}{2} \|x(T) - x^*\|^2 \\ & - \frac{1}{2} \|\dot{x}(0) + \gamma(0)(x(0) - x^*)\|^2 - \lambda(0)g(x(0)) + \frac{\dot{\gamma}(0)}{2} \|x(0) - x^*\|^2 \\ & + \zeta \int_0^T (g(x(t)) - g(x^*)) dt \leq (\lambda(T) - \lambda(0))g(x^*). \end{aligned}$$

By neglecting the nonnegative terms in the left-hand side of the inequality above and by using that $g(x(T)) \geq g(x^*)$, it yields

$$\zeta \int_0^T (g(x(t)) - g(x^*)) dt \leq \frac{1}{2} \|v_0 + \gamma(0)(u_0 - x^*)\|^2 - \frac{\dot{\gamma}(0)}{2} \|u_0 - x^*\|^2 + \lambda(0)(g(u_0) - g(x^*)).$$

The conclusion follows by using

$$g(u_0) - g(x^*) \leq \frac{1}{2\beta} \|u_0 - x^*\|^2,$$

which is a consequence of the descent lemma (see [29, Lemma 1.2.3] and notice that $\nabla g(x^*) = 0$), and the inequality

$$g\left(\frac{1}{T} \int_0^T x(t) dt\right) - g(x^*) \leq \frac{1}{T} \int_0^T (g(x(t)) - g(x^*)) dt,$$

which holds since g is convex. ■

Remark 18 Under assumption (A6), we obtain in the above theorem (only) the convergence of the function g along the ergodic trajectory to a global minimum value. If one is interested also in the (weak) convergence of the trajectory to a minimizer of g , this follows via Theorem 6 when λ, γ are assumed to fulfill (A1) (notice that if x converges weakly to a minimizer of g , then from the Cesaro-Stolz Theorem one also obtains the weak convergence of the ergodic trajectory $T \mapsto \frac{1}{T} \int_0^T x(t)dt$ to the same minimizer).

For $a, a', \rho, \rho' \geq 0$ and $b, b' > 0$ fulfilling the inequalities $b'^2 b > \frac{1}{\beta}$ and $0 \leq \rho \leq b'$ one can prove that the functions

$$\lambda(t) = \frac{1}{ae^{-\rho t} + b} \text{ and } \gamma(t) = a'e^{-\rho' t} + b',$$

verify assumption (A1) in Theorem 6 for $0 < \theta \leq b'^2 b \beta - 1$ and assumption (A6) in Theorem 17 for $0 < \zeta \leq \frac{bb'}{(a+b)^2}$. Hence, for this choice of the relaxation and damping functions, both convergence of both the objective function g along the ergodic trajectory to its global minimum value and (weak) convergence of the trajectory to a minimizer of g are guaranteed.

Acknowledgements. The authors are thankful to the handling editor and two anonymous reviewers for comments and remarks which substantially improved the quality of the paper.

References

- [1] B. Abbas, H. Attouch, *Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, Optimization 64(10), 2223–2252, 2015
- [2] B. Abbas, H. Attouch, B.F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and its Applications 161(2), 331–360, 2014
- [3] S. Adly, H. Attouch, A. Cabot, *Finite time stabilization of nonlinear oscillators subject to dry friction*, in: P. Alart, O. Maissenueve, R.T. Rockafellar (eds.), Nonsmooth Mechanics and Analysis, Advances in Mechanics and Mathematics 12, 289–304, 2006
- [4] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization 38(4), 1102–1119, 2000
- [5] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM Journal on Optimization 14(3), 773–782, 2004
- [6] F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9(1-2), 311, 2001
- [7] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, Journal de Mathématiques Pures et Appliquées (9) 81(8), 747–779, 2002
- [8] A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differential'nye Uravneniya 30(9), 1475–1486, 1994; translation in Differential Equations 30(9), 1365–1375, 1994

- [9] H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, in: Optimization (Namur, 1998), Lecture Notes in Economics and Mathematical Systems 481, Springer, Berlin, 25–35, 2000
- [10] H. Attouch, Z. Chbani, *Fast inertial dynamics and FISTA algorithms in convex optimization. Perturbation aspects*, arXiv:1507.01367, 2015
- [11] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations 248(6), 1315–1344, 2010
- [12] H. Attouch, P.-E. Maingé, *Asymptotic behavior of second-order dissipative evolution equations combining potential with non-potential effects*, ESAIM: Control, Optimisation and Calculus of Variations 17(3), 836–857, 2011
- [13] H. Attouch, M. Marques Alves, B.F. Svaiter, *A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $O(1/n^2)$* , arXiv:1502.04286v1, to appear in Journal of Convex Analysis
- [14] H. Attouch, X. Goudou, P. Redont, *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*, Communications in Contemporary Mathematics 2(1), 1–34, 2000
- [15] H. Attouch, J. Peypoquet, P. Redont, *Fast convergence of an inertial gradient-like system with vanishing viscosity*, arXiv:1507.04782, 2015
- [16] H. Attouch, B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49(2), 574–598, 2011
- [17] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011
- [18] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and its Applications 119(2), 235–259, 2003
- [19] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010
- [20] R.I. Boţ, E.R. Csetnek, *A dynamical system associated with the fixed points set of a non-expansive operator*, Journal of Dynamics and Differential Equations, DOI: 10.1007/s10884-015-9438-x, 2015
- [21] R.I. Boţ, E.R. Csetnek, *Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems*, arXiv:1503.01871
- [22] R.I. Boţ, E.R. Csetnek, C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Applied Mathematics and Computation 256, 472–487, 2015
- [23] A. Cabot, *Stabilization of oscillators subject to dry friction: finite time convergence versus exponential decay results*, Transactions of the American Mathematical Society 360(1), 103–121, 2008
- [24] A. Cabot, H. Engler, S. Gadat, *On the long time behavior of second order differential equations with asymptotically small dissipation*, Transactions of the American Mathematical Society 361(11), 5983–6017, 2009

- [25] A. Cabot, H. Engler, S. Gadat, *Second-order differential equations with asymptotically small dissipation and piecewise flat potentials*, Proceedings of the Seventh Mississippi State-UAB Conference on Differential Equations and Computational Simulations, 33–38, Electronic Journal of Differential Equations Conference 17, 2009
- [26] E. Ghadimi, H.R. Feyzmahdavian, M. Johansson, *Global convergence of the Heavy-ball method for convex optimization*, arXiv:1412.7457
- [27] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris, 1991
- [28] A. Moudafi, M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, Journal of Computational and Applied Mathematics 155, 447–454, 2003
- [29] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, Dordrecht, 2004
- [30] N. Ogura, I. Yamada, *Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping*, Numerical Functional Analysis and Optimization 23(1-2), 113–137, 2002
- [31] J.-C. Pesquet, N. Pustelnik, *A parallel inertial proximal optimization method*, Pacific Journal of Optimization 8(2), 273–305, 2012
- [32] R.T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific Journal of Mathematics 33(1), 209–216, 1970
- [33] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14(5), 877–898, 1976
- [34] S. Simons, *From Hahn-Banach to Monotonicity*, Springer, Berlin, 2008
- [35] W. Su, S. Boyd, E.J. Candes, *A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights*, arXiv:1503.01243, 2015