Approaching monotone inclusion problems via second order dynamical systems with linear and anisotropic damping

Radu Ioan Boţ* and Ernö Robert Csetnek**

University of Vienna
Faculty of Mathematics
Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria
**E-mail: radu.bot@univie.ac.at*
***E-mail: ernoe.robert.csetnek@univie.ac.at*

We investigate second order dynamical systems of the from $\ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)B(x(t)) = 0$, where $\Gamma : \mathcal{H} \to \mathcal{H}$ is an elliptic bounded self-adjoint linear operator defined on a real Hilbert space $\mathcal{H}$, $B : \mathcal{H} \to \mathcal{H}$ is a cocoercive operator and $\lambda : [0, +\infty) \to (0, +\infty)$ is a relaxation function depending on time. We prove via Lyapunov analysis that the generated trajectories weakly converge to a zero of the operator B. This opens the gate towards approaching through a second order dynamical system the problem of determining zeros of the sum of a maximally monotone operator and a cocoercive one, which captures as particular case the minimization of the sum of a nonsmooth convex function with a smooth convex one. Finally, when B is the gradient of a smooth convex function, we prove a rate of $\mathcal{O}(1/t)$ for the convergence of the function values along the ergodic trajectory to its minimum value.

Keywords: Dynamical systems; Lyapunov analysis; Monotone inclusions; Convex optimization problems; Continuous forward-backward method.

1. Introduction and preliminaries

Let $\mathcal{H}$ a real Hilbert space endowed with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The problem of approaching the minimization of a potential function $f : \mathcal{H} \to \mathbb{R}$, supposed to be convex and differentiable, has been considered by several authors see [3, 7, 8, 11]). These investigations addressed either the convergence of the generated trajectories to a critical point of f or the convergence of the function along the trajectories to its global minimum value. We recall in this context the heavy ball with friction dynamical system

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

which is a nonlinear oscillator with constant damping parameter $\gamma > 0$. When $\mathcal{H} = \mathbb{R}^2$, this system describes 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 [11]). The time discretization of (1) leads 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, [3–5, 17, 19, 21]).

The minimization of f over a nonempty, convex and closed set $C \subseteq \mathcal{H}$ has been approached in the same spirit in [7, 8], by considering the gradient-projection second order dynamical system

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

where $P_C : \mathcal{H} \to C$ denotes the projection onto the set C and $\eta > 0$. These investigations have been further expanded in [8] to more general systems of the form

$$\ddot{x} + \gamma\dot{x} + x - Tx = 0, \tag{3}$$

where $T : \mathcal{H} \to \mathcal{H}$ is a nonexpansive operator. When $\gamma^2 > 2$ the trajectory of (3) has been shown to converge weakly to an element in the fixed points set of T, provided it is nonempty.

The dynamical system which we investigate in this paper reads

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)B(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{4}$$

where $u_0, v_0 \in \mathcal{H}$, $\Gamma : \mathcal{H} \to \mathcal{H}$ is a bounded self-adjoint linear operator, which is γ-elliptic, that is, $\langle \Gamma u, u\rangle \geq \gamma\|u\|^2$ for all $u \in \mathcal{H}$ (with $\gamma > 0$) and $B : \mathcal{H} \to \mathcal{H}$ is a β-*cocoercive* operator for $\beta > 0$, that is $\beta\|Bx - By\|^2 \leq \langle x - y, Bx - By\rangle$ for all $x, y \in \mathcal{H}$. This obviously implies that B is $1/\beta$-Lipschitz continuous, that is $\|Bx-By\| \leq \frac{1}{\beta}\|Bx-By\|$ for all $x, y \in \mathcal{H}$. The elliptic operator Γ induces an anisotropic damping, a similar construction being already used in [3] in the context of minimizing a convex and smooth function.

The assumption which we make on the function λ is

(λ1) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous, monotonically increasing on $[0, +\infty)$ with $\lim_{t\to+\infty} \lambda(t) < \beta\gamma^2$.

According to the above assumption, $\dot{\lambda}(t)$ exists for almost every $t \geq 0$ and $\dot{\lambda}$ is Lebesgue integrable on each interval $[0, b]$ for $0 < b < +\infty$. Since λ is monotonically increasing, as λ is assumed to take only positive values, (λ1) yields the existence of a lower bound $\underline{\lambda}$ and a positive real number θ

such that for every $t \in [0, +\infty)$ one has

$$0 < \underline{\lambda} \le \lambda(t) \le \frac{\beta\gamma^2}{1+\theta}. \tag{5}$$

Thus, the existence and uniqueness of global solutions in $C^2([0, +\infty); \mathcal{H})$ for (4) follows from the global version of the Picard-Lindelöf Theorem.

We begin our investigations by showing that under mild assumptions on the relaxation function λ the trajectory $x(t)$ converges weakly as $t \to +\infty$ to a zero of the operator B, provided it has a nonempty set of zeros. 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 (22). Dynamical systems of implicit type have been already considered in the literature in [1, 2, 9, 12, 14–16]. We further specialize these investigations to the minimization of the sum of a nonsmooth convex function with a smooth convex function one. This allows us to recover and improve results given in [7, 8] in the context of studying the dynamical system (2). Whenever B is the gradient of a smooth convex function, we show that the function values of the latter converge along the ergodic trajectories generated by (4) to its minimum value with a rate of convergence of $\mathcal{O}(1/t)$.

2. Convergence of the trajectories

In this section we address the convergence properties of the trajectories generated by the dynamical system (4) and use to this end as essential ingredient the continuous version of the Opial Lemma that we state below (see, for example, [2, Lemma 5.3], [1, Lemma 1.10]).

Lemma 2.1. *Let $S \subseteq \mathcal{H}$ be a nonempty set and $x : [0, +\infty) \to \mathcal{H}$ a given map. Assume that*

(i) for every $x^* \in S$, $\lim_{t\to+\infty} \|x(t) - x^*\|$ 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 $t \to +\infty$.

We come now to the main result of this section.

Theorem 2.1. *Let $B : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator for $\beta > 0$ such that $\operatorname{zer} B := \{u \in \mathcal{H} : Bu = 0\} \neq \emptyset$, $\Gamma : \mathcal{H} \to \mathcal{H}$ be a γ-elliptic operator, $\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling $(\lambda 1)$ and $u_0, v_0 \in \mathcal{H}$. Let*

$x \in C^2([0,+\infty);\mathcal{H})$ be the unique global solution of (4). 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) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} B(x(t) = 0$;

(iii) $x(t)$ converges weakly to an element in $\operatorname{zer} B$ as $t \to +\infty$.

Proof. (i) Take an arbitrary $x^* \in \operatorname{zer} B$ 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 (4), we get for every $t \in [0,+\infty)$

$$\begin{aligned}&\ddot{h}(t) + \gamma\dot{h}(t) + \\ &\lambda(t)\left\langle x(t) - x^*, B(x(t))\right\rangle + \left\langle x(t) - x^*, \Gamma(\dot{x}(t)) - \gamma\dot{x}(t)\right\rangle = \|\dot{x}(t)\|^2.\end{aligned} \tag{6}$$

Now we introduce the function $p : [0,+\infty) \to \mathbb{R}$,

$$p(t) = \frac{1}{2}\left\langle (\Gamma - \gamma \operatorname{Id})(x(t) - x^*), x(t) - x^*\right\rangle, \tag{7}$$

where Id denotes the identity on $\mathcal{H}$. Since $\langle(\Gamma - \gamma \operatorname{Id})u, u\rangle \geq 0$ for all $u \in \mathcal{H}$, it holds

$$p(t) \geq 0 \text{ for all } t \geq 0. \tag{8}$$

Moreover, $\dot{p}(t) = \langle(\Gamma - \gamma \operatorname{Id})(\dot{x}(t)), x(t) - x^*\rangle$, which combined with (6), the cocoercivity of B and the fact that $Bx^* = 0$ yields for every $t \in [0,+\infty)$

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

Taking into account (4) one obtains for every $t \in [0,+\infty)$

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

hence

$$\begin{aligned}&\ddot{h}(t) + \gamma\dot{h}(t) + \\ &\frac{\beta}{\lambda(t)}\|\ddot{x}(t)\|^2 + \frac{2\beta}{\lambda(t)}\left\langle \ddot{x}(t), \Gamma(\dot{x}(t))\right\rangle + \frac{\beta}{\lambda(t)}\|\Gamma(\dot{x}(t))\|^2 + \dot{p}(t) \leq \|\dot{x}(t)\|^2.\end{aligned} \tag{9}$$

We have

$$\gamma\|u\| \leq \|\Gamma u\| \text{ for all } u \in \mathcal{H}, \tag{10}$$

which combined with (9) yields for every $t \in [0,+\infty)$

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

By taking into account that for almost every $t \in [0, +\infty)$

$$\frac{1}{\lambda(t)} \frac{d}{dt} \big(\langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \big) \geq \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \gamma \frac{\dot{\lambda}(t)}{\lambda^2(t)} \|\dot{x}(t)\|^2, \tag{11}$$

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

$$\beta \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \left(\frac{\beta\gamma^2}{\lambda(t)} + \beta\gamma \frac{\dot{\lambda}(t)}{\lambda^2(t)} - 1 \right) ||\dot{x}(t)||^2 +$$

$$\ddot{h}(t) + \gamma \dot{h}(t) + \dot{p}(t) + \frac{\beta}{\lambda(t)} ||\ddot{x}(t)||^2 \leq 0. \tag{12}$$

By using now assumption $(\lambda 1)$ and (5) we obtain that the following inequality holds for almost every $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma \dot{h}(t) + \dot{p}(t) + \beta \frac{d}{dt} \left(\frac{1}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \right) + \theta ||\dot{x}(t)||^2 + \frac{1+\theta}{\gamma^2} ||\ddot{x}(t)||^2 \leq 0. \tag{13}$$

This implies that the function $t \mapsto \dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle$ is monotonically decreasing. Hence there exists a real number M such that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle \leq M, \tag{14}$$

which yields, together with (8) and $(\lambda 1)$, that for every $t \in [0, +\infty)$ $\dot{h}(t) + \gamma h(t) \leq M$. This implies that

$$h \text{ is bounded} \tag{15}$$

and, consequently,

$$\text{the trajectory } x \text{ is bounded.} \tag{16}$$

On the other hand, from (14), by taking into account (8), $(\lambda 1)$ and (5), it follows that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \frac{1+\theta}{\gamma} \|\dot{x}(t)\|^2 \leq M,$$

hence $\langle x(t) - x^*, \dot{x}(t) \rangle + \frac{1+\theta}{\gamma} \|\dot{x}(t)\|^2 \leq M$. This inequality, in combination with (16), yields

$$\dot{x} \text{ is bounded,} \tag{17}$$

which further implies that

$$\dot{h} \text{ is bounded.} \tag{18}$$

Integrating the inequality (13) we obtain that there exists a real number $N \in \mathbb{R}$ such that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle +$$
$$\theta \int_0^t ||\dot{x}(s)||^2 ds + \frac{1+\theta}{\gamma^2} \int_0^t ||\ddot{x}(s)||^2 ds \leq N.$$

From here, via (18) and (8), we conclude that $\dot{x}(\cdot), \ddot{x}(\cdot) \in L^2([0, +\infty); \mathcal{H})$. Finally, from (4) and ($\lambda 1$) 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 [2, Lemma 5.2] together with (i) lead to $\lim_{t \to +\infty} \dot{x}(t) = 0$.

Further, 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 [2, Lemma 5.2] and (i) we get $\lim_{t \to +\infty} B(x(t)) = 0$, while the fact that $\lim_{t \to +\infty} \ddot{x}(t) = 0$ follows from (4), and ($\lambda 1$).

(iii) As seen in the proof of part (i), the function $t \mapsto \dot{h}(t) + \gamma h(t) + p(t) + \frac{\beta}{\lambda(t)} \langle \dot{x}(t), \Gamma(\dot{x}(t)) \rangle$ is monotonically decreasing, thus from (i), (ii), (8), (Γ) and ($\lambda 1$) we deduce that $\lim_{t \to +\infty}(\gamma h(t) + p(t))$ exists and it is a real number.

In the following we consider the scalar product defined by $\langle\langle x, y \rangle\rangle = \frac{1}{\gamma}\langle \Gamma x, y \rangle$ and the corresponding induced norm $|||x|||^2 = \frac{1}{\gamma}\langle \Gamma x, x \rangle$. Taking into account the definition of p, we have that $\lim_{t \to +\infty} \frac{1}{2}|||x(t) - x^*|||^2$ exists and it is a real number.

Let $\overline{x}$ be a weak sequential cluster point of x, that is, there exists a sequence $t_n \to +\infty$ (as $n \to +\infty$) such that $(x(t_n))_{n \in \mathbb{N}}$ converges weakly to $\overline{x}$. Since B is a maximally monotone operator (see for instance [13, 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 $\lim_{n \to +\infty} B(x(t_n)) = 0$, we conclude that $B\overline{x} = 0$, hence $\overline{x} \in \operatorname{zer} B$.

The conclusion follows by applying the Opial Lemma in the Hilbert space $(\mathcal{H}, \langle\langle \cdot, \cdot \rangle\rangle)$, by noticing that a sequence $(x_n)_{n \geq 0}$ converges weakly to $\overline{x} \in \mathcal{H}$ in $(\mathcal{H}, \langle\langle \cdot, \cdot \rangle\rangle)$ if and only if $(x_n)_{n \geq 0}$ converges weakly to $\overline{x}$ in $(\mathcal{H}, \langle \cdot, \cdot \rangle)$.

□

A standard instance of a cocoercive operator defined on a real Hilbert spaces is the one that can be represented as $B = \operatorname{Id} - T$, where $T : \mathcal{H} \to \mathcal{H}$ is a *nonexpansive operator*, that is, a 1-Lipschitz continuous operator. It is easy to see that in this case B is $1/2$-cocoercive. For this particular choice of the operator B, the dynamical system (4) becomes

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\big(x(t) - T(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{19}$$

while assumption $(\lambda 1)$ reads

$(\lambda 2)$ $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous, monotonically increasing on $[0, +\infty)$ with $\lim_{t\to+\infty} \lambda(t) < \frac{\gamma^2}{2}$.

Theorem 2.1 gives rise to the following result.

Corollary 2.1. *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive operator such that* $\operatorname{Fix} T = \{u \in \mathcal{H} : Tu = u\} \neq \emptyset$*, $\Gamma : \mathcal{H} \to \mathcal{H}$ be a γ-elliptic operator,* $\lambda : [0, +\infty) \to (0, +\infty)$ *be a function fulfilling* $(\lambda 2)$ *and $u_0, v_0 \in \mathcal{H}$. Let $x \in$* $C^2([0, +\infty); \mathcal{H})$ *be the unique global solution of* (19)*. Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, (\operatorname{Id} - T)x \in L^2([0, +\infty); \mathcal{H})$*;*

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty} (\operatorname{Id} - T)(x(t)) = 0$*;*

(iii) $x(t)$ converges weakly to a point in $\operatorname{Fix} T$ *as $t \to +\infty$.*

Remark 2.1. Taking $\Gamma = \gamma \operatorname{Id}$ for $\gamma > 0$ and $\lambda(t) = 1$ for all $t \in [0, +\infty)$, the dynamical system (19) turns out to be

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

The convergence of the trajectories generated by (20) has been studied in [8, Theorem 3.2] under the condition $\gamma^2 > 2$. In this case $(\lambda 2)$ is obviously fulfilled. However, different to [8], we allow in Corollary 2.1 an anisotropic damping through the use of the elliptic operator Γ and also a variable relaxation function λ depending on time.

Finally, we discuss a consequence of the above corollary applied to second order dynamical systems governed by averaged operators. The operator $R : \mathcal{H} \to \mathcal{H}$ is said to be *α-averaged* for $\alpha \in (0, 1)$, if there exists a nonexpansive operator $T : \mathcal{H} \to \mathcal{H}$ such that $R = (1 - \alpha) \operatorname{Id} + \alpha T$. An important representative of this class are the *firmly nonexpansive* operators which are obtained for $\alpha = \frac{1}{2}$.

We consider the dynamical system

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

and make the assumption

($\lambda 3$) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous, monotonically increasing on $[0, +\infty)$ with $\lim_{t\to+\infty} \lambda(t) < \frac{\gamma^2}{2\alpha}$.

Corollary 2.2. *Let* $R : \mathcal{H} \to \mathcal{H}$ *be an* α*-averaged operator for* $\alpha \in (0, 1)$ *such that* $\operatorname{Fix} R \neq \emptyset$*,* $\Gamma : \mathcal{H} \to \mathcal{H}$ *be a* γ*-elliptic operator,* $\lambda : [0, +\infty) \to (0, +\infty)$ *be a function fulfilling* ($\lambda 3$) *and* $u_0, v_0 \in \mathcal{H}$*. Let* $x \in C^2([0, +\infty); \mathcal{H})$ *be the unique global solution of* (21)*. Then the following statements are true:*

(i) the trajectory x *is bounded and* $\dot{x}, \ddot{x}, (\operatorname{Id} - R)x \in L^2([0, +\infty); \mathcal{H})$*;*

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = \lim_{t\to+\infty}(\operatorname{Id} - R)(x(t)) = 0$*;*

(iii) $x(t)$ *converges weakly to a point in* $\operatorname{Fix} R$ *as* $t \to +\infty$*.*

Proof. Since R is α-averaged, there exists a nonexpansive operator $T : \mathcal{H} \to \mathcal{H}$ such that $R = (1-\alpha)\operatorname{Id} + \alpha T$. Corollary 2.1 leads to the conclusion, by taking into account also $\operatorname{Fix} R = \operatorname{Fix} T$. □

3. Forward-backward second order dynamical systems

In this section we address the structured monotone inclusion problem

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

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ is a β-cocoercive operator for $\beta > 0$ via a second-order forward-backward dynamical system with anisotropic damping and variable relaxation parameter.

For readers convenience we recall at the beginning some standard notions and results in monotone operator theory which will be used in the following (see also [13]). For an arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\operatorname{Gr} A = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its graph. We use also the notation $\operatorname{zer} A = \{x \in \mathcal{H} : 0 \in Ax\}$ for the set of zeros of A. We say that A is monotone, if $\langle x - y, u - v\rangle \geq 0$ for all $(x, u), (y, v) \in \operatorname{Gr} A$. 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 $J_A = (\operatorname{Id} + A)^{-1}$. If A is maximally monotone, then $J_A : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (see [13, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [13, Proposition 23.2]) $p \in J_{\gamma A}x$ if and only if $(p, \gamma^{-1}(x-p)) \in \operatorname{Gr} A$.

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

For $\eta \in (0, 2\beta)$, we approach the monotone inclusion problem to solve via the dynamical system

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\Big[x(t) - J_{\eta A}\Big(x(t) - \eta B(x(t))\Big)\Big] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{22}$$

and make the following assumption for the relaxation function

($\lambda 4$) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous, monotonically increasing on $[0, +\infty)$ with $\lim_{t\to+\infty} \lambda(t) < \frac{(4\beta-\eta)\gamma^2}{4\beta}$.

Theorem 3.1. *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ be β-cocoercive operator for $\beta > 0$ such that $\operatorname{zer}(A+B) \neq \emptyset$. Let $\eta \in (0, 2\beta)$, $\Gamma : \mathcal{H} \to \mathcal{H}$ be a γ-elliptic operator, $\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling ($\lambda 4$), $u_0, v_0 \in \mathcal{H}$ and $x \in C^2([0, +\infty); \mathcal{H})$ be the unique global solution of (22). Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, \big(\operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)\big)x \in L^2([0, +\infty); \mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = 0$;

(iii) $x(t)$ converges weakly to a point in $\operatorname{zer}(A+B)$ as $t \to +\infty$;

(iv) if $x^* \in \operatorname{zer}(A+B)$, then $B(x(\cdot)) - Bx^* \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t\to+\infty} B(x(t)) = Bx^*$ and B is constant on $\operatorname{zer}(A+B)$;

(v) if A or B is uniformly monotone, then $x(t)$ converges strongly to the unique point in $\operatorname{zer}(A+B)$ as $t \to +\infty$.

Proof. (i)-(iii) It is immediate that the dynamical system (22) can be written in the form

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\big(x(t) - R(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{23}$$

where $R = J_{\eta A} \circ (\operatorname{Id} - \eta B)$. According to [13, Corollary 23.8 and Remark 4.24(iii)], $J_{\eta A}$ is 1/2-cocoercive. Moreover, by [13, Proposition 4.33], $\operatorname{Id} - \eta B$ is $\eta/(2\beta)$-averaged. Combining this with [20, Theorem 3(b)], we derive that R is $\frac{2\beta}{4\beta-\eta}$-averaged. The statements (i)-(iii) follow now from Corollary 2.2 by noticing that $\operatorname{Fix} R = \operatorname{zer}(A+B)$ (see [13, Proposition 25.1(iv)]).

(iv) The fact that B is constant on $\operatorname{zer}(A+B)$ follows from the cocoercivity of B and the monotonicity of A.

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

$$-B(x(t)) - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \in A\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t)\right), \tag{24}$$

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

$$\begin{aligned} 0 \leq & \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*, -B(x(t)) + Bx^* \right\rangle \\ & + \left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t)) + x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \right\rangle. \end{aligned} \tag{25}$$

After using the cocoercivity of B we obtain for every $t \in [0,+\infty)$

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

For evaluating the last term of the above inequality we consider the functions $h : [0,+\infty) \to \mathbb{R}, h(t) = \frac{1}{2}\|x(t) - x^*\|^2$ and $p : [0,+\infty) \to \mathbb{R}$, $p(t) = \frac{1}{2}\langle(\Gamma - \gamma\operatorname{Id})(x(t) - x^*), x(t) - x^*\rangle$, already used in the proof of Theorem 2.1. For every $t \in [0,+\infty)$ we have $\langle x(t) - x^*, \ddot{x}(t)\rangle = \ddot{h}(t) - \|\dot{x}(t)\|^2$ and $\dot{p}(t) = \langle x(t) - x^*, \Gamma(\dot{x}(t))\rangle - \gamma\langle x(t) - x^*, \dot{x}(t)\rangle = \langle x(t) - x^*, \Gamma(\dot{x}(t))\rangle - \gamma\dot{h}(t)$, hence

$$\begin{aligned} & \left\langle x(t) - x^*, -\frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{1}{\eta\lambda(t)}\Gamma(\dot{x}(t)) \right\rangle = \\ & -\frac{1}{\eta\lambda(t)}\left(\ddot{h}(t) + \gamma\dot{h}(t) + \dot{p}(t) - \|\dot{x}(t)\|^2\right). \end{aligned} \tag{26}$$

From here, for every $t \in [0,+\infty)$ it holds

$$\begin{aligned} \frac{\beta}{2}\|B(x(t)) - Bx^*\|^2 \leq & \frac{1}{2\beta}\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{1}{\lambda(t)}\Gamma(\dot{x}(t))\right\|^2 \\ & - \frac{1}{\eta\lambda(t)}\left(\ddot{h}(t) + \gamma\dot{h}(t) + \dot{p}(t) - \|\dot{x}(t)\|^2\right). \end{aligned} \tag{27}$$

By taking into account ($\lambda 4$) we obtain a lower bound $\underline{\lambda}$ such that for every $t \in [0, +\infty)$ one has $0 < \underline{\lambda} \leq \lambda(t)$. By multiplying (27) with $\lambda(t)$ we obtain for every $t \in [0, +\infty)$ that

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

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

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

As $\dot{x}, \ddot{x} \in L^2([0, +\infty); \mathcal{H})$, $h(T) \geq 0, p(T) \geq 0$ for every $T \in [0, +\infty)$ and $\lim_{T \to +\infty} \dot{h}(T) = 0$, it follows that $B(x(\cdot)) - Bx^* \in L^2([0, +\infty); \mathcal{H})$.

Further, we have

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

hence, in light of [2, Lemma 5.2], it follows that $\lim_{t \to +\infty} B(x(t)) = Bx^*$.

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

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

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

which combined with the inequality $\langle x(t)-x^*, B(x(t))-Bx^*\rangle \geq 0$ yields

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

As λ is 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 $\lim_{t\rightarrow+\infty}\phi_A\left(\left\|\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{1}{\lambda(t)}\Gamma(\dot{x}(t))+x(t)-x^*\right\|\right)=0$ and the properties of the function ϕ_A allow to conclude that $\frac{1}{\lambda(t)}\ddot{x}(t)+\frac{1}{\lambda(t)}\Gamma(\dot{x}(t))+x(t)-x^*$ converges strongly to 0 as $t\rightarrow+\infty$. By using again the boundedness of λ and (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 $\langle x(t)-x^*, B(x(t))-Bx^*\rangle \geq \phi_B(\|x(t)-x^*\|)\ \forall t\in[0,+\infty)$ converge to $+\infty$ and by using that x is bounded and $\lim_{t\rightarrow+\infty}(B(x(t)-Bx^*)=0$. $\square$

Remark 3.1. The statements in Theorem 3.1 remain valid also for $\eta:=2\beta$. Under this assumption, the cocoercivity of B implies that $\mathrm{Id}-\eta B$ is nonexpansive, hence $R=J_{\eta A}\circ(\mathrm{Id}-\eta B)$ used in the proof is nonexpansive, too, and so the statements in (i)-(iii) follow from Corollary 2.1. The statements (iv) and (v) can be proved in the similar way for $\eta=2\beta$, too.

We close this section by approaching from the perspective of second order dynamical systems 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 $1/\beta$-Lipschitz continuous gradient for $\beta>0$.

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 $\partial f(x)=\{u\in\mathcal{H}: f(y)\geq f(x)+\langle u,y-x\rangle\ \forall y\in\mathcal{H}\}$. When seen as a set-valued mapping, it is a maximally monotone operator (see [22]) and its resolvent is given by $J_{\eta\partial f}=\mathrm{prox}_{\eta f}$ (see [13]), where $\mathrm{prox}_{\eta f}:\mathcal{H}\rightarrow\mathcal{H}$,

$$\mathrm{prox}_{\eta f}(x)=\operatorname*{argmin}_{y\in\mathcal{H}}\left\{f(y)+\frac{1}{2\eta}\|y-x\|^2\right\}, \tag{28}$$

denotes the proximal point operator of f and $\eta > 0$. According to [13, Definition 10.5], f is said to be uniformly convex with modulus function $\phi : [0,+\infty) \to [0,+\infty]$, 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 $x, y \in \operatorname{dom} f := \{x \in \mathcal{H} : f(x) < +\infty\}$. If this inequality holds for $\phi = (\nu/2)|\cdot|^2$ for $\nu > 0$, then f is said to be ν-strongly convex.

The second order dynamical system we consider in order to approach the minimizers of $f + g$ reads

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\left[x(t) - \operatorname{prox}_{\eta f}\Big(x(t) - \eta\nabla g(x(t))\Big)\right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{29}$$

Corollary 3.1. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\} \neq \emptyset$. Let $\eta \in (0, 2\beta]$, $\Gamma : \mathcal{H} \to \mathcal{H}$ be a γ-elliptic operator, $\lambda : [0,+\infty) \to (0,+\infty)$ be a function fulfilling ($\lambda 4$), $u_0, v_0 \in \mathcal{H}$ and $x \in C^2([0,+\infty); \mathcal{H})$ be the unique global solution of (29). Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, \big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0,+\infty); \mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = 0$;

(iii) $x(t)$ converges weakly to a minimizer of $f + g$ as $t \to +\infty$;

(iv) if x^* is a minimizer of $f + g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0,+\infty); \mathcal{H})$, $\lim_{t\to+\infty} \nabla g(x(t)) = \nabla g(x^*)$ 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 $t \to +\infty$.

Proof. The statements are direct consequences of the corresponding ones from Theorem 3.1 (see also Remark 3.1), by choosing $A := \partial f$ and $B := \nabla g$, by taking into account that $\operatorname{zer}(\partial f + \nabla g) = \operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\}$ and by making use of the Baillon-Haddad Theorem, which says that ∇g is $1/\beta$-Lipschitz if and only if ∇g is β-cocoercive (see [13, Corollary 18.16]). 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 [13, Example 22.3(iii)]). □

Remark 3.2. Let us consider again the setting in Remark 2.1, namely, when $\Gamma = \gamma \operatorname{Id}$ for $\gamma > 0$ and $\lambda(t) = 1$ for every $t \in [0,+\infty)$. 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 (29) attached in this setting to the minimization of g over C becomes

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) - P_C\big(x(t) - \eta\nabla g(x(t))\big) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{30}$$

where P_C denotes the projection onto the set C.

In [8, Theorem 3.1] it has been shown that whenever $\gamma^2 > 2$ and $0 < \eta \leq 2\beta$ the trajectories of (30) asymptotically converge to o a minimizer of g over C, provided this exists. In this setting, ($\lambda 4$) trivially holds. Moreover, in order to verify ($\lambda 4$) when $\lambda(t) = 1$ for every $t \in [0, +\infty)$, one needs to equivalently assume that $\gamma^2 > \frac{4\beta}{4\beta - \eta}$. As $\frac{4\beta}{4\beta-\eta} \leq 2$, this provides a slight improvement over [8, Theorem 3.1] in what concerns the choice of γ. We refer the reader also to [7] for an analysis of the convergence rates of trajectories of the dynamical system (30) 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 3.1, Remark 3.1 and Corollary 3.1). This allowed us to guarantee for the generated trajectories the existence of the limit $\lim_{t\to+\infty} \|x(t) - x^*\|^2$, where x^* denotes a solution of the problem under investigation. However, when dealing with convex optimization problems, one can go also beyond this classical restriction concerning the choice of the step size (see also [1, Section 4.2]). This is proved in the following corollary, which is valid under the assumption

($\lambda 5$) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous, monotonically increasing on $[0, +\infty)$ and there exist $a, \theta, \theta' > 0$ such that for $t \geq 0$

$$\frac{1}{\beta}\left(\theta' + \frac{a}{2}\|\Gamma - \gamma \operatorname{Id}\|\right) \leq \lambda(t) \leq \frac{\gamma^2}{\eta\theta + \frac{\eta}{\beta} + \frac{\eta}{2a}\|\Gamma - \gamma \operatorname{Id}\| + 1}.$$

Corollary 3.2. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that* $\operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\} \neq \emptyset$. *Let be $\eta > 0$, $\Gamma : \mathcal{H} \to \mathcal{H}$ be a γ-elliptic operator, $\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling ($\lambda 5$), $u_0, v_0 \in \mathcal{H}$ and $x \in C^2([0, +\infty); \mathcal{H})$ be the unique global solution of (29). Then the following statements are true:*

(i) the trajectory x is bounded and $\dot{x}, \ddot{x}, \big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0, +\infty); \mathcal{H})$;

(ii) $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \ddot{x}(t) = 0$;

(iii) $x(t)$ *converges weakly to a minimizer of* $f+g$ *as* $t \to +\infty$;

(iv) if x^* *is a minimizer of* $f+g$, *then* $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0,+\infty);\mathcal{H})$, $\lim_{t\to+\infty} \nabla g(x(t)) = \nabla g(x^*)$ *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* $t \to +\infty$.

Proof. Consider an arbitrary element $x^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\} = \operatorname{zer}(\partial f + \nabla g)$. Consider the function $q : [0,+\infty) \to \mathbb{R}$, $q(t) = g(x(t)) - g(x^*) - \langle \nabla g(x^*), x(t) - x^* \rangle$. Using similar techniques as in Theorem 3.1 it follows that the function

$$t \mapsto \frac{d}{dt}\left(\frac{1}{\eta}h + q\right)(t) + \gamma\left(\frac{1}{\eta}h + q\right)(t) + \frac{1}{\eta}p(t) + \frac{1}{\eta}\left(\frac{1}{\lambda(t)}\langle \dot{x}(t), \Gamma(\dot{x}(t))\rangle\right) \tag{31}$$

is monotonically decreasing. Arguing as in the proof of Theorem 2.1, it follows that $\frac{1}{\eta}h + q$, h, q, $x, \dot{x}, \dot{h}, \dot{q}$ are bounded, $\dot{x}, \ddot{x}$ and $\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0,+\infty);\mathcal{H})$ and $\lim_{t\to+\infty} \dot{x}(t) = 0$. Since $\frac{d}{dt}\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)x \in L^2([0,+\infty);\mathcal{H})$, we derive from [2, Lemma 5.2] that $\lim_{t\to+\infty}\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)(x(t)) = 0$. As $\ddot{x}(t) = -\Gamma(\dot{x}(t)) - \lambda(t)\big(\operatorname{Id} - \operatorname{prox}_{\eta f} \circ (\operatorname{Id} - \eta\nabla g)\big)(x(t))$ for every $t \in [0,+\infty)$, we obtain that $\lim_{t\to+\infty} \ddot{x}(t) = 0$. Since $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0,+\infty);\mathcal{H})$, by applying again [2, Lemma 5.2], it yields $\lim_{t\to+\infty} \nabla g(x(t)) = \nabla g(x^*)$. In this way the statements (i), (ii) and (iv) are shown.

(iii) Since the function in (31) is monotonically decreasing, from (i), (ii) and (iv) it follows that the limit $\lim_{t\to+\infty}\left(\gamma\left(\frac{1}{\eta}h + q\right)(t) + \frac{1}{\eta}p(t)\right)$ exists and it is a real number. Consider again the renorming of the space already used in the proof of Theorem 2.1(iii).

As x^* has been chosen as an arbitrary minimizer of $f+g$ and taking into account the definition of p and the new norm, we conclude that for all $x^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$ the limit $\lim_{t\to+\infty} E(t,x^*) \in \mathbb{R}$, 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 [14] (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 $x^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x)+g(x)\}$ and $t_n \to +\infty$

(as $n \to +\infty$) be such that $(x(t_n))_{n\in\mathbb{N}}$ converges weakly to $\overline{x}$. Since $(x(t_n), \nabla g(x(t_n))) \in \operatorname{Gr}(\nabla g)$, $\lim_{n\to+\infty} \nabla g(x(t_n)) = \nabla g(x^*)$ and $\operatorname{Gr}(\nabla g)$ is sequentially closed in the weak-strong topology, we obtain $\nabla g(\overline{x}) = \nabla g(x^*)$.

From (24) written for $t = t_n$, $A = \partial f$ and $B = \nabla g$, by letting n converge to $+\infty$ and by using that $\operatorname{Gr}(\partial f)$ is sequentially closed in the weak-strong topology, we obtain $-\nabla g(x^*) \in \partial f(\overline{x})$. This, combined with $\nabla g(\overline{x}) = \nabla g(x^*)$ delivers $-\nabla g(\overline{x}) \in \partial f(\overline{x})$, hence $\overline{x} \in \operatorname{zer}(\partial f + \nabla g) = \operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\}$.

Next we show that $x(\cdot)$ has at most one weak sequential cluster point, which will actually guarantee that it has exactly one weak sequential cluster point. This will imply 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 $t_n \to +\infty$ (as $n \to +\infty$) and $t_n' \to +\infty$ (as $n \to +\infty$) such that $(x(t_n))_{n\in\mathbb{N}}$ converges weakly to x_1^* (as $n \to +\infty$) and $(x(t_n'))_{n\in\mathbb{N}}$ converges weakly to x_2^* (as $n \to +\infty$). Since $x_1^*, x_2^* \in \operatorname{argmin}_{x\in\mathcal{H}}\{f(x) + g(x)\}$, we have $\lim_{t\to+\infty} E(t, x_1^*) \in \mathbb{R}$ and $\lim_{t\to+\infty} E(t, x_2^*) \in \mathbb{R}$, hence $\exists \lim_{t\to+\infty}(E(t, x_1^*) - E(t, x_2^*)) \in \mathbb{R}$. We obtain

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

which, when expressed by means of the sequences $(t_n)_{n\in\mathbb{N}}$ and $(t_n')_{n\in\mathbb{N}}$, leads to $\frac{1}{\eta}\langle\langle x_1^*, x_2^* - x_1^*\rangle\rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_1^*\rangle = \frac{1}{\eta}\langle\langle x_2^*, x_2^* - x_1^*\rangle\rangle + \langle \nabla g(x_2^*) - \nabla g(x_1^*), x_2^*\rangle$. 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) This follows in analogy to the proof of the corresponding statement of Theorem 3.1(v) written for $A = \partial f$ and $B = \nabla g$. □

Remark 3.3. When $\Gamma = \gamma \operatorname{Id}$ for $\gamma > 0$, in order to verify the left-hand side of the second statement in assumption ($\lambda 5$) one can take $\theta' := \beta \inf_{t\geq 0} \lambda(t)$. Thus, the bounds in ($\lambda 5$) amounts in this case to the existence of $\theta > 0$ such that $\lambda(t) \leq \frac{\gamma^2}{\eta\theta + \frac{\eta}{\beta} + 1}$. Whenever $\lambda(t) = 1$ for every $t \in [0, +\infty)$, this is verified if and only if $\gamma^2 > \frac{\eta}{\beta} + 1$. In other words, ($\lambda 5$) permits 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 [8].

4. Convergence rate for the function values

In this section we furnish a rate for the convergence of a convex and (Fréchet) differentiable function with Lipschitz continuous gradient $g : \mathcal{H} \to \mathbb{R}$ along the ergodic trajectory generated by

$$\begin{cases} \ddot{x}(t) + \Gamma(\dot{x}(t)) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0 \end{cases} \tag{32}$$

to the minimum value of g. To this end we make the following assumption

($\lambda 6$) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous and there exists $\zeta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$0 < \zeta \le \gamma\lambda(t) - \dot{\lambda}(t). \tag{33}$$

The following result is in the spirit of a convergence rate given for the objective function values on a sequence iteratively generated by an inertial-type algorithm recently obtained in [18, Theorem 1].

Theorem 4.1. *Let $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $1/\beta$-Lipschitz continuous gradient for $\beta > 0$ such that* $\operatorname{argmin}_{x\in\mathcal{H}} g(x) \neq \emptyset$. *Let $\Gamma : \mathcal{H} \to \mathcal{H}$ be a γ-elliptic operator, $\lambda : [0, +\infty) \to (0, +\infty)$ be a function fulfilling ($\lambda 6$), $u_0, v_0 \in \mathcal{H}$ and $x \in C^2([0, +\infty); \mathcal{H})$ be the unique global solution of* (32).

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

$$\begin{aligned} 0 \le & g\left(\frac{1}{T}\int_0^T x(t)dt\right) - g(x^*) \le \\ & \frac{1}{2\zeta T}\left[\|v_0 + \gamma(u_0 - x^*)\|^2 + \left(\gamma\|\Gamma - \gamma \operatorname{Id}\| + \frac{\lambda(0)}{\beta}\right)\|u_0 - x^*\|^2\right]. \end{aligned}$$

Proof. Let be $x^* \in \operatorname{argmin}_{x\in\mathcal{H}} g(x)$ and $T > 0$. Consider again the function $p : [0, +\infty) \to \mathbb{R}, p(t) = \frac{1}{2}\langle(\Gamma - \gamma \operatorname{Id})(x(t) - x^*), x(t) - x^*\rangle$ which was defined in (7). By using (32), the formula for the derivative of p, the positive semidefinitness of $\Gamma - \gamma \operatorname{Id}$, the convexity of g and ($\lambda 6$) we get for almost every $t \in [0, +\infty)$

$$\begin{aligned} & \frac{d}{dt}\left(\frac{1}{2}\|\dot{x}(t) + \gamma(x(t) - x^*)\|^2 + \gamma p(t) + \lambda(t)g(x(t))\right) \\ \le & -\zeta(g(x(t)) - g(x^*)) + \dot{\lambda}(t)g(x^*) \end{aligned}$$

After integration, by neglecting the nonnegative terms and by using that $g(x(T)) \geq g(x^*)$, it yields

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

The conclusion follows by using $p(0) = \frac{1}{2}\langle(\Gamma - \gamma \operatorname{Id})(u_0 - x^*), u_0 - x^*\rangle \leq \frac{1}{2}\|\Gamma - \gamma \operatorname{Id}\|\|u_0 - x^*\|^2$, $g(u_0) - g(x^*) \leq \frac{1}{2\beta}\|u_0 - x^*\|^2$, which is a consequence of the descent lemma 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. $\square$

Remark 4.1. Under assumption (λ6) on the relaxation function λ, 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 2.1 when λ is assumed to fulfill (λ1) (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). Take $a \geq 0$, $b > \frac{1}{\beta\gamma^2}$ and $0 \leq \rho \leq \gamma$. Then $\lambda(t) = \frac{1}{ae^{-\rho t}+b}$ is an example of a relaxation function which verifies assumption (λ1) and assumption (λ6) (with $0 < \zeta \leq \frac{\gamma b}{(a+b)^2}$).

Acknowledgments

The research of Ernö Robert Csetnek has been supported by FWF (Austrian Science Fund), Lise Meitner Programme, project M 1682-N25.

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] 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

[4] 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
[5] 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
[6] 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
[7] A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differentsial'nye Uravneniya 30(9), 1475–1486, 1994; translation in Differential Equations 30(9), 1365–1375, 1994
[8] H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, in: Optimization (Namur, 1998), 25–35, in: Lecture Notes in Economics and Mathematical Systems 481, Springer, Berlin, 2000
[9] H. Attouch, M. Marques Alves, B.F. Svaiter, *A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity* $O(1/n^2)$, Journal of Convex Analysis 23, 2016
[10] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations 248(6), 1315-1344, 2010
[11] 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
[12] 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
[13] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011
[14] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and its Applications 119(2), 235–259, 2003
[15] R.I. Boţ, E.R. Csetnek, *A dynamical system associated with the fixed*

points set of a nonexpansive operator, Journal of Dynamics and Differential Equations, DOI: 10.1007/s10884-015-9438-x, 2015

[16] R.I. Boţ, E.R. Csetnek, *Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems*, Journal of Mathematical Analysis and Applications 435(2), 1688-1700, 2016

[17] R.I. Boţ, E.R. Csetnek, C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems*, Applied Mathematics and Computation 256, 472–487, 2015

[18] E. Ghadimi, H.R. Feyzmahdavian, M. Johansson, *Global convergence of the Heavy-ball method for convex optimization*, arXiv:1412.7457

[19] A. Moudafi, M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, Journal of Computational and Applied Mathematics 155, 447–454, 2003

[20] 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

[21] J.-C. Pesquet, N. Pustelnik, *A parallel inertial proximal optimization method*, Pacific Journal of Optimization 8(2), 273–305, 2012

[22] R.T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific Journal of Mathematics 33(1), 209–216, 1970

[23] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Control and Optimization 14(5), 877–898, 1976