Convergence rates for forward-backward dynamical systems associated with strongly monotone inclusions

Radu Ioan Boţ *   Ernö Robert Csetnek †

July 5, 2016

Abstract. We investigate the convergence rates of the trajectories generated by implicit first and second-order dynamical systems associated to the determination of the zeros of the sum of a maximally monotone operator and a monotone and Lipschitz continuous one in a real Hilbert space. We show that these trajectories strongly converge with exponential rate to a zero of the sum, provided the latter is strongly monotone. We derive from here convergence rates for the trajectories generated by dynamical systems associated to the minimization of the sum of a proper, convex and lower semicontinuous function with a smooth convex one provided the objective function fulfills a strong convexity assumption. In the particular case of minimizing a smooth and strongly convex function, we prove that its values converge along the trajectory to its minimum value with exponential rate, too.

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

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

1 Introduction and preliminaries

The main topic of this paper is the investigation of convergence rates for implicit dynamical systems associated with monotone inclusion problems of the form

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in Ax^* + Bx^*, \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $B : \mathcal{H} \to \mathcal{H}$ is a monotone and $\frac{1}{\beta}$-Lipschitz continuous operator for $\beta > 0$ and $A + B$ is ρ-strongly monotone for $\rho > 0$. Dynamical systems of implicit type have been already considered in the literature in [1,2,7,9,13,15–18].

We deal in a first instance with the first-order dynamical system with variable relaxation parameters

$$\begin{cases} \dot{x}(t) = \lambda(t) \left[J_{\eta A}\Big(x(t) - \eta B(x(t))\Big) - x(t) \right] \\ x(0) = x_0, \end{cases} \tag{2}$$

*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.

where $x_0 \in \mathcal{H}$, $\lambda : [0,+\infty) \to [0,\infty)$ is a Lebesgue measurable function and $J_{\eta A}$ denotes the resolvent of the operator ηA for $\eta > 0$.

We notice that Abbas and Attouch considered in [1, Section 5.2] the dynamical system of same type

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\mu\Phi}\big(x(t) - \mu B(x(t))\big) \\ x(0) = x_0 \end{cases} \tag{3}$$

in connection to the determination of the zeros of $\partial\Phi + B$, where $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function, $B : \mathcal{H} \to \mathcal{H}$ is a cocoercive operator, $\partial\Phi$ denotes the convex subdifferential of Φ and $\text{prox}_{\mu\Phi}$ denotes the proximal point operator of $\mu\Phi$.

Before that, Antipin in [7] and Bolte in [15] studied the convergence of the trajectories generated by

$$\begin{cases} \dot{x}(t) + x(t) = P_C\big(x(t) - \mu\nabla g(x(t))\big) \\ x(0) = x_0 \end{cases} \tag{4}$$

to a minimizer of the smooth and convex function $g : \mathcal{H} \to \mathbb{R}$ over the nonempty, convex and closed set $C \subseteq \mathcal{H}$, where $\mu > 0$ and P_C denotes the projection operator on the set C.

In the second part of the paper we approach the monotone inclusion (1) via the second-order dynamical system with variable damping and relaxation parameters

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\Big[x(t) - J_{\eta A}\Big(x(t) - \eta B(x(t))\Big)\Big] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{5}$$

where $u_0, v_0 \in \mathcal{H}$, $\lambda : [0,+\infty) \to [0,\infty)$ and $\gamma : [0,+\infty) \to [0,\infty)$ are Lebesgue measurable functions, and $\eta > 0$.

Second-order dynamical systems of the form

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) - Tx(t) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{6}$$

for $\gamma > 0$ and $T : \mathcal{H} \to \mathcal{H}$ a nonexpansive operator, have been treated by Attouch and Alvarez in [8] in connection to the problem of approaching the fixed points of T, see also [12].

For the minimization of the smooth and convex function $g : \mathcal{H} \to \mathbb{R}$ over the nonempty, convex and closed set $C \subseteq \mathcal{H}$, a continuous in time second-order gradient-projection approach has been considered in [7, 8], having as starting point the dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) - P_C(x(t) - \eta\nabla g(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{7}$$

with constant damping parameter $\gamma > 0$ and constant step size $\eta > 0$.

For an exhaustive asymptotic analysis of the first and second-order dynamical systems (2) and (5), in case B is cocoercive, we refer the reader to [16] and [18], respectively. According to the above-named works, one can expect under mild assumptions on the relaxation and, in the second-order setting, on the damping functions, that the generated trajectories converge to a zero of $A + B$. The main scope of this paper is to show that when weakening the assumptions on B to monotonicity and Lipschitz continuity, however, provided that $A+B$ is strongly monotone, the trajectories converge strongly to the unique zero of $A+B$ with an exponential rate. Exponential convergence rates have been obtained also by Antipin in [7] for the dynamical systems (4) and (7), by imposing for the smooth function g supplementary strong convexity assumptions.

We transfer the results obtained for both first and second-order dynamical systems to optimization problems of the form

$$\min_{x\in\mathcal{H}} f(x)+g(x), \tag{8}$$

where $f:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ is a proper, convex and lower semicontinuous function, $g:\mathcal{H}\to\mathbb{R}$ is a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\beta>0$ and $f+g$ is ρ-strongly convex for $\rho>0$, by taking into consideration that its set of minimizers coincides with the solution set of the monotone inclusion problem

$$\text{find } x^*\in\mathcal{H} \text{ such that } 0\in\partial f(x^*)+\nabla g(x^*).$$

When further particularizing this context to the one of solving minimization problems like

$$\min_{x\in\mathcal{H}} g(x), \tag{9}$$

where $g:\mathcal{H}\to\mathbb{R}$ is a ρ-strongly convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\rho>0$ and $\beta>0$, we show that the values of g converge along the trajectories generated by the corresponding first and second-order dynamical systems to its minimum value also with exponential rate.

Dynamical systems approaching monotone inclusions and optimization problems enjoy much attention since the seventies of the last century, not only due to their intrinsic importance in areas like differential equations and applied functional analysis, but also because they have been recognized as a valuable tool for discovering and studying numerical algorithms for optimization problems obtained by time discretization of the continuous dynamics. The dynamic approach to iterative methods in optimization can furnish deep insights into the expected behaviour of the method and the techniques used in the continuous case can be adapted to obtain results for the discrete algorithm. For instance, Theorem 11 in this paper can be seen as the continuous counterpart of [21, Theorem 4], where recently a linear rate of convergence for the values of a convex and smooth function on a sequence iteratively generated by an inertial gradient-type algorithm has been provided. We also notice that the relaxation function $\lambda:[0,+\infty)\to[0,\infty)$ in the considered first- and second-order systems can be seen as the continuous counterpart of the sequences of relaxation parameters in the corresponding discrete forward-backward schemes.

The rest of this section is devoted to some notations and definitions used in the paper. We denote by $\mathcal{H}$ a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $\|\cdot\|=\sqrt{\langle\cdot,\cdot\rangle}$. 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 $J_A=(\operatorname{Id}+A)^{-1}$, where $\operatorname{Id}:\mathcal{H}\to\mathcal{H}$ denotes the identity operator on $\mathcal{H}$. If A is maximally monotone, then $J_A:\mathcal{H}\to\mathcal{H}$ is single-valued and maximally monotone (see [14, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma>0$ we have (see [14, Proposition 23.2])

$$p\in J_{\gamma A}x \text{ if and only if } (p,\gamma^{-1}(x-p))\in\operatorname{Gr} A. \tag{10}$$

The operator A is said to be *ρ-strongly monotone* for $\rho > 0$, if

$$\langle x - y, u - v\rangle \geq \rho\|x - y\|^2 \ \forall (x, u), (y, v) \in \operatorname{Gr} A. \tag{11}$$

As in [2,13], we consider the following definition of an absolutely continuous function.

Definition 1 (see, for instance, [2,13]) A function $x : [0, b] \to \mathcal{H}$ (where $b > 0$) is said to be absolutely continuous if one of the following equivalent properties holds:

(i) there exists an integrable function $y : [0, b] \to \mathcal{H}$ 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 is Lebesgue integrable on $[0, b]$;

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

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

2 Converges rates for first-order dynamical systems

The starting point of the investigations we carry out in this section is the first-order dynamical system (2) that we formulated in relation to the monotone inclusion problem (1). We say that $x : [0, +\infty) \to \mathcal{H}$ is a *strong global solution* of (2), if the following properties are satisfied:

(i) $x : [0, +\infty) \to \mathcal{H}$ is *locally absolutely continuous*, that is, absolutely continuous on each interval $[0, b]$ for $0 < b < +\infty$;

(ii) For almost every $t \in [0, +\infty)$ it holds $\dot{x}(t) = \lambda(t)\left[J_{\eta A}\Big(x(t) - \eta B(x(t))\Big) - x(t)\right]$;

(iii) $x(0) = x_0$.

In case λ is locally integrable, the existence and uniqueness of strong global solutions of the system (2) follow from the Cauchy-Lipschitz-Picard Theorem, by noticing that the operator $T = J_{\eta A} \circ (\operatorname{Id} - \eta B) - \operatorname{Id}$ is Lipschitz continuous (see also [16, Section 2]).

The following result can bee seen as the continuous counterpart of [14, Proposition 25.9], where it is shown that the sequence iteratively generated by the forward-backward algorithm linearly converges to the unique solution of (1), provided that one of the two involved operators is strongly monotone.

Theorem 1 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $B : \mathcal{H} \to \mathcal{H}$ a monotone and $\frac{1}{\beta}$-Lipschitz continuous operator for $\beta > 0$ such that $A + B$ is ρ-strongly monotone for $\rho > 0$ and x^* be the unique point in $\operatorname{zer}(A + B)$. Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a Lebesgue measurable function such that there exist real numbers $\underline{\lambda}$ and $\overline{\lambda}$ fulfilling*

$$0 < \underline{\lambda} \leq \inf_{t\geq 0} \lambda(t) \leq \sup_{t\geq 0} \lambda(t) \leq \overline{\lambda}.$$

Chose $\alpha > 0$ and $\eta > 0$ such that

$$\alpha < 2\rho\beta^2\underline{\lambda} \text{ and } \frac{1}{\beta} + \frac{\overline{\lambda}}{2\alpha} \leq \rho + \frac{1}{\eta}.$$

If $x_0 \in \mathcal{H}$ *and* $x : [0,+\infty) \to \mathcal{H}$ *is the unique strong global solution of the dynamical system* (2)*, then for every* $t \in [0,+\infty)$ *one has*

$$\|x(t)-x^*\|^2 \le \|x_0 - x^*\|^2 \exp(-Ct),$$

where $C := \frac{2\rho\underline{\lambda} - \frac{\alpha}{\beta^2}}{2\rho + \frac{1}{\eta}} > 0$.

Proof. Notice that B is a maximally monotone operator (see [14, Corollary 20.25]) and, since B has full domain, $A+B$ is maximally monotone, too (see [14, Corollary 24.4]). Therefore, due to the strong monotonicity of $A+B$, $\text{zer}(A+B)$ is a singleton (see [14, Corollary 23.37]).

A direct consequence of (2) and of the definition of the resolvent is the inclusion

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

which holds for almost every $t \in [0,+\infty)$. Combining it with $0 \in (A+B)(x^*)$ and the strong monotonicity of $A+B$, it yields for almost every $t \in [0,+\infty)$

$$\rho\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|^2 \le$$
$$\left\langle \frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*, -\frac{1}{\eta\lambda(t)}\dot{x}(t) - B(x(t)) + B\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right)\right\rangle.$$

By using the notation $h(t) = \frac{1}{2}\|x(t)-x^*\|^2$ for $t \in [0,+\infty)$, the Cauchy-Schwarz inequality, the Lipschitz property of B and the fact that $\dot{h}(t) = \langle x(t)-x^*, \dot{x}(t)\rangle$, we deduce that for almost every $t \in [0,+\infty)$

$$\begin{aligned}
\rho\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|^2 \le & -\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{1}{\lambda(t)}\left\langle \dot{x}(t), B\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t))\right\rangle \\
& - \frac{1}{\eta\lambda(t)}\dot{h}(t) + \left\langle x(t)-x^*, B\left(\frac{1}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t))\right\rangle \\
\le & -\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{1}{\beta\lambda^2(t)}\|\dot{x}(t)\|^2 - \frac{1}{\eta\lambda(t)}\dot{h}(t) \\
& + \frac{1}{\beta\lambda(t)}\|x(t)-x^*\|\|\dot{x}(t)\| \\
\le & -\frac{1}{\eta\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{1}{\beta\lambda^2(t)}\|\dot{x}(t)\|^2 - \frac{1}{\eta\lambda(t)}\dot{h}(t) \\
& + \frac{\alpha}{\beta^2\lambda(t)}h(t) + \frac{1}{2\alpha\lambda(t)}\|\dot{x}(t)\|^2.
\end{aligned}$$

As

$$\rho\left\|\frac{1}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|^2 = \frac{\rho}{\lambda^2(t)}\|\dot{x}(t)\|^2 + \frac{2\rho}{\lambda(t)}\dot{h}(t) + 2\rho h(t),$$

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

$$\left(\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)}\right)\dot{h}(t) + \left(2\rho - \frac{\alpha}{\beta^2\lambda(t)}\right)h(t) +$$
$$\left(\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{2\alpha\lambda(t)}\right)\|\dot{x}(t)\|^2 \le 0.$$

However, the way in which the involved parameters were chosen imply for almost every $t \in [0, +\infty)$ that

$$\left(\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)}\right)\dot{h}(t) + \left(2\rho - \frac{\alpha}{\beta^2\lambda(t)}\right)h(t) \leq 0 \tag{12}$$

or, equivalently,

$$\dot{h}(t) + \frac{2\rho\lambda(t) - \frac{\alpha}{\beta^2}}{2\rho + \frac{1}{\eta}}h(t) \leq 0.$$

This further implies

$$\dot{h}(t) + Ch(t) \leq 0$$

for almost every $t \in [0, +\infty)$. By multiplying this inequality with $\exp(Ct)$ and integrating from 0 to T, where $T \geq 0$, one easily obtains the conclusion. ■

Remark 2

(a) By time rescaling arguments one could consider $\lambda(t) = 1$ for every $t \geq 0$ and, consequently, investigate the asymptotic properties of the system

$$\begin{cases} \dot{x}(t) + M(x(t)) = 0 \\ x(0) = x_0, \end{cases} \tag{13}$$

where $M : \mathcal{H} \to \mathcal{H}$ is defined by $M = \operatorname{Id} - J_{\eta A} \circ (\operatorname{Id} - \eta B)$. In the hypotheses of Theorem 1 the operator M satisfies the following inequality for all $x \in \mathcal{H}$:

$$\left(2\rho + \frac{1}{\eta}\right)\langle Mx, x - x^*\rangle \geq \left(\rho - \frac{\alpha}{2\beta^2}\right)\|x - x^*\|^2 + \left(\rho + \frac{1}{\eta} - \frac{1}{\beta} - \frac{1}{2\alpha}\right)\|Mx\|^2. \tag{14}$$

This follows by using the same arguments as used in the proof of Theorem 1, namely the definition of the resolvent operator, the inclusion $0 \in (A + B)(x^*)$ and the strong monotonicity of $A+B$. Coming back to the system (13), the exponential convergence rate for the trajectory is further obtained by applying the Gronwall Lemma in the inequality

$$\left(2\rho + \frac{1}{\eta}\right)\langle \dot{x}(t), x(t) - x^*\rangle + \left(\rho - \frac{\alpha}{2\beta^2}\right)\|x(t) - x^*\|^2 \leq 0,$$

which is nothing else than relation (12) in the proof of Theorem 1.

(b) Notice that by chosing the involved parameters as in Theorem 1, relation (14) yields the inequality

$$\left(2\rho + \frac{1}{\eta}\right)\langle Mx, x - x^*\rangle \geq \left(\rho - \frac{\alpha}{2\beta^2}\right)\|x - x^*\|^2 \ \forall x \in \mathcal{H},$$

where $Mx^* = 0$. Thus the operator M satisfies a strong monotone property in the sense of Pazy (see relation (11.2) in Theorem 11.2 in [23]). However, the hypotheses of Theorem 1 do not imply in general the strong monotonicity of the operator M in the sense of (11), thus the result presented in Theorem 1 does not fall into the framework of the classical result concerning exponential convergence rates for the semigroup generated by a strongly monotone operator as presented in [20, Theorem 3.9].

Further, we discuss some situations when the operator M is strongly monotone in the classical sense (see (11)). We start with two trivial cases. The first one is $Ax = 0$ for every $x \in \mathcal{H}$ and B is strongly monotone. The second one is $Bx = 0$ for every $x \in \mathcal{H}$ and A is strongly monotone, in which case $J_{\eta A}$ is a contraction (see [14, Proposition 23.11]), hence $M = \mathrm{Id} - J_{\eta A}$ is strongly monotone. Other situations follow in the framework of [14, Proposition 25.9]: i) if A is strongly monotone, B is β-cocoercive (that is $\langle x-y, Bx-By\rangle \geq \beta\|Bx-By\|^2$ for all $x, y \in \mathcal{H}$) and $\eta < 2\beta$; ii) if B is θ-strongly monotone and β^{-1}-Lipschitz continuous, $\theta\beta \leq 1$ and $\eta < 2\theta\beta^2$.

We come now to the convex optimization problem (8) and notice that, since $\operatorname{argmin}(f+g) = \operatorname{zer}(\partial(f+g)) = \operatorname{zer}(\partial f + \nabla g)$, one can approach this set by means of the trajectories of the dynamical system (2) written for $A = \partial f$ and $B = \nabla g$. Here, $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$, defined by

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

if $f(x) \in \mathbb{R}$ and $\partial f(x) = \emptyset$, otherwise, denotes the *convex subdifferential* of f, which is a maximally monotone operator, provided that f is proper, convex and lower semicontinuous (see [25]). We notice that, for $\eta > 0$, the resolvent of $\eta\partial f$ is given by $J_{\eta\partial f} = \operatorname{prox}_{\eta f}$ (see [14]), where $\operatorname{prox}_{\eta f} : \mathcal{H} \to \mathcal{H}$,

$$\operatorname{prox}_{\eta f}(x) = \operatorname*{argmin}_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\eta}\|y - x\|^2 \right\}, \tag{15}$$

denotes the *proximal point operator* of ηf. This being said, the dynamical system (2) becomes

$$\left\{ \begin{array}{l} \dot{x}(t) = \lambda(t)\left[\operatorname{prox}_{\eta f}\Big(x(t) - \eta\nabla g(x(t))\Big) - x(t)\right] \\ x(0) = x_0. \end{array} \right. \tag{16}$$

The following result is a direct consequence of Theorem 1. Let us also notice that $f + g$ is said to be ρ-strongly convex for $\rho > 0$, if $f + g - \frac{\rho}{2}\|\cdot\|^2$ is a convex function. In this situation $\partial(f + g) = \partial f + \nabla g$ is a ρ-strongly monotone operator (see [14, Example 22.3(iv)].)

Theorem 3 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function, $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\beta > 0$ such that $f + g$ is ρ-strongly convex for $\rho > 0$ and x^* be the unique minimizer of $f + g$ over $\mathcal{H}$. Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a Lebesgue measurable function such that there exist real numbers $\underline{\lambda}$ and $\overline{\lambda}$ fulfilling*

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t) \leq \sup_{t \geq 0} \lambda(t) \leq \overline{\lambda}.$$

Chose $\alpha > 0$ and $\eta > 0$ such that

$$\alpha < 2\rho\beta^2\underline{\lambda} \ \textit{and} \ \frac{1}{\beta} + \frac{\overline{\lambda}}{2\alpha} \leq \rho + \frac{1}{\eta}.$$

If $x_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ is the unique strong global solution of the dynamical system (16), then for every $t \in [0. + \infty)$ one has

$$\|x(t) - x^*\|^2 \leq \|x_0 - x^*\|^2 \exp(-Ct),$$

where $C := \frac{2\rho\underline{\lambda} - \frac{\alpha}{\beta^2}}{2\rho + \frac{1}{\eta}} > 0$.

Remark 4 The explicit discretization of (16) with respect to the time variable t, with step size $h_n > 0$ and initial point $x_0 \in \mathcal{H}$, yields the following iterative scheme:

$$\frac{x_{n+1} - x_n}{h_n} = \lambda_n \left[\text{prox}_{\eta f}\left(x_n - \gamma \nabla g x_n\right) - x_n\right] \ \forall n \geq 0.$$

For $h_n = 1$ this becomes

$$x_{n+1} = x_n + \lambda_n \left[\text{prox}_{\eta f}\left(x_n - \gamma \nabla g x_n\right) - x_n\right] \ \forall n \geq 0, \tag{17}$$

which is the classical forward-backward algorithm with relaxation parameters $(\lambda_n)_{n\geq 0}$ for finding the minimizers of $f + g$ (see [14]). For this iterative scheme it is known, at least in the situation $\lambda_n = 1$ for all $n \geq 0$, that linear convergence is achieved in case at least one of the functions f and g satisfies a strong convexity assumption.

In the last part of this section we approach the convex minimization problem (9) via the first-order dynamical system

$$\begin{cases} \dot{x}(t) + \lambda(t)\nabla g(x(t)) = 0 \\ x(0) = x_0. \end{cases} \tag{18}$$

The following result quantifies the rate of convergence of g to its minimum value along the trajectories generated by (18).

Theorem 5 *Let $g : \mathcal{H} \to \mathbb{R}$ be a ρ-strongly convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\rho > 0$ and $\beta > 0$ and x^* be the unique minimizer of g over $\mathcal{H}$. Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a Lebesgue measurable and locally integrable function such that there exists a real number $\underline{\lambda} \in \mathbb{R}$ fulfilling*

$$0 < \underline{\lambda} \leq \inf_{t\geq 0} \lambda(t).$$

Chose $\alpha > 0$ such that

$$\alpha \leq 2\underline{\lambda}\beta\rho^2.$$

If $x_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ is the unique strong global solution of the dynamical system (18), then for every $t \in [0, +\infty)$ one has

$$0 \leq \frac{\rho}{2}\|x(t)-x^*\|^2 \leq g(x(t))-g(x^*) \leq (g(x_0)-g(x^*))\exp(-\alpha t) \leq \frac{1}{2\beta}\|x_0-x^*\|^2 \exp(-\alpha t).$$

Proof. The second inequality is a consequence of the strong convexity of the function g. Further, we recall that according to the descent lemma, which is valid for an arbitrary differentiable function with Lipschitz continuous gradient (see [22, Lemma 1.2.3]), we have

$$g(u) \leq g(v) + \langle \nabla g(v), u - v\rangle + \frac{1}{2\beta}\|u - v\|^2 \ \forall u, v \in \mathcal{H}.$$

By setting in the previous relation, for every $t \in [0, +\infty)$, $u := x(t)$ and $v := x^*$ and by taking into account that $\nabla g(x^*) = 0$, we obtain

$$g(x(t)) - g(x^*) \leq \frac{1}{2\beta}\|x(t) - x^*\|^2. \tag{19}$$

From here, the last inequality in the conclusion follows automatically.

Using the strong convexity of g we have for every $t \in [0, +\infty)$ that

$$\rho\|x(t) - x^*\|^2 \leq \langle x(t) - x^*, \nabla g(x(t))\rangle \leq \|x(t) - x^*\|\|\nabla g(x(t))\|,$$

thus

$$\rho\|x(t) - x^*\| \leq \|\nabla g(x(t))\|. \tag{20}$$

Finally, from the first equation in (18), (19), (20) and using the way in which α was chosen, we obtain for almost every $t \in [0, +\infty)$

$$\begin{aligned}
\frac{d}{dt}\big(g(x(t)) - g(x^*)\big) + \alpha(g(x(t)) - g(x^*)) &= \langle \dot{x}(t), \nabla g(x(t))\rangle + \alpha(g(x(t)) - g(x^*)) \\
&\leq -\lambda(t)\|\nabla g(x(t))\|^2 + \frac{\alpha}{2\beta}\|x(t) - x^*\|^2 \\
&\leq \left(-\lambda(t) + \frac{\alpha}{2\beta\rho^2}\right)\|\nabla g(x(t))\|^2 \\
&\leq 0.
\end{aligned}$$

By multiplying this inequality with $\exp(\alpha t)$ and integrating from 0 to T, where $T \geq 0$, one easily obtains also the third inequality. ■

3 Converges rates for second-order dynamical systems

The starting point of the investigations we go through in this section is again the monotone inclusion problem (1), however, this time approached via the second-order dynamical system (5). We say that $x : [0, +\infty) \to \mathcal{H}$ is a *strong global solution* of (5), if the following properties are satisfied:

(i) $x, \dot{x} : [0, +\infty) \to \mathcal{H}$ are locally absolutely continuous;

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

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\left[x(t) - J_{\eta A}\Big(x(t) - \eta B(x(t))\Big)\right] = 0;$$

(iii) $x(0) = u_0, \dot{x}(0) = v_0$.

By assuming that λ, γ are locally integrable, the existence and uniqueness of strong global solutions of the system (5) follow from the Cauchy-Lipschitz-Picard Theorem applied in a product space (see also [18]).

The following result will be useful when deriving the convergence rates.

Lemma 6 *Let $h, \gamma, b_1, b_2, b_3, u : [0, +\infty) \to \mathbb{R}$ be given functions such that h, γ, b_2, u are locally absolutely continuous and $\dot{h}$ is locally absolutely continuous, too. Assume that*

$$h(t), b_2(t), u(t) \geq 0 \ \forall t \in [0, +\infty)$$

and that there exists $\underline{\gamma} > 1$ such that

$$\gamma(t) \geq \underline{\gamma} > 1 \ \forall t \in [0, +\infty).$$

Further, assume that for almost every $t \in [0, +\infty)$ one has

$$\gamma(t) + \dot{\gamma}(t) \leq b_1(t) + 1, \tag{21}$$

$$b_2(t) + \dot{b}_2(t) \leq b_3(t) \tag{22}$$

and

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + b_1(t)h(t) + b_2(t)\dot{u}(t) + b_3(t)u(t) \leq 0. \tag{23}$$

Then there exists $M > 0$ such that the following statements hold:

(i) if $1 < \underline{\gamma} < 2$, then for almost every $t \in [0, +\infty)$

$$0 \leq h(t) \leq \left(h(0) + \frac{M}{2 - \underline{\gamma}}\right) \exp(-(\underline{\gamma} - 1)t);$$

(ii) if $2 < \underline{\gamma}$, then for almost every $t \in [0, +\infty)$

$$0 \leq h(t) \leq h(0)\exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2}\exp(-t) \leq \left(h(0) + \frac{M}{\underline{\gamma} - 2}\right)\exp(-t);$$

(iii) if $\underline{\gamma} = 2$, then for almost every $t \in [0, +\infty)$

$$0 \leq h(t) \leq (h(0) + Mt)\exp(-t).$$

Proof. We multiply the inequality (23) with $\exp(t)$ and use the identities

$$\begin{aligned}
\exp(t)\ddot{h}(t) &= \frac{d}{dt}\big(\exp(t)\dot{h}(t)\big) - \exp(t)\dot{h}(t) \\
\exp(t)\dot{u}(t) &= \frac{d}{dt}(\exp(t)u(t)) - \exp(t)u(t) \\
\exp(t)\dot{h}(t) &= \frac{d}{dt}\big(\exp(t)h(t)\big) - \exp(t)h(t)
\end{aligned}$$

in order to derive for almost every $t \in [0, +\infty)$ the inequality

$$\begin{aligned}
&\frac{d}{dt}\big(\exp(t)\dot{h}(t)\big) + (\gamma(t) - 1)\frac{d}{dt}\big(\exp(t)h(t)\big) + \\
&\exp(t)h(t)(b_1(t) + 1 - \gamma(t)) + b_2(t)\frac{d}{dt}\big(\exp(t)u(t)\big) + (b_3(t) - b_2(t))\exp(t)u(t) \leq 0.
\end{aligned}$$

By using also

$$\begin{aligned}
(\gamma(t) - 1)\frac{d}{dt}\big(\exp(t)h(t)\big) &= \frac{d}{dt}\Big((\gamma(t) - 1)\exp(t)h(t)\Big) - \dot{\gamma}(t)\exp(t)h(t) \\
b_2(t)\frac{d}{dt}\big(\exp(t)u(t)\big) &= \frac{d}{dt}\big(b_2(t)\exp(t)u(t)\big) - \dot{b}_2(t)\exp(t)u(t)
\end{aligned}$$

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

$$\begin{aligned}
&\frac{d}{dt}\big(\exp(t)\dot{h}(t)\big) + \frac{d}{dt}\Big((\gamma(t) - 1)\exp(t)h(t)\Big) + \frac{d}{dt}\big(b_2(t)\exp(t)u(t)\big) + \\
&\big(b_1(t) + 1 - \gamma(t) - \dot{\gamma}(t)\big)\exp(t)h(t) + \big(b_3(t) - b_2(t) - \dot{b}_2(t)\big)\exp(t)u(t) \leq 0.
\end{aligned}$$

The hypotheses regarding the parameters involved imply that the function

$$t \to \exp(t)\dot{h}(t) + (\gamma(t) - 1)\exp(t)h(t) + b_2(t)\exp(t)u(t)$$

is monotonically decreasing, hence there exists $M > 0$ such that

$$\exp(t)\dot{h}(t) + (\gamma(t) - 1)\exp(t)h(t) + b_2(t)\exp(t)u(t) \leq M.$$

Since $u(t), b_2(t) \geq 0$ we get

$$\dot{h}(t) + (\gamma(t) - 1)h(t) \leq M\exp(-t),$$

hence

$$\dot{h}(t) + (\underline{\gamma} - 1)h(t) \leq M\exp(-t)$$

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

$$\frac{d}{dt}\big(\exp((\underline{\gamma} - 1)t)h(t)\big) \leq M\exp((\underline{\gamma} - 2)t),$$

for every $t \in [0, +\infty)$, from which the conclusion follows easily by integration. ■

We come now to the first main result of this section.

Theorem 7 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $B : \mathcal{H} \to \mathcal{H}$ a monotone and $\frac{1}{\beta}$-Lipschitz continuous operator for $\beta > 0$ such that $A + B$ is ρ-strongly monotone for $\rho > 0$ and x^* be the unique point in $\operatorname{zer}(A + B)$. Chose $\alpha, \delta \in (0, 1)$ and $\eta > 0$ such that $\delta\beta\rho < 1$ and $\frac{1}{\eta} = \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta} - \rho > 0$.*

Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling for every $t \in [0, +\infty)$

$$\theta(t) := \lambda(t)\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}}$$
$$\leq \lambda(t)\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} + \lambda^2(t)\left(\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}\right)^2$$

and such that there exists a real number $\underline{\lambda}$ with the property that

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t)$$

and

$$2 < \theta := \underline{\lambda}\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}}.$$

Further, let $\gamma : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling

$$\frac{1 + \sqrt{1 + 4\theta(t)}}{2} \leq \gamma(t) \leq 1 + \lambda(t)\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} \text{ for every } t \in [0, +\infty) \tag{24}$$

and

$$\dot{\gamma}(t) \leq 0 \text{ and } \frac{d}{dt}\left(\frac{\gamma(t)}{\lambda(t)}\right) \leq 0 \text{ for almost every } t \in [0, +\infty). \tag{25}$$

Let $u_0, v_0 \in \mathcal{H}$ and $x : [0,+\infty) \to \mathcal{H}$ be the unique strong global solution of the dynamical system (5).

Then $\gamma(t) \geq \underline{\gamma} := \frac{1+\sqrt{1+4\theta}}{2} > 2$ for every $t \in [0,+\infty)$ and there exists $M > 0$ such that for every $t \in [0,+\infty)$

$$\begin{aligned} 0 \leq \|x(t)-x^*\|^2 &\leq \|u_0 - x^*\|^2 \exp(-(\underline{\gamma}-1)t) + \frac{M}{\underline{\gamma}-2}\exp(-t) \\ &\leq \left(\|u_0 - x^*\|^2 + \frac{M}{\underline{\gamma}-2}\right)\exp(-t). \end{aligned}$$

Proof. From the definition of the resolvent we have for almost every $t \in [0,+\infty)$

$$\begin{aligned} &B\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t)) - \frac{1}{\eta\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t) \in \\ &(A+B)\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right). \end{aligned} \tag{26}$$

We combine this with $0 \in (A+B)x^*$, the strong monotonicity of $A+B$, the Lipschitz continuity of B and, by also using the Cauchy-Schwartz inequality, we get for almost every $t \in [0,+\infty)$

$$\begin{aligned} &\frac{\rho}{\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2 + \frac{2\rho}{\lambda(t)}\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\rangle + \rho\|x(t)-x^*\|^2 \\ =&\rho\left\|\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*\right\|^2 \\ \leq&\left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*, B\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t))\right\rangle \\ &-\left\langle \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t) - x^*, \frac{1}{\eta\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\eta\lambda(t)}\dot{x}(t)\right\rangle \\ =&\frac{1}{\lambda(t)}\left\langle \ddot{x}(t)+\gamma(t)\dot{x}(t), B\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t))\right\rangle \\ &+\left\langle x(t)-x^*, B\left(\frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + x(t)\right) - B(x(t))\right\rangle \\ &-\frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2 - \frac{1}{\eta\lambda(t)}\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\rangle \\ \leq&\frac{1}{\beta\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2 - \frac{1}{\eta\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2 \\ &+\frac{1}{4\rho\beta^2\alpha\lambda^2(t)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2 + \rho\alpha\|x(t)-x^*\|^2 - \frac{1}{\eta\lambda(t)}\langle x(t)-x^*, \ddot{x}(t)+\gamma(t)\dot{x}(t)\rangle. \end{aligned}$$

Using again the notation $h(t) = \frac{1}{2}\|x(t)-x^*\|^2$, we have for almost every $t \in [0,+\infty)$

$$\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2 = \|\ddot{x}(t)\|^2 + \gamma^2(t)\|\dot{x}(t)\|^2 + \gamma(t)\frac{d}{dt}(\|\dot{x}(t)\|^2) \tag{27}$$

and

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

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

$$\begin{aligned}
&\left(\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)} \right) \|\ddot{x}(t)\|^2 \\
&+ \left[\gamma^2(t) \left(\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)} \right) - \frac{2\rho}{\lambda(t)} - \frac{1}{\eta\lambda(t)} \right] \|\dot{x}(t)\|^2 \\
&+ \gamma(t) \left(\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)} \right) \frac{d}{dt} \left(\|\dot{x}(t)\|^2 \right) \\
&+ \left(\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)} \right) \ddot{h}(t) + \gamma(t) \left(\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)} \right) \dot{h}(t) + 2\rho(1-\alpha)h(t) \le 0.
\end{aligned}$$

The hypotheses imply that

$$\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)} = \frac{1}{\lambda^2(t)} \left(\rho + \frac{1}{\eta} - \frac{1}{\beta} - \frac{1}{4\rho\beta^2\alpha} \right) > 0,$$

hence the first term in the left hand side of the above inequality can be neglected and we obtain for almost every $t \in [0, +\infty)$ that

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + b_1(t)h(t) + b_2(t)\frac{d}{dt}(\|\dot{x}(t)\|^2) + b_3(t)\|\dot{x}(t)\|^2 \le 0, \tag{28}$$

where

$$b_1(t) := \lambda(t)\frac{2\rho(1-\alpha)}{2\rho + \frac{1}{\eta}} > 0$$

$$b_2(t) := \gamma(t)\frac{\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)}}{\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)}} = \frac{\gamma(t)}{\lambda(t)} \frac{\rho + \frac{1}{\eta} - \frac{1}{\beta} - \frac{1}{4\rho\beta^2\alpha}}{2\rho + \frac{1}{\eta}} > 0$$

and

$$b_3(t) := \frac{\gamma^2(t) \left(\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)} \right) - \frac{2\rho}{\lambda(t)} - \frac{1}{\eta\lambda(t)}}{\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)}}.$$

This shows that (23) in Lemma 6 for $u := \|\dot{x}(\cdot)\|^2$ is fulfilled. In order to apply Lemma 6, we have only to prove that (21) and (22) are satisfied, as every other assumption in this statement is obviously guaranteed.

A simple calculation shows that

$$b_3(t) \ge b_2(t) \Longleftrightarrow \gamma^2(t) - \gamma(t) \ge \frac{\frac{2\rho}{\lambda(t)} + \frac{1}{\eta\lambda(t)}}{\frac{\rho}{\lambda^2(t)} + \frac{1}{\eta\lambda^2(t)} - \frac{1}{\beta\lambda^2(t)} - \frac{1}{4\rho\beta^2\alpha\lambda^2(t)}} = \theta(t), \tag{29}$$

which is true according to (24), thus $b_3(t) \ge b_2(t)$ for every $t \in [0, +\infty)$. On the other hand (see (25)),

$$\dot{b}_2(t) \le 0$$

for almost every $t \in [0, +\infty)$, from which (22) follows.

Further, again by using (24), observe that

$$1 + b_1(t) = 1 + \lambda(t)\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha} \right) \frac{1}{\delta}} \ge \gamma(t)$$

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

$$\dot{\gamma}(t) \leq 0$$

for almost every $t \in [0, +\infty)$, shows that (21) is also fulfilled.

The conclusion follows from Lemma 6(ii), by noticing that $\underline{\gamma} > 2$, as $\theta > 2$. ■

Remark 8 One can notice that when $\dot{\gamma}(t) \leq 0$ for almost every $t \in [0, +\infty)$, the second assumption in (25) is fulfilled provided that $\dot{\lambda}(t) \geq 0$ for almost every $t \in [0, +\infty)$.

Further, we would like to point out that one can oviously chose $\lambda(t) = \underline{\lambda}$ and $\gamma(t) = \gamma$ for every $t \in [0, +\infty)$, where

$$\begin{aligned} 2 < \theta &:= \underline{\lambda}\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}} \\ &\leq \underline{\lambda}\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} + \underline{\lambda}^2\left(\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}\right)^2 \end{aligned}$$

and

$$\frac{1 + \sqrt{1+4\theta}}{2} \leq \gamma \leq 1 + \underline{\lambda}\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}.$$

When considering the convex optimization problem (8), the second-order dynamical system (5) written for $A = \partial f$ and $B = \nabla g$ becomes

$$\left\{ \begin{array}{l} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)\left[x(t) - \operatorname{prox}_{\eta f}\left(x(t) - \eta\nabla g(x(t))\right)\right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{array} \right. \tag{30}$$

Theorem 7 gives rise to the following result.

Theorem 9 *Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function, $g : \mathcal{H} \to \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$-Lipschitz continuous gradient for $\beta > 0$ such that $f + g$ is ρ-strongly convex for $\rho > 0$ and x^* be the unique minimizer of $f + g$ over $\mathcal{H}$. Chose $\alpha, \delta \in (0, 1)$ and $\eta > 0$ such that $\delta\beta\rho < 1$ and $\frac{1}{\eta} = \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta} - \rho > 0$.*

Let $\lambda : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling for every $t \in [0, +\infty)$

$$\begin{aligned} \theta(t) &:= \lambda(t)\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}} \\ &\leq \lambda(t)\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}} + \lambda^2(t)\left(\frac{2\rho(1-\alpha)}{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}\right)^2 \end{aligned}$$

and such that there exists a real number $\underline{\lambda}$ with the property that

$$0 < \underline{\lambda} \leq \inf_{t \geq 0} \lambda(t)$$

and

$$2 < \theta := \underline{\lambda}\frac{\delta}{1-\delta}\frac{\rho + \left(\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}\right)\frac{1}{\delta}}{\frac{1}{\beta} + \frac{1}{4\rho\beta^2\alpha}}.$$

Further, let $\gamma : [0, +\infty) \to [0, +\infty)$ *be a locally absolutely continuous function fulfilling* (24) *and* (25), $u_0, v_0 \in \mathcal{H}$ *and* $x : [0, +\infty) \to \mathcal{H}$ *be the unique strong global solution of the dynamical system* (30).

Then $\gamma(t) \geq \underline{\gamma} := \frac{1+\sqrt{1+4\theta}}{2} > 2$ *for every* $t \in [0, +\infty)$ *and there exists* $M > 0$ *such that for every* $t \in [0, +\infty)$

$$\begin{aligned} 0 \leq \|x(t) - x^*\|^2 &\leq \|u_0 - x^*\|^2 \exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2}\exp(-t) \\ &\leq \left(\|u_0 - x^*\|^2 + \frac{M}{\underline{\gamma} - 2}\right)\exp(-t). \end{aligned}$$

Remark 10 The explicit discretization of (30) 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 $x_0 := u_0 \in \mathcal{H}$ and $x_1 := v_0 \in \mathcal{H}$, yields the following iterative scheme

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

For $h_n = 1$ this becomes

$$x_{n+1} = \left(1 - \frac{\lambda_n}{1+\gamma_n}\right)x_n + \frac{\lambda_n}{1+\gamma_n}\operatorname{prox}_{\eta f}\left(x_n - \eta\nabla g(x_n)\right) + \frac{\lambda_n}{1+\gamma_n}(x_n - x_{n-1}) \ \forall n \geq 1, \tag{31}$$

which is the forward-backward algorithm for minimizing $f + g$ with relaxation variables and inertial effects. The suitable control of the inertial term by means of the variable parameters λ_n and γ_n can increase the speed of convergence of the algorithm (31). Thus, the function λ used in the dynamical system gives rise to the sequence of relaxation variables $(\lambda_n)_{n\geq 1}$ in (31), which have been considered in the literature in order to gain more freedom in the choice of the parameters involved in the numerical scheme and to accelerate the algorithm.

Finally, we approach the convex minimization problem (9) via the second-order dynamical system

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

and provide an exponential rate of convergence of g to its minimum value along the generated trajectories. The following result can be seen as the continuous counterpart of [21, Theorem 4], where recently a linear rate of convergence for the values of g on a sequence iteratively generated by an inertial-type algorithm has been obtained.

Theorem 11 *Let* $g : \mathcal{H} \to \mathbb{R}$ *be a* ρ*-strongly convex and (Fréchet) differentiable function with* $\frac{1}{\beta}$*-Lipschitz continuous gradient for* $\rho > 0$ *and* $\beta > 0$ *and* x^* *be the unique minimizer of* g *over* $\mathcal{H}$.

Let $\alpha : [0, +\infty) \to \mathbb{R}$ be a Lebesgue measurable function such that there exists $\underline{\alpha} > 1$ with

$$\inf_{t \geq 0} \alpha(t) \geq \max \left\{ \underline{\alpha}, \frac{2}{\beta^2 \rho^2} - 1 \right\} \tag{33}$$

and $\lambda : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling for every $t \in [0, +\infty)$

$$\frac{\alpha(t)}{\beta \rho^2} \leq \lambda(t) \leq \frac{\beta}{2} \big(\alpha(t) + \alpha^2(t) \big). \tag{34}$$

Further, let $\gamma : [0, +\infty) \to [0, +\infty)$ be a locally absolutely continuous function fulfilling

$$\frac{1 + \sqrt{1 + 8 \frac{\lambda(t)}{\beta}}}{2} \leq \gamma(t) \leq 1 + \alpha(t) \text{ for every } t \in [0, +\infty) \tag{35}$$

and (25).

Let $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \to \mathcal{H}$ be the unique strong global solution of the dynamical system (32).

Then $\gamma(t) \geq \underline{\gamma} := \frac{1 + \sqrt{1 + 8 \frac{\underline{\alpha}}{\beta^2 \rho^2}}}{2} > 2$ and there exists $M > 0$ such that for every $t \in [0, +\infty)$

$$\begin{aligned} 0 \leq \frac{\rho}{2} \|x(t) - x^*\|^2 \leq g(x(t)) - g(x^*) &\leq (g(u_0) - g(x^*)) \exp(-(\underline{\gamma} - 1)t) + \frac{M}{\underline{\gamma} - 2} \exp(-t) \\ &\leq \left(g(u_0) - g(x^*) + \frac{M}{\underline{\gamma} - 2} \right) \exp(-t) \leq \left(\frac{1}{2\beta} \|u_0 - x^*\|^2 + \frac{M}{\underline{\gamma} - 2} \right) \exp(-t). \end{aligned}$$

Proof. One has for almost every $t \in [0, +\infty)$

$$\frac{d}{dt} g(x(t)) = \langle \dot{x}(t), \nabla g(x(t)) \rangle$$

and (see [16, Remark 1(b)])

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

Further, by using (19), (20) and the first equation in (32), we derive for almost every $t \in [0, +\infty)$

$$\begin{aligned} &\frac{d^2}{dt^2} \big(g(x(t)) - g(x^*) \big) + \gamma(t) \frac{d}{dt} \big(g(x(t)) - g(x^*) \big) + \alpha(t) \big(g(x(t)) - g(x^*) \big) \\ \leq\ & -\lambda(t) \|\nabla g(x(t))\|^2 + \frac{\alpha(t)}{2\beta\rho^2} \|\nabla g(x(t))\|^2 + \frac{1}{\beta} \|\dot{x}(t)\|^2 \\ =\ & -\frac{1}{2\lambda(t)} \|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \frac{\lambda(t)}{2} \|\nabla g(x(t))\|^2 + \frac{\alpha(t)}{2\beta\rho^2} \|\nabla g(x(t))\|^2 + \frac{1}{\beta} \|\dot{x}(t)\|^2. \end{aligned}$$

Taking into account (27) we obtain for almost every $t \in [0, +\infty)$

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

According to the choice of the parameters involved, we have

$$\frac{\lambda(t)}{2} - \frac{\alpha(t)}{2\beta\rho^2} \geq 0,$$

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

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

This shows that (23) in Lemma 6 for $u := \|\dot{x}(\cdot)\|^2$,

$$b_1(t) := \alpha(t),$$

$$b_2(t) := \frac{\gamma(t)}{2\lambda(t)}$$

and

$$b_3(t) := \frac{\gamma^2(t)}{2\lambda(t)} - \frac{1}{\beta}$$

is fulfilled. By combining (35) and the first condition in (25) one obtains (21), while, by combining (35) and the second condition in (25) one obtains (22).

Furthermore, by taking into account the Lipschitz property of ∇g and the strong convexity of g, it yields

$$\rho\beta \leq 1.$$

From (34), (33) and $\underline{\alpha} > 1$ we obtain

$$\frac{\lambda(t)}{\beta} \geq \underline{\alpha}\frac{1}{\beta^2\rho^2} > 1 \text{ for every } t \in [0, +\infty),$$

which combined with (35) leads to $\underline{\gamma} > 2$.

The conclusion follows from Lemma 6(ii), the strong convexity of g and (19). ■

Remark 12 In Theorem 11 one can obviously chose $\alpha(t) = \alpha$, where $\alpha = \frac{2}{\beta^2\rho^2} - 1$, if $\beta\rho < 1$, or $\alpha = 1 + \varepsilon$, with $\varepsilon > 0$, otherwise, $\lambda(t) = \lambda$ and $\gamma(t) = \gamma$ for every $t \in [0, +\infty)$, where

$$\frac{\alpha}{\beta\rho^2} \leq \lambda \leq \frac{\beta}{2}\big(\alpha + \alpha^2\big)$$

and

$$\frac{1 + \sqrt{1 + 8\frac{\lambda}{\beta}}}{2} \leq \gamma \leq 1 + \alpha.$$

Acknowledgement The authors are grateful to an anonymous referee for suggestions which improved the presentation of the manuscript.

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), 3–11, 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(1), 139–180, 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, 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, B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49(2), 574–598, 2011

[14] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011

[15] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and its Applications 119(2), 235–259, 2003

[16] 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

[17] 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

[18] R.I. Boţ, E.R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM Journal on Control and Optimization, 54(3), 1423–1443, 2016

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

[20] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematics Studies No. 5, Notas de Matemática (50), North-Holland/Elsevier, New York, 1973

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

[22] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, Dordrecht, 2004

[23] A. Pazy, *Semigroups of nonlinear contractions and their asymptotic behaviour*, in Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. III (Heriot-Watt Univ., Edinburgh), R.J. Knops (ed.), pp. 36–134, Res. Notes in Math., 30, Pitman, Boston, Mass.-London, 1979

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

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

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