Second order dynamical systems with penalty terms associated to monotone inclusions

Radu Ioan Boţ [*] Ernö Robert Csetnek [†] Szilárd Csaba László [‡]

January 8, 2018

Abstract. In this paper we investigate in a Hilbert space setting a second order dynamical system of the form

$$\ddot{x}(t)+\gamma(t)\dot{x}(t)+x(t)-J_{\lambda(t)A}\big(x(t)-\lambda(t)D(x(t))-\lambda(t)\beta(t)B(x(t))\big)=0,$$

where $A:\mathcal{H}\rightrightarrows\mathcal{H}$ is a maximal monotone operator, $J_{\lambda(t)A}:\mathcal{H}\longrightarrow\mathcal{H}$ is the resolvent operator of $\lambda(t)A$ and $D,B:\mathcal{H}\to\mathcal{H}$ are cocoercive operators, and $\lambda,\beta:[0,+\infty)\to(0,+\infty)$, and $\gamma:[0,+\infty)\to(0,+\infty)$ are step size, penalization and, respectively, damping functions, all depending on time. We show the existence and uniqueness of strong global solutions in the framework of the Cauchy-Lipschitz-Picard Theorem and prove ergodic asymptotic convergence for the generated trajectories to a zero of the operator $A+D+N_C$, where $C=\operatorname{zer}B$ and N_C denotes the normal cone operator of C. To this end we use Lyapunov analysis combined with the celebrated Opial Lemma in its ergodic continuous version. Furthermore, we show strong convergence for trajectories to the unique zero of $A+D+N_C$, provided that A is a strongly monotone operator.

Key Words. dynamical systems, Lyapunov analysis, monotone inclusions, penalty schemes

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

1 Introduction and preliminaries

Consider the bilevel optimization problem

$$\inf_{x\in\operatorname{argmin}\psi}\{f(x)+g(x)\},\tag{1}$$

where $f:\mathcal{H}\longrightarrow\mathbb{R}\cup\{+\infty\}$ is a proper convex and lower semicontinuous function, $g,\psi:\mathcal{H}\longrightarrow\mathbb{R}$ are convex and (Fréchet) differentiable functions both with Lipschitz continuous gradients, and $\operatorname{argmin}\psi$ denotes the set of global minimizers of ψ, assumed to be nonempty.

By making use of the indicator function of $\operatorname{argmin}\psi$, (1) can be rewritten as

$$\inf_{x\in\mathcal{H}}\{f(x)+g(x)+\delta_{\operatorname{argmin}\psi}(x)\}.\tag{2}$$

[*]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at. Research partially supported by FWF (Austrian Science Fund), project I 2419-N32.

[†]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.

[‡]Technical University of Cluj-Napoca, Department of Mathematics, Str. Memorandumului nr. 28, 400114 Cluj-Napoca, Romania, e-mail: laszloszilszi@yahoo.com

Obviously, $x \in \operatorname{argmin} \psi$ is an optimal solution of (2) if and only if $0 \in \partial(f + g + \delta_{\operatorname{argmin} \psi})(x)$, which can be split in

$$0 \in \partial f(x) + \nabla g(x) + \partial \delta_{\operatorname{argmin} \psi}(x), \tag{3}$$

provided a suitable qualification condition which guarantees the subdifferential sum rule holds, like for instance $0 \in \operatorname{int}(\operatorname{dom} f - \operatorname{argmin} \psi)$.

Using that $\partial \delta_{\operatorname{argmin} \psi}(x) = N_{\operatorname{argmin} \psi}(x)$ and $\operatorname{argmin} \psi = \operatorname{zer} \nabla \psi$, (3) is nothing else than

$$0 \in \partial f(x) + \nabla g(x) + N_{\operatorname{zer} \nabla \psi}(x). \tag{4}$$

This motivates us to investigate the following inclusion problem

$$0 \in Ax + Dx + N_C(x), \tag{5}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $D : \mathcal{H} \longrightarrow \mathcal{H}$ is a L_D^{-1}-cocoercive operator, and $B : \mathcal{H} \longrightarrow \mathcal{H}$ is a L_B^{-1}-cocoercive operator with $L_D, L_B > 0$, $C = \operatorname{zer} B$, assumed to be nonempty, and N_C denotes the normal cone operator of the set C. We recall that by the classical Baillon-Haddad Theorem, the gradient of a convex and (Fréchet) differentiable function is L-Lipschitz continuous, for $L > 0$, if and only if it is L^{-1}-cocoercive, see for instance [12, Corollary 18.16]).

In [17], a first order dynamical system has been assigned to the monotone inclusion (5), and it has been shown that the generated trajectories converge to a solution of it. In this paper, we assign to (5) the following second order dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + x(t) = J_{\lambda(t)A}\big(x(t) - \lambda(t)D(x(t)) - \lambda(t)\beta(t)B(x(t))\big) \\ x(0) = u_0,\, \dot{x}(0) = v_0, \end{cases} \tag{6}$$

where $u_0, v_0 \in \mathcal{H}$ and $\gamma, \lambda, \beta : [0, +\infty) \longrightarrow (0, +\infty)$. Dynamical systems governed by resolvents of maximally monotone operators have been considered in [1, 2], and then further developed in [18, 20].

The study of second order dynamical systems is motivated by the fact that the presence of the acceleration term $\ddot{x}(t)$ can lead to better convergence properties of the trajectories. Time discretizations of second order dynamical systems give usually rise to numerical algorithms with inertial terms which have been shown to have improved convergence properties (see [25]). The geometric damping function γ which acts on the velocity can in some situations accelerate the asymptotic properties of the orbits, as emphasized for example in [30].

For $B = 0$ and λ is constant, the differential equation (6) becomes the second order forward-backward dynamical system investigated in [20] in relation to the monotone inclusion problem

$$0 \in Ax + Dx.$$

On the other hand, when particularized to the monotone inclusion system (4), the differential equation (6) reads

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + x(t) = \operatorname{prox}_{\lambda(t)f}\big(x(t) - \lambda(t)\nabla g(x(t)) - \lambda(t)\beta(t)\nabla\psi(x(t))\big) \\ x(0) = u_0,\, \dot{x}(0) = v_0, \end{cases} \tag{7}$$

where we made used of the fact that the resolvent of the subdifferential of a proper, convex and lower semicontinuous function is the proximal point operator of the latter. In case $f = 0$, γ is constant and λ is also constant and identical to 1, (7) leads to the differential equation that has been investigated in [8] and [19].

The first part of the paper is devoted to the proof of the existence and uniqueness of (locally) absolutely continuous trajectories generated by the dynamical system (6); an important ingredient for this analysis is the Cauchy-Lipschitz-Picard Theorem (see [24, 29]). The proof of the convergence of the trajectories to a solution of (5) is the main result of the manuscript. Provided that a condition expressed in terms of the Fitzpatrick function of the cocoercive operator B is fulfilled, we prove weak ergodic convergence of the orbits. Furthermore, we show that, if the operator A is strongly monotone, then one obtains even strong (non-ergodic) convergence for the generated trajectories.

In the remaining of this section, we explain the notations we used up to this point and will use throughout the paper (see [15, 12, 28]).

The real Hilbert space $\mathcal{H}$ is endowed with *inner product* $\langle\cdot,\cdot\rangle$ and associated *norm* $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. The *normal cone* of of a set $S \subseteq \mathcal{H}$ is defined by $N_S(x) = \{u \in \mathcal{H} : \langle y - x, u\rangle \leq 0 \ \forall y \in S\}$, if $x \in S$ and $N_S(x) = \emptyset$ for $x \notin S$. The following characterization of the elements of the normal cone of a nonempty set by means of its support function will be used several times in the paper: for $x \in S$, $u \in N_S(x)$ if and only if $\sigma_S(u) = \langle x, u\rangle$, where $\sigma_S : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is defined by $\sigma_S(u) = \sup_{y\in S}\langle y, u\rangle$.

Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator. We denote by $\operatorname{Gr} A = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its *graph*, by $\operatorname{dom} A = \{x \in \mathcal{H} : Ax \neq \emptyset\}$ its *domain* and by $\operatorname{ran} A = \{u \in \mathcal{H} : \exists x \in \mathcal{H} \text{ s.t. } u \in Ax\}$ its *range*. The notation $\operatorname{zer} A = \{x \in \mathcal{H} : 0 \in Ax\}$ stands for the *set of zeros* of the operator A. We say that A is *monotone* if $\langle x - y, u - v\rangle \geq 0$ for all $(x, u), (y, v) \in \operatorname{Gr} A$. Further, 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 following characterization of the zeros of a maximally monotone operator will be crucial in the asymptotic analysis of (6): if A is maximally monotone, then

$$z \in \operatorname{zer} A \text{ if and only if } \langle u - z, w\rangle \geq 0 \text{ for all } (u, w) \in \operatorname{Gr} A.$$

The *resolvent* of A, $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by $p \in J_A(x)$ if and only if $x \in p + Ap$. Moreover, if A is maximally monotone, then $J_A : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (cf. [12, Proposition 23.7 and Corollary 23.10]). We will also use the *Yosida approximation* of the operator A, which is defined for $\alpha > 0$ by $A_\alpha = \frac{1}{\alpha}(\operatorname{Id} - J_{\alpha A})$, where $\operatorname{Id} : \mathcal{H} \to \mathcal{H}, \operatorname{Id}(x) = x$ for all $x \in \mathcal{H}$, is the *identity operator* on $\mathcal{H}$.

The notion of *Fitzpatrick function* associated to a monotone operator A will be important in the formulation of the condition under which the convergence of the trajectories is achieved. It is defined as

$$\varphi_A : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}},\ \varphi_A(x, u) = \sup_{(y,v)\in \operatorname{Gr} A} \{\langle x, v\rangle + \langle y, u\rangle - \langle y, v\rangle\},$$

and it is a convex and lower semicontinuous function. Introduced by Fitzpatrick in [23], this notion played in the last years a crucial role in the investigation of maximality of monotone operators by means of convex analysis specific tools (see [12, 13, 15, 28] and the references therein). We notice that, if A is maximally monotone, then φ_A is proper and

$$\varphi_A(x, u) \geq \langle x, u\rangle \ \forall (x, u) \in \mathcal{H} \times \mathcal{H},$$

with equality if and only if $(x, u) \in \operatorname{Gr} A$. We refer the reader to [13] for explicit formulae of Fitzpatrick functions associated to particular classes of monotone operators.

Let $\gamma > 0$ be arbitrary. A single-valued operator $A : \mathcal{H} \to \mathcal{H}$ is said to be *γ-cocoercive*, if $\langle x - y, Ax - Ay\rangle \geq \gamma\|Ax - Ay\|^2$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$, and *$\gamma$-Lipschitz continuous*, if $\|Ax - Ay\| \leq \gamma\|x - y\|$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$.

For a proper, convex and lower semicontinuous function $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, 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 and its resolvent is given by $J_{\gamma\partial f} = \text{prox}_{\gamma f}$ (see [12]), where $\text{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$,

$$\text{prox}_{\gamma f}(x) = \underset{y\in\mathcal{H}}{\text{argmin}} \left\{ f(y) + \frac{1}{2\gamma}\|y-x\|^2 \right\}, \tag{8}$$

denotes the proximal point operator of f.

2 Existence and uniqueness of the trajectory

We start by specifying which type of solutions are we considering in the analysis of the dynamical system (6).

Definition 1. We say that $x : [0,+\infty) \to \mathcal{H}$ is a strong global solution of (6), if the following properties are satisfied:

(i) $x, \dot{x} : [0,+\infty) \to \mathcal{H}$ are locally absolutely continuous, in other words, absolutely continuous on each interval $[0,b]$ for $0 < b < +\infty$;

(ii) $\ddot{x}(t) + \gamma(t)\dot{x}(t) + \big(x(t) - J_{\lambda(t)A}(x(t) - \lambda(t)D(x(t)) - \lambda(t)\beta(t)B(x(t)))\big) = 0$ for almost every $t \geq 0$;

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

For proving existence and uniqueness of the strong global solutions of (6), we use the Cauchy-Lipschitz-Picard Theorem for absolutely continues trajectories (see for example [24, Proposition 6.2.1], [29, Theorem 54]). The key argument is that one can rewrite (6) as a particular first order dynamical system in a suitably chosen product space (see also [3]).

To this end we make the following assumption:

$$(H1): \ \gamma, \lambda, \beta : [0,+\infty) \longrightarrow (0,+\infty) \text{ are continuous on each interval } [0,b], \text{ for } 0 < b < +\infty,$$

which also describes the framework in which we will carry out the convergence analysis in the forthcoming sections.

Theorem 2. *Suppose that γ, λ and β satisfy $(H1)$. Then for every $u_0, v_0 \in \mathcal{H}$ there exists a unique strong global solution of* (6).

Proof. Define $X : [0,+\infty) \longrightarrow \mathcal{H} \times \mathcal{H}$ as $X(t) = (x(t), \dot{x}(t))$. Then (6) is equivalent to

$$\begin{cases} \dot{X}(t) = F(t, X(t)) \\ X(0) = (u_0, v_0), \end{cases} \tag{9}$$

where $F(t,u,v) = \big(v, -\gamma(t)v - u + J_{\lambda(t)A}(u - \lambda(t)D(u) - \lambda(t)\beta(t)B(u))\big)$.

First we show that $F(t,\cdot,\cdot)$ is Lipschitz continuous with a Lipschitz constant $L(t) \in L^1_{loc}([0,+\infty))$, for every $t \geq 0$. Indeed,

$$\begin{aligned} \|F(t,u,v) - F(t,\bar{u},\bar{v})\| &= \sqrt{\|v-\bar{v}\|^2 + \|\gamma(t)(\bar{v}-v) + (\bar{u}-u) + (J_{\lambda(t)A}(s) - J_{\lambda(t)A}(\bar{s}))\|^2} \\ &\leq \sqrt{\|v-\bar{v}\|^2 + 2\|\gamma(t)(\bar{v}-v) + (\bar{u}-u)\|^2 + 2\|J_{\lambda(t)A}(s) - J_{\lambda(t)A}(\bar{s})\|^2} \\ &\leq \sqrt{(1+4\gamma^2(t))\|v-\bar{v}\|^2 + 4\|(\bar{u}-u)\|^2 + 2\|J_{\lambda(t)A}(s) - J_{\lambda(t)A}(\bar{s})\|^2}, \end{aligned}$$

where $s = u - \lambda(t)D(u) - \lambda(t)\beta(t)B(u)$ and $\bar{s} = \bar{u} - \lambda(t)D(\bar{u}) - \lambda(t)\beta(t)B(\bar{u})$.

By using the nonexpansivness of $J_{\lambda(t)A}$ we get

$$\begin{aligned}\|J_{\lambda(t)A}(s) - J_{\lambda(t)A}(\bar{s})\| &\leq \|(u-\bar{u}) + \lambda(t)(D(\bar{u}) - D(u)) + \lambda(t)\beta(t)(B(\bar{u}) - B(u)\| \\ &\leq (1 + \lambda(t)L_D + \lambda(t)\beta(t)L_B)\|u-\bar{u}\|.\end{aligned}$$

Hence,

$$\begin{aligned}\|F(t,u,v) - F(t,\bar{u},\bar{v})\| &\leq \sqrt{(1+4\gamma^2(t))\|v-\bar{v}\|^2 + (4 + 2(1+\lambda(t)L_D + \lambda(t)\beta(t)L_B)^2)\|u-\bar{u}\|^2} \\ &\leq \sqrt{5 + 4\gamma^2(t) + 2(1+\lambda(t)L_D + \lambda(t)\beta(t)L_B)^2}\sqrt{\|u-\bar{u}\|^2 + \|v-\bar{v}\|^2} \\ &\leq (\sqrt{5} + 2\gamma(t) + \sqrt{2}(1+\lambda(t)L_D + \lambda(t)\beta(t)L_B))\|(u,v)-(\bar{u},\bar{v})\|.\end{aligned}$$

Since $\gamma, \lambda, \beta \in L^1_{loc}([0,+\infty))$, it follows that

$$L(t) := \sqrt{5} + 2\gamma(t) + \sqrt{2}(1+\lambda(t)L_D + \lambda(t)\beta(t)L_B)$$

is also locally integrable on $[0,+\infty)$.

Next we show that $F(\cdot,u,v) \in L^1_{loc}([0,+\infty),\mathcal{H}\times\mathcal{H})$ for all $u,v \in \mathcal{H}$. We fix $u,v \in \mathcal{H}$ and $b > 0$, and notice that

$$\begin{aligned}\int_0^b \|F(t,u,v)\|dt &= \int_0^b \sqrt{\|v\|^2 + \|\gamma(t)v + u - J_{\lambda(t)A}(u - \lambda(t)D(u) - \lambda(t)\beta(t)B(u))\|^2}dt \\ &\leq \int_0^b \sqrt{(1+2\gamma^2(t))\|v\|^2 + 4\|u\|^2 + 4\|J_{\lambda(t)A}(u - \lambda(t)D(u) - \lambda(t)\beta(t)B(u))\|^2}dt.\end{aligned}$$

According to $(H1)$, there exist positive numbers $\underline{\lambda}$ and $\underline{\beta}$ such that $0 < \underline{\lambda} \leq \lambda(t)$ and $0 < \underline{\beta} \leq \beta(t)$ for all $t \in [0,b]$. Hence

$$\begin{aligned}&\|J_{\lambda(t)A}(u - \lambda(t)D(u) - \lambda(t)\beta(t)B(u))\| = \\ &\|J_{\lambda(t)A}(u - \lambda(t)D(u) - \lambda(t)\beta(t)B(u)) - J_{\lambda(t)A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u)) + J_{\lambda(t)A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\| \leq \\ &(\lambda(t) - \underline{\lambda})\|D(u)\| + (\lambda(t)\beta(t) - \underline{\lambda}\underline{\beta})\|B(u)\| + \|J_{\lambda(t)A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\|.\end{aligned}$$

In addition,

$$\begin{aligned}&\|J_{\lambda(t)A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\| = \\ &\|J_{\lambda(t)A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u)) - J_{\underline{\lambda}A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u)) + J_{\underline{\lambda}A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\| \leq \\ &(\lambda(t) - \underline{\lambda})\|A_{\underline{\lambda}}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\| + \|J_{\underline{\lambda}A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\|,\end{aligned}$$

where the last inequality follows from the Lipschitz property of the resolvent operator as a function of the step size, which basically follows by combining [22, Proposition 2.6] and [12, Proposition 23.28] (see also [2, Proposition 3.1]). Hence,

$$\begin{aligned}\int_0^b \|F(t,u,v)\|dt \leq &\int_0^b \Big((1+\sqrt{2}\gamma(t))\|v\| + 2\|u\| + 2(\lambda(t) - \underline{\lambda})\|D(u)\| + 2(\lambda(t)\beta(t) - \underline{\lambda}\underline{\beta})\|B(u)\|\Big)\,dt \\ &+ \int_0^b \left(2(\lambda(t) - \underline{\lambda})\|A_{\underline{\lambda}}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\| + 2\|J_{\underline{\lambda}A}(u - \underline{\lambda}D(u) - \underline{\lambda}\underline{\beta}B(u))\|\right)dt.\end{aligned}$$

Hence, $F(\cdot,u,v) \in L^1_{loc}([0,+\infty),\mathcal{H}\times\mathcal{H})$ for all $u,v \in \mathcal{H}$. The conclusion of the theorem follows by applying the Cauchy-Lipschitz-Picard theorem to the first order dynamical system (9).

□

3 Some preparatory lemmas

In this section we provide some preparatory lemmas which will be used when proving the convergence of the trajectories generated by the dynamical system (6). We start by recalling two central results; see for example [2, Lemma 5.1] and [2, Lemma 5.2], respectively.

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

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

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

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

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

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

Lemma 5. *Suppose that $(H1)$ holds and let x be the unique strong global solution of (6). Take $(x^*, w) \in \operatorname{Gr}(A + D + N_C)$ such that $w = v + Dx^* + p$, where $v \in Ax^*$ and $p \in N_C(x^*)$. For every $t \geq 0$ consider the function $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$. Then the following inequality holds for almost every $t \geq 0$:*

$$\begin{aligned}&\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t)\left(\frac{1}{L_D} - \lambda(t)\right)\|D(x(t)) - Dx^*\|^2 - \|\dot{x}(t)\|^2 \leq \\ &\lambda(t)\beta(t)\left(\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right) + \lambda^2(t)\|Dx^* + v\|^2 + \lambda(t)\langle w, x^* - x(t)\rangle + \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2.\end{aligned} \tag{10}$$

Proof. We have $\dot{h}(t) = \langle \dot{x}(t), x(t) - x^*\rangle$ and $\ddot{h}(t) = \langle \ddot{x}(t), x(t) - x^*\rangle + \|\dot{x}(t)\|^2$ for every $t \geq 0$. By using the definition of the resolvent, the differential equation in (6) can be written for almost every $t \geq 0$ as

$$x(t) - \lambda(t)D(x(t)) - \lambda(t)\beta(t)B(x(t)) \in \ddot{x}(t) + \gamma(t)\dot{x}(t) + x(t) + \lambda(t)A(\ddot{x}(t) + \gamma(t)\dot{x}(t) + x(t))$$

or, equivalently,

$$-\frac{1}{\lambda(t)}\ddot{x}(t) - \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) - D(x(t)) - \beta(t)B(x(t)) \in A(\ddot{x}(t) + \gamma(t)\dot{x}(t) + x(t)). \tag{11}$$

Since $v \in Ax^*$ and A is monotone, we get for almost every $t \geq 0$

$$\left\langle v + \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + D(x(t)) + \beta(t)B(x(t)), x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\right\rangle \geq 0.$$

It follows that

$$\begin{aligned}&\lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + v, x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\rangle \geq \\ &\langle \ddot{x}(t) + \gamma(t)\dot{x}(t), -x^* + \ddot{x}(t) + \gamma(t)\dot{x}(t) + x(t)\rangle = \ddot{h}(t) + \gamma(t)\dot{h}(t) + \|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \|\dot{x}(t)\|^2\end{aligned} \tag{12}$$

for almost every $t \geq 0$. Hence, for almost every $t \geq 0$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) - \|\dot{x}(t)\|^2 \leq \lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + v, x^* - x(t)\rangle$$
$$+\lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + v, -\ddot{x}(t) - \gamma(t)\dot{x}(t)\rangle - \|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 \leq$$
$$\lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + v, x^* - x(t)\rangle + \frac{\lambda^2(t)}{4}\|D(x(t)) + \beta(t)B(x(t)) + v\|^2,$$

and from here, by using mean inequalities,

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) - \|\dot{x}(t)\|^2 \leq \lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + v, x^* - x(t)\rangle +$$
$$\frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 + \lambda^2(t)\|Dx^* + v\|^2 + \lambda^2(t)\|D(x(t)) - Dx^*\|^2.$$

Since $v = w - Dx^* - p$, we obtain for the first summand of the term on the right-hand side of the above inequality for every $t \geq 0$ the following estimate

$$\lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + v, x^* - x(t)\rangle =$$
$$\lambda(t)\langle D(x(t)) + \beta(t)B(x(t)) + w - Dx^* - p, x^* - x(t)\rangle =$$
$$\lambda(t)\langle D(x(t)) - Dx^*, x^* - x(t)\rangle + \lambda(t)\langle w, x^* - x(t)\rangle +$$
$$\lambda(t)\beta(t)\left[\langle B(x(t)), x^*\rangle + \left\langle \frac{p}{\beta(t)}, x(t)\right\rangle - \langle B(x(t)), x(t)\rangle - \left\langle \frac{p}{\beta(t)}, x^*\right\rangle\right] \leq$$
$$-\frac{\lambda(t)}{L_D}\|D(x(t)) - Dx^*\|^2 + \lambda(t)\langle w, x^* - x(t)\rangle + \lambda(t)\beta(t)\left(\sup_{u \in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right).$$

Hence, for almost every $t \geq 0$, we have

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) - \|\dot{x}(t)\|^2 \leq \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 + \lambda^2(t)\|Dx^* + v\|^2 + \lambda^2(t)\|D(x(t)) - Dx^*\|^2$$
$$- \frac{\lambda(t)}{L_D}\|D(x(t)) - Dx^*\|^2 + \lambda(t)\langle w, x^* - x(t)\rangle$$
$$+ \lambda(t)\beta(t)\left(\sup_{u \in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right),$$

which is nothing else than the desired conclusion. □

Lemma 6. *Suppose that $(H1)$ holds and let x be the unique strong global solution of (6). Take $x^* \in C \cap \operatorname{dom} A$ and $v \in Ax^*$. For every $t \geq 0$ consider the function $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$. Then for every $\epsilon > 0$ the following inequality holds for almost every $t \geq 0$:*

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{1 + 2\epsilon}{2 + 2\epsilon}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \|\dot{x}(t)\|^2 + \frac{\epsilon\lambda(t)\beta(t)}{1 + \epsilon}\langle B(x(t)), x(t) - x^*\rangle \leq$$
$$\lambda(t)\beta(t)\left(\frac{1 + \epsilon}{2}\lambda(t)\beta(t) - \frac{1}{(1 + \epsilon)L_B}\right)\|B(x(t))\|^2 + \lambda(t)\langle D(x(t)) + v, x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\rangle. \quad (13)$$

Proof. Let be $\epsilon > 0$ fixed. According to (12) in the proof of the above lemma, we have for almost every $t \geq 0$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \|\dot{x}(t)\|^2 \leq \lambda(t)\beta(t)\langle B(x(t)), x^* - x(t)\rangle +$$
$$\lambda(t)\beta(t)\langle B(x(t)), -\ddot{x}(t) - \gamma(t)\dot{x}(t)\rangle +$$
$$\lambda(t)\langle D(x(t)) + v, x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\rangle.$$

Since B is $\frac{1}{L_B}$-cocoercive and $Bx^*=0$ we have $\langle B(x(t)), x^*-x(t)\rangle \le -\frac{1}{L_B}\|B(x(t))\|^2$, hence

$$\lambda(t)\beta(t)\langle B(x(t)), x^*-x(t)\rangle \le -\frac{\lambda(t)\beta(t)}{(1+\epsilon)L_B}\|B(x(t))\|^2 + \frac{\epsilon}{1+\epsilon}\lambda(t)\beta(t)\langle B(x(t)), x^*-x(t)\rangle,$$

for every $t \ge 0$. Consequently,

$$\ddot{h}(t)+\gamma(t)\dot{h}(t)+\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2-\|\dot{x}(t)\|^2+\frac{\epsilon\lambda(t)\beta(t)}{1+\epsilon}\langle B(x(t)), x(t)-x^*\rangle \le -\frac{\lambda(t)\beta(t)}{(1+\epsilon)L_B}\|B(x(t))\|^2$$
$$+\lambda(t)\beta(t)\langle B(x(t)), -\ddot{x}(t)-\gamma(t)\dot{x}(t)\rangle+\lambda(t)\langle D(x(t))+v, x^*-\ddot{x}(t)-\gamma(t)\dot{x}(t)-x(t)\rangle,$$

which, combined with

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

implies for almost every $t \ge 0$ relation (13). $\square$

Lemma 7. *Suppose that* $(H1)$ *holds and let* x *be the unique strong global solution of* (6)*. Furthermore, suppose that* $\limsup_{t\longrightarrow+\infty}\lambda(t)\beta(t) < \frac{1}{L_B}$*. Take* $x^* \in C \cap \operatorname{dom} A$ *and* $v \in Ax^*$*. For every* $t \ge 0$ *consider the function* $h(t)=\frac{1}{2}\|x(t)-x^*\|^2$*. Then there exist* $a,b,c>0$ *and* $t_0>0$ *such that for almost every* $t \ge t_0$ *the following inequality holds:*

$$\ddot{h}(t)+\gamma(t)\dot{h}(t)+c\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2+a\lambda(t)\beta(t)\Big(\langle B(x(t)), x(t)-x^*\rangle+\|B(x(t))\|^2\Big) \le$$
$$\left(b\lambda^2(t)-\frac{\lambda(t)}{L_D}\right)\|D(x(t))-Dx^*\|^2+\lambda(t)\langle Dx^*+v, x^*-x(t)\rangle+b\lambda^2(t)\|Dx^*+v\|^2+\|\dot{x}(t)\|^2. \tag{14}$$

Proof. Let be $\epsilon>0$. According to the previous lemma, (13) holds for almost every $t\ge 0$. We estimate the last summand in the right-hand side of (13) by using the mean inequality and the cocoercieveness of D. For every $t \ge 0$ we obtain

$$\lambda(t)\langle D(x(t))+v, x^*-\ddot{x}(t)-\gamma(t)\dot{x}(t)-x(t)\rangle \le$$
$$\frac{\epsilon}{4(1+\epsilon)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2+\frac{\lambda^2(t)(1+\epsilon)}{\epsilon}\|D(x(t))+v\|^2+\lambda(t)\langle D(x(t))+v, x^*-x(t)\rangle =$$
$$\frac{\epsilon}{4(1+\epsilon)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2+\frac{\lambda^2(t)(1+\epsilon)}{\epsilon}\|D(x(t))+v\|^2+$$
$$\lambda(t)\langle D(x(t))-Dx^*, x^*-x(t)\rangle+\lambda(t)\langle Dx^*+v, x^*-x(t)\rangle \le$$
$$\frac{\epsilon}{4(1+\epsilon)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2+\frac{\lambda^2(t)(1+\epsilon)}{\epsilon}\|D(x(t))+v\|^2+$$
$$-\frac{\lambda(t)}{L_D}\|D(x(t))-Dx^*\|^2+\lambda(t)\langle Dx^*+v, x^*-x(t)\rangle,$$

which, combined with $\|D(x(t))+v\|^2 \le 2\|D(x(t))-Dx^*\|^2+2\|Dx^*+v\|^2$, implies

$$\lambda(t)\langle D(x(t))+v, x^*-\ddot{x}(t)-\gamma(t)\dot{x}(t)-x(t)\rangle \le \frac{\epsilon}{4(1+\epsilon)}\|\ddot{x}(t)+\gamma(t)\dot{x}(t)\|^2+$$
$$\left(\frac{2\lambda^2(t)(1+\epsilon)}{\epsilon}-\frac{\lambda(t)}{L_D}\right)\|D(x(t))-Dx^*\|^2+\frac{2\lambda^2(t)(1+\epsilon)}{\epsilon}\|Dx^*+v\|^2+\lambda(t)\langle Dx^*+v, x^*-x(t)\rangle.$$

Using the above estimate in (13), we obtain for almost every $t \geq 0$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{1+2\epsilon}{2+2\epsilon}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \|\dot{x}(t)\|^2 + \frac{\epsilon\lambda(t)\beta(t)}{1+\epsilon}\langle B(x(t)), x(t) - x^*\rangle \leq$$
$$\lambda(t)\beta(t)\left(\frac{1+\epsilon}{2}\lambda(t)\beta(t) - \frac{1}{(1+\epsilon)L_B}\right)\|B(x(t))\|^2 + \frac{\epsilon}{4(1+\epsilon)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 +$$
$$\left(\frac{2\lambda^2(t)(1+\epsilon)}{\epsilon} - \frac{\lambda(t)}{L_D}\right)\|D(x(t)) - Dx^*\|^2 + \frac{2\lambda^2(t)(1+\epsilon)}{\epsilon}\|Dx^* + v\|^2 + \lambda(t)\langle Dx^* + v, x^* - x(t)\rangle$$

or, equivalently

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{2+3\epsilon}{4(1+\epsilon)}\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \|\dot{x}(t)\|^2 + \frac{\epsilon\lambda(t)\beta(t)}{1+\epsilon}\left(\langle B(x(t)), x(t) - x^*\rangle + \|B(x(t))\|^2\right) \leq$$
$$\lambda(t)\beta(t)\left(\frac{1+\epsilon}{2}\lambda(t)\beta(t) - \frac{1}{(1+\epsilon)L_B} + \frac{\epsilon}{1+\epsilon}\right)\|B(x(t))\|^2 + \left(\frac{2\lambda^2(t)(1+\epsilon)}{\epsilon} - \frac{\lambda(t)}{L_D}\right)\|D(x(t)) - Dx^*\|^2$$
$$+ \frac{2\lambda^2(t)(1+\epsilon)}{\epsilon}\|Dx^* + v\|^2 + \lambda(t)\langle Dx^* + v, x^* - x(t)\rangle.$$

Since $\limsup_{t\longrightarrow+\infty}\lambda(t)\beta(t) < \frac{1}{L_B}$, there exists $t_0 > 0$ such that

$$\frac{1+\epsilon}{2}\lambda(t)\beta(t) - \frac{1}{(1+\epsilon)L_B} + \frac{\epsilon}{1+\epsilon} < \frac{1+\epsilon}{2L_B} - \frac{1}{(1+\epsilon)L_B} + \frac{\epsilon}{1+\epsilon}$$

for every $t \geq t_0$. Further, we notice that

$$\frac{1+\epsilon}{2L_B} - \frac{1}{(1+\epsilon)L_B} + \frac{\epsilon}{1+\epsilon} \leq 0$$

for every $\epsilon \in \left(0, \sqrt{(1+L_B)^2+1} - (1+L_B)\right]$. By choosing ϵ_0 from this interval and defining

$$a := \frac{\epsilon_0}{1+\epsilon_0}, b := \frac{2(1+\epsilon_0)}{\epsilon_0} \text{ and } c := \frac{2+3\epsilon_0}{4(1+\epsilon_0)},$$

the conclusion follows. □

Remark 8. In the proof of the above theorem, the choice $\epsilon_0 \leq \sqrt{(1+L_B)^2+1} - (1+L_B) < \sqrt{2} - 1$, implies that $a < 1 - \frac{1}{\sqrt{2}}$ and $\frac{1}{2} < c < \frac{3}{4} - \frac{\sqrt{2}}{8}$.

Lemma 9. *Suppose that $(H1)$ holds and let x be the unique strong global solution of* (6). *Furthermore, suppose that $\limsup_{t\longrightarrow+\infty}\lambda(t)\beta(t) < \frac{1}{L_B}$ and $\lim_{t\longrightarrow+\infty}\lambda(t) = 0$. Take $(x^*, w) \in \operatorname{Gr}(A + D + N_C)$ such that $w = v + Dx^* + p$, where $v \in Ax^*$ and $p \in N_C(x^*)$. For every $t \geq 0$ consider the function $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$. Then there exist $a, b, c > 0$ and $t_1 > 0$ such that for almost every $t \geq t_1$ the following inequality holds:*

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + c\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + a\lambda(t)\beta(t)\left(\frac{1}{2}\langle B(x(t)), x(t) - x^*\rangle + \|B(x(t))\|^2\right) \leq$$
$$\frac{a\lambda(t)\beta(t)}{2}\left(\sup_{u\in C}\varphi_B\left(u, \frac{2p}{a\beta(t)}\right) - \sigma_C\left(\frac{2p}{a\beta(t)}\right)\right) + b\lambda^2(t)\|Dx^* + v\|^2 + \lambda(t)\langle w, x^* - x(t)\rangle + \|\dot{x}(t)\|^2. \tag{15}$$

Proof. According to Lemma 7, there exist $a, b, c > 0$ and $t_0 > 0$ such that for almost every $t \geq t_0$ the inequality (14) holds. Since $\lim_{t \longrightarrow +\infty} \lambda(t) = 0$, there exists $t_1 \geq t_0$ such that $\lambda(t) \leq \frac{1}{bL_D}$, hence $b\lambda^2(t) - \frac{\lambda(t)}{L_D} \leq 0$ for every $t \geq t_1$. Consequently, we can omit for every $t \geq t_1$ the term $\left(b\lambda^2(t) - \frac{\lambda(t)}{L_D}\right) \|D(x(t)) - Dx^*\|^2$ in (14) and obtain that the inequality

$$\begin{aligned}&\ddot{h}(t) + \gamma(t)\dot{h}(t) + c\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + a\lambda(t)\beta(t)\Big(\langle B(x(t)), x(t) - x^*\rangle + \|B(x(t))\|^2\Big) \leq \\ &\lambda(t)\langle Dx^* + v, x^* - x(t)\rangle + b\lambda^2(t)\|Dx^* + v\|^2 + \|\dot{x}(t)\|^2 \end{aligned} \tag{16}$$

holds for almost every $t \geq t_1$.

Since $Dx^* + v = w - p$, we have for every $t \geq 0$

$$\begin{aligned}&\frac{a\lambda(t)\beta(t)}{2}\langle B(x(t)), x^* - x(t)\rangle + \lambda(t)\langle Dx^* + v, x^* - x(t)\rangle = \\ &\frac{a\lambda(t)\beta(t)}{2}\left(\langle B(x(t)), x^*\rangle + \left\langle \frac{2p}{a\beta(t)}, x(t)\right\rangle - \langle B(x(t)), x(t)\rangle - \left\langle \frac{2p}{a\beta(t)}, x^*\right\rangle\right) + \lambda(t)\langle w, x^* - x(t)\rangle \leq \\ &\frac{a\lambda(t)\beta(t)}{2}\left(\sup_{u\in C}\varphi_B\left(u, \frac{2p}{a\beta(t)}\right) - \sigma_C\left(\frac{2p}{a\beta(t)}\right)\right) + \lambda(t)\langle w, x^* - x(t)\rangle.\end{aligned}$$

On the other hand, (16) can be equivalently written for almost every $t \geq t_1$ as

$$\begin{aligned}&\ddot{h}(t) + \gamma(t)\dot{h}(t) + c\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + a\lambda(t)\beta(t)\Big(\frac{1}{2}\langle B(x(t)), x(t) - x^*\rangle + \|B(x(t))\|^2\Big) \leq \\ &\frac{a\lambda(t)\beta(t)}{2}\langle B(x(t)), x^* - x(t)\rangle + \lambda(t)\langle Dx^* + v, x^* - x(t)\rangle + b\lambda^2(t)\|Dx^* + v\|^2 + \|\dot{x}(t)\|^2,\end{aligned}$$

hence

$$\begin{aligned}&\ddot{h}(t) + \gamma(t)\dot{h}(t) + c\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + a\lambda(t)\beta(t)\left(\frac{1}{2}\langle B(x(t)), x(t) - x^*\rangle + \|B(x(t))\|^2\right) \leq \\ &\frac{a\lambda(t)\beta(t)}{2}\left(\sup_{u\in C}\varphi_B\left(u, \frac{2p}{a\beta(t)}\right) - \sigma_C\left(\frac{2p}{a\beta(t)}\right)\right) + b\lambda^2(t)\|Dx^* + v\|^2 + \lambda(t)\langle w, x^* - x(t)\rangle + \|\dot{x}(t)\|^2.\end{aligned}$$

□

4 Main result: the convergence of the trajectories

For the proof of the convergence of the trajectories generated by the dynamical system (6) we will utilize the following assumptions:

$(H2)$: $A + N_C$ is maximally monotone and $\operatorname{zer}(A + D + N_C) \neq \emptyset$;

$(H3)$: $\lambda \in L^2([0, +\infty)) \setminus L^1([0, +\infty))$ and $\lim_{t\longrightarrow+\infty} \lambda(t) = 0$;

(H_{fitz}) : For every $p \in \operatorname{ran} N_C$, $\int_0^{+\infty} \lambda(t)\beta(t)\left(\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right) dt < +\infty$.

Remark 10. (i) The first assumption in $(H2)$ is fulfilled when a regularity condition which ensures the maximality of the sum of two maximally monotone operators holds. This is a widely studied topic in the literature; we refer the reader to [12, 14, 15, 28] for such conditions, including the classical Rockafellar's condition expressed in terms of the domains of the involved operators.

(ii) With respect to (H_{fitz}), we would like to remind that a similar condition formulated in terms of the Fitzpatrick function has been considered for the first time in [16] in the discrete setting. Its continuous version has been introduced in [17] and further used also in [5].

This class of conditions, widely used in the context of penalization approaches, has its origin in [7]. Here, in the particular case $C = \operatorname{argmin} \psi$, where $\psi : \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with Lipschitz continuous gradient and such that $\min \psi = 0$, the condition

(H): For every $p \in \operatorname{ran} N_C$, $\int_0^{+\infty} \lambda(t)\beta(t) \left[\psi^*\left(\frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right] dt < +\infty$.

has been used in the asymptotic analysis of a coupled dynamical system with multiscale aspects. The function $\psi^* : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, $\psi^*(u) = \sup_{x \in \mathcal{H}}\{\langle u, x\rangle - \psi(x)\}$, denotes the Fenchel conjugate of ψ.

According to [13], it holds

$$\varphi_{\nabla\psi}(x, u) \leq \psi(x) + \psi^*(u) \ \forall (x, u) \in \mathcal{H} \times \mathcal{H}. \tag{17}$$

Since $\psi(x) = 0$ for $x \in C$, condition (H_{fitz}) applied to $B = \nabla\psi$ is fulfilled, provided that (H) is fulfilled. For several particular situations where (H_{fitz}) is verified (in its continuous or discrete version) we refer the reader to [7, 10, 9, 11, 27, 26].

For proving the convergence of the trajectories generated by the dynamical system (6) we will also make use of the following ergodic version of the continuous Opial Lemma (see [7, Lemma 2.3]).

Lemma 11. *Let $S \subseteq \mathcal{H}$ be a nonempty set, $x : [0, +\infty) \to \mathcal{H}$ a given map and $\lambda : [0, +\infty) \to (0, +\infty)$ such that $\int_0^{+\infty} \lambda(t) = +\infty$. Define $\tilde{x} : [0, +\infty) \to \mathcal{H}$ by*

$$\tilde{x}(t) = \frac{1}{\int_0^t \lambda(s)ds} \int_0^t \lambda(s)x(s)ds.$$

Assume that

(i) for every $z \in S$, $\lim_{t\to+\infty} \|x(t) - z\|$ exists;

(ii) every weak sequential cluster point of the map $\tilde{x}$ belongs to S.

Then there exists $x_\infty \in S$ such that $w - \lim_{t\to+\infty} \tilde{x}(t) = x_\infty$.

We can state now the main result of this paper.

Theorem 12. *Suppose that $(H1) - (H3)$ and (H_{fitz}) hold, and let x be the unique strong global solution of (6). Furthermore, suppose that $\limsup_{t\longrightarrow+\infty} \lambda(t)\beta(t) < \frac{1}{L_B}$, γ is locally absolutely continuous and for almost every $t \geq 0$ it holds $\gamma(t) \geq \sqrt{2}$ and $\dot{\gamma}(t) \leq 0$. Let $\tilde{x} : [0, +\infty) \longrightarrow \mathcal{H}$ be defined by*

$$\tilde{x}(t) = \frac{1}{\int_0^t \lambda(s)ds} \int_0^t \lambda(s)x(s)ds.$$

Then the following statements hold:

(i) for every $x^* \in \operatorname{zer}(A + D + N_C)$, $\|x(t) - x^*\|$ converges as $t \longrightarrow +\infty$; in addition, $\dot{x}, \ddot{x} \in L^2([0, +\infty), \mathcal{H})$, $\lambda(\cdot)\beta(\cdot)\|B(x(\cdot))\|^2 \in L^1([0, +\infty))$, $\int_0^{+\infty} \lambda(t)\beta(t)\langle B(x(t)), x(t) - x^*\rangle dt < +\infty$, and $\lim_{t\to+\infty} \dot{x}(t) = \lim_{t\to+\infty} \dot{h}(t) = 0$, where $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$;

(ii) $\tilde{x}(t)$ converges weakly as $t \longrightarrow +\infty$ to an element in $\operatorname{zer}(A + D + N_C)$;

(iii) if, additionally, A is strongly monotone, then $x(t)$ converges strongly as $t \longrightarrow +\infty$ to the unique element of $\operatorname{zer}(A + D + N_C)$.

Proof. (i) Let be $x^* \in \operatorname{zer}(A+D+N_C)$, thus $(x^*,0) \in \operatorname{Gr}(A+D+N_C)$ and $0 = v + Dx^* + p$ for $v \in Ax^*$ and $p \in N_C(x^*)$. According to Lemma 9 and Remark 8, there exist $a,b,c>0$, with $a < 1 - \frac{1}{\sqrt{2}}$ and $\frac{1}{2} < c < \frac{3}{4} - \frac{\sqrt{2}}{8}$, and $t_1 > 0$ such that for almost every $t \geq t_1$ it holds

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + c\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + a\lambda(t)\beta(t)\left(\frac{1}{2}\langle B(x(t)), x(t)-x^*\rangle + \|B(x(t))\|^2\right) \leq$$
$$\frac{a\lambda(t)\beta(t)}{2}\left(\sup_{u\in C}\varphi_B\left(u, \frac{2p}{a\beta(t)}\right) - \sigma_C\left(\frac{2p}{a\beta(t)}\right)\right) + b\lambda^2(t)\|Dx^*+v\|^2 + \|\dot{x}(t)\|^2.$$

On the other hand, since

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

it holds

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + c\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 - \|\dot{x}(t)\|^2 \geq$$
$$\frac{d}{dt}\left(\dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2\right) + (c\gamma^2(t) - c\dot{\gamma}(t) - 1)\|\dot{x}(t)\|^2 + c\|\ddot{x}(t)\|^2$$

for every $t \geq 0$.

By combining these two inequalities, we obtain for almost every $t \geq t_1$

$$\frac{d}{dt}\left(\dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2\right) + (c\gamma^2(t) - c\dot{\gamma}(t) - 1)\|\dot{x}(t)\|^2 + c\|\ddot{x}(t)\|^2 +$$
$$a\frac{\lambda(t)\beta(t)}{2}\langle B(x(t)), x(t)-x^*\rangle + a\lambda(t)\beta(t)\|B(x(t))\|^2 \leq$$
$$\frac{a\lambda(t)\beta(t)}{2}\left(\sup_{u\in C}\varphi_B\left(u, \frac{2p}{a\beta(t)}\right) - \sigma_C\left(\frac{2p}{a\beta(t)}\right)\right) + b\lambda^2(t)\|Dx^*+v\|^2. \tag{18}$$

Since $\gamma(t) \geq \sqrt{2}$ and $c > \frac{1}{2}$ one has $c\gamma^2(t) - c\dot{\gamma}(t) - 1 \geq 2c - 1 > 0$ for almost every $t \geq 0$. By using that $\langle B(x(t)), x(t)-x^*\rangle \geq \frac{1}{L_B}\|B(x(t))\|^2$ for every $t \geq 0$ and by neglecting the non-negative terms on the left-hand side of (18), we get for almost every $t \geq t_1$

$$\frac{d}{dt}\left(\dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2\right) \leq$$
$$\frac{a\lambda(t)\beta(t)}{2}\left(\sup_{u\in C}\varphi_B\left(u, \frac{2p}{a\beta(t)}\right) - \sigma_C\left(\frac{2p}{a\beta(t)}\right)\right) + b\lambda^2(t)\|Dx^*+v\|^2.$$

Further, by integration, we easily derive that there exists $M > 0$ such that for every $t \geq 0$

$$\dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2 \leq M. \tag{19}$$

Hence, $\dot{h}(t) + \gamma(t)h(t) \leq M$, which leads to $\dot{h}(t) + \sqrt{2}h(t) \leq M$ for all $t \geq 0$. Consequently, $\frac{d}{dt}(h(t)e^{\sqrt{2}t}) \leq Me^{\sqrt{2}t}$; therefore, by integrating this inequality from 0 to $T > 0$, one obtains

$$h(T) \leq \frac{M}{\sqrt{2}} - \frac{M}{\sqrt{2}}e^{\sqrt{2}(-T)} + h(0)e^{\sqrt{2}(-T)},$$

which shows that h is bounded, hence x is bounded. Combining this with

$$\langle \dot{x}(t), x(t)-x^*\rangle + c\sqrt{2}\|\dot{x}(t)\|^2 \leq M \ \forall t \geq 0,$$

which is a consequence of (19), we derive that $\dot{x}$ is bounded, too.

In conclusion, $t \mapsto \dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2$ is bounded from below. By taking into account relation (18) and applying Lemma 3, we obtain

$$\lim_{t \longrightarrow +\infty} \left(\dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2\right) \in \mathbb{R} \tag{20}$$

and

$$\int_0^{+\infty} \|\dot{x}(t)\|^2 dt, \int_0^{+\infty} \|\ddot{x}(t)\|^2 dt, \int_0^{+\infty} \lambda(t)\beta(t)\langle B(x(t)), x(t) - x^*\rangle dt, \int_0^{+\infty} \lambda(t)\beta(t)\|B(x(t))\|^2 \in \mathbb{R}.$$

Since

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

for every $t \geq 0$ and the function on the right-hand side of the above inequality belongs to $L^1([0,+\infty))$, according to Lemma 4 one has $\lim_{t\to+\infty} \dot{x}(t) = 0$. Further, the equality $\dot{h}(t) = \langle \dot{x}(t), x(t) - x^*\rangle$ leads to $-\|\dot{x}(t)\|\|x(t) - x^*\| \leq \dot{h}(t) \leq \|\dot{x}(t)\|\|x(t) - x^*\|$ for every $t \geq 0$. Since $\lim_{t\to+\infty} \dot{x}(t) = 0$ and $\|x(\cdot) - x^*\|$ is bounded, one obtains $\lim_{t\to+\infty} \dot{h}(t) = 0$.

From $\lim_{t\longrightarrow+\infty}(\dot{h}(t) + \gamma(t)h(t) + c\gamma(t)\|\dot{x}(t)\|^2) \in \mathbb{R}$, $\lim_{t\longrightarrow+\infty} \dot{h}(t) = 0$ and $\lim_{t\longrightarrow+\infty} c\gamma(t)\|\dot{x}(t)\|^2 = 0$, one obtains that the limit $\lim_{t\longrightarrow+\infty} \gamma(t)h(t)$ exists and it is a finite number. On the other hand, since the limit $\lim_{t\longrightarrow+\infty} \gamma(t) \geq \sqrt{2}$ exists and it is a positive number, one can conclude that $\lim_{t\longrightarrow+\infty} h(t)$ exists and it is finite. Consequently, $\|x(t) - x^*\|$ converges as $t \longrightarrow +\infty$.

(ii) We show that every weak sequential limit point of $\tilde{x}$ belongs to $\operatorname{zer}(A + D + N_C)$. Indeed, let x_0 be a weak sequential limit point of $\tilde{x}$; thus, there exists a sequence $(s_n)_{n\geq 0}$ with $s_n \longrightarrow +\infty$ and $\tilde{x}(s_n) \longrightarrow x_0$ as $n \longrightarrow +\infty$.

Take an arbitrary $(x^*, w) \in \operatorname{Gr}(A + D + N_C)$ with $w = v + Dx^* + p$, $v \in Ax^*$ and $p \in N_C(x^*)$. Since $\lim_{t\longrightarrow+\infty} \lambda(t) = 0$, there exists $t_2 > 0$ such that for every $t \geq t_2$ one has $\lambda(t)\left(\frac{1}{L_D} - \lambda(t)\right) \geq 0$. From (10) we obtain

$$\begin{aligned}\ddot{h}(t) + \gamma(t)\dot{h}(t) \leq{}& \\ \lambda(t)\beta(t)\left(\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right) &+ \lambda^2(t)\|Dx^* + v\|^2 + \\ \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 &+ \|\dot{x}(t)\|^2 + \lambda(t)\langle w, x^* - x(t)\rangle.\end{aligned}$$

for every $t \geq t_2$. By integrating from t_2 to $T > t_2$, we get from here

$$\int_{t_2}^T \ddot{h}(t) + \gamma(t)\dot{h}(t)dt \leq L + \left\langle w, \left(\int_{t_2}^T \lambda(t)dt\right)x^* - \int_{t_2}^T \lambda(t)x(t)dt\right\rangle,$$

where

$$\begin{aligned}L :={}& \int_{t_2}^T \lambda(t)\beta(t)\left(\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right)dt \\ &+ \int_{t_2}^T \left(\lambda^2(t)\|Dx^* + v\|^2 + \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 + \|\dot{x}(t)\|^2\right)dt.\end{aligned}$$

Since $\gamma(t)\dot{h}(t) \geq \frac{d}{dt}(\gamma(t)h(t))$ and $\gamma(T)h(T) \geq 0$, we obtain

$$-\gamma(t_2)h(t_2) \leq L + \left\langle w, \left(\int_{t_2}^{T} \lambda(t)dt\right) x^* - \int_{t_2}^{T} \lambda(t)x(t)dt \right\rangle - \dot{h}(T) + \dot{h}(t_2),$$

hence

$$\frac{-\gamma(t_2)h(t_2)}{\int_0^T \lambda(t)dt} \leq \frac{L_1 - \dot{h}(T)}{\int_0^T \lambda(t)dt} + \left\langle w, x^* - \frac{\int_0^T \lambda(t)x(t)dt}{\int_0^T \lambda(t)dt} \right\rangle,$$

where $L_1 := L + \dot{h}(t_2) + \left\langle w, \int_0^{t_2} \lambda(t)x(t)dt - \left(\int_0^{t_2} \lambda(t)dt\right) x^* \right\rangle \in \mathbb{R}$.

We choose in the above inequality $T = s_n$ for those n for which $s_n > t_2$, let n converge to $+\infty$ and so, by taking into account that $\int_0^{+\infty} \lambda(t)dt = +\infty$ and $\dot{h}$ is bounded, we obtain

$$\langle w, x^* - x_0 \rangle \geq 0.$$

Since $(x^*, w) \in \mathrm{Gr}(A + D + N_C)$ was arbitrary chosen, it follows that $x_0 \in \mathrm{zer}(A + D + N_C)$. Hence, by Lemma 11, $\tilde{x}(t)$ converges weakly as $t \longrightarrow +\infty$ to an element in $\mathrm{zer}(A + D + N_C)$.

(iii) Assume now that A is strongly monotone, i.e. there exists $\eta > 0$ such that

$$\langle u^* - v^*, x - y \rangle \geq \eta\|x - y\|^2, \text{ for all } (x, u^*), (y, v^*) \in \mathrm{Gr}(A).$$

Let x^* be the unique element of $\mathrm{zer}(A + D + N_C)$. Then $0 = v + Dx^* + p$, for $v \in Ax^*$ and $p \in N_C(x^*)$.

Since $v \in Ax^*$ and A is $\eta-$strongly monotone, from (11) we obtain for almost every $t \geq 0$

$$\left\langle v + \frac{1}{\lambda(t)}\ddot{x}(t) + \frac{\gamma(t)}{\lambda(t)}\dot{x}(t) + D(x(t)) + \beta(t)B(x(t)), x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t) \right\rangle \geq \\ \eta\|x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\|^2.$$

By repeating the arguments in the proof of Lemma 5, we easily derive for almost every $t \geq 0$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) - \|\dot{x}(t)\|^2 + \eta\lambda(t)\|x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\|^2 \leq \\ \left(\lambda^2(t) - \frac{\lambda(t)}{L_D}\right)\|D(x(t)) - Dx^*\|^2 + \lambda(t)\beta(t)\left(\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right) + \\ \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 + \lambda^2(t)\|v + Dx^*\|^2.$$

Since $\lim_{t\longrightarrow+\infty} \lambda(t) = 0$, there exists $t_3 > 0$ such that $\lambda^2(t) - \frac{\lambda(t)}{L_D} \leq 0$ for every $t \geq t_3$. Thus for almost every $t \geq t_3$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \eta\lambda(t)\|x^* - \ddot{x}(t) - \gamma(t)\dot{x}(t) - x(t)\|^2 \leq \\ \lambda(t)\beta(t)\left(\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right) + \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 + \lambda^2(t)\|v + Dx^*\|^2 + \|\dot{x}(t)\|^2.$$

Combining this inequality with

$$\|x^* - x(t)\|^2 \leq 2\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 + 2\|x^* - x(t) - \ddot{x}(t) - \gamma(t)\dot{x}(t)\|^2,$$

we derive for almost every $t \geq t_3$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \frac{\eta\lambda(t)}{2}\|x^* - x(t)\|^2 \leq$$
$$\eta\lambda(t)\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 +$$
$$\lambda(t)\beta(t)\left(\sup_{u\in C}\varphi_B\left(u, \frac{p}{\beta(t)}\right) - \sigma_C\left(\frac{p}{\beta(t)}\right)\right) + \frac{\lambda^2(t)\beta^2(t)}{2}\|B(x(t))\|^2 + \lambda^2(t)\|v + Dx^*\|^2 + \|\dot{x}(t)\|^2.$$

By using that $\gamma(t)\dot{h}(t) \geq \frac{d}{dt}(\gamma(t)h(t))$ and $\|\ddot{x}(t) + \gamma(t)\dot{x}(t)\|^2 \leq 2\|\ddot{x}(t)\|^2 + 2\gamma^2(t)\|\dot{x}(t)\|^2$ for every $t \geq 0$, by integrating from t_3 to $T > t_3$, by letting T converge to $+\infty$, and by using (i), we obtain

$$\int_0^{+\infty} \lambda(t)\|x^* - x(t)\|^2 dt < +\infty.$$

Since $\lambda \in L^2([0, +\infty)) \setminus L^1([0, +\infty))$, it follows that $\|x(t) - x^*\| \longrightarrow 0$ as $t \longrightarrow +\infty$. □

Remark 13. We want to emphasize that unlike other papers addressing the asymptotic analysis of second order dynamical systems, and where the damping function $\gamma(t)$ was assumed to be strictly greater than $\sqrt{2}$ for all $t \geq 0$ (see [4, 20]), in Theorem 12 we allow for it take this value, too.

We close the paper by formulating Theorem 12 in the context of the optimization problem (1), where we also use the relation between the assumptions (H) and (H_{fitz}) (see Remark 10).

Corollary 14. *Consider the optimization problem* (1)*. Suppose that its system of optimality conditions*

$$0 \in \partial f(x) + \nabla g(x) + N_{\operatorname{zer}\nabla\psi}(x)$$

is solvable, that $(H1)-(H3)$ *and* (H) *hold, and let* x *be the unique strong global solution of* (7)*. Furthermore, suppose that* $\limsup_{t\longrightarrow+\infty}\lambda(t)\beta(t) < \frac{1}{L_B}$*,* γ *is locally absolutely continuous and for almost every* $t \geq 0$ *it holds* $\gamma(t) \geq \sqrt{2}$ *and* $\dot{\gamma}(t) \leq 0$*. Let* $\tilde{x} : [0, +\infty) \longrightarrow \mathcal{H}$ *be defined by*

$$\tilde{x}(t) = \frac{1}{\int_0^t \lambda(s)ds}\int_0^t \lambda(s)x(s)ds.$$

Then the following statements hold:

(i) for every $x^* \in \operatorname{zer}(\partial f + \nabla g + N_{\operatorname{zer}\nabla\psi})$*,* $\|x(t) - x^*\|$ *converges as* $t \longrightarrow +\infty$*; in addition,* $\dot{x}, \ddot{x} \in L^2([0, +\infty), \mathcal{H})$*,* $\lambda(\cdot)\beta(\cdot)\|\nabla\psi(x(\cdot))\|^2 \in L^1([0, +\infty))$*,* $\int_0^{+\infty}\lambda(t)\beta(t)\langle\nabla\psi(x(t)), x(t) - x^*\rangle dt < +\infty$*, and* $\lim_{t\to+\infty}\dot{x}(t) = \lim_{t\to+\infty}\dot{h}(t) = 0$*, where* $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$*;*

(ii) $\tilde{x}(t)$ *converges weakly as* $t \longrightarrow +\infty$ *to an element in* $\operatorname{zer}(\partial f + \nabla g + N_{\operatorname{zer}\nabla\psi})$*, which is also an optimal solution of* (1)*;*

(iii) if, additionally, f *is strongly convex, then* $x(t)$ *converges strongly as* $t \longrightarrow +\infty$ *to the unique element of* $\operatorname{zer}(\partial f + \nabla g + N_{\operatorname{zer}\nabla\psi})$*, which is the unique optimal solution of* (1)*.*

Remark 15. For unconstrained optimization problems (which corresponds to the situation when $\psi(x) = 0$ for all $x \in \mathcal{H}$) one can obtain convergence rates of $\mathcal{O}(t^{-2})$ for the objective along the trajectories by choosing as damping function $\gamma(t) = 1/t$, for $t > 0$ (see for instance [6]). This is a setting which is not covered by our analysis, however it is a topic which we might be of interest also in the context of dynamical systems with penalty terms associated to bilevel programming problems.

Remark 16. Time discretization of the dynamical system (6) with step size $\lambda_k > 0$, damping variable $\gamma_k > 0$, penalty parameter $\beta_k > 0$ leads to the numerical scheme

$$\frac{x^{k+1} - 2x^k + x^{k-1}}{h_k^2} + \gamma_k \frac{x^{k+1} - x^k}{h_k} + x^k = J_{\lambda_k A}\big(x^k - \lambda_k D x^k - \lambda_k \beta_k B x^k\big) \ \forall k \geq 1,$$

where $h_k > 0$ is the size of the time step and $x^0 = u_0$ and $x^1 = v_0$ are the initial points. For $h_k = 1$ this becomes

$$x^{k+1} = \left(1 - \frac{1}{1+\gamma_k}\right) x^k + \frac{1}{1+\gamma_k} J_{\lambda_k A}\big(x^k - \lambda_k D x^k - \lambda_k \beta_k B x^k\big) + \frac{1}{1+\gamma_k}(x^k - x^{k-1}) \ \forall k \geq 1,$$

which is a relaxed forward-backward algorithm of penalty type with inertial terms for solving the inclusion (5). For the convergence analysis of inertial forward-backward penalty algorithms in the particular case of convex optimization problems like (1) we refer the reader to [21].

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, 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

[4] H. Attouch, F. Alvarez, *The heavy ball with friction dynamical system for convex constrained minimization problems*, in: Optimization (Namur, 1998), Lecture Notes in Economics and Mathematical Systems 481, Springer, Berlin, 25–35, 2000

[5] H. Attouch, A. Cabot, M.-O. Czarnecki, *Asymptotic behavior of nonautonomous monotone and subgradient evolution equations*, Transactions of the American Mathematical Society 370(2), 755–790, 2018

[6] H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, Mathematical Programming, DOI 10.1007/s10107-016-0992-8

[7] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of coupled dynamical systems with multiscale aspects*, Journal of Differential Equations 248(6), 1315–1344, 2010

[8] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects*, Journal of Differential Equations 262(3), 2745–2770, 2017

[9] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Prox-penalization and splitting methods for constrained variational problems*, SIAM Journal on Optimization 21(1), 149–173, 2011

[10] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM Journal on Optimization 21(4), 1251–1274, 2011

[11] S. Banert, R.I. Boţ, *Backward penalty schemes for monotone inclusion problems*, Journal of Optimization Theory and Applications 166(3), 930–948, 2015

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

[13] H.H. Bauschke, D.A. McLaren, H.S. Sendov, *Fitzpatrick functions: inequalities, examples and remarks on a problem by S. Fitzpatrick*, Journal of Convex Analysis 13(3-4), 499–523, 2006

[14] J.M. Borwein, *Maximal monotonicity via convex analysis*, Journal of Convex Analysis 13(3-4), 561–586, 2006

[15] J.M. Borwein, J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010

[16] R.I. Boţ, E.R. Csetnek, *Forward-backward and Tseng's type penalty schemes for monotone inclusion problems*, Set-Valued and Variational Analysis 22, 313–331, 2014

[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, 1688–1700, 2016

[18] 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 29(1), 155–168, 2017

[19] R.I. Boţ, E.R. Csetnek, *Second order dynamical systems associated to variational inequalities*, Applicable Analysis 96(5), 799–809, 2017

[20] 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

[21] R.I. Boţ, E.R. Csetnek, N. Nimana, *An inertial proximal-gradient penalization scheme for constrained convex optimization problems*, Vietnam Journal of Mathematics, DOI 10.1007/s10013-017-0256-9

[22] 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

[23] S. Fitzpatrick, *Representing monotone operators by convex functions*, in: Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988), Proceedings of the Centre for Mathematical Analysis 20, Australian National University, Canberra, 59–65, 1988

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

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

[26] N. Noun, J. Peypouquet, *Forward-backward penalty scheme for constrained convex minimization without inf-compactness*, Journal of Optimization Theory and Applications, 158(3), 787–795, 2013

[27] J. Peypouquet, *Coupling the gradient method with a general exterior penalization scheme for convex minimization*, Journal of Optimizaton Theory and Applications 153(1), 123–138, 2012

[28] S. Simons, *From Hahn-Banach to Monotonicity*, Springer, Berlin, 2008

[29] E.D. Sontag, *Mathematical control theory. Deterministic finite-dimensional systems*, Second edition, Texts in Applied Mathematics 6, Springer-Verlag, New York, 1998

[30] W. Su, S. Boyd, E.J. Candes, *A differential equation for modeling Nesterov's accelerated gradient method: theory and insights*, arXiv:1503.01243, 2015