Approaching nonsmooth nonconvex minimization through second order proximal-gradient dynamical systems

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

February 12, 2018

Abstract. We investigate the asymptotic properties of the trajectories generated by a second-order dynamical system of proximal-gradient type stated in connection with the minimization of the sum of a nonsmooth convex and a (possibly nonconvex) smooth function. The convergence of the generated trajectory to a critical point of the objective is ensured provided a regularization of the objective function satisfies the Kurdyka-Łojasiewicz property. We also provide convergence rates for the trajectory formulated in terms of the Łojasiewicz exponent.

Key Words. second order dynamical system, nonsmooth nonconvex optimization, limiting subdifferential, Kurdyka-Łojasiewicz property

AMS subject classification. 34G25, 47J25, 47H05, 90C26, 90C30, 65K10

1 Introduction

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function and let $g : \mathbb{R}^n \to \mathbb{R}$ be a (possibly nonconvex) Fréchet differentiable function with β-Lipschitz continuous gradient, i.e. there exists $\beta \geq 0$ such that $\|\nabla g(x) - \nabla g(y)\| \leq \beta \|x - y\|$ for all $x, y \in \mathbb{R}^n$. In this paper we investigate the optimization problem

$$\inf_{x\in\mathbb{R}^n} [f(x) + g(x)] \tag{1}$$

by associating to it the following second order dynamical system of implicit-type

$$\begin{cases} \ddot{x}(t) + \gamma\dot{x}(t) + x(t) = \operatorname{prox}_{\lambda f}\big(x(t) - \lambda\nabla g(x(t))\big) \\ x(0) = u_0,\, \dot{x}(0) = v_0, \end{cases} \tag{2}$$

where $u_0, v_0 \in \mathbb{R}^n$, $\gamma, \lambda \in (0, +\infty)$ and

$$\operatorname{prox}_{\lambda f} : \mathbb{R}^n \to \mathbb{R}^n, \quad \operatorname{prox}_{\lambda f}(x) = \underset{y\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ f(y) + \frac{1}{2\lambda}\|y - x\|^2 \right\}, \tag{3}$$

denotes the proximal point operator of λf.

[*]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 partially supported by FWF (Austrian Science Fund), project P 29809-N32 and by an Advanced Fellowship STAR-UBB of Babeş-Bolyai University, Cluj Napoca.

[‡]Technical University of Cluj-Napoca, Department of Mathematics, Memorandumului 28, Cluj-Napoca, Romania, e-mail: szilard.laszlo@math.utcluj.ro. This work was supported by a grant of Ministry of Research and Innovation, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2016-0190, within PNCDI III.

Dynamical systems of proximal-gradient type associated to optimization problems have been intensively treated in the literature. In [16], Bolte studied the convergence of the trajectories of the first order dynamical system

$$\left\{\begin{array}{l} \dot{x}(t)+x(t)=\operatorname{proj}_C\big(x(t)-\lambda\nabla g(x(t))\big)\\ x(0)=x_0, \end{array}\right. \tag{4}$$

where $g:\mathbb{R}^n\to\mathbb{R}$ is a convex smooth function, $C\subseteq\mathbb{R}^n$ is a nonempty, closed and convex set, $x_0\in\mathbb{R}^n$, and proj_C denotes the projection operator on the set C. The trajectory of (4) has been proved to converge to a minimizer of the optimization problem

$$\inf_{x\in C} g(x), \tag{5}$$

provided the latter is solvable. We refer also to the work of Antipin [7] for further results related to (4).

The following extension of the dynamical system (4)

$$\left\{\begin{array}{l} \dot{x}(t)+x(t)=\operatorname{prox}_{\lambda f}\big(x(t)-\lambda\nabla g(x(t))\big)\\ x(0)=x_0, \end{array}\right. \tag{6}$$

where $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a proper, convex and lower semicontinuous function, $g:\mathbb{R}^n\to\mathbb{R}$ is a convex smooth function and $x_0\in\mathbb{R}^n$, has been recently considered by Abbas and Attouch [1] in relation to the optimization problem (1). In case (1) is solvable, the trajectory generated by (6) has been proved to converge to a global minimizer of it.

In connection with the optimization problem (5), the second order projected-gradient system

$$\left\{\begin{array}{l} \ddot{x}(t)+\gamma\dot{x}(t)+x(t)=\operatorname{proj}_C(x(t)-\lambda\nabla g(x(t)))\\ x(0)=u_0, \dot{x}(0)=v_0, \end{array}\right. \tag{7}$$

with damping parameter $\gamma>0$ and step size $\lambda>0$, has been considered in [7,8]. The system (7) becomes in case $C=\mathbb{R}^n$ the so-called "heavy ball method with friction". This nonlinear oscillator with damping is, in case $n=2$, a simplified version of the differential system describing the motion of a heavy ball that rolls over the graph of g and keeps rolling under its own inertia until friction stops it at a critical point of g (see [14]).

Implicit dynamical systems related to both optimization problems and monotone inclusions have been considered in the literature also by Attouch and Svaiter in [15], Attouch, Abbas and Svaiter in [2] and Attouch, Alvarez and Svaiter in [9]. These investigations have been continued and extended in [21–24].

The aim of this manuscript is to study the asymptotic properties of the trajectory generated by the second order dynamical system (2) under convexity assumptions for f and by allowing g to be nonconvex. In the same setting, a first order dynamical system of type (6) attached to (1) has been recently studied in [25]. An asymptotic analysis for a gradient-like second order dynamical system (which corresponds to 7 when $C=\mathbb{R}^n$) has been made in [29] (see also the recent review [30]) in the analytic setting.

The main results of the current work are Theorem 16, where we prove convergence of the trajectories to a critical point of the objective function of (1), provided a regularization of it satisfies the Kurdyka-Łojasiewicz property, and Theorem 20, where convergences rates by means of the Łojasiewicz exponent are provided for both the trajectory and the velocity. The convergence analysis relies on methods and techniques of real algebraic geometry introduced by Łojasiewicz [32] and Kurdyka [31] and extended to the nonsmooth setting by Attouch, Bolte and Svaiter [13] and Bolte, Sabach and Teboulle [17].

The explicit discretization of (2) with respect to the time variable t, with step size $h_k>0$, damping variable $\gamma_k>0$ and initial points $x_0:=u_0$ and $x_1:=v_0$ yields the iterative scheme

$$\frac{x_{k+1}-2x_k+x_{k-1}}{h_k^2}+\gamma_k\frac{x_{k+1}-x_k}{h_k}+x_k=\operatorname{prox}_{\lambda f}\big(x_k-\lambda\nabla g(x_k)\big)\ \forall k\geq 1.$$

For $h_k = 1$ this becomes

$$x_{k+1} = \left(1 - \frac{1}{1+\gamma_k}\right) x_k + \frac{1}{1+\gamma_k} \operatorname{prox}_{\lambda f}\left(x_k - \lambda \nabla g(x_k)\right) + \frac{1}{1+\gamma_k}(x_k - x_{k-1}) \ \forall k \geq 1,$$

which is a relaxed proximal-gradient algorithm for minimizing $f+g$ with inertial effects. For inertial-type algorithms we refer the reader to [3–5]. The dynamical system investigated in this paper can be seen as a continuous counterpart of the inertial-type algorithms presented in [26] and [34].

2 Preliminaries

In this section we introduce some basic notions and present preliminary results that will be used in the sequel. The finite-dimensional spaces considered in the manuscript are endowed with the Euclidean norm topology. The *domain* of the function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined by $\operatorname{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$. We say that f is *proper*, if $\operatorname{dom} f \neq \emptyset$. For the following generalized subdifferential notions and their basic properties we refer to [33, 35]. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. For $x \in \operatorname{dom} f$, the *Fréchet (viscosity) subdifferential* of f at x is defined as

$$\hat{\partial} f(x) = \left\{ v \in \mathbb{R}^n : \liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y-x \rangle}{\|y-x\|} \geq 0 \right\}.$$

For $x \notin \operatorname{dom} f$, we set $\hat{\partial} f(x) := \emptyset$. The *limiting (Mordukhovich) subdifferential* is defined at $x \in \operatorname{dom} f$ by

$$\partial f(x) = \{v \in \mathbb{R}^n : \exists x_k \to x, f(x_k) \to f(x) \text{ and } \exists v_k \in \hat{\partial} f(x_k), v_k \to v \text{ as } k \to +\infty\},$$

while for $x \notin \operatorname{dom} f$, we set $\partial f(x) := \emptyset$. Notice the inclusion $\hat{\partial} f(x) \subseteq \partial f(x)$ for each $x \in \mathbb{R}^n$.

In case f is convex, these notions coincide with the *convex subdifferential*, which means that $\hat{\partial} f(x) = \partial f(x) = \{v \in \mathbb{R}^n : f(y) \geq f(x) + \langle v, y - x\rangle \ \forall y \in \mathbb{R}^n\}$ for all $x \in \operatorname{dom} f$.

We will use the following closedness criterion concerning the graph of the limiting subdifferential: if $(x_k)_{k \geq 0}$ and $(v_k)_{k \geq 0}$ are sequences in $\mathbb{R}^n$ such that $v_k \in \partial f(x_k)$ for all $k \geq 0$, $(x_k, v_k) \to (x, v)$ and $f(x_k) \to f(x)$ as $k \to +\infty$, then $v \in \partial f(x)$.

The Fermat rule reads in this nonsmooth setting as: if $x \in \mathbb{R}^n$ is a local minimizer of f, then $0 \in \partial f(x)$. Notice that in case f is continuously differentiable around $x \in \mathbb{R}^n$ we have $\partial f(x) = \{\nabla f(x)\}$. We denote by

$$\operatorname{crit}(f) = \{x \in \mathbb{R}^n : 0 \in \partial f(x)\}$$

the set of *(limiting)-critical points* of f. We also mention the following subdifferential rule: if $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper and lower semicontinuous and $h : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, then $\partial(f+h)(x) = \partial f(x) + \nabla h(x)$ for all $x \in \mathbb{R}^n$.

Definition 1 (see, for instance, [2, 15]) A function $x : [0, +\infty) \to \mathbb{R}^n$ is said to be locally absolutely continuous, if is absolutely continuous on every interval $[0, T]$, $T > 0$, that is, one of the following equivalent properties holds:

(i) there exists an integrable function $y : [0, T] \to \mathbb{R}^n$ such that

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

(ii) x is continuous and its distributional derivative is Lebesgue integrable on $[0, T]$;

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

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

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

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

Further, we recall the following result of Brézis [27].

Lemma 2 *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function. Let $x \in L^2([0, T], \mathbb{R}^n)$, $T > 0$, be absolutely continuous such that $\dot{x} \in L^2([0, T], \mathbb{R}^n)$ and $x(t) \in \operatorname{dom} f$ for almost every $t \in [0, T]$. Assume that there exists $\xi \in L^2([0, T], \mathbb{R}^n)$ such that $\xi(t) \in \partial f(x(t))$ for almost every $t \in [0, T]$. Then the function $t \longrightarrow f(x(t))$ is absolutely continuous and for every t such that $x(t) \in \operatorname{dom} \partial f$ we have*

$$\frac{d}{dt} f(x(t)) = \langle \dot{x}(t), h \rangle, \, \forall h \in \partial f(x(t)).$$

The following central results will be used when proving the convergence of the trajectories generated by the dynamical system (2); 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 $\lim_{t \to +\infty} F(t) = 0$.

3 Existence and uniqueness of the trajectories

Existence and uniqueness of the trajectories of (2) are obtained in the framework of the global version of the Cauchy-Lipschitz Theorem (see for instance [12, Theorem 17.1.2(b)]), by rewriting (2) as a first order dynamical system in a suitable product space and by employing the Lipschitz continuity of the proximal operator and of the gradient.

Theorem 5 *For every starting points $u_0, v_0 \in \mathbb{R}^n$, the dynamical system (2) has a unique global solution $x \in C^2([0, +\infty), \mathbb{R}^n)$.*

Proof. By making use of the notation $X(t) = (x(t), \dot{x}(t))$, the system (2) can be rewritten as

$$\begin{cases} \dot{X}(t) = F(X(t)) \\ X(0) = (u_0, v_0), \end{cases} \tag{8}$$

where $F : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n \times \mathbb{R}^n$, $F(u, v) = \big(v, \operatorname{prox}_{\lambda f}\big(u - \lambda \nabla g(u)\big) - \gamma v - u\big)$.

We prove the existence and uniqueness of a global solution of (8) by using the Cauchy-Lipschitz Theorem. To this aim it is enough to show that F is globally Lipschitz continuous. Let be $(u,v),(\overline{u},\overline{v}) \in \mathbb{R}^n \times \mathbb{R}^n$. We have

$$\begin{aligned}\|F(u,v)-F(\overline{u},\overline{v})\| =& \Big\|\Big(v-\overline{v}, \operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)+\gamma(\overline{v}-v)+(\overline{u}-u)\Big)\Big\| \\ =& \sqrt{\|v-\overline{v}\|^2+\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)+\gamma(\overline{v}-v)+(\overline{u}-u)\|^2}.\end{aligned}$$

We have

$$\begin{aligned}&\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)+\gamma(\overline{v}-v)+(\overline{u}-u)\|^2 = \\ &\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)\|^2+\gamma^2\|\overline{v}-v\|^2+\|\overline{u}-u\|^2+ \\ &2\gamma\langle\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big),\overline{v}-v\rangle+ \\ &2\langle\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big),\overline{u}-u\rangle+ \\ &2\gamma\langle\overline{v}-v,\overline{u}-u\rangle.\end{aligned}$$

By the nonexpansiveness of $\operatorname{prox}_{\lambda f}$ and the β-Lipschitz property of ∇g we have

$$\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)\| \leq \|(u-\overline{u})-\lambda(\nabla g(u)-\nabla g(\overline{u})\| \leq (1+\lambda\beta)\|u-\overline{u}\|.$$

On the other hand,

$$\begin{aligned}2\gamma\langle\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big),\overline{v}-v\rangle \leq \\ \gamma\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)\|^2+\gamma\|\overline{v}-v\|^2 \leq \\ \gamma(1+\lambda\beta)^2\|u-\overline{u}\|^2+\gamma\|\overline{v}-v\|^2,\end{aligned}$$

$$\begin{aligned}2\langle\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big),\overline{u}-u\rangle \leq \\ \|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)\|^2+\|\overline{u}-u\|^2 \leq \\ (1+(1+\lambda\beta)^2)\|u-\overline{u}\|^2\end{aligned}$$

and

$$2\gamma\langle\overline{v}-v,\overline{u}-u\rangle \leq \gamma\|\overline{v}-v\|^2+\gamma\|u-\overline{u}\|^2.$$

Consequently,

$$\begin{aligned}\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)+\gamma(\overline{v}-v)+(\overline{u}-u)\|^2 \leq \\ (\gamma+2)((1+\lambda\beta)^2+1)\|u-\overline{u}\|^2+(\gamma^2+2\gamma)\|\overline{v}-v\|^2,\end{aligned}$$

which leads to

$$\|F(u,v)-F(\overline{u},\overline{v})\| \leq \sqrt{(\gamma+1)^2\|\overline{v}-v\|^2+(\gamma+2)((1+\lambda\beta)^2+1)\|u-\overline{u}\|^2} \leq L_1\|(u,v)-(\overline{u},\overline{v})\|,$$

where $L_1 := \sqrt{\max\big((\gamma+1)^2,(\gamma+2)((1+\lambda\beta)^2+1)\big)}$.

Consequently, F is globally Lipschitz continuous, which implies that (8) has a global solution $X \in C^1([0,+\infty),\mathbb{R}^n\times\mathbb{R}^n)$. This shows that $x \in C^2([0,+\infty),\mathbb{R}^n)$. ■

Remark 6 Another Lipschitz constant can be obtained by using the inequalities:

$$\begin{aligned}2\gamma\langle\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big),\overline{v}-v\rangle \leq \\ 2\gamma\|\operatorname{prox}_{\lambda f}\big(u-\lambda\nabla g(u)\big)-\operatorname{prox}_{\lambda f}\big(\overline{u}-\lambda\nabla g(\overline{u})\big)\|\|\overline{v}-v\| \leq 2\gamma(1+\lambda\beta)\|u-\overline{u}\|\|\overline{v}-v\|,\end{aligned}$$

$$2\langle \operatorname{prox}_{\lambda f}\left(u-\lambda\nabla g(u)\right)-\operatorname{prox}_{\lambda f}\left(\overline{u}-\lambda\nabla g(\overline{u})\right),\overline{u}-u\rangle \leq$$
$$2\|\operatorname{prox}_{\lambda f}\left(u-\lambda\nabla g(u)\right)-\operatorname{prox}_{\lambda f}\left(\overline{u}-\lambda\nabla g(\overline{u})\right)\|\|\overline{u}-u\| \leq 2(1+\lambda\beta)\|\overline{u}-u\|^2,$$

$$2\gamma\langle \overline{v}-v,\overline{u}-u\rangle \leq 2\gamma\|\overline{u}-u\|\|\overline{v}-v\|,$$

and

$$2\|\overline{u}-u\|\|\overline{v}-v\| \leq \|\overline{u}-u\|^2+\|\overline{v}-v\|^2.$$

In this case one obtains the Lipschitz constant

$$L_2 := \sqrt{\max((\gamma+1)^2+\gamma\lambda\beta,(2+\lambda\beta)^2+\gamma(2+\lambda\beta))}.$$

Remark 7 Considering again the setting of the proof of Theorem 5, from Remark 1(b) it follows that $\ddot{X}$ exists almost everywhere on $[0,+\infty)$ and that for almost every $t\in[0,+\infty)$ one has

$$\|\ddot{X}(t)\| \leq L_1\|\dot{X}(t)\| = \sqrt{\max\left((\gamma+1)^2,(\gamma+2)((1+\lambda\beta)^2+1)\right)}\|\dot{X}(t)\|.$$

Hence, $\sqrt{\|\ddot{x}(t)\|^2+\|x^{(3)}(t)\|^2} \leq \sqrt{\max\left((\gamma+1)^2,(\gamma+2)((1+\lambda\beta)^2+1)\right)}\sqrt{\|\dot{x}(t)\|^2+\|\ddot{x}(t)\|^2}$, for almost every $t\in[0,+\infty)$, or, equivalently,

$$\begin{aligned}\|x^{(3)}(t)\|^2 \leq& \max\left((\gamma+1)^2,(\gamma+2)((1+\lambda\beta)^2+1)\right)\|\dot{x}(t)\|^2+\\ &(\max\left((\gamma+1)^2,(\gamma+2)((1+\lambda\beta)^2+1)\right)-1)\|\ddot{x}(t)\|^2.\end{aligned} \tag{9}$$

Similarly, by using L_2, one obtains for almost every $t\in[0,+\infty)$

$$\begin{aligned}\|x^{(3)}(t)\|^2 \leq& \max((\gamma+1)^2+\gamma\lambda\beta,(2+\lambda\beta)^2+\gamma(2+\lambda\beta))\|\dot{x}(t)\|^2+\\ &(\max((\gamma+1)^2+\gamma\lambda\beta,(2+\lambda\beta)^2+\gamma(2+\lambda\beta))-1)\|\ddot{x}(t)\|^2.\end{aligned} \tag{10}$$

Remark 8 Obviously, $L_1>2$ and $L_2>2$. One can easily verify that $L_2\leq L_1$, provided $\gamma\leq\sqrt{3}$. Moreover, if $\gamma\leq\sqrt{3}$, then

$$L_2=\sqrt{(2+\lambda\beta)^2+\gamma(2+\lambda\beta)}.$$

However, for $\gamma>\sqrt{3}$, one may have $L_2>L_1$ and also $L_2<L_1$. Indeed, for $\gamma=2$ and $\lambda\beta=\frac{1}{10}$, it holds

$$L_2=\sqrt{9,2}>3=L_1,$$

while for $\gamma=2$ and $\lambda\beta=1$ it holds

$$L_2=\sqrt{15}<\sqrt{20}=L_1.$$

4 Asymptotic analysis

In this section we will address the asymptotic behaviour of the trajectory generated by the second order dynamical system (2). We begin the analysis with some technical results.

Lemma 9 *Suppose that $f+g$ is bounded from bellow and $\gamma,\lambda>0$ satisfy the following set of conditions:*

$$(\rho)\quad\begin{cases} A=-\dfrac{1}{2}\dfrac{\gamma}{\lambda}+\dfrac{\beta}{2}(L^2+2\gamma^2+1)<0\\ B=-\dfrac{1}{2L^2}\dfrac{\gamma}{\lambda}+\dfrac{\beta}{2}(L^2+\gamma^2+1)<0\\ C=-\dfrac{(2L^2+1)}{(L^2+1)^2}\gamma^2+3\beta\gamma\lambda-1<0,\end{cases}$$

where $L := \min(L_1, L_2)$, and L_1, L_2 were defined as

$$L_1 = \sqrt{\max\left((\gamma+1)^2, (\gamma+2)((1+\lambda\beta)^2+1)\right)}$$

and

$$L_2 = \sqrt{\max((\gamma+1)^2 + \gamma\lambda\beta, (2+\lambda\beta)^2 + \gamma(2+\lambda\beta))}.$$

For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Then the following statements are true

(a) $\dot{x} \in L^2([0,+\infty), \mathbb{R}^n)$ and $\lim_{t \longrightarrow +\infty} \dot{x}(t) = 0$;

(b) $\ddot{x} \in L^2([0,+\infty), \mathbb{R}^n)$ and $\lim_{t \longrightarrow +\infty} \ddot{x}(t) = 0$;

(c) $\exists \lim_{t \longrightarrow +\infty} (f+g)(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) \in \mathbb{R}$.

Proof. Let $T > 0$. Since $x \in C^2([0,T], \mathbb{R}^n)$, we have $x, \dot{x}, \ddot{x} \in L^2([0,T], \mathbb{R}^n)$. Further, by the β-Lipschitz property of ∇g we have $\nabla g \in L^2([0,T], \mathbb{R}^n)$. Moreover, (9) ensures that $x^{(3)} \in L^2([0,T], \mathbb{R}^n)$.

According to (2), we have $\ddot{x}(t) + \gamma\dot{x}(t) + x(t) = \operatorname{prox}_{\lambda f}\left(x(t) - \lambda\nabla g(x(t))\right)$ for all $t \in [0,+\infty)$, hence

$$-\frac{1}{\lambda}\ddot{x}(t) - \frac{\gamma}{\lambda}\dot{x}(t) - \nabla g(x(t)) \in \partial f(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)). \tag{11}$$

On the other hand, $\xi(t) = -\frac{1}{\lambda}\ddot{x}(t) - \frac{\gamma}{\lambda}\dot{x}(t) - \nabla g(x(t)) \in L^2([0,T], \mathbb{R}^n)$, hence by Lemma 2 we have that $t \longrightarrow f(\ddot{x}(t) + \gamma\dot{x}(t) + x(t))$ is absolutely continuous and

$$\frac{d}{dt} f(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) = \left\langle x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t), -\frac{1}{\lambda}\ddot{x}(t) - \frac{\gamma}{\lambda}\dot{x}(t) - \nabla g(x(t))\right\rangle \tag{12}$$

for almost every $t \in [0,T]$.

Obviously,

$$\frac{d}{dt} g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) = \left\langle x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t), \nabla g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t))\right\rangle \tag{13}$$

for almost every $t \in [0,T]$. By summing up the last two equalities we get

$$\begin{aligned}
&\frac{d}{dt}(f+g)(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) = \\
&\left\langle x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t), \nabla g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) - \nabla g(x(t)) - \frac{1}{\lambda}\ddot{x}(t) - \frac{\gamma}{\lambda}\dot{x}(t)\right\rangle = \\
&-\frac{1}{2\lambda}\frac{d}{dt}\|\ddot{x}(t)\|^2 - \frac{1+\gamma^2}{2\lambda}\frac{d}{dt}\|\dot{x}(t)\|^2 - \frac{\gamma}{\lambda}\|\dot{x}(t)\|^2 - \frac{\gamma}{\lambda}\|\ddot{x}(t)\|^2 - \frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle + \\
&\left\langle x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t), \nabla g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) - \nabla g(x(t))\right\rangle
\end{aligned}$$

for almost every $t \in [0,T]$. It is easy to check that $\langle x^{(3)}(t), \dot{x}(t)\rangle = \frac{1}{2} \cdot \frac{d^2}{dt^2}\|\dot{x}(t)\|^2 - \|\ddot{x}(t)\|^2$ for almost every $t \in [0,+\infty)$. Let $c \in (0,1)$. We have

$$-\frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle = -c\frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle - (1-c)\frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle$$

and

$$-(1-c)\frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle \leq \left(a\|x^{(3)}(t)\|^2 + b\|\dot{x}(t)\|^2\right),$$

where $ab = \frac{\gamma^2(1-c)^2}{4\lambda^2}$, hence by using (9) and (10) one obtains that for almost every $t \in [0, +\infty)$

$$-(1-c)\frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle \leq (aL^2 + b)\|\dot{x}(t)\|^2 + a(L^2-1)\|\ddot{x}(t)\|^2.$$

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

$$-\frac{\gamma}{\lambda}\langle x^{(3)}(t), \dot{x}(t)\rangle \leq -\frac{c\gamma}{2\lambda} \cdot \frac{d^2}{dt^2}\|\dot{x}(t)\|^2 + (aL^2+b)\|\dot{x}(t)\|^2 + \left(aL^2 + \frac{c\gamma}{\lambda} - a\right)\|\ddot{x}(t)\|^2. \tag{14}$$

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

$$\begin{aligned}\Big\langle x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t), \nabla g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) - \nabla g(x(t))\Big\rangle &\leq \\ \beta\|\ddot{x}(t) + \gamma\dot{x}(t)\|\|x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t)\| &\leq \\ \beta(\|\ddot{x}(t) + \gamma\dot{x}(t)\|\|x^{(3)}(t)\| + \|\ddot{x}(t) + \gamma\dot{x}(t)\|\|\gamma\ddot{x}(t) + \dot{x}(t)\|) &\leq \\ \frac{\beta}{2}(2\|\ddot{x}(t) + \gamma\dot{x}(t)\|^2 + \|x^{(3)}(t)\|^2 + \|\gamma\ddot{x}(t) + \dot{x}(t)\|^2) &= \\ \frac{\beta}{2}\big((2+\gamma^2)\|\ddot{x}(t)\|^2 + (2\gamma^2+1)\|\dot{x}(t)\|^2 + \|x^{(3)}(t)\|^2 + 6\gamma\langle\ddot{x}(t), \dot{x}(t)\rangle\big) &= \\ \frac{\beta}{2}\left((2+\gamma^2)\|\ddot{x}(t)\|^2 + (2\gamma^2+1)\|\dot{x}(t)\|^2 + \|x^{(3)}(t)\|^2 + 3\gamma\frac{d}{dt}\|\dot{x}(t)\|^2\right)&.\end{aligned}$$

By using (9) and (10) one obtains for almost every $t \in [0, T]$

$$\|x^{(3)}(t)\|^2 \leq L^2\|\dot{x}(t)\|^2 + (L^2-1)\|\ddot{x}(t)\|^2,$$

hence

$$\begin{aligned}\Big\langle x^{(3)}(t) + \gamma\ddot{x}(t) + \dot{x}(t), \nabla g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) - \nabla g(x(t))\Big\rangle &\leq \\ \frac{\beta}{2}(L^2+\gamma^2+1)\|\ddot{x}(t)\|^2 + \frac{\beta}{2}(L^2+2\gamma^2+1)\|\dot{x}(t)\|^2 + 3\frac{\beta}{2}\gamma\frac{d}{dt}\|\dot{x}(t)\|^2&.\end{aligned}$$

Consequently, for almost every $t \in [0, T]$ we have

$$\frac{d}{dt}(f+g)(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) + \frac{1}{2\lambda}\frac{d}{dt}\|\ddot{x}(t)\|^2 + \frac{(1+\gamma^2) - 3\lambda\beta\gamma}{2\lambda}\frac{d}{dt}\|\dot{x}(t)\|^2 + \frac{c\gamma}{2\lambda} \cdot \frac{d^2}{dt^2}\|\dot{x}(t)\|^2 \leq$$
$$\left((c-1)\frac{\gamma}{\lambda} + aL^2 - a + \frac{\beta}{2}(L^2+\gamma^2+1)\right)\|\ddot{x}(t)\|^2 + \left(-\frac{\gamma}{\lambda} + aL^2 + b + \frac{\beta}{2}(L^2+2\gamma^2+1)\right)\|\dot{x}(t)\|^2.$$

Recall that a, b and c have been arbitrarily chosen such that $c \in (0,1)$ and $ab = \frac{\gamma^2(1-c)^2}{4\lambda^2}$.
We chose

$$c := \frac{L^2}{L^2+1}, \ a := \frac{\gamma}{2(L^2+1)L^2\lambda} \text{ and } b := \frac{L^2\gamma}{2(L^2+1)\lambda}.$$

Then, for almost every $t \in [0, T]$ we have

$$\frac{d}{dt}\left[(f+g)(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) + \frac{1}{2\lambda}\|\ddot{x}(t)\|^2 + \frac{c^2\gamma^2 - C}{2\lambda}\|\dot{x}(t)\|^2 + \frac{2c\gamma}{2\lambda}\langle\ddot{x}(t), \dot{x}(t)\rangle\right] \leq$$
$$A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2. \tag{15}$$

By integration we get

$$(f+g)(\ddot{x}(T)+\gamma\dot{x}(T)+x(T))+\frac{1}{2\lambda}\|\ddot{x}(T)\|^2+\frac{c^2\gamma^2-C}{2\lambda}\|\dot{x}(T)\|^2+\frac{2c\gamma}{2\lambda}\langle\ddot{x}(T),\dot{x}(T)\rangle\leq$$
$$(f+g)(\ddot{x}(0)+\gamma\dot{x}(0)+x(0))+\frac{1}{2\lambda}\|\ddot{x}(0)\|^2+\frac{c^2\gamma^2-C}{2\lambda}\|\dot{x}(0)\|^2+\frac{2c\gamma}{2\lambda}\langle\ddot{x}(0),\dot{x}(0)\rangle+$$
$$A\int_0^T\|\dot{x}(t)\|^2dt+B\int_0^T\|\ddot{x}(t)\|^2dt.$$

In other words,

$$(f+g)(\ddot{x}(T)+\gamma\dot{x}(T)+x(T))+\frac{1}{2\lambda}\|\ddot{x}(T)+c\gamma\dot{x}(T)\|^2-\frac{C}{2\lambda}\|\dot{x}(T)\|^2\leq$$
$$(f+g)(\ddot{x}(0)+\gamma\dot{x}(0)+x(0))+\frac{1}{2\lambda}\|\ddot{x}(0)+c\gamma\dot{x}(0)\|^2-\frac{C}{2\lambda}\|\dot{x}(0)\|^2+$$
$$A\int_0^T\|\dot{x}(t)\|^2dt+B\int_0^T\|\ddot{x}(t)\|^2dt. \tag{16}$$

By using that $A<0$, $B<0$, $C<0$ and $f+g$ is bounded from below, and by taking into account that $T>0$ has been arbitrary chosen, we obtain that $\dot{x},\ \ddot{x}\in L^2([0,+\infty),\mathbb{R}^n)$. Moreover, from (9) we obtain that $x^{(3)}\in L^2([0,+\infty),\mathbb{R}^n)$.

Now, by using Lemma 4 and the fact that for almost every $t\in[0,+\infty)$ we have

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

and

$$\frac{d}{dt}\|\ddot{x}(t)\|^2=2\langle\ddot{x}(t),x^{(3)}(t)\rangle\leq\|\ddot{x}(t)\|^2+\|x^{(3)}(t)\|^2,$$

we obtain that $\lim_{t\longrightarrow+\infty}\dot{x}(t)=0$ and $\lim_{t\longrightarrow+\infty}\ddot{x}(t)=0$.

Since $T>0$ has been arbitrary chosen, we get from (15) that for almost every $t\in[0,+\infty)$

$$\frac{d}{dt}\left[(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+\frac{1}{2\lambda}\|\ddot{x}(t)\|^2+\frac{c^2\gamma^2-C}{2\lambda}\|\dot{x}(t)\|^2+\frac{c\gamma}{\lambda}\langle\ddot{x}(t),\dot{x}(t)\rangle\right]\leq 0.$$

Now using Lemma 3 we obtain that the limit

$$\lim_{t\longrightarrow+\infty}\left[(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+\frac{1}{2\lambda}\|\ddot{x}(t)\|^2+\frac{c^2\gamma^2-C}{2\lambda}\|\dot{x}(t)\|^2+\frac{c\gamma}{\lambda}\langle\ddot{x}(t),\dot{x}(t)\rangle\right]$$

exists and is finite. Since

$$\lim_{t\longrightarrow+\infty}\left[\frac{1}{2\lambda}\|\ddot{x}(t)\|^2+\frac{c^2\gamma^2-C}{2\lambda}\|\dot{x}(t)\|^2+\frac{c\gamma}{\lambda}\langle\ddot{x}(t),\dot{x}(t)\rangle\right]=0,$$

one obtains that

$$\lim_{t\longrightarrow+\infty}(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))\in\mathbb{R}.$$

■

Remark 10 The choice $\gamma\lambda\beta\leq\frac{1}{3}$ guarantees that $C<0$. Moreover, in this case $B>A$. Indeed,

$$B-A=\frac{\gamma}{2\lambda}\left(1-\frac{1}{L^2}-\gamma\lambda\beta\right)\geq\frac{\gamma}{2\lambda}\left(\frac{2}{3}-\frac{1}{L^2}\right)>0.$$

Corollary 11 *Suppose that $f+g$ is bounded from bellow and $\sqrt{3} \geq \gamma > 0, \lambda > 0$ satisfy the following condition*

$$-\frac{1}{(2+\lambda\beta)^2+\gamma(2+\lambda\beta)}\frac{\gamma}{\lambda}+\beta((2+\lambda\beta)^2+\gamma(2+\lambda\beta)+\gamma^2+1)<0.$$

For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty),\mathbb{R}^n)$ be the unique global solution of (2)*. Then the following statements are true*

(a) $\dot{x} \in L^2([0,+\infty),\mathbb{R}^n)$ *and* $\lim_{t\longrightarrow+\infty} \dot{x}(t) = 0$*;*

(b) $\ddot{x} \in L^2([0,+\infty),\mathbb{R}^n)$ *and* $\lim_{t\longrightarrow+\infty} \ddot{x}(t) = 0$*;*

(c) $\exists \lim_{t\longrightarrow+\infty}(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t)) \in \mathbb{R}$.

Proof. The condition $\gamma \leq \sqrt{3}$ ensures that $L = \sqrt{(2+\lambda\beta)^2+\gamma(2+\lambda\beta)}$, hence

$$2B = -\frac{1}{(2+\lambda\beta)^2+\gamma(2+\lambda\beta)}\frac{\gamma}{\lambda}+\beta((2+\lambda\beta)^2+\gamma(2+\lambda\beta)+\gamma^2+1)<0.$$

Under these auspicies, it can proved that $\gamma\lambda\beta \leq \frac{1}{3}$, hence, according to the previous remark, $C<0$ and $A<0$. The statement follows from Lemma 9. ■

Lemma 12 *Assume that $f+g$ is bounded from below and γ, λ satisfy the set of conditions (ρ). For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty),\mathbb{R}^n)$ be the unique global solution of* (2)*. Then the set of limit points of x, which we denote by $\omega(x)$, is a subset of the set of critical points of $f+g$. In other words,*

$$\omega(x) := \{\overline{x} \in \mathbb{R}^n : \exists t_k \longrightarrow \infty \text{ such that } x(t_k) \longrightarrow \overline{x},\ k \longrightarrow +\infty\} \subseteq \operatorname{crit}(f+g).$$

Proof. Let $\overline{x} \in \omega(x)$ and $t_k \longrightarrow +\infty$ such that $x(t_k) \longrightarrow \overline{x}$, $k \longrightarrow +\infty$. We have to show that $0 \in \partial(f+g)(\overline{x})$. From (11) we have for every $k \geq 0$

$$-\frac{1}{\lambda}\ddot{x}(t_k)-\frac{\gamma}{\lambda}\dot{x}(t_k)-\nabla g(x(t_k)) \in \partial f(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k))$$

hence,

$$\begin{aligned} v_k = -\frac{1}{\lambda}\ddot{x}(t_k)-\frac{\gamma}{\lambda}\dot{x}(t_k)-\nabla g(x(t_k))+\nabla g(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k)) \in \\ \partial f(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k))+\nabla g(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k)) = \\ \partial(f+g)(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k)) = \partial(f+g)(u_k), \end{aligned}$$

where $u_k := \ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k)$.

According to Lemma 9, $\lim_{k\longrightarrow+\infty}\dot{x}(t_k) = 0$ and $\lim_{k\longrightarrow+\infty}\ddot{x}(t_k) = 0$. Further, ∇g is continuous, hence $\lim_{k\longrightarrow+\infty}[-\nabla g(x(t_k))+\nabla g(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k))] = -\nabla g(\overline{x})+\nabla g(\overline{x}) = 0$. Consequently,

$$\lim_{k\longrightarrow+\infty}(u_k, v_k) = (\overline{x}, 0).$$

We show that $\lim_{k\longrightarrow+\infty}(f+g)(u_k) = (f+g)(\overline{x})$. Since f is lower semicontinuous, one has

$$\liminf_{k\longrightarrow+\infty} f(u_k) \geq f(\overline{x}).$$

Further we have for every $k \geq 0$

$$u_k = \ddot{x}(t_k) + \gamma\dot{x}(t_k) + x(t_k) = \text{prox}_{\lambda f}\left(x(t_k) - \lambda\nabla g(x(t_k))\right) =$$

$$\underset{y\in\mathbb{R}^n}{\text{argmin}}\left[f(y) + \frac{1}{2\lambda}\|y - (x(t_k) - \lambda\nabla g(x(t_k))\|^2\right] =$$

$$\underset{y\in\mathbb{R}^n}{\text{argmin}}\left[f(y) + \frac{1}{2\lambda}\|y - x(t_k)\|^2 + \langle y - x(t_k), \nabla g(x(t_k))\rangle + \frac{\lambda}{2}\|\nabla g(x(t_k))\|^2\right] =$$

$$\underset{y\in\mathbb{R}^n}{\text{argmin}}\left[f(y) + \frac{1}{2\lambda}\|y - x(t_k)\|^2 + \langle y - x(t_k), \nabla g(x(t_k))\rangle\right].$$

Hence, for every $k \geq 0$ we have

$$f(u_k) + \frac{1}{2\lambda}\|u_k - x(t_k)\|^2 + \langle u_k - x(t_k), \nabla g(x(t_k))\rangle \leq f(\overline{x}) + \frac{1}{2\lambda}\|\overline{x} - x(t_k)\|^2 + \langle \overline{x} - x(t_k), \nabla g(x(t_k))\rangle.$$

Taking the limit superior as $k \longrightarrow +\infty$, we obtain

$$\limsup_{k\longrightarrow+\infty} f(u_k) \leq f(\overline{x}).$$

This shows that $\lim_{k\longrightarrow+\infty} f(u_k) = f(\overline{x})$ and, since g is continuous, we obtain

$$\lim_{k\longrightarrow+\infty}(f+g)(u_k) = (f+g)(\overline{x}).$$

By the closedness criterion of the graph of the limiting subdifferential it follows that

$$0 \in \partial(f+g)(\overline{x}).$$

■

Lemma 13 *Assume that $f + g$ is bounded from below and γ, λ satisfy the set of conditions (ρ), and let the constants L, A, B and C be defined as in Lemma 9. For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Consider the function*

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\},\ H(u,v,w) = (f+g)(u) + \frac{1}{2\lambda}\|u-v\|^2 - \frac{C}{2\lambda}\|w\|^2.$$

Then the following statements are true

(H_1) for almost every $t \in [0,+\infty)$ it holds

$$\frac{d}{dt}\left(H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t))\right) \leq 0$$

and the limit

$$\lim_{t\longrightarrow+\infty} H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t))$$

exists and is finite, where $c = \frac{L^2}{L^2+1}$;

(H_2) for almost every $t \in [0,+\infty)$ and for every $a \geq 0$ we have

$$w(t) = \left(-\nabla g(x(t)) + \nabla g(\ddot{x}(t) + \gamma\dot{x}(t) + x(t)) - \frac{1}{\lambda}a\gamma\dot{x}(t)), -\frac{1}{\lambda}(\ddot{x}(t) + (1-a)\gamma\dot{x}(t)), -\frac{C}{\lambda}\dot{x}(t)\right) \in \partial H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma a\dot{x}(t) + x(t), \dot{x}(t))$$

and

$$\|w(t)\| \leq \left(\beta + \frac{1}{\lambda}\right)\|\ddot{x}(t)\| + \frac{\beta\lambda\gamma + (2a+1)\gamma - C}{\lambda}\|\dot{x}(t)\|;$$

(H_3) for $\overline{x} \in \omega(x)$ and $t_k \longrightarrow +\infty$ such that $x(t_k) \longrightarrow \overline{x}$ as $k \longrightarrow +\infty$, and for every $a \geq 0$ we have

$$H(\ddot{x}(t_k) + \gamma\dot{x}(t_k) + x(t_k), a\gamma\dot{x}(t) + x(t_k), \dot{x}(t_k)) \longrightarrow H(\overline{x}, \overline{x}, 0) \text{ as } k \longrightarrow +\infty.$$

Proof. (H_1). From (15) we have that for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}\left[(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+\frac{1}{2\lambda}\|\ddot{x}(t)+c\gamma\dot{x}(t)\|^2-\frac{C}{2\lambda}\|\dot{x}(t)\|^2\right] \leq A\|\dot{x}(t)\|^2+B\|\ddot{x}(t)\|^2.$$

Taking into account that $A < 0$, $B < 0$, we obtain that for almost every $t \in [0, +\infty)$

$$\frac{d}{dt}\left(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t), \gamma(1-c)\dot{x}(t)+x(t), \dot{x}(t))\right) =$$
$$\frac{d}{dt}\left[(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+\frac{1}{2\lambda}\|\ddot{x}(t)+c\gamma\dot{x}(t)\|^2-\frac{C}{2\lambda}\|\dot{x}(t)\|^2\right] \leq 0.$$

By Lemma 3 it follows that the limit

$$\lim_{t\longrightarrow+\infty} H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t), \gamma(1-c)\dot{x}(t)+x(t), \dot{x}(t)) \in \mathbb{R}$$

exists.

(H_2). From (11) we have that for every $t \in [0, +\infty)$ it holds

$$-\frac{1}{\lambda}\ddot{x}(t)-\frac{\gamma}{\lambda}\dot{x}(t)-\nabla g(x(t)) \in \partial f(\ddot{x}(t)+\gamma\dot{x}(t)+x(t)),$$

hence

$$-\frac{1}{\lambda}\ddot{x}(t)-\frac{\gamma}{\lambda}\dot{x}(t)-\nabla g(x(t))+\nabla g(\ddot{x}(t)+\gamma\dot{x}(t)+x(t)) \in \partial(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t)).$$

Since for every $(u, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$

$$\partial H(u,v,w) = \left(\partial(f+g)(u)+\frac{1}{\lambda}(u-v)\right) \times \left\{-\frac{1}{\lambda}(u-v)\right\} \times \left\{-\frac{C}{\lambda}w\right\},$$

we get

$$\partial H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t), \gamma a\dot{x}(t)+x(t), \dot{x}(t)) =$$
$$\left(\partial(f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))+\frac{1}{\lambda}(\ddot{x}(t)+(1-a)\gamma\dot{x}(t))\right) \times \left\{-\frac{1}{\lambda}(\ddot{x}(t)+(1-a)\gamma\dot{x}(t))\right\} \times \left\{-\frac{C}{\lambda}\dot{x}(t)\right\},$$

consequently,

$$w(t) = \left(-\nabla g(x(t))+\nabla g(\ddot{x}(t)+\gamma\dot{x}(t)+x(t))-\frac{1}{\lambda}a\gamma\dot{x}(t)), -\frac{1}{\lambda}(\ddot{x}(t)+(1-a)\gamma\dot{x}(t)), -\frac{C}{\lambda}\dot{x}(t)\right) \in$$
$$\partial H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t), \gamma a\dot{x}(t)+x(t), \dot{x}(t))$$

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

From the $\beta-$Lipschitz continuity of ∇g we get for every $t \in [0, +\infty)$

$$\|w(t)\| \leq \left(\beta+\frac{1}{\lambda}\right)\|\ddot{x}(t)+\gamma\dot{x}(t)\|+2\frac{a\gamma}{\lambda}\|\dot{x}(t)\|-\frac{C}{\lambda}\|\dot{x}(t)\| \leq$$
$$\left(\beta+\frac{1}{\lambda}\right)\|\ddot{x}(t)\|+\frac{\beta\lambda\gamma+(2a+1)\gamma-C}{\lambda}\|\dot{x}(t)\|.$$

(H_3). Let $a \geq 0$, $\overline{x} \in \omega(x)$ and $t_k \longrightarrow +\infty$ such that $x(t_k) \longrightarrow \overline{x}$ as $k \longrightarrow +\infty$. According to the proof of Lemma 12 it holds $(f+g)(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k)) \longrightarrow (f+g)(\overline{x})$ as $k \longrightarrow +\infty$. Further, from Lemma 9 we have $\ddot{x}(t_k) \longrightarrow 0$ and $\dot{x}(t_k) \longrightarrow 0$ as $k \longrightarrow +\infty$. Hence,

$$H(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k), a\gamma\dot{x}(t_k)+x(t_k), \dot{x}(t_k)) = (f+g)(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k))+$$
$$\frac{1}{2\lambda}\|\ddot{x}(t_k)+(1-a)\gamma\dot{x}(t_k)\|^2-\frac{C}{2\lambda}\|\dot{x}(t_k)\|^2 \longrightarrow (f+g)(\overline{x}) = H(\overline{x}, \overline{x}, 0) \text{ as } k \longrightarrow +\infty.$$

■

Lemma 14 *Assume that $f+g$ is bounded from below and γ, λ satisfy the set of conditions (ρ), and let the constants L, A, B and C be defined as in Lemma 9. For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Consider the function*

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\},\ H(u,v,w) = (f+g)(u) + \frac{1}{2\lambda}\|u-v\|^2 - \frac{C}{2\lambda}\|w\|^2.$$

Suppose that x is bounded and let $a \geq 0$. Then the following statements are true

(a) $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x}) \subseteq \operatorname{crit}(H) = \{(u,u,0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : u \in \operatorname{crit}(f+g)\}$;

(b) $\lim_{t \longrightarrow +\infty} \operatorname{dist}((\ddot{x}(t)+\gamma\dot{x}(t)+x(t), a\gamma\dot{x}(t)+x(t), \dot{x}(t)), \omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})) = 0$;

(c) H is finite and constant on $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})$;

(d) $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})$ *is nonempty, compact and connected.*

Proof. (a) By definition,

$$\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x}) =$$
$$\{(\overline{x},\overline{y},\overline{z}) \in (\mathbb{R}^n)^3 : \exists t_k \to +\infty \text{ s. t. } (\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k), a\gamma\dot{x}(t_k)+x(t_k), \dot{x}(t_k)) \to (\overline{x},\overline{y},\overline{z}),\ k \to +\infty\}.$$

According to Lemma 9, $\ddot{x}(t_k) \longrightarrow 0$, $\dot{x}(t_k) \longrightarrow 0$ as $t_k \longrightarrow +\infty$, hence

$$\begin{aligned} \omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x}) = \omega(x,x,0) = \\ \{(\overline{x},\overline{x},0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \exists t_k \to +\infty \text{ such that } x(t_k) \longrightarrow \overline{x},\ k \to +\infty\} = \\ \{(\overline{x},\overline{x},0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \overline{x} \in \omega(x)\}. \end{aligned}$$

According to Lemma 12,

$$\{(\overline{x},\overline{x},0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \overline{x} \in \omega(x)\} \subseteq \{(\overline{x},\overline{x},0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \overline{x} \in \operatorname{crit}(f+g)\} = \operatorname{crit}(H).$$

(b) Obviously

$$0 \leq \lim_{t \longrightarrow +\infty} \operatorname{dist}((\ddot{x}(t)+\gamma\dot{x}(t)+x(t), a\gamma\dot{x}(t)+x(t), \dot{x}(t)), \omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})) \leq$$
$$\lim_{t_k \longrightarrow +\infty} \operatorname{dist}((\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k), a\gamma\dot{x}(t_k)+x(t_k), \dot{x}(t_k)), \omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})) = 0.$$

(c) According to Lemma 9,

$$\lim_{t \longrightarrow +\infty} (f+g)(\ddot{x}(t)+\gamma\dot{x}(t)+x(t)) = l \in \mathbb{R}.$$

Let $(\overline{x},\overline{x},0) \in \omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})$. Then there exists $t_k \longrightarrow +\infty$ such that $(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k), a\gamma\dot{x}(t_k)+x(t_k), \dot{x}(t_k)) \longrightarrow (\overline{x},\overline{x},0)$ as $k \longrightarrow +\infty$. From Lemma 13($\mathrm{H}_3$) one has

$$H(\overline{x},\overline{x},0) = \lim_{t_k \longrightarrow +\infty} H(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k), a\gamma\dot{x}(t_k)+x(t_k), \dot{x}(t_k)) =$$
$$\lim_{t_k \longrightarrow +\infty} [(f+g)(\ddot{x}(t_k)+\gamma\dot{x}(t_k)+x(t_k)) + \frac{1}{2\lambda}\|\ddot{x}(t_k)+(1-a)\gamma\dot{x}(t_k)\|^2 - \frac{C}{2\lambda}\|\dot{x}(t_k)\|^2] = l.$$

Hence, H takes on $\omega(\ddot{x}+\gamma\dot{x}+x, a\gamma\dot{x}+x, \dot{x})$ the constant value l.

Finally, (d) is a classical result from [28]. We also refer the reader to the proof of Theorem 4.1 in [6], where it is shown that the properties of $\omega(x)$ of being nonempty, compact and connected are generic for bounded trajectories fulfilling $\lim_{t \to +\infty} \dot{x}(t) = 0$ (see also [17] for a discrete version of this result). ■

The convergence of the trajectory generated by the dynamical system (2) will be shown in the framework of functions satisfying the *Kurdyka-Łojasiewicz property.* For $\eta \in (0, +\infty]$, we denote by Θ_η the class of concave and continuous functions $\varphi : [0, \eta) \to [0, +\infty)$ such that $\varphi(0) = 0$, φ is continuously differentiable on $(0, \eta)$, continuous at 0 and $\varphi'(s) > 0$ for all $s \in (0, \eta)$. In the following definition (see [11, 17]) we use the *distance function* to a set, defined for $A \subseteq \mathbb{R}^n$ as $\operatorname{dist}(x, A) = \inf_{y \in A} \|x - y\|$ for all $x \in \mathbb{R}^n$.

Definition 2 (*Kurdyka-Łojasiewicz property*) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. We say that f satisfies the *Kurdyka-Łojasiewicz (KL) property* at $\overline{x} \in \operatorname{dom} \partial f = \{x \in \mathbb{R}^n : \partial f(x) \neq \emptyset\}$ if there exist $\eta \in (0, +\infty]$, a neighborhood U of $\overline{x}$ and a function $\varphi \in \Theta_\eta$ such that for all x in the intersection

$$U \cap \{x \in \mathbb{R}^n : f(\overline{x}) < f(x) < f(\overline{x}) + \eta\}$$

the following inequality holds

$$\varphi'(f(x) - f(\overline{x})) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

If f satisfies the KL property at each point in $\operatorname{dom} \partial f$, then f is called a *KL function.*

The origins of this notion go back to the pioneering work of Łojasiewicz [32], where it is proved that for a real-analytic function $f : \mathbb{R}^n \to \mathbb{R}$ and a critical point $\overline{x} \in \mathbb{R}^n$ (that is $\nabla f(\overline{x}) = 0$), there exists $\theta \in [1/2, 1)$ such that the function $|f - f(\overline{x})|^\theta \|\nabla f\|^{-1}$ is bounded around $\overline{x}$. This corresponds to the situation when $\varphi(s) = C(1 - \theta)^{-1} s^{1-\theta}$. The result of Łojasiewicz allows the interpretation of the KL property as a re-parametrization of the function values in order to avoid flatness around the critical points. Kurdyka [31] extended this property to differentiable functions definable in an o-minimal structure. Further extensions to the nonsmooth setting can be found in [11, 18–20].

One of the remarkable properties of the KL functions is their ubiquity in applications, according to [17]. To the class of KL functions belong semi-algebraic, real sub-analytic, semiconvex, uniformly convex and convex functions satisfying a growth condition. We refer the reader to [10, 11, 13, 17–20] and the references therein for more details regarding all the classes mentioned above and illustrating examples.

An important role in our convergence analysis will be played by the following uniformized KL property given in [17, Lemma 6].

Lemma 15 *Let $\Omega \subseteq \mathbb{R}^n$ be a compact set and let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that f is constant on Ω and f satisfies the KL property at each point of Ω. Then there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for all $\overline{x} \in \Omega$ and for all x in the intersection*

$$\{x \in \mathbb{R}^n : \operatorname{dist}(x, \Omega) < \varepsilon\} \cap \{x \in \mathbb{R}^n : f(\overline{x}) < f(x) < f(\overline{x}) + \eta\} \tag{17}$$

the following inequality holds

$$\varphi'(f(x) - f(\overline{x})) \operatorname{dist}(0, \partial f(x)) \geq 1. \tag{18}$$

We state the first main result of the paper.

Theorem 16 *Assume that $f + g$ is bounded from below and γ, λ satisfy the set of conditions (ρ), and let the constants L, A, B and C be defined as in Lemma 9. For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (2). Consider the function*

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\},\ H(u, v, w) = (f + g)(u) + \frac{1}{2\lambda}\|u - v\|^2 - \frac{C}{2\lambda}\|w\|^2.$$

Suppose that x is bounded and H is a KL function. Then the following statements are true *(a)* $\dot{x} \in L^1([0, +\infty), \mathbb{R}^n)$;

(b) $\ddot{x} \in L^1([0, +\infty), \mathbb{R}^n)$;

(c) there exists $\overline{x} \in \operatorname{crit}(f+g)$ *such that* $\lim_{t \longrightarrow +\infty} x(t) = \overline{x}$.

Proof. Let be $c := \frac{L^2}{L^2+1}$. Consider an arbitrary $(\overline{x}, \overline{x}, 0) \in \omega(\ddot{x} + \gamma\dot{x} + x, (1-c)\gamma\dot{x} + x, \dot{x})$. Then one has

$$\lim_{t \longrightarrow +\infty} \left(H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) \right) = H(\overline{x}, \overline{x}, 0).$$

Case I. There exists $\overline{t} \geq 0$ such that

$$H(\ddot{x}(\overline{t}) + \gamma\dot{x}(\overline{t}) + x(\overline{t}), \gamma(1-c)\dot{x}(\overline{t}) + x(\overline{t}), \dot{x}(\overline{t})) = H(\overline{x}, \overline{x}, 0).$$

We have for almost every $t \in [0, +\infty)$ that

$$\frac{d}{dt} \left[H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) \right] \leq A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2 \leq 0.$$

Hence, for every $t \geq \overline{t}$ it holds

$$H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) \leq H(\overline{x}, \overline{x}, 0).$$

On the other hand

$$H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) \geq$$
$$\lim_{t \longrightarrow +\infty} \left(H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) \right) = H(\overline{x}, \overline{x}, 0),$$

hence

$$H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) = H(\overline{x}, \overline{x}, 0)$$

for every $t \geq \overline{t}$.

Consequently,

$$\frac{d}{dt} \left[H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t) \right] = 0$$

for every $t \geq \overline{t}$, which means that

$$0 \leq A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2 \leq 0$$

for every $t \geq \overline{t}$.

But $A < 0$ and $B < 0$, hence $\dot{x}(t) = 0$ and $\ddot{x}(t) = 0$ on $[\overline{t}, +\infty)$. This leads to $\dot{x}, \ddot{x} \in L^1([0, +\infty), \mathbb{R}^n)$ and to the fact hat $x(t) = \overline{x}$ is constant on $[\overline{t}, +\infty)$.

Case II. For every $t \geq 0$

$$H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) > H(\overline{x}, \overline{x}, 0).$$

Let $\Omega = \omega(\ddot{x} + \gamma\dot{x} + x, (1-c)\gamma\dot{x} + x, \dot{x})$. According to Lemma 14, H is constant and finite on Ω and Ω is nonempty, compact and connected. Since H is a KL function, by Lemma 15, there exist $\varepsilon, \eta > 0$ and a concave function $\varphi \in \Theta_\eta$ such that for every $(\overline{x}, \overline{x}, 0) \in \Omega$ and every

$$\begin{aligned} (x, y, z) \in \{(u, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : \operatorname{dist}((u, v, w), \Omega) < \varepsilon\} \cap \\ \{(u, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n : H(\overline{x}, \overline{x}, 0) < H(u, v, w) < H(\overline{x}, \overline{x}, 0) + \eta\} \end{aligned} \tag{19}$$

the following inequality holds

$$\varphi'(H(x,y,z) - H(\overline{x},\overline{x},0))\,\mathrm{dist}((0,0,0),\partial H(x,y,z)) \geq 1. \tag{20}$$

Since

$$\lim_{t\longrightarrow+\infty} (H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t))) = H(\overline{x},\overline{x},0)$$

and

$$H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) > H(\overline{x},\overline{x},0),$$

there exists $t_1 > 0$ such that

$$H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) < H(\overline{x},\overline{x},0) + \eta\,\forall t \geq t_1.$$

Since $\lim_{t\longrightarrow+\infty} \mathrm{dist}((\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)),\Omega) = 0$, there exists $t_2 \geq 0$ such that

$$\mathrm{dist}((\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)),\Omega) < \epsilon,\ \forall t \geq t_2.$$

Hence, for every $t \geq T = \max(t_1,t_2)$ we have

$$\begin{aligned}&\varphi'(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) - H(\overline{x},\overline{x},0))\cdot\\ &\quad \mathrm{dist}((0,0,0),\partial H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t))) \geq 1.\end{aligned}$$

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

$$\mathrm{dist}((0,0,0),\partial H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t))) \leq \|w(t)\|,$$

where

$$w(t) = \left(-\nabla g(x(t)) + \nabla g(\ddot{x}(t)+\gamma\dot{x}(t)+x(t)) - \frac{1}{\lambda}(1-c)\gamma\dot{x}(t)), -\frac{1}{\lambda}(\ddot{x}(t)+c\gamma\dot{x}(t)), -\frac{C}{\lambda}\dot{x}(t)\right)$$

since, according to Lemma 13 (H_2),

$$w(t) \in \partial H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)).$$

Further,

$$\|w(t)\| \leq \left(\beta + \frac{1}{\lambda}\right)\|\ddot{x}(t)\| + \frac{\beta\lambda\gamma + (3-2c)\gamma - C}{\lambda}\|\dot{x}(t)\|$$

which leads to

$$\varphi'(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) - H(\overline{x},\overline{x},0))\,(s\|\ddot{x}(t)\| + p\|\dot{x}(t)\|) \geq 1\ \forall t \in [T,+\infty),$$

where $s := \beta + \frac{1}{\lambda} > 0$ and $p := \frac{\beta\lambda\gamma+(3-2c)\gamma-C}{\lambda} > 0$.

We have

$$\begin{aligned}&\frac{d}{dt}\varphi(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) - H(\overline{x},\overline{x},0)) =\\ &\quad\varphi'(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) - H(\overline{x},\overline{x},0))\cdot\\ &\qquad\qquad\frac{d}{dt}H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t))\end{aligned}$$

and since

$$\frac{d}{dt}H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t)) \leq A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2 \leq 0$$

and

$$\varphi'(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t))-H(\overline{x},\overline{x},0))\geq\frac{1}{s\|\ddot{x}(t)\|+p\|\dot{x}(t)\|}$$

we get for every $t\in[T,+\infty)$

$$\frac{d}{dt}\varphi(H(\ddot{x}(t)+\gamma\dot{x}(t)+x(t),\gamma(1-c)\dot{x}(t)+x(t),\dot{x}(t))-H(\overline{x},\overline{x},0))\leq\frac{A\|\dot{x}(t)\|^2+B\|\ddot{x}(t)\|^2}{s\|\ddot{x}(t)\|+p\|\dot{x}(t)\|}\leq 0. \quad (21)$$

Since φ is bounded from below, similarly as in the proof of Lemma 9, we obtain that

$$\frac{\|\dot{x}(\cdot)\|^2}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|},\frac{\|\ddot{x}(\cdot)\|^2}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}\in L^1([0,+\infty),\mathbb{R}).$$

By using the arithmetical-geometrical mean inequality we have

$$\sqrt{\frac{\|\dot{x}(\cdot)\|^2}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}\cdot\frac{\|\ddot{x}(\cdot)\|^2}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}}=\frac{\|\dot{x}(\cdot)\|\|\ddot{x}(\cdot)\|}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}\in L^1([0,+\infty),\mathbb{R}).$$

Hence,

$$\|\dot{x}(\cdot)\|+\|\ddot{x}(\cdot)\|=p\frac{\|\dot{x}(\cdot)\|^2}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}+s\frac{\|\ddot{x}(\cdot)\|^2}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}+(s+p)\frac{\|\dot{x}(\cdot)\|\|\ddot{x}(\cdot)\|}{s\|\ddot{x}(\cdot)\|+p\|\dot{x}(\cdot)\|}\in L^1([0,+\infty),\mathbb{R}).$$

This shows that $\dot{x},\ddot{x}\in L^1([0,+\infty),\mathbb{R}^n)$, hence, according to Lemma 3, there exists $\lim_{t\longrightarrow+\infty}x(t)=\overline{x}$. ■

Remark 17 Similar regularizations of the objective function as the one considered in this section have been used in [25] for studying first order dynamical systems, but also in [26, 34], in the investigation of non-relaxed forward-backward methods involving inertial and memory effects in the nonconvex setting.

Remark 18 Since the class of semi-algebraic functions is closed under addition (see for example [17]) and $(u,v)\mapsto\alpha\|u-v\|^2$ and $w\mapsto\alpha'\|w\|^2$ are semi-algebraic for $\alpha,\alpha'>0$, the conclusion of the previous theorem holds if the condition H is a KL function is replaced by the assumption that $f+g$ is semi-algebraic.

Remark 19 Assume that $\gamma,\lambda>0$ fulfill the set of conditions (ρ) and that $f+g$ is coercive, that is

$$\lim_{\|u\|\to+\infty}(f+g)(u)=+\infty.$$

For $u_0,v_0\in\mathbb{R}^n$, let $x\in C^2([0,+\infty),\mathbb{R}^n)$ be the unique global solution of (2). Then x is bounded.

Indeed, notice that $f+g$ is bounded from below, being a proper, lower semicontinuous and coercive function (see for example [35]). From (16) it follows that $\ddot{x}(T)+\gamma\dot{x}(T)+x(T)$ is contained for every $T\geq 0$ in a lower level set of $f+g$, which is a bounded set due to the coercivity assumption. Combining this fact with Lemma 9 one can easily derive that x is bounded.

5 Convergence rates

In the context of optimization problems involving KL functions, it is known (see [10, 18, 32]) that convergence rates of the trajectory can be formulated in terms of the so-called Łojasiewicz exponent.

Definition 3 *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. The function f is said to fulfill the Łojasiewicz property, if for every $\overline{x} \in \operatorname{crit} f$ there exist $K, \epsilon > 0$ and $\theta \in (0,1)$ such that*

$$|f(x) - f(\overline{x})|^\theta \leq K\|x^*\| \text{ for every } x \text{ fulfilling } \|x - \overline{x}\| < \epsilon \text{ and every } x^* \in \partial f(x).$$

The number θ is called the Łojasiewicz exponent of f at the critical point $\overline{x}$.

In the following theorem we obtain convergence rates for both the trajectory generated (2) and its velocity (see, also, [10, 18]).

Theorem 20 *Assume that $f + g$ is bounded from below and γ, λ satisfy the set of conditions (ρ), and let the constants L, A, B and C be defined as in Lemma 9. For $u_0, v_0 \in \mathbb{R}^n$, let $x \in C^2([0,+\infty), \mathbb{R}^n)$ be the unique global solution of (2). Consider the function*

$$H : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\},\ H(u,v,w) = (f+g)(u) + \frac{1}{2\lambda}\|u - v\|^2 - \frac{C}{2\lambda}\|w\|^2.$$

Suppose that x is bounded and let $\overline{x} \in \operatorname{crit}(f+g)$ be such that $\lim_{t\longrightarrow+\infty} x(t) = \overline{x}$ and H fulfills the Łojasiewicz property at $(\overline{x}, \overline{x}, 0) \in \operatorname{crit} H$ with Łojasiewicz exponent θ.

Then, there exist $a_1, a_2, a_3, a_4 > 0$ and $t_0 > 0$ such that for every $t \in [t_0, +\infty)$ the following statements are true

(a) if $\theta \in (0, \frac{1}{2})$, then x converges in finite time;

(b) if $\theta = \frac{1}{2}$, then $\|x(t) - \overline{x}\| \leq a_1 e^{-a_2 t}$ and $\|\dot{x}(t)\| \leq a_1 e^{-a_2 t}$;

(c) if $\theta \in (\frac{1}{2}, 1)$, then $\|x(t) - \overline{x}\| \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}}$ and $\|\dot{x}(t)\| \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}}$.

Proof. Let be $s := \beta + \frac{1}{\lambda} > 0$ and $p := \frac{\beta\lambda\gamma + (3-2c)\gamma - C}{\lambda} > 0$, as defined in Lemma 13. The function $g : [0, +\infty) \longrightarrow \mathbb{R}$, $g(r) = \frac{A + Br^2}{p + (s+p)r + sr^2}$ attains at $r_0 = \frac{(sA - pB) - \sqrt{(sA-pB)^2 + (s+p)^2 AB}}{(s+p)B} > 0$ its maximum. Hence, for $m := \max\left(\frac{B}{s}, g(r_0)\right) < 0$, it holds

$$A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2 \leq m(s\|\ddot{x}(t)\| + p\|\dot{x}(t)\|)(\|\dot{x}(t)\| + \|\ddot{x}(t)\|)$$

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

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

$$\sigma(t) := \int_t^{+\infty} (\|\dot{x}(s)\| + \|\ddot{x}(t)\|) ds.$$

Let $t \in [0, +\infty)$ be fixed. For $T \geq t$ we have

$$\|x(t) - \overline{x}\| = \left\| x(T) - \overline{x} - \int_t^T \dot{x}(s) ds \right\| \leq \|x(T) - \overline{x}\| + \int_t^T \|\dot{x}(s)\| ds.$$

By taking the limit as $T \longrightarrow +\infty$ we obtain

$$\|x(t) - \overline{x}\| \leq \int_t^{+\infty} \|\dot{x}(s)\| ds \leq \sigma(t). \tag{22}$$

Further, for $T \geq t$ we have

$$\|\dot{x}(t)\| = \left\| \dot{x}(T) - \int_t^T \ddot{x}(s) ds \right\| \leq \|\dot{x}(T)\| + \int_t^T \|\ddot{x}(s)\| ds.$$

By taking the limit as $T \longrightarrow +\infty$ we obtain

$$\|\dot{x}(t)\| \le \int_t^{+\infty} \|\ddot{x}(s)\| ds \le \sigma(t). \tag{23}$$

We have seen in the proof of Theorem 16 that, if there exists $\bar{t} \ge 0$ such that

$$H(\ddot{x}(\bar{t}) + \gamma\dot{x}(\bar{t}) + x(\bar{t}), \gamma(1-c)\dot{x}(\bar{t}) + x(\bar{t}), \dot{x}(\bar{t})) = H(\overline{x}, \overline{x}, 0),$$

then x is constant on $[\bar{t}, +\infty)$, hence the conclusion follows automatically.

On the other hand, if for every $t \ge 0$ one has

$$H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) > H(\overline{x}, \overline{x}, 0),$$

then, according to the proof of Theorem 16 and (21), there exists $t_0 \ge 0$ such that for every $t \in [t_0, +\infty)$

$$K\frac{d}{dt}(H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) - H(\overline{x}, \overline{x}, 0))^{1-\theta} \le \frac{A\|\dot{x}(t)\|^2 + B\|\ddot{x}(t)\|^2}{s\|\ddot{x}(t)\| + p\|\dot{x}(t)\|},$$

and

$$\|(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), (1-c)\gamma\dot{x}(t) + x(t), \dot{x}(t)) - (\overline{x}, \overline{x}, 0)\| < \epsilon.$$

Hence, for every $t \in [t_0, +\infty)$

$$M(\|\dot{x}(t)\| + \|\ddot{x}(t)\|) + \frac{d}{dt}(H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) - H(\overline{x}, \overline{x}, 0))^{1-\theta} \le 0, \tag{24}$$

and

$$\|(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), (1-c)\gamma\dot{x}(t) + x(t), \dot{x}(t)) - (\overline{x}, \overline{x}, 0)\| < \epsilon,$$

where $M := -\frac{m}{K} > 0$. If we integrate (24) on the interval $[t, T]$, where $T \ge t \ge t_0$, we obtain

$$M\int_t^T (\|\dot{x}(s)\| + \|\ddot{x}(s)\|)ds + (H(\ddot{x}(T) + \gamma\dot{x}(T) + x(T), \gamma(1-c)\dot{x}(T) + x(T), \dot{x}(T)) - H(\overline{x}, \overline{x}, 0))^{1-\theta} \le$$
$$(H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) - H(\overline{x}, \overline{x}, 0))^{1-\theta},$$

hence

$$M\sigma(t) \le (H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) - H(\overline{x}, \overline{x}, 0))^{1-\theta} \quad \forall t \ge t_0.$$

Since θ is the Łojasiewicz exponent of H at the point $(\overline{x}, \overline{x}, 0) \in \operatorname{crit} H$, we have

$$|H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)) - H(\overline{x}, \overline{x}, 0)|^{\theta} \le K\|x^*\|,$$

for every $t \in [t_0, +\infty)$ and every

$$x^* \in \partial H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t)).$$

According to Lemma 13(H_2), there exists some $\widetilde{x}^* \in \partial H(\ddot{x}(t) + \gamma\dot{x}(t) + x(t), \gamma(1-c)\dot{x}(t) + x(t), \dot{x}(t))$ such that for almost every $t \in [t_0, +\infty)$

$$\|\widetilde{x}^*(t)\| \le s\|\ddot{x}(t)\| + p\|\dot{x}(t)\| \le N(\|\ddot{x}(t)\| + \|\dot{x}(t)\|),$$

where $N = \max(s, p)$. Hence,

$$M\sigma(t) \le (KN(\|\ddot{x}(t)\| + \|\dot{x}(t)\|))^{\frac{1-\theta}{\theta}}$$

for almost every $t \in [t_0, +\infty)$. But $\dot{\sigma}(t) = -\|\ddot{x}(t)\| - \|\dot{x}(t)\|$, consequently, there exists $\alpha > 0$ such that for almost every $t \in [t_0, +\infty)$

$$\dot{\sigma}(t) \leq -\alpha(\sigma(t))^{\frac{\theta}{1-\theta}}. \tag{25}$$

If $\theta = \frac{1}{2}$, then $\dot{\sigma}(t) \leq -\alpha(\sigma(t))$ for almost every $t \in [t_0, +\infty)$. By multiplying with $e^{\alpha t}$ and integrating on $[t_0, t]$, we get that there exist $a_1, a_2 > 0$ such that

$$\sigma(t) \leq a_1 e^{-a_2 t} \ \forall t \in [t_0, +\infty),$$

hence, by (22) and (23), we get

$$\|x(t) - \overline{x}\| \leq a_1 e^{-a_2 t} \text{ and } \|\dot{x}(t)\| \leq a_1 e^{-a_2 t} \ \forall t \in [t_0, +\infty),$$

which proves (b).

Assume now that $0 < \theta < \frac{1}{2}$. By using (25) we obtain

$$\frac{d}{dt}\left(\sigma(t)\right)^{\frac{1-2\theta}{1-\theta}} = \frac{1-2\theta}{1-\theta}\left(\sigma(t)\right)^{\frac{-\theta}{1-\theta}}\dot{\sigma}(t) \leq -\alpha\frac{1-2\theta}{1-\theta},$$

for almost every $t \in [t_0, +\infty)$.

By integration we get

$$\left(\sigma(t)\right)^{\frac{1-2\theta}{1-\theta}} \leq -\overline{\alpha}t + \overline{\beta} \ \forall t \in [t_0, +\infty),$$

where $\overline{\alpha} > 0$. Hence, there exists $T \geq 0$ such that $\sigma(T) \leq 0 \ \forall t \in [T, +\infty)$, which implies that x is constant on $[T, +\infty)$.

Assume now that $\frac{1}{2} < \theta < 1$. By using (25) we obtain

$$\frac{d}{dt}\left(\sigma(t)\right)^{\frac{1-2\theta}{1-\theta}} = \frac{1-2\theta}{1-\theta}\left(\sigma(t)\right)^{\frac{-\theta}{1-\theta}}\dot{\sigma}(t) \geq \alpha\frac{2\theta-1}{1-\theta}$$

for almost every $t \in [t_0, +\infty)$.

By integration we get

$$\sigma(t) \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}} \ \forall t \in [t_0, +\infty),$$

where $a_3, a_4 > 0$.

From (22) and (23) we have

$$\|x(t) - \overline{x}\| \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}} \text{ and } \|\dot{x}(t)\| \leq (a_3 t + a_4)^{-\frac{1-\theta}{2\theta-1}} \ \forall t \in [t_0, +\infty),$$

which proves (c). ■

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, 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, J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming 116(1-2) Series B, 5–16, 2009

[11] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality*, Mathematics of Operations Research 35(2), 438–457, 2010

[12] H. Attouch, G. Buttazo, G. Michaille, Variational Analysis in Sobolev ans BV Spaces: Applications to PDEs and Optimization, Second Edition, MOS-SIAM Series on Optimization, Phiadelphia, 2014.

[13] H. Attouch, J. Bolte, B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137(1-2) Series A, 91–129, 2013

[14] H. Attouch, X. Goudou, P. Redont, *The heavy ball with friction method. I. The continuous dynamical system: global exploration of the local minima of a real-valued function by asymptotic analysis of a dissipative dynamical system*, Communications in Contemporary Mathematics 2(1), 1–34, 2000

[15] H. Attouch, B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49(2), 574-598, 2011

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

[17] J. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming Series A (146)(1–2), 459–494, 2014

[18] J. Bolte, A. Daniilidis, A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205–1223, 2006

[19] J. Bolte, A. Daniilidis, A. Lewis, M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization 18(2), 556–572, 2007

[20] J. Bolte, A. Daniilidis, O. Ley, L. Mazet, *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society 362(6), 3319–3363, 2010

[21] 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

[22] 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

[23] 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

[24] R.I. Boţ, E.R. Csetnek, *Convergence rates for forward-backward dynamical systems associated with strongly monotone inclusions*, Journal of Mathematical Analysis and Applications 457(2), 1135–1152, 2018

[25] R.I. Boţ, E.R. Csetnek, *A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function*, to appear in ESAIM: Control, Optimisation and Calculus of Variations, arXiv:1507.01416, 2015

[26] R.I. Boţ, E.R. Csetnek, S. László, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO Journal on Computational Optimization 4, 3–25, 2016

[27] H. Brézis, Operateurs Maximaux Monotones et Semi-Groupes De Contractions Dans Les Espaces De Hilbert, North-Holland Mathematics Studies 5, 1973

[28] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathé- matiques Appliquées 17, Masson, Paris, 1991

[29] A. Haraux, M.A. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonliniarities*, Journal of Differential Equations 144(2), 313–320, 1998

[30] A. Haraux, M.A. Jendoubi, *The Convergence Problem for Dissipative Autonomous Systems. Classical Methods and Recent Advances*, SpringerBriefs in Mathematics, Springer, Cham, 2015

[31] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Annales de l'institut Fourier (Grenoble) 48(3), 769–783, 1998

[32] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris, 87–89, 1963

[33] B. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*, Springer-Verlag, Berlin, 2006

[34] P. Ochs, Y. Chen, T. Brox, T. Pock, *iPiano: Inertial proximal algorithm for non-convex optimization*, SIAM Journal on Imaging Sciences 7(2), 1388–1419, 2014

[35] R.T. Rockafellar, R.J.-B. Wets, *Variational Analysis*, Fundamental Principles of Mathematical Sciences 317, Springer-Verlag, Berlin, 1998