A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function

Radu Ioan Boţ [*] Ernö Robert Csetnek [†]

February 28, 2017

Abstract. We address the minimization of the sum of a proper, convex and lower semicontinuous with a (possibly nonconvex) smooth function from the perspective of an implicit dynamical system of forward-backward type. The latter is formulated by means of the gradient of the smooth function and of the proximal point operator of the nonsmooth one. The trajectory generated by the dynamical system is proved to asymptotically converge to a critical point of the objective, provided a regularization of the latter satisfies the Kurdyka-Łojasiewicz property. Convergence rates for the trajectory in terms of the Łojasiewicz exponent of the regularized objective function are also provided.

Key Words. dynamical systems, continuous forward-backward method, nonsmooth optimization, limiting subdifferential, Kurdyka-Łojasiewicz property

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

1 Introduction

In this paper we approach the solving of the optimization problem

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

where $f:\mathbb{R}^n\to\mathbb{R}\cup\{+\infty\}$ is a proper, convex, lower semicontinuous function and $g:\mathbb{R}^n\to\mathbb{R}$ a (possibly nonconvex) Fréchet differentiable function with β-Lipschitz continuous gradient for $\beta\geq 0$, i.e., $\|\nabla g(x)-\nabla g(y)\|\leq\beta\|x-y\|\ \forall x,y\in\mathbb{R}^n$, by associating to it the implicit dynamical system

$$\left\{\begin{array}{l}\dot{x}(t)+x(t)=\operatorname{prox}_{\eta f}\Big(x(t)-\eta\nabla g(x(t))\Big)\\ x(0)=x_0,\end{array}\right. \tag{2}$$

where $\eta>0$, $x_0\in\mathbb{R}^n$ is chosen arbitrary and $\operatorname{prox}_{\eta f}:\mathbb{R}^n\to\mathbb{R}^n$, defined by

$$\operatorname{prox}_{\eta f}(y)=\operatorname*{argmin}_{u\in\mathbb{R}^n}\left\{f(u)+\frac{1}{2\eta}\|u-y\|^2\right\}$$

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

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

is the *proximal point operator* of ηf.

Due to the Lipschitz property of the proximal point operator and of the gradient of the differentiable function, the existence and uniqueness of strong global solutions of the dynamical system (2) is ensured in the framework of the Cauchy-Lipschitz Theorem.

The asymptotic analysis of the trajectories is carried out in the setting of functions satisfying the *Kurdyka-Łojasiewicz* property (so-called *KL functions*). To this large class belong functions with different analytic features. The convergence analysis relies on methods and techniques of real algebraic geometry introduced by Łojasiewicz [38] and Kurdyka [37] and developed recently in the nonsmooth setting by Attouch, Bolte and Svaiter [13] and Bolte, Sabach and Teboulle [24].

The approach for proving asymptotic convergence for the trajectories generated by (2) towards a critical point of the objective function of (1), expressed as a zero of the limiting (Mordukhovich) subdifferential, use three main ingredients (see [6] for the continuous case and also [13, 24] for a similar approach in the discrete setting). Namely, we show a sufficient decrease property along the trajectories of a regularization of the objective function, the existence of a subgradient lower bound for the trajectories and, finally, we obtain convergence by making use of the Kurdyka-Łojasiewicz property of the objective function. The case when the objective function is semi-algebraic follows as particular case of our analysis. We close our investigations by establishing convergence rates for the trajectories expressed in terms of the Łojasiewicz exponent of the regularized objective function.

In the context of minimizing a (nonconvex) smooth function (which corresponds to the case when in (1) f is taken equal to zero), several first- and second-order gradient type dynamical systems have been investigated by Łojasiewicz [38], Simon [42], Haraux and Jendoubi [35], Alvarez, Attouch, Bolte and Redont [6, Section 4], Bolte, Daniilidis and Lewis [21, Section 4], etc. In the aforementioned papers, the convergence of the trajectories is obtained in the framework of KL functions.

In what concerns implicit dynamical systems of the same type as (2), let us mention that Bolte has studied in [20] the asymptotic convergence of the trajectories of

$$\left\{ \begin{array}{l} \dot{x}(t) + x(t) = \operatorname{proj}_C \big(x(t) - \eta\nabla g(x(t))\big) \\ x(0) = x_0. \end{array} \right. \tag{3}$$

where $g : \mathbb{R}^n \to \mathbb{R}^n$ is convex and differentiable with Lipschitz continuous gradient and proj_C denotes the projection operator on the nonempty, closed and convex set $C \subseteq \mathbb{R}^n$, towards a minimizer of g over C. This corresponds to the case when in (1) f is the indicator function of the set C, namely, $f(x) = 0$, for $x \in C$ and $+\infty$, otherwise. We refer also to the work of Antipin [7] for more statements and results concerning the dynamical system (3). The approach of (1) by means of (2), stated as a generalization of (3), has been recently considered by Abbas and Attouch in [1, Section 5.2] in the full convex setting. Implicit dynamical systems related to both optimization problems and monotone inclusions have been considered in the literature also by Attouch and Svaiter in [16], Attouch, Abbas and Svaiter in [2] and Attouch, Alvarez and Svaiter in [10]. These investigations have been continued and extended in [17, 26–29].

A further argument in favour of implicit-type dynamical systems of type (2) is that its time discretization leads to the relaxed forward-backward iterative algorithm

$$x^{k+1} = (1-\lambda_k)x^k + \lambda_k \operatorname{prox}_{\eta f}(x^k - \eta\nabla g(x^k)) \ \forall k \geq 0, \tag{4}$$

where the starting point $x^0 \in \mathbb{R}^n$ is arbitrarily chosen and $(\lambda_k)_{k\geq 0}$ is the sequence of relaxation parameters. Relaxation algorithms are important for applications, since the parameter involving the relaxation offers the opportunity to accelerate the convergence of the iterates. While in the convex setting this is a well-known and understood fact (see for example [18]), in the nonconvex setting the convergence analysis of relaxation versions is less studied. The results presented below let us hope that one can develope a convergence analysis for the relaxed forward-backward method (4) in the nonconvex setting, too, an issue which is left as an open problem and which can be subject of future investigations.

The non-relaxed variant of (4) (when $\lambda_k = 1$ for all $k \geq 0$) has been investigated also in the nonconvex setting for KL functions. We refer the reader to [11–13,24,25,30,32,33,36,40] for different treatments of iterative schemes of type (4) in the nonconvex setting. Among the concrete applications of optimization problems involving analytic features, we mention here the use of the sparsity measure in compressive sensing, constrained feasibility problems involving semi-algebraic sets [13], sparse matrix factorization problems [24], restoration of noisy blurred images by means of the Student-t distribution [30, 40] and phase retrieval [19].

2 Preliminaries

In this section we recall some notions and results which are needed throughout the paper. Let $\mathbb{N} = \{0, 1, 2, ...\}$ be the set of nonnegative integers. For $n \geq 1$, the Euclidean scalar product and the induced norm on $\mathbb{R}^n$ are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Notice that all the finite-dimensional spaces considered in the manuscript are endowed with the topology induced by the Euclidean norm.

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 [39,41]. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. If $x \in \operatorname{dom} f$, we consider the *Fréchet (viscosity) subdifferential* of f at x as the set

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

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 $x \notin \operatorname{dom} f$, one takes $\partial f(x) := \emptyset$. Therefore $\hat{\partial} f(x) \subseteq \partial f(x)$ for each $x \in \mathbb{R}^n$.

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

The graph of the limiting subdifferential fulfills the following closedness criterion: if $(x_k)_{k\in\mathbb{N}}$ and $(v_k)_{k\in\mathbb{N}}$ are sequences in $\mathbb{R}^n$ such that $v_k \in \partial f(x_k)$ for all $k \in \mathbb{N}$, $(x_k, v_k) \to (x, v)$ 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 follows: if $x \in \mathbb{R}^n$ is a local minimizer of f, then $0 \in \partial f(x)$. We denote by

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

the set of *(limiting)-critical points* of f.

When f is continuously differentiable around $x \in \mathbb{R}^n$ we have $\partial f(x) = \{\nabla f(x)\}$. We will make use of the following subdifferential sum 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 $x \in \mathbb{R}^m$.

We turn now our attention to functions satisfying the *Kurdyka-Łojasiewicz property*. This class of functions will play a crucial role in the asymptotic analysis of the dynamical system (2). 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 [12,24]) we use also 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 1 (*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 *KL function*.

The origins of this notion go back to the pioneering work of Łojasiewicz [38], 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 $\varphi(s) = Cs^{1-\theta}$, where $C > 0$. 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 [37] extended this property to differentiable functions definable in o-minimal structures. Further extensions to the nonsmooth setting can be found in [12, 21–23].

One of the remarkable properties of the KL functions is their ubiquity in applications (see [24]). 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 [11–13, 21–24] and the references therein for more on KL functions and illustrating examples.

An important role in our convergence analysis will be played by the following uniform KL property given in [24, Lemma 6].

Lemma 1 *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 that it 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 all x in the intersection*

$$\{x \in \mathbb{R}^n : \operatorname{dist}(x, \Omega) < \varepsilon\} \cap \{x \in \mathbb{R}^n : f(\overline{x}) < f(x) < f(\overline{x}) + \eta\} \tag{5}$$

the inequality

$$\varphi'(f(x) - f(\overline{x}))\,\mathrm{dist}(0, \partial f(x)) \geq 1. \tag{6}$$

holds.

In the following we recall the notion of locally absolutely continuous function and state two of its basic properties.

Definition 2 (see, for instance, [2,16]) A function $x : [0, +\infty) \to \mathbb{R}^n$ is said to be locally absolutely continuous, if it absolutely continuous on every interval $[0, T]$, where $T > 0$, which means that one of the following equivalent properties holds:

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

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

(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 2 (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 $x : [0, T] \to \mathbb{R}^n$ is absolutely continuous for $T > 0$ and $B : \mathbb{R}^n \to \mathbb{R}^n$ is L-Lipschitz continuous for $L \geq 0$, 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 2(iii). Moreover, z is differentiable almost everywhere on $[0, T]$ and the inequality $\|\dot{z}(t)\| \leq L\|\dot{x}(t)\|$ holds for almost every $t \in [0, T]$.

The following two results, which can be interpreted as continuous versions of the quasi-Fejér monotonicity for sequences, will play an important role in the asymptotic analysis of the trajectories of the dynamical system investigated in this paper. For their proofs we refer the reader to [2, Lemma 5.1] and [2, Lemma 5.2], respectively.

Lemma 3 *Suppose that $F : [0, +\infty) \to \mathbb{R}$ is locally absolutely continuous and bounded from 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 $\lim_{t\to+\infty} F(t) = 0$.

Further we recall a differentiability result involving the composition of convex functions with absolutely continuous trajectories which is due to Brézis ([31, Lemme 3.3, p. 73]; see also [14, Lemma 3.2]).

Lemma 5 *Let $f : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper, convex and lower semicontinuous function. Let $x \in L^2([0,T],\mathbb{R}^n)$ 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 $t \mapsto f(x(t))$ is absolutely continuous and for almost every t such that $x(t) \in \operatorname{dom}\partial f$ we have*

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

We close this section by recalling the following characterization of the proximal point operator of a proper, convex and lower semincontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. For every $\eta > 0$ it holds (see for example [18])

$$p = \operatorname{prox}_{\eta f}(x) \text{ if and only if } x \in p + \eta\partial f(p), \tag{7}$$

where ∂f denotes the convex subdifferential of f.

3 Asymptotic analysis

Before starting with the convergence analysis for the dynamical system (2), we would like to point out that this can be written as

$$\begin{cases} \dot{x}(t) = (\operatorname{prox}\circ(\operatorname{Id} - \eta\nabla g) - \operatorname{Id})\big(x(t)\big), \\ x(0) = x_0, \end{cases} \tag{8}$$

where $\operatorname{prox}\circ(\operatorname{Id} - \eta\nabla g) - \operatorname{Id}$ is a $(2 + \eta\beta)$-Lipschitz continuous operator. This follows from the fact that the proximal point operator of a proper, convex and lower semicontinuous function is nonexpansive, i.e., 1-Lipschitz continuous (see for example [18]). According to the global version of the Cauchy-Lipschitz Theorem (see for instance [9, Theorem 17.1.2(b)]), there exists a unique global solution $x \in C^1([0,+\infty),\mathbb{R}^n)$ of the above dynamical system.

3.1 Convergence of the trajectories

Lemma 6 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality*

$$\eta\beta(3 + \eta\beta) < 1. \tag{9}$$

For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0,+\infty),\mathbb{R}^n)$ be the unique global solution of (2). Then the following statements hold:

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

(b) $\exists \lim_{t\to+\infty}(f+g)\big(\dot{x}(t)+x(t)\big) \in \mathbb{R}$.

Proof. Let us start by noticing that in the light of the the reformulation in (8) and of Remark 2(b), $\dot{x}$ is locally absolutely continuous, hence $\ddot{x}$ exists and for almost every $t \in [0,+\infty)$ one has

$$\|\ddot{x}(t)\| \le (2+\eta\beta)\|\dot{x}(t)\|. \tag{10}$$

We fix an arbitrary $T > 0$. Due to the continuity properties of the trajectory on $[0,T]$, (10) and the Lipschitz continuity of ∇g, one has

$$x, \dot{x}, \ddot{x}, \nabla g(x) \in L^2([0,T];\mathbb{R}^n).$$

Further, from the characterization (7) of the proximal point operator we have

$$-\frac{1}{\eta}\dot{x}(t) - \nabla g(x(t)) \in \partial f(\dot{x}(t)+x(t)) \ \forall t \in [0,+\infty). \tag{11}$$

Applying Lemma 5 we obtain that the function $t \mapsto f\big(\dot{x}(t)+x(t)\big)$ is absolutely continuous and

$$\frac{d}{dt}f\big(\dot{x}(t)+x(t)\big) = \left\langle -\frac{1}{\eta}\dot{x}(t) - \nabla g(x(t)), \ddot{x}(t)+\dot{x}(t)\right\rangle$$

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

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

for almost every $t \in [0,T]$. Summing up the last two equalities, we obtain

$$\begin{aligned}
\frac{d}{dt}(f+g)\big(\dot{x}(t)+x(t)\big) &= \left\langle -\frac{1}{\eta}\dot{x}(t) - \nabla g(x(t)) + \nabla g\big(\dot{x}(t)+x(t)\big), \ddot{x}(t)+\dot{x}(t)\right\rangle \\
&= -\frac{1}{2\eta}\frac{d}{dt}\big(\|\dot{x}(t)\|^2\big) - \frac{1}{\eta}\|\dot{x}(t)\|^2 \\
&\quad + \big\langle \nabla g\big(\dot{x}(t)+x(t)\big) - \nabla g(x(t)), \ddot{x}(t)+\dot{x}(t)\big\rangle \\
&\le -\frac{1}{2\eta}\frac{d}{dt}\big(\|\dot{x}(t)\|^2\big) - \frac{1}{\eta}\|\dot{x}(t)\|^2 + \beta\|\dot{x}(t)\|\cdot\|\ddot{x}(t)+\dot{x}(t)\| \qquad (12) \\
&\le -\frac{1}{2\eta}\frac{d}{dt}\big(\|\dot{x}(t)\|^2\big) - \frac{1}{\eta}\|\dot{x}(t)\|^2 + \beta(3+\eta\beta)\|\dot{x}(t)\|^2 \qquad (13) \\
&= -\frac{1}{2\eta}\frac{d}{dt}\big(\|\dot{x}(t)\|^2\big) - \left[\frac{1}{\eta} - \beta(3+\eta\beta)\right]\|\dot{x}(t)\|^2
\end{aligned}$$

for almost every $t \in [0,T]$, where in (12) we used the Lipschitz continuity of ∇g and in (13) the inequality (10). Altogether, we conclude that for almost every $t \in [0,T]$ we have

$$\frac{d}{dt}\left[(f+g)\big(\dot{x}(t)+x(t)\big) + \frac{1}{2\eta}\|\dot{x}(t)\|^2\right] + \left[\frac{1}{\eta} - \beta(3+\eta\beta)\right]\|\dot{x}(t)\|^2 \le 0 \tag{14}$$

and by integration we get

$$\begin{aligned}
&(f+g)\big(\dot{x}(T)+x(T)\big) + \frac{1}{2\eta}\|\dot{x}(T)\|^2 + \left[\frac{1}{\eta} - \beta(3+\eta\beta)\right]\int_0^T \|\dot{x}(t)\|^2 dt \le \\
&(f+g)\big(\dot{x}(0)+x_0\big) + \frac{1}{2\eta}\|\dot{x}(0)\|^2. \qquad (15)
\end{aligned}$$

By using (9) and the fact that $f+g$ is bounded from below and by taking into account that $T>0$ has been arbitrarily chosen, we obtain

$$\dot{x} \in L^2([0,+\infty);\mathbb{R}^n). \tag{16}$$

Due to (10), this further implies

$$\ddot{x} \in L^2([0,+\infty);\mathbb{R}^n). \tag{17}$$

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

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

By applying Lemma 4, it follows that $\lim_{t\to+\infty} \dot{x}(t) = 0$ and the proof of (a) is complete. From (14), (9) and by using that $T>0$ has been arbitrarily chosen, we get

$$\frac{d}{dt}\left[(f+g)\big(\dot{x}(t)+x(t)\big) + \frac{1}{2\eta}\|\dot{x}(t)\|^2\right] \leq 0$$

for almost every $t \in [0,+\infty)$. From Lemma 3 it follows that

$$\lim_{t\to+\infty}\left[(f+g)\big(\dot{x}(t)+x(t)\big) + \frac{1}{2\eta}\|\dot{x}(t)\|^2\right]$$

exists and it is a real number, hence from $\lim_{t\to+\infty}\dot{x}(t)=0$ the conclusion follows. ■

We define the limit set of x as

$$\omega(x) = \{\overline{x} \in \mathbb{R}^n : \exists t_k \to +\infty \text{ such that } x(t_k) \to \overline{x} \text{ as } k \to +\infty\}.$$

Lemma 7 *Suppose that $f+g$ is bounded from below and $\eta>0$ fulfills the inequality* (9)*. For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0,+\infty),\mathbb{R}^n)$ be the unique global solution of* (2)*. Then*

$$\omega(x) \subseteq \operatorname{crit}(f+g).$$

Proof. Let $\overline{x} \in \omega(x)$ and $t_k \to +\infty$ be such that $x(t_k) \to \overline{x}$ as $k \to +\infty$. From (11) we have

$$\begin{aligned}-\frac{1}{\eta}\dot{x}(t_k) - \nabla g(x(t_k)) + \nabla g\big(\dot{x}(t_k)+x(t_k)\big) &\in \partial f\big(\dot{x}(t_k)+x(t_k)\big) + \nabla g\big(\dot{x}(t_k)+x(t_k)\big)\\ &= \partial(f+g)\big(\dot{x}(t_k)+x(t_k)\big)\ \forall k \in \mathbb{N}.\end{aligned} \tag{18}$$

Lemma 6(a) and the Lipschitz continuity of ∇g ensure that

$$-\frac{1}{\eta}\dot{x}(t_k) - \nabla g(x(t_k)) + \nabla g\big(\dot{x}(t_k)+x(t_k)\big) \to 0 \text{ as } k \to +\infty \tag{19}$$

and

$$\dot{x}(t_k)+x(t_k) \to \overline{x} \text{ as } k \to +\infty. \tag{20}$$

We claim that

$$\lim_{k\to+\infty}(f+g)\big(\dot{x}(t_k)+x(t_k)\big) = (f+g)(\overline{x}). \tag{21}$$

Due to the lower semicontinuity of f it holds

$$\liminf_{k\to+\infty} f\big(\dot{x}(t_k)+x(t_k)\big) \geq f(\overline{x}). \tag{22}$$

Further, since

$$\begin{aligned}\dot{x}(t_k)+x(t_k) &= \operatorname*{argmin}_{u\in\mathbb{R}^n}\left[f(u)+\frac{1}{2\eta}\left\|u-\big(x(t_k)-\eta\nabla g(x(t_k))\big)\right\|^2\right]\\ &= \operatorname*{argmin}_{u\in\mathbb{R}^n}\left[f(u)+\frac{1}{2\eta}\|u-x(t_k)\|^2+\langle u-x(t_k),\nabla g(x(t_k))\rangle\right]\end{aligned}$$

we have the inequality

$$\begin{aligned}& f\big(\dot{x}(t_k)+x(t_k)\big)+\frac{1}{2\eta}\|\dot{x}(t_k)\|^2+\langle \dot{x}(t_k),\nabla g(x(t_k))\rangle\\ \leq\ & f(\overline{x})+\frac{1}{2\eta}\|\overline{x}-x(t_k)\|^2+\langle \overline{x}-x(t_k),\nabla g(x(t_k))\rangle \ \forall k\in\mathbb{N}.\end{aligned}$$

Taking the limit as $k\to+\infty$ we derive by using again Lemma 6(a) that

$$\limsup_{k\to+\infty} f\big(\dot{x}(t_k)+x(t_k)\big) \leq f(\overline{x}),$$

which combined with (22) implies

$$\lim_{k\to+\infty} f\big(\dot{x}(t_k)+x(t_k)\big) = f(\overline{x}).$$

By using (20) and the continuity of g we conclude that (21) is true.

Altogether, from (18), (19), (20), (21) and the closedness criteria of the limiting subdifferential we obtain $0\in\partial(f+g)(\overline{x})$ and the proof is complete. ■

Lemma 8 *Suppose that $f+g$ is bounded from below and $\eta>0$ fulfills the inequality* (9). *For $x_0\in\mathbb{R}^n$, let $x\in C^1([0,+\infty),\mathbb{R}^n)$ be the unique global solution of* (2) *and consider the function*

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

Then the following statements are true:

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

$$\frac{d}{dt}H\big(\dot{x}(t)+x(t),x(t)\big) \leq -\left[\frac{1}{\eta}-(3+\eta\beta)\beta\right]\|\dot{x}(t)\|^2 \leq 0$$

and

$$\exists \lim_{t\to+\infty} H\big(\dot{x}(t)+x(t),x(t)\big)\in\mathbb{R};$$

(H_2) for every $t\in[0,+\infty)$ it holds

$$z(t) := \left(-\nabla g(x(t))+\nabla g\big(\dot{x}(t)+x(t)\big), -\frac{1}{\eta}\dot{x}(t)\right)\in\partial H\big(\dot{x}(t)+x(t),x(t)\big)$$

and

$$\|z(t)\| \leq \left(\beta+\frac{1}{\eta}\right)\|\dot{x}(t)\|;$$

(H_3) for $\overline{x} \in \omega(x)$ and $t_k \to +\infty$ such that $x(t_k) \to \overline{x}$ as $k \to +\infty$, we have $H\big(\dot{x}(t_k) + x(t_k), x(t_k)\big) \to H(\overline{x}, \overline{x})$ as $k \to +\infty$.

Proof. (H1) follows from Lemma 6. The first statement in (H2) is a consequence of (11) and the relation

$$\partial H(u, v) = \big(\partial(f + g)(u) + \eta^{-1}(u - v)\big) \times \{\eta^{-1}(v - u)\} \ \forall (u, v) \in \mathbb{R}^n \times \mathbb{R}^n, \tag{23}$$

while the second one is a consequence of the Lipschitz continuity of ∇g. Finally, (H3) has been shown as intermediate step in the proof of Lemma 7. ■

Lemma 9 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality (9). For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (2) and consider the function*

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

Suppose that x is bounded. Then the following statements are true:

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

(b) $\lim_{t\to+\infty} \operatorname{dist}\Big(\big(\dot{x}(t) + x(t), x(t)\big), \omega\big(\dot{x} + x, x\big)\Big) = 0$;

(c) $\omega\big(\dot{x} + x, x\big)$ is nonempty, compact and connected;

(d) H is finite and constant on $\omega\big(\dot{x} + x, x\big)$.

Proof. (a), (b) and (d) are direct consequences Lemma 6, Lemma 7 and Lemma 8.

Finally, (c) is a classical result from [34]. 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 $\lim_{t\to+\infty} \dot{x}(t) = 0$). ■

Remark 10 Suppose that $\eta > 0$ fulfills the inequality (9) and $f + g$ is coercive, that is

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

For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (2). Then $f + g$ is bounded from below and x is bounded.

Indeed, since $f + g$ is a proper, lower semicontinuous and coercive function, it follows that $\inf_{u\in\mathbb{R}^n}[f(u) + g(u)]$ is finite and the infimum is attained. Hence $f + g$ is bounded from below. On the other hand, from (15) it follows

$$\begin{aligned}(f + g)\big(\dot{x}(T) + x(T)\big) &\le (f + g)\big(\dot{x}(T) + x(T)\big) + \frac{1}{2\eta}\|\dot{x}(T)\|^2 \\ &\le (f + g)\big(\dot{x}(0) + x_0)\big) + \frac{1}{2\eta}\|\dot{x}(0)\|^2 \ \forall T \ge 0.\end{aligned}$$

Since the lower level sets of $f + g$ are bounded, the above inequality yields the boundedness of $\dot{x} + x$, which combined with $\lim_{t\to+\infty} \dot{x}(t) = 0$ delivers the boundedness of x.

We come now to the main result of the paper.

Theorem 11 *Suppose that $f+g$ is bounded from below and $\eta > 0$ fulfills the inequality (9). For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0,+\infty),\mathbb{R}^n)$ be the unique global solution of (2) and consider the function*

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

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

Proof. According to Lemma 9, we can choose an element $\overline{x} \in \operatorname{crit}(f+g)$ such that $(\overline{x},\overline{x}) \in \omega(\dot{x}+x,x)$. According to Lemma 8, it follows that

$$\lim_{t\to+\infty} H\big(\dot{x}(t)+x(t),x(t)\big) = H(\overline{x},\overline{x}).$$

We treat the following two cases separately.

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

$$H\big(\dot{x}(\overline{t})+x(\overline{t}),x(\overline{t})\big) = H(\overline{x},\overline{x}).$$

Since from Lemma 8(H1) we have

$$\frac{d}{dt}H\big(\dot{x}(t)+x(t),x(t)\big) \leq 0\ \forall t \in [0,+\infty),$$

we obtain for every $t \geq \overline{t}$ that

$$H\big(\dot{x}(t)+x(t),x(t)\big) \leq H\big(\dot{x}(\overline{t})+x(\overline{t}),x(\overline{t})\big) = H(\overline{x},\overline{x}).$$

Thus $H\big(\dot{x}(t)+x(t),x(t)\big) = H(\overline{x},\overline{x})$ for every $t \geq \overline{t}$. This yields by Lemma 8(H1) that $\dot{x}(t) = 0$ for almost every $t \in [\overline{t},+\infty)$, hence x is constant on $[\overline{t},+\infty)$ and the conclusion follows.

II. For every $t \geq 0$ it holds $H\big(\dot{x}(t)+x(t),x(t)\big) > H(\overline{x},\overline{x})$. Take $\Omega = \omega(\dot{x}+x,x)$.

In virtue of Lemma 9(c) and (d) and since H is a KL function, by Lemma 1, there exist positive numbers ϵ and η and a concave function $\varphi \in \Theta_\eta$ such that for all

$$\begin{aligned}(x,y) \in &\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^n : \operatorname{dist}((u,v),\Omega) < \epsilon\} \\ &\cap \{(u,v) \in \mathbb{R}^n \times \mathbb{R}^n : H(\overline{x},\overline{x}) < H(u,v) < H(\overline{x},\overline{x}) + \eta\}\end{aligned} \tag{24}$$

one has

$$\varphi'(H(x,y) - H(\overline{x},\overline{x}))\operatorname{dist}((0,0),\partial H(x,y)) \geq 1. \tag{25}$$

Let $t_1 \geq 0$ be such that $H\big(\dot{x}(t)+x(t),x(t)\big) < H(\overline{x},\overline{x}) + \eta$ for all $t \geq t_1$. Since $\lim_{t\to+\infty} \operatorname{dist}\Big(\big(\dot{x}(t)+x(t),x(t)\big),\Omega\Big) = 0$, there exists $t_2 \geq 0$ such that $\operatorname{dist}\Big(\big(\dot{x}(t) +$

$x(t), x(t)), \Omega\Big) < \epsilon$ for all $t \geq t_2$. Hence for all $t \geq T := \max\{t_1, t_2\}$, $\big(\dot{x}(t) + x(t), x(t)\big)$ belongs to the intersection in (24). Thus, according to (25), for every $t \geq T$ we have

$$\varphi'\Big(H\big(\dot{x}(t) + x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big) \operatorname{dist}\Big((0,0), \partial H\big(\dot{x}(t) + x(t), x(t)\big)\Big) \geq 1. \tag{26}$$

By applying Lemma 8(H2) we obtain for almost every $t \in [T, +\infty)$

$$(\beta + \eta^{-1})\|\dot{x}(t)\|\varphi'\Big(H\big(\dot{x}(t) + x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big) \geq 1. \tag{27}$$

From here, by using Lemma 8(H1) and that $\varphi' > 0$ and

$$\frac{d}{dt}\varphi\Big(H\big(\dot{x}(t) + x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big) =$$
$$\varphi'\Big(H\big(\dot{x}(t) + x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big)\frac{d}{dt}H\big(\dot{x}(t) + x(t), x(t)\big),$$

we deduce that for almost every $t \in [T, +\infty)$ it holds

$$\frac{d}{dt}\varphi\Big(H\big(\dot{x}(t) + x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big) \leq -\big(\beta + \eta^{-1}\big)^{-1}\left[\frac{1}{\eta} - (3 + \eta\beta)\beta\right]\|\dot{x}(t)\|. \tag{28}$$

Since φ is bounded from below, by taking into account (9), it follows $\dot{x} \in L^1([0, +\infty); \mathbb{R}^n)$. From here we obtain that $\lim_{t\to+\infty} x(t)$ exists and this closes the proof. ■

Since the class of semi-algebraic functions is closed under addition (see for example [24]) and $(u, v) \mapsto c\|u - v\|^2$ is semi-algebraic for $c > 0$, we can stat the following direct consequence of the previous theorem.

Corollary 12 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality (9). For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (2). Suppose that x is bounded and $f + g$ is semi-algebraic. Then the following statements are true:*

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

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

Remark 13 The construction of the function H, which we used in the above arguments in order to derive a descent property, has been inspired by the decrease property obtained in (14). Similar regularizations of the objective function of (1) have been considered also in [30, 40], in the context of the investigation of non-relaxed forward-backward methods involving inertial and memory effects in the nonconvex setting.

Remark 14 Some comments concerning condition (9) that involves the step size and the Lipschitz constant of the smooth function are in order. In the full convex setting, it has been proved by Abbas and Attouch in [1, Theorem 5.2(2)] that in order to obtain weak convergence of the trajectory to an optimal solution of (1) no restriction on the step size is needed. This surprising fact which is closely related to the continuous case, as for obtaining convergence for the time-discretized version (4) one usually has to assume that η is in the interval $(0, 2\beta)$. Working in the nonconvex setting, one has to even strengthen this assumption and it will be a question of future research to find out if one can relax the restriction (9). Another question of future interest will be to properly choose the step size when the Lipschitz constant of the gradient is not known, eventually by making use of backtracking strategies, as is was already done in the convex setting.

3.2 Convergence rates

In this subsection we investigate the convergence rates of the trajectories generated by the dynamical system (2). When solving optimization problems involving KL functions, convergence rates have been proved to depend on the so-called Łojasiewicz exponent (see [11,21,33,38]). The main result of this subsection refers to the KL functions which satisfy Definition 1 for $\varphi(s) = Cs^{1-\theta}$, where $C > 0$ and $\theta \in (0,1)$. We recall the following definition considered in [11].

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

$$|f(x) - f(\overline{x})|^{\theta} \le C\|x^*\| \text{ for every } x \text{ fulfilling } \|x - \overline{x}\| < \varepsilon \text{ and every } x^* \in \partial f(x). \tag{29}$$

According to [12, Lemma 2.1 and Remark 3.2(b)], the KL property is automatically satisfied at any noncritical point, fact which motivates the restriction to critical points in the above definition. The real number θ in the above definition is called *Łojasiewicz exponent* of the function f at the critical point $\overline{x}$.

Theorem 15 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality* (9)*. For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0,+\infty), \mathbb{R}^n)$ be the unique global solution of* (2) *and consider the function*

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

Suppose that x is bounded and H satisfies Definition 1 for $\varphi(s) = Cs^{1-\theta}$, where $C > 0$ and $\theta \in (0,1)$. Then there exists $\overline{x} \in \operatorname{crit}(f+g)$ such that $\lim_{t\to+\infty} x(t) = \overline{x}$. Let θ be the Lojasiewicz exponent of H at $(\overline{x},\overline{x}) \in \operatorname{crit} H$, according to the Definition 3. Then there exist $a, b, c, d > 0$ and $t_0 \ge 0$ such that for every $t \ge t_0$ 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 $\|x(t) - \overline{x}\| \le a\exp(-bt)$;*

(c) *if $\theta \in (\frac{1}{2}, 1)$, then $\|x(t) - \overline{x}\| \le (ct + d)^{-\left(\frac{1-\theta}{2\theta-1}\right)}$.*

Proof. We define for every $t \ge 0$ (see also [21])

$$\sigma(t) = \int_t^{+\infty} \|\dot{x}(s)\| ds \ \text{ for all } t \ge 0.$$

It is immediate that

$$\|x(t) - \overline{x}\| \le \sigma(t) \ \forall t \ge 0. \tag{30}$$

Indeed, this follows by noticing that for $T \ge t$

$$\begin{aligned}\|x(t) - \overline{x}\| &= \|x(T) - \overline{x} - \int_t^T \dot{x}(s)ds\| \\ &\le \|x(T) - \overline{x}\| + \int_t^T \|\dot{x}(s)\|ds,\end{aligned}$$

and by letting afterwards $T \to +\infty$.

We assume that for every $t \geq 0$ we have $H\big(\dot{x}(t)+x(t), x(t)\big) > H(\overline{x}, \overline{x})$. As seen in the proof of Theorem 11, in the other case the conclusion follows automatically. Furthermore, by invoking again the proof of above-named result, there exist $t_0 \geq 0$ and $M > 0$ such that for almost every $t \geq t_0$ (see (28))

$$M\|\dot{x}(t)\| + \frac{d}{dt}\left[\Big(H\big(\dot{x}(t)+x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big)\right]^{1-\theta} \leq 0$$

and

$$\|\big(\dot{x}(t)+x(t), x(t)\big) - (\overline{x}, \overline{x})\| < \varepsilon.$$

We derive by integration (for $T \geq t \geq t_0$)

$$M\int_t^T \|\dot{x}(s)\|ds + \left[\Big(H\big(\dot{x}(T)+x(T), x(T)\big) - H(\overline{x}, \overline{x})\Big)\right]^{1-\theta}$$

$$\leq \left[\Big(H\big(\dot{x}(t)+x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big)\right]^{1-\theta},$$

hence

$$M\sigma(t) \leq \left[\Big(H\big(\dot{x}(t)+x(t), x(t)\big) - H(\overline{x}, \overline{x})\Big)\right]^{1-\theta} \ \forall t \geq t_0. \tag{31}$$

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

$$|H\big(\dot{x}(t)+x(t), x(t)\big) - H(\overline{x}, \overline{x})|^{\theta} \leq C\|x^*\| \ \forall x^* \in \partial H\big(\dot{x}(t)+x(t), x(t)\big)$$

for every $t \geq t_0$. According to Lemma 8(H2), we can find a constant $N > 0$ and $x^*(t) \in \partial H\big(\dot{x}(t)+x(t), x(t)\big)$ such that for every $t \in [0, +\infty)$

$$\|x^*(t)\| \leq N\|\dot{x}(t)\|.$$

From the above two inequalities we derive for almost every $t \in [t_0, +\infty)$

$$|H\big(\dot{x}(t)+x(t), x(t)\big) - H(\overline{x}, \overline{x})|^{\theta} \leq C \cdot N\|\dot{x}(t)\|,$$

which combined with (31) yields

$$M\sigma(t) \leq (C \cdot N\|\dot{x}(t)\|)^{\frac{1-\theta}{\theta}}. \tag{32}$$

Since

$$\dot{\sigma}(t) = -\|\dot{x}(t)\| \tag{33}$$

we conclude that there exists $\alpha > 0$ such that for almost every $t \in [t_0, +\infty)$

$$\dot{\sigma}(t) \leq -\alpha\big(\sigma(t)\big)^{\frac{\theta}{1-\theta}}. \tag{34}$$

If $\theta = \frac{1}{2}$, then

$$\dot{\sigma}(t) \leq -\alpha\sigma(t)$$

for almost every $t \in [t_0, +\infty)$. By multiplying with $\exp(\alpha t)$ and integrating afterwards from t_0 to t, it follows that there exist $a, b > 0$ such that

$$\sigma(t) \leq a\exp(-bt) \ \forall t \geq t_0$$

and the conclusion of (b) is immediate from (30).

Assume that $0 < \theta < \frac{1}{2}$. We obtain from (34)

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

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

By integration we get

$$\big(\sigma(t)\big)^{\frac{1-2\theta}{1-\theta}} \leq -\overline{\alpha}t + \overline{\beta}\ \forall t \geq t_0,$$

where $\overline{\alpha} > 0$. Thus there exists $T \geq 0$ such that

$$\sigma(T) \leq 0\ \forall t \geq T,$$

which implies that x is constant on $[T, +\infty)$.

Finally, suppose that $\frac{1}{2} < \theta < 1$. We obtain from (34)

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

By integration one derives

$$\sigma(t) \leq (ct+d)^{-\left(\frac{1-\theta}{2\theta-1}\right)}\ \forall t \geq t_0,$$

where $c, d > 0$, and (c) follows from (30). ■

Acknowledgement We are grateful to the anonymous referees for suggestions which improved the presentation of the manuscript.

References

[1] B. Abbas, H. Attouch, *Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, Optimization 64(10), 2223–2252, 2015

[2] B. Abbas, H. Attouch, B.F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and its Applications 161(2), 331–360, 2014

[3] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM Journal on Control and Optimization 38(4), 1102–1119, 2000

[4] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM Journal on Optimization 14(3), 773–782, 2004

[5] F. Alvarez, H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis 9(1-2), 311, 2001

[6] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, Journal de Mathématiques Pures et Appliquées (9) 81(8), 747–779, 2002

[7] A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differentsial'nye Uravneniya 30(9), 1475–1486, 1994; translation in Differential Equations 30(9), 1365–1375, 1994

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

[9] H. Attouch, G. Buttazzo, G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization, Second Edition*, MOS-SIAM Series on Optimization, Philadelphia, 2014

[10] 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 Analyis 23(1), 139–180, 2016

[11] 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

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

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

[15] 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

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

[17] S. Banert, R.I. Boţ, *A forward-backward-forward differential equation and its asymptotic properties*, Journal of Convex Analysis 25(2), 2018

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

[19] H.H. Bauschke, P.L. Combettes, D.R. Luke, *Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization*, Journal of the Optical Society of America A. Optics, Image Science, and Vision 19(7), 1334–1345, 2002

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

[21] 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

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

[23] 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

[24] 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

[25] R.I. Boţ, E.R. Csetnek, *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems*, Journal of Optimization Theory and Applications 171(2), 600–616, 2016

[26] 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

[27] 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

[28] 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

[29] R.I. Boţ, E.R. Csetnek, *Convergence rates for forward-backward dynamical systems associated with strongly monotone inclusions*, Journal of Mathematical Analysis and Applications, DOI: 10.1016/j.jmaa.2016.07.007, 2016

[30] 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

[31] 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

[32] E. Chouzenoux, J.-C. Pesquet, A. Repetti, *Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function*, Journal of Optimization Theory and its Applications 162(1), 107–132, 2014

[33] P. Frankel, G. Garrigos, J. Peypouquet, *Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates*, Journal of Optimization Theory and its Applications 165(3), 874–900, 2015

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

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

[36] R. Hesse, D.R. Luke, S. Sabach, M.K. Tam, *Proximal heterogeneous block input-output method and application to blind ptychographic diffraction imaging*, SIAM Journal on Imaging Sciences 8(1), 426–457, 2015

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

[38] 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

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

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

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

[42] L. Simon, *Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems*, Annals of Mathematics (2) 118, 525–571, 1983