



Continuous Dynamics Related to Monotone Inclusions and Non-Smooth Optimization Problems

Ernö Robert Csetnek¹

Received: 16 March 2020 / Accepted: 30 June 2020 / Published online: 13 July 2020
© The Author(s) 2020

Abstract

The aim of this survey is to present the main important techniques and tools from variational analysis used for first and second order dynamical systems of implicit type for solving monotone inclusions and non-smooth optimization problems. The differential equations are expressed by means of the resolvent (in case of a maximally monotone set valued operator) or the proximal operator for non-smooth functions. The asymptotic analysis of the trajectories generated relies on Lyapunov theory, where the appropriate energy functional plays a decisive role. While the most part of the paper is related to monotone inclusions and convex optimization problems in the variational case, we present also results for dynamical systems for solving non-convex optimization problems, where the Kurdyka-Łojasiewicz property is used.

Keywords Dynamical systems · Lyapunov analysis · Krasnosel’skiĭ–Mann algorithm · Monotone inclusions · Resolvent · Proximal operator · Forward-backward algorithm · Non-smooth optimization problem · Kurdyka-Łojasiewicz property

Mathematics Subject Classification (2010) 34G25 · 47J25 · 47H05 · 90C25

1 Introduction and preliminaries

Dynamical systems approaching monotone inclusions and optimization problems enjoy much attention since the seventies of the last century (Brézis, Baillon and Bruck, see [25, 52, 53]), not only due to their intrinsic importance in areas like differential equations and applied functional analysis, but also because they have been recognized as a valuable tool for discovering and studying numerical algorithms for optimization problems obtained by time discretization of the continuous dynamics. The dynamic approach to iterative methods in optimization can furnish deep insights into the expected behavior of the method and the

Research supported by FWF (Austrian Science Fund), project P 29809-N32.

✉ Ernő Robert Csetnek
erno.e.robert.csetnek@univie.ac.at

¹ Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria

techniques used in the continuous case can be adapted to obtain results for the discrete algorithm. For more on the relations between the continuous and discrete dynamics we refer the reader to [75].

Let us mention that the discretization $\dot{x}(t) \approx \frac{1}{\gamma}(x_{n+1} - x_n)$ (with $\gamma > 0$) of the gradient flow

$$\dot{x}(t) = -\nabla g(x(t)),$$

where $g : \mathcal{H} \rightarrow \mathbb{R}$ is a smooth function, leads to the well known steepest descent method (gradient method)

$$x_{n+1} = x_n - \gamma \nabla g(x_n),$$

used for solving the minimization problem

$$\min_{x \in \mathcal{H}} g(x).$$

In the case of convex non-smooth optimization problems, the discretization of the differential inclusion

$$\dot{x}(t) \in -\partial f(x(t))$$

leads to the subgradient method $x_{n+1} \in x_n - \gamma \partial f(x_n)$, or, when we use the discretization $x_{n+1} - x_n \in \gamma \partial f(x_{n+1})$, we obtain the proximal point algorithm [79]

$$x_{n+1} = (\text{Id} + \gamma \partial f)^{-1}(x_n).$$

Let us proceed and consider continuous *implicit-type dynamical systems* associated with monotone inclusions/optimization problems, which are ordinary differential equations formulated via resolvents of maximally monotone operators. In [32], Bolte studied the convergence of the trajectories of the following dynamical system

$$\dot{x}(t) + x(t) = \text{proj}_C(x(t) - \gamma \nabla g(x(t))) \quad (1)$$

where C is a nonempty, closed and convex subset of \mathcal{H} , $x_0 \in \mathcal{H}$, and proj_C denotes the projection operator on the set C . It has been shown in the convex setting that the trajectory of (1) converges weakly to a minimizer of the optimization problem

$$\inf_{x \in C} g(x), \quad (2)$$

provided the latter is solvable. We refer also to the work of Antipin [5] for further statements and results concerning (1).

The following generalization of the dynamical system (1) has been considered by Abbas and Attouch in [1, Section 5.2]:

$$\dot{x}(t) + x(t) = \text{prox}_{\gamma f}(x(t) - \gamma B(x(t))), \quad (3)$$

where B is a cocoercive operator. According to [1], in case $\text{zer}(\partial f + B) \neq \emptyset$, the weak asymptotical convergence of the orbit x of (3) to an element in $\text{zer}(\partial f + B) \neq \emptyset$ is ensured by choosing the step-size γ in a suitable domain bounded by the parameter of cocoercivity of the operator B .

Let us also mention that dynamical systems of implicit type have been considered in the literature also by Attouch and Svaiter in [24], Attouch, Abbas and Svaiter in [2] and Attouch, Alvarez and Svaiter in [21].

For the minimization of the smooth and convex function $g : \mathcal{H} \rightarrow \mathbb{R}$ over the nonempty, convex and closed set $C \subseteq \mathcal{H}$, a continuous in time second order gradient-projection approach has been considered in [5, 7], having as starting point the dynamical system

$$\ddot{x}(t) + \gamma \dot{x}(t) + x(t) = \text{proj}_C(x(t) - \eta \nabla g(x(t))), \quad (4)$$

with constant damping parameter $\gamma > 0$ and constant step size $\eta > 0$. The system (4) becomes in case $C = \mathcal{H}$ the "heavy ball method", sometimes called also "heavy ball method with friction". This nonlinear oscillator with damping is in case $\mathcal{H} = \mathbb{R}^2$ a simplified version of the differential system describing the motion of a heavy ball that rolls over the graph of g and that keep rolling under its own inertia until friction stop it at a critical point of g (see [17]). Later, (see [13, 86] and the references therein) it was highlighted the fact that a vanishing in time damping is more suitable in the context of second order dynamics, since this will induce fast convergence results in comparison to first order dynamics.

Second order dynamics provide through discretization numerical schemes with inertial terms, where the new iterate is computed based on the previous two ones. Taking into account the history of the iterates makes the asymptotic analysis more involved and may provide better convergence rates. It is an interesting topic to take into consideration more than two iterates in order to compute the next one. In case of algorithms we refer the reader to the recent contribution [56] and to [15] for a third order gradient evolution system. In this survey we are considering first and second order dynamics.

Finally, motivated by concrete applications where non-convex objects appear, let us say a few words about this setting. In the context of minimizing a non-convex smooth function, several first- and second-order gradient type dynamical systems have been investigated by Łojasiewicz [68], Simon [83], Haraux and Jendoubi [65], Alvarez, Attouch, Bolte and Redont [4, Section 4], Bolte, Daniilidis and Lewis [33, Section 4], etc. In the aforementioned papers, the convergence of the trajectories is obtained in the framework of functions satisfying the Kurdyka-Łojasiewicz property, see Section 4 for details.

1.1 Contents of the Paper

In this manuscript we revisit the most important first and second-order dynamical systems of implicit type studied in the context of solving monotone inclusions and non-smooth optimization problems (convex and non-convex). The main focus is on the techniques which are used in the investigation of the asymptotic behaviour of the trajectories generated. Also the connection to discrete schemes will be also underlined in order to see the algorithmic impact of the dynamical systems considered. Moreover, the proofs are not in every stage of this survey complete, but we will always refer to appropriate references when something is missing. The existence and uniqueness of the trajectory usually follows from the Cauchy-Lipschitz-Picard Theorem and its variants and will be omitted in the following. However, the reader will find the corresponding details in the literature mentioned here. Let us also mention that there are no new results in this survey, apart from Theorem 15, which is a generalization to the continuous setting of the $O\left(\frac{1}{n}\right)$ rate for objective function values for prox-gradient schemes with no acceleration (see also ISTA in [29]).

The main challenge in the analysis of the dynamical systems is defining an appropriate energy functional in order to conduct the investigations in the framework of Lyapunov theory. We start with first order dynamics governed by a nonexpansive operator with the aim of approaching the set of its fixed points [39]. While at a first look, this seems to be quite far from monotone inclusions or convex optimization problems, let us underline that this offers a setting in which also the aforementioned problems can be embedded. In this setting the energy functional is usually expressed by means of the distance of the trajectory to a solution of the problem under consideration. The continuous version of the Opial Lemma plays a decisive role on the analysis. As particular cases we mention the dynamical systems

of forward-backward type, forward-backward-forward or Douglas-Rachford. The case of convex optimization problems is also considered.

Afterwards we continue with second-order dynamical systems [38]. These are motivated by the fact that by time discretization inertial terms are induced in the algorithmic schemes which are responsible for acceleration (see also the works of Polyak [76], Nesterov [71, 72], Bertsekas [30] in the discrete case). Also in this setting we start with a very general dynamical system formulated by means of a cocoercive operator and then consider particular cases for monotone inclusions and convex optimization problems. Relations to other types of dynamics considered in the literature are also mentioned.

Finally, we consider dynamical systems for non-convex and non-smooth optimization problems [41]. How to find the appropriate energy functional in this context is not trivial. The approach for proving asymptotic convergence for the trajectory towards a critical point of the objective function, uses three main ingredients (see [4] for the continuous case and also [10, 36] 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.

1.2 Notations

Let us fix a few notations used throughout the paper. Let $\mathbb{N} = \{0, 1, 2, \dots\}$ be the set of nonnegative integers. Let \mathcal{H} be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

For readers convenience let us recall some standard notions and results in monotone operator theory which will be used in the following (see also [28, 37, 84]). For an arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\text{Gr } A = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its graph. We use also the notation $\text{zer } A = \{x \in \mathcal{H} : 0 \in Ax\}$ for the set of zeros of A . We say that A is monotone, if $\langle x - y, u - v \rangle \geq 0$ for all $(x, u), (y, v) \in \text{Gr } A$. A monotone operator A is said to be maximally monotone, if there exists no proper monotone extension of the graph of A on $\mathcal{H} \times \mathcal{H}$. The resolvent of A , $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by $J_A = (\text{Id}_{\mathcal{H}} + A)^{-1}$, where $\text{Id}_{\mathcal{H}} : \mathcal{H} \rightarrow \mathcal{H}$, $\text{Id}_{\mathcal{H}}(x) = x$ for all $x \in \mathcal{H}$, is the identity operator on \mathcal{H} . Moreover, if A is maximally monotone, then $J_A : \mathcal{H} \rightarrow \mathcal{H}$ is single-valued and maximally monotone (see [28, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [28, Proposition 23.2])

$$p \in J_{\gamma A} x \text{ if and only if } (p, \gamma^{-1}(x - p)) \in \text{Gr } A. \quad (5)$$

The operator A is said to be uniformly monotone if there exists an increasing function $\phi_A : [0, +\infty) \rightarrow [0, +\infty)$ that vanishes only at 0, and $\langle x - y, u - v \rangle \geq \phi_A(\|x - y\|)$ for every $(x, u) \in \text{Gr } A$ and $(y, v) \in \text{Gr } A$. If this property is fulfilled in the special case $\phi_A(t) = \gamma t^2$, with $\gamma > 0$, we say that A is γ -strongly monotone. We consider also the class of cocoercive operators: the single valued mapping $B : \mathcal{H} \rightarrow \mathcal{H}$ is γ -cocoercive, if $\langle x - y, Bx - By \rangle \geq \gamma \|Bx - By\|^2$ for all $x, y \in \mathcal{H}$. Notice that this is nothing else than B^{-1} is γ -strongly monotone.

We recall some standard notations and facts in convex analysis. For a proper, convex and lower semicontinuous function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, its (convex) subdifferential at $x \in \mathcal{H}$ is defined as

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

When seen as a set-valued mapping, it is a maximally monotone operator (see [78]) and for $\gamma > 0$, its resolvent is given by $J_{\gamma \partial f} = \text{prox}_{\gamma f}$ (see [28]), where $\text{prox}_{\gamma f} : \mathcal{H} \rightarrow \mathcal{H}$,

$$\text{prox}_{\gamma f}(x) = \operatorname{argmin}_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma} \|y - x\|^2 \right\}, \tag{6}$$

denotes the proximal point operator of f . The function f is said to be ν -strongly convex, where $\nu > 0$, if $f - (\nu/2)\|\cdot\|^2$ is a convex function. Let us mention that if f is ν -strongly convex, then ∂f is ν -strongly monotone, see [28, Example 22.3(iv)].

2 First Order Dynamical Systems

Let $T : \mathcal{H} \rightarrow \mathcal{H}$ be a nonexpansive mapping (that is $\|Tx - Ty\| \leq \|x - y\|$ for all $x, y \in \mathcal{H}$), $\lambda : [0, +\infty) \rightarrow [0, 1]$ be a Lebesgue measurable function and $x_0 \in \mathcal{H}$. We are concerned with the following dynamical system:

$$\begin{cases} \dot{x}(t) = \lambda(t)(T(x(t)) - x(t)) \\ x(0) = x_0. \end{cases} \tag{7}$$

As in [2, 24], we consider the following definition of an absolutely continuous function.

Definition 1 (see, for instance, [2, 24]) A function $f : [0, b] \rightarrow \mathcal{H}$ (where $b > 0$) is said to be absolutely continuous if one of the following equivalent properties holds:

- (i) there exists an integrable function $g : [0, b] \rightarrow \mathcal{H}$ such that

$$f(t) = f(0) + \int_0^t g(s)ds \quad \forall t \in [0, b];$$

- (ii) f is continuous and its distributional derivative is Lebesgue integrable on $[0, b]$;
- (iii) for every $\varepsilon > 0$, there exists $\eta > 0$ such that for any finite family of intervals $I_k = (a_k, b_k) \subseteq [a, b]$ we have:

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

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 $f' = g$ by the integration formula (i).

- (b) If $f : [0, b] \rightarrow \mathcal{H}$ (where $b > 0$) is absolutely continuous and $B : \mathcal{H} \rightarrow \mathcal{H}$ is L -Lipschitz continuous (where $L \geq 0$), then the function $h = B \circ f$ is absolutely continuous. This can be easily verified by considering the characterization in Definition 1(iii). Moreover, h is almost everywhere differentiable and the inequality $\|h'(\cdot)\| \leq L\|f'(\cdot)\|$ holds almost everywhere.

We say that $f : [0, +\infty) \rightarrow \mathcal{H}$ is locally absolutely continuous if it is absolutely continuous on every compact interval $[0, b]$, with $b > 0$.

Definition 2 We say that $x : [0, +\infty) \rightarrow \mathcal{H}$ is a strong global solution of (7) if the following properties are satisfied:

- (i) $x : [0, +\infty) \rightarrow \mathcal{H}$ is locally absolutely continuous;
- (ii) $\dot{x}(t) = \lambda(t)(T(x(t)) - x(t))$ for all $t \in [0, +\infty)$;
- (iii) $x(0) = x_0$.

The existence and uniqueness of strong global solutions of (7) is a consequence of the Cauchy-Lipschitz theorem for absolutely continuous trajectories (see for example [64, Proposition 6.2.1], [85, Theorem 54]).

In order to proceed with the asymptotic analysis, we need the following preparatory results.

Lemma 2 [2, Lemma 5.1]) *Suppose that $F : [0, +\infty) \rightarrow \mathbb{R}$ 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 $\lim_{t \rightarrow \infty} F(t) \in \mathbb{R}$.

Lemma 3 [2, Lemma 5.2]) *If $1 \leq p < \infty$, $1 \leq r \leq \infty$, $F : [0, +\infty) \rightarrow [0, +\infty)$ is locally absolutely continuous, $F \in L^p([0, +\infty))$, $G : [0, +\infty) \rightarrow \mathbb{R}$, $G \in L^r([0, +\infty))$ and for almost all t*

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

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

The next result which we recall here is the continuous version of the Opial Lemma (see for example [2, Lemma 5.3], [1, Lemma 1.10]).

Lemma 4 *Let $S \subseteq \mathcal{H}$ be a nonempty set and $x : [0, +\infty) \rightarrow \mathcal{H}$ a given map. Assume that*

- (i) *for every $z \in S$, $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists;*
- (ii) *every weak sequential cluster point of the map x belongs to S .*

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

Theorem 5 *Let $T : \mathcal{H} \rightarrow \mathcal{H}$ be a nonexpansive mapping such that $\text{Fix } T \neq \emptyset$, $\lambda : [0, +\infty) \rightarrow [0, 1]$ a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that either*

$$\int_0^{+\infty} \lambda(t)(1 - \lambda(t))dt = +\infty \text{ or } \inf_{t \geq 0} \lambda(t) > 0.$$

Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (7). Then the following statements are true:

- (i) *the trajectory x is bounded and $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$;*
- (ii) *$\lim_{t \rightarrow +\infty} (T(x(t)) - x(t)) = 0$;*
- (iii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$;*
- (iv) *$x(t)$ converges weakly to a point in $\text{Fix } T$, as $t \rightarrow +\infty$.*

Proof We rely on Lyapunov analysis combined with the Opial Lemma. We take an arbitrary $y \in \text{Fix } T$. From the nonexpansiveness of T and [28, Corollary 2.14] we obtain:

$$\begin{aligned} \frac{d}{dt} \|x(t) - y\|^2 &= 2 \langle \dot{x}(t), x(t) - y \rangle = \|\dot{x}(t) + x(t) - y\|^2 - \|x(t) - y\|^2 - \|\dot{x}(t)\|^2 \\ &= \|\lambda(t)(T(x(t)) - y) + (1 - \lambda(t))(x(t) - y)\|^2 - \|x(t) - y\|^2 - \|\dot{x}(t)\|^2 \\ &= \lambda(t)\|T(x(t)) - y\|^2 + (1 - \lambda(t))\|x(t) - y\|^2 \\ &\quad - \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 - \|x(t) - y\|^2 - \|\dot{x}(t)\|^2 \\ &\leq -\lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 - \|\dot{x}(t)\|^2. \end{aligned}$$

Hence for all $t \geq 0$ we have that

$$\frac{d}{dt} \|x(t) - y\|^2 + \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 + \|\dot{x}(t)\|^2 \leq 0. \tag{8}$$

Since $\lambda(t) \in [0, 1]$ for all $t \geq 0$, from (8) it follows that $t \mapsto \|x(t) - y\|$ is decreasing, hence $\lim_{t \rightarrow +\infty} \|x(t) - y\|$ exists. From here we obtain the boundedness of the trajectory and by integrating (8) we deduce also that $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$ and

$$\int_0^{+\infty} \lambda(t)(1 - \lambda(t))\|T(x(t)) - x(t)\|^2 dt < +\infty, \tag{9}$$

thus (i) holds. Since $y \in \text{Fix } T$ has been chosen arbitrary, the first assumption in the continuous version of Opial Lemma is fulfilled.

We show in the following that $\lim_{t \rightarrow +\infty} \|T(x(t)) - x(t)\|$ exists and it is a real number. This is immediate if we show that the function $t \mapsto \frac{1}{2}\|T(x(t)) - x(t)\|^2$ is decreasing. According to Remark 1(b), the function $t \mapsto T(x(t))$ is almost everywhere differentiable and $\|\frac{d}{dt}T(x(t))\| \leq \|\dot{x}(t)\|$ holds for almost all $t \geq 0$. Moreover, by the first equation of (7) we have

$$\begin{aligned} \frac{d}{dt} \left(\frac{1}{2} \|T(x(t)) - x(t)\|^2 \right) &= \left\langle \frac{d}{dt} T(x(t)) - \dot{x}(t), T(x(t)) - x(t) \right\rangle \\ &= -\langle \dot{x}(t), T(x(t)) - x(t) \rangle + \left\langle \frac{d}{dt} T(x(t)), T(x(t)) - x(t) \right\rangle \\ &= -\lambda(t)\|T(x(t)) - x(t)\|^2 + \left\langle \frac{d}{dt} T(x(t)), T(x(t)) - x(t) \right\rangle \\ &\leq -\lambda(t)\|T(x(t)) - x(t)\|^2 + \|\dot{x}(t)\| \cdot \|T(x(t)) - x(t)\| = 0, \end{aligned}$$

hence $\lim_{t \rightarrow +\infty} \|T(x(t)) - x(t)\|$ exists and is a real number.

- (a) Firstly, let us assume that $\int_0^{+\infty} \lambda(t)(1 - \lambda(t))dt = +\infty$. This immediately implies by (9) that $\lim_{t \rightarrow +\infty} (T(x(t)) - x(t)) = 0$, thus (ii) holds. Taking into account that λ is bounded, from (7) and (ii) we deduce (iii). For the last property of the theorem we need to verify the second assumption of the Opial Lemma. Let $\bar{x} \in \mathcal{H}$ be a weak sequential cluster point of x , that is, there exists a sequence $t_n \rightarrow +\infty$ (as $n \rightarrow \infty$) such that $(x(t_n))_{n \in \mathbb{N}}$ converges weakly to \bar{x} . Applying the demiclosedness principle [28, Corollary 4.18] and (ii) we obtain $\bar{x} \in \text{Fix } T$ and the conclusion follows.
- (b) We suppose now that $\inf_{t \geq 0} \lambda(t) > 0$. From the first relation of (7) and (i) we easily deduce that $Tx - x \in L^2([0, \infty), \mathcal{H})$, hence the function $t \mapsto \frac{1}{2}\|T(x(t)) - x(t)\|^2$ belongs to $L^1([0, \infty))$. Since $\frac{d}{dt} \left(\frac{1}{2}\|T(x(t)) - x(t)\|^2 \right) \leq 0$ for almost all $t \geq 0$, we

obtain by applying Lemma 3 that $\lim_{t \rightarrow \infty} \|T(x(t)) - x(t)\|^2 = 0$, thus (ii) holds. The rest of the proof can be done in the lines of case (a) considered above. \square

Remark 6 We refer the reader to [60] for a recent contribution concerning convergence rates of the trajectories generated by (7) in case the operator T satisfies some regularity assumptions, like boundedly linear/Hölder regularity.

Remark 7 The explicit discretization of (7) with respect to the time variable t , with step size $h_n > 0$, yields for an initial point x_0 the following iterative scheme:

$$x_{n+1} = x_n + h_n \lambda_n (T x_n - x_n) \quad \forall n \geq 0.$$

By taking $h_n = 1$ this becomes

$$x_{n+1} = x_n + \lambda_n (T x_n - x_n) \quad \forall n \geq 0, \tag{10}$$

which is the classical Krasnosel’skiĭ–Mann algorithm for finding the set of fixed points of the nonexpansive operator T (see [28, Theorem 5.14]). Let us mention that the convergence of (10) is guaranteed under the condition $\sum_{n \in \mathbb{N}} \lambda_n (1 - \lambda_n) = +\infty$. Notice that in case $\lambda_n = 1$ for all $n \in \mathbb{N}$ and for an initial point x_0 different from 0, the convergence of (10) can fail, as it happens for instance for the operator $T = -\text{Id}$. In contrast to this, as pointed out in Theorem 5, the dynamical system (7) has a strong global solution and the convergence of the trajectory is guaranteed also in case $\lambda(t) = 1$ for all $t \geq 0$.

Remark 8 The conclusions of Theorem 5 remain valid for the class of averaged operators. Let $\alpha \in (0, 1)$ be fixed. We say that $R : \mathcal{H} \rightarrow \mathcal{H}$ is α -averaged if there exists a nonexpansive operator $T : \mathcal{H} \rightarrow \mathcal{H}$ such that $R = (1 - \alpha)\text{Id} + \alpha T$. In this case we consider $\lambda : [0, +\infty) \rightarrow [0, 1/\alpha]$ and the condition $\int_0^{+\infty} \lambda(t)(1 - \lambda(t))dt = +\infty$ is replaced by $\int_0^{+\infty} \lambda(t)(1 - \alpha\lambda(t))dt = +\infty$.

Remark 9 Relying on time rescaling arguments and using the link with a dynamical system governed by a cocoercive operator, it has been shown in [39, Section 4] that the conclusion of the theorem remains valid under the condition $\int_0^{+\infty} \lambda(t)dt = +\infty$, which is weaker than the ones considered above. Also the specific upper bound for λ is not needed anymore. Our option to present the results in this manner is motivated by the link with the discrete case and the fact that for the convergence rates in Theorem 10 the upper bound will be used again.

In the following we investigate the convergence rate of the trajectories of the dynamical system (7). This will be done in terms of the fixed point residual function $t \mapsto \|T(x(t)) - x(t)\|$ and of $t \mapsto \|\dot{x}(t)\|$. Notice that convergence rates for the discrete iteratively generated algorithm (10) have been investigated in [59, 62, 67].

Theorem 10 *Let $T : \mathcal{H} \rightarrow \mathcal{H}$ be a nonexpansive mapping such that $\text{Fix } T \neq \emptyset$, $\lambda : [0, +\infty) \rightarrow (0, 1)$ a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that*

$$0 < \inf_{t \geq 0} \lambda(t) \leq \sup_{t \geq 0} \lambda(t) < 1.$$

Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (7). Then for all $t \geq 0$ we have

$$t \|\dot{x}(t)\|^2 \leq t \|T(x(t)) - x(t)\|^2 \leq \frac{2}{\underline{\lambda}} \int_{t/2}^t \lambda(s)(1 - \lambda(s)) \|T(x(s)) - x(s)\|^2 ds,$$

where $\underline{\tau} = \inf_{t \geq 0} \lambda(t)(1 - \lambda(t)) > 0$ and $\lim_{t \rightarrow +\infty} \int_{t/2}^t \lambda(s)(1 - \lambda(s)) \|T(x(s)) - x(s)\|^2 ds = 0$.

Proof Define the function $f : [0, +\infty) \rightarrow [0, +\infty)$,

$$f(t) = \int_0^t \lambda(s)(1 - \lambda(s)) \|T(x(s)) - x(s)\|^2 ds.$$

According to (9) we have that $\lim_{t \rightarrow +\infty} f(t) \in \mathbb{R}$.

Since $t \mapsto \frac{1}{2} \|T(x(t)) - x(t)\|^2$ is decreasing (see the proof of Theorem 5), we have for all $t \geq 0$:

$$\begin{aligned} \|T(x(t)) - x(t)\|^2 \int_{t/2}^t \lambda(s)(1 - \lambda(s)) ds &\leq \int_{t/2}^t \lambda(s)(1 - \lambda(s)) \|T(x(s)) - x(s)\|^2 ds \\ &= f(t) - f(t/2). \end{aligned}$$

Taking into account the definition of $\underline{\tau}$, we easily derive

$$\frac{\underline{\tau}}{2} t \|T(x(t)) - x(t)\|^2 \leq \int_{t/2}^t \lambda(s)(1 - \lambda(s)) \|T(x(s)) - x(s)\|^2 ds,$$

and the conclusion follows by using again (7). □

In the following we investigate a continuous version of the forward-backward algorithm. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $\beta > 0$ and $B : \mathcal{H} \rightarrow \mathcal{H}$ be β -cocoercive. Consider the dynamical system

$$\begin{cases} \dot{x}(t) = \lambda(t) \left[J_{\gamma A} \left(x(t) - \gamma B(x(t)) \right) - x(t) \right] \\ x(0) = x_0. \end{cases} \tag{11}$$

It is immediate that (11) can be written in the form (7) for $T = J_{\gamma A} \circ (\text{Id} - \gamma B)$. According to [28, Corollary 23.8 and Remark 4.24(iii)], $J_{\gamma A}$ is $1/2$ -cocoercive. Moreover, by [28, Proposition 4.33], $\text{Id} - \gamma B$ is $\gamma/(2\beta)$ -averaged. Combining this with [74, Theorem 3(b)], we derive that R is $1/\delta$ -averaged, where $\delta := \frac{4\beta - \gamma}{2\beta}$. In the next theorem, the statements (i)-(iv) follow from Remark 8 by noticing that $\text{Fix } T = \text{zer}(A + B)$, see [28, Proposition 25.1(iv)]. For (v) and (vi) we refer to [39, Theorem 12].

Theorem 11 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator, $\beta > 0$ and $B : \mathcal{H} \rightarrow \mathcal{H}$ be β -cocoercive such that $\text{zer}(A + B) \neq \emptyset$. Let $\gamma \in (0, 2\beta)$ and set $\delta = \frac{4\beta - \gamma}{2\beta}$. Let $\lambda : [0, +\infty) \rightarrow [0, \delta]$ be a Lebesgue measurable function and $x_0 \in \mathcal{H}$. Suppose that either*

$$\int_0^{+\infty} \lambda(t)(\delta - \lambda(t)) dt = +\infty \text{ or } \inf_{t \geq 0} \lambda(t) > 0.$$

Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (11). Then the following statements are true:

- (i) *the trajectory x is bounded and $\int_0^{+\infty} \|\dot{x}(t)\|^2 dt < +\infty$;*
- (ii) $\lim_{t \rightarrow +\infty} \left[J_{\gamma A} \left(x(t) - \gamma B(x(t)) \right) - x(t) \right] = 0$;
- (iii) $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$;
- (iv) *$x(t)$ converges weakly to a point in $\text{zer}(A + B)$, as $t \rightarrow +\infty$.*

Suppose that $\inf_{t \geq 0} \lambda(t) > 0$. Then the following hold:

- (v) if $y \in \text{zer}(A + B)$, then $\lim_{t \rightarrow +\infty} B(x(t)) = By$ and B is constant on $\text{zer}(A + B)$;
- (vi) if A or B is uniformly monotone, then $x(t)$ converges strongly to the unique point in $\text{zer}(A + B)$, as $t \rightarrow +\infty$.

Remark 12 (i) The explicit discretization of (11) with respect to the time variable t , with step size $h_n > 0$ and initial point x_0 , yields the following iterative scheme:

$$\frac{x_{n+1} - x_n}{h_n} = \lambda_n \left[J_{\gamma A} \left(x_n - \gamma Bx_n \right) - x_n \right] \quad \forall n \geq 0.$$

For $h_n = 1$ this becomes

$$x_{n+1} = x_n + \lambda_n \left[J_{\gamma A} \left(x_n - \gamma Bx_n \right) - x_n \right] \quad \forall n \geq 0, \tag{12}$$

which is the classical forward-backward algorithm for finding the set of zeros of $A+B$ (see [28, Theorem 25.8]). Let us mention that the convergence of (12) is guaranteed under the condition $\sum_{n \in \mathbb{N}} \lambda_n (\delta - \lambda_n) = +\infty$.

- (ii) Let us notice that in case $A + B$ is strongly monotone, then rate of the convergence in (vi) is exponential, see [40, Theorem 1].
- (iii) Notice that strong convergence of the trajectory can be induced by Tikhonov regularization. We refer the reader to [50] where the dynamics

$$\begin{cases} \dot{x}(t) = \lambda(t) \left[J_{\gamma A} \left(x(t) - \gamma B(x(t)) + \epsilon(t)x(t) \right) - x(t) \right] \\ x(0) = x_0 \end{cases} \tag{13}$$

has been investigated under appropriate conditions imposed on the Tikhonov regularization in order to generate strong convergence of the trajectory towards the minimal norm solution. Notice that (13) already appears in [32, Section 5] in case $\lambda(t) = 1$, A is the normal cone to a nonempty, closed convex set $C \subseteq \mathcal{H}$ (that is, $J_{\gamma A}$ is the projector operator onto C) and B is the gradient of a convex and smooth function.

Remark 13 Let us mention that in case $A = \partial f$, where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function defined on a real Hilbert space \mathcal{H} , and for $\lambda(t) = 1$ for all $t \geq 0$, the dynamical system (11) has been studied in [1]. Notice that the weak convergence is obtained in [1, Theorem 5.2] for a constant step-size $\gamma \in (0, 4\beta)$.

Remark 14 In Theorem 11 above related to the dynamical system (11), the cocoercivity of the operator B plays an important role. Now let us stress the fact that in applications arising from game theory for example, the operator $(x, y) \rightarrow (\nabla_x \phi(x, y), -\nabla_y \phi(x, y))$ is not cocoercive (even in the bilinear case when $\phi(x, y) = \langle x, Ky \rangle$ with K a linear and continuous operator). Under further conditions imposed on $\phi : \mathcal{H} \times \mathcal{G} \rightarrow \mathbb{R}$ (\mathcal{G} is another real Hilbert space), like convexity in x , concavity in y and some smoothness conditions, this operator is monotone and Lipschitz. Hence the desire to consider dynamical systems and numerical schemes also in this case is well motivated.

With respect to the case when B is monotone and Lipschitz, let us mention two types of dynamics considered in the literature. First, a dynamical system of forward-backward-forward-type

$$\begin{cases} y(t) = J_{\gamma A}(x(t) - \gamma B(x(t))) \\ \dot{x}(t) + x(t) = y(t) + \lambda [B(x(t)) - B(y(t))] \\ x(0) = x_0, \end{cases} \tag{14}$$

introduced and investigated in [26]. The time discretization of this dynamics leads to the well-known Tseng algorithm [88]

$$\begin{cases} y_n = J_{\gamma A}(x_n - \gamma B(x_n)) \\ x_{n+1} = y_n + \lambda [B(x_n) - B(y_n)], \end{cases} \tag{15}$$

which plays an important role also in primal-dual type algorithms for highly structured monotone inclusion/optimization problems, see for example [57]. Let us also mention that the dynamics (14) and several discretizations have been considered in the context of pseudo-monotone variational inequalities, see [49].

Second, the dynamical system

$$\begin{cases} \dot{x}(t) + x(t) = J_{\gamma A}(x(t) - y(t)) - \dot{y}(t) \\ y(t) = \gamma B(x(t)) \\ x(0) = x_0, \end{cases} \tag{16}$$

has been investigated in [61]. By using the notation $z(t) = x(t) + y(t)$, it can be shown that the dynamics (16) is equivalent to

$$\dot{z}(t) + z(t) = \left(\frac{\text{Id} + R_{\gamma A} R_{\gamma B}}{2} \right) z(t), \tag{17}$$

where $R_{\gamma A} = 2J_{\gamma A} - \text{Id}$ is the reflected resolvent of A . Notice that this is a continuous version of the Douglas-Rachford algorithm, see [28]. Now, (17) can be written in the language of (7), since the reflected resolvent is a nonexpansive mapping. Further, let us mention that time discretization of (16) leads to the numerical scheme

$$x_{n+1} = J_{\gamma A}(x_n - \gamma B(x_n)) - \gamma (B(x_n) - B(x_{n-1})), \tag{18}$$

which compared to Tseng’s algorithm (15) has the advantage that in each iteration it requires only one forward evaluation of the single valued operator B (see also [69] for another numerical scheme with this property). Finally, we mention also [77] for continuous in time dynamics and numerical schemes for monotone inclusion problems involving monotone and Lipschitz operators.

The results presented in this section can be specialized in the context of the convex optimization problem:

$$\min_{x \in \mathcal{H}} \{f(x) + g(x)\} \tag{19}$$

where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function, $\beta > 0$ and $g : \mathcal{H} \rightarrow \mathbb{R}$ is a convex and (Fréchet) differentiable function with $1/\beta$ -Lipschitz continuous gradient (which implies that ∇g is β -cocoercive, due to the celebrated Baillon-Haddad Theorem, see [28, Corollary 8.16]). Notice that

$$\operatorname{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\} = \operatorname{zer}(\partial f + \nabla g)$$

and the system (11) becomes

$$\begin{cases} \dot{x}(t) = \lambda(t) \left[\operatorname{prox}_{\gamma f} \left(x(t) - \gamma \nabla g(x(t)) \right) - x(t) \right] \\ x(0) = x_0. \end{cases} \tag{20}$$

Let us underline the remarkable result that in case $\lambda(t) = 1$, the trajectory of (20) converges weakly to a solution of (19) with no further restriction on the positive step size γ . This has been proved in [1, Theorem 5.2] by using the energy functional

$$E(t, x^*) = \frac{1}{2\gamma} \|x(t) - x^*\|^2 + g(x(t)) - g(x^*) - \langle \nabla g(x^*), x(t) - x^* \rangle, \tag{21}$$

where x^* is a solution of (19).

For the prox-gradient scheme (see ISTA in [29]) it is known that we have the rate $O\left(\frac{1}{n}\right)$ for the functions values along the generated iterates. The following theorem can be seen as the continuous counterpart of this result.

Theorem 15 *In the above setting, we consider $\lambda(t) = 1$ and $\frac{\gamma}{\beta}(3 + \frac{\gamma}{\beta}) \leq 1$. For $x_0 \in \mathcal{H}$, let $x(\cdot) : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (20). Then for all $T > 0$ the following holds*

$$0 \leq (f + g)(\dot{x}(T) + x(T)) - (f + g)(x^*) + \frac{1}{2\gamma} \|\dot{x}(T)\|^2 \leq \frac{1}{2\gamma T} \|x_0 - x^*\|^2 \tag{22}$$

where x^* is a solution of (19).

Proof From the characterization of the proximal operator we have

$$-\frac{1}{\gamma} \dot{x}(t) - \nabla g(x(t)) \in \partial f(\dot{x}(t) + x(t)), \tag{23}$$

The convexity of f yields

$$f(x^*) \geq f(\dot{x}(t) + x(t)) + \left\langle -\frac{1}{\gamma} \dot{x}(t) - \nabla g(x(t)), x^* - \dot{x}(t) - x(t) \right\rangle. \tag{24}$$

Further, by using the convexity of g and the Descent Lemma (see [72, Lemma 1.2.3]) we derive

$$\begin{aligned} g(x^*) &\geq g(x(t)) + \langle \nabla g(x(t)), x^* - x(t) \rangle \\ &\geq g(\dot{x}(t) + x(t)) - \langle \nabla g(x(t)), \dot{x}(t) \rangle - \frac{1}{2\beta} \|\dot{x}(t)\|^2 + \langle \nabla g(x(t)), x^* - x(t) \rangle \\ &= g(\dot{x}(t) + x(t)) + \langle \nabla g(x(t)), x^* - \dot{x}(t) - x(t) \rangle - \frac{1}{2\beta} \|\dot{x}(t)\|^2 \end{aligned} \tag{25}$$

Combining (24) and (25) we obtain

$$\begin{aligned} (f + g)(x^*) &\geq (f + g)(\dot{x}(t) + x(t)) - \frac{1}{\gamma} \langle \dot{x}(t), x^* - \dot{x}(t) - x(t) \rangle - \frac{1}{2\beta} \|\dot{x}(t)\|^2 \\ &= (f + g)(\dot{x}(t) + x(t)) + \frac{2-\gamma}{2\gamma} \|\dot{x}(t)\|^2 + \frac{1}{2\gamma} \frac{d}{dt} (\|x(t) - x^*\|^2) \\ &\geq (f + g)(\dot{x}(t) + x(t)) + \frac{1}{2\gamma} \|\dot{x}(t)\|^2 + \frac{1}{2\gamma} \frac{d}{dt} (\|x(t) - x^*\|^2). \end{aligned} \tag{26}$$

Let us fix an arbitrary $T > 0$. By integration, we derive from (26)

$$\begin{aligned} &\frac{1}{2\gamma} \|x(T) - x^*\|^2 + \int_0^T \left[(f + g)(\dot{x}(t) + x(t)) - (f + g)(x^*) + \frac{1}{2\gamma} \|\dot{x}(t)\|^2 \right] dt \\ &\leq \frac{1}{2\gamma} \|x_0 - x^*\|^2. \end{aligned} \tag{27}$$

We notice that due to (20), we have $\dot{x}(t) = M(x(t))$ where $M : \mathcal{H} \rightarrow \mathcal{H}$, $M = \text{prox}_{\gamma f} \circ (\text{Id} - \gamma \nabla g) - \text{Id}$ is a $(2 + \frac{\gamma}{\beta})$ -Lipschitz continuous operator. Hence, \dot{x} is locally absolutely continuous and (see Remark 1(b)) \ddot{x} exists and for almost every $t \geq 0$ one has

$$\|\ddot{x}(t)\| \leq \left(2 + \frac{\gamma}{\beta}\right) \|\dot{x}(t)\|. \tag{28}$$

Due to the continuity properties on $[0, T]$ of the trajectory, one has

$$\begin{aligned} \dot{x} + x &\in L^2([0, T]; \mathcal{H}), \\ -\frac{1}{\gamma}\dot{x} - \nabla g(x) &\in L^2([0, T]; \mathcal{H}) \\ \ddot{x} + \dot{x} &\in L^2([0, T]; \mathcal{H}). \end{aligned}$$

From (23) and [52, Lemme 4, p. 73] (see also [16, Lemma 3.2]) we obtain that the function $t \mapsto f(\dot{x}(t) + x(t))$ is absolutely continuous and

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

Moreover, it holds

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

Summing up the last two equalities we derive

$$\begin{aligned} \frac{d}{dt} (f + g)(\dot{x}(t) + x(t)) &= \left\langle -\frac{1}{\gamma}\dot{x}(t) - \nabla g(x(t)) + \nabla g(\dot{x}(t) + x(t)), \ddot{x}(t) + \dot{x}(t) \right\rangle \\ &= -\frac{1}{2\gamma} \frac{d}{dt} (\|\dot{x}(t)\|^2) - \frac{1}{\gamma} \|\dot{x}(t)\|^2 \\ &\quad + \langle \nabla g(\dot{x}(t) + x(t)) - \nabla g(x(t)), \ddot{x}(t) + \dot{x}(t) \rangle \\ &\leq -\frac{1}{2\gamma} \frac{d}{dt} (\|\dot{x}(t)\|^2) - \frac{1}{\gamma} \|\dot{x}(t)\|^2 \\ &\quad + \frac{1}{\beta} \|\dot{x}(t)\| \cdot \|\ddot{x}(t) + \dot{x}(t)\| \tag{29} \\ &\leq -\frac{1}{2\gamma} \frac{d}{dt} (\|\dot{x}(t)\|^2) - \frac{1}{\gamma} \|\dot{x}(t)\|^2 + \frac{1}{\beta} (3 + \frac{\gamma}{\beta}) \|\dot{x}(t)\|^2 \tag{30} \\ &= -\frac{1}{2\gamma} \frac{d}{dt} (\|\dot{x}(t)\|^2) - \left[\frac{1}{\gamma} - \frac{1}{\beta} (3 + \frac{\gamma}{\beta}) \right] \|\dot{x}(t)\|^2, \end{aligned}$$

where in (29) we used the Lipschitz continuity of ∇g and in (30) the inequality (28). Altogether, we conclude that for almost every $t \geq 0$ we have

$$\frac{d}{dt} \left[(f + g)(\dot{x}(t) + x(t)) + \frac{1}{2\gamma} \|\dot{x}(t)\|^2 \right] + \left[\frac{1}{\gamma} - \frac{1}{\beta} (3 + \frac{\gamma}{\beta}) \right] \|\dot{x}(t)\|^2 \leq 0. \tag{31}$$

Due to the condition imposed on the step size, we get that

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

hence the function $t \mapsto (f + g)(\dot{x}(t) + x(t)) - (f + g)(x^*) + \frac{1}{2\gamma} \|\dot{x}(t)\|^2$ is nonincreasing. This together with (27) yields the inequality

$$\begin{aligned} & \frac{1}{2\gamma} \|x(T) - x^*\|^2 + T \left[(f + g)(\dot{x}(T) + x(T)) - (f + g)(x^*) + \frac{1}{2\gamma} \|\dot{x}(T)\|^2 \right] \\ & \leq \frac{1}{2\gamma} \|x_0 - x^*\|^2 \end{aligned}$$

and the proof is complete. □

Remark 16 Let us mention now a recent contribution related to structured optimization problems. For \mathcal{H} and \mathcal{G} real Hilbert spaces, we consider the convex minimization problem

$$\min_{x \in \mathcal{H}} f(x) + h(x) + g(Ax), \tag{32}$$

where $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : \mathcal{G} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, convex and lower semicontinuous functions, $h : \mathcal{H} \rightarrow \mathbb{R}$ is a convex and Fréchet differentiable function with Lipschitz continuous gradient and $A : \mathcal{H} \rightarrow \mathcal{G}$ is a continuous linear operator.

Due to the presence of the composition with a linear operator, dual variables appear both when writing the Fenchel dual problem and also in primal-dual numerical schemes existing in the literature for solving highly structured optimization problems. The following dynamical system introduced in [47] can be seen as a continuous version of the mentioned numerical schemes for solving (32):

$$\begin{cases} \dot{x}(t) + x(t) \in (\partial f + cA^*A + M_1(t))^{-1} (M_1(t)x(t) + cA^*z(t) - A^*y(t) - \nabla h(x(t))) \\ \dot{z}(t) + z(t) \in (\partial g + cI + M_2(t))^{-1} (M_2(t)z(t) + cA(\gamma\dot{x}(t) + x(t)) + y(t)) \\ \dot{y}(t) = cA(x(t) + \dot{x}(t)) - c(z(t) + \dot{z}(t)) \\ x(0) = x^0 \in \mathcal{H}, z(0) = z^0 \in \mathcal{G}, y(0) = y^0 \in \mathcal{G}, \end{cases} \tag{33}$$

where $c > 0$, $\gamma \in [0, 1]$, $A^* : \mathcal{G} \rightarrow \mathcal{H}$ is the adjoint operator and $M_1 : [0, +\infty) \rightarrow S_+(\mathcal{H})$ and $M_2 : [0, +\infty) \rightarrow S_+(\mathcal{G})$ (which means that for all $t \geq 0$, $M_1(t) : \mathcal{H} \rightarrow \mathcal{H}$ is linear, continuous, self-adjoint and positive semidefinite).

Let us stress the fact that the analysis of the dynamical system (33) is involved. It needs more technical results in order to show that the system is well-posed. Moreover, the asymptotic analysis requires deep machinery in order to derive an appropriate energy functional. We invite the reader to consult [47] for all these details, just mentioning that under appropriate hypotheses, $(x(t), z(t), y(t))$ converges weakly to a saddle point of the Lagrangian l , defined as $l : \mathcal{H} \times \mathcal{G} \times \mathcal{G} \rightarrow \overline{\mathbb{R}}, l(x, z, y) = f(x) + h(x) + g(z) + \langle y, Ax - z \rangle$.

There are not too many works in the literature devoted to the solving of optimization problems involving compositions with linear operators by means of continuous in time dynamics. Let us also mention a recent contribution of Attouch [6], where a different dynamical system is proposed in [6, Section 2.2] for solving block-structured optimization problems with linear constraint. We mention also [31], where a dynamical system attached to a more involved optimization problem is investigated and which is a continuous counterpart of the proximal alternating minimization algorithm AMA (see also the work of Tseng on AMA [87]).

The dynamics (33) provides through explicit time discretization a numerical algorithm which is a combination of the linearized proximal method of multipliers and the proximal ADMM algorithm.

Indeed, by writing the inclusion in an equivalent form and using the explicit discretization of with respect to the time variable t and constant step $h_k \equiv 1$ yields the iterative scheme (see [47, Remark 1])

$$\begin{cases} x_{n+1} \in \operatorname{argmin}_{x \in \mathcal{H}} \left(f(x) + \langle x - x_n, \nabla h(x_n) \rangle + \frac{c}{2} \|Ax - z_n + \frac{y_n}{c}\|^2 + \frac{1}{2} \|x - x_n\|_{M_1^n}^2 \right) \\ z_{n+1} \in \operatorname{argmin}_{z \in \mathcal{G}} \left(g(z) + \frac{c}{2} \|A(\gamma x_{n+1} + (1 - \gamma)x_n) - z + \frac{y_n}{c}\|^2 + \frac{1}{2} \|z - z_n\|_{M_2^n}^2 \right) \\ y_{n+1} = y_n + c(Ax_{n+1} - z_{n+1}) \end{cases} \quad (34)$$

where $(M_1^n)_{n \geq 0}$ and $(M_2^n)_{n \geq 0}$ are two operator sequences in $S_+(\mathcal{H})$ and $S_+(\mathcal{G})$, respectively.

The algorithm (34) is a combination of the linearized proximal method of multipliers and the proximal ADMM algorithm.

Indeed, in the case when $\gamma = 1$, (34) becomes the proximal ADMM algorithm with variable metrics from [27] (see, also, [42]). If, in addition, $h = 0$ and the operator sequences $(M_1^k)_{k \geq 0}$ and $(M_2^k)_{k \geq 0}$ are constant, then (34) becomes the proximal ADMM algorithm investigated in [81, Section 3.2] (see, also, [63]). It is known that the proximal ADMM algorithm can be seen as a generalization of the full splitting primal-dual algorithms of Chambolle-Pock (see [54]) and Condat-Vu (see [58, 89] and [43, 44]).

On the other hand, in the case when $\gamma = 0$, (34) becomes an extension of the linearized proximal method of multipliers of Chen-Teboulle (see [55], [81, Algorithm 1]).

A particular choice for the linear maps M_1 and M_2 transforms (33) into a dynamical system of primal-dual type formulated in the spirit of the full splitting paradigm. For every $t \in [0, +\infty)$, define

$$M_1(t) = \frac{1}{\tau(t)}I - cA^*A \text{ and } M_2(t) = 0,$$

where $\tau(t) > 0$ is such that $c\tau(t)\|A\|^2 \leq 1$.

In this particular setting, the dynamical system (33) can be equivalently written as (see [47, Remark 2])

$$\begin{cases} \dot{x}(t) + x(t) = \operatorname{prox}_{\tau(t)f} \left((I - c\tau(t)A^*A)x(t) + c\tau(t)A^*z(t) - \tau(t)A^*y(t) - \tau(t)\nabla h(x(t)) \right) \\ \dot{y}(t) + y(t) + c(\gamma - 1)A\dot{x}(t) = \operatorname{prox}_{cg^*} \left(cA(\gamma\dot{x}(t) + x(t)) + y(t) \right) \\ \dot{y}(t) = cA(x(t) + \dot{x}(t)) - c(z(t) + \dot{z}(t)) \\ x(0) = x^0 \in \mathcal{H}, z(0) = z^0 \in \mathcal{G}, y(0) = y^0 \in \mathcal{G}, \end{cases} \quad (35)$$

where $g^* : \mathcal{G} \rightarrow \mathbb{R} \cup \{+\infty\}$ is the Fenchel conjugate of the proper, convex and lower semi-continuous function g . Let us also mention that when $h = 0$ and $\gamma = 1$ the discretization of the dynamical system (35) leads to the primal-dual algorithm proposed by Chambolle and Pock in [54].

3 Second Order Dynamical Systems

As in the previous section, we start with a general dynamical system which will be considered then in several special instances, including the link to optimization problems as well:

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)B(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \quad (36)$$

where $u_0, v_0 \in \mathcal{H}$, $B : \mathcal{H} \rightarrow \mathcal{H}$ is a β -cocoercive operator, $\lambda : [0, +\infty) \rightarrow [0, +\infty)$ is a relaxation function in time and $\gamma : [0, +\infty) \rightarrow [0, +\infty)$ is a continuous damping parameter.

The existence and uniqueness of the trajectory follows (under appropriate conditions imposed on the parameters) from the Cauchy-Lipschitz-Picard Theorem by writing (36) as a first order dynamical system in a product space (see [38, Theorem 4]).

In order to prove the convergence of the trajectories of (36), we make the following assumptions on the relaxation function λ and the damping parameter γ , respectively:

(A1) $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ are locally absolutely continuous and there exists $\theta > 0$ such that for almost every $t \in [0, +\infty)$ we have

$$\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t) \text{ and } \frac{\gamma^2(t)}{\lambda(t)} \geq \frac{1 + \theta}{\beta}. \quad (37)$$

Due to Definition 1 and Remark 1(a), $\dot{\lambda}(t), \dot{\gamma}(t)$ exists for almost every $t \geq 0$ and $\dot{\lambda}, \dot{\gamma}$ are Lebesgue integrable on each interval $[0, b]$ for $0 < b < +\infty$. This combined with $\dot{\gamma}(t) \leq 0 \leq \dot{\lambda}(t)$ and the fact that λ, γ take only positive values yield the existence of a positive lower bound $\underline{\lambda}$ for λ and of a positive upper bound $\overline{\gamma}$ for γ . Furthermore, the second assumption in (37) provides also a positive upper bound $\overline{\lambda}$ for λ and a positive lower bound $\underline{\gamma}$ for γ .

We would also like to point out that under the conditions considered in (A1) the global version of the Cauchy-Lipschitz-Picard Theorem allows us to conclude that, for $u_0, v_0 \in \mathcal{H}$, there exists a unique trajectory $x : [0, +\infty) \rightarrow \mathcal{H}$ which is a C^2 -function and which satisfies the first relation in (36) for every $t \in [0, +\infty)$. The considerations we make in the following take into account this fact.

Let us also mention that in case $\gamma(t) = \gamma$ and $\lambda(t) = \lambda$ for every $t \in [0, +\infty)$, where $\gamma, \lambda > 0$, the assumption (A1) becomes $\gamma^2\beta > \lambda$, a condition which has been used in [20] in relation with the study of the asymptotical convergence of the dynamical system (36).

We state now the convergence result.

Theorem 17 *Let $B : \mathcal{H} \rightarrow \mathcal{H}$ be a β -cocoercive operator for $\beta > 0$ such that $\text{zer } B := \{u \in \mathcal{H} : Bu = 0\} \neq \emptyset$, $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A1) and $u_0, v_0 \in \mathcal{H}$. Let $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (36). Then the following statements are true:*

- (i) *the trajectory x is bounded and $\dot{x}, \ddot{x}, Bx \in L^2([0, +\infty); \mathcal{H})$;*
- (ii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} B(x(t)) = 0$;*
- (iii) *$x(t)$ converges weakly to an element in $\text{zer } B$ as $t \rightarrow +\infty$.*

Proof (i) Take an arbitrary $x^* \in \text{zer } B$ and consider for every $t \in [0, +\infty)$ the function $h(t) = \frac{1}{2}\|x(t) - x^*\|^2$. We have $\dot{h}(t) = \langle x(t) - x^*, \dot{x}(t) \rangle$ and $\ddot{h}(t) = \|\dot{x}(t)\|^2 +$

$\langle x(t) - x^*, \dot{x}(t) \rangle$ for every $t \in [0, +\infty)$. Taking into account (36), we get for every $t \in [0, +\infty)$

$$\ddot{h}(t) + \gamma(t)\dot{h}(t) + \lambda(t) \langle x(t) - x^*, B(x(t)) \rangle = \|\dot{x}(t)\|^2. \tag{38}$$

The cocoercivity of B and the fact that $Bx^* = 0$ yields for every $t \in [0, +\infty)$

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

Taking again into account (36) one obtains for every $t \in [0, +\infty)$

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

or, equivalently,

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

Combining this inequality with

$$\frac{\gamma(t)}{\lambda(t)} \frac{d}{dt}(\|\dot{x}(t)\|^2) = \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) - \frac{\dot{\gamma}(t)\lambda(t) - \gamma(t)\dot{\lambda}(t)}{\lambda^2(t)} \|\dot{x}(t)\|^2 \tag{39}$$

and

$$\gamma(t)\dot{h}(t) = \frac{d}{dt}(\gamma h)(t) - \dot{\gamma}(t)h(t) \geq \frac{d}{dt}(\gamma h)(t), \tag{40}$$

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

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

Now, assumption (A1) delivers for almost every $t \in [0, +\infty)$ the inequality

$$\ddot{h}(t) + \frac{d}{dt}(\gamma h)(t) + \beta \frac{d}{dt} \left(\frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \right) + \theta \|\dot{x}(t)\|^2 + \beta\bar{\lambda}^{-1} \|\ddot{x}(t)\|^2 \leq 0. \tag{41}$$

This implies that the function $t \mapsto \dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2$, which is locally absolutely continuous, is monotonically decreasing. Hence there exists a real number M such that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 \leq M, \tag{42}$$

which yields that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \underline{\gamma}h(t) \leq M.$$

By multiplying this inequality with $\exp(\underline{\gamma}t)$ and then integrating from 0 to T , where $T > 0$, one easily obtains

$$h(T) \leq h(0) \exp(-\underline{\gamma}T) + \frac{M}{\underline{\gamma}}(1 - \exp(-\underline{\gamma}T)),$$

thus

$$h \text{ is bounded} \tag{43}$$

and, consequently,

$$\text{the trajectory } x \text{ is bounded.} \quad (44)$$

On the other hand, from (42), it follows that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \beta \underline{\gamma} \bar{\lambda}^{-1} \|\dot{x}(t)\|^2 \leq M,$$

hence

$$\langle x(t) - x^*, \dot{x}(t) \rangle + \beta \underline{\gamma} \bar{\lambda}^{-1} \|\dot{x}(t)\|^2 \leq M.$$

This inequality in combination with (44) yields

$$\dot{x} \text{ is bounded,} \quad (45)$$

which further implies that

$$\dot{h} \text{ is bounded.} \quad (46)$$

Integrating the inequality (41) we obtain that there exists a real number $N \in \mathbb{R}$ such that for every $t \in [0, +\infty)$

$$\dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2 + \theta \int_0^t \|\dot{x}(s)\|^2 ds + \beta \bar{\lambda}^{-1} \int_0^t \|\ddot{x}(s)\|^2 ds \leq N.$$

From here, via (46), we conclude that $\dot{x}(\cdot), \ddot{x}(\cdot) \in L^2([0, +\infty); \mathcal{H})$. Finally, from (36) and (A1) we deduce $Bx \in L^2([0, +\infty); \mathcal{H})$ and the proof of (i) is complete.

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

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

and Lemma 3 together with (i) lead to $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$.

Further, by taking into consideration Remark 1(b), for every $t \in [0, +\infty)$ we have

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

By using again Lemma 3 and (i) we get $\lim_{t \rightarrow +\infty} B(x(t)) = 0$, while the fact that $\lim_{t \rightarrow +\infty} \ddot{x}(t) = 0$ follows from (36) and (A2).

(iii) We are going to prove that both assumptions in Opial Lemma are fulfilled. The first one concerns the existence of $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$. As seen in the proof of part (i), the function $t \mapsto \dot{h}(t) + \gamma(t)h(t) + \beta \frac{\gamma(t)}{\lambda(t)} \|\dot{x}(t)\|^2$ is monotonically decreasing, thus from (i), (ii) and (A1) we deduce that $\lim_{t \rightarrow +\infty} \gamma(t)h(t)$ exists and it is a real number. By taking also into account that $\exists \lim_{t \rightarrow +\infty} \gamma(t) \in (0, \infty)$, we obtain the existence of $\lim_{t \rightarrow +\infty} \|x(t) - x^*\|$.

We come now to the second assumption of the Opial Lemma. Let \bar{x} be a weak sequential cluster point of x , that is, there exists a sequence $t_n \rightarrow +\infty$ (as $n \rightarrow +\infty$) such that $(x(t_n))_{n \in \mathbb{N}}$ converges weakly to \bar{x} . Since B is a maximally monotone operator (see for instance [28, Example 20.28]), its graph is sequentially closed with respect to the weak-

strong topology of the product space $\mathcal{H} \times \mathcal{H}$. By using also that $\lim_{n \rightarrow +\infty} B(x(t_n)) = 0$, we conclude that $B\bar{x} = 0$, hence $\bar{x} \in \text{zer } B$ and the proof is complete. \square

A standard choice of a cocoercive operator defined on a real Hilbert spaces is $B = \text{Id} - T$, where $T : \mathcal{H} \rightarrow \mathcal{H}$ is a nonexpansive operator. As it easily follows from the nonexpansiveness of T , B is in this case $1/2$ -cocoercive. For this particular operator B the dynamical system (36) becomes

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t)(x(t) - T(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{47}$$

Theorem 17 gives rise to a corresponding result where the trajectory converges weakly to a fixed point of T , where in (A1) we use the condition $\frac{\gamma^2(t)}{\lambda(t)} \geq 2(1 + \theta)$.

Remark 18 (i) The analysis can be further extended to the situation when T in (47) is an α -averaged operator (with $0 < \alpha < 1$). In this case we use in (A1) the condition $\frac{\gamma^2(t)}{\lambda(t)} \geq 2\alpha(1 + \theta)$.

(ii) In the particular case when $\gamma(t) = \gamma > 0$ for all $t \geq 0$ and $\lambda(t) = 1$ for all $t \in [0, +\infty)$ the dynamical system (47) has been studied in [7, Theorem 3.2] under the condition $\gamma^2 > 2$, which is covered by the theory presented above. We would also like to notice that in [3] an anisotropic damping has been considered in the context of approaching the minimization of a smooth convex function via second order dynamical systems.

Let us turn now our attention to monotone inclusions and consider the dynamics:

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t) \left[x(t) - J_{\eta A} \left(x(t) - \eta B(x(t)) \right) \right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{48}$$

Further, in (A1) we use $\frac{\gamma^2(t)}{\lambda(t)} \geq \frac{2(1+\theta)}{\delta}$, where $\delta := \frac{4\beta-\eta}{2\beta}$.

The statements (i)-(iii) in the next result follow from Remark 18(ii) by taking into account that under the hypotheses, $T = J_{\eta A} \circ (\text{Id} - \eta B)$ is $1/\delta$ -averaged, see also the comments before Theorem 11. For (iv) and (v) we invite the reader to consult [38, Theorem 12]. Moreover, under appropriate hypotheses, the rate of convergence in (v) is exponential, see [40, Theorem 7].

Theorem 19 *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $B : \mathcal{H} \rightarrow \mathcal{H}$ be a β -cocoercive operator for $\beta > 0$ such that $\text{zer}(A + B) \neq \emptyset$. Let $\eta \in (0, 2\beta)$ and set $\delta := \frac{4\beta-\eta}{2\beta}$. Let $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A1) as mentioned right after (48), $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (48). Then the following statements are true:*

- (i) *the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - J_{\eta A} \circ (\text{Id} - \eta B))x \in L^2([0, +\infty); \mathcal{H})$;*
- (ii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - J_{\eta A} \circ (\text{Id} - \eta B))(x(t)) = 0$;*
- (iii) *$x(t)$ converges weakly to a point in $\text{zer}(A + B)$ as $t \rightarrow +\infty$;*
- (iv) *if $x^* \in \text{zer}(A + B)$, then $B(x(\cdot)) - Bx^* \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t \rightarrow +\infty} B(x(t)) = Bx^*$ and B is constant on $\text{zer}(A + B)$;*
- (v) *if A or B is uniformly monotone, then $x(t)$ converges strongly to the unique point in $\text{zer}(A + B)$ as $t \rightarrow +\infty$.*

Remark 20 For second order dynamical systems with vanishing damping we refer the reader to [22] where the following differential equation is investigated

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{49}$$

$A_{\lambda(t)} : \mathcal{H} \rightarrow \mathcal{H}$ being the Yosida regularization of A , defined by $A_{\lambda(t)} = \frac{1}{\lambda(t)}(\text{Id} - J_{\lambda(t)A})$. Convergence of the trajectory towards a zero of the maximally monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is reported in [22] together with convergence rates concerning the velocity and the acceleration. Let us also mention the recent contribution [19] where a Newton-like correction term is used:

$$\begin{cases} \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \frac{d}{dt}(A_{\lambda(t)}(x(t))) + A_{\lambda(t)}(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{50}$$

with $\beta > 0$. In case of convex potentials, the correction term becomes the Hessian driven damping which is responsible for attenuation of the oscillations of the trajectories and for better convergence rates for the gradients along the orbits, see also [12] and [82]. For the algorithmic consequences of (49) and (50) we invite the reader to consult [22] and [18].

In the remaining of this section we turn our attention to optimization problems of the form

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

see (19) and the setting described there.

In the following statement we approach the minimizers of $f + g$ via the second order dynamical system

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \lambda(t) \left[x(t) - \text{prox}_{\eta f} \left(x(t) - \eta \nabla g(x(t)) \right) \right] = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{51}$$

Corollary 21 *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ by a proper, convex and lower semicontinuous function and $g : \mathcal{H} \rightarrow \mathbb{R}$ be a convex and (Fréchet) differentiable function with $\frac{1}{\beta}$ -Lipschitz continuous gradient for $\beta > 0$ such that $\text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\} \neq \emptyset$. Let $\eta \in (0, 2\beta)$ and set $\delta := \frac{4\beta - \eta}{2\beta}$. Let $\lambda, \gamma : [0, +\infty) \rightarrow (0, +\infty)$ be functions fulfilling (A1) as mentioned right after (48), $u_0, v_0 \in \mathcal{H}$ and $x : [0, +\infty) \rightarrow \mathcal{H}$ be the unique strong global solution of (51). Then the following statements are true:*

- (i) *the trajectory x is bounded and $\dot{x}, \ddot{x}, (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))x \in L^2([0, +\infty); \mathcal{H})$;*
- (ii) *$\lim_{t \rightarrow +\infty} \dot{x}(t) = \lim_{t \rightarrow +\infty} \ddot{x}(t) = \lim_{t \rightarrow +\infty} (\text{Id} - \text{prox}_{\eta f} \circ (\text{Id} - \eta \nabla g))(x(t)) = 0$;*
- (iii) *$x(t)$ converges weakly to a minimizer of $f + g$ as $t \rightarrow +\infty$;*
- (iv) *if x^* is a minimizer of $f + g$, then $\nabla g(x(\cdot)) - \nabla g(x^*) \in L^2([0, +\infty); \mathcal{H})$, $\lim_{t \rightarrow +\infty} \nabla g(x(t)) = \nabla g(x^*)$ and ∇g is constant on $\text{argmin}_{x \in \mathcal{H}} \{f(x) + g(x)\}$;*
- (v) *if f or g is uniformly convex, then $x(t)$ converges strongly to the unique minimizer of $f + g$ as $t \rightarrow +\infty$.*

Remark 22 Consider again the setting in Remark 18(ii), namely, when $\gamma(t) = \gamma > 0$ for every $t \geq 0$ and $\lambda(t) = 1$ for every $t \in [0, +\infty)$. Furthermore, for C a nonempty, convex, closed subset of \mathcal{H} , let $f = \delta_C$ be the indicator function of C , which is defined as being

equal to 0 for $x \in C$ and to $+\infty$, else. The dynamical system (20) attached in this setting to the minimization of g over C becomes

$$\begin{cases} \ddot{x}(t) + \gamma \dot{x}(t) + x(t) - \text{proj}_C(x(t) - \eta \nabla g(x(t))) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{52}$$

The asymptotic convergence of the trajectories of (52) has been studied in [7, Theorem 3.1] under the conditions $\gamma^2 > 2$ and $0 < \eta \leq 2\beta$. In this case the corresponding assumption (A1) trivially holds by choosing θ such that $0 < \theta \leq (\gamma^2 - 2)/2 \leq (\delta\gamma^2 - 2)/2$. Thus, in order to verify the corresponding (A1) in case $\lambda(t) = 1$ for every $t \in [0, +\infty)$ one needs to equivalently assume that $\gamma^2 > 2/\delta$. Since $\delta \geq 1$, this provides a slight improvement over [7, Theorem 3.1] in what concerns the choice of γ . We refer the reader also to [5] for an analysis of the convergence rates of trajectories of the dynamical system (52) when g is endowed with supplementary properties.

Remark 23 By using the energy functional (21), it is possible to go beyond the condition $0 < \eta \leq 2\beta$ considered in Corollary 21. We refer the reader to [38, Corollary 16] for more details. Moreover, in [38, Remark 17] it has been pointed out that in case $\gamma(t) = \gamma > 0$ for every $t \geq 0$ and $\lambda(t) = 1$ for every $t \in [0, +\infty)$, the only condition imposed on the parameters is $\gamma^2 > \frac{\eta}{\beta} + 1$. In other words, this allows in this particular setting a more relaxed choice for the parameters, beyond the standard assumptions $0 < \eta \leq 2\beta$ and $\gamma^2 > 2$ considered in [7].

Remark 24 The explicit discretization of (51) with respect to the time variable t , with step size $h_n > 0$, relaxation variable $\lambda_n > 0$, damping variable $\gamma_n > 0$ and initial points $x_0 := u_0$ and $x_1 := v_0$ yields the following iterative scheme

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

For $h_n = 1$ this becomes

$$x_{n+1} = \left(1 - \frac{\lambda_n}{1 + \gamma_n} \right) x_n + \frac{\lambda_n}{1 + \gamma_n} \text{prox}_{\eta f}(x_n - \eta \nabla g(x_n)) + \frac{\lambda_n}{1 + \gamma_n} (x_n - x_{n-1}) \quad \forall n \geq 1, \tag{53}$$

which is a relaxed forward-backward algorithm for minimizing $f + g$ with inertial effects. In order to investigate the convergence properties of the above iterative scheme, it is natural to assume that $(\gamma_n)_{n \geq 1}$ is nonincreasing, $(\lambda_n)_{n \geq 1}$ is nondecreasing, and there exists a lower bound for $(\gamma_n^2/\lambda_n)_{n \geq 1}$. The control of the inertial term by means of the variable parameters λ_n and γ_n could increase the speed of convergence of the algorithm (53). Making use of the the sequence $(\lambda_n)_{n \geq 1}$ in (53), one obtains a relaxed forward-backward scheme, where the relaxation is usually considered in the literature in order to achieve more freedom in the choice of the parameters involved in the numerical scheme. We refer the reader to [11] where this fact has been underlined in the context of monotone inclusion problems.

Remark 25 In case f is vanishing and $\eta\lambda(t) = 1$, the dynamical system (51) becomes

$$\begin{cases} \ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla g(x(t)) = 0 \\ x(0) = u_0, \dot{x}(0) = v_0. \end{cases} \tag{54}$$

There is an intensive research in the literature mainly devoted to the vanishing damping of the form $\gamma(t) = \frac{\alpha}{t}$, with $\alpha > 0$. This is due to the fact that in case $\alpha > 3$ this provides

a fast convergence result in terms of the objective function of order $o\left(\frac{1}{t^2}\right)$. The starting point of the investigations was the work of Su, Boyd and Candès [86], followed by relevant contributions of Attouch and his co-workers (see for example [13, 14] and the references therein). This is related to the accelerated gradient method in the sense of Nesterov

$$\begin{cases} y_n &= x_n + \frac{n-1}{n+\alpha-1}(x_n - x_{n-1}) \\ x_{n+1} &= y_n - \gamma \nabla g(y_n) \end{cases}$$

where inertial terms by means of iterates accelerate the convergence behaviour of the numerical scheme. Let us mention also the recent contributions [12, 82] where inertial terms induced by the gradient of g are also considered, which are coming from the discretization of Hessian driven damping terms, see also [23] and [48].

4 First Order Dynamics for Nonconvex Optimization Problems

In this section we approach the solving of the optimization problem

$$\inf_{x \in \mathbb{R}^n} [f(x) + g(x)], \tag{55}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, convex, lower semicontinuous function and $g : \mathbb{R}^n \rightarrow \mathbb{R}$ 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

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\eta f} \left(x(t) - \eta \nabla g(x(t)) \right) \\ x(0) = x_0, \end{cases} \tag{56}$$

where $\eta > 0$ and $x_0 \in \mathbb{R}^n$ is chosen arbitrary (we consider the euclidean space \mathbb{R}^n).

Lemma 26 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality*

$$\eta\beta(3 + \eta\beta) < 1. \tag{57}$$

For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (56). Then the following statements hold:

- (a) $\dot{x} \in L^2([0, +\infty); \mathbb{R}^n)$ and $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$;
- (b) $\exists \lim_{t \rightarrow +\infty} (f + g)(\dot{x}(t) + x(t)) \in \mathbb{R}$.

Proof Let us start by noticing that \dot{x} is locally absolutely continuous, hence \ddot{x} exists and for almost every $t \in [0, +\infty)$ one has

$$\|\ddot{x}(t)\| \leq (2 + \eta\beta)\|\dot{x}(t)\|. \tag{58}$$

We fix an arbitrary $T > 0$. Due to the continuity properties of the trajectory on $[0, T]$, (28) 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 of the proximal point operator we have

$$-\frac{1}{\eta}\dot{x}(t) - \nabla g(x(t)) \in \partial f(\dot{x}(t) + x(t)) \forall t \in [0, +\infty). \tag{59}$$

Like in the proof of Theorem 15 we obtain

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

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

$$\frac{d}{dt} \left[(f + g)(\dot{x}(t) + x(t)) + \frac{1}{2\eta} \|\dot{x}(t)\|^2 \right] + \left[\frac{1}{\eta} - \beta(3 + \eta\beta) \right] \|\dot{x}(t)\|^2 \leq 0. \tag{60}$$

By integration we get

$$\begin{aligned} (f + g)(\dot{x}(T) + x(T)) + \frac{1}{2\eta} \|\dot{x}(T)\|^2 + \left[\frac{1}{\eta} - \beta(3 + \eta\beta) \right] \int_0^T \|\dot{x}(t)\|^2 dt \leq \\ (f + g)(\dot{x}(0) + x_0) + \frac{1}{2\eta} \|\dot{x}(0)\|^2. \end{aligned} \tag{61}$$

By using (57) and the fact that $f + g$ is bounded from below and by taking into account that $T > 0$ has been arbitrarily chosen, we obtain

$$\dot{x} \in L^2([0, +\infty); \mathbb{R}^n). \tag{62}$$

Due to (58), this further implies

$$\ddot{x} \in L^2([0, +\infty); \mathbb{R}^n). \tag{63}$$

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

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

By applying Lemma 3, it follows that $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$ and the proof of (a) is complete. From (60), (57) and by using that $T > 0$ has been arbitrarily chosen, we get

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

for almost every $t \in [0, +\infty)$. From Lemma 2 it follows that

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

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

For the following generalized subdifferential notions and their basic properties we refer to [70, 80]. Let $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. If $x \in \text{dom } h$, we consider the *Fréchet (viscosity) subdifferential* of h at x as the set

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

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

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

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

Notice that in case h is convex, these subdifferential notions coincide with the *convex subdifferential*, thus $\hat{\partial}h(x) = \partial_L h(x) = \{v \in \mathbb{R}^n : h(y) \geq h(x) + \langle v, y - x \rangle \ \forall y \in \mathbb{R}^n\}$ 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 $v_k \in \partial_L h(x_k)$ for all $k \in \mathbb{N}$, $(x_k, v_k) \rightarrow (x, v)$ and $h(x_k) \rightarrow h(x)$ as $k \rightarrow +\infty$, then $v \in \partial_L h(x)$.

The Fermat rule reads in this nonsmooth setting as follows: if $x \in \mathbb{R}^n$ is a local minimizer of h , then $0 \in \partial_L h(x)$. We denote by

$$\text{crit}(h) = \{x \in \mathbb{R}^n : 0 \in \partial_L h(x)\}$$

the set of (limiting)-critical points of h .

We define the limit set of the trajectory x as

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

Lemma 27 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality (57). For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (56). Then*

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

Proof Let $\bar{x} \in \omega(x)$ and $t_k \rightarrow +\infty$ be such that $x(t_k) \rightarrow \bar{x}$ as $k \rightarrow +\infty$. From the characterization of the prox operator we have

$$\begin{aligned} -\frac{1}{\eta} \dot{x}(t_k) - \nabla g(x(t_k)) + \nabla g(\dot{x}(t_k) + x(t_k)) &\in \partial f(\dot{x}(t_k) + x(t_k)) + \nabla g(\dot{x}(t_k) + x(t_k)) \\ &= \partial_L(f + g)(\dot{x}(t_k) + x(t_k)) \ \forall k \in \mathbb{N}. \end{aligned} \tag{64}$$

Lemma 26(a) and the Lipschitz continuity of ∇g ensure that

$$-\frac{1}{\eta} \dot{x}(t_k) - \nabla g(x(t_k)) + \nabla g(\dot{x}(t_k) + x(t_k)) \rightarrow 0 \text{ as } k \rightarrow +\infty \tag{65}$$

and

$$\dot{x}(t_k) + x(t_k) \rightarrow \bar{x} \text{ as } k \rightarrow +\infty. \tag{66}$$

We claim that

$$\lim_{k \rightarrow +\infty} (f + g)(\dot{x}(t_k) + x(t_k)) = (f + g)(\bar{x}). \tag{67}$$

Due to the lower semicontinuity of f it holds

$$\liminf_{k \rightarrow +\infty} f(\dot{x}(t_k) + x(t_k)) \geq f(\bar{x}). \tag{68}$$

Further, since

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

we have the inequality

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

Taking the limit as $k \rightarrow +\infty$ we derive by using again Lemma 26(a) that

$$\limsup_{k \rightarrow +\infty} f(\dot{x}(t_k) + x(t_k)) \leq f(\bar{x}),$$

which combined with (68) implies

$$\lim_{k \rightarrow +\infty} f(\dot{x}(t_k) + x(t_k)) = f(\bar{x}).$$

By using (66) and the continuity of g we conclude that (67) is true.

Altogether, from (64), (65), (66) and (67) and the closedness criteria of the limiting subdifferential we obtain $0 \in \partial_L(f + g)(\bar{x})$ and the proof is complete. \square

Lemma 28 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality (57). For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (56) and consider the function*

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

Then the following statements are true:

(H₁) for almost every $t \in [0, +\infty)$ it holds

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

and

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

(H₂) for every $t \in [0, +\infty)$ it holds

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

and

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

(H₃) for $\bar{x} \in \omega(x)$ and $t_k \rightarrow +\infty$ such that $x(t_k) \rightarrow \bar{x}$ as $k \rightarrow +\infty$, we have $H(\dot{x}(t_k) + x(t_k), x(t_k)) \rightarrow H(\bar{x}, \bar{x})$ as $k \rightarrow +\infty$.

Proof (H₁) follows from Lemma 26. The first statement in (H₂) follows from the characterization of the proximal operator and the relation

$$\partial_L H(u, v) = (\partial_L(f + g)(u) + \eta^{-1}(u - v)) \times \{\eta^{-1}(v - u)\} \forall (u, v) \in \mathbb{R}^n \times \mathbb{R}^n, \quad (69)$$

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 27. \square

Lemma 29 *Assume that the hypotheses of Lemma 28 hold. Suppose that x is bounded. Then the following statements are true:*

- (a) $\omega(\dot{x} + x, x) \subseteq \text{crit}(H) = \{(u, u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in \text{crit}(f + g)\}$;
- (b) $\lim_{t \rightarrow +\infty} \text{dist} \left((\dot{x}(t) + x(t), x(t)), \omega(\dot{x} + x, x) \right) = 0$;
- (c) $\omega(\dot{x} + x, x)$ is nonempty, compact and connected;
- (d) H is finite and constant on $\omega(\dot{x} + x, x)$.

Proof (a), (b) and (d) are direct consequences Lemma 26, Lemma 27 and Lemma 28.

Finally, (c) is a classical result from [64]. We also refer the reader to the proof of Theorem 4.1 in [4], where it is shown that the properties of $\omega(x)$ of being nonempty, compact and connected are generic for bounded trajectories fulfilling $\lim_{t \rightarrow +\infty} \dot{x}(t) = 0$. \square

Remark 30 Suppose that $\eta > 0$ fulfills the inequality (57) and $f + g$ is coercive, that is

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

For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (56). 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 (61) it follows

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

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

The class of functions satisfying the *Kurdyka-Łojasiewicz property* plays a crucial role in the asymptotic analysis of the dynamical system (56). For $\eta \in (0, +\infty]$, we denote by Θ_η the class of concave and continuous functions $\varphi : [0, \eta) \rightarrow [0, +\infty)$ 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 [9, 36]) we use also the *distance function* to a set, defined for $A \subseteq \mathbb{R}^n$ as $\text{dist}(x, A) = \inf_{y \in A} \|x - y\|$ for all $x \in \mathbb{R}^n$.

Definition 3 (*Kurdyka-Łojasiewicz property*) Let $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. We say that h satisfies the *Kurdyka-Łojasiewicz (KL) property* at $\bar{x} \in \text{dom } \partial_L h = \{x \in \mathbb{R}^n : \partial_L h(x) \neq \emptyset\}$, if there exist $\eta \in (0, +\infty]$, a neighborhood U of \bar{x} and a function $\varphi \in \Theta_\eta$ such that for all x in the intersection

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

the following inequality holds

$$\varphi'(h(x) - h(\bar{x})) \text{dist}(0, \partial_L h(x)) \geq 1.$$

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

The origins of this notion go back to the pioneering work of Łojasiewicz [68], where it is proved that for a real-analytic function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ and a critical point $\bar{x} \in \mathbb{R}^n$ (that is $\nabla h(\bar{x}) = 0$), there exists $\theta \in [1/2, 1)$ such that the function $|h - h(\bar{x})|^\theta \|\nabla h\|^{-1}$ is bounded around \bar{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 [66] extended this property to differentiable functions definable in o-minimal structures. Further extensions to the non-smooth setting can be found in [9, 33–35].

One of the remarkable properties of the KL functions is their ubiquity in applications (see [36]). 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 [8–10, 33–36] and the references therein for more on KL functions and illustrating examples.

We come now to the main result of this section.

Theorem 31 *Suppose that $f + g$ is bounded from below and $\eta > 0$ fulfills the inequality (57). For $x_0 \in \mathbb{R}^n$, let $x \in C^1([0, +\infty), \mathbb{R}^n)$ be the unique global solution of (56) and consider the function*

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

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 $\bar{x} \in \text{crit}(f + g)$ such that $\lim_{t \rightarrow +\infty} x(t) = \bar{x}$.*

Proof According to Lemma 29, we can choose an element $\bar{x} \in \text{crit}(f + g)$ such that $(\bar{x}, \bar{x}) \in \omega(\dot{x} + x, x)$. According to Lemma 28, it follows that

$$\lim_{t \rightarrow +\infty} H(\dot{x}(t) + x(t), x(t)) = H(\bar{x}, \bar{x}).$$

We treat the following two cases separately.

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

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

Since from Lemma 28(H1) we have

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

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

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

Thus $H(\dot{x}(t) + x(t), x(t)) = H(\bar{x}, \bar{x})$ for every $t \geq \bar{t}$. This yields by Lemma 28(H1) that $\dot{x}(t) = 0$ for almost every $t \in [\bar{t}, +\infty)$, hence x is constant on $[\bar{t}, +\infty)$ and the conclusion follows.

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

In virtue of Lemma 29(c) and (d) and since H is a KL function, there exist positive numbers ϵ and η and a concave function $\varphi \in \Theta_\eta$ such that for all

$$(x, y) \in \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^n : \text{dist}((u, v), \Omega) < \epsilon\} \cap \{(u, v) \in \mathbb{R}^n \times \mathbb{R}^n : H(\bar{x}, \bar{x}) < H(u, v) < H(\bar{x}, \bar{x}) + \eta\} \tag{70}$$

one has

$$\varphi'(H(x, y) - H(\bar{x}, \bar{x})) \text{dist}((0, 0), \partial_L H(x, y)) \geq 1. \tag{71}$$

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

$x(t), x(t)), \Omega) < \epsilon$ for all $t \geq t_2$. Hence for all $t \geq T := \max\{t_1, t_2\}$, $(\dot{x}(t) + x(t), x(t))$ belongs to the intersection in (70). Thus, according to (71), for every $t \geq T$ we have

$$\varphi' \left(H(\dot{x}(t) + x(t), x(t)) - H(\bar{x}, \bar{x}) \right) \text{dist} \left((0, 0), \partial_L H(\dot{x}(t) + x(t), x(t)) \right) \geq 1. \tag{72}$$

By applying Lemma 28(H2) we obtain for almost every $t \in [T, +\infty)$

$$(\beta + \eta^{-1}) \|\dot{x}(t)\| \varphi' \left(H(\dot{x}(t) + x(t), x(t)) - H(\bar{x}, \bar{x}) \right) \geq 1. \tag{73}$$

From here, by using Lemma 28(H1) and that $\varphi' > 0$ and

$$\begin{aligned} & \frac{d}{dt} \varphi \left(H(\dot{x}(t) + x(t), x(t)) - H(\bar{x}, \bar{x}) \right) = \\ & \varphi' \left(H(\dot{x}(t) + x(t), x(t)) - H(\bar{x}, \bar{x}) \right) \frac{d}{dt} H(\dot{x}(t) + x(t), x(t)), \end{aligned}$$

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

$$\frac{d}{dt} \varphi \left(H(\dot{x}(t) + x(t), x(t)) - H(\bar{x}, \bar{x}) \right) \leq - \left(\beta + \eta^{-1} \right)^{-1} \left[\frac{1}{\eta} - (3 + \eta\beta)\beta \right] \|\dot{x}(t)\|. \tag{74}$$

Since φ is bounded from below, by taking into account (57), it follows $\dot{x} \in L^1([0, +\infty); \mathbb{R}^n)$. From here we obtain that $\lim_{t \rightarrow +\infty} x(t)$ exists and this closes the proof. \square

Since the class of semi-algebraic functions is closed under addition (see for example [36]) and $(u, v) \mapsto c\|u - v\|^2$ is semi-algebraic for $c > 0$, we can state the above result in case $f + g$ is semi-algebraic.

Remark 32 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 (60). Similar regularizations of the objective function of (55) have been considered also in [45, 73], in the context of the investigation of non-relaxed forward-backward methods involving inertial and memory effects in the nonconvex setting.

Remark 33 (i) We invite the reader to consult [41, Section 3.2] for results concerning convergence rates of the trajectory, which are expressed by means of the Łojasiewicz exponent.

(ii) The optimization problem (55) can be approached in this non-convex setting also by a second order dynamical system of implicit-type

$$\begin{cases} \ddot{x}(t) + \gamma \dot{x}(t) + x(t) = \text{prox}_{\eta f} \left(x(t) - \eta \nabla g(x(t)) \right) \\ x(0) = u_0, \dot{x}(0) = v_0, \end{cases} \tag{75}$$

where $u_0, v_0 \in \mathbb{R}^n$ and $\gamma, \eta \in (0, +\infty)$. A similar analysis can be carried out for this dynamics too, see [46].

(iii) Finally, let us mention a recent contribution [51] related to a dynamical system proposed and investigated in connection with block-structured optimization problems

$$\inf_{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m} [f(x) + g(y) + H(x, y)], \tag{76}$$

where f, g are proper, lower-semicontinuous and H satisfies certain smoothness conditions. The dynamics considered in [51] can be seen as a continuous counter-part of the PALM method introduced in [36].

Acknowledgments The author is thankful to the handling editor and two reviewers for their comments which improved the presentation of the manuscript.

Funding Information Open access funding provided by University of Vienna.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Abbas, B., Attouch, H.: 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. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. *J. Optim. Theory Appl.* **161**(2), 331–360 (2014)
3. Alvarez, F.: On the minimizing property of a second order dissipative system in Hilbert spaces. *SIAM J. Control. Optim.* **38**(4), 1102–1119 (2000)
4. Alvarez, F., Attouch, H., Bolte, J., Redont, P.: A second-order gradient-like dissipative dynamical system with Hessian driven damping. Application to optimization and mechanics. *J. Math. Pures Appl.* (9) **81**(8), 747–779 (2002)
5. Antipin, A.S.: Minimization of convex functions on convex sets by means of differential equations. (Russian) *Differ. Uravnen.* **30**(9), 1475–1486 (1994). translation in *Differential Equations* 30(9), 1365–1375 (1994)
6. Attouch, H.: Fast inertial proximal ADMM algorithms for convex structured optimization with linear constraint, <https://hal.archives-ouvertes.fr/hal-02501604>, hal-02501604 (2020)
7. Attouch, H., Alvarez, F.: The Heavy Ball with Friction Dynamical System for Convex Constrained Minimization Problems. In: *Optimization* (Namur, 1998), Lecture Notes in Economics and Mathematical Systems, vol. 481, pp. 25–35. Springer, Berlin (2000)
8. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Programm.* **116**(1-2) Series B, 5–16 (2009)
9. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality. *Math. Oper. Res.* **35**(2), 438–457 (2010)
10. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. *Math. Programm.* **137**(1-2) Series A, 91–129 (2013)
11. Attouch, H., Cabot, A.: Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions. *Appl. Math. Optim.* **80**(3), 547–598 (2019)
12. Attouch, H., Chbani, Z., Fadili, J., Riahi, H.: First-order optimization algorithms via inertial systems with Hessian driven damping. arXiv:1907.10536v1 (2019)
13. Attouch, H., Chbani, Z., Peypouquet, J., Redont, P.: Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity. *Mathematical Programming* **168**(1-2), Ser B, 123–175 (2018)
14. Attouch, H., Chbani, Z., Riahi, H.: Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control Optim. Calc. Var.*, **25** (2019)
15. Attouch, H., Chbani, Z., Riahi, H.: Fast convex optimization via a third-order in time evolution equation, *Optimization*, online first
16. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of coupled dynamical systems with multiscale aspects. *J. Differ. Equ.* **248**(6), 1315–1344 (2010)
17. Attouch, H., Goudou, X., Redont, P.: 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. *Commun. Contemp. Math.* **2**(1), 1–34 (2000)

18. Attouch, H., László, S.C.: Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators. HAL-02549730 (2020)
19. Attouch, H., László, S.C.: Continuous Newton-like Inertial Dynamics for Monotone Inclusions, <https://hal.archives-ouvertes.fr/hal-02577331> (2020)
20. Attouch, H., Maingé, P.-E.: Asymptotic behavior of second-order dissipative evolution equations combining potential with non-potential effects. ESAIM: Control Optim. Calc. Var. **17**(3), 836–857 (2011)
21. Attouch, H., Marques Alves, M., Svaiter, B.F.: A dynamic approach to a proximal-Newton method for monotone inclusions in Hilbert spaces, with complexity $o(1/n^2)$. J. Convex Anal. **23**(1), 139–180 (2016)
22. Attouch, H., Peypouquet, J.: Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operator. Math. Programm. **174**, 391–432 (2019)
23. Attouch, H., Peypouquet, J., Redont, P.: Fast convex optimization via inertial dynamics with Hessian driven damping. J. Differ. Equ. **261**(10), 5734–5783 (2016)
24. Attouch, H., Svaiter, B.F.: A continuous dynamical Newton-like approach to solving monotone inclusions. SIAM J. Control. Optim. **49**(2), 574–598 (2011)
25. Baillon, J.B., Brézis, H.: Une remarque sur le comportement asymptotique des semigroupes non linéaires. Houst. J. Math. **2**(1), 5–7 (1976)
26. Banert, S., Boţ, R.I.: A forward-backward-forward differential equation and its asymptotic properties. J. Convex Anal. **25**(2), 371–388 (2018)
27. Banert, S., Boţ, R.I., Csetnek, E.R.: Fixing and extending some recent results on the ADMM algorithm, Numerical Algorithms, online first
28. Bauschke, H.H., Combettes, P.L., Analysis, C.onvex.: Monotone Operator Theory in Hilbert Spaces CMS Books in Mathematics. Springer, New York (2011)
29. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
30. Bertsekas, D.P. Nonlinear Programming, 2nd edn. Athena Scientific, Cambridge (1999)
31. Bitterlich, S., Csetnek, E.R., Wanka, G.: A Dynamical Approach to Two-Block Separable Convex Optimization Problems with Linear Constraints, arXiv:2005.09953 (2020)
32. Bolte, J.: Continuous gradient projection method in Hilbert spaces. J. Optim. Theory Appl. **119**(2), 235–259 (2003)
33. Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2006)
34. Bolte, J., Daniilidis, A., Lewis, A., Shota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
35. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. **362**(6), 3319–3363 (2010)
36. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Programm. Ser. A **146**(1–2), 459–494 (2014)
37. Borwein, J.M., Vanderwerff, J.D.: Convex Functions: Constructions, Characterizations and Counterexamples. Cambridge University Press, Cambridge (2010)
38. Boţ, R.I., Csetnek, E.R.: Second order forward-backward dynamical systems for monotone inclusion problems. SIAM J. Control. Optim. **54**(3), 1423–1443 (2016)
39. Boţ, R.I., Csetnek, E.R.: A dynamical system associated with the fixed points set of a nonexpansive operator. J. Dyn. Diff. Equ. **29**(1), 155–168 (2017)
40. Boţ, R.I., Csetnek, E.R.: Convergence rates for dynamical systems associated with monotone inclusion problems. J. Math. Anal. Appl. **457**(2), 1135–1152 (2018)
41. Boţ, R.I., Csetnek, E.R.: A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function. ESAIM Control Optim. Calc. Var. **24**(2), 463–477 (2018)
42. Boţ, R.I., Csetnek, E.R.: ADMM For monotone operators: convergence analysis and rates. Adv. Comput. Math. **45**(1), 327–359 (2019)
43. Boţ, R.I., Csetnek, E.R., Heinrich, A.: A primal-dual splitting algorithm for finding zeros of sums of maximally monotone operators. SIAM J. Optim. **23**(4), 2011–2036 (2013)
44. Boţ, R.I., Csetnek, E.R., Heinrich, A., Hendrich, C.: On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. Math. Program. **150**(2), 251–279 (2015)
45. Boţ, R.I., Csetnek, E.R., László, S.C.: An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions. EURO J. Comput. Optim. **4**, 3–25 (2016)
46. Boţ, R.I., Csetnek, E.R., László, S.C.: Approaching nonsmooth nonconvex minimization through second order proximal-gradient dynamical systems. J. Evol. Equ. **18**(3), 1291–1318 (2018)

47. Boţ, R.I., Csetnek, E.R., László, S.C.: A primal-dual dynamical approach to structured convex minimization problems, arXiv:1905.08290 (2019)
48. Boţ, R.I., Csetnek, E.R., László, S.C.: Tikhonov regularization of a second order dynamical system with Hessian driven damping, *Mathematical Programming*, online first
49. Boţ, R.I., Csetnek, E.R., Vuong, P.T.: A The Forward-Backward-Forward Method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces. *Europ. J. Oper. Res.* **287**, 49–60 (2020)
50. Boţ, R.I., Grad, S.M., Meier, D., Staudigl, M.: Inducing strong convergence of trajectories in dynamical systems associated to monotone inclusions with composite structure, arXiv:1911.04758 (2019)
51. Boţ, R.I., Kanzler, L.: A forward-backward dynamical approach for nonsmooth problems with block structure coupled by a smooth function, arXiv:2001.10051 (2020)
52. Brézis, H.: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, *North-Holland Mathematics Studies No. 5* Notas De Matemática, vol. 50. North-Holland/Elsevier, New York (1973)
53. Bruck, R.E. Jr.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. *J. Funct. Anal.* **18**, 15–26 (1975)
54. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011)
55. Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. *Math. Program.* **64**, 81–101 (1994)
56. Combettes, P.L., Glaudin, L.E.: Quasinonexpansive iterations on the affine hull of orbits: From Mann’s mean value algorithm to inertial methods. *SIAM J. Optim.* **27**, 2356–2380 (2017)
57. Combettes, P.L., Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.* **20**(2), 307–330 (2012)
58. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* **158**(2), 460–479 (2013)
59. Corman, E., Yuan, X.: A Generalized proximal point algorithm and its convergence rate. *SIAM J. Optim.* **24**(4), 1614–1638 (2014)
60. Csetnek, E.R., Eberhard, A., Tam, M.K.: Convergence Rates for Boundedly Regular Systems, arXiv:2004.00818 (2020)
61. Csetnek, E.R., Malitsky, Y., Tam, M.K.: Shadow Douglas-Rachford splitting for monotone inclusions. *Appl. Math. Optim.* **80**(3), 665–678 (2019)
62. Davis, D., Yin, W.: Convergence Rate Analysis of Several Splitting Schemes. In: Glowinski, R., Osher, S.J., Yin, W. (eds.) *Splitting Methods in Communication, Imaging, Science, and Engineering*, pp. 115–163 (2017)
63. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications in system identification and realization. *SIAM J. Matrix Anal. Appl.* **34**, 946–977 (2013)
64. Haraux, A.: *Systèmes Dynamiques Dissipatifs Et Applications* Recherches En Mathématiques Appliquées, vol. 17. Masson, Paris (1991)
65. Haraux, A., Jendoubi, M.: Convergence of solutions of second-order gradient-like systems with analytic nonlinearities. *J. Differ. Equ.* **144**(2), 313–320 (1998)
66. Kurdyka, K.: On gradients of functions definable in o-minimal structures. *Ann. l’inst. Four. (Grenoble)* **48**(3), 769–783 (1998)
67. Liang, J., Fadili, J., Peyré, G.: Convergence rates with inexact nonexpansive operators. *Math. Program.* **159**, 403–434 (2016)
68. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. *Les Équations aux Dérivées Partielles*, Éditions du Centre National de la Recherche Scientifique, Paris, pp. 87–89 (1963)
69. Malitsky, Y., Tam, M.K.: A Forward-Backward splitting method for monotone inclusions without cocoercivity. *SIAM J. Optim.* **30**(2), 1451–1472 (2020)
70. Mordukhovich, B.: *Variational Analysis and Generalized Differentiation I: Basic Theory II: Applications*, Springer, Berlin (2006)
71. Nesterov, Y.: A method for unconstrained convex minimization problem with the rate of convergence $o(1/k^2)$. *Doklady AN SSSR (transl. Sov. Math. Doct.)* **269**, 543–547 (1983)
72. Nesterov, Y.: *Introductory Lectures on Convex Optimization: a Basic Course*. Kluwer Academic Publishers, Dordrecht (2004)
73. Ochs, P., Chen, Y., Brox, T., Pock, T.: Ipiano: Inertial proximal algorithm for non-convex optimization. *SIAM J. Imaging Sci.* **7**(2), 1388–1419 (2014)
74. Ogura, N., Yamada, I.: Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping. *Numer. Funct. Anal. Optim.* **23**(1-2), 113–137 (2002)

75. Peypouquet, J., Sorin, S.: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. *J. Convex Anal.* **17**(3-4), 1113–1163 (2010)
76. Polyak, B.T.: Introduction to Optimization, (Translated from the Russian) Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York (1987)
77. Rieger, J., Tam, M.K.: Backward-Forward-Reflected-Backward Splitting for three operator monotone inclusions. *Applied Mathematics and Computation*, 381 (2020)
78. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. *Pac. J. Math.* **33**(1), 209–216 (1970)
79. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control. Optim.* **14**(5), 877–898 (1976)
80. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis Fundamental Principles of Mathematical Sciences, vol. 317. Springer, Berlin (1998)
81. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex optimization. *SIAM J. Optim.* **24**(1), 269–297 (2014)
82. Shi, B., Du, S.S., Su, W.J., Jordan, M.I.: Acceleration via symplectic discretization of high-resolution differential equations. [arXiv:1902.03694v2](https://arxiv.org/abs/1902.03694v2) (2019)
83. Simon, L.: Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems. *Ann. Math. (2)* **118**, 525–571 (1983)
84. Simons, S.: From Hahn-Banach to Monotonicity. Springer, Berlin (2008)
85. Sontag, E.D.: Mathematical control theory. Deterministic finite-dimensional systems, 2nd Edn, vol. 6. Texts in Applied Mathematics. Springer, New York (1998)
86. Su, W., Boyd, S., Candès, E.J.: A differential equation for modeling Nesterov’s accelerated gradient method: theory and insights. *J. Mach. Learn. Res.* **17**(153), 1–43 (2016)
87. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Control. Optim.* **29**(1), 119–138 (1991)
88. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control. Optim.* **38**, 431–446 (2000)
89. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.* **38**(3), 667–681 (2013)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.