

A forward-backward dynamical approach for nonsmooth problems with block structure coupled by a smooth function

Radu Ioan Bot^{*} Laura Kanzler[†]

November 11, 2020

Abstract

In this paper we aim to minimize the sum of two nonsmooth (possibly also nonconvex) functions in separate variables connected by a smooth coupling function. To tackle this problem we choose a continuous forward-backward approach and introduce a dynamical system which is formulated by means of the partial gradients of the smooth coupling function and the proximal point operator of the two nonsmooth functions. Moreover, we consider variable rates of implicitness of the resulting system. We discuss the existence and uniqueness of a solution and carry out the asymptotic analysis of its convergence behaviour to a critical point of the optimization problem, when a regularization of the objective function fulfills the Kurdyka-Łojasiewicz property. We further provide convergence rates for the solution trajectory in terms of the Łojasiewicz exponent. We conclude this work with numerical simulations which confirm and validate the analytical results.

Key Words. block-coordinate minimization, forward-backward dynamical system, asymptotic analysis, Kurdyka-Łojasiewicz property

AMS subject classification. 34G25, 37N40, 49J52, 90C26, 90C56

1 Introduction

We consider a block-structured optimization problem of the form

$$\min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} \Psi(x,y) := f(x) + g(y) + H(x,y), \quad (1)$$

where the following general assumptions on the functions f , g and H are made:

- Assumption 1.**
- $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ and $g : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ are proper and lower semicontinuous functions, where we denote $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, with $\inf_{\mathbb{R}^n} f > -\infty$ and $\inf_{\mathbb{R}^m} g > -\infty$;
 - $H \in C^1(\mathbb{R}^n \times \mathbb{R}^m)$, ∇H is Lipschitz continuous with constant L , and $\inf_{\mathbb{R}^n \times \mathbb{R}^m} H > -\infty$.

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

[†]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, email: laura.kanzler@univie.ac.at. Research supported by the SFB F65 “Taming Complexity in Partial Differential Systems” and by the Doctoral Program “Dissipation and Dispersion in Nonlinear PDEs”, project number W1245, which are funded by the Austrian Science Fund (FWF).

We approach the solving the optimization problem (1) by associating to it the dynamical system

$$\begin{cases} \dot{x}(t) + x(t) \in \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x(t) - \frac{1}{\gamma_1 L} \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right) \\ \dot{y}(t) + y(t) \in \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y(t) - \frac{1}{\gamma_2 L} \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \right) \\ (x(0), y(0)) = (x_0, y_0), \end{cases} \quad (2)$$

where $\lambda, \mu \in [0, 1]$.

The choice of these parameters have a crucial influence on the structure of the dynamical system (2) since they decide weather it is of explicit or implicit nature. Even more, their values decide how strong the derivatives \dot{x} and \dot{y} are involved on the right hand side of (2). For the extreme values we have on one hand a fully explicit system of ordinary differential inclusions if $\lambda = \mu = 1$, while for $\lambda = \mu = 0$ it is implicit. They have a major impact on the analysis, which will be explained throughout Section3 and Section4, and displayed by numerical simulations in Section5. In the following, we will refer to them as *rates of implicitity*.

Further, the proximal map $\text{prox}_{\alpha h} : \mathbb{R}^d \rightarrow \mathcal{P}(\mathbb{R}^d)$ for a proper and lower semicontinuous function $h : \mathbb{R}^d \rightarrow \bar{\mathbb{R}}$ and $\alpha > 0$ is defined as (see [31, Definition 1.22])

$$\text{prox}_{\alpha h}(u) := \text{argmin}_{v \in \mathbb{R}^d} \left\{ h(v) + \frac{1}{2\alpha} \|v - u\|^2 \right\}.$$

We chose the stepsizes $\frac{1}{\gamma_1 L}$ and $\frac{1}{\gamma_2 L}$ in dependence of the Lipschitz constant of the gradient of H for some suitable $\gamma_1, \gamma_2 > 0$.

In [14], Bolte, Sabach and Teboulle established the PALM (proximal alternating linearized minimization) algorithm for minimizing optimization problems of the form (1). This work served as key motivation to study (2), since an explicit time discretization for the case $\mu = 1$ and $\lambda = 0$ provides a direct correspondence between the ODE system and the PALM algorithm. In this special case the system (2) is of the form

$$\begin{cases} \dot{x}(t) + x(t) \in \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x(t) - \frac{1}{\gamma_1 L} \nabla_x H(x(t), y(t)) \right) \\ \dot{y}(t) + y(t) \in \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y(t) - \frac{1}{\gamma_2 L} \nabla_y H(\dot{x}(t) + x(t), y(t)) \right) \\ (x(0), y(0)) = (x_0, y_0). \end{cases} \quad (3)$$

Explicit time discretization of the first component of (3) has the form

$$x^{k+1} \in \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x^k - \frac{1}{\gamma_1 L} \nabla_x H(x^k, y^k) \right) \quad \forall k \geq 0, \quad (4)$$

where the time stepsize was set equal to one. On the other hand, explicit time discretization of the second component with stepsize one gives

$$y^{k+1} \in \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y^k - \frac{1}{\gamma_2 L} \nabla_y H(x^{k+1}, y^k) \right) \quad \forall k \geq 0. \quad (5)$$

Therefore, we obtain the PALM algorithm for starting point $(x^0, y^0) := (x_0, y_0)$ from the ODE system (3). In [14] it was proved that the limit points of $(x^k, y^k)_{k \geq 0}$ are critical points of the optimization problem (1). In addition, if the objective function satisfies the Kurdyka-Łojasiewicz property and $(x^k, y^k)_{k \geq 0}$ is bounded, then the sequence converges to a critical point of (1).

In case $\mu = \lambda = 1$, explicit time discretization with stepsize one gets

$$x^{k+1} \in \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x^k - \frac{1}{\gamma_1 L} \nabla_x H(x^k, y^k) \right), \quad y^{k+1} \in \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y^k - \frac{1}{\gamma_2 L} \nabla_y H(x^k, y^k) \right) \quad \forall k \geq 0,$$

which is the preconditioned forward-backward algorithm for solving (1). The forward-backward algorithm has been investigated in the fully nonconvex setting in [6, 17].

The approach of convex optimization problems and monotone inclusions from the perspective of dynamical systems has a long tradition, starting with the contributions of Crandall and Pazy [23] as well as Baillon, Brézis [9], [21], and Bruck [22]. In these works dynamical systems formulated as monotone inclusions and governed by subdifferential operators or, more general, maximal monotone operators in Hilbert spaces are investigated. The question of existence and uniqueness of trajectories generated by these dynamical systems is usually investigated in the framework of the Cauchy-Lipschitz Theorem, while the asymptotic convergence behavior of solutions to a minimizer of the convex optimization problem or a zero point of the governing maximal monotone operator builds on Lyapunov analysis.

In the last years, dynamical systems of implicit type approaching monotone inclusions/convex optimization problems, namely, defined by means of resolvent/proximal operators have enjoyed much attention. Here, the pioneering works are by Antipin [4], Bolte [12], and Abbas, Attouch and Svaiter [1, 2, 8]. A general approach for addressing implicit dynamical systems was considered in [15], a dynamical system of forward-backward-forward type was matter of investigation in [10], one of Douglas-Rachford type in [24], while a primal-dual dynamical system approaching structured convex minimization problems was recently introduced in [19].

In what concerns nonconvex optimization problems, we want to mention that the problem of minimizing a general smooth function has been approached from the perspective of first-order and second-order gradient type dynamical systems in Simon [32], Haraux and Jendoubi [27], and Alvarez, Attouch, Bolte and Redont [3]. In addition, proximal-gradient type dynamical systems of first-order and second-order have been investigated in [16] and [18], respectively, in relation to the minimization of the sum of a proper, convex and lower semicontinuous function and a general smooth function.

The outline of this paper is the following: after some preliminaries, we will address in Section 3 the question of existence and uniqueness of strong global solutions to (2), which will be provided for the case $\mu = 1$ and f and g are convex functions and in the framework of the global Picard-Lindelöf Theorem. Section (4) is dedicated to the asymptotic analysis of the trajectories of the general system (2) towards a critical point of the objective function of (1), expressed as a zero of the limiting subdifferential. To this end we will use three main ingredients (see [6, 14] for a similar approach in discrete and [3, 16] in continuous time): (1) we will prove sufficient decrease of a regularized objective function along the trajectories; (2) we will show existence of a subgradient lower bound of the solution trajectories; (3) we will obtain convergence of the solution by taking use of the Kurdyka-Łojasiewicz (KL) property of a regularization of the objective function. Functions satisfying the KL property build a large class of functions and include semi-algebraic functions and functions having analytic features. We close our investigations by establishing convergence rates for the trajectories expressed in terms of the Łojasiewicz exponent of the regularized objective function. We will conclude by confirming and validating the analytical results through numerical simulations.

2 Preliminaries

For $d \geq 1$, we consider on \mathbb{R}^d the Euclidean scalar product and the induced norm. These are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively, independently from the value of d , since confusion is not possible.

Let $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be a given function. Its *effective domain* is defined as $\text{dom } h := \{x \in \mathbb{R}^d : h(x) < +\infty\}$ and we say that h is *proper*, if $\text{dom } h \neq \emptyset$.

Definition 1. • Let $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be proper and lower semicontinuous. The *Fréchet subdifferential*

of h at $x \in \text{dom } h$ is defined as

$$\partial_F h(x) := \left\{ \xi \in \mathbb{R}^d : \liminf_{z \rightarrow x} \frac{h(z) - h(x) - \langle \xi, z - x \rangle}{\|z - x\|} \geq 0 \right\}.$$

For $x \notin \text{dom } h$, we set by convention $\partial_F h(x) := \emptyset$.

- The *limiting (Mordukhovich) subdifferential* of h at $x \in \text{dom } h$ (see [30]) is then defined as the sequential closure of $\partial_F h(x)$ in the following way

$$\partial h(x) := \left\{ \xi \in \mathbb{R}^d; \exists x_k \rightarrow x, h(x_k) \rightarrow h(x) \text{ and } \partial_F h(x_k) \ni \xi_k \rightarrow \xi, \text{ as } k \rightarrow \infty \right\}.$$

For $x \notin \text{dom } h$, we set by convention $\partial h(x) := \emptyset$. We denote by $\text{dom } \partial h := \{x \in \mathbb{R}^d : \partial h(x) \neq \emptyset\}$ the *domain* of the limiting subdifferential ∂h .

Remark 1. i) From the definition it follows that for all $x \in \mathbb{R}^d$ it holds $\partial_F h(x) \subseteq \partial h(x)$. While $\partial h(x)$ is closed, $\partial_F h(x)$ is convex and closed ([31, Theorem 8.6]).

ii) With $\text{crit}(h) := \{x \in \mathbb{R}^d : 0 \in \partial h(x)\}$ we denote the set of (*limiting-*)*critical points* of h . Also in our nonsmooth setting *Fermat's theorem* holds, i.e. if $x \in \mathbb{R}^d$ is a local minimizer of h , then $0 \in \partial h(x)$, i.e. $x \in \text{crit}(h)$.

iii) If h is continuously differentiable at $x \in \mathbb{R}^n$, then we have $\partial h(x) = \{\nabla h(x)\}$.

iv) For the sum of a proper and lower semicontinuous function $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ and a continuously differentiable function $k : \mathbb{R}^d \rightarrow \mathbb{R}$ it holds $\partial(h+k)(x) = \partial h(x) + \nabla k(x)$ for every $x \in \mathbb{R}^d$ ([31]).

Remark 2. i) The proximal operator of a nonconvex function is in general a set-valued map. For a proper and lower semicontinuous function $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ with $\inf h > -\infty$ we have that for every $\alpha > 0$ and every $x \in \mathbb{R}^n$ the set $\text{prox}_{\alpha h}(x)$ is nonempty and compact.

ii) In our considerations below, we will use the following property connecting the *proximal map* with the *limiting subdifferential*. For every proper and lower semicontinuous function $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ and constant $\alpha > 0$ it holds

$$p \in \text{prox}_{\alpha h}(x) \Rightarrow \frac{1}{\alpha}(x - p) \in \partial h(p) \quad \forall x \in \mathbb{R}^d. \quad (6)$$

Since we will have to deal with locally absolutely continuous solution trajectories of (2), we recall the following definition.

Definition 2. A function $x : [0, +\infty) \rightarrow \mathbb{R}^n$ is called *locally absolutely continuous* if $x : [0, T] \rightarrow \mathbb{R}^n$ is absolutely continuous for all $T > 0$, namely, there exists an integrable function $y : [0, T] \rightarrow \mathbb{R}^n$ such that

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

We point out two important properties of absolute continuous trajectories, which will be of much use later on.

Remark 3. i) An absolutely continuous function is almost everywhere (a.e.) differentiable, its derivative coincides a.e. with its distribution derivative, and by the above integration formula it is possible to recover the function from its derivative.

- ii) For an absolutely continuous function $x : [0, T] \rightarrow \mathbb{R}^d$, where $T > 0$, and a Lipschitz continuous function K with Lipschitz constant $\beta \geq 0$, the composition $z := K \circ x$ is also absolutely continuous. Furthermore, z is differentiable a.e. in $[0, T]$ and the inequality $\|\dot{z}(t)\| \leq \beta \|\dot{x}(t)\|$ holds for a.e. $t \in [0, T]$.

Next, we introduce the notion of solution we will mostly deal with.

Definition 3. We say that $z := (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ is a *strong global solution* of (2), if it satisfies the following properties:

- a) the functions x and y are locally absolutely continuous;
b)

$$\dot{x}(t) + x(t) \in \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x(t) - \frac{1}{\gamma_1 L} \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right)$$

and

$$\dot{y}(t) + y(t) \in \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y(t) - \frac{1}{\gamma_2 L} \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \right)$$

for a.e. $t \in [0, +\infty)$;

- c) the functions $t \mapsto f(\dot{x}(t) + x(t))$ and $t \mapsto g(\dot{y}(t) + y(t))$ are locally absolutely continuous;
d) $(x(0), y(0)) = (x_0, y_0)$.

Also the following two results for locally absolutely continuous functions will be crucial for the asymptotic analysis of the trajectories of (2) (see for example [2]).

Lemma 1. *Suppose that $F : [0, +\infty) \rightarrow \mathbb{R}$ is locally absolutely continuous and bounded from below. Furthermore, assume that $\exists G \in L^1([0, +\infty); \mathbb{R})$ such that for a.e. $t \in [0, +\infty)$*

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

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

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

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

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

3 Existence and uniqueness of a trajectory

The aim of this section is to provide a setting in which the existence and uniqueness of a trajectory of the dynamical system (2) is guaranteed. To this end:

- we ask the nonsmooth functions f and g to be in addition *convex*. We recall that the proximal operator of a proper, convex and lower semicontinuous function is single-valued and non-expansive (see, e.g. [11, Proposition 12.28]).

- we consider the following two constellations of the rates of implicity: $\mu = 1$, while $\lambda \in [0, 1]$ remains general, or $\lambda = 1$, while $\mu \in [0, 1]$.

The above setting allows us to rewrite the dynamical system (2) in an explicit form as

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x(t) - \frac{1}{\gamma_1 L} \nabla_x H(x(t), y(t)) \right) \\ \dot{y}(t) + y(t) = \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y(t) - \frac{1}{\gamma_2 L} \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \right) \\ (x(0), y(0)) = (x_0, y_0), \end{cases} \quad (7)$$

for the case $\mu = 1$, and as

$$\begin{cases} \dot{x}(t) + x(t) = \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x(t) - \frac{1}{\gamma_1 L} \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right) \\ \dot{y}(t) + y(t) = \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y(t) - \frac{1}{\gamma_2 L} \nabla_y H(x(t), y(t)) \right) \\ (x(0), y(0)) = (x_0, y_0), \end{cases} \quad (8)$$

for the case $\lambda = 1$.

Theorem 3. *Let f , g and H fulfill Assumption (1) and let further f and g be convex. Then for every starting point $(x_0, y_0) \in \mathbb{R}^n \times \mathbb{R}^m$ the dynamical system (2) for the cases $\mu = 1$ and $\lambda \in [0, 1]$ as well as $\lambda = 1$ and $\mu \in [0, 1]$ has a unique strong global solution $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$, which is in addition continuously differentiable.*

Proof. We will prove the statement only for the case $\mu = 1$ and $\lambda \in [0, 1]$, since for $\lambda = 1$ and $\mu \in [0, 1]$ the proof follows analogously.

To this end we denote $z(t) = (x(t), y(t))$ and notice that (7) can be equivalently written as

$$\begin{cases} \dot{z}(t) = \Gamma(z(t)) \\ z(0) = (x_0, y_0), \end{cases}$$

where $\Gamma : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ is defined as

$$\begin{pmatrix} \Gamma_1(x, y) \\ \Gamma_2(x, y) \end{pmatrix} := \begin{pmatrix} \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x - \frac{1}{\gamma_1 L} \nabla_x H(x, y) \right) - x \\ \text{prox}_{\frac{1}{\gamma_2 L} g} \left(y - \frac{1}{\gamma_2 L} \nabla_y H((1 - \lambda)(\Gamma_1(x, y) + x) + \lambda x, y) \right) - y \end{pmatrix}.$$

We will show that the operator Γ is Lipschitz continuous. Let therefore $(x, y), (\tilde{x}, \tilde{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ be arbitrarily chosen. We denote $(\Gamma_1, \Gamma_2) := \Gamma(x, y)$, $(\tilde{\Gamma}_1, \tilde{\Gamma}_2) := \Gamma(\tilde{x}, \tilde{y})$, and estimate

$$\begin{aligned} \|\Gamma_1 - \tilde{\Gamma}_1\|^2 &= \left\| \text{prox}_{\frac{1}{\gamma_1 L} f} \left(x - \frac{1}{\gamma_1 L} \nabla_x H(x, y) \right) - x - \text{prox}_{\frac{1}{\gamma_1 L} f} \left(\tilde{x} - \frac{1}{\gamma_1 L} \nabla_x H(\tilde{x}, \tilde{y}) \right) + \tilde{x} \right\|^2 \\ &\leq 2 \|x - \tilde{x}\|^2 + 2 \left\| x - \frac{1}{\gamma_1 L} \nabla_x H(x, y) + \frac{1}{\gamma_1 L} \nabla_x H(\tilde{x}, \tilde{y}) - \tilde{x} \right\|^2 \\ &\leq 6 \|x - \tilde{x}\|^2 + \frac{4}{\gamma_1^2 L^2} \|\nabla_x H(x, y) - \nabla_x H(\tilde{x}, \tilde{y})\|^2 \\ &\leq 6 \|x - \tilde{x}\|^2 + \frac{4}{\gamma_1^2} (\|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2) \\ &= \left(6 + \frac{4}{\gamma_1^2} \right) \|x - \tilde{x}\|^2 + \frac{4}{\gamma_1^2} \|y - \tilde{y}\|^2, \end{aligned}$$

where we first used the *non-expansiveness* (i.e. 1-Lipschitz continuity) of the proximal operator, while the last inequality is due to the L -Lipschitz continuity of ∇H . Similarly we obtain

$$\begin{aligned}
& \|\Gamma_2 - \tilde{\Gamma}_2\|^2 \\
& \leq 2\|y - \tilde{y}\|^2 + 2\left\|y - \frac{1}{\gamma_2 L} \nabla_y H((1-\lambda)(\Gamma_1 + x) + \lambda x, y) + \frac{1}{\gamma_2 L} \nabla_y H((1-\lambda)(\tilde{\Gamma}_1 + \tilde{x}) + \lambda \tilde{x}, \tilde{y}) + \tilde{y}\right\|^2 \\
& \leq 6\|y - \tilde{y}\|^2 + \frac{4}{\gamma_2^2 L^2} \left\| \nabla_y H((1-\lambda)(\Gamma_1 + x) + \lambda x, y) - \nabla_y H((1-\lambda)(\tilde{\Gamma}_1 + \tilde{x}) + \lambda \tilde{x}, \tilde{y}) \right\|^2 \\
& \leq 6\|y - \tilde{y}\|^2 + \frac{4}{\gamma_2^2} \left(\|(1-\lambda)[(\Gamma_1 + x) - (\tilde{\Gamma}_1 + \tilde{x})] + \lambda(x - \tilde{x})\|^2 + \|y - \tilde{y}\|^2 \right) \\
& \leq 6\|y - \tilde{y}\|^2 + \frac{4}{\gamma_2^2} \left((1-\lambda)^2 \|\Gamma_1 - \tilde{\Gamma}_1\|^2 + \|x - \tilde{x}\|^2 + \|y - \tilde{y}\|^2 \right) \\
& = \frac{4}{\gamma_2^2} (1-\lambda)^2 \|\Gamma_1 - \tilde{\Gamma}_1\|^2 + \frac{4}{\gamma_2^2} \|x - \tilde{x}\|^2 + \left(6 + \frac{4}{\gamma_2^2}\right) \|y - \tilde{y}\|^2.
\end{aligned}$$

Putting the estimates for $\|\Gamma_1 - \tilde{\Gamma}_1\|$ and $\|\Gamma_2 - \tilde{\Gamma}_2\|$ together we obtain

$$\begin{aligned}
& \|\Gamma(x, y) - \Gamma(\tilde{x}, \tilde{y})\|^2 = \|\Gamma_1 - \tilde{\Gamma}_1\|^2 + \|\Gamma_2 - \tilde{\Gamma}_2\|^2 \\
& \leq \left[6 + \frac{4}{\gamma_1^2} + \frac{4}{\gamma_2^2} + \frac{16(1-\lambda)^2}{\gamma_1^2 \gamma_2^2} + \frac{24(1-\lambda)^2}{\gamma_2^2} \right] \|x - \tilde{x}\|^2 + \left[6 + \frac{4}{\gamma_1^2} + \frac{4}{\gamma_2^2} + \frac{16(1-\lambda)^2}{\gamma_1^2 \gamma_2^2} \right] \|y - \tilde{y}\|^2 \\
& \leq \beta(\lambda, \gamma_1, \gamma_2)^2 \|(x, y) - (\tilde{x}, \tilde{y})\|^2,
\end{aligned}$$

where

$$\beta(\lambda, \gamma_1, \gamma_2) := \left(6 + \frac{4}{\gamma_1^2} + \frac{4 + 24(1-\lambda)^2}{\gamma_2^2} + \frac{16(1-\lambda)^2}{\gamma_1^2 \gamma_2^2} \right)^{\frac{1}{2}}.$$

The Lipschitz continuity of Γ guarantees, according to the global version of the Picard-Lindelöf Theorem (see, e.g. ([33, Theorem 2.2]), the existence of a unique continuously differentiable strong global solution $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ to (7). This means that the statements a), b), and d) in Definition 3 are true.

From here we have that for a.e. $t \geq 0$

$$-\gamma_1 L \dot{x}(t) - \nabla_x H(x(t), (1-\mu)(\dot{y}(t) + y(t)) + \mu y(t)) \in \partial f(\dot{x}(t) + x(t)), \quad (9)$$

$$-\gamma_2 L \dot{y}(t) - \nabla_y H((1-\lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \in \partial g(\dot{y}(t) + y(t)), \quad (10)$$

thus $\dot{x}(t) + x(t) \in \text{dom } f$ and $\dot{y}(t) + y(t) \in \text{dom } g$.

By using again that Γ is $\beta(\lambda, \gamma_1, \gamma_2)$ -Lipschitz continuous, from Remark 3 ii) it follows that the first derivative $\dot{z} = (\dot{x}, \dot{y})$ is locally absolutely continuous, the second derivatives of the partial trajectories exist for a.e. $t \in [0, +\infty)$, and

$$\|(\ddot{x}(t), \ddot{y}(t))\| \leq \beta(\lambda, \gamma_1, \gamma_2) \|(\dot{x}(t), \dot{y}(t))\| \text{ for a.e. } t \in [0, +\infty). \quad (11)$$

Let $T > 0$. Then we have $\dot{x} + x, \ddot{x} + \dot{x} \in L^2([0, T], \mathbb{R}^n)$ and $\dot{y} + y, \ddot{y} + \dot{y} \in L^2([0, T], \mathbb{R}^m)$. Using that H is continuously differentiable, it yields that $-\gamma_1 L \dot{x} - \nabla_x H(x, (1-\mu)(\dot{y} + y) + \mu y) \in L^2([0, T], \mathbb{R}^n)$ and $-\gamma_2 L \dot{y} - \nabla_y H((1-\lambda)(\dot{x} + x) + \lambda x, y) \in L^2([0, T], \mathbb{R}^m)$. Thus, according to [21, Lemma 3.3], the functions $t \mapsto f(\dot{x}(t) + x(t))$ and $t \mapsto g(\dot{y}(t) + y(t))$ are locally absolutely continuous. Thus the statement is proved.

Before closing the proof, we want to mention that in case $\lambda = 1$ and $\mu \in [0, 1]$ the constant which appears in the estimate (11) of the second derivatives of the partial trajectories reads

$$\beta(\mu, \gamma_1, \gamma_2) := \left(6 + \frac{4}{\gamma_2^2} + \frac{4 + 24(1 - \mu)^2}{\gamma_1^2} + \frac{16(1 - \mu)^2}{\gamma_1^2 \gamma_2^2} \right)^{\frac{1}{2}}.$$

□

4 Asymptotic analysis

4.1 A preparatory result

In order to prove convergence of trajectories of (2), we follow the general approach that consists of several steps. In the first step, we will find a *Lyapunov functional* with sufficient decrease of its time derivative, while in the second step we search for *subgradient lower bounds* for the derivative of the trajectory. With standard arguments from [26] one can show that the set of limit points of the trajectory is nonempty, compact and connected. Finally, global convergence of a trajectory of (2) to a critical point will be established, by requiring that a regularization of the objective function fulfills the Kurdyka-Łojasiewicz property.

We again have to require additional assumptions on the involved functions. Since the required additional assumptions on the involved functions do not necessarily include, or are not as strong as the ones from the previous section, existence of a solution trajectory to (2) itself will be presupposed. For performing asymptotic analysis it suffices to require less regularity of the solution to (2) than we were able to obtain in Theorem 3 for the special system (7). More precisely, we just have to require the existence of a *strong global solution* to the dynamical system as defined in Definition 3.

Assumption 2. There exists a locally absolutely continuous solution $z = (x, y)$ to the dynamical system (2) with absolutely continuous derivative and an almost everywhere existing second derivative fulfilling the estimate $\|(\ddot{x}(t), \ddot{y}(t))\| \leq \beta(\lambda, \gamma_1, \gamma_2)\|(\dot{x}(t), \dot{y}(t))\|$ for a.e. $t \in [0, +\infty)$ and some constant $\beta > 0$, possibly depending on λ and the stepsizes γ_1, γ_2 .

Remark 4. The statement of Assumption 2 is true for the cases when f and g are convex and $\lambda = 1$ or $\mu = 1$, as it was shown in the previous section. Nevertheless, as we will see as follows, the analysis of the asymptotic behavior of the trajectory can be carried out in a more general setting, provided that the trajectory exists.

The first ingredient of the asymptotic analysis will be a *sufficient decrease* property of the derivative of the trajectory. For obtaining it, we need to control the time derivative of f and g along trajectories and therefore to assume that f and g admit a *chain rule*.

Definition 4. Let $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be a proper and lower semicontinuous function. We say that h admits the *chain rule* if, for every locally absolutely continuous arc $z : [0, +\infty) \rightarrow \mathbb{R}^d$ fulfilling $z(t) \in \text{dom } \partial h$ for a.e. $t \geq 0$ and such that $t \mapsto h(z(t))$ is locally absolutely continuous, it holds for a.e. $t \geq 0$

$$\frac{d}{dt}h(z(t)) = \langle \xi, \dot{z}(t) \rangle \text{ for all } \xi \in \partial h(z(t)). \quad (12)$$

Remark 5. • If h is, in addition, *convex*, then it admits the chain rule (see [21, Lemma 3.3], [7]). Notice that if $z : [0, +\infty) \rightarrow \mathbb{R}^d$ is a locally absolutely continuous arc such that $z(t) \in \text{dom } h$ and there exists $\xi(t) \in \partial h(z(t))$ for a.e. $t \geq 0$, and for every $T > 0$ it holds $z \in L^2([0, T], \mathbb{R}^d)$, $\dot{z} \in L^2([0, T], \mathbb{R}^d)$ and $\xi \in L^2([0, T], \mathbb{R}^d)$, then $t \mapsto h(z(t))$ is locally absolutely continuous.

- A locally Lipschitz function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ that is *subdifferentially regular* admits a chain rule (see [25, Lemma 5.4]). We recall that a locally Lipschitz function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ is called subdifferentially regular at a point $u \in \mathbb{R}^d$, if for all $\xi \in \partial h(u)$ the inequality

$$h(v) \geq h(u) + \langle \xi, v - u \rangle + o(\|v - u\|)$$

holds as $v \rightarrow u$.

Notice that in this case, if $z : [0, +\infty) \rightarrow \mathbb{R}^d$ is a locally absolutely continuous arc, then $t \mapsto h(z(t))$ is locally absolutely continuous, too, and the limiting subdifferential of h is nothing else than *Clarke's subdifferential* of h .

- A locally Lipschitz function $h : \mathbb{R}^d \rightarrow \mathbb{R}$, which is *Whitney C^1 -stratifiable*, admits a chain rule (see [25, Theorem 5.8]). A function $h : \mathbb{R}^d \rightarrow \mathbb{R}$ is called Whitney C^p -stratifiable, if its graph admits a Whitney C^p -stratification. A Whitney C^p -stratification \mathcal{A} of a set $Q \subseteq \mathbb{R}^d$ is a partition of Q into finitely many nonempty C^p smooth manifolds, called *strata*, satisfying the following two conditions:

- For any two strata $L, M \in \mathcal{A}$ it holds:

$$L \cap \text{cl}(M) \neq \emptyset \implies L \subset \text{cl}(M).$$

- For any sequence $(u_k)_{k \geq 0}$ in a stratum M converging to a point u in a stratum L , if the corresponding *normal vectors* $v_k \in N_M(u_k)$ converge to a vector v , then $v \in N_L(u)$ holds.

A set $M \subseteq \mathbb{R}^d$ is a C^p smooth manifold if there is an integer $r \in \mathbb{N}$ such that around any point $u \in M$, there is a neighborhood U and a C^p -smooth map $F : U \rightarrow \mathbb{R}^{d-r}$ with $\nabla F(u)$ of full rank and satisfying $M \cap U = \{y \in U : F(y) = 0\}$. Then the *tangent space* $T_M(u)$ to M at u is the null space of $\nabla F(u)$, and the *normal space* $N_M(u)$ to M at u is the orthogonal space to $T_M(u)$.

Functions having as graphs semialgebraic, subanalytic or even sets that are definable in an o-minimal structure are Whitney C^∞ -stratifiable.

Theorem 4. *Let f, g and H fulfill Assumption 1 and f, g admit a chain rule. Let further $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ be a strong global solution of the dynamical system (2) fulfilling Assumption 2, where the constants are chosen such that*

$$\min\{\gamma_1, \gamma_2\} > \beta(\lambda, \gamma_1, \gamma_2) \max\left\{(1 + \lambda + \mu + \lambda^2), (1 + \lambda + \mu + \mu^2)\right\} + \frac{\lambda + \mu}{2}. \quad (13)$$

Then the following statements are true:

1. $\dot{z} \in L^2([0, +\infty); \mathbb{R}^n \times \mathbb{R}^m)$ and $\lim_{t \rightarrow +\infty} \dot{z}(t) = 0$;
2. $\exists \lim_{t \rightarrow +\infty} \Psi(\dot{z}(t) + z(t)) \in \mathbb{R}$.

Proof. Using the characterization (6) of the proximal map in the system (2), we obtain that (9)-(10) hold for a.e. $t \geq 0$. Since f and g admit a chain rule (12), it holds

$$\begin{aligned} \frac{d}{dt} f(\dot{x}(t) + x(t)) &= -\left\langle \ddot{x}(t) + \dot{x}(t), \gamma_1 L \dot{x}(t) + \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right\rangle \\ \frac{d}{dt} g(\dot{y}(t) + y(t)) &= -\left\langle \ddot{y}(t) + \dot{y}(t), \gamma_2 L \dot{y}(t) + \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \right\rangle \end{aligned}$$

for a.e. $t \geq 0$. Using the identity

$$\begin{aligned} & \frac{d}{dt}H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \\ &= \left\langle \nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), \ddot{x}(t) + \dot{x}(t) \right\rangle + \left\langle \nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), \ddot{y}(t) + \dot{y}(t) \right\rangle, \end{aligned}$$

we can compute

$$\begin{aligned} & \frac{d}{dt}\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \\ &= \frac{d}{dt}H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) + \frac{d}{dt}f(\dot{x}(t) + x(t)) + \frac{d}{dt}g(\dot{y}(t) + y(t)) \\ &= \left\langle \ddot{x}(t) + \dot{x}(t), -\gamma_1 L \dot{x}(t) - \nabla_x H(x(t), (1-\mu)(\dot{y}(t) + y(t)) + \mu y(t)) + \nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \right\rangle \\ &+ \left\langle \ddot{y}(t) + \dot{y}(t), -\gamma_2 L \dot{y}(t) - \nabla_y H((1-\lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) + \nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \right\rangle \end{aligned}$$

for a.e. $t \geq 0$. After first factorizing out the components of the inner product and then using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \frac{d}{dt}\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \\ &= -\gamma_1 L \|\dot{x}(t)\|^2 - \gamma_2 L \|\dot{y}(t)\|^2 - \gamma_1 L \left\langle \ddot{x}(t), \dot{x}(t) \right\rangle - \gamma_2 L \left\langle \ddot{y}(t), \dot{y}(t) \right\rangle \\ &+ \left\langle \ddot{x}(t) + \dot{x}(t), \nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1-\mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right\rangle \\ &+ \left\langle \ddot{y}(t) + \dot{y}(t), \nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_y H((1-\lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \right\rangle \\ &\leq -\gamma_1 L \|\dot{x}(t)\|^2 - \gamma_2 L \|\dot{y}(t)\|^2 - \frac{\gamma_1 L}{2} \frac{d}{dt} \|\dot{x}(t)\|^2 - \frac{\gamma_2 L}{2} \frac{d}{dt} \|\dot{y}(t)\|^2 \\ &+ \|\ddot{x}(t) + \dot{x}(t)\| \|\nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1-\mu)(\dot{y}(t) + y(t)) + \mu y(t))\| \\ &+ \|\ddot{y}(t) + \dot{y}(t)\| \|\nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_y H((1-\lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t))\| \end{aligned}$$

for a.e. $t \geq 0$. In the next step we apply the Lipschitz continuity of the gradient of H , which yields

$$\begin{aligned} & \frac{d}{dt}\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \\ &\leq -\gamma_1 L \|\dot{x}(t)\|^2 - \gamma_2 L \|\dot{y}(t)\|^2 - \frac{\gamma_1 L}{2} \frac{d}{dt} \|\dot{x}(t)\|^2 - \frac{\gamma_2 L}{2} \frac{d}{dt} \|\dot{y}(t)\|^2 \\ &+ L(\|\dot{x}(t)\| + \mu \|\dot{y}(t)\|) \|\ddot{x}(t) + \dot{x}(t)\| + L(\lambda \|\dot{x}(t)\| + \|\dot{y}(t)\|) \|\ddot{y}(t) + \dot{y}(t)\| \\ &\leq -\gamma_1 L \|\dot{x}(t)\|^2 - \gamma_2 L \|\dot{y}(t)\|^2 - \frac{\gamma_1 L}{2} \frac{d}{dt} \|\dot{x}(t)\|^2 - \frac{\gamma_2 L}{2} \frac{d}{dt} \|\dot{y}(t)\|^2 \\ &+ L(\|\dot{x}(t)\| + \mu \|\dot{y}(t)\|) \|\ddot{x}(t)\| + L(\lambda \|\dot{x}(t)\| + \|\dot{y}(t)\|) \|\ddot{y}(t)\| \\ &+ L(\|\dot{x}(t)\| + \mu \|\dot{y}(t)\|) \|\dot{x}(t)\| + L(\lambda \|\dot{x}(t)\| + \|\dot{y}(t)\|) \|\dot{y}(t)\| \end{aligned}$$

for a.e. $t \geq 0$. Further, we have

$$\begin{aligned} & L(\|\dot{x}(t)\| + \mu \|\dot{y}(t)\|) \|\ddot{x}(t)\| + L(\lambda \|\dot{x}(t)\| + \|\dot{y}(t)\|) \|\ddot{y}(t)\| \\ &\leq L \sqrt{\|\ddot{x}(t)\|^2 + \|\ddot{y}(t)\|^2} \sqrt{(\|\dot{x}(t)\| + \mu \|\dot{y}(t)\|)^2 + (\lambda \|\dot{x}(t)\| + \|\dot{y}(t)\|)^2} \\ &= L \|(\ddot{x}(t), \ddot{y}(t))\| \sqrt{(1 + \lambda^2) \|\dot{x}(t)\|^2 + (1 + \mu^2) \|\dot{y}(t)\|^2 + 2(\lambda + \mu) \|\dot{x}(t)\| \|\dot{y}(t)\|} \\ &\leq L \beta(\lambda, \gamma_1, \gamma_2) \|(\dot{x}(t), \dot{y}(t))\| \sqrt{(1 + \lambda + \mu + \lambda^2) \|\dot{x}(t)\|^2 + (1 + \lambda + \mu + \mu^2) \|\dot{y}(t)\|^2}, \end{aligned} \tag{14}$$

for a.e. $t \geq 0$, where in the last inequality we used Assumption 2. By defining

$$M_{\lambda,\mu} := \max \left\{ (1 + \lambda + \mu + \lambda^2)^{\frac{1}{2}}, (1 + \lambda + \mu + \mu^2)^{\frac{1}{2}} \right\},$$

we can write

$$L(\|\dot{x}(t)\| + \mu\|\dot{y}(t)\|)\|\dot{x}(t)\| + L(\lambda\|\dot{x}(t)\| + \|\dot{y}(t)\|)\|\dot{y}(t)\| \leq L\beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu}(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2)$$

for a.e. $t \geq 0$. On the other hand, it holds

$$\begin{aligned} & L(\|\dot{x}(t)\| + \mu\|\dot{y}(t)\|)\|\dot{x}(t)\| + L(\lambda\|\dot{x}(t)\| + \|\dot{y}(t)\|)\|\dot{y}(t)\| \\ &= L(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2) + L(\lambda + \mu)\|\dot{x}(t)\|\|\dot{y}(t)\| \\ &\leq L\left(1 + \frac{\lambda + \mu}{2}\right)(\|\dot{x}(t)\|^2 + \|\dot{y}(t)\|^2) \end{aligned}$$

for a.e. $t \geq 0$. Combining the estimates above we obtain for a.e. $t \geq 0$ the following inequality

$$\begin{aligned} & \frac{d}{dt}\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \\ &\leq -\frac{\gamma_1 L}{2} \frac{d}{dt}\|\dot{x}(t)\|^2 - \frac{\gamma_2 L}{2} \frac{d}{dt}\|\dot{y}(t)\|^2 \\ &\quad - L\left(\gamma_1 - \beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu} - \frac{\lambda + \mu}{2}\right)\|\dot{x}(t)\|^2 - L\left(\gamma_2 - \beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu} - \frac{\lambda + \mu}{2}\right)\|\dot{y}(t)\|^2 \end{aligned}$$

or, equivalently,

$$\begin{aligned} & \frac{d}{dt}\left(\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) + \frac{1}{2}\left\|\left(\sqrt{\gamma_1 L}\dot{x}(t), \sqrt{\gamma_2 L}\dot{y}(t)\right)\right\|^2\right) \\ &\leq -L\left(\gamma_1 - \beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu} - \frac{\lambda + \mu}{2}\right)\|\dot{x}(t)\|^2 - L\left(\gamma_2 - \beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu} - \frac{\lambda + \mu}{2}\right)\|\dot{y}(t)\|^2 \leq 0, \end{aligned} \tag{15}$$

where we used that the stepsizes γ_1 and γ_2 fulfill (13). We will use the following notation for the positive constants in the above inequality

$$m_1 := L\left(\gamma_1 - \beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu} - \frac{\lambda + \mu}{2}\right) \text{ and } m_2 := L\left(\gamma_2 - \beta(\lambda, \gamma_1, \gamma_2)M_{\lambda,\mu} - \frac{\lambda + \mu}{2}\right).$$

The decreasing property (15) serves as inspiration for the definition of the Lyapunov functional, which will be used in the next subsection to show convergence of the trajectory. Integrating (15) from 0 to T , with $T > 0$ fixed, yields

$$\begin{aligned} & \Psi(\dot{x}(T) + x(T), \dot{y}(T) + y(T)) + \frac{1}{2}\left\|\left(\sqrt{\gamma_1 L}\dot{x}(T), \sqrt{\gamma_2 L}\dot{y}(T)\right)\right\|^2 + \int_0^T [m_1\|\dot{x}(t)\|^2 + m_2\|\dot{y}(t)\|^2] dt \\ &\leq \Psi(\dot{x}(0) + x_0, \dot{y}(0) + y_0) + \frac{1}{2}\left\|\left(\sqrt{\gamma_1 L}\dot{x}(0), \sqrt{\gamma_2 L}\dot{y}(0)\right)\right\|^2. \end{aligned} \tag{16}$$

By letting T converge to $+\infty$, we can easily deduce that $\dot{x} \in L^2([0, +\infty); \mathbb{R}^n)$, $\dot{y} \in L^2([0, +\infty); \mathbb{R}^m)$, and therefore

$$\dot{z} = (\dot{x}, \dot{y}) \in L^2([0, +\infty); \mathbb{R}^n \times \mathbb{R}^m),$$

as well as

$$\frac{d}{dt} \left(\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) + \frac{1}{2} \left\| \left(\sqrt{\gamma_1 L} \dot{x}(t), \sqrt{\gamma_2 L} \dot{y}(t) \right) \right\|^2 \right) \leq 0 \text{ for a.e. } t \geq 0. \quad (17)$$

Due to Assumption 2, we also obtain $\ddot{z} \in L^2([0, +\infty); \mathbb{R}^n \times \mathbb{R}^m)$. Since for a.e. $t \in [0, \infty)$

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

it follows from Lemma 2 that

$$\lim_{t \rightarrow +\infty} \dot{z}(t) = 0 \text{ and therefore also } \lim_{t \rightarrow +\infty} \dot{x}(t) = 0 \text{ and } \lim_{t \rightarrow +\infty} \dot{y}(t) = 0.$$

From (17) and Lemma 1 we can also conclude that

$$\exists \lim_{t \rightarrow +\infty} \Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \in \mathbb{R}.$$

□

4.2 A Lyapunov functional

The major aim of this subsection is to prove that the limit set of the solution trajectory is a subset of the set of critical points of the objective function Ψ . We recall that the *limit set* of a trajectory $z : [0, +\infty) \rightarrow \mathbb{R}^m \times \mathbb{R}^n$ of the dynamical system (2) is defined as

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

The main tool in our analysis is the following *Lyapunov functional* $\mathcal{H} : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \mapsto \bar{\mathbb{R}}$,

$$\mathcal{H}[(x, y), (u, v)] := \Psi(x, y) + \frac{1}{2} \left\| \left(\sqrt{\gamma_1 L}(x - u), \sqrt{\gamma_2 L}(y - v) \right) \right\|^2, \quad (18)$$

the definition of which is obviously inspired by (17). Note that

$$\partial \mathcal{H}[(x, y), (u, v)] = \left(\partial \Psi(x, y) + L(\gamma_1(x - u), \gamma_2(y - v)) \right) \times \left\{ -L(\gamma_1(x - u), \gamma_2(y - v)) \right\}$$

for every $(x, y), (u, v) \in \mathbb{R}^m \times \mathbb{R}^n$, thus

$$\begin{aligned} & \partial \mathcal{H}[(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), (x(t), y(t))] \\ &= \left(\partial \Psi((\dot{x}(t) + x(t), \dot{y}(t) + y(t)) + L(\gamma_1 \dot{x}(t), \gamma_2 \dot{y}(t)) \right) \times \left\{ -L(\gamma_1 \dot{x}(t), \gamma_2 \dot{y}(t)) \right\} \text{ for a.e. } t \geq 0. \end{aligned}$$

Instead of investigating the trajectory z itself, we go over to $(\dot{z} + z, z)$. We can prove the following properties of the Lyapunov functional \mathcal{H} .

Theorem 5. *Let f, g and H fulfill Assumption 1 and f, g admit a chain rule. Let further $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ be a strong global solution of the dynamical system (2) fulfilling Assumption 2, where the constants are chosen such that (13) holds. Then the following statements are true:*

(i) **decrease of \mathcal{H} :** for a.e. $t \geq 0$ it holds

$$\frac{d}{dt} \mathcal{H}[(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), (x(t), y(t))] \leq -\min\{m_1, m_2\} \|(\dot{x}(t), \dot{y}(t))\|^2 \leq 0, \quad (19)$$

as well as

$$\exists \lim_{t \rightarrow \infty} \mathcal{H}[(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), (x(t), y(t))] \in \mathbb{R}. \quad (20)$$

(ii) subgradient lower bound: for a.e. $t \geq 0$ it holds

$$\begin{aligned} & \left(\left(\nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right), -L \begin{pmatrix} \gamma_1 \dot{x}(t) \\ \gamma_2 \dot{y}(t) \end{pmatrix} \right) \\ & \in \partial \mathcal{H}[(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), (x(t) + y(t))] \end{aligned} \quad (21)$$

and

$$\begin{aligned} & \left\| \left(\left(\nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right), -L \begin{pmatrix} \gamma_1 \dot{x}(t) \\ \gamma_2 \dot{y}(t) \end{pmatrix} \right) \right\| \\ & \leq L \sqrt{\max\{(1 + \gamma_1^2 + \lambda^2), (1 + \gamma_2^2 + \mu^2)\}} \|\dot{x}(t), \dot{y}(t)\|. \end{aligned} \quad (22)$$

(iii) convergence to a critical point: it

$$\omega(z) \subset \text{crit}(\Psi) \quad (23)$$

and for every $\bar{z} = (\bar{x}, \bar{y}) \in \omega(z)$ and $t_k \rightarrow +\infty$ such that $z(t_k) = (x(t_k), y(t_k)) \rightarrow \bar{z}$ as $k \rightarrow +\infty$, we have

$$\mathcal{H}[(\dot{x}(t_k) + x(t_k), \dot{y}(t_k) + y(t_k)), (x(t_k), y(t_k))] \rightarrow \mathcal{H}[(\bar{x}, \bar{y}), (\bar{x}, \bar{y})] \text{ as } k \rightarrow +\infty. \quad (24)$$

Proof. **(i) decrease of \mathcal{H} :** Relation (19) follows directly from (15), while (20) can be deduced from Theorem 4 (2).

(ii) subgradient lower bound: Adding $\nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t))$ to (9) and $\nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t))$ to (10) one obtains for a.e. $t \geq 0$

$$\begin{aligned} & -\gamma_1 L \dot{x}(t) + \nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \\ & \in \partial_x(\tilde{f} + H)(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) \end{aligned} \quad (25)$$

and, respectively,

$$\begin{aligned} & -\gamma_2 L \dot{y}(t) + \nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \\ & \in \partial_y(\tilde{g} + H)(\dot{x}(t) + x(t), \dot{y}(t) + y(t)), \end{aligned} \quad (26)$$

where $\tilde{f} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$, $\tilde{f}(x, y) := f(x)$ and $\tilde{g} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \bar{\mathbb{R}}$, $\tilde{g}(x, y) := g(y)$. Since we have

$$\partial \Psi(z) = \left(\partial_x(\tilde{f} + H)(z), \partial_y(\tilde{g} + H)(z) \right) \text{ for all } z \in \mathbb{R}^m \times \mathbb{R}^n,$$

we can conclude

$$\begin{aligned} & \left(\nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \right) \\ & \left(\nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \right) \\ & \in \partial \Psi \begin{pmatrix} \dot{x}(t) + x(t) \\ \dot{y}(t) + y(t) \end{pmatrix} + L \begin{pmatrix} \gamma_1 \dot{x}(t) \\ \gamma_2 \dot{y}(t) \end{pmatrix} \text{ for a.e. } t \geq 0, \end{aligned}$$

which is equivalent to (21). Furthermore, we can estimate the subgradient using the Lipschitz-continuity of ∇H and obtain for a.e. $t \geq 0$

$$\begin{aligned}
& \left\| \left(\begin{array}{c} \nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t)) \\ \nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t)) \end{array} \right), - \begin{pmatrix} \gamma_1 L \dot{x}(t) \\ \gamma_2 L \dot{y}(t) \end{pmatrix} \right) \right\|^2 \\
&= \|\nabla_x H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_x H(x(t), (1 - \mu)(\dot{y}(t) + y(t)) + \mu y(t))\|^2 \\
&\quad + \|\nabla_y H(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) - \nabla_y H((1 - \lambda)(\dot{x}(t) + x(t)) + \lambda x(t), y(t))\|^2 \\
&\quad + \gamma_1^2 L^2 \|\dot{x}(t)\|^2 + \gamma_2^2 L^2 \|\dot{y}(t)\|^2 \\
&\leq L^2(1 + \gamma_1^2 + \lambda^2) \|\dot{x}(t)\|^2 + L^2(1 + \gamma_2^2 + \mu^2) \|\dot{y}(t)\|^2,
\end{aligned}$$

from which statement (22) follows.

(iii) convergence to a critical point: Let $(\bar{x}, \bar{y}) \in \omega(x, y)$ and $(t_k)_{k \geq 0}$ be such that $(x(t_k), y(t_k))$ converges to (\bar{x}, \bar{y}) as $k \rightarrow +\infty$. Due to Theorem 4 it also holds $\lim_{k \rightarrow +\infty} (\dot{x}(t_k), \dot{y}(t_k)) = (0, 0)$. We claim that

$$\lim_{k \rightarrow +\infty} (\tilde{f} + H)(\dot{x}(t_k) + x(t_k), \dot{y}(t_k) + y(t_k)) = (\tilde{f} + H)(\bar{x}, \bar{y})$$

and

$$\lim_{k \rightarrow +\infty} (\tilde{g} + H)(\dot{x}(t_k) + x(t_k), \dot{y}(t_k) + y(t_k)) = (\tilde{g} + H)(\bar{x}, \bar{y}).$$

In fact, since f and g are assumed to be lower semicontinuous, we have

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

and

$$\liminf_{k \rightarrow +\infty} g(\dot{y}(t_k) + y(t_k)) \geq g(\bar{y}).$$

On the other hand, we can conclude from the definition of the dynamical system (2) that for any $k \geq 0$

$$\begin{aligned}
& \dot{x}(t_k) + x(t_k) \\
& \in \operatorname{argmin}_{u \in \mathbb{R}^n} \left\{ f(u) + \frac{\gamma_1 L}{2} \|u - x(t_k)\|^2 + \left\langle u - x(t_k), \nabla_x H(x(t_k), (1 - \mu)(\dot{y}(t_k) + y(t_k)) + \mu y(t_k)) \right\rangle \right\}
\end{aligned}$$

and

$$\begin{aligned}
& \dot{y}(t_k) + y(t_k) \\
& \in \operatorname{argmin}_{v \in \mathbb{R}^m} \left\{ g(v) + \frac{\gamma_2 L}{2} \|v - y(t_k)\|^2 + \left\langle v - y(t_k), \nabla_y H((1 - \lambda)(\dot{x}(t_k) + x(t_k)) + \lambda x(t_k), y(t_k)) \right\rangle \right\},
\end{aligned}$$

which implies

$$\begin{aligned}
& f(\dot{x}(t_k) + x(t_k)) + \frac{\gamma_1 L}{2} \|\dot{x}(t_k)\|^2 + \left\langle \dot{x}(t_k), \nabla_x H(x(t_k), (1 - \mu)(\dot{y}(t_k) + y(t_k)) + \mu y(t_k)) \right\rangle \\
& \leq f(\bar{x}) + \frac{\gamma_1 L}{2} \|\bar{x} - x(t_k)\|^2 + \left\langle \bar{x} - x(t_k), \nabla_x H(x(t_k), (1 - \mu)(\dot{y}(t_k) + y(t_k)) + \mu y(t_k)) \right\rangle,
\end{aligned}$$

and

$$\begin{aligned}
& g(\dot{y}(t_k) + y(t_k)) + \frac{\gamma_2 L}{2} \|\dot{y}(t_k)\|^2 + \left\langle \dot{y}(t_k), \nabla_y H(\dot{x}(t_k) + x(t_k), y(t_k)) \right\rangle \\
& \leq g(\bar{y}) + \frac{\gamma_2 L}{2} \|\bar{y} - y(t_k)\|^2 + \left\langle \bar{y} - y(t_k), \nabla_y H((1 - \lambda)(\dot{x}(t_k) + x(t_k)) + \lambda x(t_k), y(t_k)) \right\rangle.
\end{aligned}$$

Therefore we can conclude under consideration of the continuity of ∇H

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

as well as

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

from which the claim follows. Using again continuity of ∇H we obtain for $k \rightarrow +\infty$

$$\begin{aligned} -\gamma_1 L \dot{x}(t_k) + \nabla_x H(\dot{x}(t_k) + x(t_k), \dot{y}(t_k) + y(t_k)) - \nabla_x H(x(t_k), (1 - \mu)(\dot{y}(t_k) + y(t_k)) + \mu y(t_k)) &\rightarrow 0, \\ -\gamma_2 L \dot{y}(t_k) + \nabla_y H(\dot{x}(t_k) + x(t_k), \dot{y}(t_k) + y(t_k)) - \nabla_y H((1 - \lambda)(\dot{x}(t_k) + x(t_k)) + \lambda x(t_k), y(t_k)) &\rightarrow 0. \end{aligned}$$

Using (25) and (26), and the closedness of the graph of the limiting subdifferential, it yields

$$0 \in \partial_x(\tilde{f} + H)(\bar{x}, \bar{y}) \text{ and } 0 \in \partial_y(\tilde{g} + H)(\bar{x}, \bar{y}),$$

which is equivalent to

$$0 \in \partial \Psi(\bar{x}, \bar{y})$$

and (23) is proven. The statement in (24) follows immediately. \square

Remark 6. The limit set $\omega(z)$ of the trajectory z is not empty if $z(t), t \in [0, +\infty)$, is bounded. This follows, for example, if the objective function Ψ is *coercive*, i.e.

$$\lim_{\|z\| \rightarrow +\infty} \Psi(z) = +\infty.$$

Furthermore, in this case it also follows that Ψ is bounded from below, the infimum is attained and all its sublevel sets are bounded. Furthermore, from (16), again with the notation $z(t) = (x(t), y(t))$, we have for all $T > 0$

$$\Psi(\dot{x}(T) + x(T), \dot{y}(T) + y(T)) \leq \Psi(\dot{x}(0) + x_0, \dot{y}(0) + y_0) + \frac{1}{2} \left\| \left(\sqrt{\gamma_1 L} \dot{x}(0), \sqrt{\gamma_2 L} \dot{y}(0) \right) \right\|^2.$$

Due to the boundedness of the sublevel sets of Ψ , we can conclude from the above estimate that $\dot{z} + z$ is bounded. Since $\lim_{t \rightarrow \infty} \dot{z}(t) = 0$, it follows that z is bounded.

Corollary 6. *Let f, g and H fulfill Assumption 1 and f, g admit a chain rule. Let further $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ be a strong global solution of the dynamical system (2) assumed to be bounded and fulfilling Assumption 2, where the constants are chosen such that (13) holds. Then the following statements are true:*

1. $\omega(\dot{z} + z, z) \subseteq \text{crit}(\mathcal{H}) = \{(u, u) \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n : u \in \text{crit}(\Psi)\}$;
2. $\lim_{t \rightarrow +\infty} \text{dist}((\dot{z}(t) + z(t), z(t)), \omega(\dot{z} + z, z)) = 0$;
3. *the set $\omega(\dot{z} + z, z)$ is nonempty, compact and connected;*
4. \mathcal{H} *is finite and constant on $\omega(\dot{z} + z, z)$.*

Proof. The statements 1., 2. and 4. are direct consequences of Theorem 5. For the proof of statement 3. we refer the reader to [3, Theorem 4.1.], where it is shown that the non-emptiness, compactness and connectedness of the limit set of a trajectory is a generic property for bounded trajectories, whose derivatives converge to 0 for $t \rightarrow +\infty$. \square

Remark 7. In the special case when $\mu = \lambda = 0$ one can prove that the statements in Theorem 4, Theorem 5 and also in Corollary 6 remain true under the weaker assumption on ∇H that its partial gradients $\nabla_x H(\cdot, y)$ and $\nabla_y H(x, \cdot)$ are Lipschitz continuous with global constants L_x and L_y in y , respectively, in x . This allows us to choose in the in the two inclusions of the dynamical system (2) as stepsizes $\gamma_1 L_x$ and $\gamma_2 L_y$, respectively. This means that the dynamical system is of the form

$$\begin{cases} \dot{x}(t) + x(t) \in \text{prox}_{\frac{1}{\gamma_1 L_x} f} \left(x(t) - \frac{1}{\gamma_1 L_x} \nabla_x H(x(t), \dot{y}(t) + y(t)) \right), \\ \dot{y}(t) + y(t) \in \text{prox}_{\frac{1}{\gamma_2 L_y} g} \left(y(t) - \frac{1}{\gamma_2 L_y} \nabla_y H(\dot{x}(t) + x(t), y(t)) \right), \\ (x(0), y(0)) = (x_0, y_0). \end{cases} \quad (27)$$

and has as discrete counterpart the PALM algorithm from [14].

The essential ingredient in the asymptotic analysis is proof of a decreasing property for the function $t \mapsto \frac{d}{dt} \Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t))$. While in the proof of Theorem 4 we used to this end the Lipschitz continuity of the full gradient ∇H when estimating the time derivative of Ψ along the trajectory $(x(t) + \dot{x}(t), y(t) + \dot{y}(t))$, we now obtain

$$\begin{aligned} \frac{d}{dt} \Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) &\leq -\gamma_1 L_x \|\dot{x}(t)\|^2 - \gamma_2 L_y \|\dot{y}(t)\|^2 - \frac{\gamma_1 L_x}{2} \frac{d}{dt} \|\dot{x}(t)\|^2 - \frac{\gamma_2 L_y}{2} \frac{d}{dt} \|\dot{y}(t)\|^2 \\ &\quad + L_x \|\dot{x}(t)\| \|\ddot{x}(t) + \dot{x}(t)\| + L_y \|\dot{y}(t)\| \|\ddot{y}(t) + \dot{y}(t)\| \text{ for a.e. } t \geq 0. \end{aligned}$$

Therefore, (14) simplifies to

$$\begin{aligned} L_x \|\dot{x}(t)\| \|\ddot{x}(t) + \dot{x}(t)\| + L_y \|\dot{y}(t)\| \|\ddot{y}(t) + \dot{y}(t)\| &\leq \sqrt{\|\ddot{x}(t)\|^2 + \|\ddot{y}(t)\|^2} \sqrt{L_x^2 \|\dot{x}(t)\|^2 + L_y^2 \|\dot{y}(t)\|^2} \\ &= \max \{L_x, L_y\} \|\ddot{x}(t), \ddot{y}(t)\| \|\dot{x}(t), \dot{y}(t)\| \\ &\leq \beta(\lambda, \gamma_1, \gamma_2) \max \{L_x, L_y\} \|\dot{x}(t), \dot{y}(t)\|^2 \text{ for a.e. } t \geq 0, \end{aligned}$$

where $\beta(\lambda, \gamma_1, \gamma_2)$ describes the constant from Assumption 2. Proceeding in the same way as in the proof of Theorem 4, we obtain, instead of (15), the following decreasing property for a.e. $t \geq 0$

$$\begin{aligned} &\frac{d}{dt} \left(\Psi(\dot{x}(t) + x(t), \dot{y}(t) + y(t)) + \frac{1}{2} \left\| \left(\sqrt{\gamma_1 L_x} \dot{x}(t), \sqrt{\gamma_2 L_y} \dot{y}(t) \right) \right\|^2 \right) \\ &\leq - (L_x \gamma_1 - \max \{L_x, L_y\} \beta) \|\dot{x}(t)\|^2 - (L_y \gamma_2 - \max \{L_x, L_y\} \beta) \|\dot{y}(t)\|^2 \\ &\leq 0, \end{aligned} \quad (28)$$

for γ_1 and γ_2 chosen such that

$$\gamma_1 > \frac{\max \{L_x, L_y\} \beta(\lambda, \gamma_1, \gamma_2)}{L_x} \text{ and } \gamma_2 > \frac{\max \{L_x, L_y\} \beta(\lambda, \gamma_1, \gamma_2)}{L_y}.$$

4.3 Global convergence through the KL property

In this subsection we will show that if the Lyapunov function satisfies the *Kurdyka-Łojasiewicz (KL) property*, then one can establish *global convergence* of the trajectory to a critical point of Ψ .

To recall the definition of the KL property, let us denote by Θ_η , for $\eta \in [0, +\infty]$, the set of all concave and continuous functions $\varphi : [0, \eta) \rightarrow \mathbb{R}$, which satisfy the following properties:

- $\varphi(0) = 0$;
- $\varphi \in C^1((0, \eta))$ and φ is continuous at 0;

- $\forall s \in (0, \eta) : \varphi'(s) > 0$.

Definition 5 (Kurdyka-Łojasiewicz property). Let $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be proper and lower semicontinuous.

1. The function h is said to satisfy the Kurdyka-Łojasiewicz (KL) property at a point $\bar{u} \in \text{dom } \partial h$, if there exist $\eta \in (0, +\infty]$, a neighborhood U of \bar{u} and a function $\varphi \in \Theta_\eta$ such that for every

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

the inequality

$$\varphi'(h(u) - h(\bar{u})) \text{dist}(0, \partial h(u)) \geq 1$$

holds. The function φ is called *desingularizing function* for h at point \bar{u} .

2. If h satisfies the KL property at every point in $\text{dom } \partial h$, then h is said to be a *KL function*.

To the class of KL functions belong semi-algebraic, real sub-analytic, semiconvex, uniformly convex and convex functions satisfying a growth condition, see [6, 13, 14]. However, for our analysis we will make use of the *uniform KL property*, which follows directly from the above definition, see [14, Lemma 3.6].

Lemma 7. Let $\Omega \subset \mathbb{R}^d$ be a compact set and $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ a proper and lower semicontinuous function. Assume that h is constant on Ω and it satisfies the KL property at each point of Ω . Then there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for every $\bar{u} \in \Omega$ and every u in the intersection

$$\{u \in \mathbb{R}^d : \text{dist}(u, \Omega) < \varepsilon\} \cap \{u \in \mathbb{R}^d : h(\bar{u}) < h(u) < h(\bar{u}) + \eta\}$$

the inequality

$$\varphi'(h(u) - h(\bar{u})) \text{dist}(0, \partial h(u)) \geq 1$$

holds.

With these preparations we are now ready to state the main result:

Theorem 8. Let f, g and H fulfill Assumption 1 and f, g admit a chain rule. Let further $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ be a strong global solution of the dynamical system (2) assumed to be bounded and fulfilling Assumption 2, where the constants are chosen such that (13) holds. Moreover, suppose that the Lyapunov function \mathcal{H} , defined as in (18), is a KL function. Then the following statements hold:

1. $\dot{z} = (\dot{x}, \dot{y}) \in L^1([0, \infty); \mathbb{R}^n \times \mathbb{R}^m)$;
2. there exists $\bar{z} = (\bar{x}, \bar{y}) \in \text{crit}(\Psi)$ such that $\lim_{t \rightarrow +\infty} z(t) = \lim_{t \rightarrow +\infty} (x(t), y(t)) = (\bar{x}, \bar{y}) = \bar{z}$.

Proof. Choose $\bar{z} \in \text{crit}(\Psi)$ such that $(\bar{z}, \bar{z}) \in \omega(\dot{z} + z, z)$. From Theorem 5 and Corollary 6 we also have $\lim_{t \rightarrow +\infty} \mathcal{H}[\dot{z}(t) + z(t), z(t)] = \mathcal{H}[\bar{z}, \bar{z}]$ and $\bar{z} \in \text{crit}(\Psi)$. We distinguish between two cases:

Case I: $\exists \bar{t} \geq 0$ such that $\mathcal{H}[\dot{z}(\bar{t}) + z(\bar{t}), z(\bar{t})] = \mathcal{H}[\bar{z}, \bar{z}]$.

Since from Theorem 5 (i) we have that \mathcal{H} is non-increasing along trajectories, we obtain for all $t \geq \bar{t}$

$$\mathcal{H}[\dot{z}(t) + z(t), z(t)] \leq \mathcal{H}[\dot{z}(\bar{t}) + z(\bar{t}), z(\bar{t})] = \mathcal{H}[\bar{z}, \bar{z}],$$

and therefore

$$t \mapsto \mathcal{H}[\dot{z}(t) + z(t), z(t)]$$

is constant on $[\bar{t}, +\infty)$. Due to the subgradient estimate in Theorem 5 (ii), we obtain for a.e. $t \in [\bar{t}, +\infty)$ that $\dot{z}(t) = (\dot{x}(t), \dot{y}(t)) = 0$ and, hence, $z(t) = (x(t), y(t))$ is constant on $[\bar{t}, +\infty)$, from which the conclusion follows.

Case II: $\forall t \geq 0 : \mathcal{H}[\dot{z}(t) + z(t), z(t)] > \mathcal{H}[\bar{z}, \bar{z}]$.

We use Lemma 7 for $\Omega := \omega(\dot{z} + z, z)$ and $h := \mathcal{H}$. According to it, there exist $\varepsilon, \eta > 0$ and $\varphi \in \Theta_\eta$ such that for all $(\bar{z}, \bar{z}) \in \omega(\dot{z} + z, z)$ and all (v, w) in the intersection

$$\mathcal{A} := \{(v, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m : \text{dist}((v, w), \omega(\dot{z} + z, z)) < \varepsilon\} \cap \{(v, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m : \mathcal{H}[\bar{z}, \bar{z}] < \mathcal{H}[v, w] < \mathcal{H}[\bar{z}, \bar{z}] + \eta\}$$

one has

$$\varphi'(\mathcal{H}[v, w] - \mathcal{H}[\bar{z}, \bar{z}]) \text{dist}((0, 0), \partial\mathcal{H}[v, w]) \geq 1.$$

On one hand, since

$$\lim_{t \rightarrow +\infty} \mathcal{H}[\dot{z}(t) + z(t), z(t)] = \mathcal{H}[\bar{z}, \bar{z}],$$

there exists $t_1 \geq 0$ such that

$$\mathcal{H}[\bar{z}, \bar{z}] < \mathcal{H}[v, w] < \mathcal{H}[\bar{z}, \bar{z}] + \eta$$

for every $t \geq t_1$. On the other hand, from

$$\lim_{t \rightarrow +\infty} \text{dist}((\dot{z}(t) + z(t), z(t)), \omega(\dot{z} + z, z)) = 0,$$

there exist $t_2 \geq 0$ such that

$$\text{dist}((\dot{z}(t) + z(t), z(t)), \omega(\dot{z} + z, z)) < \varepsilon.$$

Therefore, for a.e. $t \geq T := \max\{t_1, t_2\}$ it holds

$$(\dot{z}(t) + z(t), z(t)) \in \mathcal{A},$$

and hence

$$\varphi'(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}]) \text{dist}((0, 0), \partial\mathcal{H}[\dot{z}(t) + z(t), z(t)]) \geq 1.$$

From here we can conclude using (21) and (22) that

$$1 \leq \varphi'(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}]) L \sqrt{\max\{(1 + \gamma_1^2 + \lambda^2), (1 + \gamma_2^2 + \mu^2)\}} \|\dot{z}(t)\| \text{ for a.e. } t \geq T. \quad (29)$$

This yields the following estimation for the time derivative

$$\begin{aligned} & \frac{d}{dt} \varphi(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}]) \\ &= \varphi'(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}]) \frac{d}{dt} \mathcal{H}[\dot{z}(t) + z(t), z(t)] \\ &\leq \min\{m_1, m_2\} \left(L \sqrt{\max\{(1 + \gamma_1^2 + \lambda^2), (1 + \gamma_2^2 + \mu^2)\}} \right)^{-1} \|\dot{z}(t)\| \leq 0 \text{ for a.e. } t \geq T, \end{aligned} \quad (30)$$

which is due to the decreasing property (19) of the Lyapunov-function \mathcal{H} and inequality (29). After integration, since φ is bounded from below, we can deduce

$$\dot{z} \in L^1([0, \infty); \mathbb{R}^n \times \mathbb{R}^m).$$

This yields that the limit $\lim_{t \rightarrow +\infty} z(t)$ exists, from which the conclusion follows. \square

4.4 Convergence rates

This subsection is dedicated to the derivation convergence rates of the trajectory of the dynamical system (2) to a critical point of Ψ . Convergence rate results (see [5]) can be achieved considering a KL function, which satisfies Definition 5 with desingularizing function

$$\varphi(s) = \frac{1}{c}s^{1-\theta}, \quad \theta \in [0, 1), c > 0. \quad (31)$$

Definition 6 (Łojasiewicz property). Let $h : \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ be proper and lower semicontinuous.

1. The function h is said to satisfy the Łojasiewicz property at a point $\bar{u} \in \text{dom } \partial h$, if there exist $\eta \in (0, +\infty]$, a neighborhood U of \bar{u} , $c > 0$ and $\theta \in [0, 1)$ such that for every

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

it holds

$$|h(u) - h(\bar{u})|^\theta \leq c \|\xi\| \text{ for every } \xi \in \partial h(u).$$

The exponent θ is called the *Łojasiewicz exponent*.

2. If h has the KL property and has the same Łojasiewicz exponent θ at every point in $\text{dom } \partial h$, then we say that h is a Łojasiewicz function with an exponent of θ .

It is known that analytic functions and proper lower semicontinuous semi-algebraic functions have the Łojasiewicz property (see [6, 29]). Calculus rules for the Łojasiewicz exponent have been developed in [28].

Theorem 9. *Let f, g and H fulfill Assumption 1 and f, g admit a chain rule. Let further $z = (x, y) : [0, +\infty) \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ be a strong global solution of the dynamical system (2) assumed to be bounded and fulfilling Assumption 2, where the constants are chosen such that (13) holds. Moreover, suppose that the Lyapunov function \mathcal{H} , defined as in (18), has the Łojasiewicz property. Then there exists $\bar{z} = (\bar{x}, \bar{y}) \in \text{crit}(\Psi)$ such that $\lim_{t \rightarrow +\infty} z(t) = \lim_{t \rightarrow +\infty} (x(t), y(t)) = (\bar{x}, \bar{y}) = \bar{z}$. Let $\theta \in [0, 1)$ be the Łojasiewicz exponent of \mathcal{H} at (\bar{z}, \bar{z}) . Then there exist $\alpha, \beta, \gamma, \delta > 0$ and a time $T \geq 0$ such the following statements are true:*

1. *if $\theta \in [0, \frac{1}{2})$, then $z = (x, y)$ converges in finite time;*
2. *if $\theta = \frac{1}{2}$, then $\|z(t) - \bar{z}\| \leq \alpha e^{-\beta t} \forall t \geq T$;*
3. *if $\theta \in (\frac{1}{2}, 1)$, then $\|z(t) - \bar{z}\| \leq (\gamma t + \delta)^{-\frac{1-\theta}{2\theta-1}} \forall t \geq T$.*

Proof. If there exists $\bar{t} \geq 0$ such that $\mathcal{H}[\dot{z}(\bar{t}) + z(\bar{t}), z(\bar{t})] = \mathcal{H}[\bar{z}, \bar{z}]$, then the trajectory converges in finite time, as we saw in the first case of the proof of Theorem 8.

Let us therefore assume that for all $t \geq 0$ the inequality $\mathcal{H}[\dot{z}(t) + z(t), z(t)] > \mathcal{H}[\bar{z}, \bar{z}]$ holds. We define the *arc-length* of the trajectory z from time $t \geq 0$ on by

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

We immediately obtain

$$\|\dot{z}(t)\| = -\dot{\sigma}(t) \text{ for a.e. } t \geq 0, \quad (32)$$

and by calculating for some time $\tau \geq t$

$$\|z(t) - \bar{z}\| = \left\| z(\tau) - \bar{z} - \int_t^\tau \dot{z}(s) \, ds \right\| \leq \|z(\tau) - \bar{z}\| + \int_t^\tau \|\dot{z}(s)\| \, ds,$$

and then taking the limit $\tau \rightarrow +\infty$, we obtain the following estimate for the distance of the trajectory to its limit point \bar{z}

$$\|z(t) - \bar{z}\| \leq \sigma(t) \text{ for all } t \geq 0.$$

From the considerations in (30) in the proof of Theorem 8, while taking into account the special shape of the desingularizing function (31), we get that there exists $T > 0$ such that for a.e. $t \geq T$

$$M_1 \|\dot{z}(t)\| + c \frac{d}{dt} \left(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}] \right)^{1-\theta} \leq 0,$$

where we define

$$M_1 := \min \{m_1, m_2\} \left(L \sqrt{\max \{(1 + \gamma_1^2 + \lambda^2), (1 + \gamma_2^2 + \mu^2)\}} \right)^{-1}.$$

Integration from t to \tilde{T} , where $\tilde{T} \geq T$, yields

$$M_1 \int_t^{\tilde{T}} \|\dot{z}(s)\| \, ds + c \left(\mathcal{H}[\dot{z}(\tilde{T}) + z(\tilde{T}), z(\tilde{T})] - \mathcal{H}[\bar{z}, \bar{z}] \right)^{1-\theta} \leq c \left(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}] \right)^{1-\theta},$$

which, after letting $\tilde{T} \rightarrow +\infty$, produces the estimate

$$M_1 \sigma(t) \leq c \left(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}] \right)^{1-\theta} \quad (33)$$

for a.e. $t \geq T$. In Theorem 5 (ii) we provided an element $(\xi(t), \eta(t)) \in \partial \mathcal{H}[\dot{z}(t) + z(t), z(t)]$ such that

$$\|(\xi(t), \eta(t))\| \leq M_2 \|\dot{z}(t)\| \quad \text{for a.e. } t \geq 0,$$

where we defined

$$M_2 := L \sqrt{\max \{(1 + \gamma_1^2 + \lambda^2), (1 + \gamma_2^2 + \mu^2)\}}.$$

Using that \mathcal{H} has the Łojasiewicz property θ at (\bar{z}, \bar{z}) we further have that for a.e. $t \geq T$ (after potentially increasing T in order to guarantee that $(\dot{z}(t) + z(t), z(t))$ is in the neighborhood on which the Łojasiewicz property is fulfilled)

$$\left(\mathcal{H}[\dot{z}(t) + z(t), z(t)] - \mathcal{H}[\bar{z}, \bar{z}] \right)^\theta \leq c \|(\xi(t), \eta(t))\| \leq c M_2 \|\dot{z}(t)\|.$$

Combining this with estimate (33) gives

$$M_1 \sigma(t) \leq c^{\frac{1}{\theta}} M_2^{\frac{1-\theta}{\theta}} \|\dot{z}(t)\|^{\frac{1-\theta}{\theta}} \quad \text{for a.e. } t \geq T.$$

Taking into account (32), we are able to conclude

$$\dot{\sigma}(t) \leq -C \sigma(t)^{\frac{\theta}{1-\theta}} \quad \text{for a.e. } t \geq T, \quad (34)$$

where we denote

$$C := \frac{M_1^{\frac{\theta}{1-\theta}}}{c^{\frac{1}{1-\theta}} M_2} > 0.$$

From (34) we now can deduce the speed of convergence of the trajectory. To this end we distinguish between different values for the Łojasiewicz exponent θ .

- If $\theta = \frac{1}{2}$, then (34) simplifies to

$$\dot{\sigma}(t) \leq -C\sigma(t),$$

from which exponential convergence

$$\|z(t) - \bar{z}\| \leq \sigma(t) \leq \alpha e^{-\beta t} \quad \text{for a.e. } t \geq T,$$

with $\alpha = \sigma(T)$ and $\beta = C$, follows by using Gronwall's inequality.

- If $\theta \in [0, \frac{1}{2})$, then (34) yields

$$\frac{d}{dt}\sigma(t)^{\frac{1-2\theta}{1-\theta}} = \frac{1-2\theta}{1-\theta}\sigma(t)^{-\frac{\theta}{1-\theta}}\dot{\sigma}(t) \leq -C\frac{1-2\theta}{1-\theta} \quad \text{for a.e. } t \geq T.$$

After integration we obtain

$$\sigma(t)^{\frac{1-2\theta}{1-\theta}} \leq -c_1 t + c_2 \quad \text{for all } t \geq T,$$

for some constants $c_1, c_2 > 0$. This means that there has to exist some $\bar{T} > 0$ such that $\sigma(t) = 0$ for all $t \geq \bar{T}$, which implies that $z(t)$ has to be constant on $[\bar{T}, +\infty)$.

- Last, we investigate the case $\theta \in (\frac{1}{2}, 1)$. Since $\frac{1-2\theta}{1-\theta} < 0$ we obtain similarly to the calculation above

$$\frac{d}{dt}\sigma(t)^{\frac{1-2\theta}{1-\theta}} \geq C\frac{2\theta-1}{1-\theta} \quad \text{for a.e. } t \geq T.$$

Again, by integrating the inequality above we obtain for some constants $\gamma, \delta \in \mathbb{R}$

$$\sigma(t) \leq (\gamma t + \delta)^{\frac{1-2\theta}{1-\theta}} \quad \text{for all } t \geq T.$$

□

5 Numerical experiments

In this last section we aim to illustrate the analytical results with some numerical simulations. Especially, we want to focus on the effect the parameters λ as well the stepsizes γ_1 and γ_2 have on the asymptotic behavior of the solution trajectories. As in Section 3, we consider the case $\mu = 1$, where the dynamical system can be written in the explicit form (7). The simulations were carried out in MATLAB, where we used the function `ode15s`.

Example 1. In the first experiment, we investigated the optimization problem

$$\min_{(x,y) \in \mathbb{R} \times \mathbb{R}} \Psi(x, y) := |x| + |y| + (1 - x - y)^2, \quad (35)$$

which corresponds to problem (1) in the special setting $f, g : \mathbb{R} \rightarrow \mathbb{R}$, $f = g = |\cdot|$, and $H : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $H(x, y) = (1 - x - y)^2$. The critical points of Ψ are given by the set

$$\text{crit}(\Psi) = \left\{ (\bar{x}, \bar{y}) \in \mathbb{R}_+ \times \mathbb{R}_+ : \bar{x} + \bar{y} = \frac{1}{2} \right\},$$

where all elements of this set are optimal solutions for (35). We chose $(x_0, y_0) = \left(1, \frac{1}{2}\right)$ as starting point.

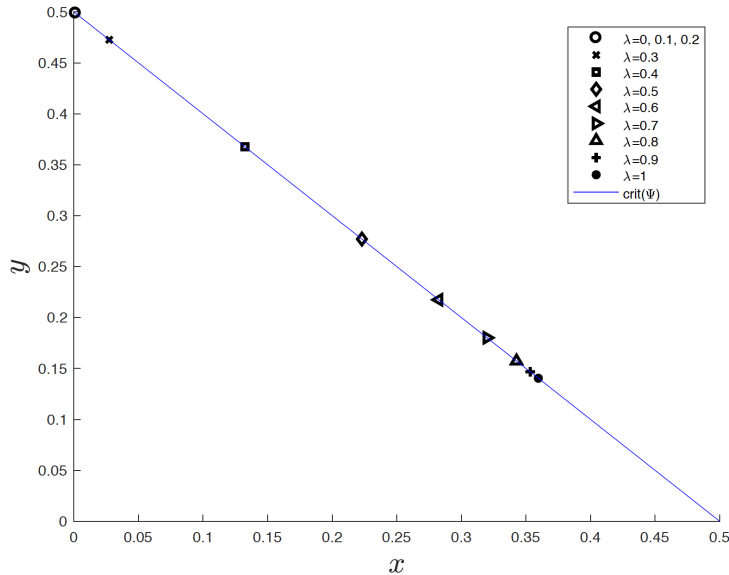


Figure 1: For $c_1 = c_2 = 1$ and $(x_0, y_0) = \left(1, \frac{1}{2}\right)$ the limits of the trajectory change according to the choice of λ . While for $\lambda \in [0, 0.3)$ (rather implicit regime), the solution converges to the critical point $(0, \frac{1}{2})$, for $\lambda \in [0.3, 1]$ the trajectory approaches other points further down the line of $\text{crit}(\Psi)$. The points are plotted after 100 time steps.

The first observation was that the rate of implicitity, which is given by the value of λ in the second component of the system (7), plays a crucial role in determining to which critical point the solution trajectory converges, provided to be in a setting where the stepsizes

$$c_1 := \gamma_1 L \text{ and } c_2 := \gamma_2 L$$

are chosen such that (13) is fulfilled for all $\lambda \in [0, 1]$ (see Figure 2). One should also remark that the rate of implicitity seems to influence the stability of convergence also in dependence of the shape of the problem. Here, for higher values of λ , this means for more explicit systems, the trajectory converges faster, while for more implicit systems, which arise when λ is close to 0, the trajectory first moves away from the limit set until the convergence starts, as one can see as a peak in the functions for the x -coordinates. For an overview over the limit points the different trajectories converge to we refer to Figure 1.

In addition to the above observations we further investigated the problem (35) in higher dimensions, namely

$$\min_{(x,y) \in \mathbb{R}^2 \times \mathbb{R}^2} \Psi(x, y) := \|x\|_1 + \|y\|_1 + (1 - x_1 - x_2 - y_1 - y_2)^2, \quad (36)$$

in order to emphasize the different convergence behavior of the solution trajectories with respect to the parameter λ . Here, $f, g : \mathbb{R}^2 \rightarrow \mathbb{R}$ are both given by the 1-norm and $H : \mathbb{R}^4 \rightarrow \mathbb{R}$ is of the form $H(x, y) = (1 - x_1 - x_2 - y_1 - y_2)^2$, where $x = (x_1, x_2)$, $y = (y_1, y_2)$. The set of critical points of Ψ are given by the set

$$\text{crit}(\Psi) = \left\{ (\bar{x}, \bar{y}) \in \mathbb{R}_+^2 \times \mathbb{R}_+^2 : \bar{x}_1 + \bar{x}_2 + \bar{y}_1 + \bar{y}_2 = \frac{1}{2} \right\},$$

where all elements of this set are optimal solutions for (36). We chose as starting value $(x_0, y_0) = \left(-1, -2, -\frac{1}{2}, -4\right)$. Again, we were able to observe that for $c_1 = c_2 = 5$, which is a setting in which

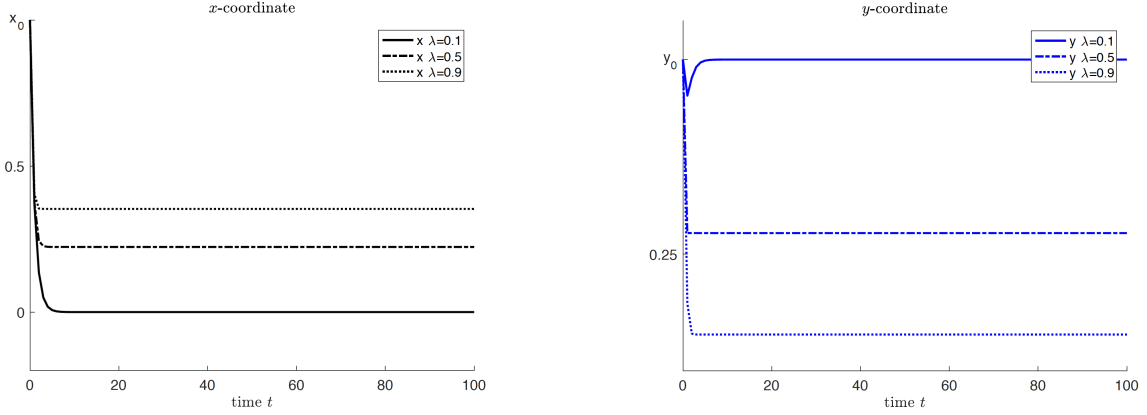


Figure 2: For $c_1 = c_2 = 1$ and $(x_0, y_0) = \left(1, \frac{1}{2}\right)$, three different solution trajectories approaching three different critical values are plotted. *On the left:* for $\lambda = 0.1$, the x -coordinate of the solution $(x(t), y(t))$ converges to $\bar{x} = 0$ (black solid), for $\lambda = 0.5$, to $\bar{x} = 0.223$ (black dashed) and, for $\lambda = 0.9$, to $\bar{x} = 0.3533$ (black dotted). *On the right:* for $\lambda = 0.1$, $y(t)$ converges to $\bar{y} = 0.5$ (blue solid), for $\lambda = 0.5$, to $\bar{y} = 0.277$, (blue dashed) and, for $\lambda = 0.9$, to $\bar{y} = 0.1467$ (blue dotted).

(13) is fulfilled for all $\lambda \in [0, 1]$, the limit points of the trajectory are determined by the choice of λ . This behaviour can be seen in Figure 3. Once more one can notice that also here the more explicit the system is (i.e. the closer λ is to 1), the more stable the trajectory converges to a limit point.

Moreover, we could observe that the validity of inequality (13), which in our special setting is of the form

$$\min\{c_1 L, c_2 L\} > \left(6 + \frac{4}{c_1^2 L^2} + \frac{4 + 24(1 - \lambda)^2}{c_2^2 L^2} + \frac{16(1 - \lambda)^2}{c_1^2 c_2^2 L^4}\right)^{\frac{1}{2}} (1 + \lambda + \lambda^2) + \frac{\lambda}{2},$$

is essential for the solution trajectory to converge. The plots of Figure 4 show that the smaller λ is chosen, the bigger the stepsizes have to be: $c_1 = c_2 = 0.5$ just provide convergence for $\lambda \in [0.3, 1]$, while for the very implicit system, i.e. for $\lambda \in [0, 0.3]$ it fails. Bigger stepsizes $c_1 = c_2 = 0.9$ assure convergence for a wider range of implicitity, i.e. for $\lambda \in [0.1, 1]$, while choosing $c_1 = c_2 = 1$ guarantees convergence for all $\lambda \in [0, 1]$.

Example 2. In the second numerical experiment we considered the minimization problem

$$\min_{(x,y) \in \mathbb{R} \times \mathbb{R}} \Psi(x, y) := |x| + L_2(y) - \frac{1}{5}(1 - x - y)^2,$$

where, for $\alpha > 0$,

$$L_\alpha : \mathbb{R} \rightarrow \mathbb{R}, \quad L_\alpha(y) := \begin{cases} \frac{y^2}{2}, & |y| \leq \alpha \\ \alpha|y| - \frac{\alpha^2}{2} & \text{otherwise,} \end{cases}$$

defines the *Huber function*. Taking $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) := |x|$, and $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(y) := L_2(y)$, as well as $H : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, $H(x, y) := (1 - x - y)^2$, we are again in the setting of problem (7). The set of critical points of Ψ is given by

$$\text{crit}(\Psi) = \left\{ \left(-\frac{1}{2}, 1\right); \left(0, -\frac{2}{3}\right); \left(\frac{5}{2}, 1\right) \right\},$$

while $\left(0, -\frac{2}{3}\right)$ is the optimal solution.

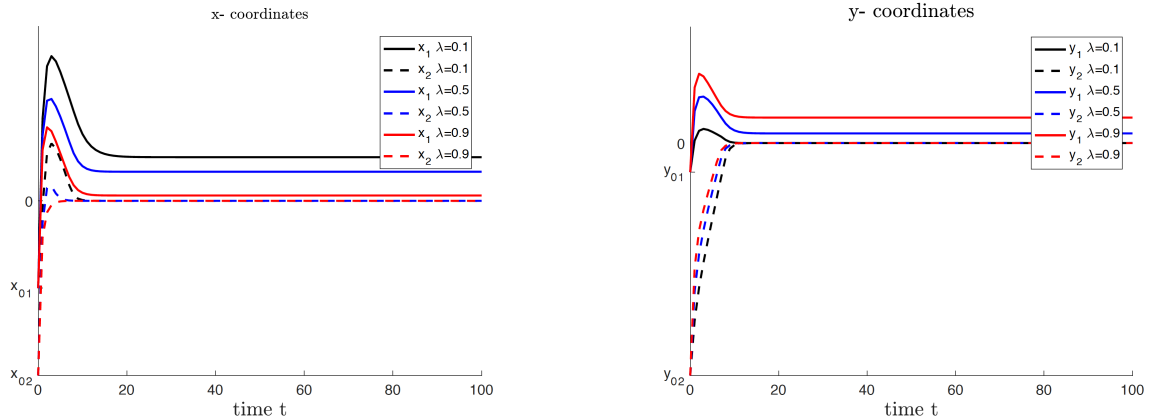


Figure 3: *On the left:* $x(t) \in \mathbb{R}^2$ converges to $\bar{x} = (0.5, 0)$, for $\lambda = 0.1$ (black), to $\bar{x} = (0.3333, 0)$, for $\lambda = 0.5$ (blue), and to $(0.0612, 0)$, for $\lambda = 0.9$ (red). *On the right:* $y(t) \in \mathbb{R}^2$ converges to $\bar{y} = (0, 0)$, for $\lambda = 0.1$ (black), to $\bar{y} = (0.1666, 0)$, for $\lambda = 0.5$ (blue), and to $\bar{y} = (0.4388, 0)$ for $\lambda = 0.9$ (red).

In Figure 5 again the x - and y -solution trajectories for three dynamical systems with different rates of implicitness are plotted. As starting value $(x_0, y_0) = (1, -1)$ was chosen. The Lipschitz constant of H allows to choose as stepsizes $c_1 = c_2 = 0.3$ in order to verify (13). One can observe that the values of λ do not affect the convergence of the x -trajectories at all, while for the y -trajectories the implicitness has influence on the stability of convergence. For the relatively explicit system with $\lambda = 0.9$ one can see that the solution first moves away from the limit point, before the convergence starts. The trajectory corresponding to $\lambda = 0.5$ reaches the limit point the fastest.

6 Conclusion and further perspectives

In this paper we presented a way to tackle the minimization of the sum of two nonsmooth, possibly also nonconvex, functions in separate variables coupled by a smooth function via a continuous time approach of forward-backward type. The associated dynamical system was defined in a general way by leaving the rate of implicitness in both components of the system variable.

We first aimed to show existence of a unique solution, which we could achieve in the setting of asking convexity for the nonsmooth part of the objective function as well as assuming that one of the two components of the system is of explicit structure. Further, we were able to show the convergence of the solution trajectory - assuming it exists - in a more general setting, where convexity of the nonsmooth part of the objective function is not needed. We could further weaken the assumptions on the rates of implicitness of the dynamical system in the sense that they, together with the stepsizes, have to fulfill technical but not unreasonable inequalities, which were arising in order to obtain sufficient decrease of a Lyapunov functional, which had the role of being the key ingredient in proving convergence of the trajectory to a critical point of the objective function. Another not less important tool for the asymptotic analysis was the Kurdyka-Łojasiewicz (KL) property, which we asked a regularization of the just mentioned Lyapunov functional to have. This allowed us also to prove global convergence for the trajectories, but also convergence rates in the case where the Lyapunov functional is KL with a desingularization function of exponential structure. These convergence rates can be computed explicitly depending on the Łojasiewicz exponent. Numerical simulations of the last section underlay the analytical results and further give evidence of the necessity of the above mentioned inequalities

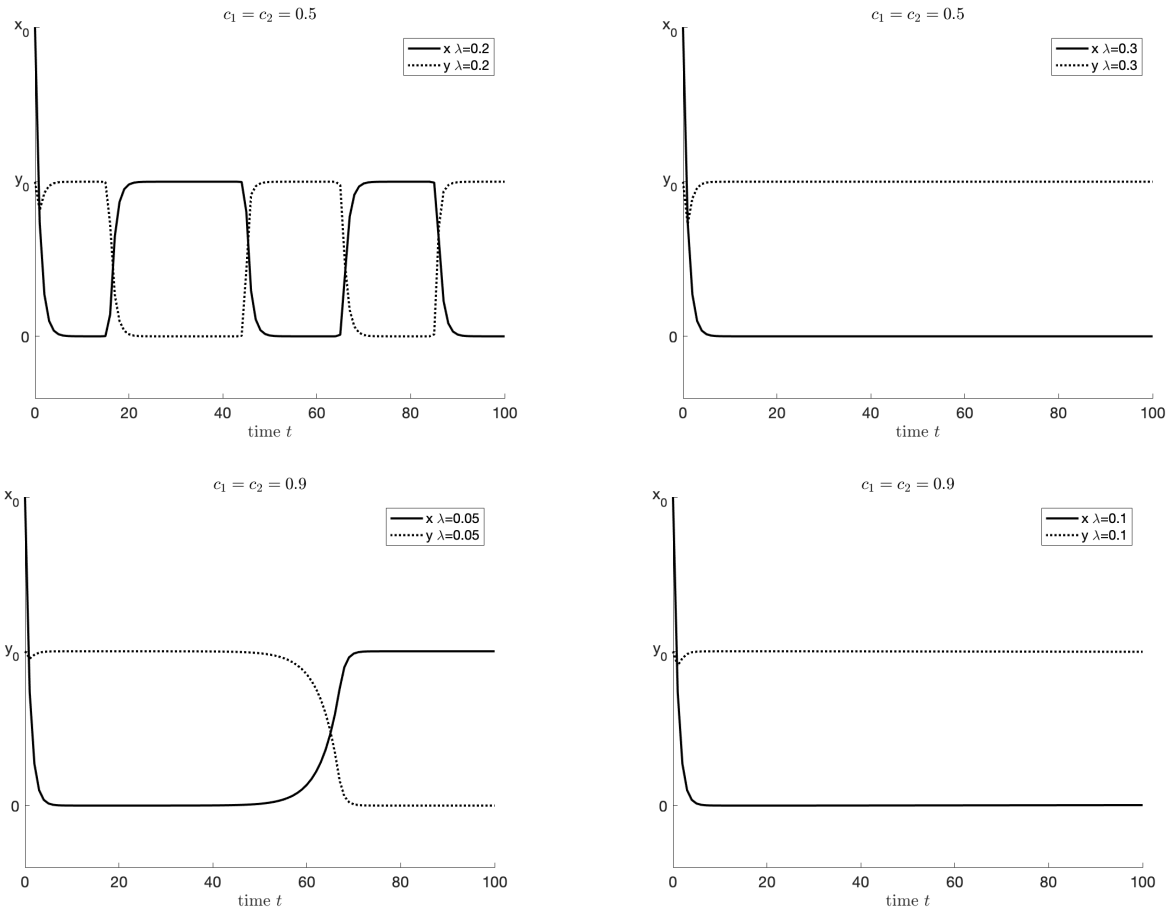


Figure 4: *Top row:* on the left, $(x(t), y(t))$ moves periodically between the critical points $(0, 0.5)$ and $(0.5, 0)$ for $c_1 = c_2 = 0.5$ and $\lambda = 0.2$. On the right one can see the trajectory converging for $c_1 = c_2 = 0.5$ and $\lambda = 0.3$. *Bottom row:* on the left, $(x(t), y(t))$ shows periodic behavior for $c_1 = c_2 = 0.9$ and $\lambda = 0.05$, while on the right, for $\lambda = 0.1$, the trajectory converges.

involving the stepsizes and the rate of implicitity to be valid.

We want to notice that several results, including those which state that the limit points of the trajectory are critical points of the objective function, are valid also in infinite dimensional Hilbert spaces. It is the use of the KL property for the proof of the global convergence and the derivation of the convergence rates which makes it important to work in finite dimensional spaces.

This work describes a first step in solving minimization problems of this special structure (1) via the approach of analyzing a system of differential inclusions using the special shape of the objective function. Above summary of results, achieved in this general setting, naturally leaves some further directions, which can be addressed in future investigation.

We already observed in the introduction that for specific choices of implicitity we can reconstruct, by performing explicit time-discretization with time-stepsize one, the PALM algorithm [14], which originally served as key motivation to study this problem. Considering other configurations of parameters as well as different types of discretization methods of ODEs could give rise to new numerical algorithms, maybe with improved performance to already existing schemes. Interesting applications resulting in an optimization problem of structure (1) are usually matrix factorization problems, where

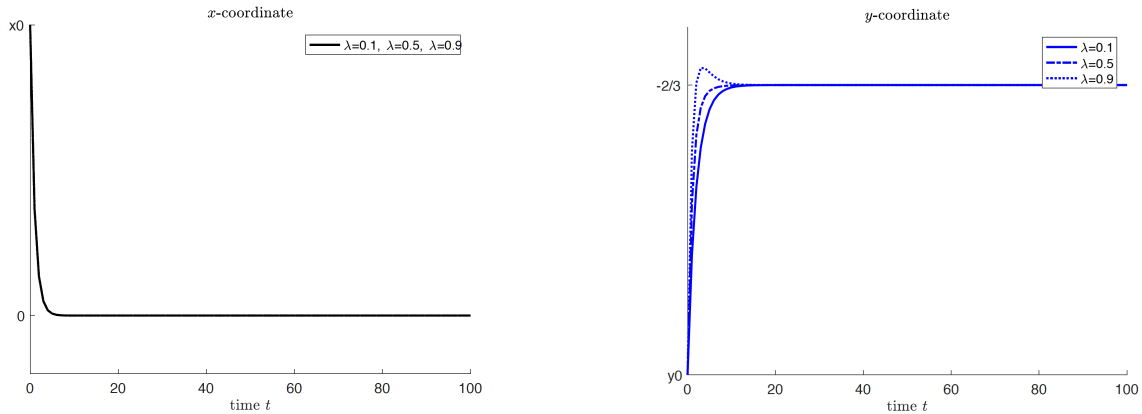


Figure 5: *On the left:* $x(t)$ converges to $\bar{x} = 0$, for $\lambda = 0.1$, $\lambda = 0.5$, $\lambda = 0.9$. *On the right:* $y(t)$ converges to $\bar{y} = -\frac{2}{3}$ with different convergence behavior, for $\lambda = 0.1$, $\lambda = 0.5$, $\lambda = 0.9$

in [14] numerical simulations were performed.

Further, the question of improving the convergence and of generating a mechanism for detecting different minimizers by adding inertia to the system would be a clear next step, which would result in investigating a second order system of differential inclusions. Another natural extension of our above analysis would be to develop a continuous time approach and to derive from here new iterative schemes for problems of type (1) with additional structure. Recently, in [20], a full splitting algorithm was investigated in the fully nonconvex setting for such a problem, where one of the nonsmooth components consists of a composition of a nonsmooth function with a linear operator.

Acknowledgements. The authors are thankful to two anonymous reviewers for their remarks and recommendations which improved the quality of the paper.

References

- [1] B. Abbas, H. Attouch, *Dynamical systems and forward-backward algorithms associated with the sum of a convex subdifferential and a monotone cocoercive operator*, Optimization 64(10), 2223–2252 (2015)
- [2] B. Abbas, H. Attouch, B. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, Journal of Optimization Theory and Applications 161, 331–360 (2014)
- [3] F. Alvarez, H. Attouch, J. Bolte, P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, Journal de Mathématiques Pures et Appliquées (9) 81(8), 747–779 (2002)
- [4] A.S. Antipin, *Minimization of convex functions on convex sets by means of differential equations*, (Russian) Differential'nye Uravneniya 30(9), 1475–1486 (1994)
- [5] H. Attouch, J. Bolte. *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Mathematical Programming 116(1) Series B, 5–16 (2007)

- [6] H. Attouch, J. Bolte, B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137(1-2) Series A, 91–129 (2013)
- [7] H. Attouch, G. Buttazzo and G. Michaille, *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization*, MOS-SIAM Series on Optimization, Philadelphia (2014)
- [8] H. Attouch, B. F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49, 574–598 (2011)
- [9] J.B. Baillon, H. Brézis, *Une remarque sur le comportement asymptotique des semigroupes non linéaires*, Houston Journal of Mathematics 2(1), 5–7 (1976)
- [10] S. Banert, R.I. Boţ, *A forward-backward-forward differential equation and its asymptotic properties*, Journal of Convex Analysis 25(2), 371–388 (2018)
- [11] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer (2017)
- [12] J. Bolte, *Continuous gradient projection method in Hilbert spaces*, Journal of Optimization Theory and its Applications 119(2), 235–259 (2003)
- [13] J. Bolte, A. Daniilidis, A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM Journal on Optimization 17(4), 1205–1223 (2006)
- [14] J. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming 146 (1-2) Series A, 459–494 (2014)
- [15] R.I. Boţ, E.R. Csetnek, *A dynamical system associated with the fixed points set of a nonexpansive operator*, Journal of Dynamics and Differential Equations 29(1), 155–168 (2017)
- [16] R.I. Boţ, E.R. Csetnek, *A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function*, ESAIM: Control, Optimisation and Calculus of Variations 24(2), 463–477 (2018)
- [17] R.I. Boţ, E.R. Csetnek, S.C. László, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO Journal on Computational Optimization 4(1), 3–25 (2016)
- [18] R.I. Boţ, E.R. Csetnek, S.C. László, *Approaching nonsmooth nonconvex minimization through second order proximal-gradient dynamical systems*, Journal of Evolution Equations 18(3), 1291–1318 (2018)
- [19] R.I. Boţ, E.R. Csetnek, S.C. László, *A primal-dual dynamical approach to structured convex minimization problems*, Journal of Differential Equations 269(5), 10717–10757 (2020)
- [20] R.I. Boţ, E.R. Csetnek, D.-K. Nguyen, *A proximal minimization algorithm for structured nonconvex and nonsmooth problems*, SIAM Journal on Optimization 29(2), 1300–1328 (2019)
- [21] H. Brézis, *Operateurs Maximaux Monotones: Et Semi-Groupes De Contractions Dans Les Espaces De Hilbert*, North-Holland Mathematics Studies 5, North Holland (1973)

- [22] R.E. Bruck, Jr., *Asymptotic convergence of nonlinear contraction semigroups in Hilbert space*, Journal of Functional Analysis 18, 15–26 (1975)
- [23] M.G. Crandall, A. Pazy, *Semi-groups of nonlinear contractions and dissipative sets*, Journal of Functional Analysis 3, 376–418 (1969)
- [24] E.R. Csetnek, Y. Malitsky, M. K. Tam, *Shadow Douglas-Rachford splitting for monotone inclusions*, Applied Mathematics and Optimization 80, 665–678 (2019)
- [25] D. Davis, D. Drusvyatskiy, S. Kakade, J. D. Lee, *Stochastic subgradient method converges on tame functions*, Foundations of Computational Mathematics (2019)
- [26] A. Haraux, *Systèmes Dynamiques Dissipatifs et Applications*, Recherches en Mathématiques Appliquées 17, Masson, Paris (1991)
- [27] A. Haraux, M. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, Journal of Differential Equations 144(2), 313–320 (1998)
- [28] G. Li, T.K. Pong, *Calculus of the exponent of the Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods*, Foundations of Computational Mathematics 18(5), 1199–1232 (2018)
- [29] S. Łojasiewicz, *Une Propriété Topologique des Sous-Ensembles Analytiques Réels*, Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique Paris, 87–89 (1963)
- [30] B. Mordukhovich, *Variational Analysis and Generalized Differentiation*, Springer, Berlin (2006)
- [31] R.T. Rockafellar, R. Wets, *Variational Analysis*, Springer, Berlin (1998)
- [32] L. Simon, *Asymptotics for a class of nonlinear evolution equations, with applications to geometric problems*, Annals of Mathematics (2)118, 525–571 (1983)
- [33] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, Graduate Studies in Mathematics 140, American Mathematical Society (2012)