A second order dynamical system with Hessian-driven damping and penalty term associated to variational inequalities

Radu Ioan Boţ *† Ernö Robert Csetnek ‡

December 17, 2017

Dedicated to the memory of Jon Borwein, who was so inspiring and motivating

Abstract. We consider the minimization of a convex objective function subject to the set of minima of another convex function, under the assumption that both functions are twice continuously differentiable. We approach this optimization problem from a continuous perspective by means of a second order dynamical system with Hessian-driven damping and a penalty term corresponding to the constrained function. By constructing appropriate energy functionals, we prove weak convergence of the trajectories generated by this differential equation to a minimizer of the optimization problem as well as convergence for the objective function values along the trajectories. The performed investigations rely on Lyapunov analysis in combination with the continuous version of the Opial Lemma. In case the objective function is strongly convex, we can even show strong convergence of the trajectories.

Key Words. dynamical systems, Lyapunov analysis, convex optimization, nonautonomous systems, Newton dynamics

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

1 Introduction

The Newton-like dynamical system

\[
\begin{aligned}
\ddot{x}(t) + \gamma \dot{x}(t) + \lambda \nabla^2 \Phi(x(t))(\dot{x}(t)) + \nabla \Phi(x(t)) &= 0 \\
x(0) &= u_0, \quad \dot{x}(0) = v_0,
\end{aligned}
\]

has been investigated by Alvarez, Attouch, Bolte and Redont in [5] in connection with the optimization problem

\[
\inf_{x \in \mathcal{H}} \Phi(x).
\]

Here, \( \mathcal{H} \) is a real Hilbert space endowed with inner product \( \langle \cdot , \cdot \rangle \) and associated norm \( \| \cdot \| = \sqrt{\langle \cdot , \cdot \rangle} \), \( u_0, v_0 \in \mathcal{H} \) are the initial data, \( \lambda, \gamma > 0 \), while \( \nabla \Phi \) and \( \nabla^2 \Phi \) denote the gradient and Hessian of the function \( \Phi : \mathcal{H} \to \mathbb{R} \), respectively. A particular feature of this dynamical system is the presence of the geometric damping that acts on the velocity through the Hessian of the function \( \Phi \).

* University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at. Research supported by FWF (Austrian Science Fund), project I 2419-N32.
† Invited Associated Professor, Babeş-Bolyai University, Faculty of Mathematics and Computer Science, Str. M. Kogălniceanu 1, Cluj-Napoca 400084, Romania.
‡ University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: ernoe.robert.csetnek@univie.ac.at. Research supported by FWF (Austrian Science Fund), Lise Meitner Programme, project M 1682-N25.
As underlined in [5], the dynamical system (1) can be seen as a mixture of the continuous Newton method
\[ \nabla^2 \Phi(x(t))(\dot{x}(t)) + \nabla \Phi(x(t)) = 0, \quad (3) \]
investigated by Alvarez and Pérez in [6], with the heavy ball with friction system
\[ \ddot{x}(t) + \gamma \dot{x}(t) + \nabla \Phi(x(t)) = 0, \quad (4) \]
studied for the first time in Polyak [35] and Antipin [7]. Thus the dynamical system (1) possesses most of the advantages of the systems (3) and (4). We refer the reader to [5, 6, 9, 18–20] for more insights on Newton-type dynamics and their motivations coming from mechanics and control theory.

The aim of this paper is to associate a second order Newton-type dynamical system to the optimization problem
\[ \inf_{x \in \text{argmin } \Psi} \Phi(x), \quad (5) \]
where \( \Phi, \Psi : \mathcal{H} \to \mathbb{R} \) are convex and twice differentiable functions.

Let us notice that solving (5) can be formulated as a variational inequality of the form
\[ \text{find } x \in \text{argmin } \Psi \text{ such that } \langle \nabla \Phi(x), y - x \rangle \geq 0 \quad \forall y \in \text{argmin } \Psi, \quad (6) \]
where \( \text{argmin } \Psi \) denotes the set of minimizers of \( \Psi \) over \( \mathcal{H} \).

In [12], Attouch and Czarnecki have assigned to (5) the nonautonomous first order dynamical system
\[ \dot{x}(t) + \nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t)) = 0, \quad (7) \]
where \( \beta : [0, +\infty) \to (0, +\infty) \) is a function of time assumed to tend to \( +\infty \) as \( t \to +\infty \), which penalizes the constrained function. Several convergence results for the trajectories generated by (7) to the solution set of (5) have been reported in [12] under the key assumption
\[ \forall p \in \text{ran } N_{\text{argmin } \Psi} \int_0^{+\infty} \beta(t) \left[ \Psi^* \left( \frac{p}{\beta(t)} \right) - \sigma_{\text{argmin } \Psi} \left( \frac{p}{\beta(t)} \right) \right] dt < +\infty, \quad (8) \]
where \( \Psi^* : \mathcal{H} \to \mathbb{R} \cup \{ +\infty \} \) is the Fenchel-Legendre transformation of \( \Psi \):
\[ \Psi^*(p) = \sup_{x \in \mathcal{H}} \{ \langle p, x \rangle - \Psi(x) \} \quad \forall p \in \mathcal{H}; \]
\( \sigma_{\text{argmin } \Psi} : \mathcal{H} \to \mathbb{R} \cup \{ +\infty \} \) is the support function of the set \( \text{argmin } \Psi \):
\[ \sigma_{\text{argmin } \Psi}(p) = \sup_{x \in \text{argmin } \Psi} \langle p, x \rangle \quad \forall p \in \mathcal{H}; \]
and \( N_{\text{argmin } \Psi} \) is the normal cone to the set \( \text{argmin } \Psi \), defined by
\[ N_{\text{argmin } \Psi}(x) = \{ p \in \mathcal{H} : \langle p, y - x \rangle \leq 0 \quad \forall y \in \text{argmin } \Psi \} \]
for \( x \in \text{argmin } \Psi \) and \( N_{\text{argmin } \Psi}(x) = \emptyset \) for \( x \not\in \text{argmin } \Psi \). Finally, \( \text{ran } N_{\text{argmin } \Psi} \) denotes the range of the normal cone \( N_{\text{argmin } \Psi} \), that is, \( p \in \text{ran } N_{\text{argmin } \Psi} \) if and only if there exists \( x \in \text{argmin } \Psi \) such that \( p \in N_{\text{argmin } \Psi}(x) \). Let us notice that for \( x \in \text{argmin } \Psi \) one has \( p \in N_{\text{argmin } \Psi}(x) \) if and only if \( \sigma_{\text{argmin } \Psi}(p) = \langle p, x \rangle \).

We present a situation where the above condition (8) is fulfilled. According to [12], if we take
\[ \psi(x) = \frac{1}{2} \inf_{y \in C} \| x - y \|^2, \]
for a nonempty, convex and closed set \( C \subseteq \mathcal{H} \), then the condition (8) is fulfilled if and only if

\[
\int_0^{+\infty} \frac{1}{\beta(t)} \, dt < +\infty,
\]

which is trivially satisfied for \( \beta(t) = (1 + t)^\alpha \) with \( \alpha > 1 \).

The paper of Attouch and Czarnecki [12] was the starting point of a considerable number of research articles devoted to this subject, including those addressing generalizations to variational inequalities formulated with maximal monotone operators (see [10,12,14,15,17,22,24,27–29,33,34]). We refer also to the above-listed references for more general formulations of the key assumption (8) and for further examples for which these conditions are satisfied.

The optimization problem (5) was approached in [24] through the second order nonautonomous dynamical system

\[
\ddot{x}(t) + \gamma \dot{x}(t) + \lambda \nabla^2 \Phi(x(t)) (\dot{x}(t)) + \lambda \beta(t) \nabla^2 \Psi(x(t)) (\dot{x}(t)) + \nabla \Phi(x(t)) = 0,
\]

(9)

where \( \gamma > 0 \) and \( \beta : [0, +\infty) \to (0, +\infty) \) is a function of time. Under assumption (8) and for \( \beta \) assuming to converge to \( +\infty \) as \( t \to +\infty \), we proved weak convergence of the generated trajectories to a minimizer of (5) as well as convergence for the objective function values along the trajectories. We refer to [13] for another variant of this system, where the objective function is penalized instead of the constraint function.

Our purpose is to combine Newton-like dynamics with systems of the form (9) in order to approach from a continuous perspective the solving of the optimization problem (5). To this end, we propose to investigate the asymptotic behavior of the dynamical system

\[
\ddot{x}(t) + \gamma \dot{x}(t) + \lambda \nabla^2 \Phi(x(t)) (\dot{x}(t)) + \lambda \beta(t) \nabla^2 \Psi(x(t)) (\dot{x}(t)) + \nabla \Phi(x(t)) + (\beta(t) + \lambda \dot{\beta}(t)) \nabla \Psi(x(t)) = 0.
\]

(10)

The advantages of a dynamical system which combines the continuous Newton dynamic with a heavy ball with friction system are emphasized in [5]. Condition (8) will be again crucial in the analysis performed. By using Lyapunov analysis in combination with the continuous version of the Opial Lemma, we prove weak convergence of the trajectories to a minimizer of the optimization problem (5) as well as convergence for the objective function values along the trajectories. In case the objective function is strongly convex, we can even show strong convergence of the trajectories.

2 Preliminaries

In this section we will introduce preliminary notions and results that will be useful throughout the paper.

The following statement can be interpreted as the continuous counterpart of the convergence result of quasi-Fejér monotone sequences. For its proofs we refer the reader to [1, Lemma 5.1].

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

\[
\dot{F}(t) \leq G(t).
\]

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

We will focus our investigations on the following second order dynamical system

\[
\begin{cases}
\ddot{x}(t) + \gamma \dot{x}(t) + \lambda \nabla^2 \Phi(x(t))(\dot{x}(t)) + \lambda \beta(t) \nabla^2 \Psi(x(t))(\dot{x}(t)) + \nabla \Phi(x(t)) + (\beta(t) + \lambda \dot{\beta}(t)) \nabla \Psi(x(t)) = 0 \\
x(0) = u_0, \dot{x}(0) = v_0,
\end{cases}
\]

(11)

where \( \gamma, \lambda > 0, u_0, v_0 \in \mathcal{H} \), provided that the following assumptions are satisfied:


\((H_P)\) \(\Psi : \mathcal{H} \to [0, +\infty)\) is convex, twice differentiable such that \(\nabla \Psi\) and \(\nabla^2 \Psi\) are Lipschitz continuous and \(\argmin \Psi = \Psi^{-1}(0) \neq \emptyset\);

\((H_P)\) \(\Phi : \mathcal{H} \to \mathbb{R}\) is convex, twice differentiable, bounded from below, such that \(\nabla \Phi\) and \(\nabla^2 \Phi\) are Lipschitz continuous and \(S := \{z \in \argmin \Psi : \Phi(z) \leq \Phi(x) \ \forall x \in \argmin \Psi\} \neq \emptyset\);

\((H_\beta)\) \(\beta : [0, +\infty) \to (0, +\infty)\) is a \(C^1\)-function with \(\lim_{t \to +\infty} \beta(t) = +\infty\) and it satisfies the growth condition \(0 \leq \beta \leq k \beta\), where \(0 < k < \frac{\theta}{1 + \lambda \gamma} \min \left\{ \frac{2 \gamma}{3}, \frac{2}{\lambda} \right\}\) for \(\theta \in (0, 1)\).

For \(u_0, v_0 \in \mathcal{H}\) we look for solutions \(x \in C^2([0, +\infty); \mathcal{H})\) of (11) satisfying the initial conditions \(x(0) = u_0\) and \(\dot{x}(0) = v_0\). The dynamical system can be equivalently written as

\[
\dot{X}(t) = F(t, X(t)),
\]

where

\[
X(t) = (x(t), \dot{x}(t))
\]

and

\[
F(t, u, v) = (v, -\gamma v - \lambda \nabla^2 \Phi(u)(v) - \lambda \beta(t) \nabla^2 \Psi(u)(v) - \nabla \Phi(u) - (\beta(t) + \lambda \dot{\beta}(t)) \nabla \Psi(u)).
\]

Due to the Lipschitz continuity of \(\nabla \Psi, \nabla^2 \Psi\) and \(\nabla \Phi, \nabla^2 \Phi\), assumed in \((H_P)\) and \((H_P)\), respectively, the existence and uniqueness of global solutions of (11) follows by using similar arguments as in [5, Theorem 2.1(i)] and by making use of the estimation obtained in (27).

**Remark 2**

(a) In case \(\Psi = 0\), the dynamical system (11) becomes

\[
\begin{align*}
\dot{x}(t) + \gamma \dot{x}(t) + \lambda \nabla^2 \Phi(x(t)) \dot{x}(t) + \nabla \Phi(x(t)) &= 0, \\
x(0) = u_0, \dot{x}(0) = v_0,
\end{align*}
\]

the convergence of which has been investigated in [5] in connection with the minimization of the function \(\Phi\) over \(\mathcal{H}\).

(b) The time discretization of second order dynamical systems leads to iterative algorithms involving inertial terms, which basically means that every new iterate is constructed in terms of the previous two iterates (see for example [3, 4]). The explicit time discretization of (11) with step size \(h_k > 0\), penalty parameter \(\beta_k > 0\) and initial points \(x^0 = u_0\) and \(x^1 = u_1\) yields for every \(k \geq 0\)

\[
\frac{x^{k+2} - 2x^{k+1} + x^k}{h_k^2} + \gamma \frac{x^{k+1} - x^k}{h_k} + \lambda \nabla^2 \Phi(x^k) \left( \frac{x^{k+1} - x^k}{h_k} \right) + \lambda \beta_k \nabla^2 \Psi(x^k) \left( \frac{x^{k+1} - x^k}{h_k} \right)
\]

\[
+ \nabla \Phi(x^k) + (\beta_k + \lambda (\beta_{k+1} - \beta_k)) \nabla \Psi(x^k) = 0.
\]

For \(h_k = 1\) one obtains the iterative scheme

\[
x^{k+2} = x^{k+1} + (1 - \gamma)(x^{k+1} - x^k) - \lambda \nabla^2 (\Phi + \beta_k \Psi)(x^k)(x^{k+1} - x^k)
\]

\[
- \nabla (\Phi + \beta_k \Psi)(x^k) - \lambda (\beta_{k+1} - \beta_k) \nabla \Psi(x^k) \ \forall k \geq 0,
\]

which one may investigate in connection to the solving of the optimization problem (5). The asymptotic behaviour of the trajectory of (11), that we will address in the next section, suggests that one can expect the convergence of the sequence of generated iterates \((x^k)_{k \in \mathbb{N}}\) to a global minimum of (5).
3 Convergence of the trajectories and of the objective function values

This section is devoted to the asymptotic analysis of the trajectory generated by the dynamical system (11). We show weak convergence of the trajectory \( x(\cdot) \) to an optimal solution of (5) as well as convergence for the objective function values along the trajectory as \( t \to +\infty \), under the following assumption:

\[
(H) \quad \forall p \in \text{ran } N_{\text{argmin } \Psi} \int_0^{+\infty} \beta(t) \left[ \Psi^* \left( \frac{p}{\beta(t)} \right) - \sigma_{\text{argmin } \Psi} \left( \frac{p}{\beta(t)} \right) \right] dt < +\infty.
\]

Remark 3. (a) Let \( \delta_{\text{argmin } \Psi} : \mathcal{H} \to \mathbb{R} \cup \{+\infty\} \) be the indicator function of argmin \( \Psi \), which is the function that takes the value 0 on the set argmin \( \Psi \) and +\( \infty \), otherwise. Due to \( \Psi \leq \delta_{\text{argmin } \Psi} \) (see (\( H_{\Psi} \)), we have

\[
\Psi^* \geq \delta^*_{\text{argmin } \Psi} = \sigma_{\text{argmin } \Psi}.
\]

(b) Considering the case when \( \Psi = 0 \) (see Remark 2(a)), we have \( N_{\text{argmin } \Psi}(x) = \{0\} \) for every \( x \in \text{argmin } \Psi = \mathcal{H} \), \( \Psi^* = \sigma_{\text{argmin } \Psi} = \delta_{\{0\}} \) and (\( H \)) trivially holds.

For \( \delta > 0 \), we consider the following energy functional that will play an important role in the analysis below:

\[
E_{\delta}(t) = \delta \left( \Phi(x(t)) + \beta(t)\Psi(x(t)) \right) + \frac{1}{2} \| \dot{x}(t) + \lambda \nabla \Phi(x(t)) + \lambda \beta(t) \nabla \Psi(x(t)) \|^2 \quad \forall t > 0. \tag{13}
\]

For its derivative we have for every \( t \in [0, +\infty) \)

\[
\dot{E}_{\delta}(t) = \delta \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle + \delta \beta(t) \langle \nabla \Psi(x(t)), \dot{x}(t) \rangle + \delta \dot{\beta}(t) \Psi(x(t)) +
\]

\[
\langle \ddot{x}(t) + \lambda \nabla^2 \Phi(x(t)) \dot{x}(t), \dot{x}(t) \rangle + \lambda \beta(t) \nabla^2 \Psi(x(t)) \dot{x}(t) + \lambda \beta(t) \nabla \Psi(x(t)) \dot{x}(t) + \lambda \beta(t) \nabla \Psi(x(t)) \dot{x}(t) +
\]

\[
= \delta \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle + \delta \beta(t) \langle \nabla \Psi(x(t)), \dot{x}(t) \rangle + \delta \dot{\beta}(t) \Psi(x(t)) +
\]

\[
\langle -\gamma \dot{x}(t) - \nabla \Phi(x(t)) - \beta(t) \nabla \Psi(x(t)), \dot{x}(t) + \lambda \nabla \Phi(x(t)) + \lambda \beta(t) \nabla \Psi(x(t)) \rangle
\]

\[
= \delta \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle + \delta \beta(t) \langle \nabla \Psi(x(t)), \dot{x}(t) \rangle + \delta \dot{\beta}(t) \Psi(x(t)) - \gamma \| \dot{x}(t) \|^2 -
\]

\[
(1 + \gamma \beta) \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle - (1 + \gamma \beta) \langle \nabla \Psi(x(t)), \dot{x}(t) \rangle - \lambda \| \nabla \Phi(x(t)) \|^2 -
\]

\[
\lambda \beta^2(t) \| \nabla \Psi(x(t)) \|^2 - 2 \lambda \beta(t) \langle \nabla \Phi(x(t)), \nabla \Psi(x(t)) \rangle
\]

\[
= (\delta - \gamma \lambda - 1) \langle \nabla \Phi(x(t)), x(t) \rangle + \beta(t)(\delta - \gamma \lambda - 1) \langle \nabla \Psi(x(t)), x(t) \rangle + \delta \dot{\beta}(t) \Psi(x(t)) -
\]

\[
\gamma \| \dot{x}(t) \|^2 - \lambda \| \nabla \Phi(x(t)) \|^2 - \lambda \beta^2(t) \| \nabla \Psi(x(t)) \|^2 - 2 \lambda \beta(t) \langle \nabla \Phi(x(t)), \nabla \Psi(x(t)) \rangle.
\]

Finally we obtain for every \( t \in [0, +\infty) \)

\[
\dot{E}_{\delta}(t) = (\delta - \gamma \lambda - 1) \langle \nabla \Phi(x(t)), \dot{x}(t) \rangle + \beta(t)(\delta - \gamma \lambda - 1) \langle \nabla \Psi(x(t)), \dot{x}(t) \rangle + \delta \dot{\beta}(t) \Psi(x(t))
\]

\[
- \gamma \| \dot{x}(t) \|^2 - \lambda \| \nabla \Phi(x(t)) \|^2 + \beta(t) \| \nabla \Psi(x(t)) \|^2. \tag{14}
\]

Further, for \( z \in \mathcal{S} \) and

\[
\varepsilon := \theta \min \left\{ \frac{2\gamma}{3}, \frac{2}{\lambda} \right\} \tag{15}
\]

we consider the functional

\[
E(t) = \frac{1}{\varepsilon} E_{1 + \gamma \lambda}(t) + \frac{\gamma}{2} \| x(t) - z \|^2 + \langle \dot{x}(t) + \lambda \nabla \Phi(x(t)) + \lambda \beta(t) \nabla \Psi(x(t)), x(t) - z \rangle \quad \forall t > 0. \tag{16}
\]
By using (14) we easily derive for every $t \in [0, +\infty)$

$$
\dot{E}(t) = \frac{1 + \frac{\gamma}{\varepsilon} \beta(t) \Psi(x(t)) + \frac{\gamma}{\varepsilon} \|\dot{x}(t)\|^2 - \frac{\lambda}{\varepsilon} \|\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))\|^2}{\varepsilon}
\leq \frac{1 + \frac{\gamma}{\varepsilon} \beta(t) \Psi(x(t)) - \langle \nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t)), x(t) - z \rangle}{\varepsilon} - \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) ||\dot{x}(t)||^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) ||\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))||^2.
\leq \frac{1 + \frac{\gamma}{\varepsilon} \beta(t) \Psi(x(t)) + \beta(t) \nabla \Psi(x(t)), x(t) - z}{\varepsilon} - \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) ||\dot{x}(t)||^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) ||\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))||^2.
$$

(iii) \quad \lim_{t \to \infty} \int_0^t (\Phi(s) - \Phi(z) + \left(1 - \frac{\varepsilon (1 + \gamma \lambda)}{\varepsilon} \right) \beta(s) \Psi(x(s))) \, ds \in \mathbb{R};

(iv) \quad \dot{x}, \nabla \Phi(x) + \beta \nabla \Psi(x) \in L^2([0,\infty); \mathcal{H}).

**Proof.** Take an arbitrary $z \in S$. Relying on the convexity of the functions $\Phi$ and $\Psi$, the fact that $z \in \arg \min \Psi$ (hence $\Psi(z) = 0$) and the non-negativity of $\beta$ and $\Psi$ we obtain for every $t \in [0, +\infty)$

$$
-(\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t)), x(t) - z) \leq \Phi(z) - \Phi(x(t)) - \beta(t) \Psi(x(t)).
$$

From here and (17) we derive

$$
\dot{E}(t) \leq \frac{1 + \frac{\gamma}{\varepsilon} \beta(t) \Psi(x(t)) + \Phi(z) - \Phi(x(t)) - \beta(t) \Psi(x(t))}{\varepsilon} - \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) ||\dot{x}(t)||^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) ||\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))||^2.
$$

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

By using the growth condition on $\beta$ we get

$$
\dot{E}(t) \leq \left(\frac{k + \frac{\gamma}{\varepsilon} \beta(t) \Psi(x(t)) + \Phi(z) - \Phi(x(t))}{\varepsilon} - 1\right) \beta(t) \Psi(x(t)) + \Phi(z) - \Phi(x(t)) - \beta(t) \Psi(x(t))
\leq \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) ||\dot{x}(t)||^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) ||\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))||^2 \leq - \beta(t) \Psi(x(t)) + \Phi(z) - \Phi(x(t))
\leq \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) ||\dot{x}(t)||^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) ||\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))||^2,
$$

\[ \text{Lemma 4} \quad \text{Assume that (H}_\Phi, (H_\Psi), (H_\beta) \text{ and (H) hold and let } x : [0, +\infty) \to \mathcal{H} \text{ be the trajectory generated by the dynamical system (11). Then for every } z \in S \text{ the following statements are true:}
\]

(i) $\int_0^{t \to \infty} \beta(t) \Psi(x(t)) \, dt < +\infty$;

(ii) $\exists \lim_{t \to \infty} \int_0^t (\nabla \Phi(z), x(s) - z) \, ds \in \mathbb{R}$;

(iii) $\exists \lim_{t \to \infty} \int_0^t (\Phi(x(s)) - \Phi(z) + \left(1 - \frac{k(1 + \gamma \lambda)}{\varepsilon}\right) \beta(s) \Psi(x(s))) \, ds \in \mathbb{R}$;

(iv) $\dot{x}, \nabla \Phi(x) + \beta \nabla \Psi(x) \in L^2([0, +\infty); \mathcal{H})$. 


where
\[
\tilde{\beta}(t) := \left(1 - k \frac{1 + \gamma \lambda}{\varepsilon}\right) \beta(t) > 0,
\]  
(21)
due to \((H_{\beta})\) and (15).

Furthermore,
\[
\Phi(z) - \Phi(x(t)) \leq \langle -\nabla \Phi(z), x(t) - z \rangle,
\]  
(22)

thus
\[
\dot{E}(t) \leq -\tilde{\beta}(t) \Psi(x(t)) - \langle \nabla \Phi(z), x(t) - z \rangle
- \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) \|\dot{x}(t)\|^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) \|\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))\|^2
\]  
(23)

for every \(t \in [0, +\infty)\). Since \(z\) is an optimal solution of (5), the first order optimality condition delivers
\[
0 \in \partial(\Phi + \delta_{\text{argmin } \Psi})(z) = \nabla \Phi(z) + N_{\text{argmin } \Psi}(z),
\]  
(24)
hence
\[
-\nabla \Phi(z) \in N_{\text{argmin } \Psi}(z) \subseteq \text{ran } N_{\text{argmin } \Psi}.
\]  
(25)

From here and by using the Young-Fenchel inequality we obtain for every \(t \in [0, +\infty)\)
\[
-\tilde{\beta}(t) \Psi(x(t)) - \langle \nabla \Phi(z), x(t) - z \rangle
= \tilde{\beta}(t) \left(\nabla \Phi(z) - \left(\frac{\nabla \Phi(z)}{\beta(t)} \right), x(t) - z \right)
= \tilde{\beta}(t) \left(-\Psi(x(t)) + \left(\frac{\nabla \Phi(z)}{\beta(t)} \right), x(t) \right) - \sigma_{\text{argmin } \Psi} \left(\frac{\nabla \Phi(z)}{\beta(t)} \right)
\leq \tilde{\beta}(t) \left(\Psi^{*} \left(\frac{-\nabla \Phi(z)}{\beta(t)} \right) - \sigma_{\text{argmin } \Psi} \left(\frac{-\nabla \Phi(z)}{\beta(t)} \right) \right).
\]  
(26)

Thus, from (23) and (26) we obtain for every \(t \in [0, +\infty)\)
\[
\dot{E}(t) \leq \tilde{\beta}(t) \left(\Psi^{*} \left(\frac{-\nabla \Phi(z)}{\beta(t)} \right) - \sigma_{\text{argmin } \Psi} \left(\frac{-\nabla \Phi(z)}{\beta(t)} \right) \right)
- \left(\frac{\gamma}{\varepsilon} - \frac{3}{2}\right) \|\dot{x}(t)\|^2 - \frac{\lambda}{\varepsilon} \left(1 - \frac{\varepsilon \lambda}{2}\right) \|\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))\|^2
\]  
(27)
\[
\leq \tilde{\beta}(t) \left(\Psi^{*} \left(\frac{-\nabla \Phi(z)}{\beta(t)} \right) - \sigma_{\text{argmin } \Psi} \left(\frac{-\nabla \Phi(z)}{\beta(t)} \right) \right),
\]  
(28)

where the last inequality follows from (15) and the fact that \(\theta \in (0, 1)\).

By integrating the last inequality from 0 to \(T\) \((T > 0)\) and by taking into account \((H), (16), (13)\) and the fact that \(\Phi\) and \(\Psi\) are bounded from below, it yields that there exists \(M > 0\) such that
\[
\frac{\gamma}{2} \|x(T) - z\|^2 + \langle \dot{x}(T) + \lambda \nabla \Phi(x(T)) + \lambda \beta(T) \nabla \Psi(x(T)), x(T) - z \rangle
+ \frac{1}{2\varepsilon} \|\dot{x}(T) + \lambda \nabla \Phi(x(T)) + \lambda \beta(T) \nabla \Psi(x(T))\|^2 \leq M \forall T \geq 0.
\]  
(29)

Combining this with (18) and the fact that \(\Phi\) and \(\Psi\) are bounded from below, one can easily see that there exists \(M' > 0\) such that
\[
\frac{d}{dT} \left(\frac{1}{2} \|x(T) - z\|^2\right) + \gamma \left(\frac{1}{2} \|x(T) - z\|^2\right) \leq M' \forall T \geq 0.
\]  
7
A direct application of the Gronwall Lemma implies that
\[ x \text{ is bounded.} \] (30)
Further, this yields via (29) that
\[ \dot{x} + \lambda \nabla \Phi(x) + \lambda \beta \nabla \Psi(x) \text{ is bounded.} \] (31)
From (30), (31), (16) and (13) we conclude that
\[ E \text{ is bounded from below.} \] (32)
Moreover, from (20), (26) and (22) we obtain for every \( t \in [0, +\infty) \)
\[
\dot{E}(t) + \tilde{\beta}(t) \left( -\Psi^* \left( \frac{-\nabla \Phi(z)}{\beta(t)} \right) + \sigma_{\text{argmin}} \Psi \left( \frac{-\nabla \Phi(z)}{\beta(t)} \right) \right) \\
\leq \dot{E}(t) + \tilde{\beta}(t) \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle \\
\leq \dot{E}(t) + \Phi(x(t)) - \Phi(z) + \tilde{\beta}(t) \Psi(x(t)) \\
\leq 0.
\]
(33)
(i) Consider the function \( F : [0, +\infty) \to \mathbb{R} \) defined by
\[ F(t) = \int_0^t \left( -\tilde{\beta}(s) \Psi(x(s)) - \langle \nabla \Phi(z), x(s) - z \rangle \right) ds \forall t \geq 0. \]
Making again use of (see (33))
\[-\tilde{\beta}(s) \Psi(x(s)) - \langle \nabla \Phi(z), x(s) - z \rangle \geq \dot{E}(s) \forall s \in [0, +\infty) \]
and (32), we easily derive that \( F \) is bounded from below. Moreover, from (26) it follows that for every \( t \in [0, +\infty) \)
\[ \dot{F}(t) \leq \tilde{\beta}(t) \left( \Psi^* \left( \frac{-\nabla \Phi(z)}{\beta(t)} \right) - \sigma_{\text{argmin}} \Psi \left( \frac{-\nabla \Phi(z)}{\beta(t)} \right) \right). \]
Notice that according to (H), the function on the right-hand side of this inequality is \( L^1 \)-integrable on \([0, +\infty)\), hence a direct application of Lemma 1 yields that \( \lim_{t \to +\infty} F(t) \) exists and is a real number. Thus
\[ \exists \lim_{t \to +\infty} \int_0^t \left( \tilde{\beta}(s) \Psi(s) + \langle \nabla \Phi(z), x(s) - z \rangle \right) ds \in \mathbb{R}. \] (34)
Since \( \Psi \geq 0 \), we obtain for every \( t \in [0, +\infty) \)
\[ -\frac{\tilde{\beta}(t)}{2} \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle \geq \frac{\tilde{\beta}(t)}{2} \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle \]
and from here, similarly to (26),
\[ -\frac{\tilde{\beta}(t)}{2} \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle \leq \frac{\tilde{\beta}(t)}{2} \left( \Psi^* \left( \frac{-2\nabla \Phi(z)}{\beta(t)} \right) - \sigma_{\text{argmin}} \Psi \left( \frac{-2\nabla \Phi(z)}{\beta(t)} \right) \right). \]
Thus, for every \( t \in [0, +\infty) \) it holds
\[
\dot{E}(t) + \frac{\tilde{\beta}(t)}{2} \left( -\Psi^* \left( \frac{-2\nabla \Phi(z)}{\beta(t)} \right) + \sigma_{\text{argmin}} \Psi \left( \frac{-2\nabla \Phi(z)}{\beta(t)} \right) \right) \\
\leq \dot{E}(t) + \frac{\tilde{\beta}(t)}{2} \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle \\
\leq \dot{E}(t) + \tilde{\beta}(t) \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle \\
\leq 0.
\]
Following the same technique as in the proof of (34), it yields that
\[
\exists \lim_{t \to +\infty} \int_0^t \left( \frac{\beta(s)}{2} \Psi(s) + (\nabla \Phi(z), x(s) - z) \right) ds \in \mathbb{R}. \tag{35}
\]

Finally, from (34), (35) and (21) we obtain (i).

(ii) Follows from (i), (34) and (21).

(iii) Follows from (33) and (26), by following the same arguments as for proving statement (34).

(iv) Follows by integrating (27), by taking into account (H), (32), (15) and the fact that \( \theta \in (0, 1) \).

\[\blacksquare\]

**Remark 5** The assumption \( \lim_{t \to +\infty} \beta(t) = +\infty \) has not been used in the above proof. However, it will play an important role in the arguments used below.

For the asymptotic analysis of the trajectories generated by the dynamical system (11), the continuous version of the Opial Lemma that we state as follows will be crucial.

**Lemma 6** Let \( S_\infty \) be a nonempty subset of the real Hilbert space \( \mathcal{H} \) and \( x : [0, +\infty) \to \mathcal{H} \) a given function. Assume that

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

(ii) every weak limit point of \( x \) belongs to \( S_\infty \).

Then there exists \( x_\infty \in S_\infty \) such that \( x(t) \) converges weakly to \( x_\infty \) as \( t \to +\infty \).

We state now the main theorem of the paper.

**Theorem 7** Assume that \( (H_\Psi), (H_\Phi), (H_\beta) \) and \( (H) \) hold and let \( x : [0, +\infty) \to \mathcal{H} \) be the trajectory generated by the dynamical system (11). Then the following statements are true:

(i) \( \Phi(x(t)) \) converges to the optimal objective value of (5) as \( t \to +\infty \);

(ii) \( \lim_{t \to +\infty} \beta(t) \Psi(x(t)) = \lim_{t \to +\infty} \Psi(x(t)) = 0 \);

(iii) \( \int_0^{+\infty} \beta(t) \Psi(x(t)) dt < +\infty \);

(iv) \( \dot{x}, \nabla \Phi(x) + \beta \nabla \Psi(x) \in L^2([0, +\infty); \mathcal{H}) \);

(v) \( \lim_{t \to +\infty} (\dot{x}(t) + \nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))) = 0 \);

(vi) there exists \( x_\infty \in S \) such that \( x(t) \) converges weakly to \( x_\infty \) as \( t \to +\infty \).

**Proof.** Fix an arbitrary \( z \in S \) and consider the energy functional defined in (13) for
\[
\delta_1 = (1 + \sqrt{\lambda \gamma})^2 = 1 + \lambda \gamma + 2 \sqrt{\lambda \gamma}.
\]
A simple computation (see (14)) shows that
\[
E_{\delta_1}(t) = (1 + \sqrt{\lambda \gamma})^2 \beta(t) \Psi(x(t)) - \left\| \sqrt{\gamma} \dot{x}(t) - \sqrt{\lambda}(\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))) \right\|^2.
\]
Taking into account the growth condition on \( \beta \), Lemma 4(i) and the fact that \( E_{\delta_1} \) is bounded from below, we obtain from Lemma 1 that
\[
\exists \lim_{t \to +\infty} E_{\delta_1}(t) \in \mathbb{R}. \tag{36}
\]
Similarly, consider
\[\delta_2 = (1 - \sqrt{\lambda \gamma})^2 = 1 + \lambda \gamma - 2\sqrt{\lambda \gamma}.\]
We have
\[\dot{E}_{\delta_2}(t) = (1 - \sqrt{\lambda \gamma})^2 \beta(t)\Psi(x(t)) - \left\| \sqrt{\lambda} \dot{x}(t) + \sqrt{\lambda}(\nabla \Phi(x(t)) + \beta(t) \nabla \Psi(x(t))) \right\|^2\]
and
\[\exists \lim_{t \to +\infty} E_{\delta_2}(t) \in \mathbb{R}. \tag{37}\]
From (36) and (37) we get
\[\exists \lim_{t \to +\infty} (E_{\delta_1}(t) - E_{\delta_2}(t)) \in \mathbb{R}.\]
This implies by the definition of the energy functional that
\[\exists \lim_{t \to +\infty} (\Phi(x(t)) + \beta(t)\Psi(x(t))) \in \mathbb{R}. \tag{38}\]
From here we deduce
\[\lim_{t \to +\infty} \Psi(x(t)) = 0. \tag{39}\]
Indeed, since \(\Phi\) is bounded from below and, by taking into account the inequality
\[\Phi(x(t)) \leq \Phi(z) + \langle \nabla \Phi(x(t)), x(t) - z \rangle,\]
the fact that \(x\) is bounded and \(\lim_{t \to +\infty} \beta(t) = +\infty\), we have
\[\lim_{t \to +\infty} \frac{\Phi(x(t))}{\beta(t)} = 0.\]
The statement (39) follows now from the last relation and (38).
In the following we appeal to [12, Lemma 3.4] for the inequality
\[\liminf_{t \to +\infty} \Phi(x(t)) \geq \Phi(z). \tag{40}\]
Indeed, this was obtained in [12, Lemma 3.4] for a first order dynamical system. However, the statement remains true for (11) too, since a careful look at the proof of Lemma 3.4 in [12] reveals that the main arguments used to obtain this conclusion are the statements in Lemma 4(ii), the fact that the trajectory \(x\) is bounded, the weak lower semicontinuity of \(\Psi\), equation (39), the inequality (22) and relation (25).
From (40) and (38) we have \(\lim_{t \to +\infty} (\Phi(x(t)) + \beta(t)\Psi(x(t))) \geq \Phi(z)\). We claim that
\[\lim_{t \to +\infty} (\Phi(x(t)) + \beta(t)\Psi(x(t))) = \Phi(z). \tag{41}\]
Let us assume that \(\lim_{t \to +\infty} (\Phi(x(t)) + \beta(t)\Psi(x(t))) > \Phi(z)\). Then there exist \(\eta > 0\) and \(t_0 \geq 0\) such that for every \(t \geq t_0\) we have
\[\Phi(x(t)) + \beta(t)\Psi(x(t)) > \Phi(z) + \eta. \tag{42}\]
Hence, for every \(t \geq t_0\)
\[\eta < \Phi(x(t)) - \Phi(z) + \left(1 - \frac{k(1 + \gamma \lambda)}{\varepsilon}\right) \beta(t)\Psi(x(t)) + \frac{k(1 + \gamma \lambda)}{\varepsilon} \beta(t)\Psi(x(t)). \tag{43}\]
Integrating the last inequality and taking into account Lemma 4(i) and (iii) we obtain a contradiction. In conclusion, (41) holds.
Since $\Phi(x(t)) \leq \Phi(x(t)) + \beta(t)\Psi(x(t))$ for every $t \in [0, +\infty)$, according to (41) it follows that $\lim_{t \to +\infty} \Phi(x(t)) \leq \Phi(z)$, which combined with (40) imply (i).

(ii) Follows from (39), (41) and the fact that $\lim_{t \to +\infty} \Phi(x(t)) = \Phi(z)$.

(iii)-(iv) The statements have been proved in Lemma 4.

(v) From (14), Lemma 4(i) and Lemma 1, we derive that

$$\exists \lim_{t \to +\infty} E_{1+\lambda t}(t) \in \mathbb{R}. \quad (44)$$

Combining this with (13) and (38), we obtain that

$$\exists \lim_{t \to +\infty} \|\dot{x}(t) + \nabla \Phi(x(t)) + \beta(t)\nabla \Psi(x(t))\| \in \mathbb{R}. \quad (45)$$

Further, by using (16), (45), (44), (30) and (v) we conclude that

$$\exists \lim_{t \to +\infty} \|x(t) - z\| \in \mathbb{R}. \quad (46)$$

Since $z \in S$ was arbitrary chosen, the first statement of the Opial Lemma is true.

We prove now that the second condition in Lemma 6 is fulfilled, too. Let $(t_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers such that $\lim_{n \to +\infty} t_n = +\infty$ and $x(t_n)$ converges weakly to $x_\infty$ as $n \to +\infty$. By using the weak lower semicontinuity of $\Psi$ and (ii) we obtain

$$0 \leq \Psi(x_\infty) \leq \liminf_{n \to +\infty} \Psi(x(t_n)) = 0,$$

hence $x_\infty \in \argmin \Psi$. Moreover, the weak lower semicontinuity of $\Phi$ and (i) yield

$$\Phi(x_\infty) \leq \liminf_{n \to +\infty} \Phi(x(t_n)) = \Phi(z),$$

thus $x_\infty \in S$. Therefore, both conditions of Lemma 6 are fulfilled and the conclusion follows.

Finally, we consider the situation when the objective function of (5) is strongly convex. In this case, the trajectory generated by (11) converges strongly to the unique optimal solution of (5).

**Theorem 8** Assume that $(H_\Psi)$, $(H_\Phi)$, $(H_\beta)$ and $(H)$ hold and let $x : [0, +\infty) \to \mathcal{H}$ be the trajectory generated by the dynamical system (11). If $\Phi$ is strongly convex, then $x(t)$ converges strongly to the unique optimal solution of (5) as $t \to +\infty$.

**Proof.** Let $\mu > 0$ be such that $\Phi$ is $\mu$-strongly convex. In this case the optimization problem (5) has a unique optimal solution, which we denote by $z$.

We replace (22) with the stronger inequality

$$\frac{\mu}{2} \|x(t) - z\|^2 + \Phi(z) - \Phi(x(t)) \leq \langle -\nabla \Phi(z), x(t) - z \rangle \quad \forall t \in [0, +\infty) \quad (47)$$

and obtain (see the proof of Lemma 4 and (33)) for every $t \in [0, +\infty)$

$$\dot{E}(t) + \frac{\mu}{2} \|x(t) - z\|^2 + \tilde{\beta}(t) \left( -\Psi^* \left( -\nabla \Phi(z) \over \tilde{\beta}(t) \right) + \sigma_{\argmin \Psi} \left( -\nabla \Phi(z) \over \tilde{\beta}(t) \right) \right)$$

$$\leq \dot{E}(t) + \frac{\mu}{2} \|x(t) - z\|^2 + \tilde{\beta}(t) \Psi(x(t)) + \langle \nabla \Phi(z), x(t) - z \rangle$$

$$\leq \dot{E}(t) + \Phi(x(t)) - \Phi(z) + \tilde{\beta}(t) \Psi(x(t))$$

$$\leq 0. \quad (48)$$

11
Taking into account \((H)\) and that \(E\) is bounded from below (see (32)), by integration of the above inequality we obtain that there exists a constant \(C > 0\) such that

\[
\frac{\mu}{2} \int_0^T \|x(t) - z\|^2 dt \leq C \forall T \geq 0.
\]

According to (46), \(\lim_{t \to +\infty} \|x(t) - z\|\) exists, thus \(\|x(t) - z\|\) converges to 0 as \(t \to +\infty\) and the proof is complete.

\[\blacksquare\]

References


