AN ASYMPTOTIC VISCOSITY SELECTION RESULT FOR THE REGULARIZED NEWTON DYNAMIC

BOUSHRA ABBAS

I3M UMR CNRS 5149, UNIVERSITÉ MONTPELLIER II,
34095 MONTPELLIER, FRANCE

Abstract. Let $\Phi : \mathcal{H} \longrightarrow \mathbb{R}\cup\{+\infty\}$ be a closed convex proper function on a real Hilbert space $\mathcal{H}$, and $\partial\Phi : \mathcal{H} \rightrightarrows \mathcal{H}$ its subdifferential. For any control function $\epsilon : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ which tends to zero as t goes to $+\infty$, and λ a positive parameter, we study the asymptotic behavior of the trajectories of the regularized Newton dynamical system

$$\begin{aligned} &\upsilon(t) \in \partial\Phi(x(t)) \\ &\lambda\dot{x}(t) + \dot{\upsilon}(t) + \upsilon(t) + \varepsilon(t)\,x(t) = 0. \end{aligned}$$

Assuming that $\varepsilon(t)$ tends to zero moderately as t goes to $+\infty$, we show that the term $\varepsilon(\cdot)\,x(\cdot)$ asymptotically acts as a Tikhonov regularization, which forces the trajectories to converge to a particular equilibrium. Precisely, when $C = \operatorname{argmin}\Phi \neq \emptyset$, and $\varepsilon(\cdot)$ is a "slow" control, i.e., $\int_0^{+\infty} \varepsilon(t)\,dt = +\infty$, then each trajectory of the system converges weakly, as t goes to $+\infty$, to the element of minimal norm of the closed convex set C. When Φ is a convex differentiable function whose gradient is Lipschitz continuous, we show that the strong convergence property is satisfied. Then we examine the effect of other types of regularizing methods.

1. Introduction

Throughout this paper, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle\cdot,\cdot\rangle$, and $\|x\|^2 = \langle x, x\rangle$ for any $x \in \mathcal{H}$. Given $\Phi : \mathcal{H} \longrightarrow \mathbb{R}\cup\{+\infty\}$ a closed convex proper function, we will analyze some asymptotic viscosity selection properties for the regularized Newton dynamic governed by Φ.

Let us first recall some basic facts about this dynamical system. Given λ a positive constant, the Regularized Newton dynamic ((RN) for short) attached to solving the minimization problem

$$(\mathcal{P}) \quad \min_{x\in\mathcal{H}} \Phi(x)$$

is written as follows

$$\upsilon(t) \in \partial\Phi(x(t)) \tag{1}$$

$$\lambda\dot{x}(t) + \dot{\upsilon}(t) + \upsilon(t) = 0, \tag{2}$$

where the subdifferential of Φ at $x \in \operatorname{dom}\Phi$ is classically defined by

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

When Φ is a smooth function, (RN) is equivalent to

$$\lambda\dot{x}(t) + \nabla^2\Phi(x(t)) + \nabla\Phi(x(t)) = 0$$

where λ acts as a Levenberg-Marquard regularization parameter of the continuous Newton equation, whence the terminology. This dynamical system has been first introduced in [9], [8]. Its extension to the case of two potentials gives rise to a new class of forward-backward algorithms, see [1], [2], [7]. In [9], for a general closed convex and proper function Φ, it is shown that the Cauchy problem for the (x, υ) system (1)-(2) admits a unique strong global solution. In addition, under the sole assumption that $C = \operatorname{argmin}\Phi \neq \emptyset$, for any orbit of (1)-(2), $x(t)$ converges weakly to an element of C, as t goes to $+\infty$.

In many applications, a particular stationary solution is more interesting than others due to physical, economic or design considerations. When we have the global convergence of trajectories, one could let the trajectory reach a particular target equilibrium by appropriately adjusting the initial conditions. Nevertheless, in many practical situations it is not possible to have an accurate control of the initial state. An

Date: April 29, 2015.

alternative approach consists in introducing a term into the system which forces convergence to the desired stationary solution, independently of the initial state. Such a term should vanish at infinity in order to recover, at least asymptotically, an equilibrium point of (1)-(2).

The above discussion motivates the introduction of the following abstract evolution system:

$$v(t) \in \partial\Phi(x(t)) \tag{3a}$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v(t) + \varepsilon(t)\, x(t) = 0, \tag{3b}$$

where $\epsilon : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is an open-loop control function, that tends to zero as t goes to $+\infty$.

Let us briefly describe our approach. Following a similar device as in [9], setting $\mu = \frac{1}{\lambda}$, and introducing the new unknown function $y(\cdot) = x(\cdot) + \mu v(\cdot)$, we can equivalently rewrite (3a)-(3b) as

$$\begin{aligned} &x(t) = \mathrm{prox}_{\mu\Phi}(y(t)), \\ &\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\,\mathrm{prox}_{\mu\Phi}(y(t)) = 0 \end{aligned}$$

where $\mathrm{prox}_{\mu\Phi}$ is the proximal mapping associated to $\mu\Phi$. Recall that $\mathrm{prox}_{\mu\Phi} = (I + \mu\partial\Phi)^{-1}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\Phi$, and $\nabla\Phi_\mu = \frac{1}{\mu}(I - (I + \mu\partial\Phi)^{-1})$ is its Yosida approximation of index $\mu > 0$. As a key point of our analysis, we notice that $\mathrm{prox}_{\mu\Phi}$ is a gradient vector field, namely $\mathrm{prox}_{\mu\Phi} = \nabla\psi$, with

$$\psi(y) = \mu\,(\Phi^*)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right), \tag{4}$$

where Φ^* is the Fenchel conjugate of Φ. Doing so, we can reformulate our dynamic in the form

$$x(t) = (I + \mu\partial\Phi)^{-1}(y(t)), \tag{5a}$$

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \epsilon(t)\,\nabla\mu\psi(y(t)) = 0. \tag{5b}$$

Equation (5b) is a particular case of the multiscale dynamic

$$\dot{y}(t) + \partial\Theta(y(t)) + \epsilon(t)\,\partial\Psi(y(t)) \ni 0 \tag{6}$$

where Θ and Ψ are two convex potential functions. Following [4], Ψ will be referred to as the "viscosity function". A detailed study of the asymptotic behavior of the orbits of (6) can be found in [6], [14], [15], [16], [18]. Following [6] and [14], we focus our attention on the case where the parametrization $t \mapsto \epsilon(t)$ satisfies the following "slow" decay property

$$\int_0^{+\infty} \varepsilon(t)\, dt = +\infty.$$

This condition expresses that $\varepsilon(\cdot)$ does not tend to zero too rapidly, which allows the term $\varepsilon(\cdot)\, x(\cdot)$ to be effective asymptotically. In that case, we will show an asymptotic selection property. Precisely, in Theorem 3.2, under some additional moderate growth property on $\epsilon(\cdot)$, we will show that, for any trajectory (x, v) of (3a)-(3b), $x(\cdot)$ converges weakly to the minimizer of Φ which also minimizes ψ over all minima of Φ. Then we show that this element is nothing but the element of minimal norm of the solution set argminΦ, i.e.,

$$x(t) \rightharpoonup \mathrm{proj}_{\mathrm{argmin}\Phi} 0 \quad \text{as } t \to +\infty.$$

Thus we recover the classical Tikhonov viscosity selection principle, which consists in selecting the solution of minimal norm.

This result can be viewed as an asymptotic selection property: by using such a slow control ε, one can force all the trajectories to converge to the same equilibrium, which here is the equilibrium of minimal norm. This makes a sharp contrast with the non controlled situation, or fast control, where the limits of the trajectories depend on the initial data, and are in general difficult to identify.

The paper is organized as follows: we first show the existence and uniqueness of a strong global solution to the Cauchy problem (3a)-(3b). Then, we study the asymptotic convergence as t goes to $+\infty$ of the trajectories of (3a)-(3b). In our main result, Theorem 3.2, under the key assumption that $\epsilon(\cdot)$ is a "slow control", i.e., $\int_0^{+\infty} \varepsilon(t)\, dt = +\infty$, and has moderate growth, we show the weak convergence of the trajectories toward the optimal solution of problem $(\mathcal{P})$ of minimum norm. When Φ is a convex differentiable function whose gradient is Lipschitz continuous, we show that the convergence holds for the strong topology. Finally, we examine some variants of this principle of hierarchical minimization.

2. Existence and Uniqueness of Global Solutions

We consider the Cauchy problem for the differential inclusion system (3a)-(3b)

$$\upsilon(t) \in \partial\Phi(x(t)) \tag{7a}$$

$$\lambda\dot{x}(t) + \dot{\upsilon}(t) + \upsilon(t) + \varepsilon(t)\,x(t) = 0 \tag{7b}$$

$$x(0) = x_0,\ \upsilon(0) = \upsilon_0 \tag{7c}$$

First, we are going to define a notion of strong solution to the above system. Then, we shall reformulate this system with the help of the Minty representation of $\partial\Phi$. Finally, we shall prove the existence and uniqueness of a strong solution to system (7a)–(7c), by applying the Cauchy–Lipschitz theorem to this equivalent formulation.

2.1. **Definition of strong solutions.** We say that the pair $(x(\cdot), \upsilon(\cdot))$ is a strong global solution of (7a)–(7c) iff the following properties are satisfied:
i) $x(\cdot), \upsilon(\cdot) : [0, +\infty[\longrightarrow \mathcal{H}$ are absolutely continuous on each interval $[0, b]$, $0 < b < +\infty$;
ii) $\upsilon(t) \in \partial\Phi(x(t))$ for all $t \in [0, +\infty[$;
iii) $\lambda\dot{x}(t) + \dot{\upsilon}(t) + \upsilon(t) + \varepsilon(t)\,x(t) = 0$ for almost all $t \in [0, +\infty[$;
iv) $x(0) = x_0,\ \upsilon(0) = \upsilon_0$.

2.2. **Equivalent formulation as a classical differential equation.** In order to solve system (7a)–(7c) we use Minty's device. Set

$$\mu = \frac{1}{\lambda}.$$

Let us rewrite inclusion (7a) by using the following equivalences: for any $t \in [0, +\infty[$

$$\upsilon(t) \in \partial\Phi(x(t)) \Leftrightarrow \tag{8}$$

$$x(t) + \mu\upsilon(t) \in x(t) + \mu\partial\Phi(x(t)) \tag{9}$$

$$x(t) = (I + \mu\partial\Phi)^{-1}(x(t) + \mu\upsilon(t)). \tag{10}$$

Let us introduce the new unknown function $y : [0, +\infty[\longrightarrow \mathcal{H}$ which is defined for $t \in [0, +\infty[$ by

$$y(t) := x(t) + \mu\upsilon(t), \tag{11}$$

and rewrite the system (7a)–(7c) with the help of (x, y). From (10) and (11)

$$x(t) = (I + \mu\partial\Phi)^{-1}(y(t)),$$
$$\upsilon(t) = \frac{1}{\mu}\Big(y(t) - (I + \mu\partial\Phi)^{-1}(y(t))\Big).$$

Equivalently,

$$x(t) = \operatorname{prox}_{\mu\Phi}(y(t)); \tag{12}$$

$$\upsilon(t) = \nabla\Phi_\mu(y(t)), \tag{13}$$

where $\operatorname{prox}_{\mu\Phi}$ is the proximal mapping associated to $\mu\Phi$. Recall that $\operatorname{prox}_{\mu\Phi} = (I + \mu\partial\Phi)^{-1}$ is the resolvent of index $\mu > 0$ of the maximal monotone operator $\partial\Phi$, and $\nabla\Phi_\mu$ is its Yosida approximation of index $\mu > 0$.

Let us show how (7b) can be reformulated as a classical differential equation with respect to $y(\cdot)$. First, let us rewrite (7b) as

$$\dot{x}(t) + \mu\dot{\upsilon}(t) + \mu\upsilon(t) + \mu\varepsilon(t)\,x(t) = 0. \tag{14}$$

Differentiating (11), and using (14) we obtain

$$\dot{y}(t) = \dot{x}(t) + \mu\dot{\upsilon}(t) \tag{15}$$

$$= -\mu\upsilon(t) - \mu\varepsilon(t)\,x(t). \tag{16}$$

From (12), (13), and (16) we deduce that

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\operatorname{prox}_{\mu\Phi}(y(t)) = 0.$$

Finally, the (x, y) system can be written as

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{17a}$$

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\,\text{prox}_{\mu\Phi}(y(t)) = 0. \tag{17b}$$

Conversely, if $y(\cdot)$ is a solution of (17b), then $(x(\cdot), \upsilon(\cdot))$ with $x(t) = \text{prox}_{\mu\Phi}(y(t))$, $\upsilon(t) = \nabla\Phi_\mu(y(t))$ is a solution of (7a)-(7c). Let us stress the fact that the operators $\text{prox}_{\mu\Phi} : \mathcal{H} \longrightarrow \mathcal{H}$, $\nabla\Phi_\mu : \mathcal{H} \longrightarrow \mathcal{H}$ are everywhere defined and Lipschitz continuous, which makes this system relevant to the Cauchy–Lipschitz theorem.

2.3. **Global existence and uniqueness results.** Let us state our main result of existence and uniqueness for the system (7a)–(7c).

Theorem 2.1. *Suppose that $\Phi : \mathcal{H} \longrightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function, and $\lambda > 0$ is a positive constant. Let $\epsilon : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ be a nonnegative locally integrable function, and $(x_0, \upsilon_0) \in \mathcal{H} \times \mathcal{H}$ be such that $\upsilon_0 \in \partial\Phi(x_0)$. Then the following properties hold:*

i) there exists a unique strong global solution $(x(\cdot), \upsilon(\cdot)) : [0, +\infty[\longrightarrow \mathcal{H} \times \mathcal{H}$ of the Cauchy problem (7a)-(7c);

ii) the solution pair $(x(\cdot), \upsilon(\cdot))$ of (7a)-(7c) can be represented as follows: for any $t \in [0, +\infty[$,

$$x(t) = \text{prox}_{\mu\Phi}(y(t)); \tag{18}$$

$$\upsilon(t) = \nabla\Phi_\mu(y(t)), \tag{19}$$

where $y(\cdot) : [0, +\infty[\longrightarrow \mathcal{H}$ is the unique strong global solution of the Cauchy problem

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\,\text{prox}_{\mu\Phi}(y(t)) = 0, \tag{20a}$$

$$y(0) = x_0 + \mu\upsilon_0. \tag{20b}$$

Proof. Let us first prove the existence and uniqueness of a strong global solution of the Cauchy problem (20a)-(20b). The Cauchy problem (20a)–(20b) can be equivalently written in abstract form, as the following non-autonomous differential system

$$\begin{cases} \dot{y}(t) = F(t, y(t)); \\ y(0) = x_0 + \mu\upsilon_0, \end{cases} \tag{21}$$

with

$$F(t, y) = G(t, y) + K(t, y), \tag{22}$$

$$G(t, y) = -\mu\nabla\Phi_\mu(y), \tag{23}$$

$$K(t, y) = -\mu\epsilon(t)\,\text{prox}_{\mu\Phi}(y). \tag{24}$$

In order to apply the Cauchy–Lipschitz theorem to (21), let us first examine the Lipschitz continuity properties of $F(t, \cdot)$.

(a) Take arbitrary $y_i \in \mathcal{H}$, $i = 1, 2$. The Yosida approximation $\nabla\Phi_\mu$ is $\frac{1}{\mu}$-Lipschitz continuous (see [12]), and hence, for any $t \geq 0$, $G(t, \cdot) : \mathcal{H} \longrightarrow \mathcal{H}$ is nonexpansive, i.e.,

$$\|G(t, y_2) - G(t, y_1)\| \leq \|y_2 - y_1\|. \tag{25}$$

By the nonexpansive property of the resolvent operators we have

$$\|K(t, y_2) - K(t, y_1)\| \leq \mu\epsilon(t)\,\|y_2 - y_1\|. \tag{26}$$

Hence,

$$\|F(t, y_2) - F(t, y_1)\| \leq (1 + \mu\epsilon(t))\,\|y_2 - y_1\|. \tag{27}$$

Since $\varepsilon(\cdot)$ is nonegative and locally integrable, (27) shows that the Lipschitz constant $L_F(t) = (1 + \mu\varepsilon(t))$ of $F(t, \cdot)$ satisfies

$$L_F(\cdot) \in L^1([0, b]) \quad \text{for any } 0 < b < +\infty. \tag{28}$$

(b) Let us show that

$$\forall y \in \mathcal{H},\ \forall b > 0,\ F(\cdot, y) \in L^1([0, b]; \mathcal{H}). \tag{29}$$

Returning to the definition (22) of F, we deduce that

$$\|F(t,y)\| \leq \mu \|\nabla\Phi_\mu(y)\| + \mu\epsilon(t)\left\|\mathrm{prox}_{\mu\Phi}(y)\right\|.$$

By assumption, $\epsilon : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is a nonnegative locally integrable function, which gives (29). From (27) and (29), by Cauchy-Lipschitz theorem (see [17], [21] for the nonautonomous version used here), we deduce the existence and uniqueness of a strong global solution of the Cauchy problem (21), and hence of (20a)-(20b).

(2) Let us return to the initial problem (7a)-(7c). Given $y(\cdot) : [0,+\infty[\longrightarrow \mathcal{H}$ which is the unique strong solution of Cauchy problem (20a)-(20b), let us define $x(\cdot), \upsilon(\cdot) : [0,+\infty[\longrightarrow \mathcal{H}$ by

$$x(t) = \mathrm{prox}_{\mu\Phi}(y(t)), \qquad \upsilon(t) = \nabla\Phi_\mu(y(t)). \tag{30}$$

(a) Let us show that $x(\cdot), \upsilon(\cdot)$ are absolutely continuous on each bounded interval, and satisfy (7a)-(7c). Let us give arbitrary $y_1 \in \mathcal{H}$, $y_2 \in \mathcal{H}$. By the nonexpansive property of the resolvents, we have

$$\left\|\mathrm{prox}_{\mu\Phi}(y_2) - \mathrm{prox}_{\mu\Phi}(y_1)\right\| \leq \|y_2 - y_1\|. \tag{31}$$

Assuming that $s,t \in [0,b]$, by taking $y_1 = y(s)$, $y_2 = y(t)$ in (31), and owing to the definition of the absolute continuity property, we deduce that $x(t) = \mathrm{prox}_{\mu\Phi}(y(t))$ is absolutely continuous on $[0,b]$ for any $b > 0$. As a linear combination of two absolutely continuous functions, the same property holds true for $\upsilon(t) = \lambda(y(t) - x(t))$.
Moreover, for any $t \in [0,+\infty]$

$$\upsilon(t) \in \partial\Phi(x(t)), \qquad y(t) = x(t) + \mu\upsilon(t).$$

Differentiation of the above equation shows that, for almost every $t > 0$,

$$\dot{x}(t) + \mu\dot{\upsilon}(t) = \dot{y}(t).$$

On the other hand, owing to $\upsilon(t) = \nabla\Phi_\mu(y(t))$, $x(t) = \mathrm{prox}_{\mu\Phi}(y(t))$, (20a) can be equivalently written as

$$\dot{y}(t) + \mu\upsilon(t) + \mu\epsilon(t)x(t) = 0.$$

Combining the two above equations, we obtain

$$\dot{x}(t) + \mu\dot{\upsilon}(t) + \mu\upsilon(t) + \mu\epsilon(t)x(t) = 0.$$

From $\mu = \frac{1}{\lambda}$, we conclude that $(x(\cdot), \upsilon(\cdot))$ is a solution of system (7a)-(7b).
Regarding the initial condition, we observe that

$$y(0) = x_0 + \mu\upsilon_0 \tag{32}$$
$$= x(0) + \mu\upsilon(0), \tag{33}$$

with $\upsilon_0 \in \partial\Phi(x_0)$ and $\upsilon(0) \in \partial\Phi(x(0))$. Hence

$$x(0) = x_0 = (I + \mu\partial\Phi)^{-1}(x_0 + \mu\upsilon_0).$$

After simplification, we obtain $\upsilon(0) = \upsilon_0$.

(b) Let us now prove the uniqueness. Suppose that

$$x(\cdot), \upsilon(\cdot) : [0,+\infty[\longrightarrow \mathcal{H}$$

is a solution pair of (7a)-(7c). Defining $\mu = \frac{1}{\lambda}$ and

$$y(t) = x(t) + \mu\upsilon(t) \tag{34}$$

we conclude that $y(\cdot)$ is absolutely continuous, $y_0 = x_0 + \mu\upsilon_0$, and for any $t \in [0,+\infty[$

$$x(t) = (I + \mu\partial\Phi)^{-1}(y(t)), \quad \upsilon(t) = \nabla\Phi_\mu(y(t)). \tag{35}$$

Since the functions involved in the definition (34) of $y(\cdot)$, namely $x(\cdot)$ and $\upsilon(\cdot)$, are differentiable for almost all $t \in [0,+\infty[$, we have for almost $t \in [0,+\infty[$

$$\begin{aligned} \dot{y}(t) &= \dot{x}(t) + \mu\dot{\upsilon}(t) \\ &= -\mu(\dot{\upsilon}(t) + \upsilon(t) + \epsilon(t)x(t)) + \mu\dot{\upsilon}(t). \end{aligned}$$

Since $\upsilon(t) = \nabla\Phi_\mu(y(t))$, we finally obtain

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\,\mathrm{prox}_{\mu\Phi}(y(t)) = 0.$$

Moreover

$$y_0 = x_0 + \mu \upsilon_0.$$

Arguing as before, by the Cauchy–Lipschitz theorem, the solution $y(\cdot)$ of the above system is uniquely determined, and locally absolutely continuous. Thus, by (35), $x(\cdot)$ and $\upsilon(\cdot)$ are uniquely determined. □

3. Asymptotic analysis and convergence properties

In this section, we study the asymptotic behavior, as $t \to +\infty$, of the trajectories of system (7a)-(7b). Let us recall our standing assumption, namely the parametrization $\epsilon(\cdot)$ is supposed to be nonnegative, and locally integrable. In view of the asymptotic analysis, we also suppose that $\epsilon(t) \to 0$ as $t \to \infty$, and satisfies the "slow" decay property

$$\int_0^{+\infty} \varepsilon(t)\, dt = +\infty.$$

By Theorem 2.1, for any given Cauchy data $\upsilon_0 \in \partial\Phi(x_0)$, the above properties guarantee the existence and uniqueness of a global solution of system (7a)-(7b)-(7c). From now on in this section, $(x(\cdot), \upsilon(\cdot)) : [0, +\infty[\longrightarrow \mathcal{H} \times \mathcal{H}$ is the solution of (7a)-(7b)-(7c). We first study the asymptotic behavior, as $t \to +\infty$, of the trajectories of the associated system (20a)

$$\dot{y}(t) + \mu \nabla \Phi_\mu(y(t)) + \mu \epsilon(t) \operatorname{prox}_{\mu\Phi}(y(t)) = 0$$

whose existence is guaranteed by Theorem 2.1. The central point of our analysis is to reformulate this system as a multi-scale gradient system, which will allow us to use the known results concerning the asymptotic behavior, and the hierarchical selection property for such systems.

3.1. Preliminary results.

Let us state some definitions and classical properties that will be useful (see [5], [11], [12], [20] for an extended presentation of these notions):

Definition 3.1. Let f and g be functions from $\mathcal{H}$ to $\mathbb{R} \cup \{+\infty\}$. The *infimal convolution* (or *epi-sum*) of f and g is the function $f \Box g : \mathcal{H} \to [-\infty, +\infty]$ which is defined by

$$f \Box g(x) = \inf_{\xi \in \mathcal{H}} \left(f(\xi) + g(x - \xi) \right).$$

Definition 3.2. Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$, $\gamma \in \mathbb{R}_{++}$. The Moreau envelope of f of parameter γ is defined by

$$f_\gamma = f \Box \left(\frac{1}{2\gamma} \|\cdot\|^2 \right).$$

Definition 3.3. Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a convex lower semicontinuous proper function, and let $x \in \mathcal{H}$. Then $\operatorname{prox}_f x$ is the unique point in $\mathcal{H}$ that satisfies

$$f_1(x) = \min_{\xi \in \mathcal{H}} \left(f(\xi) + \frac{1}{2} \|x - \xi\|^2 \right) = f\left(\operatorname{prox}_f x\right) + \frac{1}{2} \left\|x - \operatorname{prox}_f x\right\|^2.$$

The operator $\operatorname{prox}_f : \mathcal{H} \to \mathcal{H}$ is called the proximity operator, or proximal mapping of f.

Definition 3.4. Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$. The conjugate (or Legendre-Fenchel transform, or Fenchel conjugate) of f is

$$f^* : \mathcal{H} \to \mathbb{R} \cup \{+\infty\} : u \mapsto \sup_{x \in \mathcal{H}} \left(\langle x, u \rangle - f(x) \right).$$

Remark 3.1. Let $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper then

$$f^*(0) = -\inf_{\mathcal{H}} f.$$

f is lower semi continuous and convex if and only if

$$f = f^{**}.$$

Remark 3.2. Let f and g be functions from $\mathcal{H}$ to $\mathbb{R} \cup \{+\infty\}$. Then

$$(f \Box g)^* = f^* + g^*.$$

Conversely, if one of the functions (f or g) is continuous at a point of the domain of the other, then

$$(f + g)^* = f^* \Box g^*.$$

Remark 3.3. a) Let $\varphi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be proper, and $\gamma \in \mathbb{R}_{++}$. Set $f = \varphi + \frac{1}{2\gamma} \left\|\cdot\right\|^2$. Then $\forall u \in \mathcal{H}$

$$f^*(u) = \frac{\gamma}{2} \left\|u\right\|^2 - \varphi_\gamma(\gamma u).$$

b) Let C be a nonempty subset of $\mathcal{H}$, and let $f = \delta_C + \left\|\cdot\right\|^2/2$, where δ_C is the indicator function of the set C ($\delta_C(x) = 0$ for $x \in \mathcal{H}$, $+\infty$ outwards). Then

$$f^* = \left(\left\|\cdot\right\|^2 - d_C^2\right)/2,$$

where d_c is the distance function to the set C. For the proof, set $\varphi = \delta_C$ and $\gamma = 1$ in Remark 3.3.

In the next lemma, we show that the proximal mapping can be written as the gradient of a convex differentiable function. This result will play a crucial role in our analysis.

Lemma 3.1. *Let $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex lower semicontinuous function, and let $\mu > 0$. Then, the proximal mapping $prox_{\mu\Phi} : \mathcal{H} \to \mathcal{H}$ can be written as the gradient*

$$prox_{\mu\Phi} = \nabla\psi$$

of the convex continuously differentiable function $\psi : \mathcal{H} \to \mathbb{R}$ which is defined, for any $y \in \mathcal{H}$, by

$$\psi(y) = \mu\left(\Phi^*\right)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right)$$

where $\Phi^*$is the Fenchel conjugate of Φ.

Proof. For any $y \in \mathcal{H}$, set

$$x = \text{prox}_{\mu\Phi}(y).$$

By definition of the proximal mapping, we have the following equivalent formulations

$$\begin{aligned}
&x = (I + \mu\partial\Phi)^{-1}(y)\\
&y \in (I + \mu\partial\Phi)(x)\\
&y - x \in \mu\partial\Phi(x)\\
&\frac{1}{\mu}(y - x) \in \partial\Phi(x).
\end{aligned}$$

From $(\partial\Phi)^{-1} = \partial\Phi^*$ and the above equality, we successively obtain

$$\begin{aligned}
&x \in \partial\Phi^*\left(\frac{1}{\mu}(y - x)\right);\\
&\frac{1}{\mu}y \in \frac{1}{\mu}(y - x) + \frac{1}{\mu}\partial\Phi^*\left(\frac{1}{\mu}(y - x)\right) = \left(I + \frac{1}{\mu}\partial\Phi^*\right)\left(\frac{1}{\mu}(y - x)\right);\\
&\frac{1}{\mu}(y - x) = \left(I + \frac{1}{\mu}\partial\Phi^*\right)^{-1}\left(\frac{1}{\mu}y\right);\\
&x = \mu\left(\frac{1}{\mu}y - \left(I + \frac{1}{\mu}\partial\Phi^*\right)^{-1}\left(\frac{1}{\mu}y\right)\right);\\
&x = \mu\left(I - \text{prox}_{\frac{1}{\mu}\Phi^*}\right)\left(\frac{1}{\mu}y\right).
\end{aligned}$$

By the definition of the Yosida approximation of index $\frac{1}{\mu} > 0$ of the maximal monotone operator $\partial\Phi^*$

$$x = \nabla\left[\left(\Phi^*\right)_{\frac{1}{\mu}}\right]\left(\frac{1}{\mu}y\right).$$

By the classical derivation chain rule, we deduce that the the proximal mapping $\text{prox}_{\mu\Phi}y : \mathcal{H} \to \mathcal{H}$ is the gradient of the convex continuously differentiable function $\psi : \mathcal{H} \to \mathbb{R}$ which is defined, for any $y \in \mathcal{H}$, by

$$\psi(y) = \mu\left(\Phi^*\right)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right).$$

□

Let us further analyze the function ψ, and give equivalent formulations which come with different proofs of the above lemma. By Definition 3.2 of the Moreau envelope of Φ of parameter μ

$$\Phi_\mu(y) = \inf_{x\in\mathcal{H}}\left\{\Phi(x) + \frac{1}{2\mu}\|y-x\|^2\right\}, \tag{36}$$

and since the Yosida approximation of the subdifferential of Φ is the Fréchet derivative of Moreau envelope,

$$\nabla\Phi_\mu(y) = \frac{1}{\mu}\left(y - J_\mu^{\partial\Phi}y\right)$$

$$\mu\nabla\Phi_\mu(y) = y - J_\mu^{\partial\Phi}y.$$

By definition, we have $\text{prox}_{\mu\Phi}(y) = J_\mu^{\partial\Phi}y = (I+\mu\partial\Phi)^{-1}(y)$. Hence

$$\mu\nabla\Phi_\mu(y) = y - \text{prox}_{\mu\Phi}(y).$$

$$\begin{aligned}\text{prox}_{\mu\Phi}(y) &= y - \mu\nabla\Phi_\mu(y)\\ &= \nabla\left(\frac{1}{2}\|\cdot\|^2 - \mu\Phi_\mu\right)(y).\end{aligned}$$

Let us make the link with the previous formulation of the prox as a gradient, and show that, for any $y\in\mathcal{H}$

$$\frac{1}{2}\|y\|^2 - \mu\Phi_\mu(y) = \mu\Phi^*_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right);$$

By (36), we have

$$\begin{aligned}\frac{1}{2}\|y\|^2 - \mu\Phi_\mu(y) &= \frac{1}{2}\|y\|^2 - \mu\inf_{x\in\mathcal{H}}\left\{\Phi(x) + \frac{1}{2\mu}\|y-x\|^2\right\}\\ &= \frac{1}{2}\|y\|^2 - \mu\inf_{x\in\mathcal{H}}\left\{\Phi(x) + \frac{1}{2\mu}\|y\|^2 - \frac{1}{\mu}\langle y,x\rangle + \frac{1}{2\mu}\|x\|^2\right\}\\ &= \sup_{x\in\mathcal{H}}\left\{\langle y,x\rangle - \mu\left(\Phi(x) + \frac{1}{2\mu}\|x\|^2\right)\right\}\\ &= \mu\sup_{x\in\mathcal{H}}\left\{\left\langle\frac{1}{\mu}y,x\right\rangle - \left(\Phi(x) + \frac{1}{2\mu}\|x\|^2\right)\right\}.\end{aligned}$$

By using Remark 3.2 concerning the conjugate of a sum, we obtain

$$\begin{aligned}\frac{1}{2}\|y\|^2 - \mu\Phi_\mu(y) &= \mu\left(\Phi + \frac{1}{2\mu}\|\cdot\|^2\right)^*\left(\frac{1}{\mu}y\right)\\ &= \mu\left(\Phi^*\square\frac{\mu}{2}\|\cdot\|^2\right)\left(\frac{1}{\mu}y\right)\\ &= \mu\left(\Phi^*\right)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right).\end{aligned}$$

So we obtain the same function ψ as given in Lemma 3.1. Note that by Remark 3.3, the equivalent (dual) formulation of ψ given by $\psi(y) = \frac{1}{2}\|y\|^2 - \mu\Phi_\mu(y)$, which is written as a d.c. function, is actually a convex function.

3.2. **Asymptotic hierarchical minimization.** Let us study the asymptotic behavior of the trajectories of system (7a)-(7b). We consider the equivalent system (20a), which, by Lemma 3.1, can be formulated as follows:

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{37}$$

$$\dot{y}(t) + \nabla\Theta(y(t)) + \epsilon(t)\nabla\Psi(y(t)) = 0, \tag{38}$$

where, for any $y\in\mathcal{H}$

$$\Theta(y) := \mu\Phi_\mu(y); \tag{39}$$

$$\Psi(y) = \mu^2\left(\Phi^*\right)_{\frac{1}{\mu}}\left(\frac{1}{\mu}y\right). \tag{40}$$

Note that Θ and Ψ are two convex continuously differentiable functions. We are within the framework of the multiscale gradient system $(\mathrm{MAG})_\varepsilon$, with a positive control $t \mapsto \varepsilon(t)$ that converges to 0 as $t \to \infty$,

$$(\mathrm{MAG})_\varepsilon \qquad \dot{y}(t) + \partial\Theta(y(t)) + \varepsilon(t)\partial\Psi(y(t)) \ni 0,$$

that has been considered by Attouch-Czarnecki in [6]. Let us recall this general abstract result, that we formulate with notations adapted to our setting. Since Θ enters $(\mathrm{MAG})_\varepsilon$ only by its subdifferential, it is not a restrictive assumption to assume this potential to be nonnegative, with its infimal value equal to zero (substracting the infimal value does not affect the subdifferential).

Theorem 3.1. *(Attouch-Czarnecki, [6]) Let*

- $\Theta : \mathcal{H} \to \mathbb{R}^+ \cup \{+\infty\}$ *be a closed convex proper function, such that* $C = \mathrm{argmin}\Theta = \Theta^{-1}(0) \neq \emptyset$.
- $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ *be a closed convex proper function, such that* $S = \mathrm{argmin}\{\Psi|\mathrm{argmin}\Theta\} \neq \emptyset$.

Let us assume that,

- $(\mathcal{H}_1)_\varepsilon$ $\forall p \in R(N_C)$, $\displaystyle\int_0^{+\infty} \Theta^*(\varepsilon(t)p) - \sigma_C(\varepsilon(t)p)dt < +\infty$.
- $(\mathcal{H}_2)_\varepsilon$ $\varepsilon(\cdot)$ *is a non increasing function of class* C^1, *such that* $\lim_{t\to+\infty} \varepsilon(t) = 0$, $\displaystyle\int_0^{+\infty} \varepsilon(t)dt = +\infty$, *and for some* $k \geq 0$, $-k\varepsilon^2 \leq \dot{\varepsilon}$.

Let $y(\cdot)$ *be a strong solution of* $(\mathrm{MAG})_\varepsilon$. *Then:*

(*i*) *weak convergence* $\exists y_\infty \in S = \mathrm{argmin}\{\Psi|\mathrm{argmin}\Theta\}, \qquad w - \lim_{t\to+\infty} y(t) = y_\infty;$

(*ii*) *minimizing properties* $\lim_{t\to+\infty} \Theta(y(t)) = 0;$

$\lim_{t\to+\infty} \Psi(y(t)) = \min \Psi|_{\mathrm{argmin}\Theta};$

(*iii*) $\forall z \in S \lim_{t\to+\infty} \|y(t) - z\|$ *exists* ;

(*iv*) *estimations* $\displaystyle\lim_{t\to+\infty} \frac{1}{\varepsilon(t)}\Theta(y(t)) = 0;$

$$\int_0^{+\infty} \Theta(y(t))dt < +\infty;$$

$$\limsup_{\tau\to+\infty} \int_0^{\tau} \varepsilon(t) \left(\Psi(y(t)) - \min \Psi|_{\mathrm{argmin}\Theta}\right) dt < +\infty.$$

By specializing this result to our setting, we will obtain the weak convergence of $y(\cdot)$ to a particular minimizer of Φ, which is the solution of a hierarchical minimization property. The convergence of $x(\cdot)$ is less immediate, and will follow from an energetical argument.

Analysis of the condition $(\mathcal{H}_1)_\epsilon$: The condition

$$(\mathcal{H}_1)_\varepsilon \qquad \forall p \in R(N_C), \qquad \int_0^{+\infty} \Theta^*(\varepsilon(t)p) - \sigma_C(\varepsilon(t)p)dt < +\infty,$$

plays a crucial role in our asymptotic analysis. Before proceeding in the discussion of this hypothesis, we recall some classical notions from convex analysis, that will be useful.

- σ_C is the support function of C,
$$\sigma_C(x^*) = \sup_{c\in C} \langle x^*, c\rangle .$$
- $N_C(x)$ is the normal cone to C at x,
$$N_C(x) = \{x^* \in \mathcal{H} : \langle x^*, c - x\rangle \leq 0 \text{ for all } c \in C\} \text{ if } x \in C, \text{ and } \emptyset \text{ otherwise.}$$
- $R(N_C)$ is the range of N_C, i.e. $p \in R(N_C)$ if and only if $p \in N_C(x)$ for some $x \in C$.
- Note that $\delta_C^* = \sigma_C$ where δ_C is the indicator function of C,
$$\delta_C := \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Observe that in $(\mathcal{H}_1)_\varepsilon$, all the terms in the integral are nonnegative. Indeed, since Θ is bounded from above by the indicator function of the set C, i.e. $\Theta \leq \delta_C$ (recall that $\Theta = 0$ on C), the reverse inequality holds for their Fenchel conjugates, whence

$$\Theta^* (\epsilon (t) p) - \sigma_C (\epsilon (t) p) \geq 0 \qquad \forall p \in \mathcal{H}.$$

Thus, Hypothesis $(\mathcal{H}_1)_\epsilon$ means that, for all $p \in R(N_C)$ the nonnegative function

$$t \mapsto [\Theta^* (\epsilon (t) p) - \sigma_C (\epsilon (t) p)]$$

is integrable on $(0, +\infty)$. For more clarity, let us discuss the following special case: Suppose that

$$\Theta (x) \geq \frac{r}{2} \mathrm{dis}^2 (x, C),$$

for some $r > 0$. Then $\Theta^* (x) \leq \frac{1}{2r} \|x\|^2 + \sigma_C (x)$ and

$$\Theta^* (z) - \sigma_C (z) \leq \frac{1}{2r} \|z\|^2 .$$

Hence, in this situation $(\mathcal{H}_1)_\epsilon$ is satisfied if the following condition on $\epsilon(\cdot)$ is satisfied:

$$\int_0^{+\infty} \epsilon^2 (t) < +\infty.$$

In this situation, the moderate growth condition on $\epsilon(\cdot)$, can be formulated as

$$\epsilon(\cdot) \in L^2(0, +\infty) \setminus L^1(0, +\infty).$$

Let us return to the general situation, and summarize our results in the following theorem, which is our main statement.

Theorem 3.2. *Let $\Phi : \mathcal{H} \to \mathbb{R}^+ \cup \{+\infty\}$ be a closed convex function, such that $C = \mathrm{argmin}\Phi = \Phi^{-1}(0) \neq \emptyset$. Let us assume that,*

- $(\mathcal{H}_1)_\varepsilon$ $\forall p \in R(N_C)$, $\displaystyle\int_0^{+\infty} \mu\Phi^*(\frac{1}{\mu}\varepsilon(t)p) - \sigma_C(\varepsilon(t)p) + \frac{1}{2}\|\varepsilon(t)p\|^2 dt < +\infty$;
- $(\mathcal{H}_2)_\varepsilon$ *$\varepsilon(\cdot)$ is a nonincreasing function of class C^1, Lipschitz continuous on $[0, +\infty[$, and such that* $\lim_{t\to+\infty} \varepsilon(t) = 0$, $\displaystyle\int_0^{+\infty} \varepsilon(t)dt = +\infty$, *and for some $k \geq 0$, $-k\varepsilon^2 \leq \dot{\varepsilon}$.*

Then, for any trajectory $(x(\cdot), \upsilon(\cdot)) : [0, +\infty[\longrightarrow \mathcal{H} \times \mathcal{H}$ solution of (7a)-(7b), with $y(t) = x(t) + \mu\upsilon(t)$

(i) *weak convergence* $\quad w - \lim_{t\to+\infty} y(t) = \mathrm{proj}_{\mathrm{argmin}\Phi} 0.$

Let us further assume that $\Phi(0) < +\infty$. Then

(ii) *weak convergence* $\quad w - \lim\limits_{t\to+\infty} x(t) = w - \lim\limits_{t\to+\infty} y(t) = \mathrm{proj}_{\mathrm{argmin}\Phi} 0;$

(iii) *strong convergence* $\quad s - \lim\limits_{t\to+\infty} \upsilon(t) = 0,$ *and hence* $s - \lim\limits_{t\to+\infty} x(t) - y(t) = 0.$

Proof. Let us apply Theorem 3.1 with $\Theta(y) = \mu\Phi_\mu(y)$, and $\Psi(y) = \mu^2 (\Phi^*)_{\frac{1}{\mu}} \left(\frac{1}{\mu} y\right)$, which are two convex continuously differentiable functions. By the general properties of the Moreau enveloppe, the function Θ is still nonnegative, and

$$\mathrm{argmin}\Theta = \mathrm{argmin}\Phi_\mu = \mathrm{argmin}\Phi = C.$$

Moreover, on C we have $\Theta = 0$. Let us particularize the conditions $(\mathcal{H}_1)_\varepsilon$ and $(\mathcal{H}_2)_\varepsilon$ to our setting. Let us first compute Θ^*, the conjugate of Θ. For any $z \in \mathcal{H}$

$$\begin{aligned}
\Theta^*(z) &= \sup_y \{\langle z, y\rangle - \mu\Phi_\mu(y)\} \\
&= \mu \sup_y \left\{\langle \frac{1}{\mu} z, y\rangle - \Phi_\mu(y)\right\} \\
&= \mu \left(\Phi_\mu\right)^* \left(\frac{1}{\mu} z\right) \\
&= \mu \left(\Phi^* \left(\frac{1}{\mu} z\right) + \frac{\mu}{2} \|\frac{1}{\mu} z\|^2\right) \\
&= \mu \Phi^* \left(\frac{1}{\mu} z\right) + \frac{1}{2}\|z\|^2.
\end{aligned}$$

Hence conditions $(\mathcal{H}_1)_\varepsilon$ and $(\mathcal{H}_2)_\varepsilon$ of Theorem 3.1 are satisfied. As a consequence, we obtain the convergence of $y(\cdot)$ to a solution y_∞ of the constrained minimization problem

$$\min \{\Psi(x) : \ x \in C\}. \tag{41}$$

Let us write the first-order optimality condition satisfied by y_∞. We have

$$\nabla\Psi(y_\infty) + N_C(y_\infty) \ni 0. \tag{42}$$

Since $\nabla\Psi = \mu \mathrm{prox}_{\mu\Phi}$, equivalently

$$\mu \mathrm{prox}_{\mu\Phi}(y_\infty) + N_C(y_\infty) \ni 0. \tag{43}$$

Noticing that $y_\infty \in C$, and that $z = \mathrm{prox}_{\mu\Phi} z$ for $z \in C = \mathrm{argmin}\Phi$, we obtain

$$\mu y_\infty + N_C(y_\infty) \ni 0. \tag{44}$$

By definition of N_C, equivalently, the following property is satisfied

$$\langle 0 - y_\infty, c - y_\infty\rangle \leq 0 \quad \forall c \in C.$$

Since $y_\infty \in C$, this is the condition of the obtuse angle that characterizes the projection of the origin on C. Thus

$$y_\infty = \mathrm{proj}_C(0). \tag{45}$$

We have obtained that $y(\cdot)$ converges weakly to the element of minimal norm of the solution set C, that's item (i). In order to pass from the convergence of $y(\cdot)$ to the convergence of $x(\cdot)$ we use the relation (12)

$$x\,(t) = \mathrm{prox}_{\mu\Phi}\,(y\,(t))$$

that links the two variables.

In a finite dimensional setting, we can conclude the strong convergence of $x(\cdot)$ thanks to the continuity of the proximal mapping, and using again the fact that the set C of minimizers of Φ is invariant by the proximal mapping $\mathrm{prox}_{\mu\Phi}$, i.e.,

$$\mathrm{prox}_{\mu\Phi}(y) = y \quad \text{for all } y \in C = \mathrm{argmin}\Phi.$$

In an infinite dimensional setting, we are going to use the particular structure of our dynamical system, and an energetical argument to show that

$$x(t) - y(t) \to 0 \text{ strongly as } t \to +\infty. \tag{46}$$

This will result from the finite energy property

$$\int_0^\infty \|\dot{y}(t)\|^2 dt < +\infty. \tag{47}$$

To obtain (47), take the scalar product of equation (38) with $\dot{y}\,(t)$. We obtain

$$\|\dot{y}\,(t)\,\|^2 + \frac{d}{dt}\left(\Theta\,(y\,(t))\right) + \epsilon\,(t)\,\frac{d}{dt}\left(\Psi\,(y\,(t))\right) = 0. \tag{48}$$

After integration by parts we obtain, for any $T>0$

$$\int_0^T \|\dot{y}(t)\|^2 dt + \Theta(y(T)) - \Theta(y(0)) + \epsilon(T)\Psi(y(T)) - \epsilon(0)\Psi(y(0)) - \int_0^T \dot{\epsilon}(t)\Psi(y(t))\,dt = 0. \tag{49}$$

By assumption $\Phi(0) < +\infty$. By Remark 3.1 $\Phi^{**} = \Phi$, we equivalently have $\inf \Phi^* > -\infty$, and hence

$$\inf \Psi = \inf \mu^2 (\Phi^*)_{\frac{1}{\mu}} = \mu^2 \inf \Phi^* > -\infty.$$

Set $m := \inf \Psi$. From (49), and $\dot{\epsilon}(t) \leq 0$ (recall that $\varepsilon(\cdot)$ is a nonincreasing function) we deduce that

$$\int_0^T \|\dot{y}(t)\|^2 dt \leq \Theta(y(0)) + \epsilon(0)\Psi(y(0)) + |m|\epsilon(T) + m\int_0^T \dot{\epsilon}(t)dt. \tag{50}$$

Since the above majorization is valid for any $T>0$, and ϵ is bounded (it decreases to zero), we obtain (47).
We now observe that $y(\cdot)$ is Lipschitz continuous on $[0,+\infty[$. This follows from equation (38), and the following argument. Since $y(\cdot)$ is converging weakly, it is bounded. By the Lipschitz continuity of the operators $\nabla\Theta = \mu\nabla\Phi_\mu$ and $\nabla\Psi = \mu\text{prox}_{\mu\Phi}$ (which are therefore bounded on bounded sets), and by equation (38), we deduce that $\dot{y}(\cdot)$ is bounded, and hence $y(\cdot)$ is Lipschitz continuous on $[0,+\infty[$.
Using again equation (38), and the Lipschitz continuity properties of $\nabla\Theta$ and $\nabla\Psi$, y, and ε (for this last property note that for some $c>0$, $-c \leq -k\varepsilon^2 \leq \dot{\varepsilon} \leq 0$), we deduce that $t \mapsto \dot{y}(t)$ is Lipschitz continuous on $[0,+\infty[$. Hence $\dot{y}(\cdot)$ belongs to $L^2([0,+\infty[;\mathcal{H})$, and is Lipschitz continuous. By a classical result this implies

$$\lim_{t\to+\infty} \dot{y}(t) = 0.$$

Returning to (38), and noticing that $\varepsilon(t)\nabla\Psi(y(t)) \to 0$, we obtain

$$\lim_{t\to+\infty} \nabla\Theta(y(t)) = 0.$$

Since $\nabla\Theta(y(t)) = y(t) - \text{prox}_{\mu\Phi}(y(t)) = y(t) - x(t)$, we finally obtain that $x(t) - y(t)$ converges strongly to zero as $t \longrightarrow +\infty$, which clearly implies that $x(\cdot)$ and $y(\cdot)$ converge weakly to the same limit, which is the solution of a hierarchical minimization problem. □

Example: Let us return to our model situation where

$$\Phi(x) \geq \frac{r}{2}\text{dis}^2(x,C),$$

for some $r>0$. Then $\Phi^*(x) \leq \frac{1}{2r}\|x\|^2 + \sigma_C(x)$ and

$$\Phi^*(z) - \sigma_C(z) \leq \frac{1}{2r}\|z\|^2.$$

After elementary computation, one can verify that, in this situation, $(\mathcal{H}_1)_\epsilon$ is satisfied if the following condition on $\epsilon(\cdot)$ is satisfied:

$$\int_0^{+\infty} \epsilon^2(t) < +\infty.$$

Thus, in this situation, the moderate growth condition on $\epsilon(\cdot)$, can be formulated as

$$\epsilon(\cdot) \in L^2(0,+\infty) \setminus L^1(0,+\infty).$$

3.3. **Strong convergence.** Let us now examine the strong convergence properties of the trajectories. Let us first consider the variable $y(\cdot)$. Following [6, Theorem 2.2], and equation (38), the strong convergence of $y(\cdot)$ will result from the strong monotonicity property of $\nabla\Psi = \mu\text{prox}_{\mu\Phi}$. We recall that $\nabla\Psi$ is said to be strongly monotone if there exists some $\alpha>0$ such that for any $x \in \text{dom}\nabla\Psi$, and $y \in \text{dom}\nabla\Psi$

$$\langle \nabla\Psi(x) - \nabla\Psi(y), x-y\rangle \geq \alpha\|x-y\|^2.$$

This property turns out to be equivalent to a regularity property for Φ, as stated in the following Lemma.

Lemma 3.2. *Let Φ be a convex differentiable function whose gradient is L-Lipschitz continuous for some $L>0$. Then, for any $\mu>0$ such that $\mu L<1$, the proximal mapping $prox_{\mu\Phi}$ is strongly monotone.*

Proof. Take y_i, $i = 1,2$. By definition of $\text{prox}_{\mu\Phi}(y_i)$, $\text{prox}_{\mu\Phi}(y_i) + \mu\nabla\Phi(\text{prox}_{\mu\Phi}(y_i)) = y_i$. Taking the difference of the two equations, and multiplying scalarly by $y_2 - y_1$, we obtain

$$\left\langle \text{prox}_{\mu\Phi}(y_2) - \text{prox}_{\mu\Phi}(y_1), y_2 - y_1 \right\rangle + \mu \left\langle \nabla\Phi(\text{prox}_{\mu\Phi}(y_2) - \nabla\Phi(\text{prox}_{\mu\Phi}(y_1), y_2 - y_1 \right\rangle = \|y_2 - y_1\|^2.$$

Then use the Cauchy-Schwarz inequality, the L-Lipschitz continuity of $\nabla\Phi$, and the fact that the proximal mapping is nonexpansive to obtain

$$\left\langle \text{prox}_{\mu\Phi}(y_2) - \text{prox}_{\mu\Phi}(y_1), y_2 - y_1 \right\rangle \geq (1 - \mu L)\|y_2 - y_1\|^2.$$

Conversely, one can easily establish that the strong monotonicity of the proximal mapping implies that Φ is a convex differentiable function whose gradient is Lipschitz continuous. ☐

We can now complete Theorem 3.2 as follows.

Theorem 3.3. *Let us make the assumptions of Theorem 3.2, and assume moreover that Φ is a convex differentiable function whose gradient is L-Lipschitz continuous for some $L > 0$. Then for $\mu L < 1$, we have the strong convergence property of $x(\cdot)$ and $y(\cdot)$ to the element of minimal norm of $C = \text{argmin}\Phi \neq \emptyset$.*

Proof. By Theorem 3.2 item *(iii)*, $x(t) - y(t)$ converges strongly to zero as $t \longrightarrow +\infty$. Hence we just need to prove that $y(\cdot)$ converges strongly . Since $\mu L < 1$, by Lemma 3.2, the operator $\nabla\Psi = \mu\text{prox}_{\mu\Phi}$ is strongly monotone. Thus we are in the situation examined in [6, Theorem 2.2], which gives the strong convergence property. Another equivalent approach consists in noticing that by Theorem 3.1, we have $w - \lim_{t\to+\infty} y(t) = \text{proj}_{\text{argmin}\Phi}0$ and $\Psi(y(t)) \to \Psi(\text{proj}_{\text{argmin}\Phi}0)$. From this we easily deduce that the strong convexity of Ψ implies the strong convergence of $y(\cdot)$. We recover our result by noticing that the strong convexity of Ψ is equivalent to the strong monotonicity of its gradient, i.e., of the proximal mapping. ☐

4. Other viscosity selection principles

Let us now examine the more general situation

$$\upsilon(t) \in \partial\Phi(x(t)) \tag{51a}$$

$$\lambda\dot{x}(t) + \dot{\upsilon}(t) + \upsilon(t) + \varepsilon(t)\,\partial g(x(t)) \ni 0, \tag{51b}$$

where g is a convex viscosity function.
Using the Minty transform, this system can be equivalently written as

$$x(t) = \text{prox}_{\mu\Phi}(y(t)) \tag{52a}$$

$$\dot{y}(t) + \mu\nabla\Phi_\mu(y(t)) + \mu\epsilon(t)\,\partial g(\text{prox}_{\mu\Phi}(y(t))) \ni 0. \tag{52b}$$

In order to recover exactly the Tikhonov approximation, we look for some g such that, for all $y \in \mathcal{H}$

$$\partial(\mu g)(\text{prox}_{\mu\Phi}(y)) = y.$$

Equivalently

$$(I + \mu\partial\Phi)^{-1} = (\partial(\mu g))^{-1}.$$

We obtain

$$I + \mu\partial\Phi = \partial(\mu g),$$

that is, for all $y \in \mathcal{H}$

$$\mu g(y) = \frac{1}{2}\|y\|^2 + \mu\Phi(y).$$

Thus, by taking $g(y) = \frac{1}{2\mu}\|y\|^2 + \Phi(y)$, and $\Theta(y) = \mu\Phi_\mu(y)$, equation (52b) can be equivalently written as

$$\dot{y}(t) + \nabla\Theta(y(t)) + \epsilon(t)\,y(t) = 0. \tag{53}$$

Equation (53) is a particular case of the (SDC) system (steepest descent with control)

$$\text{(SDC)} \qquad \dot{y}(t) + \partial\Theta(y(t)) + \epsilon(t)\,y(t) = 0.$$

Concerning the case $\int_0^{+\infty} \epsilon(t)\,dt = +\infty$, the first general convergence result was in [19] (based on previous work by [13]), and also requires $\epsilon(\cdot)$ to be nonincreasing, and converges to zero for $t \longrightarrow +\infty$. Under these conditions, each trajectory of (53) converges strongly to the point of minimal norm in $C = \arg\min\Theta = \arg\min\Phi$. In [15], it is proved that the convergence result still holds without assuming $\epsilon(\cdot)$ to be nonincreasing.

When $g(\xi) = \frac{1}{2\mu}\|\xi\|^2 + \Phi(\xi)$, the dynamical system (51a)-(51b) becomes

$$\upsilon(t) \in \partial\Phi(x(t)) \tag{54a}$$

$$\lambda \dot{x}(t) + \dot{\upsilon}(t) + \upsilon(t) + \varepsilon(t)\left(\frac{1}{\mu}x(t) + \upsilon(t)\right) = 0. \tag{54b}$$

Equivalently

$$\upsilon(t) \in \partial\Phi(x(t)) \tag{55a}$$

$$\dot{x}(t) + \mu\dot{\upsilon}(t) + \mu(1+\varepsilon(t))\upsilon(t) + \varepsilon(t)x(t) = 0. \tag{55b}$$

As in the preceding section, by application of the Cauchy-Lipschitz theorem (recall that Θ is differentiable, and its gradient is Lipschitz continuous), we can show that (53) admits a strong global solution $y(\cdot)$. Then, $(x(\cdot), \upsilon(\cdot))$ with $x(\cdot) = \mathrm{prox}_{\mu\Phi}(y(\cdot))$ and $\upsilon(\cdot) = \nabla\Phi_\mu(y(\cdot))$ is a strong solution of (55a)-(55b).

Let us summarize our result in the following theorem.

Theorem 4.1. *Let $\Phi : \mathcal{H} \longrightarrow \mathbb{R}\cup\{+\infty\}$ be a closed, convex, proper function, such that $C = \arg\min\Phi \neq \emptyset$. Let us assume that*

(i) $\lim_{t\longrightarrow+\infty} \epsilon(t) = 0$,

(ii) $\int_0^{+\infty} \epsilon(t)\,dt = +\infty$.

Then, for any trajectory $(x(\cdot), \upsilon(\cdot)) : [0, +\infty[\longrightarrow \mathcal{H}\times\mathcal{H}$ solution of (55a)- (55b), with $y(t) = x(t) + \mu\upsilon(t)$ the following strong convergence propery holds:

$$s - \lim_{t\longrightarrow+\infty} x(t) = s - \lim_{t\longrightarrow+\infty} y(t) = \mathop{Proj}_{\arg\min\Phi} 0.$$

Proof. We are in the situation examined in [15, Theorem 2], which gives the strong convergence of each trajectory $y(\cdot)$ of (53) towards the point of minimal norm in $C = \arg\min\Phi$. In order to pass from the convergence of $y(\cdot)$ to the convergence of $x(\cdot)$ we use the relation (52a)

$$x(t) = \mathrm{prox}_{\mu\Phi}(y(t))$$

that links the two variables. From the continuity property of the proximal mapping for the strong topology (indeed, it is a nonexpansive mapping), we deduce that each trajectory $x(\cdot)$ of (55a)- (55b) converges strongly to $\mathrm{prox}_{\mu\Phi}\left(\mathop{\mathrm{Proj}}_{\arg\min\Phi} 0\right) = \mathop{\mathrm{Proj}}_{\arg\min\Phi} 0$. To obtain this last equality, we use the fact that the set C of minimizers of Φ is invariant by the proximal mapping $\mathrm{prox}_{\mu\Phi}$, i.e.,

$$\mathrm{prox}_{\mu\Phi}(y) = y \quad \text{for all } y \in C = \mathrm{argmin}\Phi.$$

□

5. Perspective

Let us list some interesting questions to be examined in the future:

(1) Examine the discrete, algorithmical version, and the corresponding asymptotic selection property for the forward-backward algorithm.
(2) Study the case where $\lambda(t)$ depends on t in an open-loop form, as in [9].
(3) Study the case where the Levenberg-Marquart regularization term is given in a closed-loop form, $\lambda(t) = \alpha(\|\dot{x}(t)\|)$ as in [8].
(4) Examine these questions for the related dynamical systems which have been considered in [1].

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*, (2014) http://dx.doi.org/10.1080/02331934.2014.971412.

[2] B. Abbas, H. Attouch, B. F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, **161** (2014), No. 2, pp. 331-360.

[3] F. Alvarez, A. Cabot, Asymptotic selection of viscosity equilibria of semilinear evolution equations by the introduction of a slowly vanishing term, *Discrete and Continuous Dynamical Systems*, **15** (2006), pp. 921-938.

[4] H. Attouch, Viscosity solutions of minimization problems, *SIAM J. Optimization*, **6** (1996), pp. 769-806.

[5] H. Attouch, G. Buttazzo, G. Michaille, Variational analysis in Sobolev and BV spaces. Applications to PDE's and optimization, *MPS/SIAM Series on Optimization*, **6**, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, Second edition, 2014, 793 pages.

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

[7] H. Attouch, J. Peypouquet, P. Redont, Backward-forward algorithms for structured monotone inclusions in Hilbert spaces, to appear.

[8] H. Attouch, P. Redont, B. F. Svaiter, Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces, *J. Optim. Theory Appl.*, 157 (2013), No. 3, pp. 624–650.

[9] H. Attouch, B.F. Svaiter, A continuous dynamical Newton-like approach to solving monotone inclusions, *SIAM J. Control Optim.* **49** (2011), pp. 574–598.

[10] J. B. Baillon, R. Cominetti, A convergence result for nonautonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming, *Journal of Functional Analysis*, **187** (2001), pp. 263–273.

[11] H.H. Bauschke, P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York (2011).

[12] H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North-Holland/Elsevier, New-York, 1973.

[13] F.E. Browder, Nonlinear Operators and Nonlinear Equations of Evolution in Banach Spaces, Proc. Sympos. Pure Math., vol. 18 (part 2), Amer. Math. Soc., Providence, RI, 1976.

[14] A. Cabot, The steepest descent dynamical system with control. Applications to constrained minimization, *ESAIM: Control, Optimization and Calculus of Variations*, **10** (2004), pp. 243-258.

[15] R. Cominetti, J. Peypouquet, S. Sorin, Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization, J. Differential Equations, **245** (2008), pp. 3753–3763.

[16] H. Furuya, K. Miyashiba, N. Kenmochi, Asymptotic behavior of solutions to a class of nonlinear evolution equations, *Journal of Differential Equations*, **62** (1986), pp. 73–94.

[17] A. Haraux, Systèmes dynamiques dissipatifs et applications. RMA 17, Masson, Paris, (1991).

[18] S. Hirstoaga, Approximation et résolution de problèmes d'équilibre, de point fixe et d'inclusion monotone, PhD thesis, UPMC Paris 6, (2006).

[19] S. Reich, Nonlinear evolution equations and nonlinear ergodic theorems, Nonlinear Anal. 1 (1976) 319–330.

[20] R.T Rockafellar, R. J-B. Wets, Variational Analysis, Grundlehren der mathematischen Wissenschaften **317**, Springer-Verlag, (1998).

[21] E.D. Sontag, Mathematical Control Theory, second edition. Springer-Verlag, New-York (1998).