



HAL
open science

A continuous dynamical Newton-like approach to solving monotone inclusions

Hedy Attouch, Benar Fux Svaiter

► **To cite this version:**

Hedy Attouch, Benar Fux Svaiter. A continuous dynamical Newton-like approach to solving monotone inclusions. *SIAM Journal on Control and Optimization*, 2011, 49 (3), pp.574-598. 10.1137/100784114. hal-00803137

HAL Id: hal-00803137

<https://hal.science/hal-00803137v1>

Submitted on 21 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A CONTINUOUS DYNAMICAL NEWTON-LIKE APPROACH TO SOLVING MONOTONE INCLUSIONS

H. ATTOUCH*[†] B. F. SVAITER [‡] [§]

March 8, 2011

Abstract

We introduce non-autonomous continuous dynamical systems which are linked to the Newton and Levenberg-Marquardt methods. They aim at solving inclusions governed by maximal monotone operators in Hilbert spaces. Relying on the Minty representation of maximal monotone operators as lipschitzian manifolds, we show that these dynamics can be formulated as first-order in time differential systems, which are relevant to the Cauchy-Lipschitz theorem. By using Lyapunov methods, we prove that their trajectories converge weakly to equilibria. Time discretization of these dynamics gives algorithms providing new insight into Newton's method for solving monotone inclusions.

2010 Mathematics Subject Classification: 34G25, 47J25, 47J30, 47J35, 49M15, 49M37, 65K15, 90C25, 90C53.

Keywords: Maximal monotone operators, Newton-like algorithms, Levenberg-Marquardt algorithms, non-autonomous differential equations, absolutely continuous trajectories, dissipative dynamical systems, Lyapunov analysis, weak asymptotic convergence, numerical convex optimization.

1 Introduction

Let H be a real Hilbert space and $T : H \rightrightarrows H$ be a maximal monotone operator. The space H is endowed with the scalar product $\langle \cdot, \cdot \rangle$, with $\|x\|^2 = \langle x, x \rangle$ for any $x \in H$. Our objective is to design continuous and discrete Newton-like dynamics attached to solving the equation

$$\text{find } x \in H \text{ such that } 0 \in Tx. \tag{1}$$

*I3M UMR CNRS 5149, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier, France (attouch@math.univ-montp2.fr)

[†]Partially supported by ANR-08-BLAN-0294-03.

[‡]IMPA, Estrada Dona Castorina 110, 22460 - 320 Rio de Janeiro, Brazil (benar@impa.br)

[§]Partially supported by CNPq grants 480101/2008-6, 303583/2008-8, FAPERJ grant E-26/102.821/2008 and PRONEX-Optimization

When T is a C^1 operator with derivative T' , the classical Newton method generates sequences $(x_k)_{k \in \mathbb{N}}$ verifying

$$T(x_k) + T'(x_k)(x_{k+1} - x_k) = 0. \quad (2)$$

When the current iterate is far from the solution it is convenient to introduce a step-size Δt_k , and consider

$$T(x_k) + T'(x_k) \left(\frac{x_{k+1} - x_k}{\Delta t_k} \right) = 0. \quad (3)$$

Unless restrictive assumptions on T are made, this is not a well-posed equation. The Levenberg-Marquardt method consists in solving the regularized problem

$$T(x_k) + \left(\lambda_k I + T'(x_k) \right) \left(\frac{x_{k+1} - x_k}{\Delta t_k} \right) = 0, \quad (4)$$

where I is the identity operator on H , and $(\lambda_k)_{k \in \mathbb{N}}$ is a sequence of positive real numbers. When T is the gradient of a convex potential, this algorithm can be viewed as an interpolation between Newton's method and the gradient method (when λ_k is close to zero the algorithm is close to Newton's method, for λ_k large it is close to a gradient method). This algorithm has a natural interpretation as a time discretized version of the continuous dynamical system

$$\lambda(t)\dot{x}(t) + T'(x(t))\dot{x}(t) + T(x(t)) = 0, \quad (5)$$

where $\dot{x}(t) = \frac{dx}{dt}(t)$ is the derivative at time t of the mapping $t \mapsto x(t)$ (we use the two notations, indifferently), and $t \mapsto \lambda(t)$ is a positive real-valued function (we shall make precise the assumptions on $\lambda(\cdot)$ very soon).

By using the classical derivation rule for the composition of smooth mappings $\frac{d}{dt}T(x(t)) = T'(x(t))\dot{x}(t)$, we can rewrite (5) as follows: find (x, v) solution of the differential-algebraic system

$$\begin{cases} v(t) = T(x(t)), \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0. \end{cases} \quad (6)$$

Let us now consider a general maximal monotone operator $T : H \rightrightarrows H$ (one may consult Brézis [8], Zeidler [29] for a detailed presentation of the theory of maximal monotone operators in Hilbert spaces). Let us notice that the operator T is possibly multivalued, and its domain $\text{dom}T \subset H$ may be a proper subset of H . The corresponding differential-algebraic inclusion system,

$$\begin{cases} v(t) \in T(x(t)), \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0, \end{cases} \quad (7)$$

involves an inclusion instead of an equality in the first equation. In order to solve this system we are going to reformulate it with the help of the Minty representation of maximal monotone operators, see [20]. This representation makes use of $J_\mu^T =$

$(I + \mu T)^{-1}$ the resolvent of index $\mu > 0$ of T , and of $T_\mu = \frac{1}{\mu}(I - J_\mu^T)$ the Yosida approximation of index $\mu > 0$ of T . For any $t \in [0, +\infty)$ set $\mu(t) = \frac{1}{\lambda(t)}$, and introduce the new unknown function $z : [0, +\infty) \rightarrow H$ which is defined by

$$z(t) = x(t) + \mu(t)v(t). \quad (8)$$

Let us rewrite (7) with the help of (x, z) . One first obtains (see section 2)

$$x(t) = J_{\mu(t)}^T(z(t)) \quad (9)$$

$$v(t) = T_{\mu(t)}(z(t)). \quad (10)$$

In our context, this is the Minty representation of maximal monotone operators. This representation fits well our study. Indeed, the second equation of (7) can be reformulated as a classical differential equation with respect to $z(\cdot)$ (see section 2), which gives

$$x(t) = J_{\mu(t)}^T(z(t)) \quad (11)$$

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0. \quad (12)$$

As a nice feature of system ((11)-(12)), let us stress the fact that the operators $J_\mu^T : H \rightarrow H$ and $T_\mu : H \rightarrow H$ are Lipschitz continuous, which makes this system relevant to the Cauchy-Lipschitz theorem.

All along the paper, we shall pay particular attention to the case $\lambda(t) \rightarrow 0$ as $t \rightarrow +\infty$ (equivalently $\mu(t) \rightarrow +\infty$ as $t \rightarrow +\infty$). In that case, one may expect obtaining rates of convergence close to Newton's method.

The paper is organized as follows: In section 2, assuming $\lambda(\cdot)$ to be locally absolutely continuous, for any given Cauchy data $x(0) = x_0$, $v(0) = v_0 \in T(x_0)$, we prove the existence and uniqueness of a strong global solution to system (7). In section 3, we study the asymptotic behavior of the trajectories of this system. Assuming that $\lambda(t)$ does not converge too rapidly to zero as $t \rightarrow +\infty$ (with, roughly speaking, as a critical size, $\lambda(t) = e^{-t}$), we prove that, for each trajectory $(x(t), v(t))$ of system (7), $x(t)$ converges weakly to a zero of T , and $v(t)$ converges strongly to zero. In the autonomous case $\lambda(t) \equiv \lambda_0$, we make the link with some classical results concerning semi-groups of contractions generated by maximal monotone operators. In section 4, we specialize our study to the subdifferential case $T = \partial f$, with f convex lower semicontinuous, showing the optimizing properties of the trajectories. In section 5, we examine the case $\lambda(t) = \lambda_0 e^{-t}$, which is the closest situation to Newton's dynamic allowed by our study. In section 6, we give some elementary examples aiming at illustrating these dynamics. In section 7, we finally give an application to numerical convex optimization.

Our approach, which can be traced back to the Levenberg-Marquardt regularization procedure, seems original. In the case of convex optimization, it bears interesting connections with the second-order continuous dynamic approach developed by Alvarez, Attouch, Bolte, and Redont in [3], see also [4], [6] (Newton's dynamic is regularized by adding an inertial term, and a viscous damping term, which provides a second-order dissipative dynamical system with Hessian-driven damping.) Another interesting regularization method (based on the regularization of the objective

function) has been developed by Alvarez and Pérez in [2]. Among the rich literature concerning Newton's method and its links with continuous dynamical systems and optimization, let us also mention Chen, Nashed and Qi [12], and Ramm [24].

As a rather striking feature of our approach, we can develop a Newton-like method in a fairly general nonsmooth multivalued setting, namely for solving inclusions governed by maximal monotone operators in Hilbert spaces. This offers interesting perspectives concerning applications ranging from optimal control to variational inequalities and PDE's.

2 Existence and uniqueness of global solutions

We consider the Cauchy problem for the differential inclusion system

$$v(t) \in T(x(t)), \quad (13)$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0, \quad (14)$$

$$x(0) = x_0, v(0) = v_0 \in T(x_0). \quad (15)$$

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 maximal monotone operators. Finally, we shall prove the existence and uniqueness of a strong solution to system ((13)-(14)-(15)) by applying the Cauchy-Lipschitz theorem to this equivalent formulation.

2.1 Definition of strong solutions

Let us first recall some notions concerning vector-valued functions of a real variable (see Appendix of [8]).

Definition 2.1. *Given $b \in \mathbb{R}^+$, a function $f : [0, b] \rightarrow H$ is said to be absolutely continuous if one the following equivalent properties holds :*

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

$$f(t) = f(0) + \int_0^t g(s)ds \quad \text{for all } t \in [0, b] ;$$

ii) f is continuous and its distributional derivative belongs to the Lebesgue space $L^1([0, b]; H)$.

iii) for every $\epsilon > 0$, there exists some $\eta > 0$ such that for any finite family of intervals $I_k = (a_k, b_k)$

$$I_k \cap I_j = \emptyset \text{ for } i \neq j \quad \text{and} \quad \sum |b_k - a_k| \leq \eta \implies \sum \|f(b_k) - f(a_k)\| \leq \epsilon.$$

Moreover, an absolutely continuous function is almost everywhere differentiable, its derivative almost everywhere coincide with its distributional derivative, and one can recover the function from its derivative $f' = g$ by integration formula *i*). Note that the crucial property which makes the theory of absolutely continuous functions, as described above, work with vector-valued functions, is the fact that the image space H is reflexive, which is the case here (H is a Hilbert space).

Definition 2.2. We say that a pair $(x(\cdot), v(\cdot))$ is a strong global solution of ((13)-(14)-(15)) if the following properties *i*), *ii*), *iii*) and *iv*) are satisfied:

$$i) x(\cdot), v(\cdot) : [0, +\infty) \rightarrow H \text{ are continuous,} \quad (16)$$

$$\text{and absolutely continuous on each bounded interval } [0, b], \quad 0 < b < +\infty; \quad (17)$$

$$ii) v(t) \in T(x(t)) \text{ for all } t \in [0, +\infty); \quad (18)$$

$$iii) \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0 \text{ for almost all } t \in [0, +\infty); \quad (19)$$

$$iv) x(0) = x_0, v(0) = v_0. \quad (20)$$

This last condition makes sense because of the continuity property of $x(\cdot)$ and $v(\cdot)$. Let us now make our standing assumption on function $\lambda(\cdot)$:

$$\lambda : [0, +\infty) \rightarrow (0, +\infty) \text{ is continuous,} \quad (21)$$

$$\text{and absolutely continuous on each interval } [0, b], \quad 0 < b < +\infty. \quad (22)$$

Hence $\dot{\lambda}(t)$ exists for almost every $t > 0$, and $\dot{\lambda}(\cdot)$ is Lebesgue integrable on each bounded interval $[0, b]$. We stress the fact that we assume $\lambda(t) > 0$, for any $t \geq 0$. By continuity of $\lambda(\cdot)$, this implies that, for any $b > 0$, there exists some positive finite lower and upper bounds for $\lambda(\cdot)$ on $[0, b]$, i.e., for any $t \in [0, b]$

$$0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty. \quad (23)$$

This fact will be of importance for proving existence of strong solutions.

2.2 Equivalent formulation involving a classical differential equation

In order to solve system ((13)-(14)-(15)) we use Minty's device. Let us rewrite the maximal monotone inclusion (13) by using the following equivalences: for any $t \in [0, +\infty)$

$$v(t) \in T(x(t)) \Leftrightarrow \quad (24)$$

$$x(t) + \frac{1}{\lambda(t)}v(t) \in x(t) + \frac{1}{\lambda(t)}T(x(t)) \Leftrightarrow \quad (25)$$

$$x(t) = \left(I + \frac{1}{\lambda(t)}T \right)^{-1} \left(x(t) + \frac{1}{\lambda(t)}v(t) \right). \quad (26)$$

Set $\mu(t) = \frac{1}{\lambda(t)}$. Let us introduce the new unknown function $z : [0, +\infty) \rightarrow H$ which is defined for any $t \in [0, +\infty)$ by

$$z(t) = x(t) + \frac{1}{\lambda(t)}v(t) = x(t) + \mu(t)v(t) \quad (27)$$

and rewrite system ((13)-(14)-(15)) with the help of (x, z) . From (26) and (27)

$$\begin{aligned} x(t) &= (I + \mu(t)T)^{-1}(z(t)) \\ v(t) &= \frac{1}{\mu(t)} (z(t) - (I + \mu(t)T)^{-1}(z(t))). \end{aligned}$$

Equivalently,

$$x(t) = J_{\mu(t)}^T(z(t)) \quad (28)$$

$$v(t) = T_{\mu(t)}(z(t)), \quad (29)$$

where $J_{\mu}^T = (I + \mu T)^{-1}$ and $T_{\mu} = \frac{1}{\mu} (I - J_{\mu}^T)$ are respectively the resolvent and the Yosida approximation of index $\mu > 0$ of T . Indeed, this is Minty's representation of maximal monotone operators, see [20]. In a finite dimensional setting, this technique has been developed by Rockafellar in [26]: he shows that a maximal monotone operator can be represented as a lipschitzian manifold, which allows him to define second-order derivatives of convex lower semicontinuous functions.

Let us show how (14) can be reformulated as a classical differential equation with respect to $z(\cdot)$. First, let us rewrite (14) as

$$\dot{x}(t) + \mu(t)\dot{v}(t) + \mu(t)v(t) = 0. \quad (30)$$

Differentiating (27) and using (30) we obtain

$$\dot{z}(t) = \dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) \quad (31)$$

$$= -\mu(t)v(t) + \dot{\mu}(t)v(t). \quad (32)$$

From (29) and (32) we deduce that

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0. \quad (33)$$

Finally, the equivalent (x, z) system can be written as

$$x(t) = J_{\mu(t)}^T(z(t)) \quad (34)$$

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0. \quad (35)$$

As a nice feature of system ((34)-(35)), let us stress the fact that the operators $J_{\mu}^T : H \rightarrow H$ and $T_{\mu} : H \rightarrow H$ are everywhere defined and Lipschitz continuous, which makes this system relevant to the Cauchy-Lipschitz theorem.

2.3 Global existence and uniqueness results

System ((34)-(35)) involves time-dependent operators $J_{\mu(t)}^T$ and $T_{\mu(t)}$. In order to establish existence results for the corresponding evolution equations, let us study the regularity properties of the mappings $\mu \mapsto J_{\mu}^T x$ and $\mu \mapsto T_{\mu} x$.

Proposition 2.3. *For any $\lambda > 0$, $\mu > 0$ and any $x \in H$, the following properties hold:*

$$i) J_{\lambda}^T x = J_{\mu}^T \left(\frac{\mu}{\lambda} x + \left(1 - \frac{\mu}{\lambda}\right) J_{\lambda}^T x \right); \quad (36)$$

$$ii) \|J_{\lambda}^T x - J_{\mu}^T x\| \leq |\lambda - \mu| \|T_{\lambda} x\|. \quad (37)$$

As a consequence, for any $x \in H$ and any $0 < \delta < \Lambda < +\infty$, the function $\mu \mapsto J_{\mu}^T x$ is Lipschitz continuous on $[\delta, \Lambda]$. More precisely, for any λ, μ belonging to $[\delta, \Lambda]$

$$\|J_{\lambda}^T x - J_{\mu}^T x\| \leq |\lambda - \mu| \|T_{\delta} x\|. \quad (38)$$

Proof. i) Equality (36) is known as the resolvent equation. It is a classical result, see for example [8].

ii) For any $\lambda > 0$, $\mu > 0$ and any $x \in H$, by using successively the resolvent equation and the contraction property of the resolvents, we have

$$\begin{aligned} \|J_\lambda^T x - J_\mu^T x\| &= \|J_\mu^T \left(\frac{\mu}{\lambda} x + \left(1 - \frac{\mu}{\lambda}\right) J_\lambda^T x \right) - J_\mu^T x\| \\ &\leq \left\| \left(1 - \frac{\mu}{\lambda}\right) (x - J_\lambda^T x) \right\| \\ &\leq |\lambda - \mu| \|T_\lambda x\|. \end{aligned}$$

Using that $\lambda \mapsto \|T_\lambda x\|$ is decreasing, see ([8], Proposition 2.6), we obtain (38). \square

Theorem 2.4. *Let $\lambda : [0, +\infty) \rightarrow (0, +\infty)$ be a continuous function which is absolutely continuous on each bounded interval $[0, b]$, $b > 0$. Set $\mu(t) = \frac{1}{\lambda(t)}$. Let $(x_0, v_0) \in H \times H$ be such that $v_0 \in T(x_0)$. Then the following properties hold:*

1. *there exists a unique strong global solution $(x(\cdot), v(\cdot)) : [0, +\infty) \rightarrow H \times H$ of the Cauchy problem ((13)-(14)-(15));*
2. *the solution pair $(x(\cdot), v(\cdot))$ of ((13)-(14)-(15)) can be represented as: for any $t \in [0, +\infty)$,*

$$x(t) = J_{\mu(t)}^T(z(t)) \tag{39}$$

$$v(t) = T_{\mu(t)}(z(t)), \tag{40}$$

where $z(\cdot) : [0, +\infty) \rightarrow H$ is the unique strong solution of the Cauchy problem

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0, \tag{41}$$

$$z(0) = x_0 + \mu(0)v_0. \tag{42}$$

Proof. 1) Let us first prove existence and uniqueness of a strong global solution of the Cauchy problem ((41)-(42)). By the definition of $\mu(t) = \frac{1}{\lambda(t)}$, we have $\dot{\mu}(t) = -\frac{\dot{\lambda}(t)}{\lambda^2(t)}$. Thus, (41) can be written as

$$\dot{z}(t) + \left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)}\right) \frac{1}{\lambda(t)} T_{\frac{1}{\lambda(t)}}(z(t)) = 0, \tag{43}$$

which is equivalent to

$$\dot{z}(t) = F(t, z(t)) \tag{44}$$

with

$$F(t, z) = \theta(t)G(t, z), \tag{45}$$

$$\theta(t) = -\left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)}\right), \tag{46}$$

$$G(t, z) = \frac{1}{\lambda(t)} T_{\frac{1}{\lambda(t)}}(z). \tag{47}$$

In order to apply the Cauchy-Lipschitz theorem to (44), let us first examine the Lipschitz continuity properties of F .

a) For any $t \geq 0$, $G(t, \cdot) : H \rightarrow H$ is a contraction, i.e., for any $z_i \in H, i = 1, 2$

$$\|G(t, z_2) - G(t, z_1)\| \leq \|z_2 - z_1\|, \quad (48)$$

and, as a consequence,

$$\|F(t, z_2) - F(t, z_1)\| \leq |\theta(t)| \|z_2 - z_1\|. \quad (49)$$

By the definition of $\theta(\cdot)$

$$|\theta(t)| \leq 1 + \left| \frac{\dot{\lambda}(t)}{\lambda(t)} \right|. \quad (50)$$

Since $\dot{\lambda}(\cdot)$ is locally integrable and $\lambda(\cdot)$ is bounded away from zero on any bounded interval, (50) shows that

$$\theta(\cdot) \in L^1([0, b]) \text{ for any } 0 < b < +\infty. \quad (51)$$

b) Let us show that

$$\forall z \in H, \forall b > 0, F(\cdot, z) \in L^1([0, b]). \quad (52)$$

By (23), for any $t \in [0, b]$, we have $0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty$. Returning to the definition (45) of F , we deduce that

$$\|F(t, z)\| \leq \left(1 + \frac{|\dot{\lambda}(t)|}{\lambda_{b, \min}} \right) \frac{1}{\lambda_{b, \min}} \|T_{\frac{1}{\lambda_{b, \max}}} z\|. \quad (53)$$

Using again the local integrability of $\dot{\lambda}(\cdot)$ we obtain (52).

From properties (49), (51), and (52), we deduce the existence and uniqueness of a strong global solution of differential equation (44), with given Cauchy data. To that end, we use the version of the Cauchy-Lipschitz theorem relying on the integrability of $t \mapsto F(t, x)$, and involving absolutely continuous trajectories, see for example ([17], Proposition 6.2.1.), ([28], Theorem 54).

2) Let us now return to the initial problem. Given $z(\cdot) : [0, +\infty) \rightarrow H$ which is the unique strong solution of Cauchy problem ((41)-(42)), let us define $x(\cdot), v(\cdot) : [0, +\infty) \rightarrow H$ by

$$x(t) = J_{\mu(t)}^T(z(t)), \quad v(t) = T_{\mu(t)}(z(t)). \quad (54)$$

a) Let us show that $x(\cdot), v(\cdot)$ are absolutely continuous on each bounded interval, and satisfy system ((13)-(14)-(15)). Let us give arbitrary $z_1 \in H, z_2 \in H$ and $\mu_1 > 0, \mu_2 > 0$. Combining Proposition 2.3 and the contraction property of the resolvents, we obtain

$$\|J_{\mu_2}^T(z_2) - J_{\mu_1}^T(z_1)\| \leq \|J_{\mu_2}^T(z_2) - J_{\mu_2}^T(z_1)\| + \|J_{\mu_2}^T(z_1) - J_{\mu_1}^T(z_1)\| \quad (55)$$

$$\leq \|z_2 - z_1\| + |\mu_2 - \mu_1| \|T_{\mu_1} z_1\|. \quad (56)$$

Assuming that $s, t \in [0, b]$, by taking $z_1 = z(s)$, $z_2 = z(t)$ and $\mu_1 = \mu(s)$, $\mu_2 = \mu(t)$ in (56), and with the same notations as before (for any $t \in [0, b]$, $0 < \lambda_{b, \min} \leq \lambda(t) \leq \lambda_{b, \max} < +\infty$), setting more briefly $\Lambda = \lambda_{b, \max}$, we obtain

$$\|J_{\mu(t)}^T(z(t)) - J_{\mu(s)}^T(z(s))\| \leq \|z(t) - z(s)\| + |\mu(t) - \mu(s)| \|T_{\mu(t)} z(t)\| \quad (57)$$

$$\leq \|z(t) - z(s)\| + |\mu(t) - \mu(s)| \|T_{\frac{1}{\Lambda}} z(t)\|. \quad (58)$$

Noticing that $\|T_{\frac{1}{\Lambda}} z(t)\| \leq \|T_{\frac{1}{\Lambda}}(0)\| + \Lambda \|z(t)\|$ remains bounded on $[0, b]$, and owing to the absolute continuity property of $z(\cdot)$ and $\mu(\cdot)$, we deduce that $x(t) = J_{\mu(t)}^T(z(t))$ is absolutely continuous on $[0, b]$ for any $b > 0$. The same property holds true for $v(t) = T_{\mu(t)} z(t) = \lambda(t)(z(t) - x(t))$, because $\lambda(\cdot)$ is absolutely continuous on $[0, b]$ for any $b > 0$, and the product of two absolutely continuous functions is still absolutely continuous (see [9], Corollaire VIII.9). Indeed this last property is a straight consequence of Definition (2.1; *iii*) of absolute continuity.

Moreover, for any $t \in [0, +\infty)$

$$v(t) \in T(x(t)), \quad z(t) = x(t) + \mu(t) v(t).$$

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

$$\dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) = \dot{z}(t).$$

On the other hand, owing to $v(t) = T_{\mu(t)} z(t)$, (41) can be equivalently written as

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t)) v(t) = 0.$$

Combining the two above equations we obtain

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

As $\mu(t) = \lambda(t)^{-1}$, we conclude that $(x(\cdot), v(\cdot))$ is a solution of system ((13)-(14)-(15)).

Regarding the initial condition, let us observe that

$$z(0) = x_0 + \mu(0)v_0, \quad (59)$$

$$= x(0) + \mu(0)v(0), \quad (60)$$

with $v_0 \in T(x_0)$ and $v(0) \in T(x(0))$. Hence

$$x(0) = x_0 = (I + \mu(0)T)^{-1}(x_0 + \mu(0)v_0).$$

Returning to (59), after simplification, we obtain $v(0) = v_0$.

b) Let us now prove uniqueness. Suppose that

$$(x(\cdot), v(\cdot)) : [0, +\infty) \rightarrow H \times H$$

is a solution pair of ((13)-(14)-(15)). Defining $\mu(t) = \lambda(t)^{-1}$ and

$$z(t) = x(t) + \mu(t)v(t) \quad (61)$$

we conclude that $z(\cdot)$ is absolutely continuous (we use again that the product of two absolutely continuous functions is still absolutely continuous), $z_0 = x_0 + \mu v_0$, and for any $t \in [0, +\infty)$,

$$x(t) = (I + \mu(t)T)^{-1}(z(t)), \quad v(t) = T_{\mu(t)}(z(t)). \quad (62)$$

Differentiating (61) almost everywhere (the usual derivation rule for the product of two functions holds true), and using (14), we conclude that for almost all $t \in [0, +\infty)$

$$\begin{aligned} \dot{z}(t) &= \dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) \\ &= -\mu(t)(\dot{v}(t) + v(t)) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) \\ &= (-\mu(t) + \dot{\mu}(t))v(t). \end{aligned}$$

Since $v(t) = T_{\mu(t)}(z(t))$, we finally obtain

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0.$$

Moreover

$$z_0 = x_0 + \mu v_0.$$

Arguing as before, by the Cauchy-Lipschitz theorem, $z(\cdot)$ is uniquely determined and locally absolutely continuous. Thus, by (62), $x(\cdot)$ and $v(\cdot)$ are uniquely determined. \square

Remark 2.5. Assuming that $\lambda(\cdot)$ is Lipschitz continuous on bounded sets, one can easily derive from equation (53) that $z(\cdot)$ is also Lipschitz continuous on bounded sets, and by (57) the same holds true for $x(\cdot)$ and $v(\cdot)$.

3 Asymptotic analysis and convergence properties

In this section, we study the asymptotic behavior, as $t \rightarrow +\infty$, of the trajectories of Newton-like differential inclusion system ((13)-(14)). Let us recall our standing assumption, namely $\lambda : [0, +\infty) \rightarrow (0, +\infty)$ is continuous, and absolutely continuous on each bounded interval. By Theorem 2.4, for any given Cauchy data $v_0 \in T(x_0)$, this property guarantees the existence and uniqueness of a strong global solution (see Definition 2.2) of system ((13)-(14)-(15)). From now on in this section, $(x(\cdot), v(\cdot)) : [0, +\infty) \rightarrow H \times H$ is the strong global solution of ((13)-(14)-(15)).

3.1 Properties of trajectories

Let us establish some properties of the $(x(\cdot), v(\cdot))$ trajectory which will be useful for the study of its asymptotic behavior.

Proposition 3.1. *For almost all $t \in [0, +\infty)$, the following properties hold:*

$$\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0; \quad (63)$$

$$\langle \dot{x}(t), v(t) \rangle = - [\lambda(t)\|\dot{x}(t)\|^2 + \langle \dot{x}(t), \dot{v}(t) \rangle] \leq -\lambda(t)\|\dot{x}(t)\|^2 \leq 0; \quad (64)$$

$$\langle v(t), \dot{v}(t) \rangle = - [\|\dot{v}(t)\|^2 + \lambda(t)\langle \dot{x}(t), \dot{v}(t) \rangle] \leq -\|\dot{v}(t)\|^2 \leq 0; \quad (65)$$

$$\lambda(t)^2\|\dot{x}(t)\|^2 + \|\dot{v}(t)\|^2 \leq \|v(t)\|^2. \quad (66)$$

Proof. For almost all $t \in [0, +\infty)$, $\dot{x}(t)$ and $\dot{v}(t)$ are well defined, thus

$$\langle \dot{x}(t), \dot{v}(t) \rangle = \lim_{h \rightarrow 0} \frac{1}{h^2} \langle x(t+h) - x(t), v(t+h) - v(t) \rangle.$$

By (13), we have $v(t) \in T(x(t))$. Since $T : H \rightrightarrows H$ is monotone

$$\langle x(t+h) - x(t), v(t+h) - v(t) \rangle \geq 0.$$

Dividing by h^2 and passing to the limit preserves the inequality, which yields (63). Let us now use (14)

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0.$$

Equations (64), (65) follow by taking the inner product of both sides of (14) by $\dot{x}(t)$ and $\dot{v}(t)$ respectively, using the positivity of $\lambda(t)$, and (63). In order to obtain the last inequality, let us rewrite (14) as $\lambda(t)\dot{x}(t) + \dot{v}(t) = -v(t)$. By taking the square norm of these two quantities, using (63), and the positivity of $\lambda(t)$, we obtain (66). \square

The following results are direct consequences of Proposition 3.1.

Corollary 3.2. *The following properties hold:*

1. $t \mapsto \|v(t)\|$ is a decreasing function from $[0, +\infty)$ into $[0, +\infty)$;
2. $t \mapsto v(t)$ is Lipschitz continuous on $[0, +\infty)$ with constant $\|v_0\|$;
3. for any $0 < b < +\infty$, $t \mapsto x(t)$ is Lipschitz continuous on $[0, b]$, with constant

$$\frac{\|v_0\|}{\inf_{t \in [0, b]} \lambda(t)}.$$

Moreover, if $\lambda(\cdot)$ is bounded away from 0, then $t \mapsto x(t)$ is Lipschitz continuous on $[0, +\infty)$.

Proof. By (65), for almost all $t \in [0, +\infty)$,

$$\frac{d}{dt} \frac{1}{2} \|v(t)\|^2 = \langle \dot{v}(t), v(t) \rangle \leq -\|\dot{v}(t)\|^2 \leq 0.$$

Therefore, $t \mapsto \|v(t)\|$ is a decreasing function, which proves item 1. Item 2 is a straight consequence of inequality (66) which, combined with the decreasing property of $t \mapsto \|v(t)\|$, yields

$$\|\dot{v}(t)\| \leq \|v_0\|. \tag{67}$$

As a straight consequence of inequality (66) we also obtain

$$\lambda(t)^2 \|\dot{x}(t)\|^2 \leq \|v(t)\|^2,$$

which, combined with the decreasing property of $t \mapsto \|v(t)\|$, yields item 3. \square

Let us make more precise the decreasing properties of $\|v(\cdot)\|$.

Corollary 3.3. *The following properties hold:*

1. for almost all $t \in [0, +\infty)$

$$-\|v(t)\|^2 \leq \frac{1}{2} \frac{d}{dt} (\|v(t)\|^2) \leq -\|\dot{v}(t)\|^2;$$

2. $e^{-t}\|v_0\| \leq \|v(t)\| \leq \|v_0\|$ for any $t \in [0, +\infty)$;

3. $\|\dot{v}(\cdot)\| \in L^2([0, +\infty))$.

Proof. To prove the first inequality of item 1, use (14) to obtain

$$\frac{d}{dt} \frac{1}{2} \|v(t)\|^2 = \langle \dot{v}(t), v(t) \rangle = -\langle \lambda(t)\dot{x}(t) + v(t), v(t) \rangle.$$

Then combine this inequality with (64) of Proposition 3.1 to obtain

$$\frac{d}{dt} \frac{1}{2} \|v(t)\|^2 \geq -\|v(t)\|^2.$$

The second inequality of item 1 follows directly from (65) of Proposition 3.1.

Items 2 and 3 follow from item 1 by integration arguments. Just notice that, by integration of the differential inequality $\frac{d}{dt}(\phi) + 2\phi \geq 0$ with $\phi(t) = \|v(t)\|^2$, we obtain $\phi(t) \geq e^{-2t}\phi(0)$, and hence $e^{-t}\|v_0\| \leq \|v(t)\|$. \square

Note that item 2 of Corollary 3.3 shows that, if $v_0 \neq 0$, then in “finite time” we do not have $v(t) = 0$. The best we can hope is that $\|v(t)\|$ decreases like e^{-t} .

3.2 Convergence properties

In this section, as a standing assumption, we assume that the set of equilibria is non empty: $T^{-1}(0) \neq \emptyset$. For studying the convergence properties of the trajectories of system ((13)-(14)), we make use of the following Lyapunov functions. Suppose that

$$\hat{x} \in T^{-1}(0) \neq \emptyset, \tag{68}$$

and define for any $t \geq 0$

$$g(t) := \frac{1}{2} \|x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t)\|^2; \tag{69}$$

$$h(t) := \frac{1}{2} \|\lambda(t)(x(t) - \hat{x}) + v(t)\|^2; \tag{70}$$

$$u(t) = \frac{1}{2} \|x(t) - \hat{x}\|^2 + \frac{1}{\lambda(t)} \langle x(t) - \hat{x}, v(t) \rangle. \tag{71}$$

Lemma 3.4. *If $\lambda(\cdot)$ is non-increasing, then $\lim_{t \rightarrow +\infty} v(t) = 0$. Moreover $t \mapsto \|v(t)\|$ is a decreasing function, and $v(\cdot) \in L^2([0, +\infty); H)$.*

Proof. Differentiating $h(\cdot)$, and using (14) we obtain, for almost all $t \in [0, +\infty)$,

$$\begin{aligned} \frac{d}{dt}h(t) &= \langle \lambda(t)(x(t) - \hat{x}) + v(t), \lambda(t)\dot{x}(t) + \dot{v}(t) \rangle \\ &\quad + \dot{\lambda}(t)\langle \lambda(t)(x(t) - \hat{x}) + v(t), x(t) - \hat{x} \rangle \\ &= - \langle \lambda(t)(x(t) - \hat{x}) + v(t), v(t) \rangle + \dot{\lambda}(t)[\lambda(t)\|x(t) - \hat{x}\|^2 + \langle v(t), x(t) - \hat{x} \rangle]. \end{aligned}$$

By monotonicity of T , and $0 \in T(\hat{x})$, $v(t) \in T(x(t))$, we have

$$\langle x(t) - \hat{x}, v(t) \rangle \geq 0. \quad (72)$$

Using inequality (72) and the decreasing property of $\lambda(\cdot)$, we deduce that, for almost all $t \in [0, +\infty)$,

$$\frac{d}{dt}h(t) + \|v(t)\|^2 \leq 0.$$

By integration with respect to t of the above inequality, and using that $h(t)$ is non-negative, we deduce that $\|v(\cdot)\|^2 \in L^1([0, +\infty))$. Combining this property with the fact that $t \mapsto \|v(t)\|$ is a decreasing function (Corollary 3.2), we conclude that $\lim_{t \rightarrow +\infty} v(t) = 0$. \square

Lemma 3.5. *Suppose that, for almost all $t \in [0, +\infty)$*

$$\lambda(t) + \dot{\lambda}(t) \geq 0.$$

Then, $x(\cdot)$ is a bounded trajectory.

Proof. Differentiating $u(\cdot)$ and using (14) we obtain

$$\begin{aligned} \frac{d}{dt}u(t) &= \langle x(t) - \hat{x}, \dot{x}(t) \rangle + \frac{1}{\lambda(t)} \langle x(t) - \hat{x}, \dot{v}(t) \rangle \\ &\quad + \frac{1}{\lambda(t)} \langle \dot{x}(t), v(t) \rangle - \frac{\dot{\lambda}(t)}{\lambda(t)^2} \langle x(t) - \hat{x}, v(t) \rangle \\ &= \frac{1}{\lambda(t)} \langle x(t) - \hat{x}, \lambda(t)\dot{x}(t) + \dot{v}(t) \rangle + \frac{1}{\lambda(t)} \langle \dot{x}(t), v(t) \rangle - \frac{\dot{\lambda}(t)}{\lambda(t)^2} \langle x(t) - \hat{x}, v(t) \rangle \\ &= -\frac{1}{\lambda(t)^2} \left(\lambda(t) + \dot{\lambda}(t) \right) \langle x(t) - \hat{x}, v(t) \rangle + \frac{1}{\lambda(t)} \langle \dot{x}(t), v(t) \rangle. \end{aligned}$$

Using the assumption $\lambda(t) + \dot{\lambda}(t) \geq 0$, together with inequalities (64) and (72), we deduce that, for almost all $t \in [0, +\infty)$, $\frac{d}{dt}u(t) \leq 0$. Therefore, $u(\cdot)$ is decreasing, which yields, for all $t \geq 0$

$$\frac{1}{2}\|x(t) - \hat{x}\|^2 \leq u(t) \leq u(0).$$

As a consequence, $\|x(\cdot)\|$ remains bounded, with an upper bound which can be easily expressed in terms of $\|v_0\|$ and $\|x_0 - \hat{x}\|$. \square

Corollary 3.6. *Suppose that, for almost all $t > 0$*

$$0 \geq \dot{\lambda}(t) \geq -\lambda(t).$$

Then, $v(t) \rightarrow 0$ as $t \rightarrow +\infty$, $x(\cdot)$ is bounded, and every sequential weak cluster point of $x(\cdot)$, as $t \rightarrow +\infty$, is a zero of T .

Proof. By Lemma 3.4, $v(t) \rightarrow 0$ as $t \rightarrow +\infty$, and, by Lemma 3.5, $x(\cdot)$ is bounded. On the other hand, by (13), for all $t \geq 0$, $v(t) \in T(x(t))$. From the sequential closedness property of the graph of T in $(w - H) \times H$, see [8] Proposition 2.5, we infer that whenever, $t_k \rightarrow +\infty$ and $x(t_k)$ converges weakly to some x_∞ , then $0 \in T(x_\infty)$. \square

To prove the weak convergence of $x(\cdot)$ we need additional assumptions. Note that the assumption of Lemma 3.5 can be equivalently written as

$$\frac{\dot{\lambda}(t)}{\lambda(t)} \geq -1.$$

Theorem 3.7. *Suppose that $\lambda(\cdot)$ is bounded from above on $[0, +\infty)$, and*

$$\liminf_{t \rightarrow +\infty} \frac{\dot{\lambda}(t)}{\lambda(t)} > -1. \quad (73)$$

Then $v(t) \rightarrow 0$, and $x(t)$ converges weakly to a zero of T , as t goes to $+\infty$.

Proof. Differentiating g and using (14) we have, for almost all $t > 0$,

$$\frac{d}{dt}g(t) = \left\langle \dot{x}(t) + \frac{1}{\lambda(t)}\dot{v}(t), x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t) \right\rangle \quad (74)$$

$$- \frac{\dot{\lambda}(t)}{\lambda(t)^2} \left\langle v(t), x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t) \right\rangle \quad (75)$$

$$= - \left(\frac{1}{\lambda(t)} + \frac{\dot{\lambda}(t)}{\lambda(t)^2} \right) \langle v(t), x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t) \rangle \quad (76)$$

$$= - \frac{1}{\lambda(t)} \left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \right) \langle v(t), x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t) \rangle. \quad (77)$$

Let us examine this last term. By (72), $\langle v(t), x(t) - \hat{x} \rangle \geq 0$ which clearly implies

$$\langle v(t), x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t) \rangle \geq \frac{1}{\lambda(t)}\|v(t)\|^2 \geq 0.$$

On the other hand, assumption (73) on $\lambda(\cdot)$ yields the existence of some $\epsilon > 0$ such that for t large enough

$$1 + \frac{\dot{\lambda}(t)}{\lambda(t)} \geq \epsilon.$$

Combining the last two inequalities with (77), we deduce that, for t large enough

$$\frac{d}{dt}g(t) + \epsilon \left\| \frac{1}{\lambda(t)}v(t) \right\|^2 \leq 0. \quad (78)$$

Since $\lambda(\cdot)$ is upper-bounded, we obtain $\|v(\cdot)\|^2 \in L^1([0, +\infty))$. Since $t \mapsto \|v(t)\|$ is decreasing, we conclude that

$$v(t) \rightarrow 0 \text{ as } t \rightarrow +\infty. \quad (79)$$

Define for $t \in [0, +\infty)$

$$\psi(t) := \left\| \frac{1}{\lambda(t)} v(t) \right\|^2. \quad (80)$$

Since $g \geq 0$, from (78) we obtain

$$\psi(\cdot) \in L^1([0, +\infty)). \quad (81)$$

Direct calculation yields

$$\frac{d}{dt} \psi(t) = -2 \frac{\dot{\lambda}(t)}{\lambda(t)^3} \|v(t)\|^2 + 2 \frac{1}{\lambda(t)^2} \langle v(t), \dot{v}(t) \rangle.$$

Since $\langle v(t), \dot{v}(t) \rangle \leq 0$, we have

$$\frac{d}{dt} \psi(t) \leq -2 \frac{\dot{\lambda}(t)}{\lambda(t)} \psi(t).$$

Using the assumption $\dot{\lambda}(t)/\lambda(t) \geq -1 + \epsilon \geq -1$, it follows that, for t large enough,

$$\frac{d}{dt} \psi(t) \leq 2\psi(t).$$

Equivalently

$$\frac{d}{dt} \left\{ \psi(t) - 2 \int_0^t \psi(s) ds \right\} \leq 0,$$

which means that the function $t \mapsto \Psi(t) := \psi(t) - 2 \int_0^t \psi(s) ds$ is decreasing. Since $\psi(\cdot)$ is nonnegative and $\psi(\cdot) \in L^1([0, +\infty))$ (by (81)), it follows that $\Psi(\cdot)$ is bounded from below. Thus $\lim_{t \rightarrow +\infty} \Psi(t)$ exists. From $\psi(\cdot) \in L^1([0, +\infty))$ and the definition of Ψ we deduce that $\lim_{t \rightarrow +\infty} \psi(t)$ exists. Using again $\psi(\cdot) \in L^1([0, +\infty))$, we finally obtain

$$\psi(t) \rightarrow 0 \text{ as } t \rightarrow +\infty. \quad (82)$$

Let us now return to g . By (78), $t \mapsto g(t)$ is decreasing. Since g is nonnegative, there exists $\lim_{t \rightarrow +\infty} g(t)$. By (82), $\|v(t)\|/\lambda(t) \rightarrow 0$ as $t \rightarrow +\infty$. Since

$$\left| \sqrt{2g(t)} - \|x(t) - \hat{x}\| \right| \leq \frac{1}{\lambda(t)} \|v(t)\| \rightarrow 0$$

we conclude that, for any $\hat{x} \in T^{-1}(0)$, there exists $\lim_{t \rightarrow +\infty} \|x(t) - \hat{x}\|$. On the other hand, from $v(t) \in T(x(t))$, $v(t) \rightarrow 0$ as $t \rightarrow +\infty$ (79), and the sequential closedness property of the graph of T in $(w - H) \times H$, we have that if $x(t_n) \rightharpoonup \bar{x}$ weakly in H for some $t_n \rightarrow +\infty$, then \bar{x} is a zero of T . We can now use Opial's Lemma [22], that we recall below. Taking $S = T^{-1}(0)$ in Opial's Lemma, we conclude that $x(t)$ converges weakly to a zero of T , as $t \rightarrow +\infty$.

Lemma 3.8. *Let H be a Hilbert space and $x : [0, +\infty) \rightarrow H$ a function such that there exists a nonempty set $S \subset H$ verifying:*

- (i) *for every $\hat{x} \in S$, $\lim_{t \rightarrow +\infty} \|x(t) - \hat{x}\|$ exists;*
- (ii) *if, for some $t_n \rightarrow +\infty$, $x(t_n) \rightharpoonup \bar{x}$ weakly in H , then $\bar{x} \in S$.*

Then

$$w - \lim_{t \rightarrow +\infty} x(t) = x_\infty \text{ exists for some element } x_\infty \in S.$$

□

3.3 The case λ constant

In this section, $\lambda > 0$ is assumed to be a positive constant. In this particular situation, we can revisit the results of the preceding section with the help of the theory of semi-groups of contractions. Given $T : H \rightrightarrows H$ a maximal monotone operator, we consider the Cauchy problem for the differential inclusion system

$$v(t) \in T(x(t)), \tag{83}$$

$$\lambda \dot{x}(t) + \dot{v}(t) + v = 0, \tag{84}$$

$$x(0) = x_0, v(0) = v_0 \in T(x_0). \tag{85}$$

Relying on the results of Section 3.2, Lemma 3.4 and Theorem 3.7, let us summarize the asymptotic properties as $t \rightarrow +\infty$ of the trajectories of system ((83)-(84)).

Theorem 3.9. *Let us assume that $T^{-1}(0)$ is non-empty. Then, for any trajectory $(x(\cdot), v(\cdot))$ of system ((83)-(84)) the following properties hold:*

- i) *$v(t) \rightarrow 0$ strongly in H as $t \rightarrow +\infty$. Moreover $v \in L^2([0, +\infty); H)$ and $\|v(t)\|$ is a decreasing function of t .*
- ii) *$x(t) \rightharpoonup \bar{x}$ weakly in H as $t \rightarrow +\infty$, with $\bar{x} \in T^{-1}(0)$.*

Let us show an other approach to the asymptotic analysis of system ((83)-(84)) which is based on the equivalent formulation

$$\dot{z}(t) + \mu T_\mu(z(t)) = 0, \tag{86}$$

with formulae expressing $x(t)$ and $v(t)$ in terms of $z(t)$

$$x(t) = J_\mu^T(z(t)), \tag{87}$$

$$v(t) = T_\mu(z(t)). \tag{88}$$

As a key ingredient in the asymptotic analysis of (86) we will use that the operator T_μ is μ -cocoercive. An operator $A : H \rightarrow H$ is said to be θ -cocoercive for some positive constant θ if for all x, y belonging to H

$$\langle Ay - Ax, y - x \rangle \geq \theta \|Ay - Ax\|^2. \tag{89}$$

When θ can be taken equal to one, the operator is said to be firmly nonexpansive. Note that A θ -cocoercive implies that A is $\frac{1}{\theta}$ -Lipschitz continuous, the converse

statement (and hence equivalence) being true when A is the gradient of a convex differentiable function (Baillon-Haddad's theorem). In our context, this notion plays an important role because of the following property (see [8], Proposition 2.6, and [30] with further examples of cocoercive mappings):

Proposition 3.10. *Let $T : H \rightrightarrows H$ be a maximal monotone operator. Then, for any positive constant μ , the Yosida approximation T_μ of index μ of T is μ -cocoercive and μT_μ is firmly nonexpansive.*

A classical result from Baillon and Brezis [7] states that a general maximal monotone operator generates trajectories which converge weakly in the ergodic sense. Indeed, Bruck [11] proved that weak convergence holds when T is maximal monotone and demipositive. This last property is satisfied by two important classes of maximal monotone operators, namely the subdifferentials of closed convex functions, and the cocoercive operators. One can consult [23] for a recent account on this subject. Let us state the convergence result in the cocoercive case.

Proposition 3.11. *Let $T : H \rightarrow H$ be a maximal monotone operator which is cocoercive. Let us assume that $T^{-1}(0)$ is non-empty. Then, for any trajectory $z(\cdot)$ of the classical differential equation*

$$\dot{z}(t) + T(z(t)) = 0$$

the following properties hold: as $t \rightarrow +\infty$

- i) $z(t)$ converges weakly in H to some element $\bar{z} \in T^{-1}(0)$;*
- ii) $\dot{z}(t)$ converges strongly in H to zero.*

We can now give a proof of theorem 3.9 which is based on the cocoercive property: By Proposition 3.11, using (86), and the cocoercive property of μT_μ , we deduce that $z(t)$ converges weakly to some element $\bar{z} \in T_\mu^{-1}(0) = T^{-1}(0)$ and $\dot{z}(t)$ converges strongly to zero, as $t \rightarrow +\infty$. From $v(t) = T_\mu(z(t)) = -\frac{1}{\mu}\dot{z}(t)$, we deduce that $v(t)$ converges strongly to zero, and from $-\dot{z}(t) = \mu T_\mu(z(t)) = z(t) - J_\mu^T(z(t))$ and $x(t) = J_\mu^T(z(t))$, we finally obtain that $x(t)$ converges weakly in H (with the same limit \bar{z} as $z(\cdot)$).

Strong convergence of the trajectories requires further information about T . Regarding this last property, demiregularity of operator T plays a key role (see [29], Definition 27.1):

Definition 3.12. *An operator $T : H \rightrightarrows H$ is demiregular if, for every sequence $(x_n, z_n)_{n \in \mathbb{N}}$ with $z_n \in Tx_n$, the following property holds:*

$$\begin{cases} x_n \rightharpoonup x \text{ weakly} \\ z_n \rightarrow z \text{ strongly} \end{cases} \Rightarrow x_n \rightarrow x \text{ strongly} . \quad (90)$$

The wealth and applicability of this notion is illustrated through the following examples (one can consult [5] for further examples):

Proposition 3.13. *Let $T : H \rightrightarrows H$ be a maximal monotone operator. Suppose that one of the following holds.*

1. T is strongly monotone, i.e., there exists some $\alpha > 0$ such that $T - \alpha I$ is monotone.
2. $T = \partial f$, where $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex, lower semicontinuous, proper function whose lower level sets are boundedly compact.
3. There exists some $\mu > 0$ such that J_μ^T is compact, i.e., for every bounded set $C \subset H$, the closure of $J_\mu^T(C)$ is compact.
4. $T : H \rightarrow H$ is single-valued with a single-valued continuous inverse.

Then T is demiregular.

We can now state a strong convergence result for trajectories of system ((83)-(84)).

Theorem 3.14. *Let us assume that $T : H \rightrightarrows H$ is a maximal monotone operator with $T^{-1}(0)$ non-empty, and that one of the following properties is satisfied.*

- a) T is demiregular; or
- b) $T^{-1}(0)$ has a non empty interior.

Then, for any trajectory $(x(\cdot), v(\cdot))$ of system ((83)-(84)) the following properties hold:

- i) $x(t) \rightarrow \bar{x}$ strongly in H as $t \rightarrow +\infty$, with $\bar{x} \in T^{-1}(0)$;
- ii) $v(t) \rightarrow 0$ strongly in H as $t \rightarrow +\infty$.

Proof. By theorem 3.9, we already know that $x(t) \rightharpoonup \bar{x}$ weakly in H as $t \rightarrow +\infty$, with $\bar{x} \in T^{-1}(0)$, and that $v(t) \rightarrow 0$ strongly in H as $t \rightarrow +\infty$. Hence we just need to prove that strong convergence of $x(t)$ holds.

- a) Let us assume that T is demiregular. We have $v(t) \in T(x(t))$ and $v(t) = T_\mu(z(t)) = -\frac{1}{\mu}\dot{z}(t)$. By theorem 3.9, we have $v(t) \rightarrow 0$ strongly in H , and $x(t) \rightharpoonup \bar{x}$ weakly in H . Demiregularity of T implies that $x(t) \rightarrow \bar{x}$ strongly in H as $t \rightarrow +\infty$.
- b) Let us now suppose that $T^{-1}(0)$ has a non empty interior. The following equivalences hold

$$Tz \ni 0 \Leftrightarrow z + \mu Tz \ni z \Leftrightarrow J_\mu^T(z) = z \Leftrightarrow \mu T_\mu(z) = 0. \quad (91)$$

Hence $(\mu T_\mu)^{-1}(0) = T^{-1}(0)$, and $(\mu T_\mu)^{-1}(0)$ has a non empty interior. Theorem 3.13 of Brézis [8] tells us that each trajectory of the equation $\dot{z}(t) + \mu T_\mu(z(t)) = 0$ converges strongly in H as $t \rightarrow +\infty$. From $x(t) = J_\mu^T(z(t))$, and by continuity of J_μ^T , we deduce that $x(t) \rightarrow \bar{x}$ strongly in H as $t \rightarrow +\infty$. \square

Remark 3.15. In the subdifferential case, an alternative proof of theorem 4.1 in the case λ constant, would consist in relying on the equivalent formulation of the dynamic

$$\dot{z}(t) + \mu \nabla f_\mu(z(t)) = 0$$

where $(\partial f)_\mu = \nabla f_\mu$, and f_μ is the Moreau-Yosida approximation of index μ of f . Applying classical convergence results valid for general gradient systems, see Bruck [11], Güler [16], one can infer that $f_\mu(z(t)) \rightarrow \inf_H f_\mu = \inf_H f$. From

$$\inf_H f \leq f(J_\mu^T(z(t))) \leq f_\mu(z(t))$$

and $x(t) = J_\mu^T(z(t))$ we obtain that $f(x(t))$ tends to $\inf_H f$ as $t \rightarrow +\infty$.

4 T subdifferential. Links with convex optimization

Let us now suppose that $T = \partial f$ is the subdifferential of a convex lower semicontinuous proper function $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$. By a classical result, T is a maximal monotone operator. The system ((13)-(14)-(15)) reads as follows

$$v(t) \in \partial f(x(t)), \quad (92)$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0, \quad (93)$$

$$x(0) = x_0, \quad v(0) = v_0. \quad (94)$$

We assume that standing assumptions (21)-(22) are satisfied by $\lambda(\cdot)$. Let us establish some optimizing properties of the trajectories generated by this dynamical system, and show that it is a descent method. We set $S = (\partial f)^{-1}(0) = \operatorname{argmin}_H f$ which, unless specified, may be possibly empty.

Theorem 4.1. *Suppose that $T = \partial f$, where $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function. Let $t \in [0, +\infty) \mapsto (x(t), v(t)) \in H \times H$ be the strong global solution of system ((92)-(93)-(94)). Then, the following hold:*

i) for any $0 < b < +\infty$, the real-valued function $t \mapsto f(x(t))$ is Lipschitz continuous on $[0, b]$. Hence $f(x(\cdot))$ is almost everywhere differentiable on $[0, +\infty)$, and for almost all $t \in [0, +\infty)$

$$\frac{d}{dt}f(x(t)) = \langle \dot{x}(t), v(t) \rangle = -\lambda(t)\|\dot{x}(t)\|^2 - \langle \dot{x}(t), \dot{v}(t) \rangle \leq -\lambda(t)\|\dot{x}(t)\|^2;$$

ii) $t \mapsto f(x(t))$ is a non increasing function;

Assuming moreover that $t \mapsto \lambda(t)$ is non increasing, then

iii) $f(x(t))$ decreases to $\inf_H f$ as $t \uparrow +\infty$;

iv) if f is bounded from below, then $\|v(\cdot)\| \in L^2([0, +\infty))$ and $v(t) \rightarrow 0$ as $t \rightarrow +\infty$.

Proof. *i)* Suppose that $0 \leq t_1 < t_2 < +\infty$, and let

$$v_i = v(t_i), \quad x_i = x(t_i), \quad i = 1, 2.$$

Since $v_i \in \partial f(x_i)$, $i = 1, 2$ we have

$$\begin{aligned} f(x_1) + \langle x_2 - x_1, v_1 \rangle &\leq f(x_2) \\ f(x_2) + \langle x_1 - x_2, v_2 \rangle &\leq f(x_1). \end{aligned}$$

Therefore

$$\langle x_2 - x_1, v_1 \rangle \leq f(x_2) - f(x_1) \leq \langle x_2 - x_1, v_2 \rangle. \quad (95)$$

By Corollary 3.2 (1.), $t \mapsto \|v(t)\|$ is a decreasing function. From (95) we deduce that

$$|f(x_2) - f(x_1)| \leq \|x_2 - x_1\| \|v_0\|. \quad (96)$$

By Corollary 3.2 (3.), for any $0 < b < +\infty$, $t \mapsto x(t)$ is Lipschitz continuous on $[0, b]$. Combining this property with (96), we conclude that, for any $0 < b < +\infty$, $t \mapsto f(x(t))$ is Lipschitz continuous on $[0, b]$.

Suppose now that $x(\cdot)$ is differentiable at t_1 . Let us divide (95) by $t_2 - t_1 > 0$ and take the limit $t_2 \rightarrow t_1^+$. Since $v(\cdot)$ is continuous, see Corollary 3.2 (2.), it follows that

$$\frac{d}{dt}f(x(t_1)) = \langle \dot{x}(t_1), v_1 \rangle,$$

that is, for almost all $t \in [0, +\infty)$

$$\frac{d}{dt}f(x(t)) = \langle \dot{x}(t), v(t) \rangle.$$

Replacing $v(t)$ by $v(t) = -(\lambda(t)\dot{x}(t) + \dot{v}(t))$, as given by (93), in the above formula, we obtain

$$\begin{aligned} \frac{d}{dt}f(x(t)) &= \langle \dot{x}(t), v(t) \rangle \\ &= -\langle \dot{x}(t), \lambda(t)\dot{x}(t) + \dot{v}(t) \rangle \\ &= -\lambda(t)\|\dot{x}(t)\|^2 - \langle \dot{x}(t), \dot{v}(t) \rangle \\ &\leq -\lambda(t)\|\dot{x}(t)\|^2, \end{aligned}$$

the last inequality being a consequence of $\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0$, see (63).

ii) The function $t \mapsto f(x(t))$ is Lipschitz continuous, and hence absolutely continuous, on each bounded interval $[0, b]$, $0 < b < +\infty$. Moreover, its derivative is less or equal than zero for almost all $t \in [0, +\infty)$. This classically implies that $f(x(\cdot))$ is non increasing.

iii) Define, for $y \in \text{dom}f$

$$\phi_y(t) = \left[f(y) - (f(x(t)) + \langle y - x(t), v(t) \rangle) \right] + \frac{\lambda(t)}{2} \|y - x(t)\|^2. \quad (97)$$

Note that, since $f(\cdot)$ is convex and $v(t) \in \partial f(x(t))$, we have $\phi_y(t) \geq 0$ for all $t \geq 0$. Moreover, since $t \mapsto f(x(t))$, $t \mapsto x(t)$, $t \mapsto v(t)$ are locally Lipschitz continuous (by item *i*) and Corollary 3.2), the function $\phi_y(\cdot)$ is also locally Lipschitz continuous and, in particular, absolutely continuous on compact sets.

Using item *i*) and (93) we deduce that, for almost all $t \in [0, +\infty)$,

$$\begin{aligned} \frac{d\phi_y}{dt}(t) &= -\langle y - x(t), \dot{v}(t) \rangle + \lambda(t)\langle x(t) - y, \dot{x}(t) \rangle + \frac{\dot{\lambda}(t)}{2} \|y - x(t)\|^2 \\ &= \langle x(t) - y, \lambda(t)\dot{x}(t) + \dot{v}(t) \rangle + \frac{\dot{\lambda}(t)}{2} \|y - x(t)\|^2 \\ &= \langle y - x(t), v(t) \rangle + \frac{\dot{\lambda}(t)}{2} \|y - x(t)\|^2, \end{aligned}$$

which combined with the convexity of $f(\cdot)$, and the assumption on $\lambda(\cdot)$ being non increasing, yields (for almost all $t \in [0, +\infty)$)

$$\frac{d\phi_y}{dt}(t) \leq f(y) - f(x(t)).$$

Let us integrate this inequality with respect to t . Using that $t \mapsto f(x(t))$ is a non increasing function (see item *ii*), and that ϕ_y is non-negative, we deduce that, for any $t \geq 0$

$$\begin{aligned} -\phi_y(0) &\leq \phi_y(t) - \phi_y(0) \\ &\leq \int_0^t f(y) - f(x(s)) \, ds \leq t[f(y) - f(x(t))]. \end{aligned} \quad (98)$$

Hence, for any $t > 0$

$$f(x(t)) \leq f(y) + \frac{\phi_y(0)}{t}.$$

Passing to the limit as $t \rightarrow +\infty$ in the above inequality yields

$$\lim_{t \rightarrow +\infty} f(x(t)) \leq f(y).$$

This being true for any $y \in \text{dom} f$, we finally obtain item *iii*)

$$\lim_{t \rightarrow +\infty} f(x(t)) = \inf_H f.$$

iv) By item *i*),

$$\frac{d}{dt} f(x(t)) \leq -\lambda(t) \|\dot{x}(t)\|^2 \leq 0.$$

Since $f(\cdot)$ has been supposed to be bounded from below, after integration we obtain

$$\int_0^{+\infty} \lambda(t) \|\dot{x}(t)\|^2 \, dt < +\infty. \quad (99)$$

Since $\lambda(\cdot)$ is assumed to be non increasing, we have, for any $t \geq 0$,

$$\lambda(t)^2 \|\dot{x}(t)\|^2 \leq \lambda(0) (\lambda(t) \|\dot{x}(t)\|^2),$$

which, combined with (99), yields $\lambda(\cdot) \|\dot{x}(\cdot)\| \in L^2([0, +\infty))$. Then, use item 3) of Corollary 3.3, (93), and the triangle inequality, to obtain $\|v(\cdot)\| \in L^2([0, +\infty))$. Since $t \mapsto \|v(t)\|$ is a decreasing function, we conclude that $v(t) \rightarrow 0$ as $t \rightarrow +\infty$. \square

Remark 4.2. Let us now assume that $S = (\partial f)^{-1}(0) \neq \emptyset$. By taking y equal to the projection of x_0 onto S in (97), and using (98), we obtain

$$\int_0^{+\infty} (f(x(s)) - \inf_H f) \, ds \leq C, \quad (100)$$

and

$$f(x(t)) - \inf_H f \leq \frac{C}{t}, \quad (101)$$

where

$$C = \phi_y(0) \leq \frac{\lambda(0)}{2} \|y - x_0\|^2 - \langle y - x_0, v_0 \rangle \quad (102)$$

$$\leq \frac{\lambda(0)}{2} \text{dist}(x_0, S)^2 + \|v_0\| \text{dist}(x_0, S). \quad (103)$$

Remark 4.3. By theorem 4.1 item *iv*) we have $v(t) \rightarrow 0$ as $t \rightarrow +\infty$. Moreover, for any $t \geq 0$, $v(t) \in \partial f(x(t))$. From the sequential closedness property of ∂f in $(w - H) \times H$, it follows that any sequential weak cluster point of the trajectory $x(\cdot)$ belongs to $S = (\partial f)^{-1}(0)$. By contraposition, we deduce that, if S is empty, then for any trajectory of system ((92)-(93)) there is explosion, i.e., $\lim_{t \rightarrow +\infty} \|x(t)\| = +\infty$.

5 The case $\lambda(t) = \lambda_0 e^{-t}$

In this section, we discuss the case

$$\lambda(t) = \lambda_0 e^{-t}, \quad (104)$$

with $\lambda_0 > 0$, a positive given parameter. For this choice of $\lambda(\cdot)$, for any $t \geq 0$

$$0 > \dot{\lambda}(t) = -\lambda(t).$$

By Corollary 3.6, it follows that the trajectory $x(\cdot)$ is bounded, $v(t)$ converges to 0 as $t \rightarrow +\infty$, and every sequential weak cluster point of $x(\cdot)$ is a zero of T . It is possible to have a closed formula for $x(t)$, $v(t)$ and to estimate how fast is the convergence of $v(t)$ to 0. As in (27), let us define $z(\cdot)$ by

$$z(t) = x(t) + \frac{1}{\lambda(t)}v(t) = x(t) + \frac{e^t}{\lambda_0}v(t).$$

Setting $\mu(t) = \frac{1}{\lambda(t)} = \frac{e^t}{\lambda_0}$, we have $\dot{\mu}(t) = \mu(t)$, which, by (41), implies $\dot{z}(t) = 0$ for all $t \geq 0$. Hence, for all $t \geq 0$,

$$x(t) + \frac{e^t}{\lambda_0}v(t) = z(0) = x_0 + \frac{1}{\lambda_0}v_0$$

which, in view of the inclusion $v(t) \in T(x(t))$ is equivalent to

$$x(t) = J_{e^t/\lambda_0}^T(z(0)), \quad v(t) = T_{e^t/\lambda_0}(z(0)). \quad (105)$$

The next proposition is a direct consequence of the above equation.

Proposition 5.1. *Let $\lambda(\cdot)$ be given by (104). Assume that $T^{-1}(0)$ is non-empty and let x_0^* be the orthogonal projection of $x_0 + \lambda_0^{-1}v_0$ onto $T^{-1}(0)$. Then, the following properties hold:*

i) $\forall t \geq 0$, $\|x(t) - (x_0 + \lambda_0^{-1}v_0)\| \leq \|x_0^ - (x_0 + \lambda_0^{-1}v_0)\|$;*

ii) $\forall t \geq 0$, $\|v(t)\| \leq \lambda_0 e^{-t} \|x_0^ - (x_0 + \lambda_0^{-1}v_0)\|$;*

iii) $\lim_{t \rightarrow +\infty} x(t) = x_0^$.*

Proof. To simplify the proof, set $z_0 = x_0 + \lambda_0^{-1}v_0$.

i) To prove the first inequality, take $x^* \in T^{-1}(0)$. By monotonicity of T , and $z(t) = x(t) + \frac{1}{\lambda(t)}v(t) = z_0$, we have

$$0 \leq \frac{1}{\lambda} \langle x^* - x(t), 0 - v(t) \rangle = \langle x^* - x(t), x(t) - z_0 \rangle.$$

Thus,

$$\begin{aligned}\|x^* - z_0\|^2 &= \|x^* - x(t)\|^2 + 2\langle x^* - x(t), x(t) - z_0 \rangle + \|x(t) - z_0\|^2 \\ &\geq \|x(t) - z_0\|^2.\end{aligned}$$

This being true for any $x^* \in T^{-1}(0)$, passing to the infimum with respect to $x^* \in T^{-1}(0)$ establishes the formula.

ii) By (105) and item i)

$$\begin{aligned}\|v(t)\| &= \|T_{e^t/\lambda_0}(z_0)\| \\ &= \lambda_0 e^{-t} \|x(t) - z_0\| \\ &\leq \lambda_0 e^{-t} \|x_0^* - (x_0 + \lambda_0^{-1}v_0)\|.\end{aligned}$$

iii) We have $x(t) = J_{e^t/\lambda_0}^T(z_0)$, which, equivalently, can be written as

$$\lambda_0 e^{-t} (x(t) - z_0) + T(x(t)) \ni 0.$$

Noticing that $\lambda_0 e^{-t} \rightarrow 0$ as $t \rightarrow +\infty$, by using the classical asymptotic properties of the Tikhonov approximation, see for example Browder [10], we obtain

$$\lim_{t \rightarrow +\infty} x(t) = \text{proj}_{T^{-1}(0)} z_0 = x_0^*.$$

□

Note that $\|v(t)\| \leq c e^{-t}$, which, as an asymptotical behavior, is almost as good as the “pure” Newton’s continuous dynamic.

6 Examples

The following elementary examples are intended to illustrate the asymptotic behavior of the trajectories of our system.

6.1 Linear monotone operators

Given $a, b > 0$, let $T = \nabla f$, with $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ being defined by

$$f(\xi_1, \xi_2, \xi_3) = \frac{a}{2} \xi_1^2 + \frac{b}{2} \xi_2^2.$$

The corresponding solution of system ((13)-(14)-(15)) with $\lambda > 0$ constant, and Cauchy data $x_0 = (\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3)$ is given by

$$x(t) = \left(\bar{\xi}_1 \exp\left(-\frac{a}{a+\lambda}t\right), \exp\left(-\frac{b}{b+\lambda}t\right) \bar{\xi}_2, \bar{\xi}_3 \right).$$

Consider now the same system ((13)-(14)-(15)) with $\lambda(t) = \lambda_0 \exp(-t)$. The solution is given by

$$x(t) = \left(\frac{\lambda_0 + a}{\lambda_0 + a \exp(t)} \bar{\xi}_1, \frac{\lambda_0 + b}{\lambda_0 + b \exp(t)} \bar{\xi}_2, \bar{\xi}_3 \right).$$

By contrast with the steepest descent continuous dynamic, note the effect of Newton’s direction term, which makes trajectories close to straight lines.

6.2 Discontinuous monotone operators

a) Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = |x|$, $T = \partial f$ and $x_0 = 1$. Then, the solution of system ((13)-(14)-(15)) with $\lambda > 0$ constant is given by

$$x(t) = \begin{cases} 1 - t/\lambda, & 0 \leq t \leq \lambda \\ 0, & \lambda < t \end{cases} \quad v(t) = \begin{cases} 1, & 0 \leq t \leq \lambda \\ \exp(\lambda - t), & \lambda < t \end{cases}$$

b) Let us now take $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = \max\{|x|, x^2\}$. Then, for $x \geq 0$ we have

$$\partial f(x) = \begin{cases} \{2x\}, & x > 1, \\ [1, 2], & x = 1 \\ \{1\} & 0 < x < 1 \\ [-1, 1], & x = 0 \end{cases}$$

and $\partial f(x) = -\partial f(-x)$ for $x < 0$. Define

$$t_1 = \left(\frac{\lambda}{2} + 1\right) \log 2, \quad t_2 = \left(\frac{\lambda}{2} + 2\right) \log 2, \quad t_3 = \left(\frac{\lambda}{2} + 2\right) \log 2 + \lambda.$$

The solution of system ((13)-(14)-(15)) with λ constant, $T = \partial f$, and $x_0 = 2$ is given by

$$x(t) = \begin{cases} 2 \exp\left(-\frac{2}{\lambda+2}t\right), & 0 \leq t \leq t_1 \\ 1, & t_1 \leq t \leq t_2 \\ 1 - \frac{t-t_2}{\lambda}, & t_2 \leq t \leq t_3 \\ 0, & t_3 \leq t \end{cases} \quad v(t) = \begin{cases} 4 \exp\left(-\frac{2}{\lambda+2}t\right), & 0 \leq t \leq t_1 \\ 2 \exp(t_1 - t), & t_1 \leq t \leq t_2 \\ 1, & t_2 \leq t \leq t_3 \\ \exp(t_3 - t), & t_3 \leq t \end{cases}$$

6.3 Antisymmetric linear operators

As a benchmark case, in which many of the nice features attached to subdifferential of convex functions fail to be satisfied, let us consider

$$H = \mathbb{R} \times \mathbb{R}, \quad T = \text{rot}\left(0, \frac{\pi}{2}\right), \quad T(x_1, x_2) = (-x_2, x_1).$$

Clearly, T is a maximal monotone operator with $T^* = -T$ and $\langle Tx, x \rangle = 0$ for all $x \in H$. Take $\lambda > 0$ constant.

Setting $X(t) = x_1(t) + ix_2(t)$, system ((13)-(14)-(15)) can be formulated in \mathbb{C} as

$$(\lambda + i)\dot{X}(t) + iX(t) = 0.$$

Integration of this system (X_0 being the Cauchy data in \mathbb{C} corresponding to $x_0 \in \mathbb{R} \times \mathbb{R}$) yields

$$X(t) = X_0 \exp\left(-\frac{1+i\lambda}{1+\lambda^2}t\right),$$

which clearly implies $x(t) \rightarrow 0$ as $t \rightarrow +\infty$. By contrast, trajectories generated by T , which are solutions of $\dot{x}(t) + T(x(t)) = 0$, fail to converge to 0 (indeed they converge to 0 in the ergodic sense).

7 Application: A Levenberg-Marquardt algorithm for convex minimization

In this section, H is a real Hilbert space and $f : H \rightarrow \mathbb{R}$ is a C^2 convex function with a non-empty set of minimizers (not necessarily reduced to a single element). The results obtained so far suggest that, when taking $x_0 \in H$, $\lambda > 0$ and $\{t_k\}$ a sequence of strictly positive steps, the sequence (x_k) defined by the algorithm

$$\nabla f(x_k) + (\nabla^2 f(x_k) + \lambda I) \left(\frac{x_{k+1} - x_k}{t_k} \right) = 0$$

is convergent, if the t_k 's are chosen appropriately.

One may consult [1], [13], [15], [21], [27] and references therein for an overview on such Newton-like methods. In [23], one can find a survey on the rich connections between continuous evolution equations generated by maximal monotone operators and their discrete time versions. Previous *global* convergence analyses of Quasi-Newton methods required boundedness of level sets and were restricted to criticality of all cluster points of the generated sequence, which, for convex objective functions, implies optimality of these cluster points.

To simplify the exposition, we use the following equivalent formulation:

$$x_{k+1} = x_k + t_k s_k, \quad s_k = -(\nabla^2 f(x_k) + \lambda I)^{-1} \nabla f(x_k). \quad (106)$$

We assume that each t_k is chosen following Armijo's rule: we pick some $\beta \in (0, 1/2)$ and

$$t_k = \max \{t \in \{1, 1/2, 1/4, \dots\} \mid f(x_k + t s_k) \leq f(x_k) + \beta t \langle s_k, \nabla f(x_k) \rangle\}. \quad (107)$$

Our aim is to prove the following new result.

Theorem 7.1. *Let us assume that $\nabla^2 f$ is Lipschitz continuous. Then, for any initial data $x_0 \in H$, the sequence $\{x_k\}$ generated by algorithm ((106)-(107)) converges weakly to a minimizer of f .*

Let us denote by $L > 0$ the Lipschitz constant of $\nabla^2 f$ (with respect to the operator norm).

Proposition 7.2. *Given $x \in H$, set $s = -(\nabla^2 f(x) + \lambda I)^{-1} \nabla f(x)$. Then, for any $t \in [0, 1]$ the following inequality holds:*

$$f(x + ts) \leq f(x) + \frac{1}{2} t \langle s, \nabla f(x) \rangle + \frac{t^2 \|s\|^2}{2} \left[-\lambda + \frac{tL \|s\|}{3} \right].$$

Proof. Since $\nabla^2 f$ is Lipschitz continuous with constant L , we have

$$f(x + ts) \leq f(x) + t \langle s, \nabla f(x) \rangle + \frac{t^2}{2} \langle \nabla^2 f(x) s, s \rangle + \frac{Lt^3}{6} \|s\|^3.$$

By the definition of s , we have

$$\langle \nabla^2 f(x) s, s \rangle = \langle (\nabla^2 f(x) + \lambda I) s, s \rangle - \lambda \|s\|^2 \quad (108)$$

$$= -\langle \nabla f(x), s \rangle - \lambda \|s\|^2. \quad (109)$$

Combining the above equations, we conclude that

$$f(x + ts) \leq f(x) + \left(t - \frac{t^2}{2}\right) \langle s, \nabla f(x) \rangle + \frac{t^2 \|s\|^2}{2} \left[-\lambda + \frac{tL\|s\|}{3}\right]. \quad (110)$$

On the other hand, by (109) and the convexity of f , we have $\langle s, \nabla f(x) \rangle \leq 0$. Since $t \in [0, 1]$, this immediately implies

$$\left(t - \frac{t^2}{2}\right) \langle s, \nabla f(x) \rangle \leq \frac{t}{2} \langle s, \nabla f(x) \rangle.$$

Combining these two last inequalities gives the desired conclusion. \square

Proposition 7.3. *If $t_k < 1$ then*

$$\frac{1}{2} \geq t_k \geq \frac{3\lambda}{2L\|s_k\|}, \quad \text{and} \quad \|s_k\| \geq \frac{3\lambda}{L}.$$

Proof. If $t \in [0, 1]$ and $t \leq \frac{3\lambda}{L\|s_k\|}$, by using Proposition 7.2 we conclude that

$$f(x_k + ts_k) \leq f(x_k) + \beta t \langle s_k, \nabla f(x_k) \rangle.$$

Therefore, if $t_k < 1$, we must have

$$1 \geq 2t_k \geq \frac{3\lambda}{L\|s_k\|}.$$

\square

End of the proof of theorem 7.1. By (106)-(107), for all $k \in \mathbb{N}$

$$f(x_{k+1}) \leq f(x_k) + \beta t_k \langle s_k, \nabla f(x_k) \rangle \leq f(x_k) - \beta t_k \lambda \|s_k\|^2.$$

By summing these inequalities, and taking $x^* \in H$ such that $f(x^*) = \inf_H f$, we obtain

$$f(x^*) \leq f(x_0) - \sum_{k=0}^{+\infty} \beta t_k \lambda \|s_k\|^2. \quad (111)$$

Define

$$I = \{k \in \mathbb{N} \mid t_k < 1\}.$$

By Proposition 7.3

$$f(x_0) - f(x^*) \geq \sum_{k \in I} \beta t_k \lambda \|s_k\|^2 \geq \sum_{k \in I} \frac{3\beta\lambda^2}{2L} \|s_k\| \geq \sum_{k \in I} \frac{9\beta\lambda^3}{2L^2}.$$

As a consequence, I is finite. On the other hand, from (111), and $t_k = 1$ for $k \notin I$

$$\sum_{k \notin I} \beta \lambda \|s_k\|^2 \leq f(x_0) - f(x^*).$$

Since I is finite, this implies $\sum_k \|s_k\|^2 < +\infty$.

We can now prove that the sequence $\{x_k\}$ weakly converges. Set

$$r_k = \nabla f(x_{k+1}) + \lambda(x_{k+1} - x_k).$$

If $k \notin I$ (i.e., $t_k = 1$), by (106) and Taylor's formula, we easily deduce that

$$\|r_k\| \leq \frac{L}{2} \|s_k\|^2.$$

Therefore $\sum_k \|r_k\| < +\infty$, and the convergence of $\{x_k\}$ follows from Rockafellar's theorem on the proximal point method with summable error, see [25].

References

- [1] F. Alvarez, J. Bolte and J. Munier, *A unifying local convergence result for Newton's method in riemannian manifolds*, Foundations of Computational Mathematics, **8** (2008), pp. 197–226.
- [2] F. Alvarez and J. M. Pérez, *A dynamical system associated with Newton's method for parametric approximations of convex minimization problems*, Appl. Math. Optim., **38** (1998), pp. 193–217.
- [3] F. Alvarez, H. Attouch, J. Bolte and P. Redont, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., **81** (2002), pp. 747–779.
- [4] H. Attouch, J. Bolte and P. Redont, *Optimizing properties of an inertial dynamical system with geometric damping. Link with proximal methods*, Control and Cybernetics, **31** (2002), pp. 643–657.
- [5] H. Attouch, L. Bruceno and P. L. Combettes, *A parallel splitting method for coupled monotone inclusions*, SIAM J. Control Optim., **48** (2010), pp. 3246–3270.
- [6] H. Attouch and P. Redont, *The second-order in time continuous Newton method*, Approximation, optimization and mathematical economics (Pointe-à-Pitre, 1999), Physica, Heidelberg, 2001.
- [7] J.-B. Baillon and H. Brézis, *Une remarque sur le comportement asymptotique des semi-groupes non linéaires*, Houston J. Math., **2** (1976), pp. 5–7.
- [8] H. Brézis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland/Elsevier, New-York, 1973.
- [9] H. Brézis, *Analyse fonctionnelle*, Masson, Paris, 1983.
- [10] F. E. Browder, *Existence and approximation of solutions of nonlinear variational inequalities*, Proc. Nat. Acad. Sci. USA, **56** (1966), pp. 1080–1086.

- [11] R. E. Bruck, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, J. Funct. Anal., **18** (1975), pp. 15-26.
- [12] X. Chen, Z. Nashed and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operators equations*, SIAM Journal on Numerical Analysis, **38** (2000), pp. 1200-1216.
- [13] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Minimization*, Prentice-Hall, Englewood Cliffs, NJ, 1983. Reprinted by SIAM publications, 1993.
- [14] J. C. Dunn and D. P. Bertsekas, *Efficient dynamic programming implementations of Newton's method for unconstrained optimal control problems*, J. of Optimization Theory and Applications, **63** (1999), pp. 23-38.
- [15] O. P. Ferreira and B. F. Svaiter, *Kantorovich's Theorem on Newton's Method on Riemannian Manifolds*, Journal of Complexity, **18** (2002), pp. 304-329.
- [16] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control Optim., **29** (1991), pp. 403-419.
- [17] A. Haraux, *Systèmes dynamiques dissipatifs et applications*, RMA 17, Masson, Paris, 1991.
- [18] A. N. Iusem, B. F. Svaiter and J. X. Da Cruz Neto, *Central paths, generalized proximal point methods, and Cauchy trajectories in riemannian manifolds*, SIAM J. Control Optim., **37** (1989), pp. 566-588.
- [19] J. Jacobsen, O. Lewis and B. Tennis, *Approximations of continuous Newton's method: An extension of Cayley's problem*, Electronic J. of Differential Equations, **15** (2007), pp. 163-173.
- [20] G. J. Minty, *Monotone (nonlinear) operators in Hilbert spaces*, Duke Mathematical Journal, **29** (1962), pp. 341-346.
- [21] J. Nocedal and S. Wright, *Numerical optimization*, Springer Series in Operations Research, Springer-Verlag, New-York, 1999.
- [22] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., **73** (1967), pp. 591-597.
- [23] J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. of Convex Analysis, **17** (2010), pp. 1113-1163.
- [24] A. G. Ramm, *Dynamical systems method for solving operator equations*, Mathematics in Science and Engineering, **208**, Elsevier, 2006.
- [25] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., **14** (1976), pp. 877-898.

- [26] R. T. Rockafellar, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. Henri Poincaré, **2** (1985), pp. 167–184.
- [27] M. V. Solodov and B. F. Svaiter, *A globally convergent inexact Newton method for systems of monotone equations*, Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Edited by M. Fukushima and L. Qi, Kluwer Academic Publishers, (1999), pp. 355-369.
- [28] E. D. Sontag, *Mathematical control theory*, second edition, Springer-Verlag, New-York, 1998.
- [29] E. Zeidler, *Nonlinear functional analysis and its applications, Part II: Monotone operators*, Springer-Verlag, New-York, 1990.
- [30] D. L. Zhu and P. Marcotte, *Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities*, J. Optim., **6** (1996), pp. 714-726.