A CONTINUOUS DYNAMICAL NEWTON-LIKE APPROACH TO SOLVING MONOTONE INCLUSIONS

H. ATTOUCH[*] [†] B. F. SVAITER [‡] [§]

January, 25, 2010

Abstract

We introduce non-autonomous continuous dynamical systems which are linked to Newton and Levenberg-Marquardt methods. They aim at solving inclusions governed by maximal monotone operators in Hilbert spaces. Relying on 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 Cauchy-Lipschitz theorem. By using Lyapunov asymptotical analysis, we prove that their trajectories converge weakly to equilibria. Time discretization of these dynamics gives algorithms providing new insight on 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 algorithm, Levenberg-Marquardt algorithm, 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 $\langle ., . \rangle$, 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 $T^{'}$, classical Newton method generates sequences $(x_k)_{k\in\mathbb{N}}$ verifying

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

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

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

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

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

where I is the identity operator on H, and $(\lambda_k)_{n\in\mathbb{N}}$ is a sequence of positive real numbers. When T derives from a convex potential, this algorithm can be viewed as an interpolation between Newton's method and 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 gradient method). This algorithm has a natural interpretation as a time discretized version of the continuous dynamic

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

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 $\lambda(.)$ 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

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

Let us now consider a general maximal monotone operator $T : H \rightrightarrows H$, which is possibly multivalued, non everywhere defined (one may consult Brezis [8] for a detailed presentation of the theory of maximal monotone operators in Hilbert spaces). In order to solve the corresponding differential-algebraic inclusion system,

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

which involves an inclusion instead of an equality in the first equation, we use the following device. Let us rewrite the nonsmooth multivalued part of (7), namely

$v(t) \in T(x(t))$, by using the following equivalences:

$$v(t) \in T(x(t)) \Leftrightarrow \tag{8}$$

$$x(t) + \frac{1}{\lambda(t)} v(t) \in x(t) + \frac{1}{\lambda(t)} T(x(t)) \Leftrightarrow \tag{9}$$

$$x(t) = \left(I + \frac{1}{\lambda(t)} T \right)^{-1} \left(x(t) + \frac{1}{\lambda(t)} v(t) \right). \tag{10}$$

Set $\mu(t) = \frac{1}{\lambda(t)}$. Let us introduce the new unknown function

$$z(t) = x(t) + \frac{1}{\lambda(t)} v(t) = x(t) + \mu(t) v(t) \tag{11}$$

and rewrite (7) with the help of (x, z). From (10) and (11)

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

Equivalently, denoting by $J_\mu^T = (I + \mu T)^{-1}$ the resolvent of index $\mu > 0$ of T, and by $T_\mu = \frac{1}{\mu} \left(I - J_\mu^T \right)$ its Yosida approximation of index $\mu > 0$,

$$x(t) = J_{\mu(t)}^T (z(t)) \tag{12}$$

$$v(t) = T_{\mu(t)}(z(t)). \tag{13}$$

In our context, this is Minty representation of maximal monotone operators, see [20]. In a finite dimensional setting, this technic has been developped by Rockafellar in [26]: he shows that a maximal monotone operator can be represented as a lipschitzian manifold, which allows him to define the second derivatives of nonsmooth functions. This representation fits well our study. Indeed, let us show that the second equation of (7) can be reformulated as a classical differential equation with respect to $z(\cdot)$. First, let us rewrite (7) as

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

Differentiating (11) and using (14) we obtain

$$\dot{z}(t) = \dot{x}(t) + \mu(t)\dot{v}(t) + \dot{\mu}(t)v(t) \tag{15}$$

$$= -\mu(t)v(t) + \dot{\mu}(t)v(t). \tag{16}$$

From (16) and $v(t) = T_{\mu(t)}(z(t))$ we deduce that

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0. \tag{17}$$

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

$$x(t) = J_{\mu(t)}^T (z(t)) \tag{18}$$

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0. \tag{19}$$

As a nice feature of system (18, 19), let us stress the fact that the operators $J_\mu^T : H \to H$ and $T_\mu : H \to H$ are Lipschitz continuous, which makes this system relevant to Cauchy-Lipschitz theorem.
All along the paper, we shall pay particular attention to the case $\lambda(t) \to 0$ as $t \to +\infty$ (equivalently $\mu(t) \to +\infty$ as $t \to +\infty$). In that case, one may expect obtaining rates of convergence close to Newton method.

The paper is organized as follows: In section 2, assuming $\lambda(.)$ to be locally absolutely continuous, for any given Cauchy data $x(0) = x_0,\ v(0) = v_0 \in T(x_0)$, we prove existence and uniqueness of a strong global solution to the initial system (7). In section 3, this result is completed by establishing some general properties of its trajectories. In section 4, we study the asymptotic behavior of trajectories of this system. Assuming that $\lambda(t)$ does not converge too rapidly to zero as $t \to +\infty$ (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 section 5, 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 6, we consider the autonomous case $\lambda(t) \equiv \lambda_0$, and make the link with some classical results concerning semi-groups of contractions generated by maximal monotone operators. In section 7, we examine the case $\lambda(t) = \lambda_0 e^{-t}$, which is the closest situation to Newton dynamic allowed by our study. In section 8, we give some elementary examples aiming at illustrating the dynamic. In section 9, we finally give an application to numerical convex optimization.

Our approach, which can be traced back to 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 [6] (Newton 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.) An other interesting regularization method (based on regularization of the objective function) has been developed by Alvarez and Pérez in [2]. Among the rich literature concerning Newton 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's 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

$$v(t) \in T(x(t)), \tag{20}$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0, \tag{21}$$

$$x(0) = x_0,\ v(0) = v_0 \in T(x_0). \tag{22}$$

We shall successively define a notion of strong solution for this system and then prove existence and uniqueness of a strong solution by relying on the equivalent formulation (18,19).

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 $f : [0,b] \to H$ is said to be absolutely continuous if one the following equivalent properties holds :*

i) there exists an integrable function $g : [0,b] \to H$ 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 distribution 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 eveywhere coincide with its distribution 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 an Hilbert space).

Definition 2.2. *We say that a pair $(x(\cdot), v(\cdot))$ is a strong global solution of ((20), (21), (22)) if the following properties i), ii), iii) and iv) are satisfied:*

$$i)\ x(\cdot), v(\cdot) : [0,+\infty) \to H \text{ are continuous,} \tag{23}$$

$$\text{and absolutely continuous on each bounded interval } [0,b],\ 0 < b < +\infty; \tag{24}$$

$$ii)\ v(t) \in T(x(t)) \quad \text{for all } t \in [0,+\infty); \tag{25}$$

$$iii)\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0 \quad \text{for almost all } t \in [0,+\infty); \tag{26}$$

$$iv)\ x(0) = x_0, v(0) = v_0. \tag{27}$$

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

$$\lambda : [0,+\infty) \to (0,+\infty) \text{ is continuous,} \tag{28}$$

$$\text{and absolutely continuous on each interval } [0,b],\ 0 < b < +\infty. \tag{29}$$

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 $t \in [0,b]$

$$0 < \lambda_{b,min} \leq \lambda(t) \leq \lambda_{b,max} < +\infty. \tag{30}$$

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

2.2 Global existence and uniqueness results

System (18, 19) involves time-dependant operators $J^T_{\mu(t)}$ and $T_{\mu(t)}$. In order to establish existence results for the corresponding evolution equations, let us make precise the regularity properties of mappings $\mu \mapsto J^T_\mu x$ and $\mu \mapsto T_\mu x$.

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

$$i)\ J^T_\lambda x = J^T_\mu \left(\frac{\mu}{\lambda}x + \left(1 - \frac{\mu}{\lambda}\right) J^T_\lambda x\right); \tag{31}$$

$$ii)\ \|J^T_\lambda x - J^T_\mu x\| \leq |\lambda - \mu|\ \|T_\lambda x\|. \tag{32}$$

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

$$\|J^T_\lambda x - J^T_\mu x\| \leq |\lambda - \mu|\ \|T_\delta x\|. \tag{33}$$

Proof. i) Equality (31) is known as the resolvent equation, see [8]. Its proof is straightforward: By definition of $\xi = J^T_\lambda x$, we have

$$\xi + \lambda T\xi \ni x.$$

After multiplication by $\frac{\mu}{\lambda}$ one obtains

$$\tfrac{\mu}{\lambda}\xi + \mu T\xi \ni \tfrac{\mu}{\lambda}x.$$

By adding ξ to the two members of the above equality, one gets

$$\xi + \mu T\xi \ni \tfrac{\mu}{\lambda}x - \tfrac{\mu}{\lambda}\xi + \xi,$$

that is

$$\xi = J^T_\mu \left(\tfrac{\mu}{\lambda}x + \left(1 - \tfrac{\mu}{\lambda}\right) J^T_\lambda x\right),$$

which is the desired equality.

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

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

Using that $\lambda \mapsto \|T_\lambda x\|$ is decreasing, see ([8], Proposition 2.6), we obtain (33). □

Theorem 2.4. *Let $\lambda : [0, +\infty) \to (0, +\infty)$ be a continuous function which is absolutely continuous on each bounded interval $[0, b]$. 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 $(x(.), v(.))$ of the Cauchy problem ((20), (21) (22));*

2. setting $\mu(t) = \frac{1}{\lambda(t)}$*, the solution pair* $(x(.), v(.))$ *of ((20), (21), (22)) can be represented as*

$$x(t) = J^T_{\mu(t)}(z(t)) \tag{34}$$

$$v(t) = T_{\mu(t)}(z(t)), \tag{35}$$

where $z(.)$ *is the unique strong solution of the Cauchy problem*

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))T_{\mu(t)}(z(t)) = 0, \tag{36}$$

$$z(0) = x_0 + \mu(0)v_0. \tag{37}$$

Proof. 1) Let us first prove existence of a strong global solution of the Cauchy problem ((20), (21), (22)). To that end, we consider system (36)-(37) where T_μ is the Yosida regularization of index $\mu = \frac{1}{\lambda}$ of T

$$T_\mu = \frac{1}{\mu}\left[I - (I + \mu T)^{-1}\right] .$$

Equation (36) can be written as

$$\dot{z}(t) + \left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)}\right)\frac{1}{\lambda(t)} T_{\frac{1}{\lambda(t)}}(z(t)) = 0, \tag{38}$$

which is equivalent to

$$\dot{z}(t) = F(t, z(t)) \tag{39}$$

with

$$F(t) = \theta(t)G(t, z), \tag{40}$$

$$\theta(t) = -\left(1 + \frac{\dot{\lambda}(t)}{\lambda(t)}\right), \tag{41}$$

$$G(t, z) = \frac{1}{\lambda(t)} T_{\frac{1}{\lambda(t)}}(z). \tag{42}$$

In order to apply Cauchy-Lipschitz theorem to equation (39), let us first examine the Lipschitz continuity properties of F.

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

$$\|G(t, z_2) - G(t, z_1)\| \leq \|z_2 - z_1\|, \tag{43}$$

and, as a consequence,

$$\|F(t, z_2) - F(t, z_1)\| \leq |\theta(t)|\|z_2 - z_1\|. \tag{44}$$

Moreover, for any $0 < b < +\infty$

$$|\theta(t)| = |1 + \frac{\dot{\lambda}(t)}{\lambda(t)}| \in L^1([0, b]). \tag{45}$$

This follows from the local integrability of $\dot{\lambda}(\cdot)$ and from the fact that $\lambda(\cdot)$ is bounded away from zero on any interval $[0, b]$ with $0 < b < +\infty$.

b) Let us show that

$$\forall z \in H, \ \forall b > 0, \ F(., z) \in L^1([0, b]). \tag{46}$$

Recall that, by (30), for any $t \in [0, b]$, we have $0 < \lambda_{b,min} \leq \lambda(t) \leq \lambda_{b,max} < +\infty$. Returning to the definition (40) of F, we deduce that

$$\|F(t, z)\| \leq \left(1 + \frac{|\dot{\lambda}(t)|}{\lambda_{b,min}}\right) \frac{1}{\lambda_{b,min}} \|T_{\frac{1}{\lambda_{b,max}}} z\|. \tag{47}$$

Using again the the local integrability of $\dot{\lambda}(\cdot)$ we obtain (46).

As a general result, one can deduce from properties (44), (45), and (46) existence and uniqueness of a strong global solution of the differential equation (39), with given Cauchy data. To that end, we need to use a version of Cauchy-Lipschitz theorem relying on the integrability of $t \mapsto F(t, x)$, and involving absolutely continuous trajectories, see ([17], Proposition 6.2.1.). For the convenience of the reader, let us give the main lines of the proof using a fixed point argument. One first prove existence and uniqueness of a local solution of system ((36), (37)). Choose $b > 0$ sufficiently small such that

$$\int_0^b |\theta(t)| dt < 1. \tag{48}$$

This is always possible because, by property (45), $\theta(\cdot)$ is integrable on each bounded interval of $[0 + \infty)$.

Then introduce the Banach space $E = C([0, b], H)$ being equipped with the sup norm $\|z\|_E = \sup_{t \in [0,b]} \|z(t)\|$, and define $\Psi : C([0, b], H) \to C([0, b], H)$ by

$$\Psi(z)(t) = z_0 + \int_0^t F(s, z(s)) ds,$$

where $z_0 = z(0) = x_0 + \mu(0)v_0$ is the given Cauchy data. This integral makes sense, the mapping $t \mapsto F(t, z(t))$ being Legesgue integrable on each bounded interval $[0, b]$: First note that $t \mapsto F(t, z(t))$ is measurable because F is a Caratheodory mapping. Moreover by (44)

$$\|F(t, z(t))\| \leq \|F(t, z_0)\| + |\theta(t)| \|z(t) - z_0\| \tag{49}$$

$$\leq \|F(t, z_0)\| + |\theta(t)| \left(\|z\|_E + \|z_0\|\right). \tag{50}$$

By (46) $F(., z_0) \in L^1([0, b])$, and by (45) $|\theta(t)| \in L^1([0, b])$, which gives the result. Consider now two elements z_1 and z_2 belonging to $E = C([0, b], H)$. We have

$$\Psi(z_2)(t) - \Psi(z_1)(t) = \int_0^t \theta(s) \left(G(s, z_2(s)) - G(s, z_1(s))\right) ds$$

so that, by (43)

$$\|\Psi(z_2) - \Psi(z_1)\|_E \leq \left(\int_0^b |\theta(t)| dt\right) \|z_1 - z_2\|_E.$$

Thanks to (48), we can apply the Banach-Picard fixed point theorem, and the fixed point of Ψ is the unique solution of the ODE on $[0, b]$.
Passing from a local to a global solution is a standard argument: Take a maximal solution $z(.)$ of (39) which is defined on some interval $[0, b_{max}[$. Assume that $b_{max} < +\infty$, and show that $\dot{z} \in L^1([0, b_{max}[)$. By definition of $z(\cdot)$

$$
\begin{aligned}
\|z(t)\| &\leq \|z_0\| + \int_0^t \|F(s, z(s))\| ds \\
&\leq \|z_0\| + \int_0^t \left(\|F(s, z_0)\| + |\theta(s)| \|z(s) - z_0)\|\right) ds \\
&\leq C + \int_0^t |\theta(s)| \|z(s)\| ds.
\end{aligned}
$$

From $|\theta(\cdot)| \in L^1([0, b_{max}])$, applying Gronwall's lemma, we deduce that $z(\cdot)$ remains bounded on $[0, b_{max}[$. By (49) the mapping $t \mapsto F(t, z(t))$ belongs to $L^1([0, b_{max}])$ and, by equation (39) $\dot{z}(\cdot) \in L^1([0, b_{max}])$. Thus $z(t)$ admits a limit as t goes to b_{max}, which allows to extend the solution beyond b_{max}, a contradiction. Hence $b_{max} = +\infty$.
Let us finally notice that

$$
z(t) = z_0 + \int_0^t F(s, z(s)) ds,
$$

i.e., $z(\cdot)$ is the primitive of a locally integrable function, which by definition (2.1), $i)$ gives that $z(\cdot)$ is absolutely continuous on each bounded interval. Collecting these results together, we finally obtain the existence of a unique strong global solution $z : [0, +\infty) \mapsto H$ of system (36)-(37).

2) Define $x(.),\ v(.) : [0, +\infty) \to H$ by

$$
x(t) = J^T_{\mu(t)}(z(t)), \qquad v(t) = T_{\mu(t)}(z(t)). \tag{51}
$$

a) Let us show that $x(\cdot)$, $v(\cdot)$ are absolutely continuous on each bounded interval and satisfy system (20)-(21-(22). 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

$$
\begin{aligned}
\|J^T_{\mu_2}(z_2) - J^T_{\mu_1}(z_1)\| &\leq \|J^T_{\mu_2}(z_2) - J^T_{\mu_2}(z_1)\| + \|J^T_{\mu_2}(z_1) - J^T_{\mu_1}(z_1)\| & (52) \\
&\leq \|z_2 - z_1\| + |\mu_2 - \mu_1| \ \|T_{\mu_1} z_1\|. & (53)
\end{aligned}
$$

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 (53), 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

$$
\begin{aligned}
\|J^T_{\mu(t)}(z(t)) - J^T_{\mu(s)}(z(s))\| &\leq \|z(t) - z(s)\| + |\mu(t) - \mu(s)| \ \|T_{\mu(t)} z(t)\| & (54) \\
&\leq \|z(t) - z(s)\| + |\mu(t) - \mu(s)| \ \|T_{\frac{1}{\Lambda}}(z(t)\|. & (55)
\end{aligned}
$$

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 $z(.)$ and $\mu(.)$, we deduce that $x(t) = J^T_{\mu(t)}(z(t))$

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 $\lambda(.)$ 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 $t \in [0, +\infty)$

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

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 $v(t) = T_{\mu(t)}(z(t))$, equation (36) 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 (20)-(21).
Regarding the initial condition, let us observe that

$$\begin{aligned} z(0) &= x_0 + \mu(0)v_0, && (56)\\ &= x(0) + \mu(0)v(0), && (57) \end{aligned}$$

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 (56), after simplification, we obtain $v(0) = v_0$.

b) Let us now prove uniqueness. Suppose that

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

is a solution pair of ((20), (21), (22)). Defining $\mu(t) = \lambda(t)^{-1}$ and

$$z(t) = x(t) + \mu(t)v(t) \tag{58}$$

we conclude that $z(.)$ 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 $t \in [0, +\infty)$,

$$x(t) = (I + \mu(t)T)^{-1}(z(t)), \quad v(t) = T_{\mu(t)}(z(t)). \tag{59}$$

Therefore, differentiating (58) almost everywhere (the usual derivation rule for the product of two functions holds true), and using equation (21), we conclude that almost everywhere

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

Using $v(t) = T_{\mu(t)}(z(t))$, 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 Cauchy-Lipschitz theorem, $z(.)$ is uniquely determined and locally absolutely continuous. Thus, by (59), $x(.)$ and $v(.)$ are uniquely determined. □

Remark 2.5. *Assuming that $\lambda(.)$ is Lipschitz continuous on bounded sets, one can easily derive from equation (47) that $z(.)$ is also Lipschitz continuous on bounded sets, and by (54) the same holds true for $x(.)$ and $v(.)$.*

3 Properties of trajectories

Let us establish some properties of trajectories of system (20)-(21) which will be useful for studying their asymptotical behaviour. As a standing assumption, we assume that $\lambda : [0, +\infty) \to (0, +\infty)$ is continuous, and absolutely continuous on each bounded interval $[0, b]$. Let us recall that, by theorem 2.4, for any given Cauchy data $v_0 \in T(x_0)$, this property guarantees existence and uniqueness of a strong global solution of system (20)-(21)-(22).
From now on in this section, $x(\cdot),\ v(\cdot) : [0, +\infty) \to H$ is a strong global solution of (20)-(21). This means, in particular, that $x(\cdot)$ and $v(\cdot)$ are locally absolutely continuous from $[0, +\infty)$ to H and equation (21) is satisfied for almost every $t > 0$.

Proposition 3.1. *For almost every $t > 0$ the following properties hold:*

$$\langle \dot{x}(t), \dot{v}(t) \rangle \geq 0; \tag{60}$$

$$\langle \dot{x}(t), v(t) \rangle = -\left[\lambda(t)\|\dot{x}(t)\|^2 + \langle \dot{x}(t), \dot{v}(t) \rangle \right] \leq -\lambda(t)\|\dot{x}(t)\|^2 \leq 0; \tag{61}$$

$$\langle v(t), \dot{v}(t) \rangle = -\left[\|\dot{v}(t)\|^2 + \lambda(t)\langle \dot{x}(t), \dot{v}(t) \rangle \right] \leq -\|\dot{v}(t)\|^2 \leq 0; \tag{62}$$

$$\lambda(t)^2\|\dot{x}(t)\|^2 + \|\dot{v}(t)\|^2 \leq \|v(t)\|^2. \tag{63}$$

Proof. For almost every $t > 0$, $\dot{x}(t)$ and $\dot{v}(t)$ are well defined, thus

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

By equation (20), 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 (60).
Let us now use (21)

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

Equations (61), (62) follow by taking the inner product of both sides of (21) by $\dot{x}(t)$ and $\dot{v}(t)$ respectively, using the positivity of $\lambda(t)$ and (60). In order to obtain the last inequality, let us rewrite (21) as $\lambda(t)\dot{x}(t) + \dot{v}(t) = -v(t)$. By taking the square norm of theses two quantities, using (60), and the positivity of $\lambda(t)$, we obtain (63). $\square$

Let us enunciate some straight consequences of inequation (63).

Corollary 3.2. *The following properties hold:*

1. $t \mapsto \|v(t)\|$ *is decreasing;*
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 $\lambda(.)$ *is bounded away from* 0*, then* $t \mapsto x(t)$ *is Lipschitz continuous on* $[0, +\infty)$*.*

Proof. By (62)

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

Hence $t \mapsto \|v(t)\|$ is decreasing, which proves first item. The second item is a straight consequence of inequality (63)

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

which, combined with the decreasing property of $\|v(t)\|$, yields

$$\|\dot{v}(t)\| \leq \|v_0\|. \tag{64}$$

As a straight consequence of inequality (63) we also obtain

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

which combined with (64) yields last item. $\square$

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

Corollary 3.3. *The following properties hold:*

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

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

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 (21) 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 (61) 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 (62) of Proposition 3.1.
Items 2 and 3 follow from item 1 by integration arguments. Just notice that integration of the differential inequality $\frac{d}{dt}(\phi) + 2\phi \geq 0$ with $\phi(t) = \|v(t)\|^2$ yields $\phi(t) \geq e^{-2t}\phi(0)$, and hence $e^{-t}\|v_0\| \leq \|v(t)\|$. □

Note that item 2 of the above corollary 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}.

4 Asymptotic convergence analysis

In this section, as a standing assumption we assume that the set equilibria is non empty: $T^{-1}(0) \neq \emptyset$. The asymptotic convergence analysis relies on using the following Lyapunov functions. Suppose that

$$\hat{x} \in T^{-1}(0) \neq \emptyset \tag{65}$$

and define for any $t \geq 0$

$$g(t) := \frac{1}{2}\|x(t) - \hat{x} + \frac{1}{\lambda(t)}v(t)\|^2; \tag{66}$$

$$h(t) := \frac{1}{2}\|\lambda(t)(x(t) - \hat{x}) + v(t)\|^2; \tag{67}$$

$$u(t) = \frac{1}{2}\|x(t) - \hat{x}\|^2 + \frac{1}{\lambda(t)}\langle x(t) - \hat{x}, v(t)\rangle. \tag{68}$$

Lemma 4.1. *If $\lambda(\cdot)$ is non-increasing, then $\lim_{t\to+\infty} v(t) = 0$.*

Proof. Differentiating $h(\cdot)$, and using (21) we obtain

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

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

$$\langle x(t) - \hat{x}, v(t)\rangle \geq 0. \tag{69}$$

Using inequality (69) and the decreasing property of $\lambda(\cdot)$, we deduce that

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

After integration with respect to t of the above inequality, 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 $\|v(t)\|$ is a decreasing function of t (see Corollary 3.2), we conclude that $\lim_{t\to+\infty} v(t) = 0$. $\square$

Lemma 4.2. *Suppose that, for almost every $t > 0$*

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

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

Proof. Differentiating $u(\cdot)$ and using (21) 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 assumption $\lambda(t)+\dot\lambda(t) \geq 0$, together with the non-negativity of $\langle x(t) - \hat{x}, v(t)\rangle$ (see 69), and inequality (61), we deduce that $u(\cdot)$ is decreasing. Hence

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

which clearly implies that $\|x(t)\|$ remains bounded, with an upper bound which can be easily expressed in terms of $\|v_0\|$ and $\|x_0 - \hat{x}\|$. $\square$

Corollary 4.3. *Suppose that, for almost every $t > 0$*

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

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

Proof. By Lemma 4.1, $v(t) \to 0$ as $t \to +\infty$, and, by Lemma 4.2, $x(\cdot)$ is bounded. On the other hand, by (20), for all $t \geq 0$, $v(t) \in T(x(t))$. From the closure property of the graph of T in $w - H \times H$, see for example ([8]; Proposition 2.5), we infer that whenever $x(t_k)$ weakly converges to some x_∞, then $0 \in T(x_\infty)$. $\square$

To prove weak convergence of $x(.)$ we need additional assumptions. Note that the assumption of Lemma 4.1 can be equivalently written as

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

Theorem 4.4. *If $\lambda(\cdot)$ is bounded from above on $[0,+\infty)$ and*

$$\liminf_{t\to+\infty} \frac{\dot\lambda(t)}{\lambda(t)} > -1$$

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

Proof. Differentiating g and using (21) we have

$$\frac{d}{dt}g(t) = \left\langle \dot x(t) + \frac{1}{\lambda(t)}\dot v(t), x(t) - \hat x + \frac{1}{\lambda(t)}v(t)\right\rangle \tag{70}$$

$$-\frac{\dot\lambda(t)}{\lambda(t)^2}\left\langle v(t), x(t)-\hat x + \frac{1}{\lambda(t)}v(t)\right\rangle \tag{71}$$

$$= -\left(\frac{1}{\lambda(t)} + \frac{\dot\lambda(t)}{\lambda(t)^2}\right)\langle v(t), x(t)-\hat x + \frac{1}{\lambda(t)}v(t)\rangle \tag{72}$$

$$= -\frac{1}{\lambda(t)}\left(1+\frac{\dot\lambda(t)}{\lambda(t)}\right)\langle v(t), x(t)-\hat x + \frac{1}{\lambda(t)}v(t)\rangle. \tag{73}$$

Let us examine this last term. By (69), $\langle v(t), x(t)-\hat x\rangle \geq 0$ 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, by assumption on $\lambda(\cdot)$, there exists some $\epsilon>0$ such that for t large enough

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

Combining the two last inequalities with (73), we deduce that, for t large enough

$$\frac{d}{dt}g(t) + \varepsilon\left\|\frac{1}{\lambda(t)}v(t)\right\|^2 \leq 0. \tag{74}$$

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

$$v(t)\to 0 \text{ as } t\to+\infty. \tag{75}$$

Define

$$\psi(t) := \left\|\frac{1}{\lambda(t)}v(t)\right\|^2. \tag{76}$$

Since $g\geq 0$, from (74) we obtain

$$\psi(\cdot)\in L^1([0,+\infty)). \tag{77}$$

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

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

Since $\psi(\cdot) \in L^1([0,+\infty))$, it follows that

$$\left(\frac{d}{dt}\psi(t)\right)^+ \in L^1([0,+\infty)).$$

Combining this property with non-negativity of $\psi(\cdot)$, by a standard argument we infer that $\lim_{t\to+\infty}\psi(t)$ exists. Using again $\psi(\cdot) \in L^1([0,+\infty))$, we finally obtain

$$\psi(t) \to 0 \;\text{ as }\; t \to +\infty. \tag{78}$$

Let us now return to g. By (74), $t \mapsto g(t)$ is decreasing. Hence, there exists $\lim_{t\to+\infty} g(t)$. By (78), $\|v(t)\|/\lambda(t) \to 0$ as $t \to +\infty$. Since

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

we conclude that, for any $\hat{x} \in T^{-1}(0)$, there exists $\lim_{t\to+\infty}\|x(t)-\hat{x}\|$. On the other hand, from $v(t) \in T(x(t))$, $v(t) \to 0$ as $t \to +\infty$ (see (75)), and the closedness property of T in $w-H \times s-H$, we have that every weak limit of $x(t)$ is a zero of T. We now rely on Opial's Lemma (see [22]), that we recall below for convenience of the reader. Taking $S = T^{-1}(0)$ in Opial's lemma, we conclude that $x(t)$ converges weakly to a zero of T, as $t \to +\infty$.

Lemma 4.5. *Let S be a non empty subset of H and $x : [0,+\infty) \to H$ a map. Suppose that*

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

(ii) *every weak limit point of the map x belongs to S.*

Then

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

□

5 T subdifferential. Links with convex optimization

Let us now suppose that $T = \partial f$ is the subdifferential of a convex lower semicontinuous proper function $f : H \to \mathbb{R}\cup\{+\infty\}$. By a classical result, T is a maximal monotone operator. The system ((20), (21)) (22)) reads as follows

$$v(t) \in \partial f(x(t)), \tag{79}$$

$$\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) = 0, \tag{80}$$

$$x(0) = x_0,\; v(0) = v_0. \tag{81}$$

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

Theorem 5.1. *Suppose that $T = \partial f$, where $f : H \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous proper function. Then, for any strong global solution trajectory $t \in [0, +\infty) \mapsto (x(t), v(t)) \in H \times H$ of system (79), (80), the following hold:*
i) the function

$$[0, +\infty) \to \mathbb{R}, \quad t \mapsto f(x(t))$$

is Lispchitz continuous, and for almost every $t > 0$

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

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 $v(t) \to 0$ as $t \to +\infty$.

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

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

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

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

Therefore

$$\langle x_2 - x_1, v_1 \rangle \leq f(x_2) - f(x_1) \leq \langle x_2 - x_1, v_2 \rangle. \tag{82}$$

By Corollary 3.2 (1.), we know that $t \mapsto \|v(t)\|$ is a decreasing function. From (82) we deduce that

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

Hence, $t \mapsto f(x(t))$ is globally Lipschitz continuous on $[0, +\infty)$.
Suppose now that $x(\cdot)$ is differentiable at t_1. Let us divide (82) by $t_2 - t_1 > 0$ and take the limit $t_2 \to t_1^+$. 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 every $t > 0$

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

Replacing $v(t)$ by $v(t) = -(\lambda(t)\dot{x}(t) + \dot{v}(t))$, as given by (80), 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 (60).

ii) The function $t \mapsto f(x(t))$ is Lipschitz continuous, with a derivative which is less or equal than zero for almost every t. This classically implies that $f(x(\cdot))$ is non increasing.

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

$$\phi_y(t) = \Big[f(y) - (f(x(t)) + \langle y - x(t), v(t)\rangle)\Big] + \frac{\lambda(t)}{2}\|y - x(t)\|^2. \tag{83}$$

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 Lispchitz continuous and, in particular, absolutely continuous on compact sets.
Using item *i*) and (80) we deduce that for almost every $t > 0$

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

which combined with convexity of $f(\cdot)$, and the assumption on $\lambda(\cdot)$ being non increasing, yields (for almost every $t > 0$)

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

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 $t \geq 0$

$$\begin{aligned}-\phi_y(0) \leq& \phi_y(t) - \phi_y(0) \\ &\leq \int_0^t f(y) - f(x(s))\, ds \leq t[f(y) - f(x(t))].\end{aligned} \tag{84}$$

Hence, for any $t > 0$

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

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

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

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

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

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

$$\int_0^{+\infty} \lambda(t)\|\dot{x}(t)\|^2 \, dt < +\infty. \tag{85}$$

Since $\lambda(\cdot)$ is assumed to be non increasing, we have

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

which, combined with (85), yields $\lambda(\cdot)\|\dot{x}(\cdot)\| \in L^2([0,+\infty))$. Then, use item 3) of Corollary 3.3, (80), and triangle inequality, to obtain $\|v(\cdot)\| \in L^2([0,+\infty))$. Since $\|v(\cdot)\|$ is a decreasing function of t, this immediately implies $v(t) \to 0$ as $t \to +\infty$. $\square$

Remark Let us now assume that $S = (\partial f)^{-1}(0) \neq \emptyset$. Let us consider a trajectory of system (79)-(80)-(81) with Cauchy data x_0 and v_0. By taking y equal to the projection of x_0 onto S in (83), and using (84), we obtain

$$\int_0^{+\infty} (f(x(s)) - \inf_H f)ds \leq C, \tag{86}$$

and

$$f(x(t)) - \inf_H f \leq \frac{C}{t}, \tag{87}$$

where

$$C = \phi_y(0) \leq \frac{\lambda(0)}{2}\|y - x_0\|^2 - \langle y - x_0, v_0\rangle \tag{88}$$

$$\leq \frac{\lambda(0)}{2}\mathrm{dist}(x_0, S)^2 + \|v_0\|\mathrm{dist}(x_0, S). \tag{89}$$

6 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 sections 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

$$v(t) \in T(x(t)), \tag{90}$$

$$\lambda\dot{x}(t) + \dot{v}(t) + v = 0, \tag{91}$$

$$x(0) = x_0, \; v(0) = v_0 \in T(x_0). \tag{92}$$

Relying on the results of section 4, theorem 4.4, let us summarize the asymptotic behaviour as $t \to +\infty$ of trajectories of system ((90), (91)) in the following statement.

Theorem 6.1. *Let us assume that $T^{-1}(0)$ is non-empty. Then, for any trajectory $x(\cdot)$, $v(\cdot)$ of system ((90), (91)) the following properties hold:*
i) $v(t) \to 0$ strongly in H as $t \to +\infty$. 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 $t \to +\infty$, with $\bar{x} \in T^{-1}(0)$.

Let us show an other approach to the asymptotic analysis of system ((90), (91)) which is based on the equivalent formulation

$$\dot{z}(t) + \ \mu T_\mu(z(t)) = 0, \tag{93}$$

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

$$x(t) = J_\mu^T(z(t)), \tag{94}$$

$$v(t) = T_\mu(z(t)). \tag{95}$$

As a key ingredient in the asymptotic analysis of (93) we will use that the operator T_μ is μ-cocoercive. An operator $A : H \to H$ is said to be θ-cocoercive for some positive constant θ if for all x, y belonging to H

$$\langle Ay - Ax, y - x \rangle \geq \theta \|Ay - Ax\|^2. \tag{96}$$

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 theorem). In our context, this notion plays an important role because of the following property:

Proposition 6.2. *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.*

Proof. By definition of J_μ^T and T_μ one has $x = J_\mu^T x + \mu T_\mu x$ and $T_\mu x \in T(J_\mu^T x)$. By monotonicity of T one infers

$$\begin{aligned}\langle T_\mu y - T_\mu x, y - x \rangle &= \langle T_\mu y - T_\mu x, J_\mu^T y - J_\mu^T x \rangle + \mu \langle T_\mu y - T_\mu x, T_\mu y - T_\mu x \rangle \\ &\geq \mu \|T_\mu y - T_\mu x\|^2.\end{aligned}$$

Multiplying this last inequality by $\mu > 0$, one obtains

$$\langle \mu T_\mu y - \mu T_\mu x, y - x \rangle \geq \|\mu T_\mu y - \mu T_\mu x\|^2,$$

which expresses that 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 in two particular important situations, namely the subdifferential of a closed convex function case, and the cocoercive case. For convenience of the reader we give a self-contained proof of this last result.

Proposition 6.3. *Let $T : H \to H$ 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 $t \to +\infty$
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.*

Proof. Let $z(\cdot)$ be a trajectory of the classical differential equation $\dot{z}(t)+T(z(t)) = 0$. Recalling that T is Lipschitz continuous, by Cauchy-Lipschitz theorem, $z(\cdot)$ is defined on $[0, +\infty)$. Take an arbitrary $\hat{z} \in T^{-1}(0)$ and consider the function

$$h(t) := \frac{1}{2}\|z(t) - \hat{z}\|^2.$$

Differentiating h, and using $\dot{z}(t) = -T(z(t))$, we obtain

$$\dot{h}(t) = \langle z(t) - \hat{z}, \dot{z}(t)\rangle \tag{97}$$

$$= -\langle z(t) - \hat{z}, T(z(t))\rangle. \tag{98}$$

Equivalently

$$\dot{h}(t) + \langle T(z(t)) - T\hat{z}, z(t) - \hat{z}\rangle = 0.$$

Then use the θ-cooercivity of T, for some $\theta > 0$, to obtain

$$\dot{h}(t) + \theta\|T(z(t))\|^2 \leq 0.$$

Equivalently

$$\dot{h}(t) + \theta\|\dot{z}(t)\|^2 \leq 0. \tag{99}$$

Hence, for any $\hat{z} \in T^{-1}(0)$, the limit of $\|z(t) - \hat{z}\|$ exists, as $t \to +\infty$. To complete the proof with the help of Opial lemma, we just need to prove that every weak limit point of the trajectory belongs to $T^{-1}(0)$. This will follow from the strong convergence property of $\dot{z}(t)$ in H to zero and the demi-closure property of T. Strong convergence of $\dot{z}(t)$ to zero is a direct consequence of (99) which implies that $\dot{z}(.)$ belongs to $L^2[0, +\infty)$, and of the decreasing property of $\dot{z}(t)$ (apply the contraction property of the semigroup generated by T to obtain that, for any $0 < s < t$ and any $h > 0$, $\|z(t + h) - z(t)\| \leq \|z(s + h) - z(s)\|$, then divide by h and let h go to zero). $\square$

We can now give a proof of theorem 6.1 which is based on the cocoercive property: By Proposition 6.3, using equation (93), $\dot{z}(t) + \mu T_\mu(z(t)) = 0$, and the cocoercive property of μT_μ, we deduce that $z(t)$ weakly converges to some element $\bar{z} \in T_\mu^{-1}(0) = T^{-1}(0)$ and $\dot{z}(t)$ strongly converges to zero, as $t \to +\infty$. From $v(t) = T_\mu(z(t)) = -\frac{1}{\mu}\dot{z}(t)$, we deduce that $v(t)$ strongly converges 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 $z(.)$).

Strong convergence of the trajectories requires further information about T. Regarding this last property, demiregularity of operator T plays a key role:

Definition 6.4. *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:*

$$\begin{cases} x_n \rightharpoonup x & \textit{weakly} \\ z_n \to z & \textit{strongly} \end{cases} \quad \Rightarrow \quad x_n \to x \quad \textit{strongly}\ . \tag{100}$$

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

Proposition 6.5. *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 $f \in \Gamma_0(H)$ and the lower level sets of f 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. *$T : H \to H$ is single-valued with a single-valued continuous inverse.*

Then T is demiregular.

We can now state an asymptotic strong convergence result for trajectories of system ((20), (21)).

Theorem 6.6. *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 ((20), (21)) the following properties hold:
i) $x(t) \to \bar{x}$ strongly in H as $t \to +\infty$, with $\bar{x} \in T^{-1}(0)$;
ii) $v(t) \to 0$ strongly in H as $t \to +\infty$.

Proof. By theorem 6.1, we already know that $x(t) \rightharpoonup \bar{x}$ weakly in H as $t \to +\infty$, with $\bar{x} \in T^{-1}(0)$ and that $v(t) \to 0$ strongly in H as $t \to +\infty$. 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 6.1, we have $v(t) \to 0$ strongly in H, and $x(t) \rightharpoonup$ weakly in H. Demiregularity of T implies at once that $x(t) \to$ strongly in H as $t \to +\infty$.
b) Let us now suppose that $T^{-1}(0)$ has a non empty interior. The following equivalences hold

$$Tz \ni 0 \Leftrightarrow z + \mu Tz \ni z \Leftrightarrow J_\mu^T(z) = z \Leftrightarrow \mu T_\mu(z) = 0. \tag{101}$$

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

Remark In the subdifferential case, an alternative proof of theorem 5.1 in the case λ constant, would consist 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 $f_\mu(z(t)) \to \inf_H f_\mu = \inf_H f$. 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 $t \to +\infty$.

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

In this section, we discuss the case

$$\lambda(t) = \lambda_0 e^{-t}, \tag{102}$$

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 4.3, it follows that the trajectory $x(\cdot)$ is bounded, $v(t)$ converges to 0 as $t \to +\infty$, and all weak limits of $x(\cdot)$ are zeroes 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 (11), 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 equation (36), 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

$$x(t) = J^T_{e^t/\lambda_0}(z(0)), \qquad v(t) = T_{e^t/\lambda_0}(z(0)). \tag{103}$$

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

Proposition 7.1. *Let $\lambda(\cdot)$ be given by (102). 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, for any $t \geq 0$ the following properties hold:*
i) $\|x(t) - (x_0 + \lambda_0^{-1} v_0)\| \leq \|x_0^* - (x_0 + \lambda_0^{-1} v_0)\|$;

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

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

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 equation (103) 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 $x(t) = J^T_{e^t/\lambda_0}(z_0)$, which equivalently be written as

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

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

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

□

Note that $\|v(t)\| \leq c\, e^{-t}$, which, as an asymptotical behavior, is almost as good as “pure” Newton.

8 Examples

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

8.1 Linear monotone operators

Given $a, b > 0$, let $T = \nabla f$, with $f : \mathbb{R}^3 \to \mathbb{R}$ 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 (20), (21), (22) 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 (20), (21), (22) 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 the Newton direction term, which makes trajectories close to straight lines.

8.2 Discontinuous monotone operators

a) Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$, $T = \partial f$ and $x_0 = 1$. Then, the solution of system (20), (21), (22) with $\lambda > 0$ constant is given by

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

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

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

System (20), (21), (22) with λ constant $T = \partial f$ and $x_0 = 2$ is

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

8.3 Antisymmetric linear operators

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

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

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 (20), (21), (22) can be formulated as

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

Integration of this system yields

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

which clearly implies $x(t) \to 0$ as $t \to +\infty$. 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).

9 Application: A Levenberg-Marquardt algorithm for convex minimization

In this section, H is a real Hilbert space and $f : H \to \mathbb{R}$ 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

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

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

One may consult [13], [21], 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 analysis of Quasi-Newton methods required boundedness of level sets and where 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:

$$x_{k+1} = x_k + t_k s_k, \quad s_k = -(\nabla^2 f(x_k) + \lambda I)^{-1} \nabla f(x_k). \tag{104}$$

We assume that each t_k is chosen as follows: we pick some $\beta \in (0, 1/2)$ and

$$t_k = \max\left\{t \in \{1, 1/2, 1/4 \ldots\} \mid f(x_k + t s_k) \leq f(x_k) + \beta t \langle s_k, \nabla f(x_k)\rangle\right\}. \tag{105}$$

Our aim is to prove the following new result.

Theorem 9.1. *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 (104)-(105) converges weakly to a minimizer of f.*

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

Proposition 9.2. *Given $x \in H$, set $s = -\left(\nabla^2 f(x) + \lambda I\right)^{-1} \nabla f(x)$. 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

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

By definition of s we have

$$\begin{aligned}\left\langle \nabla^2 f(x)s, s\right\rangle &= \left\langle \left(\nabla^2 f(x) + \lambda I\right) s, s\right\rangle - \lambda\|s\|^2 &(106)\\ &= -\left\langle \nabla f(x), s\right\rangle - \lambda\|s\|^2. &(107)\end{aligned}$$

Combining the above equations we conclude that

$$f(x+ts) \le f(x) + \left(t - \frac{t^2}{2}\right)\langle s, \nabla f(x)\rangle + \frac{t^2\|s\|^2}{2}\left[-\lambda + \frac{tL\|s\|}{3}\right]. \tag{108}$$

On the other hand, by equation (107) and convexity of f, we have $\langle s, \nabla f(x)\rangle \le 0$. Since $t \in [0,1]$, this immediately implies

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

Combining these two last inequalities gives the desired conclusion. □

Proposition 9.3. *If* $t_k < 1$ *then*

$$\frac{1}{2} \ge t_k \ge \frac{3\lambda}{2L\|s_k\|}, \qquad \textit{and} \qquad \|s_k\| \ge \frac{3\lambda}{L}.$$

Proof. If $t \in [0,1]$ and $t \le \frac{3\lambda}{L\|s_k\|}$, by using Proposition 9.2 we conclude that

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

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

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

□

End of the proof of theorem 9.1. By (104)-(105), for all $k \in \mathbb{N}$

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

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

$$f(x^*) \le f(x_0) - \sum_{k=0}^{+\infty}\beta t_k\lambda\|s_k\|^2. \tag{109}$$

Define

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

By Proposition 9.3

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

As a consequence, I is finite. On the other hand, from (109), 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 (104) and Taylor formula, we easily deduce that

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

Therefore $\sum_k r_k < +\infty$, and 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), 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).

[24] A. G. Ramm, *Dynamical systems method for solving operator equations*, Mathematics in Science and Engineering, **208**, Elsiever, 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.