Levenberg-Marquardt dynamics associated to variational inequalities

Ernö Robert Csetnek
University of Vienna

joint work with
Radu Ioan Boţ (University of Vienna)

GDO 2018: Games, Dynamics and Optimization
Vienna, Austria
March 13-15, 2018

Motivation

For the problem

$$\text{find } x \in \mathcal{H} \text{ such that } Tx = 0$$

where

- $T : \mathcal{H} \to \mathcal{H}$ is a C^1 operator
- $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a real Hilbert space

Newton method: in general ill-posed

$$T(x_n) + T'(x_n)(x_{n+1} - x_n) = 0 \ \forall n \geq 0$$

Levenberg-Marquardt method: regularization procedure

$$T(x_n) + \left(\lambda_n \operatorname{Id} + T'(x_n)\right) \left(\frac{x_{n+1} - x_n}{\Delta t_n}\right) = 0 \ \forall n \geq 0$$

Attouch-Svaiter (2011): the Levenberg-Marquardt algorithm can be seen as a **time discretization** of the dynamical system

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

for approaching the inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Tx,$$

where $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is a (set-valued) maximally monotone operator

Remark (optimization problems) In case

- for all $x \in \mathcal{H}$
  $Tx = \partial f(x) = \{u \in \mathcal{H} : f(y) \geq f(x) + \langle u, y - x \rangle \ \forall y \in \mathcal{H}\}$
- $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lsc (lower semicontinuous),

we have

$$0 \in \partial f(\overline{x}) \Leftrightarrow \overline{x} \text{ is an optimal solution for } \min_{x \in \mathcal{H}} f(x)$$

Abbas-Attouch-Svaiter (2014): investigated the dynamical system

$$\begin{cases} v(t) \in \partial\Phi(x(t)) \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\Theta(x(t)) = 0 \\ x(0) = x_0, v(0) = v_0 \in \partial\Phi(x_0), \end{cases}$$

in connection with the optimization problem

$$\min_{x\in\mathcal{H}}\{\Phi(x) + \Theta(x)\},$$

where

- $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lsc
- $\Theta : \mathcal{H} \to \mathbb{R}$ is convex and smooth

Remark (optimality conditions):

$\overline{x}$ is an optimal solution for $\min_{x\in\mathcal{H}}\{\Phi(x)+\Theta(x)\} \Leftrightarrow 0 \in \partial\Phi(\overline{x})+\nabla\Theta(\overline{x})$

Our starting point

In connection with the bilevel optimization problem

$$\min_{x \in \operatorname{argmin} \Psi} \{\Phi(x) + \Theta(x)\},$$

where

- $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lsc
- $\Psi, \Theta : \mathcal{H} \to \mathbb{R}$ are convex and smooth

we investigate the dynamical system

$$\begin{cases} v(t) \in \partial\Phi(x(t)) \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\Theta(x(t)) + \beta(t)\nabla\Psi(x(t)) = 0 \\ x(0) = x_0, v(0) = v_0 \in \partial\Phi(x_0). \end{cases}$$

Remark (optimality conditions): in case Φ continuous

$$0 \in \partial\Phi(\overline{x}) + \nabla\Theta(\overline{x}) + N_{\operatorname{argmin} \Psi}(\overline{x}), \text{ that is}$$

find $\overline{x} \in \operatorname{argmin} \Psi$, $\overline{v} \in \partial\Phi(\overline{x}) : \ \langle \overline{v} + \nabla\Theta(\overline{x}), y - \overline{x} \rangle \geq 0 \ \forall y \in \operatorname{argmin} \Psi$

$$\min_{x \in \operatorname{argmin} \Psi} \{\Phi(x) + \Theta(x)\} \quad (*)$$

Conditions on the functions involved:

(H_Ψ) $\Psi : \mathcal{H} \to [0, +\infty)$ is convex, smooth, $\nabla\Psi$ is Lipschitz continuous and $\operatorname{argmin} \Psi = \Psi^{-1}(0) \neq \emptyset$;

(H_Θ) $\Theta : \mathcal{H} \to \mathbb{R}$ is convex, smooth, $\nabla\Theta$ is Lipschitz continuous;

(H_Φ) $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is convex, lsc and $S \neq \emptyset$

S is the set of optimal solutions of the optimization problem $(*)$.

$$\begin{cases} v(t) \in \partial\Phi(x(t)) \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\Theta(x(t)) + \beta(t)\nabla\Psi(x(t)) = 0 \\ x(0) = x_0, v(0) = v_0 \in \partial\Phi(x_0) \end{cases}$$

Definition We say that the pair (x, v) is a **solution of the dynamical system**, if:

(i) $x, v : [0, +\infty) \to \mathcal{H}$ are locally absolutely continuous functions;

(ii) $v(t) \in \partial\Phi(x(t))$ for every $t \in [0, +\infty)$;

(iii) $\lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\Theta(x(t)) + \beta(t)\nabla\Psi(x(t)) = 0$ for almost every $t \in [0, +\infty)$;

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

Existence and uniqueness of the solutions under:

(H_λ^1) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous;

(H_β^1) $\beta : [0, +\infty) \to [0, +\infty)$ is locally integrable.

Sketch of the proof:

$$v(t) \in \partial\Phi(x(t)) \Leftrightarrow z(t) \in x(t) + \mu(t)\partial\Phi(x(t)) = (\mathrm{Id} + \mu(t)\partial\Phi)(x(t)),$$

where

$$z(t) = x(t) + \mu(t)v(t) \text{ and } \mu(t) = (\lambda(t))^{-1} \quad \forall t \geq 0.$$

Thus

$$x(t) = (\mathrm{Id} + \mu(t)\partial\Phi)^{-1}(z(t)) = J_{\mu(t)\partial\Phi}(z(t)) \text{ (\textbf{resolvent operator})}$$

$$v(t) = \frac{1}{\mu(t)}(\mathrm{Id} - J_{\mu(t)\partial\Phi})(z(t)) = (\partial\Phi)_{\mu(t)}(z(t)) \text{ (\textbf{Yosida regularization})}$$

Moreover,

$$\begin{aligned}\dot{z}(t) &= \dot{x}(t) + \dot{\mu}(t)v(t) + \mu(t)\dot{v}(t) \\ &= \dot{\mu}(t)v(t) - \mu(t)v(t) - \mu(t)\nabla\Theta(x(t)) - \beta(t)\mu(t)\nabla\Psi(x(t))\end{aligned}$$

Hence: if (x, v) is a solution of the dynamical system, then

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))(\partial\Phi)_{\mu(t)}(z(t))$$

$$+\mu(t)\nabla\Theta(J_{\mu(t)\partial\Phi}(z(t))) + \beta(t)\mu(t)\nabla\Psi(J_{\mu(t)\partial\Phi}(z(t))) = 0.$$

Vice versa, if z is a solution of the above dynamical system, then (x, v) is a solution of the original system, where:

$$x(t) = J_{\mu(t)\partial\Phi}(z(t))$$

$$v(t) = (\partial\Phi)_{\mu(t)}(z(t)).$$

It is sufficient to prove existence and uniqueness for

$$\begin{cases} \dot{z}(t) = f(t, z(t)) \\ z(0) = z_0, \end{cases}$$

where $z_0 = x_0 + \mu(0)v_0$ and $f : [0, +\infty) \times \mathcal{H} \to \mathcal{H}$ is defined by

$$f(t, w) = (\dot{\mu}(t) - \mu(t))(\partial\Phi)_{\mu(t)}(w) - \mu(t)\nabla\Theta\Big(J_{\mu(t)\partial\Phi}(w)\Big) -$$

$$\beta(t)\mu(t)\nabla\Psi\Big(J_{\mu(t)\partial\Phi}(w)\Big).$$

We prove

- $\|f(t, w_1) - f(t, w_2)\| \leq \left(1 + \frac{|\dot{\lambda}(t)|}{\lambda(t)} + \frac{L_{\nabla\Theta}}{\lambda(t)} + L_{\nabla\Psi}\frac{\beta(t)}{\lambda(t)}\right)\|w_1 - w_2\|$
- $L_f : [0, +\infty) \to \mathbb{R}, L_f(t) = 1 + \frac{|\dot{\lambda}(t)|}{\lambda(t)} + \frac{L_{\nabla\Theta}}{\lambda(t)} + L_{\nabla\Psi}\frac{\beta(t)}{\lambda(t)},$
- $L_f(\cdot) \in L^1([0, b])$ for any $0 < b < +\infty$.
- $\forall w \in \mathcal{H},\ \forall b > 0,\quad f(\cdot, w) \in L^1([0, b], \mathcal{H})$.
- apply the Cauchy-Lipschitz Theorem, see Haraux 1991

- By considering the **time discretization** $\dot{z}(t) \approx \frac{z_{n+1}-z_n}{h_n}$ of the above dynamical system
- and by taking μ constant,

from

$$\dot{z}(t) + (\mu(t) - \dot{\mu}(t))(\partial\Phi)_{\mu(t)}(z(t))$$
$$+\mu(t)\nabla\Theta(J_{\mu(t)\partial\Phi}(z(t))) + \beta(t)\mu(t)\nabla\Psi(J_{\mu(t)\partial\Phi}(z(t))) = 0$$

we obtain the iterative scheme

$$(\forall n \geq 0) \begin{cases} x_n = \operatorname{prox}_{\mu\Phi}(z_n) \ \ (= (\mathrm{Id} + \mu\partial\Phi)^{-1}(z_n)) \\ z_{n+1} = (1 - h_n)z_n + h_n(x_n - \mu\nabla\Theta(x_n) - \mu\beta_n\nabla\Psi(x_n)). \end{cases}$$

For $h_n = 1$ this yields the following algorithm

$$(\forall n \geq 0) \ x_{n+1} = \operatorname{prox}_{\mu\Phi}\left(x_n - \mu\nabla\Theta(x_n) - \mu\beta_n\nabla\Psi(x_n)\right).$$

The convergence of the above algorithm has been investigated by **Attouch-Czarnecki-Peypouquet 2011** in case $\Theta = 0$.

Asymptotic analysis

(H_λ^2) $\lambda : [0,+\infty) \to (0,+\infty)$ is locally absolutely continuous and $\dot\lambda(t) \le 0$ for almost every $t \in [0,+\infty)$;

(H_β^2) $\beta : [0,+\infty) \to (0,+\infty)$ is measurable and bounded from above on each interval $[0,b]$, $0 < b < +\infty$;

$$(H)\ \forall p \in \operatorname{ran} N_{\operatorname{argmin}\Psi}\ \int_0^{+\infty} \beta(t)\Big[\Psi^*\Big(\frac{p}{\beta(t)}\Big) - \sigma_{\operatorname{argmin}\Psi}\Big(\frac{p}{\beta(t)}\Big)\Big]dt < +\infty$$

$$(\widetilde{H})\ \partial(\Phi + \Theta + \delta_{\operatorname{argmin}\Psi}) = \partial\Phi + \nabla\Theta + N_{\operatorname{argmin}\Psi},$$

where

- $\Psi^*(p) = \sup_{x\in\mathcal{H}}\{\langle p,x\rangle - \Psi(x)\}\ \forall p \in \mathcal{H}$;
- $\sigma_{\operatorname{argmin}\Psi}(p) = \sup_{x\in\operatorname{argmin}\Psi}\langle p,x\rangle$ for all $p \in \mathcal{H}$;
- $\delta_{\operatorname{argmin}\Psi} : \mathcal{H} \to \mathbb{R}\cup\{+\infty\}$ takes the value 0 on the set $\operatorname{argmin}\Psi$ and $+\infty$, otherwise.

- The condition $\dot{\lambda}(t) \leq 0$ appear also in **Attouch-Svaiter 2011**
- $\forall p \in \operatorname{ran} N_{\operatorname{argmin} \Psi} \int_0^{+\infty} \beta(t)[\Psi^*(\frac{p}{\beta(t)}) - \sigma_{\operatorname{argmin} \Psi}(\frac{p}{\beta(t)})]dt < +\infty;$
  has been introduced by **Attouch-Czarnecki 2010** for

$$0 \in \dot{x}(t) + \partial\Phi(x(t)) + \beta(t)\partial\Psi(x(t)),$$

in connection with the optimization problem

$$\inf_{x \in \operatorname{argmin} \Psi} \Phi(x).$$

Example: $\psi(x) = \frac{1}{2} \inf_{y \in C} \|x - y\|^2$, for a nonempty, convex and closed set $C \subseteq \mathcal{H}$. Then (H) holds if and only if

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

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

The condition

$$(\widetilde{H})\ \partial(\Phi + \Theta + \delta_{\operatorname{argmin}\Psi}) = \partial\Phi + \nabla\Theta + N_{\operatorname{argmin}\Psi}$$

holds if

$$(H')\quad 0 \in \operatorname{sqri}(\operatorname{dom}\Phi - \operatorname{argmin}\Psi),$$

where

$\operatorname{sqri} M := \{x \in M : \cup_{\lambda>0}\lambda(M - x)$ is a closed linear subspace of $\mathcal{H}\}$.

The condition (H') is fulfilled, if

- Φ is continuous at a point in $\operatorname{dom}\Phi \cap \operatorname{argmin}\Psi$
- or $\operatorname{int}(\operatorname{argmin}\Psi) \cap \operatorname{dom}\Phi \neq \emptyset$

For $z \in S$ and $p \in N_{\operatorname{argmin} \Psi}(z)$ such that $-p - \nabla\Theta(z) \in \partial\Phi(z)$, define $g_z, h_z : [0, +\infty) \to [0, +\infty)$ as

$$g_z(t) = \Phi(z) - \Phi(x(t)) + \langle v(t), x(t) - z \rangle$$

and

$$h_z(t) = \Theta(z) - \Theta(x(t)) + \langle \nabla\Theta(x(t)), x(t) - z \rangle.$$

Lemma:

- $\exists \lim_{t \to +\infty} \left(\frac{\lambda(t)}{2} \|x(t) - z\|^2 + g_z(t) \right) \in [0, +\infty)$;
- $\int_0^{+\infty} \beta(t) \Psi(x(t)) dt < +\infty$;
- $\exists \lim_{t \to +\infty} \int_0^t \Big((\Phi + \Theta)(x(s)) - (\Phi + \Theta)(z) \Big) ds \in \mathbb{R}$;

The key relation needed in the proof:

$$\frac{d}{dt}\left(\frac{\lambda(t)}{2}\|x(t)-z\|^2+g_z(t)\right)+\beta(t)\left(-\Psi^*\left(\frac{p}{\beta(t)}\right)+\sigma_{\operatorname{argmin}\Psi}\left(\frac{p}{\beta(t)}\right)\right)$$
$$\leq \frac{d}{dt}\left(\frac{\lambda(t)}{2}\|x(t)-z\|^2+g_z(t)\right)+\beta(t)\Psi(x(t))+\langle -p, x(t)-z\rangle$$
$$\leq \frac{d}{dt}\left(\frac{\lambda(t)}{2}\|x(t)-z\|^2+g_z(t)\right)+(\Phi+\Theta)(x(t))-(\Phi+\Theta)(z)+\beta(t)\Psi(x(t)$$
$$\leq 0$$

and **Lemma (Brézis)** Suppose that

- $f:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ proper, convex and lower semicontinuous
- $x,\dot{x}\in L^2([0,T],\mathbb{R}^m)$
- $\xi(t)\in\partial f(x(t))$ for almost every $t\in[0,T]$
- $\xi\in L^2([0,T],\mathbb{R}^m)$

Then
$$\frac{d}{dt}f(x(t))=\langle\dot{x}(t),h\rangle\ \forall h\in\partial f(x(t)).$$

(H^3_λ) $\lambda : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous,

$\dot{\lambda}(t) \leq 0$ for almost every $t \in [0, +\infty)$ and $\lim_{t \to +\infty} \lambda(t) > 0$;

(H^3_β) $\beta : [0, +\infty) \to (0, +\infty)$ is locally absolutely continuous,

$0 \leq \dot{\beta}(t) \leq k\beta(t)$ a.e. $t \in [0, +\infty)$ and $\lim_{t \to +\infty} \beta(t) = +\infty$.

Lemma

- $x(t)$ is bounded
- $\lim_{t \to +\infty} \Psi(x(t)) = 0$.

Idea: for $E_1 : [0, +\infty) \to \mathbb{R}$ defined for every $t \in [0, +\infty)$ by

$$E_1(t) = \frac{(\Phi + \Theta)(x(t))}{\beta(t)} + \Psi(x(t))$$

we derive

$$\dot{E}_1(t) \leq -\frac{\dot{\beta}(t)}{\beta^2(t)} \inf_{t \geq 0} (\Phi + \Theta)(x(t)).$$

$$\min_{x \in \operatorname{argmin} \Psi} \{\Phi(x) + \Theta(x)\} \quad (*)$$

$$\begin{cases} v(t) \in \partial\Phi(x(t)) \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + \nabla\Theta(x(t)) + \beta(t)\nabla\Psi(x(t)) = 0 \\ x(0) = x_0, v(0) = v_0 \in \partial\Phi(x_0). \end{cases}$$

Theorem

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

(ii) $\dot{x} \in L^2([0, +\infty); \mathcal{H})$;

(iii) $\langle \dot{x}, \dot{v} \rangle \in L^1([0, +\infty))$;

(iv) $(\Phi + \Theta)(x(t))$ **converges to the optimal objective value of** $(*)$ **as** $t \to +\infty$**;**

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

(vi) $x(t)$ **converges weakly to an optimal solution of** $(*)$ **as** $t \to +\infty$**.**

(vii) if $\Phi + \Theta$ **is strongly convex**, then $x(t)$ **converges strongly** to the unique optimal solution of $(*)$ as $t \to +\infty$.

Open questions/perspective

- derivation of **second order Levenberg-Marquardt dynamics** and their asymptotic analysis
- nonconvex optimization problems
- generalizations to monotone inclusions (like in **Attouch-Svaiter 2011**):

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in A\overline{x} + B\overline{x} + N_{\text{zer}\, C}(\overline{x}),$$

where
  - $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator
  - $B, C : \mathcal{H} \to \mathcal{H}$ are cocoercive (or monotone and Lipschitz continuous).

The dynamical system to be investigated should be

$$\begin{cases} v(t) \in A(x(t)) \\ \lambda(t)\dot{x}(t) + \dot{v}(t) + v(t) + B(x(t)) + \beta(t)C(x(t)) = 0. \end{cases}$$

References

[1] H. Attouch, M.-O. Czarnecki, *Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects*, Journal of Differential Equations 262(3), 2745–2770, 2017

[2] H. Attouch, M.-O. Czarnecki, J. Peypouquet, *Prox-penalization and splitting methods for constrained variational problems*, SIAM Journal on Optimization 21(1), 149–173, 2011

[3] H. Attouch, B.F. Svaiter, *A continuous dynamical Newton-like approach to solving monotone inclusions*, SIAM Journal on Control and Optimization 49(2), 574–598, 2011

[4] R.I. Boț, E.R. Csetnek, *Forward-backward and Tseng's type penalty schemes for monotone inclusion problems*, Set-Valued and Variational Analysis 22, 313–331, 2014

[5] R.I. Boț, E.R. Csetnek, *Levenberg-Marquardt dynamics associated to variational inequalities*, Set-Valued and Variational Analysis 25, 569–589, 2017