Second Order Dynamics with Closed-Loop Damping

Ernö Robert Csetnek
University of Vienna

joint work with
Hedy Attouch (University of Montpellier)
and Radu Boț (University of Vienna)

OWOS (One World Optimization Seminar)
18. October 2021

Introduction and motivation

Consider the optimization problem

$$\min_{x \in \mathcal{H}} f(x),$$

where

- $f : \mathcal{H} \to \mathbb{R}$ is convex and smooth, $\mathcal{H}$ is a real Hilbert space
- $\operatorname{argmin} f \neq \emptyset$

Second order dynamics with vanishing damping: Su, Boyd and Candès 2014, Attouch, Chbani, Peypouquet, Redont 2016

$$(\mathrm{AVD})_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \; t \geq t_0 > 0.$$

- fast convergence: $f(x(t)) - \min f = o\left(\frac{1}{t^2}\right)$ (in case $\alpha > 3$)
- weak convergence of $x(t)$ to an element in $\operatorname{argmin} f$ $(\alpha > 3)$
- discretization leads to Nesterov type scheme (inertial)

$$\begin{cases} y^k & = & x^k + \frac{k-1}{k+\alpha-1}(x^k - x^{k-1}) \\ x^{k+1} & = & y^k - \gamma \nabla f(y^k) \end{cases}$$

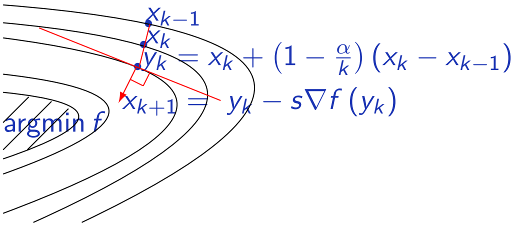


Figure: Nesterov accelerated gradient method

Dynamics with geometrical Hessian driven damping (Attouch, Peypouquet, Redont 2016):

$$(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta} \quad \ddot{x}(t)+\frac{\alpha}{t}\dot{x}(t)+\beta\nabla^2 f(x(t))\dot{x}(t)+\nabla f(x(t))=0$$

- natural relations to Newton and Levenberg-Marquardt iterative methods
- may induce a stabilization of the trajectories
- $\frac{d}{dt}\nabla f(x(t))=\nabla^2 f(x(t))\dot{x}(t)$, hence discretization leads to inertia involving $\nabla f(x^k)-\nabla f(x^{k-1})$, see also symplectic discretizations, recently investigated by Shi, Du, Jordan, Su 2019, Attouch, Chbani, Fadili, Riahi 2019
- fast convergence rates for the functions values: $o\left(\frac{1}{t^2}\right)$
- fast decay of the gradient along the trajectories: $\int_{t_0}^{\infty} t^2\|\nabla f(x(t))\|^2 dt<+\infty$ for $\alpha\geq 3$ and $\beta>0$
- weak convergence of the trajectories to a minimizer of f

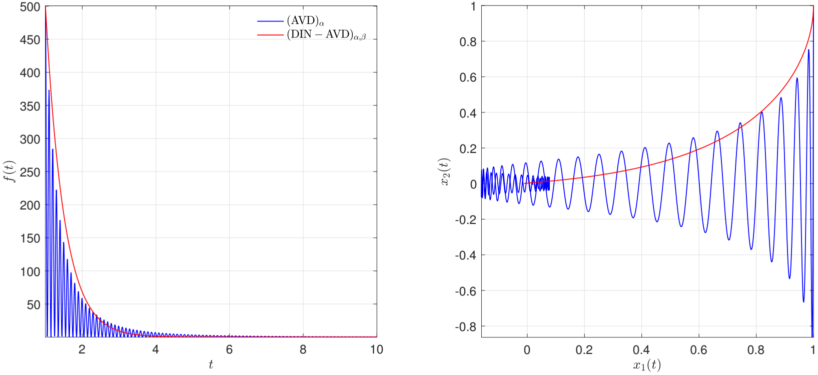


Figure: Evolution of the objective (left) and trajectories (right) for $(\mathrm{AVD})_\alpha$ ($\alpha = 3.1$) and $(\mathrm{DIN}-\mathrm{AVD})_{\alpha,\beta}$ ($\alpha = 3.1, \beta = 1$) on an ill-conditioned quadratic problem in $\mathbb{R}^2$: $f(x_1, x_2) = \frac{1}{2}(x_1^2 + 1000x_2^2)$

Link with the regularized Newton method

To overcome the ill-posed character of the continuous Newton method, Attouch and Svaiter studied the first-order system

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

- continuous version of the Levenberg-Marquardt (acts as a regularization of the Newton method)
- when $A = \nabla f$ we obtain

$$\Big(\gamma(t)\,\mathrm{Id} + \beta\nabla^2 f(x(t))\Big)\dot{x}(t) + \nabla f(x(t)) = 0.$$

The dynamics

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0$$

can be seen as an inertial (accelerated) version of the above system.

Attouch-Redont-Svaiter: closed-loop version of the above results

$$\begin{cases} v(t) \in A(x(t)) \\ \|v(t)\|^p \dot{x}(t) + \dot{v}(t) + v(t) = 0 \end{cases}$$

For optimization problems:

$$\|\nabla f(x(t))\|^p \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = 0.$$

This suggests to consider second order dynamics where the damping coefficient $\gamma(t)$ is a closed-loop control of ∇f:

$$\ddot{x}(t) + \gamma(t) \dot{x}(t) + \beta(t) \nabla^2 f(x(t)) \dot{x}(t) + b(t) \nabla f(x(t)) = 0,$$

Lin-Jordan 2020: investigated

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0,$$

where γ, β and b are defined by the following formulas:.

$$\begin{cases} |\lambda(t)|^p \|\nabla f(x(t))\|^{p-1} = \theta \\ a(t) = \frac{1}{4}\left(\int_0^t \sqrt{\lambda(s)}ds + c\right)^2 \\ \gamma(t) = 2\frac{\dot{a}(t)}{a(t)} - \frac{\ddot{a}(t)}{\dot{a}(t)} \\ \beta(t) = \left(\frac{\dot{a}(t)}{a(t)}\right)^2 \\ b(t) = \frac{\dot{a}(t)(\dot{a}(t)+\ddot{a}(t))}{a(t)} \end{cases}$$

Open loop/closed loop (design of the damping)

- Open-loop damping, non autonomous dynamic: the damping term involves coefficients which are given a priori as functions of time, example

$$(\mathrm{AVD})_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0.$$

- Closed-loop damping, adaptive methods, autonomous dynamic: the damping is a feedback of the current state of the system:

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \beta(t)\nabla^2 f(x(t))\dot{x}(t) + b(t)\nabla f(x(t)) = 0,$$

where $\gamma(t)$ involves $\nabla f(x(t))$, or even $\dot{x}(t)$.

Our investigations

$$\text{(ADIGE-V)} \quad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)),$$

- $\phi : \mathcal{H} \to \mathbb{R}_+$ is a convex continuous function
- $\partial\phi : \mathcal{H} \to 2^{\mathcal{H}}$ is the convex subdifferential
- ADIGE-V: Autonomous Damped Inertial Gradient Equation, V: the damping term is a closed-loop control of the velocity.

This model encompasses several classical situations:

- $\phi(u) = \frac{\gamma}{2}\|u\|^2$ corresponds to the Heavy Ball with Friction

$$\text{(HBF)} \quad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) = 0$$

introduced by B. Polyak, studied by Attouch–Goudou–Redont (exploration of local minima), Alvarez (convex case), Haraux-Jendoubi (analytic case), Bégout–Bolte–Jendoubi (convergence based on the Kurdyka-Lojasiewicz property)

- The case $\phi(u) = r\|u\|$ corresponds to the dry friction effect:

$$\ddot{x}(t) + r\frac{\dot{x}(t)}{\|\dot{x}(t)\|} + \nabla f(x(t)) = 0.$$

  Finite time stabilization property of the trajectories, which is satisfied generically with respect to the initial data: Adly–Attouch–Cabot, Amann–Diaz, see Adly-Attouch for recent developements.

- Take $\phi(u) = \frac{r}{p}\|u\|^p$ with $p \geq 1$, $r > 0$:

$$\ddot{x}(t) + r\|\dot{x}(t)\|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0.$$

  We will pay particular attention to the role played by the parameter p in the asymptotic convergence analysis.

  We will see that the case $p = 2$ separates the weak damping ($p > 2$) from the strong damping ($p < 2$).

We investigate also the dynamical system

$$\text{(ADIGE-VGH)} \ddot{x}(t) + \partial\phi\Big(\dot{x}(t) + \beta\nabla f(x(t)\Big) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) \ni 0,$$

where the damping term $\partial\phi\Big(\dot{x}(t) + \beta\nabla f(x(t)\Big)$ involves both $\dot{x}(t)$ and $\nabla f(x(t))$.

- When $\beta = 0$, we recover the closed loop controlled system

$$\ddot{x}(t) + \partial\phi\Big(\dot{x}(t)\Big) + \nabla f(x(t)) = 0$$

- When $\phi(u) = \frac{\gamma}{2}\|u\|^2$, we obtain the system

$$\ddot{x}(t) + \gamma\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + (1 + \gamma\beta)\nabla f(x(t)) = 0,$$

introduced by Alvarez-Attouch-Bolte-Redont.

- Take $\phi(u) = \frac{r}{p}\|u\|^p$ with $p \geq 1$, $r > 0$:

$$\ddot{x}(t) + r\|\dot{x}(t) + \beta\nabla f(x(t)\|^{p-2}(\dot{x}(t) + \beta\nabla f(x(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0$$

Precise setting and results

We consider the differential inclusion

$$\text{(ADIGE-V)} \qquad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)),$$

where ϕ is a convex damping potential:

- ϕ is a nonnegative convex continuous function;
- $\phi(0) = 0 = \min_{\mathcal{H}} \phi$;
- the minimal section of $\partial\phi$ is bounded on the bounded sets, that is, for any $R > 0$

$$\sup_{\|u\| \leq R} \|(\partial\phi)^0(u)\| < +\infty.$$

- $(\partial\phi)^0(u)$ is the element of minimal norm of the closed convex non empty set $\partial\phi(u)$.

If $\mathcal{H}$ is finite dimensional, then property (iii) is automatically satisfied. Indeed, in this case, $\partial\phi$ is bounded on the bounded sets.

Existence and uniqueness of the trajectory

The trajectory $x : [0, +\infty[\to \mathcal{H}$ is said to be a strong global solution of (ADIGE-V) if it satisfies the following properties:

- $x \in \mathcal{C}^1([0, +\infty[; \mathcal{H})$,
- $\dot{x} \in \mathrm{Lip}(0, T; \mathcal{H})$, $\ddot{x} \in L^\infty(0, T; \mathcal{H})$ for all $T > 0$,
- for almost all $t > 0$, $\quad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t))$.

Theorem

- $f : \mathcal{H} \to \mathbb{R}$ is a differentiable function, ∇f is Lipschitz continuous on the bounded subsets of $\mathcal{H}$, $\inf_{\mathcal{H}} f > -\infty$
- $\phi : \mathcal{H} \to \mathbb{R}_+$ is a damping potential

Then, for any $x_0, x_1 \in \mathcal{H}$, there exists a unique strong global solution $x : [0, +\infty[\to \mathcal{H}$ of (ADIGE-V), that is

$$\begin{cases} 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)) \\ x(0) = x_0,\ \dot{x}(0) = x_1. \end{cases}$$

• Regularize the differential inclusion:

For each $\lambda > 0$, we consider the approximate evolution equation

$$\ddot{x}_\lambda(t) + \nabla\phi_\lambda(\dot{x}_\lambda(t)) + \nabla f(x_\lambda(t)) = 0.$$

The Moreau envelope is the function $\phi_\lambda : \mathcal{H} \to \mathbb{R}$ defined by:

$$\phi_\lambda(u) = \min_{\xi\in\mathcal{H}} \left\{ \phi(\xi) + \frac{1}{2\lambda}\|u-\xi\|^2 \right\}.$$

The function ϕ_λ is convex, of class $\mathcal{C}^{1,1}$.

Set

$$Z_\lambda(t) = (x_\lambda(t), \dot{x}_\lambda(t)) \in \mathcal{H}\times\mathcal{H}.$$

The above system can be written equivalently as

$$\dot{Z}_\lambda(t) + \nabla\Phi_\lambda(Z_\lambda(t)) + G(Z_\lambda(t)) = 0, \quad Z_\lambda(0) = (x_0, x_1).$$

where

$$\Phi(x,u) = \phi(u), \quad \Phi_\lambda(x,u) = \phi_\lambda(u), \quad G(x,u) = \Big(-u, \nabla f(x) \Big).$$

• use the theory of Brézis and Attouch (variational convergence): (x_λ) converges uniformly (as $\lambda \to 0$) over the bounded time intervals to a solution of (ADIGE-V).

Asymptotic analysis of

$$\text{(ADIGE-V)} \qquad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t)),$$

Questions:

- convergence of trajectories towards critical points of f
- convergence rates for $f(x(t)) - \inf_{\mathcal{H}} f$ under convexity assumptions

Without geometric assumptions on f (like convexity, Lojasiewicz property, etc.), there is no hope.

Indeed, in case $\phi(u) = \frac{\gamma}{2}\|u\|^2$ $(\gamma > 0)$, (ADIGE-V) becomes

$$\text{(HBF)} \quad \ddot{x}(t) + \gamma\dot{x}(t) + \nabla f(x(t)) = 0$$

Attouch–Goudou–Redont (2000): example of a function $f : \mathbb{R}^2 \to \mathbb{R}$ which is $\mathcal{C}^1$, coercive, gradient is Lipschitz continuous on the bounded sets, and such that the (HBF) system admits an orbit $t \mapsto x(t)$ which does not converge as $t \to +\infty$.

Remark: we will see later that the above questions are difficult to solve even in the convex case!

Preliminary energy estimates (assume $\inf_{\mathcal{H}} f > -\infty$):

- the global energy $\mathcal{E}(t) = f(x(t)) - \inf_{\mathcal{H}} f + \frac{1}{2}\|\dot{x}(t)\|^2$ is non-increasing, and

$$\sup_{t\geq 0} \|\dot{x}(t)\| < +\infty, \quad \int_0^{+\infty} \phi(\dot{x}(t))dt < +\infty.$$

- $\sup_{t\geq 0} \|\ddot{x}(t)\| < +\infty$, if x is bounded (this is fulfilled if f is coercive)
- $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$, if moreover there exists $p \geq 1$, and $r > 0$ such that, for all $u \in \mathcal{H}$, $\phi(u) \geq r\|u\|^p$.
- $\lim_{t\to+\infty} \|\ddot{x}(t)\| = 0$ (under additional assumptions)

Idea: From $\quad 0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \nabla f(x(t))$ we get

$$0 = \frac{d}{dt}\mathcal{E}(t) + \langle \partial\phi(\dot{x}(t)), \dot{x}(t)\rangle \geq \frac{d}{dt}\mathcal{E}(t) + \phi(\dot{x}(t)).$$

(convex subdifferential inequality and $\phi(0) = 0$).
Integrate...

Strongly convex case: exponential convergence rate

Assume that:

- $f : \mathcal{H} \to \mathbb{R}$ is $\mu-$strongly convex ($\mu > 0$) and $\operatorname{argmin} f = \{\overline{x}\}$
- $\phi : \mathcal{H} \to \mathbb{R}_+$ is a damping potential which is differentiable, and $\nabla\phi$ is Lipschitz continuous on the bounded subsets of $\mathcal{H}$
- (local) there exists $\alpha, \rho > 0$

$$\langle \nabla\phi(u), u \rangle \geq \alpha \|u\|^2 \text{ whenever } \|u\| \leq \rho$$

- (global) there exist $p \geq 1$, $c > 0$, s.t. $\phi(u) \geq c\|u\|^p$ for all u.

Then, for any solution trajectory $x : [0, +\infty[\to \mathcal{H}$ of (ADIGE-V), we have exponential convergence rate to zero as $t \to +\infty$ for $f(x(t)) - f(\overline{x})$, $\|x(t) - \overline{x}\|$ and the velocity $\|\dot{x}(t)\|$.

Idea: use

$$f(\overline{x}) - f(x(t)) \geq \langle \nabla f(x(t)), \overline{x} - x(t) \rangle + \frac{\mu}{2}\|x(t) - \overline{x}\|^2$$

$$f(x(t)) - f(\overline{x}) \geq \frac{\mu}{2}\|x(t) - \overline{x}\|^2.$$

to derive $\dot{h}_\epsilon(t) + C_2 h_\epsilon(t) \leq 0,$ where

$$h_\epsilon(t) := f(x(t)) - f(\overline{x}) + \frac{1}{2}\|\dot{x}(t)\|^2 + \epsilon\langle x(t) - \overline{x}, \dot{x}(t)\rangle.$$

and $\epsilon, C_2 > 0$ are suitable chosen.

Apply Gronwall inequality

$$h_\epsilon(t) \leq h_\epsilon(0)e^{-C_2 t}.$$

The case f is convex quadratic positive definite: $f(x) = \frac{1}{2}\langle Ax, x\rangle$, $A : \mathcal{H} \to \mathcal{H}$ is linear, continuous, positive definite and self-adjoint.

$$\ddot{x}(t) + \partial\phi(\dot{x}(t)) + A(x(t)) \ni 0.$$

Then, we have the following ergodic convergence result

$$\frac{1}{t}\int_0^t x(\tau)d\tau \rightharpoonup x_\infty,$$

where $0 \in \partial\phi(0) + Ax_\infty$.

When ϕ is differentiable at the origin, we have $Ax_\infty = 0$, that is $x_\infty = 0$.

When $\phi(x) = r\|x\|$, we have $\|Ax_\infty\| \leq r$.

Remark: the difference with respect to the previous case is that we do not ask for ϕ the local and global properties

Proof:

$$0 \in \dot{z}(t) + (\partial\Phi + F)(z(t)),$$

where $z(t) = (x(t), \dot{x}(t)) \in \mathcal{H} \times \mathcal{H}$, and

- $\Phi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$, $\Phi(x, u) = \phi(u)$ is convex continuous
- $F : \mathcal{H} \times \mathcal{H} \to \mathcal{H} \times \mathcal{H}$ is defined by $F(x, u) = (-u, Ax)$.

Renorm the product space $\mathcal{H} \times \mathcal{H}$ as follows:

$$\langle\langle (x_1 u_1), (x_2, u_2) \rangle\rangle := \langle Ax_1, x_2 \rangle + \langle u_1, u_2 \rangle$$

- F is linear, continuous, skew-symmetric in the renormed space
- $\partial\Phi + F$ is maximal monotone (Rockafellar's Theorem)
- we can apply the theory concerning the semi groups generated by general maximally monotone operators

$z(t)$ converges weakly and in an ergodic way to a zero $z_\infty = (x_\infty, u_\infty)$ of $\partial\Phi + F$. This means $(0, \partial\phi(u_\infty)) + (-u_\infty, Ax_\infty) = (0, 0)$. Equivalently $u_\infty = 0$ and $\partial\phi(0) + Ax_\infty \ni 0$.

Haraux, Haraux-Jendoubi, Alabau Boussouira-Privat-Trélat

Numerical example:

$\mathcal{H} = \mathbb{R}$, $f(x) = \frac{1}{2}|x|^2$, and $\phi(u) = \frac{1}{p}|u|^p$, $p > 1$. Then, (ADIGE-V) writes

$$\ddot{x}(t) + |\dot{x}(t)|^{p-2}\dot{x}(t) + x(t) = 0.$$

- $p = 2$: $x(t) = \mathcal{O}\left(e^{-t}\right)$, $\dot{x}(t) = \mathcal{O}\left(e^{-t}\right)$
- for $p > 1$, $\lim_{t\to+\infty} x(t) = 0$ and $\lim_{t\to+\infty} \dot{x}(t) = 0$ (additional analysis is needed to pass from ergodic to nonergodic)

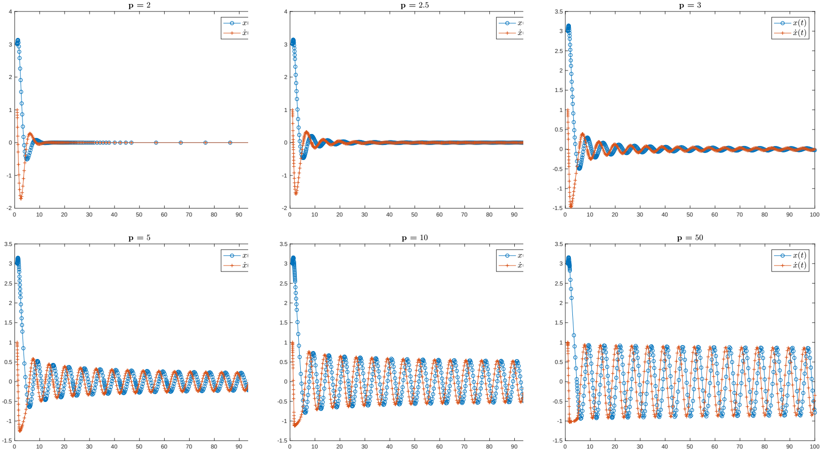


Figure: The evolution of the trajectories $x(t)$ (blue line) and $\dot{x}(t)$ (red line) for different values of $p \geq 2$.

Case $p > 2$ (weak damping): the damping $\gamma(t) := |\dot{x}(t)|^{p-2} \to 0$. As p increases, the damping effect tends to decrease, the trajectory tends to oscillate more and more, and the rate of convergence deteriorates.

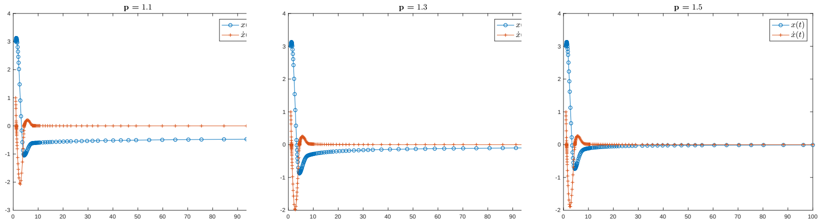


Figure: Evolution of $x(t)$ (blue) and $\dot{x}(t)$ (red) for different values of $1 < p < 2$.

Case $1 < p < 2$ (strong damping): the viscous damping

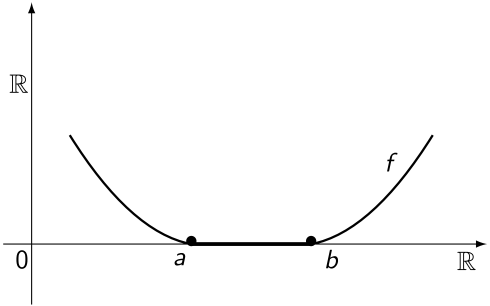
$$\gamma(t) := \frac{1}{|\dot{x}(t)|^{2-p}} \to +\infty \text{ as } t \to +\infty.$$

The trajectories exhibit small oscillations, and the velocity converges fastly to zero. When p is close to 1, the convergence of the trajectory to zero is poor, however, already a slight increase of p concisely improves the convergence of the trajectory. Indeed, when p becomes large the convergence of the trajectory improves.

Convex case: numerical example

$$\ddot{x}(t) + |\dot{x}(t)|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0.$$

Based on Haraux and Jendoubi: take $\mathcal{H} = \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}_+$ convex, $\mathcal{C}^1$, $\operatorname{argmin} f = [a, b]$ and f is coercive, *i.e.* $\lim_{|x|\to+\infty} f(x) = +\infty$.



Weak damping in the convex case: $p \geq 3$, convergence fails

Strong damping: $2 < p < 3$, convergence holds

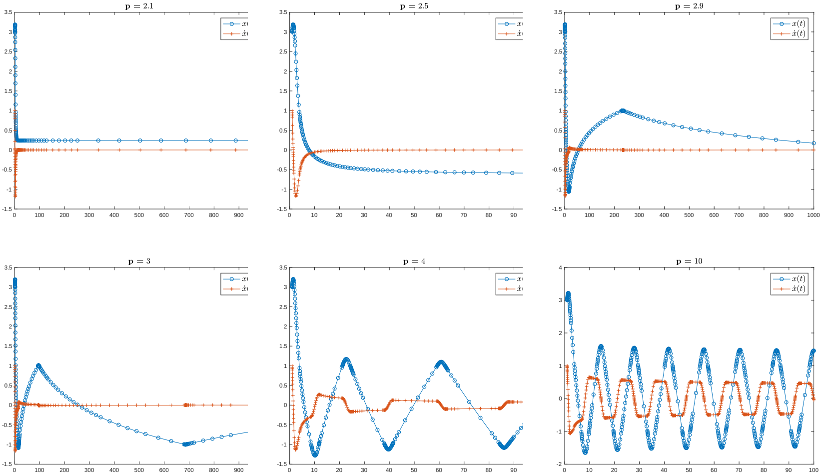


Figure: $x(t)$ (blue) and $\dot{x}(t)$ (red), $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 0$ for $|x| < 1$, $f(x) = \frac{1}{2}(x+1)^2$ for $x \leq -1$, and $f(x) = \frac{1}{2}(x-1)^2$ for $x \geq 1$.

Convergence under the Kurdyka-Lojasiewicz property

A differentiable function $G : \mathbb{R}^N \to \mathbb{R}$ has the KL property at $\overline{u} \in \mathbb{R}^N$ if there exist $r_0 > 0$, $\eta > 0$ and $\theta \in C([0, r_0), \mathbb{R}_+)$ s.t.

- $\theta(0) = 0$, $\theta \in C^1((0, r_0), \mathbb{R}_+)$ and $\theta' > 0$ on $(0, r_0)$
- $\|u - \overline{u}\| < \eta$ implies: $|G(u) - G(\overline{u})| < r_0$ and $\|\nabla(\theta \circ |G(\cdot) - G(\overline{u}|)(u)\| \geq 1$ (for $G(u) \neq G(\overline{u})$).
  θ is called desingularizing function of G at $\overline{u}$ on $B(\overline{u}, \eta)$
- G is called KL if it has the Kl property at each of its points

Many examples: semi-algebraic, real-analytic, tame, o - minimal structure, etc.

Quasi-gradient systems: Bégout–Bolte–Jendoubi, Haraux, Barta–Chill–Fašangová, Chergui, Huang

Let Γ be a nonempty closed subset of $\mathbb{R}^N$, and let $F : \mathbb{R}^N \to \mathbb{R}^N$ be a locally Lipschitz continuous mapping. We say that the first-order system

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

has a quasi-gradient structure for E on Γ, if there exist a differentiable function $E : \mathbb{R}^N \to \mathbb{R}$ and $\alpha > 0$ such that the two following conditions are satisfied:

(angle condition) $\langle \nabla E(z), F(z) \rangle \geq \alpha \|\nabla E(z)\| \|F(z)\|$ for all $z \in \Gamma$;

(rest point equivalence) $\quad \mathrm{crit}\, E \cap \Gamma = F^{-1}(0) \cap \Gamma.$

Remark: Such systems have a behavior which is very similar to those of gradient systems.

Theorem (Bégout–Bolte–Jendoubi) Let $F : \mathbb{R}^N \to \mathbb{R}^N$ be a locally Lipschitz continuous mapping. Let $z : [0, +\infty[\to \mathbb{R}^N$ be a bounded solution trajectory of

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

Take $R \geq \sup_{t\geq 0} \|z(t)\|$. Assume that F defines a quasi-gradient vector field for E_R on $\bar{B}(0, R)$, where $E_R : \mathbb{R}^N \to \mathbb{R}$ is a differentiable function. Assume further that the function E_R is (KL). Then, the following properties are satisfied:

(i) $z(t) \to z_\infty$ as $t \to +\infty$, where $z_\infty \in F^{-1}(0)$;

(ii) $\dot{z} \in L^1(0, +\infty; \mathbb{R}^N)$, $\dot{z}(t) \to 0$ as $t \to +\infty$;

(iii) $\|z(t) - z_\infty\| \leq \frac{1}{\alpha_R}\theta\Big(E_R(z(t) - E(z_\infty)\Big)$

where θ is the desingularizing function for E_R at z_∞, and α_R enters the angle condition.

$$\ddot{x}(t) + \nabla\phi(\dot{x}(t)) + \nabla f(x(t)) = 0,$$

Theorem: Let $f : \mathbb{R}^N \to \mathbb{R}$ be $\mathcal{C}^2$, ∇f is Lipschitz continuous on the bounded sets, $\inf_{\mathbb{R}^N} f > -\infty$. Let $E_\lambda : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$

$$E_\lambda(x,u) := \frac{1}{2}\|u\|^2 + f(x) + \lambda \langle \nabla f(x), u \rangle .$$

Assume: (i) E_λ satisfies the (KL) property
(ii) (local) there exists positive constants γ, δ, and $\epsilon > 0$:

$$\phi(u) \geq \gamma\|u\|^2 \text{ and } \|\nabla\phi(u)\| \leq \delta\|u\| \quad \text{for } \|u\| \leq \epsilon$$

(iii) (global) there exist $p \geq 1$, $c > 0$: $\phi(u) \geq c\|u\|^p$ for all u.

- $x(t) \to x_\infty$ as $t \to +\infty$, where $x_\infty \in \operatorname{crit} f$
- $\dot{x} \in L^1(0, +\infty; \mathbb{R}^N)$, $\dot{x}(t) \to 0$ as $t \to +\infty$
- $\|x(t) - x_\infty\| \leq \frac{1}{\alpha}\theta\Big(E_\lambda(x(t), u(t)) - E_\lambda(x_\infty, 0)\Big)$

where θ is the desingularizing function for E_λ at $(x_\infty, 0)$, and α enters the corresponding angle condition.

Idea: obtain a first order system with a quasi-gradient structure.
The Hamiltonian formulation of

$$\ddot{x}(t) + \nabla\phi(\dot{x}(t)) + \nabla f(x(t)) = 0,$$

gives the first-order differential system

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

where $z(t) = (x(t), \dot{x}(t)) \in \mathbb{R}^N \times \mathbb{R}^N$, and
$F : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ is defined by

$$F(x, u) = (-u, \nabla\phi(u)) + \nabla f(x)).$$

Take $E_\lambda : \mathbb{R}^N \to \mathbb{R}$ defined by

$$E_\lambda(x, u) := \frac{1}{2}\|u\|^2 + f(x) + \lambda \langle \nabla f(x), u \rangle .$$

- a desingularizing function of f is desingularizing of E_λ too
- it is possible to derive convergence rates for $\|x(t) - x_\infty\|$ in terms of the Lojasiewicz exponent

Closed-loop velocity control with Hessian driven damping

$$0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)).$$

The case $\phi(u) = \frac{\gamma}{2}\|u\|^2$ of a fixed viscous coefficient was first considered by Alvarez–Attouch–Bolte–Redont

The case $\phi(u) = \frac{\gamma}{2}\|u\|^2 + r\|u\|$ (viscous friction + dry friction) and Hessian damping has been considered by Adly–Attouch

By taking $\phi(u) = \frac{r}{p}\|u\|^p$, we get

$$\ddot{x}(t) + r\|\dot{x}(t)\|^{p-2}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0.$$

Existence and uniqueness

Convergence based on quasi-gradient systems and KL

Numerical illustrations

Open problems

Develop closed-loop versions of the Nesterov accelerated gradient method from a theoretical and numerical point of view.

$$(\mathrm{AVD})_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0,$$

- fast convergence: $f(x(t)) - \min f = o\left(\frac{1}{t^2}\right)$ (in case $\alpha > 3$)
- weak convergence of $x(t)$ to an element in $\operatorname{argmin} f$ $(\alpha > 3)$

Energy:

$$E(t) = t^2(f(x(t)) - \min f) + \frac{1}{2}\|\gamma(x(t) - x^*) + t\dot{x}(t)\|^2 + \frac{\xi}{2}\|x(t) - x^*\|^2$$

Closed loop:

$$\ddot{x}(t) + \gamma\|\dot{x}(t)\|^{p-2}\dot{x}(t) + \nabla f(x(t)) = 0.$$

The case $p = 2$ is the critical case separating the weak damping from the strong damping. Taking $p > 2$, with p close to 2 provides a vanishing viscosity damping coefficient, which is a specific property of the Nesterov

- Tikhonov regularization in the closed loop setting

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2 g(x(t))\dot{x}(t) + \nabla g(x(t)) + \epsilon(t)x(t) = 0$$

Attouch-Chbani-Riahi, Boţ, C., László

$$0 \in \ddot{x}(t) + \partial\phi(\dot{x}(t)) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) + \epsilon(t)x(t).$$

  - $\epsilon : [t_0, +\infty) \to [0, +\infty)$ is nonincreasing, of class C^1
  - $\lim_{t\to+\infty} \epsilon(t) = 0$

Why to use Tikhonov parametrization?

  - induces strong convergence to the minimal norm solution $\text{argmin}\{\|x\| : x \in \text{argmin}\, f\}$ (under more conditions on ϵ)
  - fast convergence rates for objective and gradient values

- Rescaling, perturbations, errors
- Nosmooth optimization problems (and the case of maximal monotone operators): Moreau envelope, Yosida regularization: in the open loop setting Attouch, Cabot, Peypouquet, László

Algorithmic consequences

Preliminary results in the paper for

$$\frac{1}{h^2}\left(x_{n+2} - 2x_{n+1} + x_n\right) + \nabla\phi\left(\frac{1}{h}(x_{n+1} - x_n)\right) + \nabla f(x_n) = 0.$$

and

$$\frac{1}{h^2}\left(x_{n+2} - 2x_{n+1} + x_n\right) + \nabla\phi\left(\frac{1}{h}(x_{n+2} - x_{n+1})\right) + \nabla f(x_{n+1}) = 0.$$

with step size $h > 0$.

This gives the proximal-gradient algorithm

$$x_{n+2} = x_{n+1} + h\text{prox}_{h\phi}\left(\frac{1}{h}(x_{n+1} - x_n) - h\nabla f(x_{n+1})\right).$$

[1] H. Attouch, R.I. Boţ, E.R. Csetnek - Fast optimization via inertial dyanmics with closed-loop damping, Journal of the European Mathematical Society, to appear (arXiv:2008.02261v3)

[2] H. Attouch, J. Peypouquet, P. Redont (2016) - Fast convex optimization via inertial dynamics with Hessian driven damping, J. Differential Equations 261(10), 5734-5783

[3] H. Attouch, B. F. Svaiter (2011), A continuous dynamical Newton-Like approach to solving monotone inclusions, SIAM J. Control Optim., 49 (2) (2011), pp. 574–598.

[4] P. Bégout, J. Bolte, M. Jendoubi, On damped second-order gradient systems, J. of Differ. Equ., 259 (7-8) (2015), pp. 3115–3143.

[5] T. Lin, M.I. Jordan, A control-theoretic perspective on optimal high-order optimization, (2019), preprint arXiv:1912.07168v1.

Thank you for your attention!