

FAST OPTIMIZATION THROUGH INERTIAL DYNAMICS WITH HESSIAN DRIVEN DAMPING.

Hedy ATTOUCH

Université Montpellier
Institut Montpelliérain Alexander Grothendieck, UMR CNRS 5149

2nd Workshop on Operator Splitting Methods in Data Analysis
Center for Computational Mathematics
Flatiron Institute, New York
March 20-22, 2019

Introduction

- $\mathcal{H}$ Hilbert space;
- $f : \mathcal{H} \rightarrow \mathbb{R}$ convex, class $\mathcal{C}^2$, $S = \operatorname{argmin} f \neq \emptyset$;
- $\alpha > 0$, $\beta \geq 0$ parameters.

Fast optimization via the inertial system with Hessian damping

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

Two types of damping:

- Hessian driven damping

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

- Asymptotic vanishing viscous damping

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

① Inertial dynamics and algorithms with Hessian damping.

- ① Convergence results. Continuous case.
- ② Convergence results. Algorithms.
- ③ Time scaling aspects.
- ④ Non-convex semi-algebraic optimization.

② Link with Newton and Levenberg-Marquardt methods.

- ① Open-loop regularization.
- ② Closed-loop regularization.

③ Nesterov-like methods. Open questions.

- ① Survey of recent results.
- ② Critical case $\alpha = 3$.
- ③ Link with the Ravine method.

1.0. Hessian driven damping. Introduction.

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

- (HBF) $\ddot{x}(t) + \Gamma\dot{x}(t) + \nabla f(x(t)) = 0$. $\Gamma = \gamma\text{Id}$, $\gamma > 0$, Polyak 1964.
 $\Gamma : \mathcal{H} \rightarrow \mathcal{H}$ anisotropic linear positive, Alvarez (SICON 2000).
- $(\text{DIN})_\beta \quad \ddot{x}(t) + \gamma\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0$.
Alvarez-A.-Bolte-Redont (JMPA 2002), A.-Maingé-Redont (DEA 2012).
- $(\text{AVD})_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0$.
Su-Boyd-Candès (NIPS '14), A-Chbani-Peypouquet-Redont (M.Prog.'18)
- $(\text{DIN} - \text{AVD})_{\alpha,\beta} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0$.
A.-Peypouquet-Redont (JDE 2016), Shi-Du-Jordan-Su (arXiv 2018).

Related dynamics: Bot-Csetnek (SVVA, 2018; Optimization, 2019).

1.0. Hessian driven damping. Introduction.

Compare $(AVD)_\alpha$ with $(DIN - AVD)_{\alpha,\beta}$

- $f(x, y) = \frac{1}{2}(x^2 + 1000y^2)$: quadratic, ill-conditioned.
- $\alpha = 3.1, \beta = 1$.
- Initial conditions: $(x(1), y(1)) = (1, 1), (\dot{x}(1), \dot{y}(1)) = (0, 0)$.

The wild oscillations of (AVD) are neutralized by (DIN-AVD).

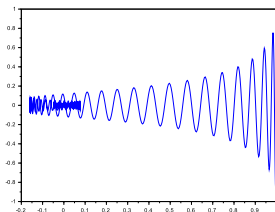
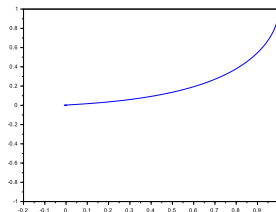


Figure: (AVD)



(DIN-AVD)

1.1. Hessian driven damping. Continuous case.

Theorem (A.-Peypouquet-Redont, JDE 2016)

Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution of $(\text{DIN} - \text{AVD})_{\alpha, \beta}$. Let $\alpha \geq 3$.

i) Suppose $\beta \geq 0$. Then $x(\cdot)$ is bounded and

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right), \quad \|\dot{x}(t) + \beta \nabla f(x(t))\| = \mathcal{O}\left(\frac{1}{t}\right).$$

ii) Suppose $\beta > 0$. Then $\int_{t_0}^{\infty} t^2 \|\nabla f(x(t))\|^2 dt < +\infty$ and

$$\lim_{t \rightarrow +\infty} \|\dot{x}(t)\| = \lim_{t \rightarrow +\infty} \|\ddot{x}(t)\| = \lim_{t \rightarrow +\infty} \|\nabla f(x(t))\| = 0.$$

Theorem (A.-Peypouquet-Redont, JDE 2016)

Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be solution of $(\text{DIN} - \text{AVD})_{\alpha, \beta}$. Take $\alpha > 3$. Then

i) $x(t) \rightarrow x_{\infty} \in \operatorname{argmin}_{\mathcal{H}} f$ as $t \rightarrow +\infty$.

ii) $f(x(t)) - \min_{\mathcal{H}} f = o\left(\frac{1}{t^2}\right)$, $\|\dot{x}(t) + \beta \nabla f(x(t))\| = o\left(\frac{1}{t}\right)$.

1.1. Hessian driven damping. Continuous case.

Perturbed dynamic

$$(\text{DIN} - \text{AVD})_{\text{pert}} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + \beta \nabla^2 f(x(t)) \dot{x}(t) + \nabla f(x(t)) = g(t).$$

Theorem (A.-Peypouquet-Redont, JDE 2016)

Let $x : [t_0, +\infty[\rightarrow \mathcal{H}$ be a solution trajectory of $(\text{DIN} - \text{AVD})_{\text{pert}}$.

Suppose $\int_{t_0}^{\infty} t \|g(t)\| dt < +\infty$. Then,

- $\alpha \geq 3, \beta \geq 0$: $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right)$.
- $\alpha > 3, \beta \geq 0$: $x(t) \rightarrow x_{\infty} \in \operatorname{argmin}_{\mathcal{H}} f$ as $t \rightarrow +\infty$.
- $\alpha \geq 3, \beta > 0$: $\int_{t_0}^{\infty} t^2 \|\nabla f(x(t))\|^2 dt < +\infty$.

1.2. Hessian driven damping. Algorithms.

Temporal scaling

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

Fully implicit discretization. $\nabla^2 f(x(t))\dot{x}(t) = \frac{d}{dt}\nabla f(x(t)).$

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{h^2} + \frac{\alpha}{kh} \frac{x_{k+1} - x_k}{h} + \frac{\beta}{h} (\nabla f(\textcolor{blue}{x}_{k+1}) - \nabla f(x_k)) + \textcolor{red}{b}_k \nabla f(\textcolor{blue}{x}_{k+1}) = 0.$$



$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k+\alpha}\right) (x_k - x_{k-1}) + \beta\sqrt{s} \left(1 - \frac{\alpha}{k+\alpha}\right) \nabla f(x_k) \\ x_{k+1} = \text{prox}_{\frac{k}{k+\alpha}(\beta\sqrt{s}+sb_k)f}(y_k). \end{cases}$$

1.2. Hessian driven damping. Algorithms.

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k+\alpha}\right) (x_k - x_{k-1}) + \beta\sqrt{s} \left(1 - \frac{\alpha}{k+\alpha}\right) \nabla f(x_k) \\ x_{k+1} = \text{prox}_{\frac{k}{k+\alpha}(\beta\sqrt{s}+sb_k)f}(y_k). \end{cases}$$

Lyapunov analysis

- $E_k := \delta_k(f(x_k) - \min_{\mathcal{H}} f) + \frac{1}{2}\|v_k\|^2$, $\delta_k := h(hkb_k - \beta)(k+1)$.
 $v_k := (\alpha - 1)(x_k - z) + k(x_k - x_{k-1} + \beta h \nabla f(x_k))$, $z \in \text{argmin } f$.
- $E_{k+1} - E_k \leq h\left((h(k+1)(k+2)b_{k+1} - hk(k+\alpha)b_k + \beta(\alpha-2))\right)(f(x_k) - \min_{\mathcal{H}} f)$

Theorem (A.-Chbani-Riahi, 2019)

Suppose $\sqrt{s}\left((k+1)(k+2)b_{k+1} - k(k+\alpha)b_k\right) + \beta(\alpha-2) \leq 0$. Then,

$$f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2 b_k}\right), \quad \sum_k k^2 b_k^2 \|\nabla f(x_{k+1})\|^2 < +\infty.$$

1.2. Hessian driven damping. Algorithms.

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k+\alpha}\right) (x_k - x_{k-1}) + \beta\sqrt{s} \left(1 - \frac{\alpha}{k+\alpha}\right) \nabla f(x_k) \\ x_{k+1} = \text{prox}_{\frac{k}{k+\alpha}(\beta\sqrt{s} + s\mathbf{b}_k)} f(y_k) \end{cases}$$

Analysis of the growth condition

$$\sqrt{s} \left((k+1)(k+2)b_{k+1} - k(k+\alpha)b_k \right) + \beta(\alpha-2) \leq 0.$$

- $b_k = 1$: OK if $\alpha > 3$.

$$f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2}\right), \quad \sum_k k^2 \|\nabla f(x_{k+1})\|^2 < +\infty.$$

- $b_k = k^\delta$: OK if $\alpha > 3 + \delta$.

$$f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^{2+\delta}}\right), \quad \sum_k k^{2+2\delta} \|\nabla f(x_{k+1})\|^2 < +\infty.$$

1.2. Hessian driven damping. Algorithms

$f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ convex, lower semicontinuous, proper.

$f_\lambda =$ Moreau envelope of f , $f_\lambda \in \mathcal{C}^{1,1}$ convex, $\min f_\lambda = \min f$.

Hessian dynamic attached to f_λ

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \beta \nabla^2 f_\lambda(x(t))\dot{x}(t) + b(t)\nabla f_\lambda(x(t)) = 0.$$

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k+\alpha}\right)(x_k - x_{k-1}) + \beta\sqrt{s}\left(1 - \frac{\alpha}{k+\alpha}\right)\nabla f_\lambda(x_k) \\ x_{k+1} = \text{prox}_{\frac{k}{k+\alpha}(\beta\sqrt{s}+sb_k)f_\lambda}(y_k). \end{cases}$$

$$\sqrt{s}\left((k+1)(k+2)b_{k+1} - k(k+\alpha)b_k\right) + \beta(\alpha-2) \leq 0.$$

$\Downarrow$

$$f_\lambda(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2 b_k}\right), \quad \sum_k k^2 b_k^2 \|\nabla f_\lambda(x_{k+1})\|^2 < +\infty.$$

1.2. Hessian driven damping. Algorithms

- $f_\lambda(x) = f(\text{prox}_{\lambda f}(x)) + \frac{1}{2\lambda} \|x - \text{prox}_{\lambda f}(x)\|^2.$
- $\nabla f_\lambda(x) = \frac{1}{\lambda} (x - \text{prox}_{\lambda f}(x)).$
- $\text{prox}_{\theta f_\lambda}(x) = \frac{\lambda}{\lambda+\theta} x + \frac{\theta}{\lambda+\theta} \text{prox}_{(\lambda+\theta)f}(x).$

Relaxed proximal inertial algorithm, $0 \leq \mu_k := \frac{\lambda(k+\alpha)}{\lambda(k+\alpha)+k(\beta\sqrt{s}+sb_k)} \leq 1.$

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k+\alpha}\right) (x_k - x_{k-1}) + \frac{\beta\sqrt{s}}{\lambda} \left(1 - \frac{\alpha}{k+\alpha}\right) (x_k - \text{prox}_{\lambda f}(x_k)) \\ x_{k+1} = \mu_k y_k + (1 - \mu_k) \text{prox}_{\frac{\lambda}{\mu_k} f}(y_k). \end{cases}$$

Theorem (A.-Chbani-Riahi, 2019)

Suppose $\sqrt{s} \left((k+1)(k+2)b_{k+1} - k(k+\alpha)b_k \right) + \beta(\alpha-2) \leq 0.$ Then $f(\text{prox}_{\lambda f}(x_k)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2 b_k}\right), \sum_k k^2 b_k^2 \|x_{k+1} - \text{prox}_{\lambda f}(x_{k+1})\|^2 < +\infty.$

1.2. Hessian driven damping. Algorithms

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

$(\text{DIN-AVD})_{\alpha,\beta}$ looks more complicated than $(\text{AVD})_{\alpha}$, but

Theorem (A.-Peypouquet-Redont, JDE 2016)

$(\text{DIN-AVD})_{\alpha,\beta}$ is equivalent to the first-order system

$$\begin{cases} \dot{x}(t) + \beta\nabla f(x(t)) - \left(\frac{1}{\beta} - \frac{\alpha}{t}\right)x(t) + \frac{1}{\beta}y(t) = 0; \\ \dot{y}(t) - \left(\frac{1}{\beta} - \frac{\alpha}{t} + \frac{\alpha\beta}{t^2}\right)x(t) + \frac{1}{\beta}y(t) = 0. \end{cases}$$

- Natural extension to $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ convex **lsc.** proper.
- Damped shocks in mechanics (A.-Maingé-Redont, DEA 2012).

Open question: discretize, compare with the previous approach.

1.3. Time scaling aspects

Model situation: $\gamma(t) = \frac{\alpha}{t}$, $\alpha \geq 3$

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

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right).$$

Change of time variable: $t = s^p$, $\xi(s) := x(s^p)$.

$$\ddot{\xi}(s) + \frac{\alpha_p}{s} \dot{\xi}(s) + p^2 s^{2(p-1)} \nabla f(\xi(s)) = 0,$$

$$\alpha_p = 1 + (\alpha - 1)p \geq \alpha \text{ for } p \geq 1, \quad b(s) = p^2 s^{2(p-1)}.$$

$$f(\xi(s)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{s^{2p}}\right).$$

Convergence rate of values: arbitrarily fast with p large.

1.3. Time scaling aspects

Discretization with fixed step size

$$(\text{IP})_{\alpha, b_k} \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{b_k f}(y_k). \end{cases}$$

Theorem (Attouch-Chbani-Riahi (HAL 2018))

Take $\alpha \geq 1$. Suppose that (b_k) satisfies the growth condition: for $k \geq 1$

$$(H_b) \quad b_{k+1} \leq \frac{k^2}{(k+1)(k+2-\alpha)} b_k.$$

Then, for any sequence (x_k) generated by the algorithm $(\text{IP})_{\alpha, b_k}$

$$f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2 b_k}\right) \quad \text{as } k \rightarrow +\infty.$$

$b_k = k^\delta$, $\delta > 0$: (H_b) is satisfied for $\alpha > 3 + \delta$. Conv. rate = $\mathcal{O}\left(\frac{1}{k^{2+\delta}}\right)$.

1.3. Time scaling aspects

Hessian damping $\gamma(t) = \frac{\alpha}{t}$, $\alpha \geq 3$, $\beta \geq 0$

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

$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^2}\right).$$

Change of time variable: $t = s^p$, $\xi(s) := x(s^p)$.

$$\ddot{\xi}(s) + \frac{\alpha_p}{s}\dot{\xi}(s) + p\beta s^{p-1}\nabla^2 f(x(t))\dot{\xi}(s) + p^2 s^{2(p-1)}\nabla f(\xi(s)) = 0,$$

$$\alpha_p = 1 + (\alpha - 1)p \geq \alpha \text{ for } p \geq 1, \quad f(\xi(s)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{s^{2p}}\right).$$

General scaling: to study

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

1.4. Non-convex semi-algebraic optimization.

$f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, semi-algebraic.

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

- $\alpha\beta > 1 \implies (DIN)$ is equivalent to a gradient system associated to a (KL) function (Alvarez-A.Bolte-Redont, JMPA 2002, Prop. 6.1) $\implies$ convergence of the trajectories and of the associated classical descent algorithms (A.-Bolte-Svaiter, Math. Prog, 2013).
- $\beta = 0$ Begout-Bolte-Jendoubi (JDE 2015) (DIN) is a **quasi-gradient** system for a modified energy function, which is still semi-algebraic $\implies$ convergence.

Questions: $\alpha > 0, \beta > 0$:

- 1) Is (DIN) a quasi-gradient system for a semi-algebraic function?
- 2) Non-autonomous case? (importance of the closed-loop approach).

2.1. Link with Newton, Levenberg-Marquardt methods

Newton method for optimization: solve $\nabla f(x) = 0$

- Discrete: $\nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k) = 0.$
- Continuous: $\nabla^2 f(x(t))\dot{x}(t) + \nabla f(x(t)) = 0.$

Levenberg-Marquardt regularization, $\lambda(t) > 0$

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

Attouch-Svaiter (SICON 2011): f convex, lsc. proper

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

Well-posed system. Convergence under appropriate tuning of $\lambda(t) \rightarrow 0$.

2.1. Link with Newton, Levenberg-Marquardt methods

A maximally monotone

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

Minty representation: $z(t) := x(t) + \mu(t)v(t)$, $\mu(t) := \frac{1}{\lambda(t)}$.

$$x(t) = J_{\mu(t)A}(z(t)) = (Id + \mu(t)A)^{-1}(z(t)) \quad \text{resolvent.}$$

$$v(t) = A_{\mu(t)}(z(t)) = \frac{1}{\mu(t)}(z(t) - J_{\mu(t)A}(z(t))) \quad \text{Yosida regularisation.}$$

Classical differential equation with respect to $z(\cdot)$

$$\begin{cases} x(t) = J_{\mu(t)A}(z(t)) \\ \dot{z}(t) + (\mu(t) - \dot{\mu}(t))A_{\mu(t)}(z(t)) = 0. \end{cases}$$

2.1. Link with Newton, Levenberg-Marquardt methods

A maximally monotone

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

$$\begin{cases} x(t) = J_{\mu(t)A}(z(t)) \\ \dot{z}(t) + (\mu(t) - \dot{\mu}(t))A_{\mu(t)}(z(t)) = 0. \end{cases}$$

Remark: $\mu(t) - \dot{\mu}(t) > 0 \iff \frac{\dot{\lambda}(t)}{\lambda(t)} > -1$.

Theorem (Attouch-Svaiter, SICON 2011)

Suppose A max. monotone, $S = A^{-1}(0) \neq \emptyset$, $\liminf_{t \rightarrow +\infty} \frac{\dot{\lambda}(t)}{\lambda(t)} > -1$.
Then $x(t)$ converges weakly to some $x_\infty \in S$, and $\|v(t)\| \rightarrow 0$.

2.1. Link with Newton, Levenberg-Marquardt methods

Closed-loop control

$$\begin{cases} v(t) \in A(x(t)) \\ \alpha(\|v(t)\|^2)\dot{x}(t) + \dot{v}(t) + v(t) = 0 \end{cases}$$

Theorem (Attouch-Redont-Svaiter, JOTA 2013)

Suppose A max. monotone, $S = A^{-1}(0) \neq \emptyset$, $\limsup_{r \rightarrow 0^+} \frac{r\dot{\alpha}(r)}{\alpha(r)} < \frac{1}{2}$.

Then $x(t)$ converges weakly to some $x_\infty \in S$, and $\|v(t)\| \rightarrow 0$.

Example: $0 < \gamma < 1$

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

2.2. Link with Newton, Levenberg-Marquardt methods

$A = \nabla f$, f convex.

Closed-loop control with Hessian damping

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

Conjecture 1

Suppose $0 < \gamma < 1$. Then $x(t)$ converges weakly to some $x_\infty \in S$, and $\|\nabla f(x(t))\| \rightarrow 0$.

Conjecture 2

Restarting methods work better with the Hessian driven damping.

2.2 Inertial dynamics & control: Adaptive restart

Strategy: maintain high velocity along the orbit.

$$(\text{AVD})_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla f(x(t)) = 0, \quad x(0) = x_0, \quad \dot{x}(0) = 0.$$

Restarting time: $T(f, x_0) = \sup\{t > 0, \forall \tau \in]0, t[, \frac{d}{d\tau}\|\dot{x}(\tau)\|^2 > 0\}$.

Before time $T(f, x_0) > 0$, $t \mapsto f(x(t))$ decreases:

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

At time $T(f, x_0)$, stop and restart, and so on.

Theorem (Su-Boyd-Candès, 2016), linear convergence

Suppose $f : \mathcal{H} \rightarrow \mathbb{R}$ strongly convex, ∇f Lipschitz continuous, $\alpha \geq 3$. Let $x_{sr}(\cdot)$ be an orbit of the speed restarting dynamic. Then

$$f(x_{sr}(t)) - \min_{\mathcal{H}} f \leq c_1 e^{-c_2 t}.$$

Question: adaptive restart for a general convex function f ?

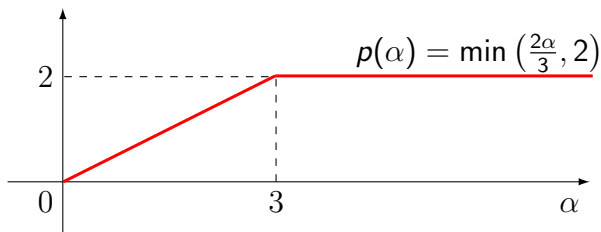
3.1 Inertial Gradient dynamic.

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

- $\alpha \geq 3$: Su-Boyd-Candès (J. of Machine Learning Research 2016)
 - $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}(\frac{1}{t^2})$ as $t \rightarrow +\infty$;
 - Link with Nesterov accelerated gradient method ($\alpha = 3$).
- $\alpha > 3$: A.-Chbani-Peypouquet-Redont (Math. Prog. 2018)
 - $f(x(t)) - \min_{\mathcal{H}} f = o(\frac{1}{t^2})$; $x(t) \rightarrow x_\infty \in S$ as $t \rightarrow +\infty$.
 - Link with Chambolle-Dossal inertial algorithm.
- $\alpha \leq 3$: A.-Chbani-Riahi (COCV 2017), Apidopoulos-Aujol-Dossal (SIOPT, 2018).
 - $f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}(\frac{1}{t^{\frac{2\alpha}{3}}})$ as $t \rightarrow +\infty$.

3.1 Inertial gradient dynamic

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



$$f(x(t)) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{t^{p(\alpha)}}\right)$$

- $\alpha \geq 3$: $\mathcal{O}\left(\frac{1}{t^2}\right)$ optimal rate, $f(x) = \|x\|^r$, $r \rightarrow +\infty$.
- $\alpha < 3$: $\mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$ optimal rate, $f(x) = \|x\|$.

3.1. Inertial approach to Nesterov accelerated method

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

Discretization: $\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2}(x_k - x_{k-1}) + \nabla f(y_k) = 0.$

Different cases for $y_k, s = h^2$

- Explicit: $y_k = x_k$, heavy ball with friction, Polyak

$$x_{k+1} = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) - s \nabla f(x_k).$$

- Implicit: $y_k = x_{k+1}$, inertial proximal algorithm, Güler, Beck-Teboulle

$$x_{k+1} = \text{prox}_{sf}(x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1})).$$

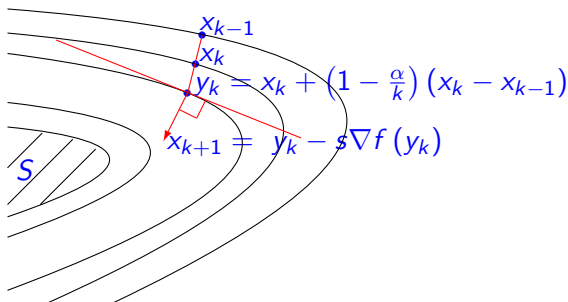
- Nesterov: $y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1})$

$$(\text{IG})_\alpha \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} = y_k - s \nabla f(y_k). \end{cases}$$

3.1. Inertial Gradient algorithm.

$$\min \{f(x) : x \in \mathcal{H}\}, \quad f : \mathcal{H} \rightarrow \mathbb{R} \text{ convex differentiable, } S = \operatorname{argmin} f \neq \emptyset.$$

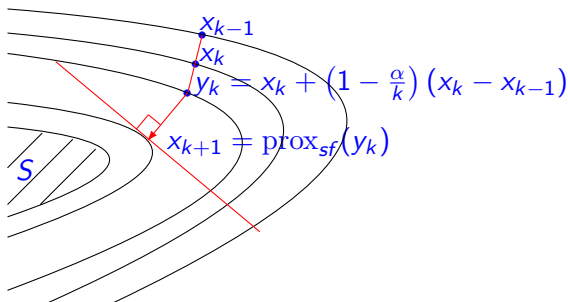
$$(\text{IG})_\alpha \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= y_k - s \nabla f(y_k) \end{cases}$$



3.1 Inertial Proximal algorithm

$\min \{f(x) : x \in \mathcal{H}\}$, $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ closed, convex, $S = \operatorname{argmin} f \neq \emptyset$.

$$(\text{IP})_{\alpha_k, \beta_k} \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} = \operatorname{prox}_{sf}(y_k). \end{cases}$$



3.1. Inertial Proximal-Gradient algorithms

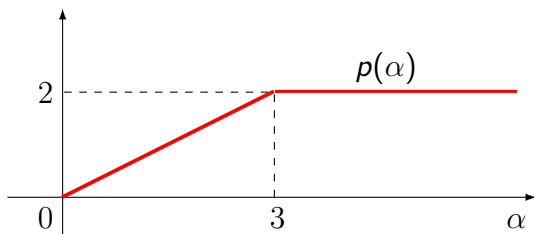
- $f : \mathcal{H} \rightarrow \mathbb{R}$ convex, $\mathcal{C}^1$, ∇f L -Lipschitz continuous; $0 < s \leq \frac{1}{L}$.
- $g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ convex, lower semicontinuous, proper.

$$(\text{IPG})_\alpha \quad \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= \text{prox}_{sg} (y_k - s \nabla f(y_k)) \end{cases}$$

- $\alpha = 3$: Nesterov ('83), Beck-Teboulle: FISTA (SIAM J. Imaging 2009)
 $(f + g)(x_k) - \min_{\mathcal{H}}(f + g) = \mathcal{O}\left(\frac{1}{k^2}\right).$
- $\alpha > 3$: $x_k \rightarrow x_\infty \in S = \text{argmin}(f + g)$, Chambolle-Dossal (JOTA 2015).
 $(f + g)(x_k) - \min_{\mathcal{H}}(f + g) = o\left(\frac{1}{k^2}\right)$, A.-Peypouquet (SIOPT 2016).
- $\alpha \leq 3$: $(f + g)(x_k) - \min_{\mathcal{H}}(f + g) = \mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right).$
Apidopoulos-Aujol-Dossal (Mat. Prog '18), A.-Chbani-Riahi (COCV '18)

3.1. Convergence rate. Global picture.

$$(\text{IPG})_{\alpha} \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= \text{prox}_{sg} (y_k - s \nabla f(y_k)) \end{cases}$$



Convergence rate of the values:

$$(f + g)(x_k) - \min_{\mathcal{H}}(f + g) = \mathcal{O}\left(\frac{1}{k^{p(\alpha)}}\right), \quad p(\alpha) = \min\left(\frac{2\alpha}{3}, 2\right).$$

3.1. General extrapolation coefficient

$$(\text{IPG})_{\alpha(\cdot)} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}); \\ x_{k+1} = \text{prox}_{sg}(y_k - s \nabla f(y_k)). \end{cases}$$

Auxiliary sequence: $t_k := 1 + \sum_{i=k}^{+\infty} \prod_{j=k}^i \alpha_j$, $\alpha_k = \frac{t_k - 1}{t_{k+1}}$

Theorem (A.-Cabot (SIOPT, 2018))

A. Suppose that the sequence (α_k) satisfies (K_0) and (K_1) .

$$(K_0) \quad \forall k \geq 1, \quad \sum_{i=k}^{+\infty} \prod_{j=k}^i \alpha_j < +\infty,$$

$$(K_1) \quad \forall k \geq 1, \quad t_{k+1}^2 - t_{k+1} - t_k^2 \leq 0.$$

Then, for any sequence (x_k) generated by algorithm $(\text{IPG})_{\alpha(\cdot)}$

$$(f + g)(x_k) - \min(f + g) = \mathcal{O}\left(\frac{1}{t_k^2}\right) \quad \text{as } k \rightarrow +\infty.$$

B. Assume moreover that $\exists m < 1$ s.t. $t_{k+1}^2 - t_k^2 \leq m t_{k+1} \forall k \geq 1$. Then

$$(f + g)(x_k) - \min_{\mathcal{H}}(f + g) = o\left(\frac{1}{\sum_{i=1}^k t_i}\right), \quad x_k \rightharpoonup \hat{x} \in \text{argmin}_{\mathcal{H}}(f + g).$$

3.1. General damping/ inertial coefficient

$$W_k := (f + g)(x_k) - \min_{\mathcal{H}}(f + g) + \frac{1}{2} \|x_k - x_{k-1}\|^2.$$

α_k	$1 - \frac{\alpha}{k}, \alpha \leq 3$	$1 - \frac{\alpha}{k}, \alpha > 3$	$1 - \frac{(\ln k)^\theta}{k}$	$1 - \frac{\alpha}{k^r}, r \in]0, 1[$	$0 \leq \alpha < 1$
W_k	$\mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$	$\mathcal{O}\left(\frac{1}{k^2}\right)$	$\mathcal{O}\left(\frac{(\ln k)^\theta}{k^2}\right)$	$\mathcal{O}\left(\frac{1}{k^{r+1}}\right)$	$\mathcal{O}\left(\frac{1}{k}\right)$

- Historical choice by Nesterov: $t_{k+1}^2 - t_{k+1} - t_k^2 = 0$. Then $\alpha_k \sim 1 - \frac{3}{k}$.
- $\alpha_k = 1 - \frac{\alpha}{k}$: $t_{k+1} = \frac{k}{\alpha-1}$. (K_1) corresponds to $\alpha \geq 3$, (K_1^+) to $\alpha > 3$.
- $0 \leq m \leq \alpha_k \leq M < 1$, $\frac{1-m}{1-M} < \frac{3}{2}$: $W_k = \mathcal{O}\left(\frac{1}{k}\right)$ (HBF).

3.2. Critical case $\alpha = 3$

Open pb. 1: convergence of the iterates in the case $\alpha = 3$.

Theorem (A.-Chbani-Riahi, COCV 2018)

$\mathcal{H} = \mathbb{R}$. Let $x : [t_0; +\infty[\rightarrow \mathbb{R}$ be a solution of $(AVD)_\alpha$ with $\alpha = 3$. Then $x(t)$ converges, as $t \rightarrow +\infty$, to a point in S .

Lemma ($\mathcal{H}$ general Hilbert space)

Let $x(\cdot)$ be a trajectory of $(AVD)_\alpha$, $\alpha \leq 3$. Suppose that for $t_2 \geq t_1$

$$x(t_1) = x(t_2) \in S = \operatorname{argmin} f.$$

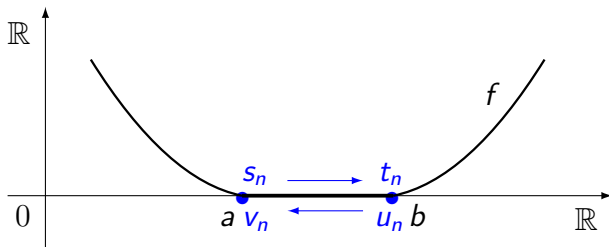
Then $t_2^{\frac{\alpha}{3}} \|\dot{x}(t_2)\| \leq t_1^{\frac{\alpha}{3}} \|\dot{x}(t_1)\|$. In particular, for $\alpha = 3$,

$$t_2 \|\dot{x}(t_2)\| \leq t_1 \|\dot{x}(t_1)\|.$$

3.2. Critical case $\alpha = 3$

Only difficult case: the trajectory passes in a and b an infinite number of times. Let us show that this is impossible.

The argument is based on **the decay of $t\|\dot{x}(t)\|$ during a loop.**



$$|v_n \dot{x}(v_n)| \leq |s_n \dot{x}(s_n)| - 2(\alpha - 1)(b - a).$$

For each loop, $t\|\dot{x}(t)\|$ decreases by a fixed positive quantity!

3.3. Link with the Ravine method.

Claim 1: (y_k) follows the Ravine method.

$$(\text{IG})_\alpha \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= y_k - s \nabla f(y_k) \end{cases}$$

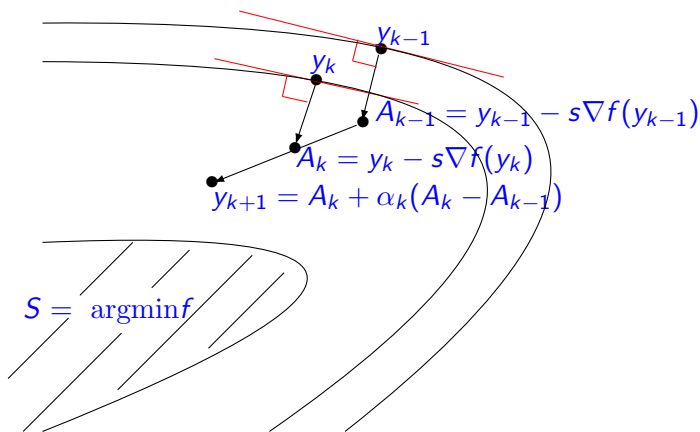
$$y_{k+1} = x_{k+1} + \left(1 - \frac{\alpha}{k+1}\right) (x_{k+1} - x_k)$$

$$y_{k+1} = y_k - s \nabla f(y_k) + \left(1 - \frac{\alpha}{k+1}\right) (y_k - s \nabla f(y_k) - (y_{k-1} - s \nabla f(y_{k-1})))$$

$$(\text{Ravine})_\alpha \begin{cases} A_k &:= y_k - s \nabla f(y_k) \\ y_{k+1} &= A_k + \left(1 - \frac{\alpha}{k+1}\right) (A_k - A_{k-1}). \end{cases}$$

3.3. Link with the Ravine method.

Ravine method: Gelfand, Tsetlin (1961), Polyak (2018).



3.3. Link with the Ravine method.

Claim 2: the continuous dynamic attached to (y_k) is $(AVD)_\alpha$.

$$y_{k+1} - y_k = -s\nabla f(y_k) + \alpha_{k+1}(y_k - y_{k-1}) - \alpha_{k+1}s(\nabla f(y_k) - \nabla f(y_{k-1})).$$

$$\frac{y_{k+1} - 2y_k + y_{k-1}}{h^2} + \nabla f(y_k) + \frac{1 - \alpha_{k+1}}{h} \frac{y_k - y_{k-1}}{h} + \alpha_{k+1}(\nabla f(y_k) - \nabla f(y_{k-1})) = 0.$$

Ansatz $y_k \approx Y(kh)$, smooth curve $Y(t)$ defined for $t \geq t_0$. Set $k = t/h$. As h goes to zero, $Y(t) \approx y_{t/h} = y_k$, $Y(t+h) \approx y_{(t+h)/h} = y_{k+1}$.

Taylor expansions:

$$\ddot{Y}(t) + \frac{\alpha}{t} \dot{Y}(t) + \nabla f(Y(t)) + o(1) = 0.$$

$h \rightarrow 0$: $Y(\cdot)$ is a solution trajectory of $(AVD)_\alpha$.

3.3. Link with the Ravine method. Convergence of (y_k) .

Claim 3: for $\alpha > 3$, (y_k) satisfies the same convergence rate as (x_k) .

$$(\text{IG})_\alpha \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= y_k - s \nabla f(y_k) \end{cases}$$

Theorem (A.-Peypouquet (SIOPT 2016), A.-Cabot (SIOPT 2018))

Let $(x_k), (y_k)$ be generated by $(\text{IG})_\alpha$ with $\alpha > 3$. Then

- $f(x_k) - \min_{\mathcal{H}} f = o(\frac{1}{k^2})$;
- $x_k \rightarrow x_\infty \in \mathcal{S}$ as $k \rightarrow +\infty$;
- $\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right)$.

Consequences:

i) $\|y_k - x_k\| \leq \|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right)$. Hence $y_k \rightarrow x_\infty \in \mathcal{S}$ as $k \rightarrow +\infty$.

The two sequences (x_k) and (y_k) have the same limit in $\mathcal{S}$.

3.3. Convergence of (y_k) .

ii) Convexity of f : $f(x_k) \geq f(y_k) + \langle \nabla f(y_k), x_k - y_k \rangle$.

$$\begin{aligned} \Rightarrow f(y_k) - \min_{\mathcal{H}} f &\leq f(x_k) - \min_{\mathcal{H}} f + \|\nabla f(y_k)\| \|y_k - x_k\| \\ &\leq o\left(\frac{1}{k^2}\right) + \frac{1}{s} \|x_{k+1} - y_k\| \|y_k - x_k\| \\ &\leq o\left(\frac{1}{k^2}\right) + \frac{1}{s} (\|x_{k+1} - x_k\| + \|x_k - y_k\|) \|x_k - y_k\| \\ &\leq o\left(\frac{1}{k^2}\right) + \frac{1}{s} (\|x_{k+1} - x_k\| + \|x_k - x_{k-1}\|) \|x_k - x_{k-1}\| \\ &\leq o\left(\frac{1}{k^2}\right) \end{aligned}$$

because $\|x_k - y_k\| \leq \|x_k - x_{k-1}\|$ and $\|x_k - x_{k-1}\| = o\left(\frac{1}{k}\right)$.

Open pb. 2: $f(y_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2}\right)$ true for $\alpha = 3$? FISTA?

3.3. Rescaled Ravine method

Modify the scaling in order to make appear the Hessian damping.

$$(\text{Ravine})_{\text{rescale}} \quad \left\{ \begin{array}{l} A_k = x_k - s \nabla f(x_k) \\ B_k = x_k - \beta \sqrt{s} \nabla f(x_k) \\ x_{k+1} = A_k + \alpha_k (B_k - B_{k-1}) \end{array} \right.$$

$$x_{k+1} = (x_k - s \nabla f(x_k)) + \alpha_k [x_k - \beta \sqrt{s} \nabla f(x_k) - x_{k-1} + \beta \sqrt{s} \nabla f(x_{k-1})].$$

$$x_{k+1} - x_k = -s \nabla f(x_k) + \alpha_k (x_k - x_{k-1}) - \beta \alpha_k \sqrt{s} (\nabla f(x_k) - \nabla f(x_{k-1})).$$

Set $h = \sqrt{s}$, time step size. Dividing by $s > 0$

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{h^2} + \nabla f(x_k) + \frac{1 - \alpha_k}{h} \frac{x_k - x_{k-1}}{h} + \beta \frac{\alpha_k}{h} (\nabla f(x_k) - \nabla f(x_{k-1})) = 0$$

3.3. Rescaled Ravine method

Claim 4: the continuous dynamic attached to the rescaled Ravine method is $(\text{DIN} - \text{AVD})_{\alpha,\beta}$.

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{h^2} + \nabla f(x_k) + \frac{1 - \alpha_k}{h} \frac{x_k - x_{k-1}}{h} + \beta \frac{\alpha_k}{h} (\nabla f(x_k) - \nabla f(x_{k-1})) = 0.$$

$x_k \approx X(kh)$, $k = t/h$. $X(t) \approx x_{t/h} = x_k$, $X(t+h) \approx x_{(t+h)/h} = x_{k+1}$.
Take $\alpha_k = 1 - \frac{\alpha}{k}$. Thus $\frac{1-\alpha_k}{h} = \frac{\alpha}{t}$. Taylor expansions

$$\frac{(x_{k+1} - x_k) - (x_k - x_{k-1}))}{h^2} \approx \ddot{X}(t) + o(1),$$

$$\nabla f(x_k) - \nabla f(x_{k-1}) \approx \nabla f(X(t)) - \nabla f(X(t-h)) = h \nabla^2 f(X(t)) \dot{X}(t) + o(h)$$

$$\ddot{X}(t) + \frac{\alpha}{t} \dot{X}(t) + \nabla f(X(t)) + \beta \nabla^2 f(X(t)) \dot{X}(t) + o(1) = 0.$$

$h \rightarrow 0$: $X(\cdot)$ is a solution trajectory of $(\text{DIN} - \text{AVD})_{\alpha,\beta}$.

Appendix 1. Hessian driven damping. Algorithms.

Inertial proximal method with Hessian damping and time scaling

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

$$\frac{x_{k+1} - 2x_k + x_{k-1}}{h^2} + \frac{\alpha}{kh} \frac{x_{k+1} - x_k}{h} + \frac{\beta}{h} (\nabla f(x_{k+1}) - \nabla f(x_k)) + b_k \nabla f(x_{k+1}) = 0.$$

$$k(x_{k+1} - 2x_k + x_{k-1}) + \alpha(x_{k+1} - x_k) + \beta hk(\nabla f(x_{k+1}) - \nabla f(x_k)) + b_k h^2 k \nabla f(x_{k+1}) = 0.$$

- $E_k := \delta_k(f(x_k) - \min_{\mathcal{H}} f) + \frac{1}{2}\|v_k\|^2.$
- $v_k := (\alpha - 1)(x_k - z) + k(x_k - x_{k-1} + \beta h \nabla f(x_k)).$

$$E_{k+1} - E_k = (\delta_{k+1} - \delta_k)(f(x_{k+1}) - m_f) + \delta_k(f(x_{k+1}) - f(x_k)) + \frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2$$

$$\frac{1}{2}\|v_{k+1}\|^2 - \frac{1}{2}\|v_k\|^2 = \langle v_{k+1} - v_k, v_{k+1} \rangle - \frac{1}{2}\|v_{k+1} - v_k\|^2.$$

$$\begin{aligned} v_{k+1} - v_k &= (\alpha - 1)(x_{k+1} - x_k) + (k + 1)(x_{k+1} - x_k + \beta h \nabla f(x_{k+1})) - k(x_k - x_{k-1} + \beta h \nabla f(x_k)) \\ &= \alpha(x_{k+1} - x_k) + k(x_{k+1} - 2x_k + x_{k-1}) + \beta hk(\nabla f(x_{k+1}) - \nabla f(x_k)) + \beta h \nabla f(x_{k+1}) \\ &= (\beta h - b_k h^2 k) \nabla f(x_{k+1}). \end{aligned}$$

Appendix 1. Hessian driven damping. Algorithms.

$$\begin{aligned}
 & \frac{1}{2} \|v_{k+1}\|^2 - \frac{1}{2} \|v_k\|^2 = -\frac{1}{2} (\beta h - b_k h^2 k)^2 \|\nabla f(x_{k+1})\|^2 \\
 & + (\beta h - b_k h^2 k) \langle \nabla f(x_{k+1}), (\alpha - 1)(x_{k+1} - z) + (k + 1)(x_{k+1} - x_k + \beta h \nabla f(x_{k+1})) \rangle \\
 & = -\left(\frac{1}{2} (\beta h - b_k h^2 k)^2 + \beta h(k + 1)(b_k h^2 k - \beta h) \right) \|\nabla f(x_{k+1})\|^2 \\
 & + (\alpha - 1)h(b_k h k - \beta) \langle \nabla f(x_{k+1}), z - x_{k+1} \rangle + (b_k h^2 k - \beta h)(k + 1) \langle \nabla f(x_{k+1}), x_k - x_{k+1} \rangle \\
 & \leq (\alpha - 1)h(b_k h k - \beta)(f(z) - f(x_{k+1})) + h(b_k h k - \beta)(k + 1)(f(x_k) - f(x_{k+1})).
 \end{aligned}$$

Take $\delta_k := h(hkb_k - \beta)(k + 1)$. $f(x_k) - f(x_{k+1})$ cancel in $E_{k+1} - E_k$

$$E_{k+1} - E_k \leq \left(\delta_{k+1} - \delta_k - (\alpha - 1)h(b_k h k - \beta) \right) (f(x_{k+1}) - f(z)).$$

$$\delta_{k+1} - \delta_k - (\alpha - 1)h(b_k h k - \beta) =$$

$$h \left((h(k + 1)b_{k+1} - \beta)(k + 2) - (hkb_k - \beta)(k + 1) - (\alpha - 1)(b_k h k - \beta) \right)$$

$$= h \left((h(k + 1)(k + 2)b_{k+1} - hk(k + \alpha)b_k + \beta(\alpha - 2)) \right)$$

$$E_{k+1} - E_k \leq h \left((h(k + 1)(k + 2)b_{k+1} - hk(k + \alpha)b_k + \beta(\alpha - 2)) \right) (f(x_k) - \min_{\mathcal{H}} f).$$

Appendix 1. Hessian driven damping. Algorithms.

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k+\alpha}\right) (x_k - x_{k-1}) + \beta\sqrt{s} \left(1 - \frac{\alpha}{k+\alpha}\right) \nabla f(x_k) \\ x_{k+1} = \text{prox}_{\frac{k(\beta h + b_k s)}{k+\alpha} f}(y_k) \end{cases}$$

Theorem (A.-Chbani-Riahi, 2019)

Suppose $h(k+1)(k+2)b_{k+1} - hk(k+\alpha)b_k + \beta(\alpha-2) \leq 0$. Then

$$f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^2 b_k}\right), \quad \sum_k k^2 b_k^2 \|\nabla f(x_{k+1})\|^2 < +\infty.$$

- $b_k = 1$: OK if $\alpha > 3$.
- $b_k = k^\delta$: OK if $\alpha > 3 + \delta$.

$$f(x_k) - \min_{\mathcal{H}} f = \mathcal{O}\left(\frac{1}{k^{2+\delta}}\right), \quad \sum_k k^{2+2\delta} \|\nabla f(x_{k+1})\|^2 < +\infty.$$

Appendix 2. Nesterov accelerated gradient method

$f : \mathcal{H} \rightarrow \mathbb{R}$ convex differentiable, ∇f L -Lipschitz, $\operatorname{argmin} f \neq \emptyset$.

- Nesterov accelerated gradient method (1983), $0 < s \leq \frac{1}{L}$

$$\text{(IG)} \quad \begin{cases} y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} &= y_k - s \nabla f(y_k) \end{cases}$$

$$\alpha_k = \frac{t_k - 1}{t_{k+1}} \quad \text{with } t_{k+1} = \frac{\sqrt{4t_k^2 + 1} + 1}{2} \quad \text{and } t_1 = 1.$$

- (α_k) increasing sequence; $\alpha_k \sim 1 - \frac{3}{k}$ as $k \rightarrow +\infty$;
- Convergence rate $\mathcal{O}(\frac{1}{k^2})$; optimal / first-order methods.
- Convergence of trajectories: open question in general.

-  B. ABBAS, H. ATTOUCH, AND B. F. SVAITER, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, JOTA, **161**(2) (2014), pp. 331–360.
-  S. ADLY, H. ATTOUCH, A. CABOT, *Finite time stabilization of nonlinear oscillators subject to dry friction*, Nonsmooth Mechanics and Analysis (ed. P. Alart, O. Maisonneuve, R.T. Rockafellar), Adv. in Math. and Mech., Kluwer , **12** (2006) pp. 289–304.
-  F. ALVAREZ, *On the minimizing property of a second-order dissipative system in Hilbert spaces*, SIAM J. Control Optim., **38**(4) (2000), pp. 1102–1119.
-  F. ALVAREZ, H. ATTOUCH, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Analysis, **9** (2001), pp. 3–11.



F. ALVAREZ, H. ATTOUCH, *Convergence and asymptotic stabilization for some damped hyperbolic equations with non-isolated equilibria*, ESAIM COCV, **6** (2001), pp. 539–552.



F. ALVAREZ, H. ATTOUCH, J. BOLTE, P. REDONT, *A second-order gradient-like dissipative dynamical system with Hessian-driven damping. Application to optimization and mechanics*, J. Math. Pures Appl., **81**(8) (2002), pp. 747–779.



V. APIDOPOULOS, J.-F. AUJOL, CH. DOSSAL, *The Differential Inclusion Modeling FISTA Algorithm and Optimality of Convergence Rate in the Case $b < 3$* , SIAM Journal on Optimization, **28**(1) 2018, pp. 551–574.



V. APIDOPOULOS, J.-F. AUJOL, CH. DOSSAL, *Convergence rate of inertial Forward-Backward algorithm beyond Nesterov's rule*, Mathematical programming, 2018,
<https://doi.org/10.1007/s10107-018-1350-9>. Preprint available on HAL.



H. ATTOUCH, G. BUTTAZZO, G. MICHAILLE, *Variational analysis in Sobolev and BV spaces. Applications to PDE's and optimization*, MPS/SIAM Series on Optimization, **6**, SIAM, 2014.

-  H. ATTOUCH, A. CABOT, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, J. Differential Equations, **263** (2017), pp. 5412-5458.
-  H. ATTOUCH, A. CABOT, *Convergence rates of inertial forward-backward algorithms*, SIAM J. Optim., **28** (1) (2018), pp. 849–874.
-  H. ATTOUCH, A. CABOT, *Convergence of damped inertial dynamics governed by Regularized maximally monotone operators*, 2017. HAL-01648383, accepted in JDE.
-  H. ATTOUCH, A. CABOT, *Convergence of a relaxed inertial proximal algorithm for maximally monotone operators*, 2018. HAL-01708905, submitted to Math Prog.



H. ATTOUCH, A. CABOT, P. REDONT, *The dynamics of elastic shocks via epigraphical regularization of a differential inclusion*, Adv. Math. Sci. Appl., **12**(1) (2002), pp. 273–306.



H. ATTOUCH, M.-O. CZARNECKI, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria*, J. Differential Equations, **179** (2002), pp. 278–310.



H. ATTOUCH, M.-O. CZARNECKI, *Asymptotic behavior of gradient-like dynamical systems involving inertia and multiscale aspects*, J. Differential Equations, **262**(3), (2017), pp. 2745–2770.

References

-  H. ATTOUCH, Z. CHBANI, J. PEYPOUQUET, P. REDONT, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity*, *Math. Program. Ser. B* 168 (2018), pp. 123–175.
-  H. ATTOUCH, Z. CHBANI, H. RIAHI, *Combining fast inertial dynamics for convex optimization with Tikhonov regularization*, *J. Math. Anal. Appl.* 457 (2018), pp. 1065–1094.
-  H. ATTOUCH, Z. CHBANI, H. RIAHI, *Rate of convergence of the Nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$* , arXiv:1706.05671v1 [math.OC] 2017, to appear in *ESAIM-COCV*.
-  H. ATTOUCH, X. GOUDOU AND P. REDONT, *The heavy ball with friction method*, *Commun. Contemp. Math.*, **2** (2000), No 1, pp. 1–34.

-  H. ATTOUCH, P.E. MAINGÉ, P. REDONT, *A second-order differential system with Hessian-driven damping; Application to non-elastic shock laws*, Differential Equations and Applications, **4** (2012), No 1, pp. 27–65.
-  H. ATTOUCH, J. PEYPOUQUET, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $\frac{1}{k^2}$* , SIAM J. Optim., **26** (2016), No. 3, pp. 1824–1834.
-  H. ATTOUCH, J. PEYPOUQUET, P. REDONT, *Fast convex minimization via inertial dynamics with Hessian driven damping*, J. Differential Equations, **261**, Issue 10, (2016), pp. 5734–5783.
-  H. ATTOUCH, J. PEYPOUQUET, P. REDONT, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM J. Optim., **24** (2014), No. 1, pp. 232–256.



H. ATTOUCH, B. F. SVAITER, *A continuous dynamical Newton-Like approach to solving monotone inclusions*, SIAM J. Control Optim., 49 (2011), No. 2, pp. 574–598.



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

References

-  J-F. AUJOL, C. DOSSAL, *Optimal rate of convergence of an ode associated to the fast gradient descent schemes for $b > 0$* , HAL Preprint, June 2017.
-  H. BAUSCHKE AND P. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert spaces*, CMS Books in Mathematics, Springer, (2011).
-  A. BECK AND M. TEBoulLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., **2** (1) (2009), pp. 183–202.
-  R. I. BOT, E. R. CSETNEK, *Second order forward-backward dynamical systems for monotone inclusion problems*, SIAM Journal on Control and Optimization 54(3) (2016), pp. 1423–1443.
-  H. BRÉZIS, *Opérateurs maximaux monotones dans les Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).

-  H. BRÉZIS, *Asymptotic behavior of some evolution systems: Nonlinear evolution equations*, Academic Press, New York, (1978), pp. 141–154.
-  R.E. BRUCK, *Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces*, J. Funct. Anal., **18** (1975), pp. 15–26.
-  A. CABOT, *Inertial gradient-like dynamical system controlled by a stabilizing term*, J. Optim. Theory Appl., 120 (2004), pp. 275–303.
-  A. CABOT, H. ENGLER, S. GADAT, *On the long time behavior of second-order differential equations with asymptotically small dissipation* Trans. AMS, **361** (2009), pp. 5983–6017.

-  A. CABOT, H. ENGLER, S. GADAT, *Second order differential equations with asymptotically small dissipation and piecewise flat potentials*, Elect. J. of Differential Equations, **17** (2009), pp. 33–38.
-  A. CABOT, P. FRANKEL, *Asymptotics for some semilinear hyperbolic equations with non-autonomous damping*, J. Differential Equations **252** (2012), pp. 294–322.
-  A. CHAMBOLLE, CH. DOSSAL, *On the convergence of the iterates of the Fast Iterative Shrinkage/Thresholding Algorithm*, Journal of Optimization Theory and Applications, 166 (2015), pp. 968–982.
-  P.-L. COMBETTES, J.-C. PESQUET, *Proximal splitting methods in signal processing*, Fixed-point algorithms for inverse problems in science and engineering, (2011), pp. 185–212.

-  J. ECKSTEIN, D. P. BERTSEKAS, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293-318.
-  J. ECKSTEIN, M. C. FERRIS, *Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control*, INFORMS J. Comput., 10 (1998), 218-235.
-  M. GHISI, M. GOBBINO, A. HARAUX, *The remarkable effectiveness of time-dependent damping terms for second order evolution equations*, SIAM J. Control Optim., 54 (2016), No. 3, 1266-1294.
-  O. GÜLER, *New proximal point algorithms for convex minimization*, SIAM J. Optim., **2** (4), (1992), 649–664.
-  O. GÜLER, *On the convergence of the proximal point algorithm for convex optimization*, SIAM J. Control Optim., 29 (1991), 403–419.

References

-  A. HARAUX, M. JENDOUBI, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, Journal of Differential Equations, **144** (2), (1998).
-  F. IUTZELER, J.M. HENDRICKX, *Generic online acceleration scheme for optimization algorithms via relaxation and inertia*, (2017), arXiv:1603.05398v3 [math.OC].
-  R. MAY, *Asymptotic for a second order evolution equation with convex potential and vanishing damping term*, Turkish Journal of Mathematics, 41 (3), (2017), pp. 681–685.
-  A. MOUDAFI, M. OLINY, *Convergence of a splitting inertial proximal method for monotone operators*, J. Comput. Appl. Math., **155** (2), (2003), pp. 447–454.
-  Y. NESTEROV, *A method of solving a convex programming problem with convergence rate $O(1/k^2)$* . Soviet Math. Doklady, **27** (1983).

References

-  Y. NESTEROV, *Introductory lectures on convex optimization*, **87**, Applied Optimization. Kluwer Academic Publishers, 2004.
-  Y. NESTEROV, *Smooth minimization of non-smooth functions*, Mathematical Programming, **103** (1) (2005), pp. 127–152.
-  Y. NESTEROV, *Gradient methods for minimizing composite objective function*, CORE Discussion Papers, 2007.
-  Z. OPIAL, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., **73** (1967), pp. 591–597.
-  B. O DONOGHUE AND E. J. CANDÈS, *Adaptive restart for accelerated gradient schemes*, Found. Comput. Math., 2013.

-  J. PEYPOUQUET, S. SORIN, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, J. Convex Anal, 17 (2010), No. 3-4, pp. 1113–1163.
-  B.T. POLYAK, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. Math. Phys., 4 (1964), pp. 1–17.
-  B.T. POLYAK, *Introduction to optimization*. New York: Optimization Software. (1987).
-  D. A. LORENZ AND THOMAS POCK, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vision, 2014.
-  A.S. NEMIROVSKY AND D.B. YUDIN, *Problem complexity and method efficiency in optimization*, Wiley, New York, 1983.

-  M. SCHMIDT, N. LE ROUX, F. BACH, *Convergence Rates of Inexact Proximal- Gradient Methods for Convex Optimization*, NIPS'11 -Grenada, Spain. 2011. <inria-00618152v3>, HAL.
-  B. SHI, S. S. DU, M. I. JORDAN, W. J. SU, *Understanding the acceleration phenomenon via high-resolution differential equations*, arXiv:submit/2440124[cs.LG] 21 Oct 2018.
-  W. SU, S. BOYD, E. J. CANDÈS, *A Differential Equation for Modeling Nesterov's Accelerated Gradient Method*, Advances in Neural Information Processing Systems **27** (NIPS 2014).