Inertial dynamical systems with vanishing damping. Fast algorithms for nonsmooth convex optimization.

Hedy ATTOUCH

Université Montpellier
Institut Montpelliérain Alexander Grothendieck, UMR CNRS 5149

Workshop "FBP 2017"
Mathematical problems with subdifferential condition

Saint-Etienne (Faculté des Sciences).

20-21 juin 2017

1a. Inertial dynamics, Asymptotic Vanishing Damping

$\mathcal{H}$ Hilbert, $\Phi : \mathcal{H} \to \mathbb{R}$ convex, differentiable, $\nabla\Phi$ loc Lip.

$$\min\left\{\Phi(x) : \quad x \in \mathcal{H}\right\}, \qquad S = \operatorname{argmin}\Phi \neq \emptyset.$$

Inertial dynamics with Asymptotic Vanishing Damping, $\alpha > 0$

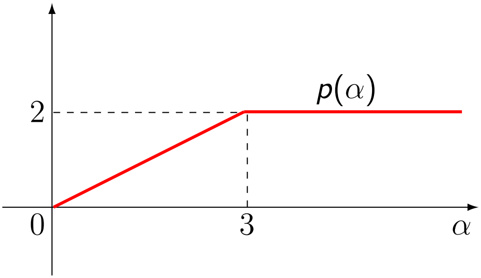
$$(\text{AVD})_\alpha \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

- $\alpha \geq 3$: Su-Boyd-Candès (NIPS, 2014):
  - $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(\frac{1}{t^2})$;
  - Link with Nesterov accelerated gradient method: $\alpha = 3$.
- $\alpha > 3$: A-Chbani-Peypouquet-Redont (Math Prog 16), May (15)
  - $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o(\frac{1}{t^2})$;
  - $x(t) \rightharpoonup x_\infty \in S \quad as \quad t \to +\infty$.
- $\alpha < 3$: A-Chbani-Riahi (2017)
  - $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$

1b. Inertial dynamics, Asymptotic Vanishing Damping

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

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{p(\alpha)}}\right), \qquad p(\alpha) = \min\left(\frac{2\alpha}{3}, 2\right).$$



1c. Inertial dynamics & Optimization: brief history

$\gamma > 0$ fixed: Heavy Ball with Friction, Polyak (87)

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

- Convergence: Haraux-Jendoubi (98) analytic, Alvarez (2000) convex.
- $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(t^{-1})$.

$\gamma(t)$ vanishing: $\lim_{t \to +\infty} \gamma(t) = 0$.

$$\ddot{x}(t) + \gamma(t)\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

- Cabot-Engler-Gadat (09)
  $$\int_{t_0}^{+\infty} \gamma(t)dt = +\infty \implies \Phi(x(t)) \to \min_{\mathcal{H}} \Phi.$$
- Cabot-Frankel (12), May (15),
  $$\gamma(t) = \frac{1}{t^\theta},\ 0 < \theta < 1 \Longrightarrow \Phi(x(t)) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{t^{1+\theta}}\right).$$
- Su-Boyd-Candès (14), A-Chbani-Peypouquet-Redont (16)
  $$\gamma(t) = \frac{\alpha}{t},\ \alpha \geq 3 \Longrightarrow \Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(t^{-2}).$$

2a. Inertial dynamics and Nesterov method.

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

Discretization: explicit /smooth Φ.

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

$$\Updownarrow$$

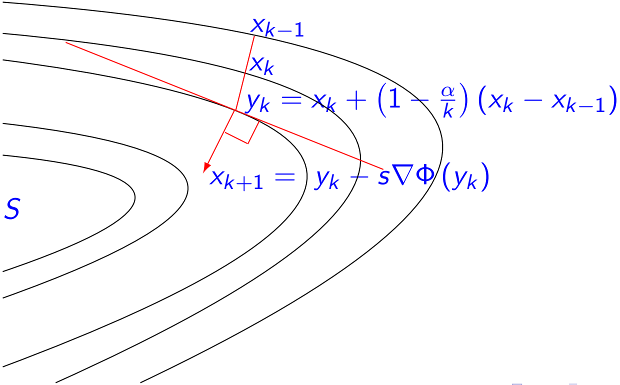
$$x_{k+1} = \left(x_k + \left(1 - \tfrac{\alpha}{k}\right)(x_k - x_{k-1})\right) - h^2\nabla\Phi(y_k).$$

Classical choice (Nesterov): $y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}), \; s = h^2$

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

Accelerated gradient method: Φ convex differentiable.

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



2c. Inertial dynamics and Nesterov method.

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

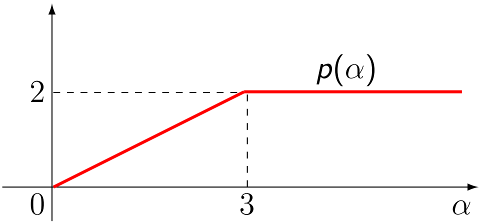
- $\alpha = 3$: Nesterov (1983, 2007)
  - $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}(\frac{1}{k^2})$;
  - Convergence of (x_k): open question.
- $\alpha > 3$: Chambolle-Dossal (2015), Attouch-Peypouquet (2016)
  - $\Phi(x_k) - \min_{\mathcal{H}} \Phi = o(\frac{1}{k^2})$;
  - $x_k \rightharpoonup \bar{x} \in S$: weak convergence.
- $\alpha \leq 3$: A-Chbani-Riahi (2017)
  - $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$;

2d. Inertial dynamics and Nesterov method.

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

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

First-order methods: $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$ worst case optimal.



3a. Sketch of the proof. Lyapunov analysis.

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

Theorem 1 (SBC, ACPR, ACR)

Let $\Phi : \mathcal{H} \to \mathbb{R}$ *be a convex continuously differentiable function such that* $\operatorname{argmin}\Phi \neq \emptyset$. *Let* $x(\cdot)$ *be a classical global solution of* $(\mathrm{AVD})_\alpha$.

- $\alpha \geq 3$: $\quad \Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}(\frac{1}{t^2})$.
- $0 < \alpha \leq 3$: $\quad \Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}(\frac{1}{t^{\frac{2\alpha}{3}}})$.

Lyapunov function, $z \in \operatorname{argmin}\Phi$, $p > 0$, $\lambda(\cdot)$, $\xi(\cdot)$ parameters

$$\mathcal{E}(t) = t^{2p}\left[\Phi(x(t)) - \min_{\mathcal{H}}\Phi\right] + \frac{1}{2}\|\lambda(t)(x(t)-z) + t^p\dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t)-z\|^2$$

$$= t^{2p}W(t) + (\lambda(t)^2 + \xi(t))h(t) + \lambda(t)t^p\dot{h}(t)$$

3b. Proof of the convergence rate.

$$\mathcal{E}(t) = t^{2p}\left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right] + \frac{1}{2}\|\lambda(t)(x(t)-z) + t^p \dot{x}(t)\|^2 + \frac{\xi(t)}{2}\|x(t)-z\|^2$$

$$\begin{aligned}
&\frac{d}{dt}\mathcal{E}(t) \leq t^p\left[2pt^{p-1} - \lambda(t))\right](\Phi(x(t)) - \min_{\mathcal{H}}\Phi) \\
&+ \left[\xi(t) + t^p\dot{\lambda}(t) - (\alpha - p)t^{p-1}\lambda(t) + \lambda(t)^2\right]\langle x(t)-z, \dot{x}(t)\rangle \\
&- t^p\left[(\alpha-p)t^{p-1} - \lambda(t)\right]\|\dot{x}(t)\|^2 + \left[\lambda(t)\dot{\lambda}(t) + \frac{\dot{\xi}(t)}{2}\right]\|x(t)-z\|^2.
\end{aligned}$$

(H$_1$): $\lambda(t) = 2pt^{p-1}$ $\Rightarrow$ first term equal to zero.

(H$_2$): $\xi(t) = 2(\alpha - 4p + 1)pt^{2(p-1)}$ $\Rightarrow$ second term equal to zero.

(H$_3$): $\alpha + 1 \geq 4p \Rightarrow \xi \geq 0$, and hence $\mathcal{E} \geq 0$.

(H$_4$): $\alpha \geq 3p \Rightarrow$ third term less or equal than zero.

(H$_5$): $1 \geq p \Rightarrow$ fourth term less or equal than zero.

Take $p = \min(1, \frac{\alpha}{3}, \frac{\alpha+1}{4}) = \min(1, \frac{\alpha}{3})$.

4. Case of a strong minimum

The convex function Φ has a strong minimum $x^* \in \mathcal{H}$, if

$$\exists \eta > 0 \; s.t. \; \Phi(x) \geq \Phi(x^*) + \frac{\eta}{2}\|x - x^*\|^2 \quad \forall x \in \mathcal{H}.$$

The convergence rate increases indefinitely with larger values of α.

Theorem 2 (SBC, ACPR, AC)

Suppose that the convex function $\Phi : \mathcal{H} \to \mathbb{R}$ has a strong minimum $x^* \in \mathcal{H}$, and let $x : [t_0, +\infty[\to \mathcal{H}$ be a solution of $(\mathrm{AVD})_\alpha$ with $\alpha > 0$. Then $x(t)$ converges strongly to x^* the unique element of $\operatorname{argmin} \Phi$, and

(i) $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{t^{\frac{2\alpha}{3}}}\right)$

(ii) $\|x(t) - x^*\|^2 = \mathcal{O}\left(t^{-\frac{2}{3}\alpha}\right)$,

(iii) $\|\dot{x}(t)\| = \mathcal{O}(t^{-\frac{1}{3}\alpha})$.

4b. Example $\Phi(x) = \frac{1}{2}\|x\|^2$. Role of Bessel functions

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

Solution of $(\text{AVD})_\alpha$ with Cauchy data $x(0) = x_0$, $\dot{x}(0) = 0$:

$$x(t) = 2^{\frac{\alpha-1}{2}}\Gamma(\frac{\alpha+1}{2})\frac{J_{\frac{\alpha-1}{2}}(t)}{t^{\frac{\alpha-1}{2}}}x_0.$$

$J_{\frac{\alpha-1}{2}}(\cdot)$: first kind Bessel function of order $\frac{\alpha-1}{2}$. For large t,

$$J_\alpha(t) = \sqrt{\frac{2}{\pi t}}\left(\cos\left(t - \frac{\pi\alpha}{2} - \frac{\pi}{4}\right) + \mathcal{O}(\tfrac{1}{t})\right).$$

Hence

$$\Phi(x(t)) - \min_{\mathcal{H}}\Phi = \mathcal{O}(t^{-\alpha}).$$

Compare with $\mathcal{O}(t^{-\frac{2}{3}\alpha})$, valid for arbitrary strongly convex functions.

5a. Nonsmooth structured optimization

Nonsmooth structured convex minimization.

$$\min\{\Phi(x)+\Psi(x): \quad x\in\mathcal{H}\}.$$

- $\mathcal{H}$ real Hilbert space.
- $\Phi:\mathcal{H}\to\mathbb{R}$ convex, differentiable, $\nabla\Phi$ L-Lipschitz continuous.
- $\Psi:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ convex, lower semicontinuous, proper.
- $S=\operatorname{argmin}(\Phi+\Psi)\neq\emptyset$.

Model examples.

- Sparse optimization:

$$\min\left\{\|Ax-b\|_2^2+\lambda\|x\|_1: \quad x\in\mathbb{R}^n\right\}.$$

- Constrained optimization ($\delta_C=0$ on C, $+\infty$ elsewhere):

$$\min\{\Phi(x)+\delta_C(x): \quad x\in\mathcal{H}\}.$$

5b. Fast splitting algorithms

Structured convex minimization.

$$\min \{\Phi(x) + \Psi(x) : \quad x \in \mathcal{H}\}.$$

- $\Phi : \mathcal{H} \to \mathbb{R}$ convex, differentiable, $\nabla\Phi$ Lipschitz continuous.
- $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, convex, lower semicontinuous.

Splitting algorithms. Basic blocks:

- Gradient (forward, explicit) w.r. smooth function Φ.
- Proximal (backward, implicit) w.r. nonsmooth function Ψ.

Combine $\rightsquigarrow$ Proximal-Gradient, Forward-Backward algorithms.

Inertial dynamics with vanishing damping $\rightsquigarrow$ fast algorithms.

- Fast minimization (in the worst case):

$$(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = o\left(\frac{1}{k^2}\right).$$

5c. Classical Proximal-Gradient: Gradient methods

$\Phi : \mathcal{H} \to \mathbb{R}$ convex, $\mathcal{C}^1$, $\nabla\Phi$ L-Lipschitz, $S = \text{argmin}\Phi \neq \emptyset$.

$$\min \{\Phi(x) : \quad x \in \mathcal{H}\}.$$

Gradient flow (continuous steepest descent).

$$\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

Brézis, Bruck, Baillon $(70 - 80)$: $x(t) \rightharpoonup x_\infty \in S$ weakly as $t \to +\infty$.

Time discretization (explicit, forward): Gradient algorithm.

- $\frac{1}{s}(x_{k+1} - x_k) + \nabla\Phi(x_k) = 0 \iff x_{k+1} = x_k - s\nabla\Phi(x_k)$.
- $0 < s < \frac{2}{L}$: $x_k \rightharpoonup x_\infty \in S$ weakly, $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}(\frac{1}{k})$.

5d. Classical Proximal-Gradient: Proximal methods

$\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ closed, convex, proper, $S = \text{argmin}\Psi \neq \emptyset$.

$$\min \{\Psi(x) : \quad x \in \mathcal{H}\}.$$

Differential inclusion governed by $\partial\Psi : \mathcal{H} \rightrightarrows \mathcal{H}$ subdifferential

$$\dot{x}(t) + \partial\Psi(x(t)) \ni 0.$$

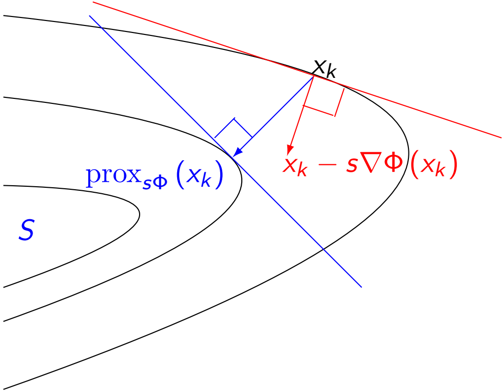
Brézis, Bruck, Baillon $(70 - 80)$: $x(t) \rightharpoonup x_\infty \in S$ weakly as $t \to +\infty$.

Time discretization (implicit, backward): Proximal algorithm.

- $\frac{1}{s}(x_{k+1} - x_k) + \partial\Psi(x_{k+1}) \ni 0 \iff x_{k+1} = \text{prox}_{s\Psi}(x_k)$.
- $\text{prox}_{s\Psi}(x) = \text{armin}_{\xi\in\mathcal{H}} \left\{s\Psi(\xi) + \frac{1}{2}\|x - \xi\|^2\right\} = (I + s\partial\Psi)^{-1}(x)$.
- $s > 0$, $x_k \rightharpoonup x_\infty \in S$ weakly, $\Psi(x_k) - \min_{\mathcal{H}} \Psi = \mathcal{O}(\frac{1}{k})$.

$$\Phi : \mathcal{H} \to \mathbb{R} \text{ convex, } \mathcal{C}^1.$$

Gradient: $x_{k+1} = x_k - s\nabla\Phi(x_k)$ Proximal: $x_{k+1} = \text{prox}_{s\Phi}(x_k)$



5f. Classical Proximal-Gradient algorithm

$$\min\left\{\Phi(x)+\Psi(x): \quad x\in\mathcal{H}\right\}.$$

- $\Phi:\mathcal{H}\to\mathbb{R}$ convex, $\mathcal{C}^1$, $\nabla\Phi$ L-Lipschitz continuous.
- $\Psi:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ proper, convex, lower semicontinuous.

Differential inclusion with two potentials.

$$\dot{x}(t)+\nabla\Phi(x(t))+\partial\Psi(x(t))\ni 0.$$

Finite difference discretization: step $s>0$, $t_k=ks$, $x_k=x(t_k)$.

Explicit /smooth function Φ. Implicit /nonsmooth function Ψ.

$$s^{-1}(x_{k+1}-x_k)+\nabla\Phi(x_k)+\partial\Psi(x_{k+1})\ni 0;$$

$$\Updownarrow$$

$$x_{k+1}=\operatorname{prox}_{s\Psi}\left(x_k-s\nabla\Phi(x_k)\right).$$

5g. Classical Proximal-Gradient algorithm

Goldstein (64), Levitin-Polyak (66), Lions-Mercier (79), Passty (79),...

- $\Phi : \mathcal{H} \to \mathbb{R}$ convex, $\mathcal{C}^1$, $\nabla\Phi$ L-Lipschitz continuous.
- $\Psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ proper, convex, lower semicontinuous.
- $0 < s < \frac{2}{L}$.

$$x_{k+1} = \text{prox}_{s\Psi}\left(x_k - s\nabla\Phi(x_k)\right).$$

- $i)$ $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = \mathcal{O}(\frac{1}{k})$;
- $ii)$ $x_k \rightharpoonup \bar{x} \in S$.

Examples

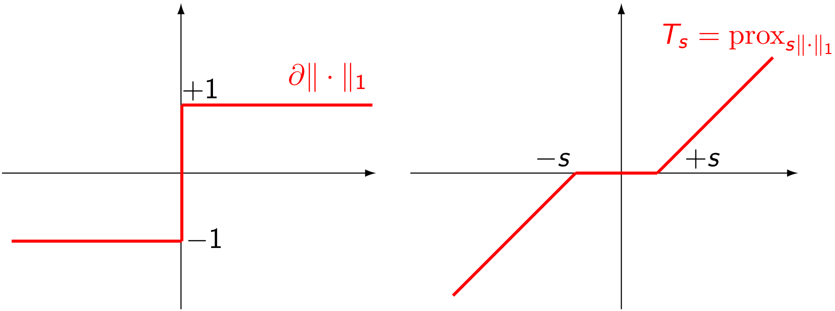
- $\Psi = \delta_C$: Gradient-Projection, $x_{k+1} = \text{proj}_C\left(x_k - s\nabla\Phi(x_k)\right)$.
- $\Psi = \|\cdot\|_1$: ISTA, Iterative Soft Thresholding Algorithm.

5h. Classical Proximal-Gradient algorithm: ISTA

ISTA: $\Psi = \lambda \| \cdot \|_1$, Iterative Soft Thresholding Algorithm.

$$\min \left\{ \Phi(x) + \lambda \|x\|_1 : \quad x \in \mathbb{R}^n \right\}.$$

$$x_{k+1} = T_s \left(x_k - \frac{s}{\lambda} \nabla \Phi(x_k) \right), \quad 0 < s < \frac{2\lambda}{L}.$$



6a. Inertial dynamics: nonsmooth potential.

Non-smooth structured convex minimization

$$\min\{\Phi(x)+\Psi(x):\ x\in\mathcal{H}\}.$$

- $\Phi:\mathcal{H}\to\mathbb{R}$ convex, differentiable, $\nabla\Phi$ Lipschitz continuous.
- $\Psi:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ closed, convex, proper;

Dynamic approach via the differential inclusion

$$\ddot{x}(t)+\frac{\alpha}{t}\dot{x}(t)+\nabla\Phi(x(t))+\partial\Psi(x(t))\ni 0.$$

- Unilateral mechanics, shocks: Paoli-Schatzman, A-Cabot-Redont. x loc. Lipschitz; $\dot{x}$ bounded variation; $\ddot{x}$ bounded measure.
- Lyapunov analysis is still valid: convex subdifferential inequalities, generalized derivation chain rule.

6b. Link with inertial dynamics: algorithms.

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) + \partial\Psi(x(t)) \ni 0.$$

Discretization: explicit /smooth Φ, implicit /nonsmooth Ψ.

$$\frac{1}{h^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\alpha}{kh^2}(x_k - x_{k-1}) + \nabla\Phi(y_k) + \partial\Psi(x_{k+1}) \ni 0.$$

$$\Updownarrow$$

$$x_{k+1} + h^2\partial\Psi(x_{k+1}) \ni \left(x_k + \left(1 - \tfrac{\alpha}{k}\right)(x_k - x_{k-1})\right) - h^2\nabla\Phi(y_k).$$

Classical choice (Nesterov): $y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1})$, $s = h^2$

$$(\mathrm{IPG})_\alpha \quad \begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}); \\ x_{k+1} = \mathrm{prox}_{s\Psi}\left(y_k - s\nabla\Phi(y_k)\right). \end{cases}$$

6c. Inertial Proximal-Gradient algorithm: $(\text{IPG})_\alpha$

Structured convex minimization. $\mathcal{H}$ real Hilbert space

$$\min\left\{\Phi(x)+\Psi(x): \quad x\in\mathcal{H}\right\}.$$

- $\Phi:\mathcal{H}\to\mathbb{R}$ convex, $\mathcal{C}^1$, $\nabla\Phi$ L-Lipschitz continuous.
- $\Psi:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ convex, lower semicontinuous, proper.
- $S=\operatorname{argmin}(\Phi+\Psi)\neq\emptyset$.

Theorem 3 (Chambolle-Dossal; A-Peypouquet; A.-Chbani-Riahi) $0<s\leq\frac{1}{L}$

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

$\alpha>3$: $(\Phi+\Psi)(x_k)-\min_{\mathcal{H}}(\Phi+\Psi)=o\left(\frac{1}{k^2}\right)$; $x_k\rightharpoonup\bar{x}\in S$.

$\alpha\leq 3$: $(\Phi+\Psi)(x_k)-\min_{\mathcal{H}}(\Phi+\Psi)=\mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$;

6d. $(\text{IPG})_\alpha$ algorithm and FISTA

$\alpha = 3$: Nesterov (1983, 2007), Beck-Teboulle (2009)

$$(\text{FISTA}) \begin{cases} y_k &= x_k + \left(1 - \frac{3}{k}\right)(x_k - x_{k-1}) \\ x_{k+1} &= \text{prox}_{s\Psi}\left(y_k - s\nabla\Phi(y_k)\right) \end{cases}$$

- $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = \mathcal{O}(\frac{1}{k^2})$;
- Convergence of (x_k): open question.

$\alpha > 3$: SBC (2014), CD 2015), AP(2016), ACR(2017)

.

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

- $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = o(\frac{1}{k^2})$;
- $x_k \rightharpoonup \bar{x} \in S$: weak convergence.

7. Newton-like inertial dynamic with Hessian damping

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

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

Proposition (A-Peypouquet-Redont, JDE 2016)

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

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

- First order system in time and space.
- In the product space: linear perturbation of a gradient system.
- Nonsmooth setting: similar results (damped shocks in mechanics).
- Time discretization gives inertial Newton-like algorithms.

8. Adaptive restart (SBC)

Strategy: maintain high velocity along the orbit.

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

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

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

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

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

Theorem (SBC), linear convergence

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

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

9. Nesterov acceleration with Tikhonov regularization.

$$(\text{AVD})_{\alpha,\epsilon} \qquad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) + \epsilon(t)x(t) = 0,$$

Theorem (A-Chbani, JMAA 2016; critical size $\epsilon(t) = \frac{1}{t^2}$)

- Fast vanishing: $\epsilon(t) = \frac{1}{t^r}$, $r > 2$, $\alpha > 3$

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi = o(t^{-2}); \quad x(t) \rightharpoonup x_\infty \in \text{argmin}_{\mathcal{H}}\Phi \ \text{ weakly.}$$

- Slow: $\epsilon(t) = \frac{1}{t^r}$, $0 < r < 2$, or $\epsilon(t) = \frac{c}{t^2}$, $\alpha > 3$, $c > \frac{4}{9}\alpha(\alpha - 3)$.

$$\liminf_{t\to\infty} \|x(t) - x^*\| = 0, \text{ where } x^* = \text{proj}_{\text{argmin}\Phi} 0.$$

- Very slow vanishing: $\epsilon(t) = \frac{1}{(\ln t)^\gamma}$, $0 < \gamma \le 1$:

$$\lim_{t\to+\infty} \frac{1}{\int_{t_0}^t \frac{\epsilon(\tau)}{\tau} d\tau} \int_{t_0}^t \frac{\epsilon(\tau)}{\tau} \|x(\tau) - x^*\| d\tau = 0.$$

10. Perspective, open questions

- $\alpha = 3$ critical?
  - Convergence of the orbits for Nesterov algorithm ($\alpha = 3$)?
  - Convergence of the values: exhibit concrete examples showing that $\alpha = 3$ is critical.
  - Numerical evidence taking α large. Variant Liang-Fadili-Peyré.
- $(\text{DIN-AVD})_{\alpha,\beta}$.
  - Algorithmic version.
  - Adaptive restart for $(\text{DIN-AVD})_{\alpha,\beta}$, without strong convexity.
- Non-convex setting, semialgebraic, Kurdyka-Lojasiewicz. For analytic potentials, the convergence theory for HBF (Haraux-Jendoubi, Bolte et al.), and DIN (AABR) still works.
- Combine Nesterov and Tikhonov.
  - Obtening both strong convergence and fast convergence.
  - Accelerate Haugazeau method (closed-loop Tikhonov method).
- Develop $(\text{AVD})_{\alpha}$ for other (splitting) methods, for maximal monotone operators, primal-dual methods.

References

B. Abbas, H. Attouch, and B. F. Svaiter, *Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces*, JOTA, **161** (2014), No 2, 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** (2000), No 4, 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.

References

F. Alvarez, H. Attouch, *Convergence and asymptotic stabilization for some damped hyperbolic equations with non-isolated equilibria*, ESAIM Control Optim. Calc. of Var., **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** (2002), No. 8, pp. 747–779.

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, Philadelphia, PA, Second edition, 2014, 793 pages.

References

- H. Attouch, A. Cabot, *Asymptotic stabilization of inertial gradient dynamics with time-dependent viscosity*, 2017. HAL-01453108.
- H. Attouch, A. Cabot, *Convergence rates of inertial forward-backward algorithms*, 2017. HAL-01453170.
- H. Attouch, A. Cabot, P. Redont, *The dynamics of elastic shocks via epigraphical regularization of a differential inclusion*, Adv. Math. Sci. Appl., **12** (2002), № 1, pp. 273–306.
- H. Attouch and M.-O. Czarnecki, *Asymptotic control and stabilization of nonlinear oscillators with non-isolated equilibria*, J. Differential Equations, **179** (2002), pp. 278–310.

References

H. Attouch, Z. Chbani, J. Peypouquet, P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing damping*, to appear in Math. Program. DOI: 10.1007/s10107-016-0992-8.

H. Attouch, X. Goudou and P. Redont, *The heavy ball with friction method*, Commun. Contemp. Math., **2** (2000), № 1, pp. 1–34.

References

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* $\frac{1}{k^2}$, to appear in SIOPT, Arxiv:1510.08740v2 [math.OC] 1 Nov 2015.

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.

References

J.-B. Baillon, *Un exemple concernant le comportement asymptotique de la solution du problème* $\frac{du}{dt} + \partial\phi(u) \ni 0$, Journal of Functional Analysis, **28** (1978), pp. 369–376.

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.

H. Brézis, *Opérateurs maximaux monotones dans les espaces de Hilbert et équations d'évolution*, Lecture Notes 5, North Holland, (1972).

References

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.

References

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.

O. Güler, *New proximal point algorithms for convex minimization*, SIAM J. Optim., **2** (4), (1992), pp. 649–664.

References

A. Haraux, M. Jendoubi, *Convergence of solutions of second-order gradient-like systems with analytic nonlinearities*, Journal of Differential Equations, **144** (2), (1998).

R. May, *Asymptotic for a second order evolution equation with convex potential and vanishing damping term*, arXiv:1509.05598.

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/k2)*. In Soviet Mathematics Doklady, **27** (1983), pp. 372–376.

Y. Nesterov, *Introductory lectures on convex optimization*, **87**, Applied Optimization. Kluwer Academic Publishers, 2004.

References

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.

References

- J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time*, Journal of Convex Analysis, **17**, (2010), pp. 1113–1163.
- 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.
- 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).