

Inertial algorithms for monotone inclusions. Combining Nesterov acceleration with relaxation.

Hedy ATTOUCH

Université Montpellier
Institut Montpelliérain Alexander Grothendieck, UMR CNRS 5149

Based on recent works with A. Cabot and J. Peypouquet.

GDO2018, Games, Dynamics and Optimization
Universität Wien, COST CA16228

March 13-15, 2018

Introduction: Inertial dynamics for monotone inclusions

\mathcal{H} Hilbert, $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ maximally monotone operator. Solve $Ax \ni 0$

- $A = \partial\Phi \quad \leftrightarrow$ convex minimization problems.
- $A = (\partial_x L, -\partial_y L) \quad \leftrightarrow$ convex-concave saddle value problems.
- $A = I - T \quad \leftrightarrow$ fixed points of nonexpansive operators.

Damped inertial continuous dynamics \leftrightarrow Accelerated algorithms.

- $A = \nabla\Phi$, Φ convex: $\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + \nabla\Phi(x(t)) = 0$, $\alpha \geq 3$.

Su-Boyd-Candès (NIPS 2014), $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = O(\frac{1}{t^2})$, link & Nesterov.

- A maximally monotone: $\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0$.

Nesterov acceleration & Yosida regularization.

A.-Peypouquet (Math Prog. 2018), A.-Cabot (JDE. 2018)

- ① Convex Minimization. Fast inertial algorithms.
- ② Monotone Inclusions. [Regularized inertial dynamics](#).
- ③ Monotone Inclusions. [Relaxed inertial proximal algorithms](#).
- ④ Combining Inertia & Relaxation: general approach.
- ⑤ An inertial ADMM algorithm.

1a. Inertial approach to Nesterov accelerated method

Asymptotic Vanishing Damping, $A = \nabla\Phi$, Φ convex, $\alpha > 0$, Su-Boyd-Candès.

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

Discretization: explicit /smooth Φ .

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

$$\begin{aligned} & \Downarrow \\ x_{k+1} &= \left(x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \right) - h^2 \nabla\Phi(y_k). \end{aligned}$$

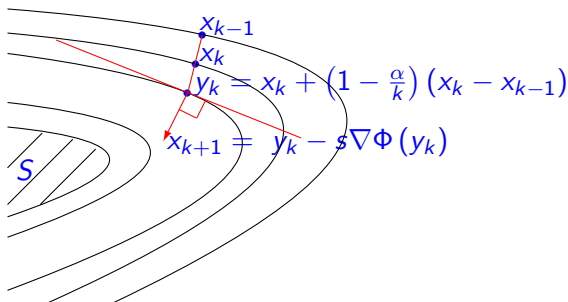
Nesterov choice: $y_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1})$, $s = h^2$

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

1b. Inertial approach to Nesterov accelerated method.

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

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



1c. Inertial approach to Nesterov accelerated method

$\min \{ \Phi(x) : x \in \mathcal{H} \}$, Φ convex, $\nabla \Phi$ L -Lipschitz, $S = \operatorname{argmin} \Phi \neq \emptyset$.

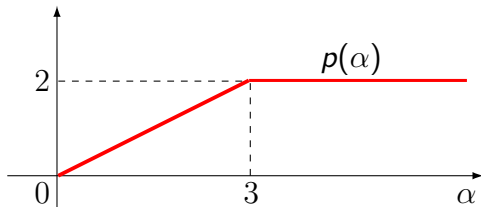
Inertial Gradient algorithm, $\alpha > 0$, $s \leq \frac{1}{L}$

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

- $\alpha = 3$: $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$, Nesterov (1983).
- $\alpha > 3$: $\Phi(x_k) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{k^2}\right)$, A.-Peypouquet (SIOPT 2016).
 $x_k \rightarrow \hat{x} \in S$, Chambolle-Dossal (JOTA 2015).
- $\alpha \leq 3$: $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$, A.-Chbani-Riahi (COCV 2018)
Apidopoulos-Aujol-Dossal (HAL, 2017).

1d. Inertial approach to Nesterov accelerated method

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



Rate of convergence of the values:

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

1e. Inertial Proximal-Gradient algorithm: $(\text{IPG})_\alpha$

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

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

- $\alpha = 3$: FISTA, Beck-Teboulle (SIAM J. Imaging 2009):
 - $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = \mathcal{O}\left(\frac{1}{k^2}\right)$.
- $\alpha > 3$: $x_k \rightarrow x_\infty \in \mathcal{S} = \text{argmin}(\Phi + \Psi)$, Chambolle-Dossal (JOTA 2015).
 - $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = o\left(\frac{1}{k^2}\right)$, A-Peypouquet (SIOPT 2016).
- $\alpha \leq 3$: Apidopoulos-Aujol-Dossal, A.-Chbani-Riahi (COCV 2018)
 - $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = \mathcal{O}\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$.

1f. General damping/ inertial coefficient

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

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

Theorem (A.-Cabot (HAL 2017), to appear in SIOPT)

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 $(\text{IPG})_{\alpha(\cdot)}$

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

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

$$(\Phi + \Psi)(x_k) - \min(\Phi + \Psi) = o\left(\frac{1}{\sum_{i=1}^k t_i}\right).$$

If $\alpha_k \in [0, 1]$ for every $k \geq 1$, then $x_k \rightarrow \hat{x} \in \text{argmin}(\Phi + \Psi)$.

1g. General damping/ inertial coefficient

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

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

- 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 $\alpha > 3$.
- $0 \leq m \leq \alpha_k \leq M < 1$, $\frac{1-m}{1-M} < \frac{3}{2}$: $W_k = o\left(\frac{1}{k}\right)$ (HBF).

2a. Inertial dynamics for maximally monotone operator

$A : \mathcal{H} \rightarrow \mathcal{H}$ λ -cocoercive ($\lambda > 0$)

$$\forall (v, w) \in \mathcal{H} \times \mathcal{H} \quad \langle Av - Aw, v - w \rangle \geq \lambda \|Av - Aw\|^2.$$

A λ -cocoercive $\implies A$ maximally monotone, $\frac{1}{\lambda}$ -Lipschitz continuous.

Heavy Ball with Friction system, $\gamma > 0$ damping coefficient.

$$(HBF) \quad \ddot{x}(t) + \gamma \dot{x}(t) + A(x(t)) = 0, \quad t \geq 0.$$

Alvarez-A.: Lecture Notes in Math. (2000), $A = I - T$, $\lambda = \frac{1}{2}$, $\gamma > \sqrt{2}$.

Theorem (A.-Maingé, ESAIM-COCV 2011)

Suppose $A : \mathcal{H} \rightarrow \mathcal{H}$ max. monotone, λ -cocoercive, $S = A^{-1}(0) \neq \emptyset$, and
 $\lambda\gamma^2 > 1$.

Then, for each solution $x(\cdot)$ of (HBF), $x(t) \rightarrow \hat{x} \in S$ as $t \rightarrow +\infty$.

2b. Inertial dynamics for maximally monotone operator

- $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ maximally monotone operator;
- $J_{\lambda A} = (I + \lambda A)^{-1}$ resolvent of index $\lambda > 0$ of A ;
- $A_{\lambda} = \frac{1}{\lambda}(I - J_{\lambda A})$ Yosida regularization of index $\lambda > 0$ of A .

Regularized Inertial Monotone System

$$\text{(RIMS)}_{\alpha, \lambda} \quad \ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0.$$

Yosida regularization and $\text{(RIMS)}_{\alpha, \lambda}$

- (i) A_{λ} Lipschitz continuous \implies Well-posed Cauchy problem.
- (ii) $A_{\lambda}^{-1}(0) = A^{-1}(0)$ \implies Preservation of the solution set.
- (iii) A_{λ} λ -cocoercive \implies A.-Maingé setting.

Tuning of $t \mapsto \lambda(t) > 0$: $\lambda \gamma^2 > 1 \implies \lambda(t) > \frac{t^2}{\alpha^2}$.

2c. Inertial dynamics for maximally monotone operator

- $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ maximally monotone, $S = A^{-1}(0) \neq \emptyset$ closed convex set.
- $A_{\lambda} : \mathcal{H} \rightarrow \mathcal{H}$ Yosida approximation of index $\lambda > 0$ of A .

$$\text{(RIMS)}_{\alpha,\lambda} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0.$$

$x(\cdot) : [t_0, +\infty[\rightarrow \mathcal{H}$ solution trajectory of $\text{(RIMS)}_{\alpha,\lambda}$.

Theorem (A.-Peypouquet, Math. Prog. 2018)

Suppose $\lambda(t) = (1 + \epsilon)\frac{t^2}{\alpha^2}$, $\alpha > 2$, $\epsilon > \frac{2}{\alpha-2}$.

Then, $x(t)$ converges weakly to an element of S as $t \rightarrow +\infty$.

Theorem (A.-Peypouquet, A.-Cabot, JDE. 2018)

Suppose $A = \partial\Phi$, $\Phi \in \Gamma_0(\mathcal{H})$, $\lambda(\cdot)$ nondecreasing, C^1 , $\lambda(t) \leq Ct^2$.

$\alpha \geq 3$: $\Phi(p(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(\frac{1}{t^2})$, $p(t) = \text{prox}_{\lambda(t)\Phi} x(t)$.

$\alpha > 3$: $x(t)$ converges weakly to an element of S , $\|x(t) - p(t)\| \rightarrow 0$.

2d. Inertial dynamics for maximally monotone operator

$A_\lambda : \mathcal{H} \rightarrow \mathcal{H}$ Yosida approximation of A of index $\lambda > 0$, $S = A^{-1}(0) \neq \emptyset$.

$$\text{(RIMS)}_{\alpha,\lambda} \quad \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0.$$

Theorem (A.-Peypouquet, Math. Prog. 2018)

Suppose $\lambda(t) = (1 + \epsilon)\frac{t^2}{\alpha^2}$, $\alpha > 2$, $\epsilon > \frac{2}{\alpha-2}$. Then,

- (i) $x(t)$ converges weakly, as $t \rightarrow +\infty$, to an element of S .
- (ii) $\|\dot{x}(t)\| = \mathcal{O}(1/t)$, $\|\ddot{x}(t)\| = \mathcal{O}(1/t^2)$ and $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty$.

Lemma (Opial)

Let $S \subset \mathcal{H}$, $S \neq \emptyset$, and let $x : [t_0, +\infty[\rightarrow \mathcal{H}$. Assume that

- (i) for every $z \in S$, $\lim_{t \rightarrow +\infty} \|x(t) - z\|$ exists;
 - (ii) every weak sequential limit point of $x(t)$, as $t \rightarrow +\infty$, belongs to S .
- Then $x(t)$ converges weakly as $t \rightarrow +\infty$ to a point in S .

2e. Inertial dynamics for maximally monotone operator

Item *i*) Opial's lemma: $S = A^{-1}(0)$, $h_z(t) := \frac{1}{2}\|x(t) - z\|^2$, $z \in S$

a) $\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \lambda(t)\|A_{\lambda(t)}(x(t))\|^2 \leq \|\dot{x}(t)\|^2.$

b) $\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \epsilon\|\dot{x}(t)\|^2 + \frac{\alpha\lambda(t)}{t}\frac{d}{dt}\|\dot{x}(t)\|^2 + \lambda(t)\|\ddot{x}(t)\|^2 \leq 0$

- a) $\dot{h}_z(t) = \langle x(t) - z, \dot{x}(t) \rangle$, $\ddot{h}_z(t) = \langle x(t) - z, \ddot{x}(t) \rangle + \|\dot{x}(t)\|^2 + (\text{RIMS})_{\alpha,\lambda} \Rightarrow$

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \langle A_{\lambda(t)}(x(t)), x(t) - z \rangle = \|\dot{x}(t)\|^2.$$

$A_{\lambda(t)}$ $\lambda(t)$ -cocoercive $\Rightarrow \langle A_{\lambda(t)}(x(t)), x(t) - z \rangle \geq \lambda(t)\|A_{\lambda(t)}(x(t))\|^2 \hookrightarrow a).$

- b) Replace $A_{\lambda(t)}(x(t)) = -\ddot{x}(t) - \frac{\alpha}{t}\dot{x}(t) \Rightarrow$

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \left(\lambda(t)\frac{\alpha^2}{t^2} - 1\right)\|\dot{x}(t)\|^2 + \alpha\frac{\lambda(t)}{t}\frac{d}{dt}\|\dot{x}(t)\|^2 + \lambda(t)\|\ddot{x}(t)\|^2 \leq 0.$$

Use $\lambda(t) = (1 + \epsilon)\frac{t^2}{\alpha^2} \hookrightarrow b).$

- Integrate b) $\Rightarrow \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty \Rightarrow \int_{t_0}^{+\infty} \dot{h}_z(t) dt < +\infty$

$\Rightarrow \lim_{t \rightarrow +\infty} h(t)$ exists. Integrate a) $\Rightarrow \int_{t_0}^{+\infty} t\lambda(t)\|A_{\lambda(t)}(x(t))\|^2 dt < +\infty.$

2f. Inertial dynamics for maximally monotone operator

Item *ii*) of Opial's lemma.

$$\int_{t_0}^{+\infty} t\lambda(t)\|A_{\lambda(t)}(x(t))\|^2 dt < +\infty \Rightarrow \int_{t_0}^{+\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\|^2 \frac{1}{t} dt < +\infty.$$

Central point: this property implies $(\star) \lim_{t \rightarrow +\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\| = 0$.

Suppose (\star) . Let $x(t_n) \rightarrow \bar{x}$ weakly. From $\lambda(t_n)A_{\lambda(t_n)}(x(t_n)) \rightarrow 0$ and $\lambda(t_n) \rightarrow +\infty$, we have $A_{\lambda(t_n)}(x(t_n)) \rightarrow 0$ strongly. Passing to the limit in

$$A_{\lambda(t_n)}(x(t_n)) \in A(x(t_n) - \lambda(t_n)A_{\lambda(t_n)}(x(t_n)))$$

and using the demi-closedness of A , we obtain $0 \in A(\bar{x})$, i.e., $\bar{x} \in S$.

Prove (\star) . Estimate the variation of $t \mapsto \lambda(t)A_{\lambda(t)}$. Resolvent equation:

$$\left\| \frac{d}{dt} (\lambda(t)A_{\lambda(t)}x(t)) \right\| \leq 2\|\dot{x}(t)\| + 2\|x(t) - z\| \frac{|\dot{\lambda}(t)|}{\lambda(t)} \leq \frac{2C+4(M+\|z\|)}{t}.$$

$w(t) := \|\lambda(t)A_{\lambda(t)}x(t)\|$ satisfies $\left| \frac{d}{dt} w(t) \right| \leq \frac{C}{t}$, and $\int_{t_0}^{+\infty} w^2(t) \frac{1}{t} dt < +\infty$.

$$\Rightarrow \left| \frac{d}{dt} w(t)^3 \right| = \left| 3w(t)^2 \frac{d}{dt} w(t) \right| \leq \frac{C}{t} w(t)^2 \in L^1(t_0, +\infty).$$

Hence $\lim w(t)$ exists, and because of $\int_{t_0}^{+\infty} w^2(t) \frac{1}{t} dt < +\infty$, this limit is zero.

2g. Inertial dynamics for maximally monotone operator

- $\mathcal{H} = \mathbb{R}^2$, $A = \text{rot}(0, \frac{\pi}{2})$, $A(x, y) = (-y, x)$.
- A linear antisymmetric: $\langle A(x, y), (x, y) \rangle = 0$ for all $(x, y) \in \mathcal{H}$.
- A maximal monotone operator, not cocoercive, $A^{-1}(0) = 0$.

Find conditions on $\lambda(t)$ ensuring the convergence of $u(t)$ to zero.

$$(\text{RIMS})_{\alpha, \lambda} \quad \ddot{u}(t) + \frac{\alpha}{t} \dot{u}(t) + A_{\lambda(t)}(u(t)) = 0, \quad u(t) = (x(t), y(t)).$$

Equivalent formulation

$\mathcal{H} = \mathbb{C}$, real Hilbert $\langle z_1, z_2 \rangle = \text{Re}(z_1 \bar{z}_2)$. $Az = iz$, $A_\lambda z = \frac{\lambda+i}{1+\lambda^2} z$.

Set $z(t) = x(t) + iy(t)$. $(\text{RIMS})_{\alpha, \lambda}$ becomes $\ddot{z}(t) + \frac{\alpha}{t} \dot{z}(t) + \frac{\lambda+i}{1+\lambda^2} z(t) = 0$.

Phase space $\mathbb{C} \times \mathbb{C}$, $Z(t) = (z(t), \dot{z}(t))^T$. First-order equivalent system

$$\dot{Z}(t) + M(t)Z(t) = 0, \quad M(t) = \begin{pmatrix} 0 & -1 \\ \frac{\lambda(t)+i}{1+\lambda(t)^2} & \frac{\alpha}{t} \end{pmatrix}.$$

2h. Inertial dynamics for maximally monotone operator

Spectral analysis

$$\dot{Z}(t) + M(t)Z(t) = 0, \quad M(t) = \begin{pmatrix} 0 & -1 \\ \frac{\lambda(t)+i}{1+\lambda(t)^2} & \frac{\alpha}{t} \end{pmatrix}.$$

$$\text{Eigenvalues of } M(t) : \quad \theta(t) = \frac{\alpha}{2t} \left\{ 1 \pm \sqrt{1 - \frac{4t^2}{\alpha^2} \frac{\lambda(t)+i}{1+\lambda(t)^2}} \right\}.$$

Case $\lambda(t) \sim t^p$

Suppose $p > 2$. The eigenvalues θ_+ and θ_- satisfy

$$\theta_+(t) \sim \frac{\alpha}{t} \quad \text{and} \quad \theta_-(t) \sim \frac{1}{\alpha t^{p-1}}.$$

- The solutions of $\dot{v}(t) + \frac{\alpha}{t}v(t) = 0$, $\alpha > 0$, converge to 0.
- The solutions of $\dot{v}(t) + \frac{1}{\alpha t^{p-1}}v(t) = 0$ do not.

To obtain the convergence results of our theorem, we are not allowed to let $\lambda(t)$ tend to infinity at a rate greater than t^2 : t^2 is a critical size for $\lambda(t)$.

2i. Inertial dynamics for maximally monotone operator

Numerical illustration, $\mathcal{H} = \mathbb{R}^2$, $A = \text{rot}(0, \frac{\pi}{2})$

- Initial condition at $t_0 = 1$ is $(10, 10)$. For second-order equations, the initial velocity is $(0, 0)$ in order not to force the system in any direction.
- When relevant, $\lambda(t) = (1 + \epsilon)t^2/\alpha^2$ with $\alpha = 10$ and $\epsilon = 1 + 2(\alpha - 2)^{-1}$. For the constant λ , we set $\lambda = 10$.

Key	Differential Equation	Distance to $(0, 0)$ at $t = 100$
(E1)	$\dot{x}(t) + Ax(t) = 0$	14.141911
(E2)	$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + Ax(t) = 0$	3.186e24
(E3)	$\dot{x}(t) + A_{\lambda(t)}(x(t)) = 0$	0.0135184
(E4)	$\dot{x}(t) + A_{\lambda}(x(t)) = 0$	0.0007827
(E5)	$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A_{\lambda(t)}x(t) = 0$	0.000323

(E4) is a first-order equation governed by the *strongly monotone* operator A_{λ} .

3a. Relaxed Inertial Proximal Algorithm

- $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ maximally monotone operator, $S = A^{-1}(0) \neq \emptyset$.
- $A_{\lambda} : \mathcal{H} \rightarrow \mathcal{H}$ Yosida approximation of A of index $\lambda > 0$.

$$\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) + A_{\lambda(t)}(x(t)) = 0, \quad t > t_0 > 0.$$

Discretization: time step $h > 0$, $t_k = kh$, $x_k = x(t_k)$, $\lambda_k = \lambda(t_k)$, $s = h^2$.

Resolvent equation

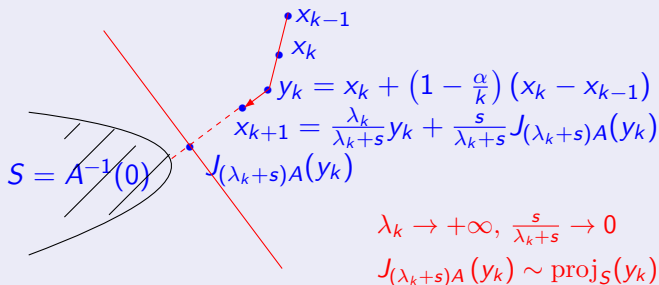
$$(A_{\lambda})_s = A_{\lambda+s} \implies (I + sA_{\lambda})^{-1} = \frac{\lambda}{\lambda+s} I + \frac{s}{\lambda+s} (I + (\lambda + s)A)^{-1}.$$

Implicit finite-difference \rightarrow Relaxed Inertial Proximal Algorithm

$$(RIPA) \quad \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} (I + (\lambda_k + s)A)^{-1} (y_k). \end{cases}$$

3b. Relaxed Inertial Proximal Algorithm

$$(RIPA) \quad \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} J_{(\lambda_k + s)A}(y_k). \end{cases}$$



3c. Relaxed Inertial Proximal Algorithm

$$(RIPA) \quad \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} J_{(\lambda_k + s)A}(y_k). \end{cases}$$

Theorem (A.-Peypouquet, Math. Prog. 2018)

Let $A : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ be a maximally monotone operator, $S = A^{-1}(0) \neq \emptyset$. Let (x_k) be a sequence generated by (RIPA) where $s > 0$, $\alpha > 2$ and

$$\lambda_k = (1 + \epsilon) \frac{s}{\alpha^2} k^2$$

for some $\epsilon > \frac{2}{\alpha-2}$ and all $k \geq 1$. Then,

- i) (x_k) converges weakly, as $k \rightarrow +\infty$, to some $\hat{x} \in S$.
- ii) (y_k) converges weakly, as $k \rightarrow +\infty$, to \hat{x} .
- iii) $\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right)$ as $k \rightarrow +\infty$, and $\sum_k k \|x_k - x_{k-1}\|^2 < +\infty$.

3d. Relaxed Inertial Proximal Algorithm

$A = \partial\Phi$, $\Phi : \mathcal{H} \rightarrow \mathbb{R} \cup +\{\infty\}$ **convex** lsc., proper, $S = \operatorname{argmin} \Phi \neq \emptyset$.

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

Let (x_k) be a sequence generated by algorithm (RIPA).

Theorem (A.-Peypouquet, A.-Cabot JDE. 2018)

Suppose that (λ_k) is a **nondecreasing** sequence, $s > 0$.

• **Case $\alpha \geq 3$:** $\Phi_{\lambda_k + s}(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}(k^{-2})$.

As a consequence, setting $p_k = \operatorname{prox}_{(\lambda_k + s)\Phi}(x_k)$, we have

$$\Phi(p_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}(k^{-2}), \quad \text{and} \quad \|x_k - p_k\|^2 = \mathcal{O}\left(\frac{\lambda_k}{k^2}\right).$$

• **Case $\alpha > 3$:** Suppose moreover that $\sup_k \frac{\lambda_k}{k^2} < +\infty$.

Then $x_k \rightarrow \hat{x} \in S$, $\Phi(p_k) - \min_{\mathcal{H}} \Phi = o(k^{-2})$, $\lim_{k \rightarrow +\infty} \|p_k - x_k\| = 0$.

4a. Combining Inertia & Relaxation: general approach

$$(H) \quad \left\{ \begin{array}{l} \bullet A : \mathcal{H} \rightarrow 2^{\mathcal{H}} \text{ is a maximally monotone operator;} \\ \bullet (\alpha_k) \text{ is a sequence of nonnegative numbers;} \\ \bullet (\mu_k) \text{ and } (\rho_k) \text{ are sequences of positive numbers.} \end{array} \right.$$

Relaxed Inertial Proximal Algorithm, (A.-Cabot, HAL 2018)

$$(RIPA) \quad \left\{ \begin{array}{l} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = (1 - \rho_k)y_k + \rho_k J_{\mu_k A}(y_k). \end{array} \right.$$

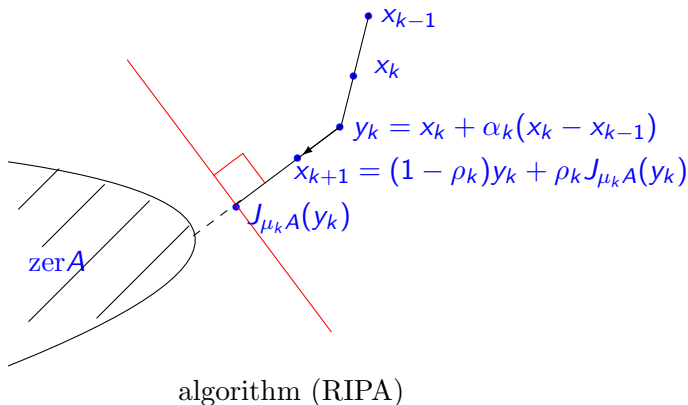
If $\rho_k = 1$ for every $k \geq 1$: (RIPA) = Inertial Proximal Algorithm

If $\alpha_k = 0$ for every $k \geq 1$: (RIPA) = Relaxed Proximal Algorithm

$$(RPA) \quad x_{k+1} = (1 - \rho_k)x_k + \rho_k J_{\mu_k A}(x_k).$$

Bot-Csetnek (SIOPT, 2016), Iutzeler-Hendrickx (arXiv 2017).

4b. Combining Inertia & Relaxation: general approach



4c. Combining Inertia & Relaxation: general approach

$$(K_0) \sum_{l=i}^{+\infty} \left(\prod_{j=i}^l \alpha_j \right) < +\infty \quad \forall i \geq 1, \quad t_i := 1 + \sum_{l=i}^{+\infty} \left(\prod_{j=i}^l \alpha_j \right), \quad \alpha_i = \frac{t_i - 1}{t_{i+1}}.$$

Theorem (A.-Cabot, HAL 2018)

Under (H), assume that $S \neq \emptyset$. Suppose that $\alpha_k \in [0, 1]$ and $\rho_k \in]0, 2]$ for every $k \geq 1$. Under (K_0) , assume that $\exists \epsilon \in]0, 1[$ s.t. for k large enough,

$$(1 - \epsilon) \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \geq \alpha_k t_{k+1} \left(1 + \alpha_k + \left[\frac{2 - \rho_k}{\rho_k} (1 - \alpha_k) - \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \right]_+ \right).$$

Then for any sequence (x_k) generated by (RIPA), we have

$$(i) \sum_{i=1}^{+\infty} \alpha_i t_{i+1} \|x_i - x_{i-1}\|^2 < +\infty, \quad \sum_{i=1}^{+\infty} \rho_i (2 - \rho_i) t_{i+1} \|\mu_i A_{\mu_i}(x_i)\|^2 < +\infty.$$

(ii) For any $z \in S$, $\lim_{k \rightarrow +\infty} \|x_k - z\|$ exists, and hence (x_k) is bounded.

Assume moreover that $\limsup_{k \rightarrow +\infty} \rho_k < 2$, and $\liminf_{k \rightarrow +\infty} \rho_k > 0$. Then

$$(iii) \lim_{k \rightarrow +\infty} \mu_k A_{\mu_k}(x_k) = 0.$$

(iv) If $\liminf_{k \rightarrow +\infty} \mu_k > 0$, then $x_k \rightharpoonup x_\infty \in S$ weakly in \mathcal{H} as $k \rightarrow +\infty$.

4d. Combining Inertia & Relaxation: general approach

α_k	$\alpha_k = 1 - \frac{\alpha}{k}$ $\alpha > 2$	$\alpha_k = 1 - \frac{\alpha}{k^q}$ $q \in]0, 1[, \alpha > 0$	$\alpha_k = 1 - \frac{\alpha}{k^q}$ $q \in]0, 1[, \alpha > 0$	$\alpha_k \equiv \alpha \in [0, 1[$
ρ_k	$\rho_k = \frac{\beta}{k^2}$ $\beta < \alpha(\alpha - 2)$	$\rho_k = \frac{\beta}{k^{2q}}$ $\beta < \alpha^2$	$\rho_k = \frac{\beta}{k^r}$ $2q < r \leq q + 1$	$\rho_k \equiv \rho \in]0, 2[$ $\rho < \frac{2(1-\alpha)^2}{2\alpha^2 - \alpha + 1}$
μ_k	$\frac{ \mu_{k+1} - \mu_k }{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k}\right)$ $\liminf \mu_k > 0$	$\frac{ \mu_{k+1} - \mu_k }{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k^q}\right)$ $\liminf \mu_k > 0$	$\frac{ \mu_{k+1} - \mu_k }{\mu_{k+1}} = \mathcal{O}\left(\frac{1}{k^{r-q}}\right)$ $\liminf \mu_k > 0$	$\liminf \mu_k > 0$

5a. An inertial proximal ADMM algorithm

Convex structured minimization with linear constraint

$$(P) \quad \min \{f(x) + g(y) : Ax - By = 0\}$$

- $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ real Hilbert spaces.
- $f : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$, $g : \mathcal{Y} \rightarrow \mathbb{R} \cup \{+\infty\}$ closed convex proper.
- $A : \mathcal{X} \rightarrow \mathcal{Z}$, $B : \mathcal{Y} \rightarrow \mathcal{Z}$ linear continuous operators.
- λ positive real parameter.

Lagrangian formulation

$$(P) \iff \min_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \max_{z \in \mathcal{Z}} \{f(x) + g(y) + \langle z, Ax - By \rangle\}$$

Maximally monotone formulation

$$(P) \iff M(x, y, z) \ni 0$$

$$M(x, y, z) = (\partial f(x) + A^t z, \partial g(y) - B^t z, By - Ax).$$

5b. An inertial proximal ADMM algorithm

Proximal splitting method for ADMM, Eckstein (1994), A.-Soueycatt (2011) .
Work in progress: A.-Bot-Fadili-Peypouquet:

$$\left\{ \begin{array}{l} u_k = x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ v_k = y_k + \left(1 - \frac{\alpha}{k}\right) (y_k - y_{k-1}) \\ w_k = z_k + \left(1 - \frac{\alpha}{k}\right) (z_k - z_{k-1}) \\ \frac{1}{\lambda_k + s} (p_k - u_k) + \partial f(p_k) + A^t (w_k + (\lambda_k + s)(Ap_k - Bv_k)) \ni 0 \\ \frac{1}{\lambda_k + s} (q_k - v_k) + \partial g(q_k) - B^t (w_k + (\lambda_k + s)(Ap_k - Bq_k)) \ni 0 \\ r_k = w_k + (\lambda_k + s)(Ap_k - Bq_k) \\ x_{k+1} = \frac{\lambda_k}{\lambda_k + s} u_k + \frac{s}{\lambda_k + s} p_k. \\ y_{k+1} = \frac{\lambda_k}{\lambda_k + s} v_k + \frac{s}{\lambda_k + s} q_k. \\ z_{k+1} = \frac{\lambda_k}{\lambda_k + s} w_k + \frac{s}{\lambda_k + s} r_k. \end{array} \right.$$

6. Recent trends

- ① Inertial splitting methods for structured monotone inclusions:
 - Forward-backward, proximal ADMM, Douglas-Rachford.
- ② Damping term = control.
 - From open-loop to closed-loop.
 - Adaptive restart. (O'Donoghue-Candès, 2013).
 - Coupling Nesterov with Tikhonov: strong & fast convergence to the minimum norm solution, A.-Chbani-Riahi (JMAA, 2018).
- ③ Geometrical damping.
 - Hessian driven damping; Combine Nesterov acceleration with Newton and Levenberg-Marquardt methods.
Alvarez-A.-Bolte-Redont (J. Math. Pures Appl., 2002)
A.-Peypouquet-Redont (SIOPT, 2014).
 - Dry friction, Adly-A.-Cabot (Adv. Mech. Math., 2006).

Annex 1. Perturbations, errors

$$(\text{IPG})_{\text{pert}} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \text{prox}_{s\Psi}(y_k - s\nabla\Phi(y_k) + sg_k). \end{cases}$$

Theorem (A.-Cabot-Chbani-Riahi, 2017)

A. Suppose that (α_k) satisfies (K_0) and (K_1) , and that (g_k) satisfies

$$(K_2) \quad \sum_{k=0}^{+\infty} t_{k+1} \|g_k\| < +\infty.$$

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

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

B. Assume moreover (K_1^+) and $\alpha_k \in [0, 1]$ for every $k \geq 1$. Then the sequence (x_k) converges weakly toward some $\bar{x} \in \text{argmin}(\Phi + \Psi)$.

C. If we assume additionally that $(K_2^+) \quad \sum_{k=1}^{+\infty} \left(\frac{1}{t_{k+1}} \sum_{i=1}^k t_{i+1}\right) \|g_k\| < +\infty$, then we have

$$(\Phi + \Psi)(x_k) - \min(\Phi + \Psi) = o\left(\frac{1}{\sum_{i=1}^k t_i}\right) \quad \text{as } k \rightarrow +\infty.$$

Annex 2. Perturbations, Tikhonov regularization

$$(\text{IFB})_{\text{Tikh}} \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}); \\ x_{k+1} = \text{prox}_{S\Psi}(y_k - s\nabla\Phi(y_k) - s\epsilon_k y_k), \end{cases}$$

Theorem (A.-Cabot-Chbani-Riahi, 2017)

Let x^* be the least norm element of $S = \text{argmin}(\Phi + \Psi)$. Suppose that

- (i) The sequence (t_k) is nondecreasing, satisfies (K_0) , (K_1) , $\sum_k \frac{1}{t_k^2} < +\infty$.
- (ii) The sequence (ϵ_k) is nonincreasing, and verifies $\sum_k \frac{\epsilon_k}{t_{k+1}} = +\infty$.

Let (x_k) be a sequence generated by the algorithm $(\text{IFB})_{\text{Tikh}}$. Then (x_k) converges strongly to x^* in the ergodic sense

$$\lim_{k \rightarrow +\infty} \left\| \frac{1}{\tau_k} \sum_{j=1}^k r_j x_j - x^* \right\| = 0, \text{ with } r_j = \frac{\epsilon_j}{t_{j+1}} \text{ and } \tau_k = \sum_{j=1}^k r_j.$$

Annex 3a. $\alpha = 3$. Convergence of trajectories.

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

Let $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ be a convex continuously differentiable function such that $S = \operatorname{argmin} \Phi \neq \emptyset$. Let $x : [t_0; +\infty[\rightarrow \mathcal{H}$ be a solution of $(AVD)_\alpha$ with $\alpha = 3$. Then $x(t)$ converges, as $t \rightarrow +\infty$, to a point in S .

For $\alpha = 3$, $x(\cdot)$ is bounded, and minimizing.

When $\operatorname{argmin} \Phi = \{x^*\}$, $x(\cdot)$ converges to its unique cluster point x^* .

When $\operatorname{argmin} \Phi = [a, b]$, there are **three possible cases**:

- $\exists T \geq t_0$ s.t. $x(t) \geq b$ for all $t \geq T$. Then b is the unique cluster point of the trajectory, which implies the convergence of $x(\cdot)$ to b . Symetrically, if $x(t) \leq a$, for all $t \geq T$, then $x(\cdot)$ converges to a .
- $\exists T \geq t_0$ s. t., for all $t \geq T$, $a \leq x(t) \leq b$. Then, $\nabla \Phi(x(t)) = 0$. Integration of $\ddot{x}(t) + \frac{\alpha}{t} \dot{x}(t) = 0$ gives $\dot{x}(t) = \frac{C}{t^\alpha}$. Since $\alpha > 1$, $\dot{x}(\cdot)$ is integrable, and hence $x(\cdot)$ converges.
- $x(\cdot)$ passes in a and b an infinite number of times.

Annex 3b. $\alpha = 3$. Convergence of trajectories

Lemma 1 (\mathcal{H} general Hilbert space)

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

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

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

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

Set $z = x(t_1) = x(t_2) \in S = \operatorname{argmin} \Phi$, take $\rho = \min(1, \frac{\alpha}{3})$, and consider

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

which is the Lyapunov function of Theorem 1. It is nonincreasing. Hence

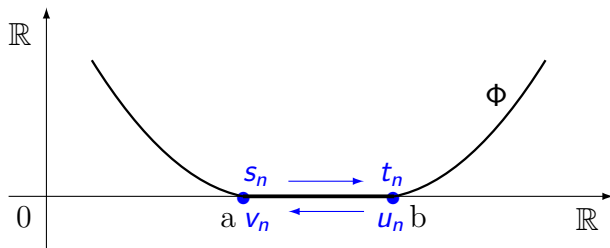
$\mathcal{E}_{\lambda, \xi}^p(t_2) \leq \mathcal{E}_{\lambda, \xi}^p(t_1)$, which equivalently gives

$$\frac{1}{2} \|t_2^p \dot{x}(t_2)\|^2 \leq \frac{1}{2} \|t_1^p \dot{x}(t_1)\|^2.$$

Annex 3c. $\alpha = 3$. Convergence of trajectories

- 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.



- $s_n \leq t_n \leq u_n \leq v_n$
- $x(s_n) = a, x(t_n) = b, a \leq x(t) \leq b$ for all $t \in [s_n, t_n]$
- $x(u_n) = b, x(v_n) = a, a \leq x(t) \leq b$ for all $t \in [u_n, v_n]$.

Annex 3d. $\alpha = 3$. Convergence of trajectories

For $t \in [s_n, t_n]$ we have $t\ddot{x}(t) + \alpha\dot{x}(t) = 0$. Equivalently

$$\frac{d}{dt}(t\dot{x}(t)) + (\alpha - 1)\dot{x}(t) = 0.$$

After integration on $[s_n, t_n]$, and taking account of the sign of \dot{x}

$$|t_n\dot{x}(t_n)| = |s_n\dot{x}(s_n)| - (\alpha - 1)(b - a).$$

Symmetrically,

$$|v_n\dot{x}(v_n)| = |u_n\dot{x}(u_n)| - (\alpha - 1)(b - a).$$

By Lemma 1

$$|u_n\dot{x}(u_n)| \leq |t_n\dot{x}(t_n)|.$$

Combining the above equalities, we obtain

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

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

Annex 4. Inertial dynamic & control: Hessian damping

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

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

Theorem (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.

Annex 5. Inertial dynamics & control: Adaptive restart

Strategy: maintain high velocity along the orbit.

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

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

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





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





Question: adaptive restart for a general convex function Φ ?




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 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** (2002), No. 8, pp. 747–779.
-  V. APIDOPOULOS, J.-F. AUJOL, CH. DOSSAL, *Convergence rate of inertial Forward- Backward algorithm beyond Nesterov's rule*, 2017, HAL-01551873.
-  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.






-  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*, 2017. HAL-01453170, accepted in SIOPT.
-  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** (2002), No 1, 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, Issue 3, (2017), pp. 2745–2770.





-  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, 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, 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, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM J. Optim., **24** (2014), No. 1, pp. 232–256.

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. BOTŽ, 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), pp. 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, pp. 1266-1294.
-  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).
-  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 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).