Fast inertial dynamics and algorithms for solving monotone inclusions

Hedy ATTOUCH

Université Montpellier
Institut Montpelliérain Alexander Grothendieck, UMR CNRS 5149

Based on a work with P. Redont and J. Peypouquet

Journées annuelles 2017 des GdR MOA et MIA
Institut de Mathématiques de Bordeaux

17-20 Octobre 2017

1. Inertial dynamic for max. monotone operators

$\mathcal{H}$ real Hilbert space, $A : \mathcal{H} \to 2^{\mathcal{H}}$ maximal monotone operator.

Solve $Ax \ni 0$

- $A = \partial\Phi$ $\hookrightarrow$ convex minimization.
- $A = (\partial_x L, -\partial_y L)$ $\hookrightarrow$ convex-concave saddle value problem.
- $A = I - T$ $\hookrightarrow$ fixed point of nonexpansive operator.

Guideline: Dissipative dynamic systems, link with algorithms.

Heavy Ball with Friction: $\ddot{x}(t) + \gamma\dot{x}(t) + A(x(t)) = 0$

- $A = \nabla\Phi$, Φ convex , $\Phi(x(t)) - \min\Phi = O(\frac{1}{t})$ (Alvarez, 2000).
- A λ-cocoercive, $\lambda\gamma^2 > 1 \Rightarrow$ convergence, (A.-Maingé, 2011).

Asymptotic Vanishing Damping: $\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + A(x(t)) = 0$

- $A = \nabla\Phi$, $\alpha \geq 3$, $\Phi(x(t)) - \min\Phi = O(\frac{1}{t^2})$ (Su-Boyd-Candès, '16).

2a. Inertial approach to Nesterov accelerated method

Asymptotic Vanishing Damping, $A = \nabla\Phi$, $\alpha \geq 3$, (Su-Boyd-Candès, 2016).

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

Discretization: explicit /smooth Φ.

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

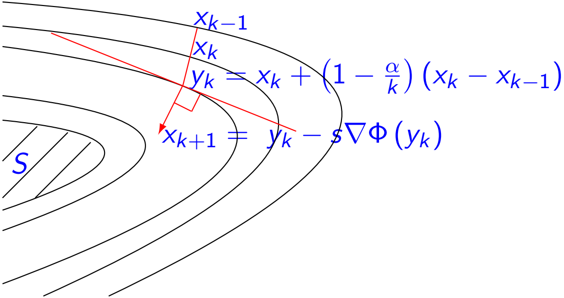
Classical choice (Nesterov): $y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1})$, $s = h^2$

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

2b. Inertial approach to Nesterov accelerated method.

$$\min\{\Phi(x):\ x\in\mathcal{H}\},\ \Phi:\mathcal{H}\to\mathbb{R} \text{ convex differentiable.}$$

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



2c. Inertial approach to Nesterov accelerated method

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

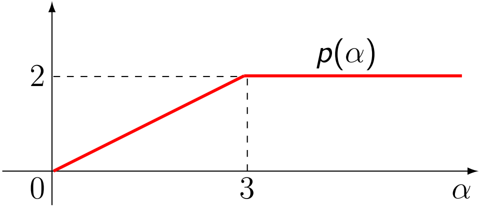
Inertial Gradient algorithm, $\alpha > 0$, $s \leq \frac{1}{L}$

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

- $\alpha = 3$: $\Phi(x_k) - \min_{\mathcal{H}} \Phi = \mathcal{O}\left(\frac{1}{k^2}\right)$, (Nesterov, 1983).
- $\alpha > 3$: $\Phi(x_k) - \min_{\mathcal{H}} \Phi = o\left(\frac{1}{k^2}\right)$, (A.-Peypouquet, SIOPT 2016).
  $x_k \rightharpoonup \bar{x} \in S$, (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, Aujol-Dossal.

2d. Inertial approach to Nesterov accelerated method

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



Rate of convergence of the values

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

$A : \mathcal{H} \to \mathcal{H}$ 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 $\Rightarrow$ A maximal monotone, $\frac{1}{\lambda}$-Lipschitz continuous.

Heavy ball with friction system, $\gamma > 0$ damping coefficient.

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

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

Theorem (A.-Maingé, ESAIM-COCV 2011)

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

$$\lambda \gamma^2 > 1.$$

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

4a. Fast inertial dynamic for max. monotone operators

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

Fast Inertial Regularized SysTem

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

Yosida regularization and (FIRST)

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

Tuning of $t \mapsto \lambda(t) > 0$ in (FIRST)

$$\lambda\gamma^2 > 1 \Rightarrow \lambda(t)\left(\frac{\alpha}{t}\right)^2 > 1.$$

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

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

$x(\cdot)$ solution trajectory of (FIRST).

Theorem (A.-Peypouquet, 2017)

Suppose $\lambda(t) = (1+\epsilon)\frac{t^2}{\alpha^2}$, $\epsilon > \frac{2}{\alpha - 2}$, $\alpha > 2$.
Then, $x(t)$ *converges weakly, as* $t \to +\infty$, *to an element of* S.

Theorem (A.-Peypouquet-Redont, 2017)

Suppose $A = \partial\Phi$, $\Phi \in \Gamma_0(\mathcal{H})$, $\lambda(\cdot)$ *nondecreasing,* $\alpha \geq 3$.
Then, $\Phi(p(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(\frac{1}{t^2})$, $p(t) = prox_{\lambda(t)\Phi}x(t)$.

4c. Fast inertial dynamic for max. monotone operator

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

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

$x(\cdot)$ solution trajectory of (FIRST).

Theorem (A.-Peypouquet, 2017)

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

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

(ii) $\|\dot{x}(t)\| = \mathcal{O}(1/t)$, $\|\ddot{x}(t)\| = \mathcal{O}(1/t^2)$ *and* $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2\, dt < +\infty$.

Proof: based on Opial's Lemma and the convergence of $h_z(\cdot)$, $z \in S$,

$$h_z(t) = \frac{1}{2}\|x(t) - z\|^2.$$

Step 1: $$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \lambda(t)\|A_{\lambda(t)}(x(t))\|^2 \leq \|\dot{x}(t)\|^2.$$

- $\dot{h}_z(t) = \langle x(t) - z, \dot{x}(t)\rangle$, $\quad \ddot{h}_z(t) = \langle x(t) - z, \ddot{x}(t)\rangle + \|\dot{x}(t)\|^2$ + (FIRST) $\Rightarrow$
$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \langle A_{\lambda(t)}(x(t)), x(t) - z\rangle = \|\dot{x}(t)\|^2.$$
- $z \in S = A^{-1}(0) = A_{\lambda(t)}^{-1}(0)$, and $A_{\lambda(t)}$ $\lambda(t)$-cocoercive $\Rightarrow$
$$\langle A_{\lambda(t)}(x(t)), x(t) - z\rangle \geq \lambda(t)\|A_{\lambda(t)}(x(t))\|^2 \hookrightarrow \text{ Step 1.}$$

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

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

Then use the assumption $\lambda(t) = (1+\epsilon)\frac{t^2}{\alpha^2}$.

Step 3: $\sup_{t\geq t_0} \|x(t)\| < +\infty.$

Rewrite Step 2 with $g(t) = \|\dot{x}(t)\|^2$. From $\lambda(t) = (1+\epsilon)\frac{t^2}{\alpha^2}$, setting $\beta = \frac{1+\epsilon}{\alpha}$

$$\ddot{h}_z(t) + \frac{\alpha}{t}\dot{h}_z(t) + \epsilon g(t) + \beta t \dot{g}(t) + \lambda(t)\|\ddot{x}(t)\|^2 \leq 0.$$

$$\Rightarrow \; t\ddot{h}_z(t) + \alpha \dot{h}_z(t) + \epsilon t g(t) + \beta t^2 \dot{g}(t) \leq 0.$$

After integration

$$t\dot{h}_z(t) + (\alpha-1)h_z(t) + \beta t^2 g(t) + (\epsilon - 2\beta)\int_{t_0}^t s g(s) ds \leq C. \qquad (*)$$

By assumption $\epsilon - 2\beta = \frac{\alpha-2}{\alpha}\left(\epsilon - \frac{2}{\alpha-2}\right) > 0$. Hence,

$$t\dot{h}_z(t) + (\alpha-1)h_z(t) \leq C.$$

After integration $h_z(t) \leq \frac{C}{\alpha-1} + \frac{D}{t^{\alpha-1}}$. Hence, $h_z(\cdot)$ is bounded, and so is $x(\cdot)$.
Set $M := \sup_{t\geq t_0} \|x(t)\| < +\infty.$

Step 4: $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty.$

$$|\dot{h}_z(t)| = |\langle x(t) - z, \dot{x}(t)\rangle| \leq \|x(t) - z\|\|\dot{x}(t)\| \leq (M + \|z\|)\|\dot{x}(t)\|, \qquad (1)$$

combined with (*) gives $\beta\big(t\|\dot{x}(t)\|\big)^2 \leq C + (M + \|z\|)\,\big(t\|\dot{x}(t)\|\big)$. Hence

$$\|\dot{x}(t)\| = \mathcal{O}(1/t). \qquad (2)$$

(FIRST) $\Rightarrow$ $\|\ddot{x}(t)\| \leq \frac{\alpha}{t}\|\dot{x}(t)\| + \|A_{\lambda(t)}(x(t))\| \leq \frac{\alpha}{t}\|\dot{x}(t)\| + \frac{M+\|z\|}{\lambda(t)}$.

Using (2) and the definition of $\lambda(t) = ct^2$, we conclude that

$$\|\ddot{x}(t)\| = \mathcal{O}(1/t^2). \qquad (3)$$

Finally, returning to (*), and using (1) and (2) we infer that

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty.$$

Step 5: Apply Opial's lemma.

(i) Prove that $\forall z \in S$ $\lim_{t\to\infty} \|x(t)-z\|$ exists . By Step 1

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

Lemma

Let $w : [t_0, +\infty[\to \mathbb{R}$ be a continuously differentiable function which is bounded from below. Given a nonnegative function θ, let us assume that, for some $\alpha > 1$

$$t\ddot{w}(t) + \alpha\dot{w}(t) + \theta(t) \leq k(t),$$

and some nonnegative function $k \in L^1(t_0, +\infty)$. Then, $[\dot{w}]_+$ belongs to $L^1(t_0, +\infty)$, and $\lim_{t\to+\infty} w(t)$ exists. Moreover, we have $\int_{t_0}^{+\infty} \theta(t)dt < +\infty$.

Take $w(t) = h_z(t), \theta(t) = t\lambda(t)\|A_{\lambda(t)}(x(t))\|^2, k(t) = t\|\dot{x}(t)\|^2$. Using Step 4, we deduce that $\lim_{t\to\infty} \|x(t)-z\|$ exists for every $z \in S$, and

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

Step 6: Opial's lemma, continued.

Since $\lambda(t) = ct^2$ $(c > 0)$, we infer

$$\int_{t_0}^{+\infty} \|\lambda(t) A_{\lambda(t)}(x(t))\|^2 \frac{1}{t} dt < +\infty. \qquad (4)$$

The central point of the proof is to show that this property implies

$$\lim_{t \to +\infty} \|\lambda(t) A_{\lambda(t)}(x(t))\| = 0. \qquad (5)$$

Suppose (5). Then the end of the proof follows easily: let $t_n \to +\infty$ such that $x(t_n) \to \bar{x}$ weakly. We have $\lambda(t_n) A_{\lambda(t_n)}(x(t_n)) \to 0$. Since $\lambda(t_n) \to +\infty$, we also have $A_{\lambda(t_n)}(x(t_n)) \to 0$ 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., $\bar{x} \in S$.

Step 7: Opial's lemma, end.

It suffices to prove (5). To obtain this result, we estimate the variation of the function $t \mapsto \lambda(t)A_{\lambda(t)}$. As a consequence of the resolvent equation

$$\|\lambda(t)A_{\lambda(t)}x(t) - \lambda(s)A_{\lambda(s)}x(s)\| \leq 2\|x(t) - x(s)\| + 2\|x(t) - z\| \frac{|\lambda(t)-\lambda(s)|}{\lambda(t)}$$

for each $z \in S$. Dividing by $t - s$, $t \neq s$, and letting s tend to t

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

According to the previous estimates and the definition of λ

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

Since $\|\lambda(t)A_{\lambda(t)}x(t)\|$ is bounded, $w(t) := \|\lambda(t)A_{\lambda(t)}x(t)\|^2$ satisfies

$$\left|\frac{d}{dt}w(t)\right| \leq \eta(t), \quad \text{and} \quad \int_{t_0}^{+\infty} w(t)\,\eta(t)\,dt < +\infty,$$

for some $\eta(t) = \frac{C}{t} \notin L^1(t_0, +\infty)$. We conclude thanks to the next Lemma.

Lemma

Let $w, \eta : [t_0, +\infty[\to [0, +\infty[$ *be absolutely continuous functions s.t.*

(i) $\int_{t_0}^{+\infty} w(t)\, \eta(t)\, dt < +\infty$;

(ii) $\eta \notin L^1(t_0, +\infty)$;

(iii) $|\frac{d}{dt} w(t)| \leq \eta(t)$ *for almost every* $t > t_0$.

Then,

$$\int_{t_0}^{+\infty} |\frac{d}{dt} w^2(t)| dt < +\infty,$$

and so $\lim_{t \to +\infty} w(t) = 0$.

4k. Fast inertial dynamic for max. monotone operator

- $\mathcal{H} = \mathbb{R}^2$, $A = \mathrm{rot}(0, \frac{\pi}{2})$, $A(x, y) = (-y, x)$.
- 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.

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

Equivalent formulation

$\mathcal{H} = \mathbb{C}$, real Hilbert $\langle z_1, z_2 \rangle = Re(z_1 \bar{z}_2)$. $Az = iz$, $A_\lambda z = \frac{\lambda + i}{1+\lambda^2} z$.
Set $z(t) = x(t) + iy(t)$. (FIRST) becomes $\ddot{z}(t) + \frac{\alpha}{t}\dot{z}(t) + \frac{\lambda + i}{1+\lambda^2} z(t) = 0$.

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

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

4l. Fast inertial dynamic for max. monotone operator

Spectral analysis

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

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

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

- 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 $\lambda(t)$.

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

- 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.

5a. Regularized Inertial Proximal Algorithm

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

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

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

Resolvent equation

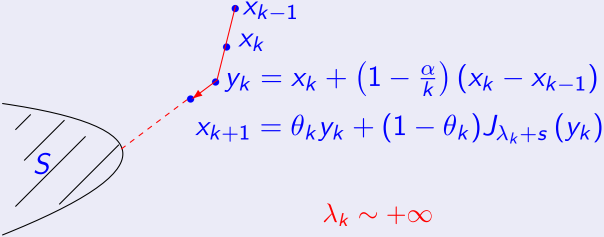
$$(A_\lambda)_s = A_{\lambda+s} \Rightarrow (I + sA_\lambda)^{-1} = \frac{\lambda}{\lambda+s} I + \frac{s}{\lambda+s}\left(I + (\lambda+s)A\right)^{-1}.$$

Implicit finite-difference → Regularized Inertial Proximal Algo

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

5b. Regularized Inertial Proximal Algorithm

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



$\lambda_k \sim +\infty$

$\theta_k \sim +1$

$J_{\lambda_k + s}(y_k) \sim \mathrm{proj}_S(y_k)$

5c. Regularized Inertial Proximal Algorithm

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

Theorem (A-Peypouquet, 2017)

Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a maximal monotone operator s.t. $S = A^{-1}(0) \neq \emptyset$. Let (x_k) be a sequence generated by (RIPA) where $\alpha > 2$ 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) $\|x_{k+1} - x_k\| = \mathcal{O}(\frac{1}{k})$ *and* $\sum_k k\|x_k - x_{k-1}\|^2 < +\infty$.

ii) *The sequence (x_k) converges weakly, as $k \to +\infty$, to some $\hat{x} \in S$.*

iii) *The sequence (y_k) converges weakly, as $k \to +\infty$, to $\hat{x}$.*

5c. Regularized Inertial Proximal Algorithm

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

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

Theorem (A-Peypouquet-Redont, 2017)

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

- Case $\alpha \geq 3$: For any sequence (x_k) generated by algorithm (RIPA)

$$\Phi_{\lambda_k + s}(x_k) - \min \Phi = \mathcal{O}(k^{-2}).$$

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

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

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

Then $x_k \rightharpoonup \hat{x} \in S$, $\Phi(p_k) - \min \Phi = o(k^{-2})$, $\lim p_k = \lim x_k$.

6a. An inertial proximal ADMM algorithm

Convex structured minimization with linear constraint

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

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

Lagrangian formulation

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

Maximal monotone formulation

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

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

6b. An inertial proximal ADMM algorithm

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

Regularized Inertial Proximal Algorithm

$$
\text{(RIPA)} \quad \begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \dfrac{\lambda_k}{\lambda_k + s} y_k + \dfrac{s}{\lambda_k + s} \left(I + (\lambda_k + s)A\right)^{-1}(y_k), \end{cases}
$$

1. Inertial prox-gradient algorithms with a general coefficient α_k. (A.-Cabot, HAL 2017).
2. Passing from open-loop to closed-loop control. Adaptive restart. (O'Donoghue and Candès, Found. Comput. Math., 2013).
3. Coupling Nesterov acceleration with Newton method. Valid for a general maximal monotone operator. (A.-Peypouquet-Redont, JDE 2016).
4. Application to saddle value problems, proximal ADMM, (A.-Soueycatt, Pacific J. Opt. 2011).
5. Coupling Nesterov with Tikhonov: strong convergence to the minimum norm solution & fast convergence, (A-Chbani-Riahi, JMMA, 2016).

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

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

Theorem 3 (A-Cabot (HAL 2017))

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* $(\mathrm{IPG})_{\alpha(\cdot)}$

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

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*

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

Annex 1b. 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).

Annex 1c. Perturbations, errors

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

Theorem 4 (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* $(\text{IPG})_{pert}$

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

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* $\overline{x} \in \operatorname{argmin}(\Phi+\Psi)$.

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

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

Annex 1d. Perturbations, Tikhonov regularization

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

Theorem 5 (A-Cabot-Chbani-Riahi, 2017)

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

(i) *The sequence (t_k) is nondecreasing, satisfies (K_0), (K_1),* $\displaystyle\sum_k \frac{1}{t_k^2} < +\infty$.

(ii) *The sequence (ϵ_k) is nonincreasing, and verifies* $\displaystyle\sum_k \frac{\epsilon_k}{t_{k+1}} = +\infty$.

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

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

Annex 2a. $\alpha = 3$. Convergence of trajectories.

Theorem 6

Let $\Phi : \mathbb{R} \to \mathbb{R}$ be a convex continuously differentiable function such that $S = \operatorname{argmin} \Phi \neq \emptyset$. Let $x : [t_0; +\infty[\to \mathcal{H}$ be a solution of $(AVD)_\alpha$ with $\alpha = 3$. Then $x(t)$ converges, as $t \to +\infty$, 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 2b. $\alpha = 3$. 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 $p = \min(1, \frac{\alpha}{3})$, and consider

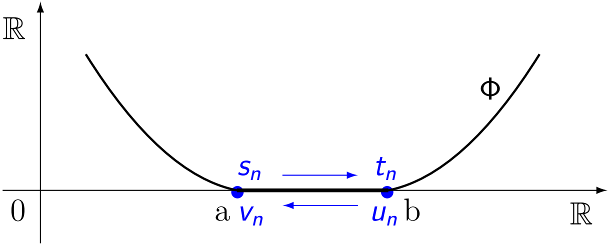
$$\mathcal{E}^p_{\lambda,\xi}(t) = t^{2p} \left[\Phi(x(t)) - \min_{\mathcal{H}} \Phi \right] + \frac{1}{2} \|\lambda(t)(x(t) - z) + t^p \dot{x}(t)\|^2 + \frac{\xi(t)}{2} \|x(t) - z\|^2$$

which is the Lyapunov function of Theorem 1. It is nonincreasing. Hence $\mathcal{E}^p_{\lambda,\xi}(t_2) \leq \mathcal{E}^p_{\lambda,\xi}(t_1)$, which equivalently gives

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

Annex 2c. $\alpha = 3$. 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 2d. $\alpha = 3$. Convergence of trajectories

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

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

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

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

Symetrically,

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

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 3. Inertial dynamic & control: 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

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 4. Inertial dynamics & control: Adaptive restart

Strategy: maintain high velocity along the orbit.

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

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 (Su-Boyd-Candès, 2016), 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

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

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 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*, J. Differential Equations, 263 (2017), pp. 5412-5458.
- H. Attouch, A. Cabot, *Convergence rates of inertial forward-backward algorithms*, 2017. HAL-01453170, SIOPT revision.
- 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, Z. Chbani, H. Riahi, *Combining fast inertial dynamics for convex optimization with Tikhonov regularization*, Journal of Mathematical Analysis and Applications, In Press, Available online 9 December 2016.

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* $\frac{1}{k^2}$, 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.

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

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.

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).