Fast convergence of a damped inertial dynamics. Link with Nesterov algorithm.

Hedy ATTOUCH

Université Montpellier 2
ACSIOM, I3M UMR CNRS 5149

Joint work with J. Peypouquet, and P. Redont

International Conference on Variational Analysis, Optimization, and quantitative Finance

in Honor of T. Rockafellar's 80th birthday
Limoges, May 18-22, 2015

1A. General presentation: dynamical system

Fast dynamical methods for convex minimization.

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

- $\mathcal{H}$ real Hilbert space; $\|x\|^2 = \langle x, x\rangle$;
- $\Phi : \mathcal{H} \to \mathbb{R}$ convex, continuously differentiable, $\mathrm{argmin}\Phi \neq \emptyset$.

Dissipative inertial system, asymptotic vanishing damping.

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

- (SBC) $\alpha \geq 3$: $\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{t^2}$;
- (APR) $\alpha > 3$: $x(t) \rightharpoonup x_\infty \in \mathrm{argmin}\Phi$ as $t \to +\infty$.

Time discretization: fast Nesterov type algorithms.

1B. General presentation: history

Heavy Ball with Friction, $\gamma > 0$.

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

- Polyak (1987), Haraux-Jendoubi (1998), A-Goudou-Redont (2000).
- Alvarez (2000), Φ convex, $x(t) \rightharpoonup x_\infty \in \text{argmin}\Phi$.

Asymptotic Vanishing Damping, $\lim_{t\to+\infty} a(t) = 0$.

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

- Cabot-Engler-Gaddat (2009)

$$\int_{t_0}^{+\infty} a(t)dt = +\infty \implies \Phi(x(t)) \to \min_{\mathcal{H}} \Phi.$$

- Su-Boyd-Candès (2014), A-Peypouquet-Redont (2015)

$$a(t) = \frac{\alpha}{t}, \ \alpha \geq 3 \Longrightarrow \Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq Ct^{-2}.$$

Contents

2A. Fast convergence of the values

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

Theorem 1 (Su-Boyd-Candès, NIPS 2014)

Suppose $\alpha \geq 3$, $t_0 > 0$, $x : [t_0, +\infty[\to \mathcal{H}$ *is an orbit of* $(\mathrm{AVD})_\alpha$. *Then,*

$$\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{t^2}.$$

Proof: for $x^* \in S = \mathrm{argmin}\Phi$, take as a Lyapunov function

$$\mathcal{E}_\alpha(t) := \tfrac{2}{\alpha-1} t^2(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + (\alpha-1)\|x(t) - x^* + \tfrac{t}{\alpha-1}\dot{x}(t)\|^2.$$

$$\dot{\mathcal{E}}_\alpha(t) + 2\frac{\alpha-3}{\alpha-1} t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \leq 0.$$

$$C = t_0^2(\Phi(x_0) - \min_{\mathcal{H}} \Phi) + (\alpha-1)^2 d^2(x_0, S) + t_0^2\|\dot{x}_0\|^2.$$

Lyapunov analysis

$$\mathcal{E}_\alpha(t) := \frac{2}{\alpha-1} t^2 (\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + (\alpha - 1)\|x(t) - x^* + \frac{t}{\alpha - 1}\dot{x}(t)\|^2.$$

Derivation of $\mathcal{E}_\alpha(\cdot)$ gives

$$\begin{aligned}
\dot{\mathcal{E}}_\alpha(t) :=& \frac{4}{\alpha-1} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + \frac{2}{\alpha-1} t^2 \langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\
&+ 2(\alpha-1)\langle x(t) - x^* + \frac{t}{\alpha-1}\dot{x}(t), \dot{x}(t) + \frac{1}{\alpha-1}\dot{x}(t) + \frac{t}{\alpha-1}\ddot{x}(t)\rangle \\
=& \frac{4}{\alpha-1} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + \frac{2}{\alpha-1} t^2 \langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \\
&+ 2(\alpha-1)\langle x(t) - x^* + \frac{t}{\alpha-1}\dot{x}(t), \frac{t}{\alpha-1}\left(\frac{\alpha}{t}\dot{x}(t) + \ddot{x}(t)\right)\rangle.
\end{aligned}$$

Then use (AVD) in this last expression to obtain

$$\dot{\mathcal{E}}_\alpha(t) = \frac{4}{\alpha-1} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) + \frac{2}{\alpha-1} t^2 \langle \nabla\Phi(x(t)), \dot{x}(t)\rangle \tag{1}$$

$$- 2t\langle x(t) - x^* + \frac{t}{\alpha-1}\dot{x}(t), \nabla\Phi(x(t))\rangle \tag{2}$$

$$= \frac{4}{\alpha-1} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) - 2t\langle x(t) - x^*, \nabla\Phi(x(t))\rangle. \tag{3}$$

By convexity of Φ

$$\Phi(x^*) \geq \Phi(x(t)) + \langle x^* - x(t), \nabla\Phi(x(t))\rangle.$$

Replacing in (3) we obtain

$$\dot{\mathcal{E}}_\alpha(t) + \left(2 - \tfrac{4}{\alpha-1}\right) t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) \leq 0.$$

$$\dot{\mathcal{E}}_\alpha(t) + 2\frac{\alpha-3}{\alpha-1} t(\Phi(x(t)) - \inf_{\mathcal{H}} \Phi) \leq 0.$$

2B. $\mathcal{O}(\frac{1}{t^2})$ as the worst possible case

$\mathcal{H} = \mathbb{R}$, $\Phi(x) = c|x|^\gamma$, $c > 0$, $\gamma > 0$, parameters.

$$\ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t) + c\gamma x(t)^{\gamma-1} = 0.$$

Nonnegative completely damped solutions of $(\mathrm{AVD})_\alpha$:

$$x(t) = \frac{1}{t^\theta}, \quad \theta > 0.$$

Replacing $x(\cdot)$ in $(\mathrm{AVD})_\alpha$ gives $\gamma > 2$, $\theta = \frac{2}{\gamma-2}$, $\alpha > \frac{\gamma}{\gamma-2}$, and

$$\Phi(x(t)) = \frac{2}{\gamma(\gamma-2)}(\alpha - \frac{\gamma}{\gamma-2})\frac{1}{t^{\frac{2\gamma}{\gamma-2}}}.$$

As $\gamma \uparrow +\infty$, $\dfrac{2\gamma}{\gamma-2} \downarrow 2$: Φ becomes very flat around its minimizer.

2C. Strong convexity: faster rate of convergence

Convergence rates increase indefinitely for larger values of α.

Theorem 2 (SBC, APR)

Suppose that $\Phi : \mathcal{H} \to \mathbb{R}$ is strongly convex. Let $x(\cdot)$ be an orbit of $(\text{AVD})_\alpha$, with $\alpha > 3$. Then

- *$x(t)$ converges strongly to the unique element $x^* \in \operatorname{argmin}\Phi$;*
- $\Phi(x(t)) - \min_{\mathcal{H}} \Phi = \mathcal{O}(t^{-\frac{2}{3}\alpha})$;
- $\|\dot{x}(t)\|^2 = \mathcal{O}(t^{-\frac{2}{3}\alpha})$;
- $\|x(t) - x^*\|^2 = \mathcal{O}(t^{-\frac{2}{3}\alpha})$.

Proof: use the Lyapunov function $\mathcal{E}_\lambda^p$ with $p = \frac{2}{3}(\alpha - 3)$, $\lambda = \frac{2}{3}\alpha$

$$\mathcal{E}_\lambda^p(t) := t^p \left(t^2(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) + \frac{1}{2}\|\lambda(x(t) - x^*) + t\dot{x}(t)\|^2 \right).$$

2D. 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.

2E. Case argmin $\Phi = \emptyset$.

Theorem 3 (APR)

Suppose $\Phi : \mathcal{H} \to \mathbb{R}$ *convex, argmin* Φ *possibly empty. Let* $x(\cdot)$ *be an orbit of* $(\mathrm{AVD})_\alpha$ *with* $\alpha > 1$. *Then,*

$$\lim_{t \to +\infty} \Phi(x(t)) = \inf_{\mathcal{H}} \Phi.$$

Moreover, if $\inf_{\mathcal{H}} \Phi > -\infty$, *then* $\lim_{t \to +\infty} \|\dot{x}(t)\| = 0$.

Fast convergence may not be satisfied in this case:

$$\Phi(x) = \frac{c}{x^\theta}, \text{ with } c = \frac{2(2\alpha + \theta(\alpha - 1))}{\theta(2+\theta)^2}.$$

Then $x(t) = t^{\frac{2}{2+\theta}}$ is solution of $(\mathrm{AVD})_\alpha$. We have $\inf_{\mathcal{H}} \Phi = 0$, and

$$\Phi(x(t)) = \frac{c}{t^{\frac{2\theta}{2+\theta}}}.$$

3A. Weak convergence of the orbits

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

Theorem 4 (APR)

Suppose $\alpha > 3$. *Let* $x : [t_0, +\infty[\to \mathcal{H}$ *be an orbit of* $(\text{AVD})_\alpha$. *Then,*

- $x(t) \rightharpoonup x^* \in \operatorname{argmin}\Phi$ *weakly as* $t \to +\infty$;
- $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0, \quad \|\dot{x}(t)\| \leq \frac{C}{t}, \quad \int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty;$
- $\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \frac{C}{t^2}, \quad \int_{t_0}^{+\infty} t\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) dt < +\infty;$
- $\lim_{t\to+\infty} \frac{1}{t^\alpha} \int_{t_0}^{t} \tau^\alpha \|\ddot{x}(\tau)\|^2 d\tau = 0.$

3B. Proof of the convergence results

Lemma (Opial)

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

(i) for every $z \in S$, $\lim_{t \to +\infty} \|x(t) - z\|$ exists;

(ii) every weak sequential cluster point of $x(\cdot)$ belongs to S.

Then, $w - \lim_{t\to+\infty} x(t) = x_\infty$ exists, for some element $x_\infty \in S$.

Lemma (differential inequality)

Let $t_0 > 0$, $\alpha > 1$, and $w : [t_0, +\infty[\to \mathbb{R}$ that satisfies

$$\dot{w}(t) + \frac{\alpha}{t} w(t) \leq g(t),$$

for some $g : [t_0, +\infty[\to \mathbb{R}^+$ such that $t \mapsto tg(t) \in L^1(t_0, +\infty)$. Then

$$w^+ \in L^1(t_0, +\infty).$$

3C. Proof of the convergence results

- *Step 1*. Given $x^* \in \operatorname{argmin}\Phi$, set $h(t) := \frac{1}{2}\|x(t) - x^*\|^2$.

$$\dot{h}(t) = \langle x(t) - x^*, \dot{x}(t)\rangle,$$
$$\ddot{h}(t) = \langle x(t) - x^*, \ddot{x}(t)\rangle + \|\dot{x}(t)\|^2.$$

Combining these two equations, and using $(\mathrm{AVD})_\alpha$, we obtain

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) = \|\dot{x}(t)\|^2 + \langle x(t) - x^*, \ddot{x}(t) + \frac{\alpha}{t}\dot{x}(t)\rangle, \qquad (4)$$
$$= \|\dot{x}(t)\|^2 + \langle x(t) - x^*, -\nabla\Phi(x(t))\rangle. \qquad (5)$$

By monotonicity of $\nabla\Phi$ and $\nabla\Phi(x^*) = 0$, we infer

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) \leq \|\dot{x}(t)\|^2. \qquad (6)$$

The next step is to prove that $\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty$.

3D. Proof of the convergence result

- *Step 2*. From

$$\dot{\mathcal{E}}_\alpha(t) + 2\frac{\alpha-3}{\alpha-1}t(\Phi(x(t)) - \min_{\mathcal{H}} \Phi) \leq 0,$$

and $\alpha > 3$, we deduce that

$$\int_{t_0}^{+\infty} t\left(\Phi(x(t)) - \min_{\mathcal{H}} \Phi\right) dt < +\infty. \tag{7}$$

Then, take the scalar product of $(\mathrm{AVD})_\alpha$ by $t^2\dot{x}(t)$, and integrate

$$\frac{1}{2}t^2\frac{d}{dt}\|\dot{x}(t)\|^2 + \alpha t\|\dot{x}(t)\|^2 + t^2\frac{d}{dt}\Phi(x(t)) \leq 0.$$

$$\tfrac{t^2}{2}\|\dot{x}(t)\|^2 + (\alpha-1)\int_{t_0}^{t}\tau\|\dot{x}(\tau)\|^2 d\tau \leq C + 2\int_{t_0}^{t}\tau(\Phi(x(\tau)) - \min_{\mathcal{H}}\Phi)d\tau.$$

$$\int_{t_0}^{+\infty} t\|\dot{x}(t)\|^2 dt < +\infty. \tag{8}$$

3E. Proof of the convergence result

- *Step 3.* Asymptotic behaviour of $\dot{x}(t)$. From

$$\frac{t^2}{2}\|\dot{x}(t)\|^2+(\alpha-1)\int_{t_0}^{t}\tau\|\dot{x}(\tau)\|^2 d\tau \leq C+2\int_{t_0}^{t}\tau(\Phi(x(\tau))-\min_{\mathcal{H}}\Phi)d\tau$$

we deduce that $\|\dot{x}(t)\| \leq \frac{C}{t}$, $\quad \lim_{t\to+\infty}\|\dot{x}(t)\| = 0$.

- *Step 4.* Let us show that $\lim_{t\to+\infty}\frac{1}{t^\alpha}\int_{t_0}^{t}\tau^\alpha\|\ddot{x}(\tau)\|^2 d\tau = 0$.
Let us return to (4)

$$\ddot{h}(t)+\frac{\alpha}{t}\dot{h}(t)+\langle x(t)-x^*,\nabla\Phi(x(t))\rangle = \|\dot{x}(t)\|^2.$$

$\nabla\Phi$ is Lipschitz continuous on bounded sets. By Baillon-Haddad theorem, it is $\frac{1}{L}$-cocoercive on a ball containing the trajectory:

$$\langle x(t)-x^*,\nabla\Phi(x(t))-\nabla\Phi(x^*)\rangle \geq \frac{1}{L}\|\nabla\Phi(x(t))-\nabla\Phi(x^*)\|^2.$$

3F. Proof of the convergence result

Combining the two above equations, and using $\nabla\Phi(x^*) = 0$, we obtain

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \frac{1}{L}\|\nabla\Phi(x(t))\|^2 \leq \|\dot{x}(t)\|^2.$$

Replacing $\nabla\Phi(x(t)) = -\ddot{x}(t) - \frac{\alpha}{t}\dot{x}(t)$

$$\ddot{h}(t) + \tfrac{\alpha}{t}\dot{h}(t) + \tfrac{1}{L}\|\ddot{x}(t) + \tfrac{\alpha}{t}\dot{x}(t)\|^2 \leq \|\dot{x}(t)\|^2.$$

Developing

$$\ddot{h}(t) + \frac{\alpha}{t}\dot{h}(t) + \frac{1}{L}\|\ddot{x}(t)\|^2 + \frac{\alpha}{Lt}\frac{d}{dt}\|\dot{x}(t)\|^2 \leq \|\dot{x}(t)\|^2.$$

Then integrate, and apply Fubini's theorem. □

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

Theorem 5 (APR)

Suppose $\alpha > 3$, and one of the following properties is satisfied by Φ:

- $\mathrm{int}(\mathrm{argmin}\ \Phi) \neq \emptyset$;
- Φ *is an even function (i.e.,* $\Phi(-x) = \Phi(x)$*);*
- Φ *is uniformly convex.*

Then, for any orbit $x(\cdot)$ *of* $(\mathrm{AVD})_\alpha$*, there exists* $x^* \in \mathrm{argmin}\Phi$ *such that*

$$x(t) \to x^* \in \mathrm{argmin}\Phi \quad \textit{strongly in } \mathcal{H} \textit{ as } t \to +\infty.$$

5A. Fast associated algorithms. Link to Nesterov.

Non-smooth structured convex minimization problem:

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

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

Optimal solutions:

$$\partial\Phi(x)+\nabla\Psi(x)\ni 0.$$

Dynamical approach via the differential inclusion

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

5B. Fast associated algorithms. Link to Nesterov.

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

$\Theta := \Phi + \Psi$, $\Theta : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ closed convex, proper.

$$\ddot{x}(t) + a(t)\dot{x}(t) + \partial\Theta(x(t)) \ni 0.$$

- $a(t) \equiv \gamma > 0$, $\dim \mathcal{H} < +\infty$, Schatzman, A-Cabot-Redont. $x(\cdot)$ loc. Lipschitz; $\dot{x}(\cdot)$ bounded variation; $\ddot{x}(\cdot)$ bounded measure. Nonuniqueness (shocks).
- $a(t) = \frac{\alpha}{t}$, $\alpha \geq 3$. Lyapunov analysis is still valid: convex subdifferential inequalites, generalized derivation chain rule.

5C. Fast associated algorithms. Link to Nesterov.

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

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

Time step $h > 0$, $t_k = kh$, $x_k = x(t_k)$. Finite difference scheme

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

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

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

5D. Fast associated algorithms. Link to Nesterov.

Proximal mapping, resolvent

$$\mathrm{prox}_{\gamma\Phi}(x) := \mathrm{argmin}_{\xi\in\mathcal{H}}\left\{\Phi(\xi) + \frac{1}{2\gamma}\|\xi - x\|^2\right\} = \left(I + \gamma\partial\Phi\right)^{-1}(x).$$

$$\begin{cases} y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}); \\ x_{k+1} = \mathrm{prox}_{h^2\Phi}\left(y_k - h^2\nabla\Psi(y_k)\right). \end{cases} \tag{9}$$

Equivalent formulation ($1 - \frac{\alpha}{k+\alpha-1} = \frac{k-1}{k+\alpha-1}$):

$$(\mathrm{AVD} - \mathrm{algo})_\alpha \quad \begin{cases} y_k = x_k + \frac{k-1}{k+\alpha-1}(x_k - x_{k-1}); \\ x_{k+1} = \mathrm{prox}_{h^2\Phi}\left(y_k - h^2\nabla\Psi(y_k)\right). \end{cases} \tag{10}$$

5E. Fast associated algorithms. Link to Nesterov.

Set $s = h^2$.

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

- Proximal inertial algo.: A-Alvarez, Moudafi-Oliny, Lorenz-Pock.
- $\alpha = 3$: Nesterov, Güler, Beck-Teboulle (FISTA)

$$(\mathrm{FISTA}) \quad \begin{cases} y_k = x_k + \frac{k-1}{k+2}(x_k - x_{k-1}); \\ x_{k+1} = \mathrm{prox}_{s\Phi}\left(y_k - s\nabla\Psi(y_k)\right). \end{cases}$$

- $\alpha \geq 3$. Recent studies Chambolle-Dossal, Su-Boyd-Candès, APR.

5F. Fast associated algorithms. Link to Nesterov.

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

Theorem 6 (Chambolle-Dossal, Su-Boyd-Candès)

- $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ closed convex proper;
- $\Psi : \mathcal{H} \to \mathbb{R}$ convex differentiable, $\nabla\Psi$ L-Lipschitz continuous;
- $S = \mathrm{argmin}(\Phi + \Psi) \neq \emptyset$, $s < \frac{1}{L}$, $\alpha > 3$.

Let (x_k) be a sequence generated by $(\mathrm{AVD}-\mathrm{algo})_\alpha$. Then,

- $x_k \rightharpoonup x^* \in \mathrm{argmin}(\Phi + \Psi)$ weakly as $k \to +\infty$.
- $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) \leq \dfrac{C}{k^2}$.
- $\sum_k k\|x_k - x_{k-1}\|^2 < +\infty$, $\quad \|x_k - x_{k-1}\| \leq \frac{C}{k}$.

5G. Fast associated algorithms. Proof of convergence

Step one. $k \mapsto \mathcal{E}_k$ is the correspondent of $\mathcal{E}_\alpha(\cdot)$:

$$\mathcal{E}_k = \frac{2s}{\alpha-1}(k+\alpha-2)^2(\Theta(x_k)-\Theta^*)+\frac{1}{\alpha-1}\|(k+\alpha-1)y_k-kx_k-(\alpha-1)x^*\|^2$$

$\mathcal{E}_k$ is a strict Lyapunov function: for any $k \in \mathbb{N}$

$$\mathcal{E}_k + \tfrac{2s}{\alpha-1}\Big((\alpha-3)(k+\alpha-2)+1\Big)(\Theta(x_k) - \inf\Theta) \leq \mathcal{E}_{k-1}.$$

Fast convergence properties

$$(\Phi+\Psi)(x_k) - \min_{\mathcal{H}}(\Phi+\Psi) \leq \frac{C}{k^2}$$

$$\sum_k k\left((\Phi+\Psi)(x_k) - \min_{\mathcal{H}}(\Phi+\Psi)\right) < +\infty.$$

Step two. Next step consists in obtening the energy estimate

$$\sum_k k\|x_k - x_{k-1}\|^2 < +\infty,$$

$=$ discrete version of the continuous energy estimate

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

Step three. The final step is to apply Opial's lemma. Using the previous estimates, it is a direct adaptation of the classical proof of the convergence of proximal-like inertial algorithms. It is a parallel argument to that using the differential inequality with $\|x_k - x^*\|^2$ instead of $\|x(t) - x^*\|^2$, and $x^* \in \operatorname{argmin}(\Phi + \Psi)$. $\square$

6. Related systems. Case $a(t) = \frac{1}{t^\gamma}$

$$\ddot{x}(t) + \frac{1}{t^\gamma}\dot{x}(t) + \nabla\Phi(x(t)) = 0.$$

Global energy : $\quad W(t) = \frac{1}{2}\|\dot{x}(t)\|^2 + \Phi(x(t)) - \min\Phi.$

$a(t) = \frac{1}{t^\gamma}$	$\gamma = 0$	$0 < \gamma < 1$	$\gamma = 1,\ a(t) = \frac{\alpha}{t},\ \alpha > 3$
$W(t) \to 0$	$\mathcal{O}(\frac{1}{t})$	$o\left(\frac{1}{t^{1+\bar{\gamma}}}\right),\ \forall \bar{\gamma} < \gamma$	$\mathcal{O}(\frac{1}{t^2})$

- $a(t) = 1,\ \gamma = 0$ (HBF), Alvarez (2000).
- $a(t) = \frac{1}{t^\gamma},\ \ 0 < \gamma < 1$, Cabot-Frankel (2012), R. May (2015).
- $a(t) = \frac{\alpha}{t},\ \alpha \geq 3$, SBC (2014), APR (2015).

7A. Related dynamics. Hessian driven damping

- $\Phi : \mathcal{H} \to \mathbb{R}$ convex, $\mathcal{C}^2$, argmin$\Phi \neq \emptyset$, $\alpha > 0$, $\beta > 0$.

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

Theorem 7 (APR)

- $\alpha > 0$: $\lim_{t\to+\infty} \Phi(x(t)) = \min \Phi$, $\qquad \lim_{t\to+\infty} \|\dot{x}(t)\| = 0.$
- $\alpha \geq 3$: *i)* $\Phi(x(t)) - \min_{\mathcal{H}} \Phi \leq \dfrac{C}{t^2}$;

  ii) $\displaystyle\int_0^\infty t^2 \|\nabla\Phi(x(t))\|^2 dt < +\infty$;

  iii) $\lim_{t\to+\infty} \|\dot{x}(t)\| = \lim_{t\to+\infty} \|\ddot{x}(t)\| = \lim_{t\to+\infty} \|\nabla\Phi(x(t))\| = 0.$
- $\alpha > 3$: $x(t)$ converges weakly to a minimizer of Φ.

7B. (DIN-AVD) with two potentials

$$\min\left\{\phi(x)+\Psi(x):\ x\in\mathcal{H}\right\},\ \Psi \text{ smooth, } \phi \text{ nonsmooth.}$$

$$(\mathrm{DIN-AVD})\ \begin{cases} \dot{x}(t)+\beta\partial\phi(x(t))-(\frac{1}{\beta}-\frac{\alpha}{t})x(t)-y(t)\ni 0; \\ \dot{y}(t)+\nabla\Psi(x(t))+\frac{1}{\beta}(\frac{1}{\beta}-\frac{\alpha}{t}+\frac{\alpha\beta}{t^2})x(t)+\frac{1}{\beta}y(t)=0. \end{cases}$$

Equivalent equation, $\phi=\Phi$ smooth:

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

Theorem 8 (APR)

Let $(x(\cdot),y(\cdot))$ be an orbit of $(\mathrm{DIN-AVD})$, $\alpha>0$. Then

$$\lim_{t\to+\infty}\Theta(x(t))=\min\Theta,\qquad \lim_{t\to+\infty}\|\dot{x}(t)\|=0.$$

7C. (DIN-AVD) algorithm with two potentials

Time step $h > 0$, $t_k = kh$, $x_k = x(t_k)$, $y_k = y(t_k)$, $(x_0, y_0) \in \mathcal{H} \times \mathcal{H}$:

$$\begin{cases} 0 & \in & \dfrac{x_{k+1} - x_k}{h} + \beta \partial \phi(x_{k+1}) - (\dfrac{1}{\beta} - \dfrac{\alpha}{kh})x_k - y_k \\ 0 & = & \dfrac{y_{k+1} - y_k}{h} + \nabla \Psi(x_{k+1}) + \dfrac{1}{\beta}(\dfrac{1}{\beta} - \dfrac{\alpha}{kh} + \dfrac{\alpha\beta}{k^2h^2})x_{k+1} + \dfrac{1}{\beta} y_{k+1} \end{cases}$$

$$\begin{cases} x_{k+1} & = & \operatorname{prox}_{\beta h \phi}\left(\left(1 + h(\dfrac{1}{\beta} - \dfrac{\alpha}{kh})\right) x_k + h y_k\right) \\ y_{k+1} & = & \dfrac{\beta}{\beta + h} y_k - \dfrac{h}{\beta + h}(\dfrac{1}{\beta} - \dfrac{\alpha}{kh} + \dfrac{\alpha\beta}{k^2h^2})x_{k+1} - \dfrac{h\beta}{\beta + h} \nabla \Psi(x_{k+1}). \end{cases}$$

First-order dynamic $/(x_k, y_k)$: $(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$.

8. Related dynamics. Adaptive restart (SBC)

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 $T(\Phi, x_0)$, $t \mapsto \Phi(x(t))$ 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 9 (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

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

Annex 1. Stochastic gradient descent algorithm (CEG)

$$\forall x \in \mathbb{R}^d \quad \Phi(x) := \int_\Omega \phi(x,\omega)\mu(d\omega).$$

$(\omega_n)_{n\geq 1}$ is a sequence of independent identically distributed variables

$$X_{n+1} = X_n - \epsilon_n \nabla\phi(X_n, \omega_{n+1}).$$

The stochastic approximation can be numerically improved:

$$X_{n+1} = X_n - \epsilon_{n+1} \frac{\sum_i^n \epsilon_i \nabla\phi(X_i, \omega_{i+1})}{\sum_i^n \epsilon_i}$$

Limit ODE ($n \to +\infty$, $\sum \epsilon_n = +\infty$, $\sum {\epsilon_n}^p < +\infty$ for some $p > 1$)

$$s\ddot{X}(s) + \dot{X}(s) + \nabla\Phi(X(s)) = 0.$$

Time rescaling $t = 2\sqrt{s}$ gives $\ddot{X}(t) + \frac{1}{t}\dot{X}(t) + \nabla\Phi(X(t)) = 0.$

Annex 2. Complexity aspects

In 1983, Nemirovsky & Yudin proved lower bounds on the complexity of first-order methods (number of subgradient calls needed to achieve a given accuracy) for convex optimization under various regularity assumptions for the objective functions. See also Nesterov (2004).

1. They constructed convex, piecewise linear functions in dimensions $n > k$, where no first-order method can have function values more accurate than $\mathcal{O}(\frac{1}{\sqrt{k}})$ after k subgradient evaluations.
2. They also constructed convex quadratic functions in dimensions $n \geq 2k$ where no first-order method can have function values more accurate than $\mathcal{O}(\frac{1}{k^2})$ after k gradient evaluations.
3. For strongly convex functions with Lipschitz continuous gradients, the known lower bounds on the complexity allow a dimension independent linear rate of convergence $\mathcal{O}(q^k)$ with $0 < q < 1$.

10. Perspective, open questions

- Convergence of the orbits for $\alpha = 3$? of Nesterov algorithm?
- Convergence of the values: exhibit concrete examples showing that $\alpha = 3$ is critical. Rate of convergence for $a(t) = \frac{\alpha}{t}$, $1 \leq \alpha < 3$?
- Find a Lyapunov function in the case $\frac{1}{t^\gamma}$, giving the rate of convergence $\frac{1}{t^{1+\gamma}}$ for $W(t)$.
- Extend to the algorithmic part the convergence properties of the continuous dynamic (strong convergence...).
- Adaptive restart for (DIN-AVD), without strong convexity.
- Compare /combine with other rapid methods: multigrid, Newton based methods, other type of friction (dry).
- Show the $\mathcal{O}(\frac{1}{t^2})$ convergence of the values for (DIN-AVD), combining Hessian driven and asymptotic vanishing damping.
- Extension to a non-convex setting: for analytic potentials, the convergence theory for HBF (HJ), and DIN (AABR) still works.
- Nonsmooth potentials, shock theory, PDE's hyperbolic equations.

$\mathcal{HAPPY\ BIRTHDAY}$

$\mathcal{TERRY}$

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, 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.

H. Attouch, X. Goudou and P. Redont, *The heavy ball with friction method. The continuous dynamical system, global exploration of the local minima of a real-valued function*, 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), № 1, pp. 27–65.

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

H. Attouch and A. Soubeyran, *Inertia and reactivity in decision making as cognitive variational inequalities*, Journal of Convex Analysis, **13** (2006), pp. 207–224.

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*, Electronic Journal 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 Fista*, HAL Id: hal-01060130 https://hal.inria.fr/hal-01060130v3 Submitted on 20 Oct 2014.

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).

- 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: A basic course*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 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.

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).