Relaxed Inertial Proximal Algorithms for Monotone Inclusions.

Hedy ATTOUCH

Université Montpellier
Institut Montpelliérain Alexander Grothendieck, UMR CNRS 5149

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

DIMACS workshop on ADMM and Proximal Splitting Methods in Optimization

Rutgers University, New Jersey
June 11-13, 2018
Introduction: Relaxed Inertial Proximal Algorithm

\( \mathcal{H} \) Hilbert, \( A : \mathcal{H} \to 2^\mathcal{H} \) maximally monotone. Fast solving \( Ax \ni 0 \).

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

Damped inertial dynamics \( \mapsto \) Accelerated methods.

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

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

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

\begin{align*}
\text{(RIPA)} & \quad \begin{cases}
  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{cases} \\
\end{align*}

A.-Peypouquet (Math Prog. 2018), A.-Cabot (HAL 2018)
**Introduction: Relaxed Inertial Proximal Algorithm**

\( \mathcal{H} \) Hilbert, \( A : \mathcal{H} \to 2^\mathcal{H} \) maximally monotone, \( S = A^{-1}(0) \neq \emptyset \).

(RIPA) \[
\begin{align*}
  y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\
  x_{k+1} &= (1 - \rho_k)y_k + \rho_k J_{\mu_k}A(y_k).
\end{align*}
\]

Find general conditions on parameters \( \alpha_k, \rho_k, \mu_k \):

- A general maximally monotone operator.
  - \( (x_k) \) converges weakly to \( \hat{x} \in S \).
  - Fast convergence of the discrete velocities \( \|x_k - x_{k-1}\| \to 0 \).

- \( A = \partial \Phi, \) \( \Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\} \) convex, lower semicontinuous, proper.
  - Fast convergence of the values in the worst case (Nesterov type).
  - Improved convergence rate under geometrical assumptions.

Then combine to obtain both type of results. Apply to FB, ADMM...
1 Inertial approach to accelerated gradient methods: survey.

2 (RIPA) algorithm for monotone inclusions: model situation.

3 (RIPA) algorithm for monotone inclusions: general case.

4 An inertial ADMM algorithm.

5 Perspectives.
1a. Inertial approach to accelerated gradient methods

\[ \Phi : \mathcal{H} \to \mathbb{R} \text{ convex differentiable, } \nabla \Phi \text{ } L\text{-Lipschitz, } \text{argmin} \Phi \neq \emptyset. \]

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

\[
\begin{align*}
\text{(IG)} \quad \begin{cases} 
    y_k &= x_k + \alpha_k (x_k - x_{k-1}) \\
    x_{k+1} &= y_k - s \nabla \Phi(y_k)
\end{cases}
\]

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

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

\[ \min \{ \Phi(x) : x \in \mathcal{H} \}, \Phi : \mathcal{H} \to \mathbb{R} \text{ convex differentiable}, \ S = \arg\min\Phi \neq \emptyset. \]

\[
\begin{align*}
(IG) \quad \left\{ \begin{array}{l}
y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\
x_{k+1} &= y_k - s \nabla \Phi(y_k)
\end{array} \right.
\end{align*}
\]
1c. Inertial approach to accelerated gradient methods

- Su-Boyd-Candès (NIPS 2014), $A = \nabla \Phi$, $\Phi$ convex, $\alpha > 0$.

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

Discretization: explicit /smooth $\Phi$.

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

\[\uparrow\]

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

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

\[(IG)_\alpha \quad \begin{cases} 
  y_k = x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\
  x_{k+1} = y_k - s \nabla \Phi(y_k).
\end{cases}\]
Inertial approach to accelerated gradient methods

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

**Inertial Gradient algorithm, \( \alpha > 0, s \leq \frac{1}{L} \)**

\[
\begin{align*}
(IG)_\alpha & \quad \left\{ \begin{array}{l}
y_k & = x_k + (1 - \frac{\alpha}{k}) (x_k - x_{k-1}) \\
x_{k+1} & = y_k - s \nabla \Phi(y_k)
\end{array} \right. \\
\end{align*}
\]

- **\( \alpha = 3 \):** \( \Phi(x_k) - \min_\mathcal{H} \Phi = \mathcal{O} \left( \frac{1}{k^2} \right) \), Nesterov (1983).

- **\( \alpha > 3 \):** \( x_k \rightharpoonup \hat{x} \in S \), Chambolle-Dossal (JOTA 2015).

\[
\Phi(x_k) - \min_\mathcal{H} \Phi = o \left( \frac{1}{k^2} \right), \text{ A.-Peypouquet (SIOPT 2016).}
\]

- **\( \alpha \leq 3 \):** \( \Phi(x_k) - \min_\mathcal{H} \Phi = \mathcal{O} \left( \frac{1}{k^{2\alpha/3}} \right) \), A.-Chbani-Riahi (COCV 2018)

Apidopoulos-Aujol-Dossal (HAL, 2017).
1e. Inertial approach to accelerated gradient methods

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

Rate of convergence of the values:

\[
\Phi(x_k) - \min_\mathcal{H} \Phi = \mathcal{O}\left(\frac{1}{kp(\alpha)}\right), \quad p(\alpha) = \min\left(\frac{2\alpha}{3}, 2\right).
\]
1f. Inertial approach to accelerated gradient methods

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

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

- $\alpha = 3$: FISTA, Beck-Teboulle (SIAM J. Imaging 2009):
  - $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = O\left(\frac{1}{k^2}\right)$.
- $\alpha > 3$: $x_k \rightharpoonup x_\infty \in S = \text{argmin}(\Phi + \Psi)$, 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)
  - $(\Phi + \Psi)(x_k) - \min_{\mathcal{H}}(\Phi + \Psi) = O\left(\frac{1}{k^{\frac{2\alpha}{3}}}\right)$. 
1g. Inertial approach to accelerated gradient methods

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

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

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

**A.** Suppose that the sequence \((\alpha_k)\) 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 \((\text{IPG})_{\alpha(\cdot)}\)

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

**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

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

If \(\alpha_k \in [0, 1]\) for every \(k \geq 1\), then \(x_k \rightharpoonup \hat{x} \in \text{argmin}(\Phi + \Psi)\).
1h. Inertial approach to accelerated gradient methods

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

<table>
<thead>
<tr>
<th>( \alpha_k )</th>
<th>( 1 - \frac{\alpha}{k} )</th>
<th>( 1 - \frac{\alpha}{k} )</th>
<th>( 1 - \frac{(\ln k)^{\theta}}{k} )</th>
<th>( 1 - \frac{\alpha}{kr} )</th>
<th>( \alpha_k \equiv \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha \leq 3 )</td>
<td>( \alpha &gt; 3 )</td>
<td>( \theta &gt; 0 )</td>
<td>( r \in ]0, 1[ )</td>
<td>( 0 &lt; \alpha &lt; 1 )</td>
<td></td>
</tr>
</tbody>
</table>

- 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 \( \alpha > 3 \).
2a. (RIPA) for monotone inclusions: model situation.

\[ A : \mathcal{H} \to \mathcal{H} \text{ } \lambda\text{-cocoercive } (\lambda > 0) \]

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

\[ A \text{ } \lambda\text{-cocoercive } \implies A \text{ maximally monotone, } \frac{1}{\lambda}\text{-Lipschitz continuous.} \]

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

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

Theorem (A.-Maingé, ESAIM-COCV 2011)

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

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

Sharp result: \( A = \text{rot}(0, \frac{\pi}{2}). \)
Harmonic oscillator

\( \mathbb{C} \) endowed with the standard real Hilbert structure \( \langle u, v \rangle = Re(\bar{u}v) \).

Consider the equation

\[
(HBF)_\gamma \quad \ddot{z}(t) + \gamma \dot{z}(t) + A(z(t)) = 0, \quad t \geq 0,
\]

where \( A : \mathbb{C} \to \mathbb{C} \)

\[
A z := (w^2 - i\gamma w)z \text{ with } \gamma > 0 \text{ and } w > 0.
\]

The operator \( A \) is \( \lambda \)-cocoercive with \( \lambda = \frac{1}{w^2 + \gamma^2} \).

A solution of \( (HBF)_\gamma \) is given by the harmonic oscillator \( z(t) = e^{iwt} \).

It appears that \( z(\cdot) \) is bounded but not convergent for any \( w > 0 \).

By letting \( w \to 0^+ \) we get \( \lambda \gamma^2 \to 1^- \implies \lambda \gamma^2 < 1 \) is not sufficient for the convergence of \( (HBF)_\gamma \) for a general \( \lambda \)-cocoercive operator.
2c. (RIPA) for monotone inclusions: model situation.

- \( A : \mathcal{H} \to 2^\mathcal{H} \) 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**

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

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

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

**Tuning of \( t \mapsto \lambda(t) > 0 \):** \( \lambda \gamma^2 > 1 \) \( \implies \) \( \lambda(t) > \frac{t^2}{\alpha^2} \).
2d. (RIPA) for monotone inclusions: model situation.

- \( A : \mathcal{H} \rightarrow 2^{\mathcal{H}} \) 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 \( A \).

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

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

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 \( S \), \( \|x(t) - p(t)\| \rightarrow 0. \)
2e. (RIPA) for monotone inclusions: model situation.

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

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

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

\[
\begin{align*}
\text{(RIPA)}_{\text{model}} \quad \left\{ \begin{array}{l}
y_k &= x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\
x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} (I + (\lambda_k + s)A)^{-1}(y_k).
\end{array} \right.
\end{align*}
\]
2f. (RIPA) for monotone inclusions: model situation.

\[(\text{RIPA})_{\text{model}} \quad \begin{cases} y_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\ x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} J(\lambda_k + s)A(y_k) \end{cases}\]

\[S = A^{-1}(0)\]

\[\lambda_k \to +\infty, \quad \frac{s}{\lambda_k + s} \to 0\]

\[J(\lambda_k + s)A(y_k) \sim \text{proj}_S(y_k)\]
2g. (RIPA) for monotone inclusions: model situation.

\[ \begin{aligned}
(RIPA)_{model} & \quad \begin{cases}
  y_k &= x_k + \left(1 - \frac{\alpha}{k}\right)(x_k - x_{k-1}) \\
  x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} J(\lambda_k + s)A(y_k).
\end{cases}
\end{aligned} \]

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

Let \( A : \mathcal{H} \to 2^\mathcal{H} \) 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 \( k \to +\infty \), to some \( \hat{x} \in S \).

ii) \( (y_k) \) converges weakly, as \( k \to +\infty \), to \( \hat{x} \).

iii) \( \|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right) \) as \( k \to +\infty \), and \( \sum_k k\|x_k - x_{k-1}\|^2 < +\infty \).
2h. (RIPA) for monotone inclusions: model situation.

\[ A = \partial \Phi, \ \Phi : \mathcal{H} \rightarrow \mathbb{R} \cup +\{\infty\} \text{ convex lsc., proper}, \ S = \arg\min \Phi \neq \emptyset. \]

\[(\text{RIPA})_{\text{model}} \left\{ \begin{array}{l}
y_k = x_k + \left(1 - \alpha \frac{1}{k}\right) (x_k - x_{k-1}) \\
x_{k+1} = \frac{\lambda_k}{\lambda_k + s} y_k + \frac{s}{\lambda_k + s} \text{prox} (\lambda_{k+s}) \Phi(y_k)
\end{array} \right. \]

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

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

Suppose that \((\lambda_k)\) 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 \(p_k = \text{prox} (\lambda_{k+s}) \Phi(x_k)\), we have

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

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

Then \(x_k \rightharpoonup \hat{x} \in S\), \(\Phi(p_k) - \min_{\mathcal{H}} \Phi = o(k^{-2})\), \(\lim_{k \to +\infty} \|p_k - x_k\| = 0\).
3a. (RIPA) for monotone inclusions: general case.

General parameters $\alpha_k, \rho_k, \mu_k$

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

3b. (RIPA) for monotone inclusions: general case.

(RIPA) \[
\begin{align*}
    y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\
    x_{k+1} &= (1 - \rho_k)y_k + \rho_k J_{\mu_k}A(y_k).
\end{align*}
\]

\((K_0)\) \[
\sum_{l=k}^{+\infty} \left( \prod_{j=k}^{l} \alpha_j \right) < +\infty \quad \text{for every } k \geq 1; \quad t_k := 1 + \sum_{l=k}^{+\infty} \left( \prod_{j=k}^{l} \alpha_j \right); \quad \alpha_k = \frac{t_k - 1}{t_{k+1}}.
\]

(L): there exists \(\epsilon \in ]0, 1[\) such that for \(k\) large enough, \[
(1 - \epsilon) \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \geq \alpha_k t_{k+1} \left( 1 + \alpha_k + \left[ \frac{2 - \rho_k}{\rho_k} (1 - \alpha_k) - \frac{2 - \rho_{k-1}}{\rho_{k-1}} (1 - \alpha_{k-1}) \right]_+ \right).
\]

Consequence: under (L), \(\alpha_k \to 1 \implies \rho_k \to 0\).
Convergence results in the case of \((\rho_k)\) bounded away from zero.

**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,

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

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

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

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

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

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

(iv) If \(\liminf_{k \to +\infty} \mu_k > 0\), then \(x_k \rightharpoonup x_\infty \in S\) weakly in \(\mathcal{H}\) as \(k \to +\infty\).
Corollary (Bertsekas-Eckstein)

Assume that $S \neq \emptyset$ and that $\rho_k \in ]0, 2]$ for every $k \geq 1$. Then, for any sequence $(x_k)$ generated by (RPA)

(RPA) \hspace{1cm} x_{k+1} = (1 - \rho_k)x_k + \rho_k J_{\mu_k} A(x_k),

we have

(i) $\sum_{i=1}^{+\infty} \frac{2 - \rho_{i-1}}{\rho_{i-1}} \|x_i - x_{i-1}\|^2 < +\infty$.

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

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

Then the following holds

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

(iv) If $\liminf_{k \to +\infty} \mu_k > 0$, then there exists $x_\infty \in S$ such that $x_k \rightharpoonup x_\infty$ weakly in $\mathcal{H}$ as $k \to +\infty$. 
Case $\rho_k = 1$ (without relaxation).

Corollary (Alvarez-A.)

Assume that $S \neq \emptyset$. Suppose that there exists $\bar{\alpha} \in [0, 1/3]$ such that $\alpha_k \in [0, \bar{\alpha}]$ for every $k \geq 1$. Then for any sequence $(x_k)$ generated by (IPA), we have

(i) $\sum_{i=1}^{+\infty} \|x_i - x_{i-1}\|^2 < +\infty$.

(ii) $\sum_{i=1}^{+\infty} \|\mu_i A_{\mu_i}(x_i)\|^2 < +\infty$.

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

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

(v) If $\lim \inf_{k \to +\infty} \mu_k > 0$, there exists $x_\infty \in S$ such that $x_k \rightharpoonup x_\infty$ weakly in $\mathcal{H}$ as $k \to +\infty$. 

3e. (RIPA) for monotone inclusions: general case.
Balance between inertia and relaxation

Suppose that \( \alpha_k \equiv \alpha \in [0, 1], \rho_k \equiv \rho \in ]0, 2[ \) for every \( k \geq 1 \), and that

\[
\frac{2-\rho}{\rho} \frac{(1 - \alpha)^2}{\alpha (1 + \alpha)} > 0.
\]

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

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

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

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

(iv) If \( \lim \inf_{k \to +\infty} \mu_k > 0 \), there exists \( x_\infty \in S \) such that \( x_k \rightharpoonup x_\infty \).

Inequation \( \iff \rho < \rho_m(\alpha) = \frac{2(1-\alpha)^2}{2\alpha^2-\alpha+1} \). \( \alpha \mapsto \rho_m(\alpha) \) decreasing on \([0, 1]\).

When the inertial effect increases \( (\alpha \nearrow) \), the relaxation effect decreases \( (\rho_m \searrow) \), and vice versa, see also Iutzeler-Hendrickx. When \( \alpha \to 0 \), the limiting value \( \rho_m(\alpha) \) is 2, in accordance with Eckstein-Bertsekas.

When \( \alpha \to 1 \), the limiting value of \( \rho_m(\alpha) \) is zero.
3g. (RIPA) for monotone inclusions: general case.

Convergence results in the case of a possibly vanishing sequence \((\rho_k)\).

**Theorem (A.-Cabot, HAL 2018)**

Let \(A : \mathcal{H} \to 2^{\mathcal{H}}\) be a maximally monotone operator such that \(S \neq \emptyset\). Suppose \(\alpha_k \in [0, 1], \rho_k \in ]0, 2], \mu_k > 0\). Suppose that \((K_0)\) and \((L)\).

Then for any sequence \((x_k)\) generated by (RIPA),

(i) There exists \(C \geq 0\) s.t. for every \(k \geq 1\), \(\|x_{k+1} - x_k\| \leq C \sum_{i=1}^{k} \left[ \left( \prod_{j=i+1}^{k} \alpha_j \right) \rho_i \right]\).

Assume moreover that \(\limsup_{k \to +\infty} \rho_k < 2\), together with

- \(\sum_{i=1}^{k} \left[ \left( \prod_{j=i+1}^{k} \alpha_j \right) \rho_i \right] = O(\rho_k t_{k+1}), \frac{\mu_{k+1} - \mu_k}{\mu_{k+1}} = O(\rho_k t_{k+1}), \rho_{k-1} t_k = O(\rho_k t_{k+1})\);
- \(\sum_{k=1}^{+\infty} \rho_k t_{k+1} = +\infty\).

Then the following holds

(ii) \(\lim_{k \to +\infty} \mu_k A_{\mu_k} (x_k) = 0\). If \(\liminf_{k \to +\infty} \mu_k > 0\), then there exists \(x_\infty \in S\) such that \(x_k \rightharpoonup x_\infty\) weakly in \(\mathcal{H}\) as \(k \to +\infty\).
3h. (RIPA) for monotone inclusions: general case.

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

\[(H) \left\{ \begin{array}{l}
\bullet (\alpha_k) \text{ is a nonnegative sequence;} \\
\bullet 0 < \rho_k \leq 1 \text{ for all } k \geq 1; \\
\bullet (\mu_k), (\rho_k\mu_k) \text{ nondecreasing sequences of positive numbers.} 
\end{array} \right.\]

\[(\text{RIPA}) \iff \left\{ \begin{array}{l}
y_k = x_k + \alpha_k(x_k - x_{k-1}) \\
x_{k+1} = y_k - \rho_k\mu_k \nabla \Phi_{\mu_k}(y_k).
\end{array} \right.\]

Moreau envelope: \[ \Phi_{\mu}(x) = \inf_{\xi \in \mathcal{H}} \left\{ \Phi(\xi) + \frac{1}{2\mu} \| x - \xi \|^2 \right\}. \]
\(\Phi_{\mu}\) convex differentiable, \(\nabla \Phi_{\mu}\) Lipschitz \(\to\) inertial gradient methods.
3i. (RIPA) for monotone inclusions: general case.

Theorem (A.-Cabot, HAL 2018), Case \( A = \partial \Phi \)

Under \((H)\), assume that the nonnegative sequence \((\alpha_k)\) satisfies \((K_0)-(K_1)\). Let \((x_k)\) be a sequence generated by the algorithm (RIPA). Then we have

(i) For every \( k \geq 1 \),

\[
\Phi_{\mu_k}(x_k) - \min \Phi \leq \frac{C}{t_k^2},
\]

with \( C = t_1^2(\Phi_{\mu_1}(x_1) - \min \Phi) + \frac{1}{\rho_1 \mu_1}(d(x_0, S)^2 + t_1^2\|x_1 - x_0\|^2) \).

As a consequence, setting \( p_k = \text{prox}_{\mu_k \Phi}(x_k) \), we have

\[
\Phi(p_k) - \min \Phi = O\left(\frac{1}{t_k^2}\right) \quad \text{and} \quad \|x_k - p_k\|^2 = O\left(\frac{\mu_k}{t_k^2}\right) \quad \text{as} \ k \to +\infty.
\]

(ii) Assume moreover \((K_1^+)\): \( \exists 0 \leq m < 1 \) s.t. \( t_{k+1}^2 - t_k^2 \leq m t_{k+1} \), and \( \alpha_k \in [0, 1] \).
Assume moreover \( \sup_k \frac{\mu_k}{\sum_{i=1}^k t_i} < +\infty \) and \( \sup_k \rho_k \mu_k < +\infty \). Then,

- \( \lim_{k \to +\infty} \|x_k - p_k\| = 0 \);
- \((x_k)\) and \( p_k \) converge weakly, as \( k \to +\infty \), to some \( x^* \in \arg\min \Phi \).
### 3j. (RIPA) for monotone inclusions: general case.

<table>
<thead>
<tr>
<th><strong>A</strong></th>
<th>maximally monotone</th>
<th>( A = \partial \Phi, \quad \Phi : \mathcal{H} \to \mathbb{R} \cup {+\infty} ) convex lsc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_k )</td>
<td>( \alpha_k = 1 - \frac{\alpha}{k} )</td>
<td>( \alpha &gt; 2 ) \quad \alpha \geq 3 \quad \alpha &gt; 3 )</td>
</tr>
<tr>
<td>( \rho_k )</td>
<td>( \rho_k = \frac{\beta}{k^2}, \quad \beta &lt; \alpha(\alpha - 2) )</td>
<td>( 0 &lt; \rho_k \leq 1 )</td>
</tr>
<tr>
<td>( \mu_k )</td>
<td>( \frac{</td>
<td>\mu_{k+1} - \mu_k</td>
</tr>
<tr>
<td>weak convergence of iterates</td>
<td>( |x_{k+1} - x_k| = O\left(\frac{1}{k}\right) )</td>
<td>( \Phi(p_k) - \min \Phi = O\left(\frac{1}{k^2}\right) ) \quad weak convergence of iterates</td>
</tr>
</tbody>
</table>

\[ (\text{RIPA}): y_k = x_k + \alpha_k(x_k - x_{k-1}), \quad x_{k+1} = (1 - \rho_k)y_k + \rho_k J_{\mu_k}A(y_k) \]
Corollary

Take the parameters $\alpha_k, \rho_k, \mu_k$ of (RIPA) as follows: for each $k \geq 1$

$\alpha_k = 1 - \frac{\alpha}{k}$, $\alpha \geq 3$; $\rho_k = \frac{\beta}{k^2}$, $\beta < \alpha(\alpha - 2)$; $\mu_k = c k^{r'}$, $r' \geq 2$, $c > 0$.

Let $(x_k)$ be generated by (RIPA). Then, we have both:

a) When $A$ is a general maximally monotone operator:
Weak convergence of $(x_k)$ to some $\hat{x} \in S$, and $\|x_{k+1} - x_k\| = \mathcal{O}\left(\frac{1}{k}\right)$.

b) When $A = \partial \Phi$, $\Phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ convex, lower semicontinuous:
Rate of convergence of the values $\mathcal{O}\left(\frac{1}{k^2}\right)$ Precisely, $\Phi(p_k) - \min \Phi = \mathcal{O}(\frac{1}{k^2})$ where $p_k = \text{prox}_{\mu_k \Phi}(x_k)$.

In the model situation, $\rho_k \mu_k$ is taken constant.
3l. (RIPA) for monotone inclusions: general case.

Relaxed Inertial Forward-Backward algorithm, A.-Cabot 2018

\[
\begin{align*}
(y_k &= x_k + \alpha_k(x_k - x_{k-1}) \\
 x_{k+1} &= (1 - \rho_k)y_k + \rho_k J_{\mu_k} A(y_k - \mu_k B(y_k))
\end{align*}
\]

We assume the following set of hypotheses

\[
\begin{cases}
A : \mathcal{H} \to 2^{\mathcal{H}} & \text{is a maximally monotone operator;}

B : \mathcal{H} \to \mathcal{H} & \text{is a } \lambda\text{-cocoercive operator for some } \lambda > 0;
\zer (A + B) := \{ x \in \mathcal{H} : Ax + Bx \ni 0 \} & \text{is nonempty;}

(\alpha_k) & \text{is a sequence of nonnegative numbers;}

(\mu_k) \text{ and } (\rho_k) & \text{are sequences of positive numbers.}
\end{cases}
\]
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.
- \(f: \mathcal{X} \to \mathbb{R} \cup \{+\infty\}, \ g: \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}\) closed convex proper.
- \(A: \mathcal{X} \to \mathcal{Z}, \ B: \mathcal{Y} \to \mathcal{Z}\) linear continuous operators.
- \(\lambda\) 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

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

\[M(x, y, z) = (\partial f(x) + A^t z, \ \partial g(y) - B^t z, \ By - Ax).\]
4b. An inertial proximal ADMM algorithm


\[
\left\{
\begin{aligned}
  u_k &= x_k + \left(1 - \frac{\alpha}{k}\right) (x_k - x_{k-1}) \\
  v_k &= y_k + \left(1 - \frac{\alpha}{k}\right) (y_k - y_{k-1}) \\
  w_k &= z_k + \left(1 - \frac{\alpha}{k}\right) (z_k - z_{k-1}) \\
  \frac{1}{\lambda_k + s} (p_k - u_k) + \partial f(p_k) + A^t (w_k + (\lambda_k + s)(Ap_k - Bv_k)) &\ni 0 \\
  \frac{1}{\lambda_k + s} (q_k - v_k) + \partial g(q_k) - B^t (w_k + (\lambda_k + s)(Ap_k - Bq_k)) &\ni 0 \\
  r_k &= w_k + (\lambda_k + s)(Ap_k - Bq_k) \\
  x_{k+1} &= \frac{\lambda_k}{\lambda_k + s} u_k + \frac{s}{\lambda_k + s} p_k \\
  y_{k+1} &= \frac{\lambda_k}{\lambda_k + s} v_k + \frac{s}{\lambda_k + s} q_k \\
  z_{k+1} &= \frac{\lambda_k}{\lambda_k + s} w_k + \frac{s}{\lambda_k + s} r_k.
\end{aligned}
\right.
\]
5. 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.

Annex 1a. Proofs

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

\begin{itemize}
    \item a) $\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \lambda(t) \|A_{\lambda(t)}(x(t))\|^2 \leq \|\dot{x}(t)\|^2$.
    \item b) $\ddot{h}_z(t) + \frac{\alpha}{t} \dot{h}_z(t) + \epsilon \|\dot{x}(t)\|^2 + \frac{\alpha \lambda(t)}{t} \frac{d}{dt} \|\dot{x}(t)\|^2 + \lambda(t) \|\ddot{x}(t)\|^2 \leq 0$
\end{itemize}

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

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

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

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

H. ATTOUCH (Univ. Montpellier) Inertial dynamics, 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 \((*)\) \(\lim_{t \to +\infty} \|\lambda(t)A_{\lambda(t)}(x(t))\| = 0.\)

Suppose \((*)\). Let \(x(t_n) \to \bar{x}\) weakly. From \(\lambda(t_n)A_{\lambda(t_n)}(x(t_n)) \to 0\) and \(\lambda(t_n) \to +\infty\), we 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., \(\bar{x} \in S.\)

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

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

\(w(t) := \|\lambda(t)A_{\lambda(t)}x(t)\|\) 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.
Annex 1c. Proofs

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

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

Spectral analysis

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

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

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

Suppose \( p > 2 \). The eigenvalues \( \theta_+ \) and \( \theta_- \) 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 \( \lambda(t) \).
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 $\lambda$, we set $\lambda = 10$.

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

(E4) is a first-order equation governed by the strongly monotone operator $A_{\lambda}$.
Annex 2. Perturbations, errors

\[ (IPG)_{pert} \left\{ \begin{array}{l}
y_k = x_k + \alpha_k(x_k - x_{k-1}) \\
x_{k+1} = \text{prox}_{s\psi}(y_k - s\nabla\Phi(y_k) + sg_k).
\end{array} \right. \]

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

**A.** Suppose that \((\alpha_k)\) satisfies \((K_0)\) and \((K_1)\), and that \((g_k)\) satisfies \((K_2)\)

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

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

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

**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 \(\bar{x} \in \text{argmin}(\Phi + \Psi)\).

**C.** If we assume additionally that \((K_2^+)\)

\[ (K_2^+) \quad \sum_{k=1}^{+\infty} \left( \frac{1}{t_{k+1}} \sum_{i=1}^{k} t_{i+1} \right) \|g_k\| < +\infty, \]

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 3. Perturbations, Tikhonov regularization

\[(\text{IFB})_{Tikh}\]

\[
\begin{align*}
y_k &= x_k + \alpha_k (x_k - x_{k-1}); \\
x_{k+1} &= \text{prox}_{s\Psi} (y_k - s\nabla\Phi(y_k) - s\epsilon_k y_k),
\end{align*}
\]

**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 \((\epsilon_k)\) is nonincreasing, and verifies \(\sum_k \frac{\epsilon_k}{t_{k+1}} = +\infty\).

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

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

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

Let $\Phi : \mathbb{R} \to \mathbb{R}$ be a convex continuously differentiable function such that $S = \text{argmin} \Phi \neq \emptyset$. 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 $\text{argmin} \Phi = \{x^*\}$, $x(\cdot)$ converges to its unique cluster point $x^*$. When $\text{argmin} \Phi = [a, b]$, 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.
Lemma 1 (H general Hilbert space)

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

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

Then

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

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

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

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

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

$$E^p_{\lambda,\xi}(t_2) \leq E^p_{\lambda,\xi}(t_1),$$

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 4c. $\alpha = 3$. 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.

\[\begin{align*}
\Phi &\quad \Phi \\
0 &\quad 0 \\
a &\quad u_n \quad v_n \\
n &\quad n \\
b &\quad b
\end{align*}\]

- $s_n \leq t_n \leq u_n \leq v_n$
- $x(s_n) = a$, $x(t_n) = b$, $a \leq x(t) \leq b$ 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 $\dot{x}$

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

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.
(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.

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

Theorem (A-Peypouquet-Redont, JDE 2016)

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

\begin{align*}
\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{align*}

- 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.

**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 \( \Phi: \mathcal{H} \to \mathbb{R} \) 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 \( \Phi \)?
References


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.


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.


