A proximal-Newton method for monotone inclusions in Hilbert spaces with complexity $\mathcal{O}(1/k^2)$.

Hedy ATTOUCH

Université Montpellier 2
ACSIOM, I3M UMR CNRS 5149

joint work with M. Marques Alves, and B.F. Svaiter

Convegno Italo-Francese, Atelier Italo-Français
Optimisation et processus dynamiques en apprentissage et problèmes inverses

Fondazione Mediaterraneo, Sestri Levante (GE), Italia

Sept. 8-12, 2014

- $\mathcal{H}$ real Hilbert space; $\|x\|^2 = \langle x, x\rangle$;
- $A : \mathcal{H} \rightrightarrows \mathcal{H}$ maximal monotone operator.

Fast methods for solving:

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax. \tag{1}$$

- $(I + \lambda A)^{-1} : \mathcal{H} \to \mathcal{H}$ resolvent of index $\lambda > 0$ of A.
- Fixed point formulation of (1): $x - (I + \lambda A)^{-1} x = 0$.

Dynamical system: control variable: $t \mapsto \lambda(t)$.

$$\dot{x}(t) + x(t) - (I + \lambda(t)A)^{-1} x(t) = 0. \tag{2}$$

1B. General presentation: dynamical approach

Differential-algebraic system:

$$
\text{(LSP)} \quad \begin{cases} \dot{x}(t) + x(t) - (I + \lambda(t)A)^{-1}x(t) = 0, \\ \lambda(t)\|(I + \lambda(t)A)^{-1}x(t) - x(t)\| = \theta. \end{cases} \tag{3}
$$

- (LSP): Large Step Proximal method.
- $(x(\cdot), \lambda(\cdot))$ variables, θ positive parameter.
- Closed-loop control: λ is taken inversely proportional to the speed.
- Asymptotic equilibration ($t \to +\infty$):

$$
\dot{x}(t) \to 0 \Rightarrow \|(I + \lambda(t)A)^{-1}x(t) - x(t)\| \to 0 \Rightarrow \lambda(t) \to +\infty.
$$

1C. General presentation: dynamical approach

Cauchy problem, main results:

$$
\text{(LSP)} \quad \begin{cases} \dot{x}(t) + x(t) - (I + \lambda(t)A)^{-1}x(t) = 0, \\ \lambda(t)\|(I + \lambda(t)A)^{-1}x(t) - x(t)\| = \theta, \\ x(0) = x_0 \in \mathcal{H}. \end{cases}
$$

1. Existence and uniqueness of (x, λ) global solution of (LSP).
2. $A^{-1}(0) \neq \emptyset$: $\lambda(t) \uparrow +\infty$, $w - \lim_{t\to+\infty} x(t) = x_\infty \in A^{-1}(0)$.
3. $A = \partial f$, $f : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ convex, lower semicontinuous, proper

$$
f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(\frac{1}{t^2}).
$$

1D. General presentation: Large step proximal method

A large-step proximal method for convex optimization:

(0) $x_0 \in \mathrm{dom}(f)$, $\sigma \in [0,1[$, $\theta > 0$ given, set $k = 1$;

(1) choose $\lambda_k > 0$, and find $x_k, v_k \in \mathcal{H}$, $\varepsilon_k \geq 0$ such that

$$v_k \in \partial_{\varepsilon_k} f(x_k), \tag{4}$$

$$\|\lambda_k v_k + x_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|x_k - x_{k-1}\|^2, \tag{5}$$

$$\lambda_k \|x_k - x_{k-1}\| \geq \theta \text{ or } v_k = 0; \tag{6}$$

(2) if $v_k = 0$ then STOP and output x_k; otherwise let $k \leftarrow k+1$ and go to step 1.

end.

Main result: $f(x_k) - \inf_{\mathcal{H}} f \leq \frac{C}{k^2}$.

Contents

1. General presentation.
2. Algebraic relationship linking λ and x.
3. Existence and uniqueness for the Cauchy problem.
4. Asymptotic behavior.
5. Link with the regularized Newton system.
6. The convex subdifferential case.
7. A large-step proximal method for convex optimization.
8. $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ proximal Newton method for convex optimization.
9. Perspective, open questions.
10. Appendix. Some examples

2A. Algebraic relationship linking λ and x

$$\lambda \|(I + \lambda A)^{-1}x - x\| = \theta. \tag{7}$$

$$\text{Set} \quad \varphi: [0, \infty[\times \mathcal{H} \to \mathbb{R}^+, \quad \varphi(\lambda, x) := \lambda \|x - (I + \lambda A)^{-1}x\|. \tag{8}$$

Some classical results on resolvents, $\lambda > 0$, $\mu > 0$, $x \in \mathcal{H}$

(1) $J_\lambda^A = (I + \lambda A)^{-1} : \mathcal{H} \to \mathcal{H}$ nonexpansive,

(2) $J_\lambda^A x = J_\mu^A \left(\frac{\mu}{\lambda} x + \left(1 - \frac{\mu}{\lambda}\right) J_\lambda^A x \right)$;

(3) $\|J_\lambda^A x - J_\mu^A x\| \leq |\lambda - \mu| \; \|A_\lambda x\|$;

(4) $\lim_{\lambda \to 0} J_\lambda^A x = \mathrm{proj}_{\overline{\mathrm{dom}(A)}} x$;

(5) $\lim_{\lambda \to +\infty} J_\lambda^A x = \mathrm{proj}_{A^{-1}(0)} x, \quad \text{if } A^{-1}(0) \neq \emptyset.$

2B. Algebraic relationship linking λ and x

Properties of $\varphi(\lambda, x) := \lambda \|x - (I + \lambda A)^{-1} x\|$

- For any $x_1, x_2 \in \mathcal{H}$ and $\lambda > 0$,

$$|\varphi(\lambda, x_1) - \varphi(\lambda, x_2)| \leq \lambda \|x_2 - x_1\|.$$

- For any $x \in \mathcal{H}$ and $0 < \lambda_1 \leq \lambda_2$,

$$\frac{\lambda_2}{\lambda_1} \varphi(\lambda_1, x) \leq \varphi(\lambda_2, x) \leq \left(\frac{\lambda_2}{\lambda_1}\right)^2 \varphi(\lambda_1, x) \tag{9}$$

  and $\varphi(\lambda, x) = 0$ if and only if $0 \in A(x)$.

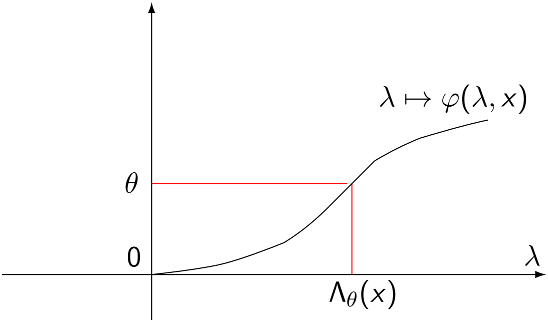
- For any $x \notin A^{-1}(0)$, $\lambda \in [0, \infty[\mapsto \varphi(\lambda, x) \in \mathbb{R}^+$ is continuous, strictly increasing, $\varphi(0, x) = 0$, and $\lim_{\lambda \to +\infty} \varphi(\lambda, x) = +\infty$.

2C. Algebraic relationship linking λ and x

$\lambda \in [0, \infty[\mapsto \varphi(\lambda, x) \in \mathbb{R}^+$ continuous, strict. increasing, $0 \uparrow +\infty$.

Definition

$$\Lambda_\theta : \mathcal{H} \setminus A^{-1}(0) \to]0, \infty[, \quad \Lambda_\theta(x) := \varphi(\cdot, x)^{-1}(\theta). \tag{10}$$



Example: $A = \mathrm{rot}(0; \frac{\pi}{2})$

3A. Existence result for the Cauchy problem

$$\begin{cases} \dot{x}(t) + x(t) - (I + \lambda(t)A)^{-1}x(t) = 0, \\ \lambda(t)\|(I + \lambda(t)A)^{-1}x(t) - x(t)\| = \theta, \\ x(0) = x_0 \in \mathcal{H} \setminus A^{-1}(0). \end{cases} \tag{11}$$

$$\Updownarrow$$

$$\dot{x}(t) + x(t) - (I + \Lambda_\theta(x(t))A)^{-1}\, x(t) = 0; \quad x(0) = x_0.$$

$$\Updownarrow$$

$$\dot{x}(t) = F(x(t)); \quad x(0) = x_0.$$

3B. Existence result for the Cauchy problem

$$F : \Omega = \mathcal{H} \setminus A^{-1}(0) \to \mathcal{H}$$
$$F(x) := J^{A}_{\Lambda_\theta(x)} x - x.$$

Continuity properties of F

- F: locally Lipschitz continuous.
- $x \in \mathcal{H} \setminus A^{-1}(0) \longmapsto \frac{1}{\Lambda_\theta(x)}$: Lipschitz continuous with constant θ.

Cauchy problem: $\dot{x}(t) = F(x(t))$; $x(0) = x_0$.

- Cauchy-Lipschitz theorem: local existence, and uniqueness.
- Global existence: estimate $0 \leq \dot{\lambda}(\cdot) \leq \lambda(\cdot)$.

$$
\text{(LSP)} \begin{cases} \dot{x}(t) + x(t) - (I + \lambda(t)A)^{-1}x(t) = 0, \\ \lambda(t)\|(I + \lambda(t)A)^{-1}x(t) - x(t)\| = \theta, \\ x(0) = x_0 \in \mathcal{H} \setminus A^{-1}(0). \end{cases}
$$

Theorem 1 (existence and uniqueness)

There exists a unique solution $(x, \lambda) : [0, +\infty[\to \mathcal{H} \times \mathbb{R}_{++}$ *of* (LSP)*;* $x(\cdot)$ *is* $\mathcal{C}^1$*, and* $\lambda(\cdot)$ *is locally Lipschitz continuous. Moreover,*

(i) $\lambda(\cdot)$ *is non-decreasing;* $\quad 0 \leq \dot{\lambda}(\cdot) \leq \lambda(\cdot)$*;*

(ii) $t \mapsto \|J^A_{\lambda(t)}x(t) - x(t)\|$ *is non-increasing.*

4. Asymptotic behavior

Theorem 2 (convergence)

Suppose that $A^{-1}(0) \neq \emptyset$. *Let* $(x, \lambda) : [0, +\infty[\to \mathcal{H} \times \mathbb{R}_{++}$ *be the unique solution of the Cauchy problem* (LSP) *with* $x_0 \in \mathcal{H} \setminus A^{-1}(0)$. *Then,*

(i) $\|\dot{x}(t)\| = \|x(t) - J^A_{\lambda(t)} x(t)\| \leq d_0/\sqrt{2t}$;
hence $\lim_{t\to+\infty} \|\dot{x}(t)\| = 0$;

(ii) $\lambda(t) \geq \frac{\theta}{d_0}\sqrt{2t}$;
hence $\lim_{t\to+\infty} \lambda(t) = +\infty$;

(iii) $w - \lim_{t\to+\infty} x(t) = x_\infty$ *exists, for some* $x_\infty \in A^{-1}(0)$,

where d_0 *is the distance from* x_0 *to* $A^{-1}(0)$.

- Weak convergence: Opial's lemma, Fejer monotone property.
- Strong convergence: A strongly monotone; $A = \partial f$, f inf-compact.

5A. Link with the regularized Newton system

$x_0 \notin A^{-1}(0)$, $(x,\lambda) : [0,+\infty[\to \mathcal{H} \times \mathbb{R}_{++}$ solution of (LSP).

$$y(t) := (I + \lambda(t)A)^{-1}x(t), \qquad v(t) := \frac{1}{\lambda(t)}(x(t) - y(t)). \tag{12}$$

Claim: $y(\cdot)$ is solution of a regularized Newton system.

Time derivation of $\lambda(t)v(t) + y(t) - x(t) = 0$, and (LSP) gives

$$\begin{cases} v(t) \in Ay(t); \\ \dot{y}(t) + \lambda(t)\dot{v}(t) + (\lambda(t) + \dot{\lambda}(t))v(t) = 0. \end{cases} \tag{13}$$

5B. Link with the regularized Newton system

Time rescaling

$$\tau(t) = \int_0^t \frac{\lambda(u) + \dot{\lambda}(u)}{\lambda(u)} du = t + \ln(\lambda(t)/\lambda(0)). \qquad (14)$$

Since $1 \leq \dfrac{\lambda(u) + \dot{\lambda}(u)}{\lambda(u)} \leq 2, \;\; t \leq \tau(t) \leq 2t.$

Set $y(t) = \tilde{y}(\tau(t)), \quad v(t) = \tilde{v}(\tau(t)).$

$$\begin{cases} \tilde{v} \in A\tilde{y}; \\ \frac{1}{\lambda \circ \tau^{-1}} \dfrac{d}{d\tau}\tilde{y} + \dfrac{d}{d\tau}\tilde{v} + \tilde{v} = 0. \end{cases} \qquad (15)$$

Regularized Newton system [AS, SICON 2011].
Levenberg-Marquardt regularization parameter $\frac{1}{\lambda \circ \tau^{-1}} \to 0$ as $\tau \to +\infty$.

6A. The subdifferential case

Suppose $\arg\min f \neq \emptyset$; $d_0 := \inf\{\|x_0 - z\| : \; z \in \arg\min f\} = \|x_0 - \bar{z}\|$.

Theorem 3 (rate of convergence)

Suppose that $f(x(0)) < +\infty$. Then,

(i) $t \mapsto f(x(t))$ *and* $t \mapsto f(y(t))$ *are non-increasing;*

(ii) *Set* $\kappa = \sqrt{\theta/d_0^3}$. *For any* $t \geq 0$

$$f(x(t)) - f(\bar{z}) \leq \frac{f(x_0) - f(\bar{z})}{\left[1 + \frac{t\kappa\sqrt{f(x_0) - f(\bar{z})}}{2 + 3\kappa\sqrt{f(x_0) - f(\bar{z})}}\right]^2} = \mathcal{O}(\frac{1}{t^2}). \tag{16}$$

6B. The subdifferential case

Differential inequality: $\frac{d}{dt}\beta \leq -c\beta^{3/2}, \quad \beta(t) := f(x(t)) - f(\bar{z}).$

Step 1: $$\frac{d}{dt}f(x(t)) \leq f(y(t)) - f(x(t)).$$

Integrate $\dot{x} + x = y$ and apply Jensen's inequality (convexity of f)

$$f(x(t+h)) \leq \int_t^{t+h} \left[e^{-h}f(x(t)) + (1-e^{-h})f(y(u))\right] \frac{e^u}{e^{t+h}-e^t}\, du.$$

$f(y(\cdot))$ non-increasing

$$f(x(t+h)) \leq e^{-h}f(x(t)) + (1-e^{-h})f(y(t))$$
$$\frac{f(x(t+h)) - f(x(t))}{h} \leq \frac{1-e^{-h}}{h}(f(y(t)) - f(x(t)).$$

6C. The subdifferential case

Step 2: $$f(y) - f(x) \leq -\frac{\kappa(f(x) - f(\bar{z}))^{3/2}}{1 + (3\kappa/2)(f(x_0) - f(\bar{z}))^{1/2}}.$$

- Since $v(t) \in \partial f(y(t))$, $\lambda(t)v(t) = x(t) - y(t)$, and $\lambda(t)^2\|v(t)\| = \theta$

$$\begin{aligned} f(x(t)) \geq f(y(t)) + \langle x(t) - y(t), v(t)\rangle &\geq f(y(t)) + \lambda(t)\|v(t)\|^2 \\ &= f(y(t)) + \sqrt{\theta}\|v(t)\|^{3/2}. \end{aligned}$$

- Since $v(t) \in \partial f(y(t))$, for any $t \geq 0$

$$\begin{aligned} f(y(t)) - f(\bar{z}) \leq \langle y(t) - \bar{z}, v(t)\rangle &\leq \|y(t) - \bar{z}\|\|v(t)\| \\ &\leq \|x(t) - \bar{z}\| \, \|v(t)\| \leq d_0\|v(t)\| \end{aligned}$$

where we have used $y(t) = J^A_{\lambda(t)}(x(t))$, $\bar{z} = J^A_{\lambda(t)}(\bar{z})$, $J^A_{\lambda(t)}$ nonexpansive, and $t \mapsto \|x(t) - \bar{z}\|$ non-increasing. Combining the above inequalities

$$f(x(t)) \geq f(y(t)) + (f(y(t) - f(\bar{z}))^{3/2}\sqrt{\theta/d_0^3}.$$

$$f(x(t)) - f(\bar{z}) \geq f(y(t)) - f(\bar{z}) + (f(y(t) - f(\bar{z}))^{3/2}\sqrt{\theta/d_0^3}. \qquad (17)$$

Convexity of $r \mapsto r^{3/2}$: If $a, b, c \geq 0$ and $a \geq b + cb^{3/2}$ then

$$b \leq a - \frac{ca^{3/2}}{1 + (3c/2)a^{1/2}}$$

Hence

$$f(y) \leq f(x) - \frac{\kappa(f(x) - f(\bar{z}))^{3/2}}{1 + (3\kappa/2)(f(x) - f(\bar{z}))^{1/2}}, \qquad (18)$$

Since $f(x(\cdot))$ is non-increasing

$$f(y) - f(x) \leq -\frac{\kappa(f(x) - f(\bar{z}))^{3/2}}{1 + (3\kappa/2)(f(x_0) - f(\bar{z}))^{1/2}}. \qquad (19)$$

6E. The subdifferential case: $y(\cdot)$ versus $x(\cdot)$

$$y(t) = (I + \lambda(t)\partial f)^{-1}x(t) = J^{\partial f}_{\lambda(t)}x(t) = \mathrm{prox}_{\lambda(t)f}x(t).$$

(1) $y(\cdot)$: solution of a regularized Newton system.

(2) $\|x(t) - J^{\partial f}_{\lambda(t)}x(t)\| = \|\dot{x}(t)\| \leq d_0/\sqrt{2t} \to 0$. Hence

$$w - \lim_{t\to+\infty} y(t) = w - \lim_{t\to+\infty} x(t) \in \arg\min f.$$

(3) $f(x(t)) \geq f_{\lambda(t)}(x(t)) \geq f(J^{\partial f}_{\lambda(t)}x(t))$. Hence

$$0 \leq f(y(t)) - \inf_{\mathcal{H}} f \leq f(x(t)) - \inf_{\mathcal{H}} f = \mathcal{O}(\frac{1}{t^2}).$$

(4) $y(\cdot) \in \mathrm{dom}\partial f$: more (space) regularity than $x(t)$.

(5) $f(y(t)) \downarrow \inf_{\mathcal{H}} f$ even if $\arg\min f = \emptyset$.

7A. A large-step prox. method for convex optimization

Algorithm 1:

(0) $x_0 \in \mathrm{dom}(f)$, $\sigma \in [0,1[$, $\theta > 0$ given, set $k = 1$;

(1) choose $\lambda_k > 0$, and find $x_k, v_k \in \mathcal{H}$, $\varepsilon_k \geq 0$ such that

$$v_k \in \partial_{\varepsilon_k} f(x_k), \tag{20}$$

$$\|\lambda_k v_k + x_k - x_{k-1}\|^2 + 2\lambda_k \varepsilon_k \leq \sigma^2 \|x_k - x_{k-1}\|^2, \tag{21}$$

$$\lambda_k \|x_k - x_{k-1}\| \geq \theta \text{ or } v_k = 0; \tag{22}$$

(2) if $v_k = 0$ STOP, output x_k; otherwise $k \leftarrow k + 1$, go to step 1.

end

- Relative error for HPE, Solodov and Svaiter, SVVA, JCA (1999).
- Large-step condition, Montero and Svaiter, SIOPT (2010, 2012).

Theorem 4 (complexity, $\varepsilon_k = 0$)

$\mathcal{D}_0 = \sup\{\|x-y\| \mid \max\{f(x), f(y)\} \leq f(x_0)\} < +\infty$, $\kappa_0 := \sqrt{\frac{\theta(1-\sigma)}{\mathcal{D}_0^3}}$.

(i) $$f(x_k) - f(\bar{x}) \leq \frac{f(x_0) - f(\bar{x})}{\left[1 + k\frac{\kappa_0\sqrt{f(x_0)-f(\bar{x})}}{2+3\kappa_0\sqrt{f(x_0)-f(\bar{x})}}\right]^2} = \mathcal{O}(1/k^2).$$

(ii) *for each $k \geq 2$ even, there exists $j \in \{k/2+1, \ldots, k\}$ such that*

$$\|v_j\| \leq \frac{4}{\sqrt[3]{\theta(1-\sigma)}} \left[\frac{f(x_0)-f(\bar{x})}{k\left[2+k\frac{\kappa_0\sqrt{f(x_0)-f(\bar{x})}}{2+3\kappa_0\sqrt{f(x_0)-f(\bar{x})}}\right]^2}\right]^{2/3} = \mathcal{O}(1/k^2).$$

8A. $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ prox. Newton method for convex optim.

$$(\mathcal{P}) \quad \text{minimize } f(x) \quad \text{s.t.} \quad x \in \mathcal{H}.$$

AS1) $f : \mathcal{H} \to \mathbb{R}$ convex, twice continuously differentiable;

AS2) $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathcal{H}$;

AS3) $\mathcal{D}_0 = \sup\{\|y - x\| \mid \max\{f(x), f(y)\} \leq f(x_0)\} < \infty$.

Algorithm 2

(0) $x_0 \in \mathcal{H}$, $0 < \sigma_\ell < \sigma_u < 1$ b given, set $k = 1$;

(1) if $\nabla f(x_{k-1}) = 0$ then **stop**. Otherwise, compute $\lambda_k > 0$ s.t.

$$\frac{2\sigma_\ell}{L} \leq \lambda_k \|(I + \lambda_k \nabla^2 f(x_{k-1}))^{-1} \lambda_k \nabla f(x_{k-1})\| \leq \frac{2\sigma_u}{L}; \tag{23}$$

(2) set $x_k = x_{k-1} - (I + \lambda_k \nabla^2 f(x_{k-1}))^{-1} \lambda_k \nabla f(x_{k-1})$;

(3) set $k \leftarrow k + 1$ and go to step 1. **end**.

8B. $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ prox. Newton method for convex optim.

Proximal method

$$y = (I + \lambda \nabla f)^{-1} x \Longleftrightarrow \lambda \nabla f(y) + y - x = 0.$$

Perform a single Newton iteration from the current iterate x

$$\lambda \nabla f(x) + (I + \lambda \nabla^2 f(x))(y - x) = 0.$$

i.e.,

$$y = x - (I + \lambda \nabla^2 f(x))^{-1} \lambda \nabla f(x).$$

Hence the

Proximal-Newton method

$$x_k = x_{k-1} - (I + \lambda_k \nabla^2 f(x_{k-1}))^{-1} \lambda_k \nabla f(x_{k-1}).$$

8C. $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ prox. Newton method for convex optim.

Theorem 5 (complexity)

Suppose AS1), AS2), AS3) *hold*, $\{x_k\}$ *generated by* Algorithm 2. *Let* $\bar{x}$ *be a solution of* $(\mathcal{P})$. *For any given tolerance* $\varepsilon > 0$ *define*

$$\kappa_0 = \sqrt{\frac{2\sigma_\ell(1-\sigma_u)}{LD_0^3}}, \quad K = \frac{2 + 3\kappa_0\sqrt{f(x_0) - f(\bar{x})}}{\kappa_0\sqrt{\varepsilon}},$$

$$J = \frac{2L^{1/6}\left(2 + 3\kappa_0\sqrt{f(x_0) - f(\bar{x})}\right)^{2/3}}{[2\sigma_\ell(1-\sigma_u)]^{1/6}\,\kappa_0^{1/3}\sqrt{\varepsilon}}.$$

Then, the following statements hold:

(a) *for any* $k \geq K$, $f(x_k) - f(\bar{x}) \leq \varepsilon$;

(b) *there exists* $j \leq 2\lceil J \rceil$ *such that* $\|\nabla f(x_j)\| \leq \varepsilon$.

8D. $\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$ prox. Newton method for convex optim.

Solving w.r. $\lambda > 0$

$$\frac{2\sigma_\ell}{L} \leq \lambda \|(I + \lambda \nabla^2 f(x_{k-1}))^{-1} \lambda \nabla f(x_{k-1})\| \leq \frac{2\sigma_u}{L};$$

The solution set is a closed interval $[\lambda_l, \lambda_u]$ s.t. $\frac{\lambda_u}{\lambda_l} \geq \sqrt{\frac{\sigma_u}{\sigma_l}}$.
Binary search in $\ln \lambda$ may be used to find λ_k.

$$\Phi(\lambda) := \lambda \|(I + \lambda \nabla^2 f(x))^{-1} \lambda \nabla f(x)\|, \quad \Psi(\mu) := \Phi(\frac{1}{\lambda}).$$

$$\frac{2\sigma_\ell}{L} \leq \Psi(\mu) \leq \frac{2\sigma_u}{L};$$

$\mu \mapsto \Psi^2(\mu)$, and $\mu \mapsto \ln(\Psi(\mu))$ are convex.

9A. Perspective, large step proximal methods

Main result

(LSP) is a large step proximal method with $\mathcal{O}(\frac{1}{k^2})$ convergence.

Related results

- Compare with Nesterov steepest descent.
- Rockafellar, SICON 1976: A is α-strongly monotone, $A^{-1}(0) = \bar{x}$,

$$\|x_{k+1} - \bar{x}\| \leq \frac{1}{1 + \alpha\lambda_k}\|x_k - \bar{x}\|.$$

  Superlinear convergence when $\lambda_k \to +\infty$.
- Att-Redont-Svaiter, JOTA 2013, regularized Newton method, Levenberg-Marquardt parameter $\mu_k = \frac{1}{\lambda_k} = \|\nabla f(x_k)\|^{\frac{1}{3}}$, yield convergence of (x_k) at the order $\frac{4}{3}$.
- Other algebraic relation: $\lambda\|(I + \lambda A)^{-1}x - x\|^\gamma = \theta$. (LSP): $\gamma = 1$.

9B. Perspective, open questions

Possible extensions

- Extension of $f(x(t)) - \inf f = \mathcal{O}(\frac{1}{t^2})$ and $f(x_k) - \inf f = \mathcal{O}(\frac{1}{k^2})$ to the non potential case: convex-concave saddle value problems? general maximal monotone operators (via Fitzpatrick function)?
- Combine (LSP) with splitting methods: forward-backward method, alternating proximal minimization?
- Implementation of the method and applications.
- Fast converging methods and second-order methods: in time (Nesterov, FISTA...), or space (Newton-like methods: LSP). Compare, combine them?
- (LSP) is linked to a regularized Newton method which still works in the nonconvex nonsmooth setting. Convergence in KL setting?

Appendix: Isotropic linear monotone operator

$\alpha > 0$ positive constant, $A = \alpha I$, i.e., for every $x \in \mathcal{H}$, $Ax = \alpha x$.

$$(\lambda A + I)^{-1} x = \frac{1}{1 + \lambda\alpha} x \tag{24}$$

$$x - (\lambda A + I)^{-1} x = \frac{\lambda\alpha}{1 + \lambda\alpha} x. \tag{25}$$

Given $x_0 \neq 0$, (LSP) can be written

$$\begin{cases} \dot{x}(t) + \frac{\alpha\lambda(t)}{1+\alpha\lambda(t)} x(t) = 0, \qquad \lambda(t) > 0, \\ \frac{\alpha\lambda(t)^2}{1+\alpha\lambda(t)} \|x(t)\| = \theta, \\ x(0) = x_0 \in \mathcal{H} \setminus A^{-1}(0). \end{cases} \tag{26}$$

Let us first integrate the above linear differential equation.

We have $x(t) = e^{-\Delta(t)} x_0$ with $\Delta(t) := \int_0^t \frac{\alpha\lambda(\tau)}{1+\alpha\lambda(\tau)} d\tau$. Hence

$$\frac{\alpha\lambda(t)^2}{1+\alpha\lambda(t)} e^{-\Delta(t)} = \frac{\theta}{\|x_0\|}. \tag{27}$$

First, check this equation at time $t = 0$. Equivalently

$$\frac{\alpha\lambda(0)^2}{1+\alpha\lambda(0)} = \frac{\theta}{\|x_0\|}. \tag{28}$$

This equation defines uniquely $\lambda(0) > 0$, because $\xi \mapsto \frac{\alpha\xi^2}{1+\alpha\xi}$ is strictly increasing from $[0,+\infty[$ onto $[0,+\infty[$. Thus, the only thing we have to prove is the existence of a positive function $t \mapsto \lambda(t)$ such that

$$h(t) := \frac{\alpha\lambda(t)^2}{1+\alpha\lambda(t)} e^{-\Delta(t)} \quad \text{is constant on } [0,+\infty[. \tag{29}$$

Writing that the derivative h' is identically zero on $[0,+\infty[$, we obtain

Appendix: Isotropic linear monotone operator

$$\lambda'(t)(\alpha\lambda(t) + 2) - \alpha\lambda(t)^2 = 0. \tag{30}$$

After integration, we obtain

$$\alpha \ln \lambda(t) - \frac{2}{\lambda(t)} = \alpha t + \alpha \ln \lambda(0) - \frac{2}{\lambda(0)}. \tag{31}$$

Let us introduce the function $g :]0, +\infty[\to \mathbb{R}$

$$g(\xi) = \alpha \ln \xi - \frac{2}{\xi}. \tag{32}$$

As t increases from 0 to $+\infty$, $g(t)$ is strictly increasing from $-\infty$ to $+\infty$. Thus, for each $t > 0$, (31) has a unique solution $\lambda(t) > 0$. Moreover, $t \to \lambda(t)$ is increasing, continuously differentiable, and $\lim_{t\to\infty} \lambda(t) = +\infty$. Returning to (31), we obtain that $\lambda(t) \approx e^t$.

Appendix: Antisymmetric linear monotone operator

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

$$(\lambda A + I)^{-1} x = \frac{1}{1+\lambda^2}\Big(\xi + \lambda\eta, \eta - \lambda\xi\Big) \tag{33}$$

$$x - (\lambda A + I)^{-1} x = \frac{\lambda}{1+\lambda^2}\Big(\lambda\xi - \eta, \lambda\eta + \xi\Big). \tag{34}$$

$$\dot{\xi}(t) + \frac{\lambda(t)}{1+\lambda(t)^2}\Big(\lambda(t)\xi(t) - \eta(t)\Big) = 0, \qquad \lambda(t) > 0, \tag{35}$$

$$\dot{\eta}(t) + \frac{\lambda(t)}{1+\lambda(t)^2}\Big(\lambda(t)\eta(t) + \xi(t)\Big) = 0, \qquad \lambda(t) > 0, \tag{36}$$

$$\frac{\lambda(t)^2}{\sqrt{1+\lambda(t)^2}}\sqrt{\xi(t)^2 + \eta(t)^2} = \theta, \tag{37}$$

Set $u(t) = \xi(t)^2 + \eta(t)^2$. After multiplying (35) by $\xi(t)$, and multiplying (36) by $\eta(t)$, then adding the results, we obtain

$$u'(t) + \frac{2\lambda(t)^2}{1+\lambda(t)^2} u(t) = 0.$$

Set

$$\Delta(t) := \int_0^t \frac{2\lambda(\tau)^2}{1+\lambda(\tau)^2} d\tau. \tag{38}$$

We have

$$u(t) = e^{-\Delta(t)} u(0). \tag{39}$$

Equation (37) becomes

$$\frac{\lambda(t)^2}{\sqrt{1+\lambda(t)^2}} e^{-\frac{\Delta(t)}{2}} = \frac{\theta}{\|x_0\|}. \tag{40}$$

First, check this equation at time $t = 0$. Equivalently

$$\frac{\lambda(0)^2}{\sqrt{1+\lambda(0)^2}} = \frac{\theta}{\|x_0\|}. \tag{41}$$

This equation defines uniquely $\lambda(0) > 0$, because the function $\rho \mapsto \frac{\rho^2}{\sqrt{1+\rho^2}}$ is strictly increasing from $[0, +\infty[$ onto $[0, +\infty[$. Thus, we just need to prove the existence of a positive function $t \mapsto \lambda(t)$ s.t.

$$h(t) := \frac{\lambda(t)^2}{\sqrt{1+\lambda(t)^2}} e^{-\frac{\Delta(t)}{2}} \quad \text{is constant on } [0, +\infty[. \tag{42}$$

Writing that the derivative h' is identically zero on $[0, +\infty[$, we obtain that $\lambda(\cdot)$ must satisfy

$$\lambda'(t)(2\lambda(t) + \lambda(t)^3) - \lambda(t)^3 = 0. \tag{43}$$

After integration of this first-order differential equation, with Cauchy data $\lambda(0)$, we obtain

$$\lambda(t) - \frac{2}{\lambda(t)} = t + \lambda(0) - \frac{2}{\lambda(0)}. \tag{44}$$

Let us introduce the function $g :]0, +\infty[\to \mathbb{R}$

$$g(\rho) = \rho - \frac{2}{\rho}. \tag{45}$$

As t increases from 0 to $+\infty$, $g(t)$ is strictly increasing from $-\infty$ to $+\infty$. Thus, for each $t > 0$, (44) has a unique solution $\lambda(t) > 0$. Moreover, the mapping $t \to \lambda(t)$ is increasing, continuously differentiable, and $\lim_{t\to\infty} \lambda(t) = +\infty$. Returning to (44), we obtain that $\lambda(t) \approx t$ as $t \to +\infty$.

References

- B. Abbas, H. Attouch, and B. F. Svaiter, Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces, *JOTA*, DOI 10.1007/s10957-013-0414-5, (2013).
- H. Attouch, Viscosity solutions of minimization problems, *SIAM J. Optim.*, **6** (1996), No. 3, pp. 769–806.
- H. Attouch, J. Bolte, and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, *Math. Program.*, **137** (2013), No. 1, pp. 91-129.
- H. Attouch, P. Redont, and B. F. Svaiter, Global convergence of a closed-loop regularized Newton method for solving monotone inclusions in Hilbert spaces, *JOTA*, **157** (2013), pp. 624–650.

References

- H. Attouch and B. F. Svaiter, A continuous dynamical Newton-like approach to solving monotone inclusions, *SIAM J. Control Optim.*, **49** (2011), pp. 574–598.
- H. Bauschke and P. Combettes, *Convex analysis and monotone operator theory*, CMS books in Mathematics, Springer, 2011.
- H. Brézis, Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert, North-Holland/Elsevier, New-York, 1973.
- R.E. Bruck, Asymptotic convergence of nonlinear contraction semigroups in Hilbert spaces, *J. Funct. Anal.*, **18** (1975), pp. 15–26.

References

- C. Gonzaga, E. W. Karas, Fine tuning Nesterov's steepest descent algorithm for differentiable convex programming, *Math. Program.*, **138** (2013), pp. 141–166.
- A. Griewank, The modification of Newton's method for unconstrained optimization by bounding cubic terms, Technical Report NA/12, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, 1981.
- B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Française Informat. Recherche Opérationnelle*, **4** (1970), (Ser. R-3), pp. 154–158.
- R. D. C. Monteiro and B. F. Svaiter, On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean, *SIAM J. Optim.*, **20** (2010), No. 6, pp. 2755–2787.

References

- R. D. C. Monteiro and B. F. Svaiter, Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems, *SIAM J. Optim.*, **22** (2012), No. 3, pp. 914–935.
- Y. Nesterov, Introductory lectures on convex optimization, A basic course, Kluwer, Boston (2004).
- Y. Nesterov and B. T. Polyak, Cubic regularization of Newton method and its global performance, *Math. Program.*, **108** (2006), (1, Ser. A), pp. 177–205.
- Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.* **73** (1967), pp. 591–597.

References

- R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.*, **1** (1976), No. 2, pp. 97–116.
- R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, **14** (1976), No. 5, pp. 877–898.
- M. V. Solodov and B. F. Svaiter, A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, **7** (1999), No. 4, pp. 323–345.
- M. V. Solodov and B. F. Svaiter, A hybrid projection-proximal point algorithm. *J. Convex Anal.*, **6** (1999), No. 1, pp. 59–70.
- M. Weiser, P. Deuflhard, and B. Erdmann, Affine conjugate adaptive Newton methods for nonlinear elastomechanics, *Optim. Methods Softw.*, **22** (2007), No. 3, pp. 413–431.