Proximal algorithms for nonconvex and nonsmooth minimization problems

Radu Ioan Boţ
(the talk relies on joint works with
Sebastian Banert, Robert Csetnek and Szilárd László)

University of Vienna
Faculty of Mathematics
Oskar-Morgenstern-Platz 1
1090 Vienna
Austria
www.mat.univie.ac.at/∼rabot

Research Seminar of the Institute for Statistics and Mathematics
Vienna University of Economics and Business
May 3, 2019

universität wien

FWF Der Wissenschaftsfonds.

Contents

The minimization of a nonsmooth plus a smooth function: the convex case

Let $\mathcal{H}$ be a real Hilbert space and

- $f : \mathcal{H} \to \overline{\mathbb{R}}$ a proper, convex, lower semicontinuous function;
- $g : \mathcal{H} \to \mathbb{R}$ a convex and Fréchet differentiable function such that ∇g is $L_{\nabla g}$-Lipschitz continuous.

Consider the convex optimization problem

$$\min_{x \in \mathcal{H}} \{f(x) + g(x)\}. \tag{1}$$

Proximal-gradient splitting

Proximal-gradient algorithm

$$(\forall n \geq 0) \;\; x_{n+1} = \operatorname{prox}_{\gamma f}(x_n - \gamma \nabla g(x_n))$$

Proximal operator

If $f \in \Gamma(\mathcal{H}) := \{k : \mathcal{H} \to \overline{\mathbb{R}} : k$ is proper, convex and lower semicontinuous$\}$ and $\gamma > 0$, then

$$\operatorname{prox}_{\gamma f}(x) := \operatorname{argmin}_{u \in \mathcal{H}} \left\{ f(u) + \frac{1}{2\gamma}\|u - x\|^2 \right\} \;\; \forall x \in \mathcal{H}.$$

Convergence of the proximal-gradient algorithm

If $\gamma \in \left(0, \frac{2}{L_{\nabla g}}\right)$, $x_0 \in \mathcal{H}$ and (1) is solvable, then $(x_n)_{n \geq 0}$ converges weakly to an optimal solution of (1).
If x^* is an optimal solution of (1) and $\gamma := \frac{1}{L_{\nabla g}}$, then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) \leq \frac{L_{\nabla g}\|x_0 - x^*\|^2}{2n} \;\; \forall n \geq 1.$$

Proximal-gradient splitting

Proximal-gradient algorithm

$$(\forall n \geq 0) \;\; x_{n+1} = \mathrm{prox}_{\gamma f}(x_n - \gamma \nabla g(x_n))$$

Proximal operator

If $f \in \Gamma(\mathcal{H}) := \{k : \mathcal{H} \to \overline{\mathbb{R}} : k$ is proper, convex and lower semicontinuous$\}$ and $\gamma > 0$, then

$$\mathrm{prox}_{\gamma f}(x) := \mathrm{argmin}_{u \in \mathcal{H}} \left\{ f(u) + \frac{1}{2\gamma}\|u - x\|^2 \right\} \;\; \forall x \in \mathcal{H}.$$

Convergence of the proximal-gradient algorithm

If $\gamma \in \left(0, \frac{2}{L_{\nabla g}}\right)$, $x_0 \in \mathcal{H}$ and (1) is solvable, then $(x_n)_{n \geq 0}$ converges weakly to an optimal solution of (1).

If x^* is an optimal solution of (1) and $\gamma := \frac{1}{L_{\nabla g}}$, then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) \leq \frac{L_{\nabla g}\|x_0 - x^*\|^2}{2n} \;\; \forall n \geq 1.$$

Proximal-gradient splitting

Proximal-gradient algorithm

$$(\forall n \geq 0) \;\; x_{n+1} = \mathrm{prox}_{\gamma f}(x_n - \gamma \nabla g(x_n))$$

Proximal operator

If $f \in \Gamma(\mathcal{H}) := \{k : \mathcal{H} \to \overline{\mathbb{R}} : k$ is proper, convex and lower semicontinuous$\}$ and $\gamma > 0$, then

$$\mathrm{prox}_{\gamma f}(x) := \mathrm{argmin}_{u \in \mathcal{H}} \left\{ f(u) + \tfrac{1}{2\gamma}\|u - x\|^2 \right\} \;\; \forall x \in \mathcal{H}.$$

Convergence of the proximal-gradient algorithm

If $\gamma \in \left(0, \frac{2}{L_{\nabla g}}\right)$, $x_0 \in \mathcal{H}$ and (1) is solvable, then $(x_n)_{n \geq 0}$ converges weakly to an optimal solution of (1).

If x^* is an optimal solution of (1) and $\gamma := \frac{1}{L_{\nabla g}}$, then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) \leq \frac{L_{\nabla g}\|x_0 - x^*\|^2}{2n} \;\; \forall n \geq 1.$$

Accelerated proximal-gradient splitting

Accelerated proximal-gradient splitting (FISTA)

$$(\forall n \geq 1) \quad \left\lfloor \begin{array}{l} x_n = \operatorname{prox}_{\frac{1}{L_{\nabla g}} f} \left(y_n - \frac{1}{L_{\nabla g}} \nabla g(y_n) \right) \\ y_{n+1} = x_n + \alpha_n (x_n - x_{n-1}) \end{array} \right.$$

Convergence of FISTA (Beck, Teboulle, 2009)

Let be $y_1 = x_0 \in \mathcal{H}$ and $\alpha_n = \frac{t_n - 1}{t_{n+1}}$ $\forall n \geq 1$, where $t_1 := 1$ and

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2} (\Leftrightarrow t_{n+1}^2 - t_{n+1} = t_n^2).$$

If x^* is an optimal solution of (1), then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) \leq \frac{2L_{\nabla g} \|x_0 - x^*\|^2}{(n+1)^2} \ \forall n \geq 1.$$

Accelerated proximal-gradient splitting

Accelerated proximal-gradient splitting (FISTA)

$$(\forall n \geq 1) \quad \left\lfloor \begin{array}{l} x_n = \operatorname{prox}_{\frac{1}{L_{\nabla g}} f}\left(y_n - \frac{1}{L_{\nabla g}} \nabla g(y_n)\right) \\ y_{n+1} = x_n + \alpha_n (x_n - x_{n-1}) \end{array} \right.$$

Convergence of FISTA (Beck, Teboulle, 2009)

Let be $y_1 = x_0 \in \mathcal{H}$ and $\alpha_n = \frac{t_n - 1}{t_{n+1}}$ $\forall n \geq 1$, where $t_1 := 1$ and

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2} (\Leftrightarrow t_{n+1}^2 - t_{n+1} = t_n^2).$$

If x^* is an optimal solution of (1), then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) \leq \frac{2L_{\nabla g}\|x_0 - x^*\|^2}{(n+1)^2} \ \forall n \geq 1.$$

Convergence of the FISTA iterates (Chambolle, Dossal, 2014)

Let be $y_1 = x_0 \in \mathcal{H}$ and $\alpha_n = \frac{t_n - 1}{t_{n+1}}$ $\forall n \geq 1$, where $t_1 := 1$ and for $a > 3$

$$t_n = \frac{n + a - 1}{a} (\Rightarrow t_{n+1}^2 - t_{n+1} \leq t_n^2).$$

Then $(x_n)_{n \geq 0}$ converges weakly to an optimal solution of (1).
If x^* is an optimal solution of (1), then

$$0 \leq (f + g)(x_n) - (f + g)(x^*) \leq \frac{L_{\nabla g} a^2 \|x_0 - x^*\|^2}{2(n + a - 1)^2} \ \forall n \geq 1.$$

(Attouch, Peypouquet, 2015)

In the hypotheses of (Chambolle, Dossal, 2014), if x^* is an optimal solution of (1), then

$$0 \leq (f + g)(x_n) - (f + g)(x^*) = o\left(\frac{1}{n^2}\right).$$

Convergence of the FISTA iterates (Chambolle, Dossal, 2014)

Let be $y_1 = x_0 \in \mathcal{H}$ and $\alpha_n = \frac{t_n - 1}{t_{n+1}}$ $\forall n \geq 1$, where $t_1 := 1$ and for $a > 3$

$$t_n = \frac{n+a-1}{a} (\Rightarrow t_{n+1}^2 - t_{n+1} \leq t_n^2).$$

Then $(x_n)_{n \geq 0}$ converges weakly to an optimal solution of (1).
If x^* is an optimal solution of (1), then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) \leq \frac{L_{\nabla g} a^2 \|x_0 - x^*\|^2}{2(n+a-1)^2} \ \forall n \geq 1.$$

(Attouch, Peypouquet, 2015)

In the hypotheses of (Chambolle, Dossal, 2014), if x^* is an optimal solution of (1), then

$$0 \leq (f+g)(x_n) - (f+g)(x^*) = o\left(\frac{1}{n^2}\right).$$

The minimization of the sum of two nonconvex functions

Consider the optimization problem

$$\min_{x \in \mathcal{H}} \{f(x) + g(x)\}. \qquad (2)$$

- $\mathcal{H}$ is a finite-dimensional real Hilbert space;
- $f : \mathcal{H} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous and bounded from below;
- $g : \mathcal{H} \to \mathbb{R}$ is Fréchet differentiable and ∇g is $L_{\nabla g}$-Lipschitz continuous.

Inertial proximal-gradient algorithm

For $0 < \underline{\alpha} \leq \alpha_n \leq \overline{\alpha}$ and $0 \leq \beta_n \leq \beta$ consider the iterative scheme:

$$(\forall n \geq 1) \;\; x_{n+1} \in \operatorname{prox}_{\alpha_n f}(x_n - \alpha_n \nabla g(x_n) + \beta_n(x_n - x_{n-1})).$$

General assumption

Let $0 < \underline{\alpha} \leq \overline{\alpha}$ and $\beta > 0$ satisfy

$$1 > \overline{\alpha} L_{\nabla g} + 2\beta \frac{\overline{\alpha}}{\underline{\alpha}}.$$

Then

$$M_1 := \frac{1 - \overline{\alpha} L_{\nabla g}}{2\overline{\alpha}} - \frac{\beta}{2\underline{\alpha}} > M_2 := \frac{\beta}{2\underline{\alpha}}.$$

The minimization of the sum of two nonconvex functions

Consider the optimization problem

$$\min_{x \in \mathcal{H}} \{f(x) + g(x)\}. \tag{2}$$

- $\mathcal{H}$ is a finite-dimensional real Hilbert space;
- $f : \mathcal{H} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous and bounded from below;
- $g : \mathcal{H} \to \mathbb{R}$ is Fréchet differentiable and ∇g is $L_{\nabla g}$-Lipschitz continuous.

Inertial proximal-gradient algorithm

For $0 < \underline{\alpha} \leq \alpha_n \leq \overline{\alpha}$ and $0 \leq \beta_n \leq \beta$ consider the iterative scheme:

$$(\forall n \geq 1) \;\; x_{n+1} \in \operatorname{prox}_{\alpha_n f}(x_n - \alpha_n \nabla g(x_n) + \beta_n (x_n - x_{n-1})).$$

General assumption

Let $0 < \underline{\alpha} \leq \overline{\alpha}$ and $\beta > 0$ satisfy

$$1 > \overline{\alpha} L_{\nabla g} + 2\beta \frac{\overline{\alpha}}{\underline{\alpha}}.$$

Then

$$M_1 := \frac{1 - \overline{\alpha} L_{\nabla g}}{2\overline{\alpha}} - \frac{\beta}{2\underline{\alpha}} > M_2 := \frac{\beta}{2\underline{\alpha}}.$$

The minimization of the sum of two nonconvex functions

Consider the optimization problem

$$\min_{x \in \mathcal{H}} \{f(x) + g(x)\}. \tag{2}$$

- $\mathcal{H}$ is a finite-dimensional real Hilbert space;
- $f : \mathcal{H} \to \overline{\mathbb{R}}$ is proper, lower semicontinuous and bounded from below;
- $g : \mathcal{H} \to \mathbb{R}$ is Fréchet differentiable and ∇g is $L_{\nabla g}$-Lipschitz continuous.

Inertial proximal-gradient algorithm

For $0 < \underline{\alpha} \leq \alpha_n \leq \overline{\alpha}$ and $0 \leq \beta_n \leq \beta$ consider the iterative scheme:

$$(\forall n \geq 1) \;\; x_{n+1} \in \operatorname{prox}_{\alpha_n f}(x_n - \alpha_n \nabla g(x_n) + \beta_n(x_n - x_{n-1})).$$

General assumption

Let $0 < \underline{\alpha} \leq \overline{\alpha}$ and $\beta > 0$ satisfy

$$1 > \overline{\alpha} L_{\nabla g} + 2\beta \frac{\overline{\alpha}}{\underline{\alpha}}.$$

Then

$$M_1 := \frac{1 - \overline{\alpha} L_{\nabla g}}{2\overline{\alpha}} - \frac{\beta}{2\underline{\alpha}} > M_2 := \frac{\beta}{2\underline{\alpha}}.$$

Fundamental inequality

$$(f+g)(x_{n+1}) + M_2\|x_n - x_{n+1}\|^2 + (M_1 - M_2)\|x_n - x_{n+1}\|^2$$
$$\leq (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2 \quad \forall n \geq 1.$$

Consequences I

If $f+g$ is bounded from below, then

- $\sum_{n\geq 1} \|x_n - x_{n-1}\|^2 < +\infty$;
- the sequence $((f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2)_{n\geq 1}$ is monotonically decreasing and convergent;
- the sequence $((f+g)(x_n))_{n\geq 0}$ is convergent.

Consequences II

If $f+g$ is coercive, i.e.

$$\lim_{\|x\|\to+\infty} (f+g)(x) = +\infty,$$

then $(x_n)_{n\geq 0}$ has a convergent subsequence to a critical point of $f+g$. In fact, every cluster point of $(x_n)_{n\geq 0}$ is a critical point of $f+g$.

Fundamental inequality

$$
\begin{aligned}
& (f+g)(x_{n+1}) + M_2\|x_n - x_{n+1}\|^2 + (M_1 - M_2)\|x_n - x_{n+1}\|^2 \\
\le\ & (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2 \quad \forall n \ge 1.
\end{aligned}
$$

Consequences I

If $f+g$ is bounded from below, then

- $\sum_{n\ge 1}\|x_n - x_{n-1}\|^2 < +\infty$;
- the sequence $((f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2)_{n\ge 1}$ is monotonically decreasing and convergent;
- the sequence $((f+g)(x_n))_{n\ge 0}$ is convergent.

Consequences II

If $f+g$ is coercive, i.e.

$$\lim_{\|x\|\to+\infty} (f+g)(x) = +\infty,$$

then $(x_n)_{n\ge 0}$ has a convergent subsequence to a critical point of $f+g$. In fact, every cluster point of $(x_n)_{n\ge 0}$ is a critical point of $f+g$.

Fundamental inequality

$$(f+g)(x_{n+1}) + M_2\|x_n - x_{n+1}\|^2 + (M_1 - M_2)\|x_n - x_{n+1}\|^2$$
$$\leq (f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2 \quad \forall n \geq 1.$$

Consequences I

If $f+g$ is bounded from below, then

- $\sum_{n\geq 1} \|x_n - x_{n-1}\|^2 < +\infty$;
- the sequence $((f+g)(x_n) + M_2\|x_{n-1} - x_n\|^2)_{n\geq 1}$ is monotonically decreasing and convergent;
- the sequence $((f+g)(x_n))_{n\geq 0}$ is convergent.

Consequences II

If $f+g$ is coercive, i.e.

$$\lim_{\|x\|\to+\infty} (f+g)(x) = +\infty,$$

then $(x_n)_{n\geq 0}$ has a convergent subsequence to a critical point of $f+g$. In fact, every cluster point of $(x_n)_{n\geq 0}$ is a critical point of $f+g$.

The limiting subdifferential of a proper and lower semicontinuous function $h : \mathcal{H} \to \overline{\mathbb{R}}$

- the Fréchet (viscosity) subdifferential at $x \in \operatorname{dom} h$:

$$\hat{\partial} h(x) = \left\{ v \in \mathcal{H} : \liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0 \right\}$$

- the limiting (Mordukhovich) subdifferential at $x \in \operatorname{dom} h$:

$$\partial h(x) = \{ v \in \mathcal{H} : \exists x_n \to x, h(x_n) \to h(x) \text{ and } \exists v_n \in \hat{\partial} h(x_n), v_n \to v \text{ as } n \to +\infty \}$$

Properties of the limiting subdifferential

- if $x \in \mathcal{H}$ is a local minimizer of h, then $x \in \operatorname{crit}(h) := \{z \in \mathcal{H} : 0 \in \partial h(z)\}$;
- if h continuously differentiable around $x \in \mathcal{H}$, then $\partial h(x) = \{\nabla h(x)\}$;
- closedness criterion: $v_n \in \partial h(x_n)$ $\forall n \geq 0$, $(x_n, v_n) \to (x, v)$ and $h(x_n) \to h(x)$ as $n \to +\infty$, then $v \in \partial h(x)$. ;
- sum formula: if $k : \mathcal{H} \to \mathbb{R}$ is continuously differentiable, then $\partial(h + k)(x) = \partial h(x) + \nabla k(x)$ for all $x \in \mathcal{H}$;
- if h is convex, then $\partial h(x) = \{v \in \mathcal{H} : h(y) \geq h(x) + \langle v, y - x \rangle \ \forall y \in \mathcal{H}\}$ $\forall x \in \operatorname{dom} h$.

The limiting subdifferential of a proper and lower semicontinuous function $h : \mathcal{H} \to \overline{\mathbb{R}}$

- the Fréchet (viscosity) subdifferential at $x \in \operatorname{dom} h$:

$$\hat{\partial} h(x) = \left\{ v \in \mathcal{H} : \liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0 \right\}$$

- the limiting (Mordukhovich) subdifferential at $x \in \operatorname{dom} h$:

$$\partial h(x) = \{v \in \mathcal{H} : \exists x_n \to x, h(x_n) \to h(x) \text{ and } \exists v_n \in \hat{\partial} h(x_n), v_n \to v \text{ as } n \to +\infty\}$$

Properties of the limiting subdifferential

- if $x \in \mathcal{H}$ is a local minimizer of h, then $x \in \operatorname{crit}(h) := \{z \in \mathcal{H} : 0 \in \partial h(z)\}$;
- if h continuously differentiable around $x \in \mathcal{H}$, then $\partial h(x) = \{\nabla h(x)\}$;
- closedness criterion: $v_n \in \partial h(x_n)$ $\forall n \geq 0$, $(x_n, v_n) \to (x, v)$ and $h(x_n) \to h(x)$ as $n \to +\infty$, then $v \in \partial h(x)$. ;
- sum formula: if $k : \mathcal{H} \to \mathbb{R}$ is continuously differentiable, then $\partial(h + k)(x) = \partial h(x) + \nabla k(x)$ for all $x \in \mathcal{H}$;
- if h is convex, then $\partial h(x) = \{v \in \mathcal{H} : h(y) \geq h(x) + \langle v, y - x \rangle \ \forall y \in \mathcal{H}\}$ $\forall x \in \operatorname{dom} h$.

Recall that

$$\sum_{n\geq 1} \|x_n - x_{n-1}\|^2 < +\infty.$$

If one can ensure that

$$\sum_{n\geq 1} \|x_n - x_{n-1}\| < +\infty,$$

then $(x_n)_{n\geq 0}$ is convergent.

Recall that

$$\sum_{n\geq 1} \|x_n - x_{n-1}\|^2 < +\infty.$$

If one can ensure that

$$\sum_{n\geq 1} \|x_n - x_{n-1}\| < +\infty,$$

then $(x_n)_{n\geq 0}$ is convergent.

The Kurdyka-Łojasiewicz property

Let $h : \mathcal{H} \to \overline{\mathbb{R}}$ be proper and lower semicontinuous. The function h is said to have the Kurdyka-Łojasiewicz (KL) property at $x \in \operatorname{dom} \partial h = \{z \in \mathcal{H} : \partial h(z) \neq \emptyset\}$ if there exist

- $\eta \in (0, +\infty]$;
- a neighborhood U of x;
- a concave and continuous function $\varphi : [0, \eta) \to [0, +\infty)$ such that $\varphi(0) = 0$, φ is continuously differentiable on $(0, \eta)$ and $\varphi'(s) > 0$ for every $s \in (0, \eta)$

such that

$$\varphi'(h(y) - h(x))\operatorname{dist}(0, \partial h(y)) = \varphi'(h(y) - h(x)) \inf\{\|v\| : v \in \partial h(y)\} \geq 1 \qquad (3)$$

for every

$$y \in U \cap \{z \in \mathcal{H} : h(x) < h(z) < h(x) + \eta\}.$$

If h has the KL property at every point in $\operatorname{dom} \partial h$, then h is called KL function.

The KL property is satisfied at every noncritical point

If $x \in \operatorname{dom} h$ is a noncritical point of h, then there exists $c > 0$ such that

$$\|y - x\| + |h(y) - h(x)| \leq c \Longrightarrow \operatorname{dist}(0, \partial h(y)) \geq c.$$

Then (3) is fulfilled for $\varphi(s) = \frac{1}{c}s$ and every

$$y \in B(x, c/2) \cap \{z \in \mathcal{H} : h(x) - c/2 < h(z) < h(x) + c/2\}.$$

The Kurdyka-Łojasiewicz property

Let $h : \mathcal{H} \to \overline{\mathbb{R}}$ be proper and lower semicontinuous. The function h is said to have the Kurdyka-Łojasiewicz (KL) property at $x \in \operatorname{dom} \partial h = \{z \in \mathcal{H} : \partial h(z) \neq \emptyset\}$
if there exist

- $\eta \in (0, +\infty]$;
- a neighborhood U of x;
- a concave and continuous function $\varphi : [0, \eta) \to [0, +\infty)$ such that $\varphi(0) = 0$, φ is continuously differentiable on $(0, \eta)$ and $\varphi'(s) > 0$ for every $s \in (0, \eta)$

such that

$$\varphi'(h(y) - h(x))\operatorname{dist}(0, \partial h(y)) = \varphi'(h(y) - h(x)) \inf\{\|v\| : v \in \partial h(y)\} \geq 1 \qquad (3)$$

for every

$$y \in U \cap \{z \in \mathcal{H} : h(x) < h(z) < h(x) + \eta\}.$$

If h has the KL property at every point in $\operatorname{dom} \partial h$, then h is called KL function.

The KL property is satisfied at every noncritical point

If $x \in \operatorname{dom} h$ is a noncritical point of h, then there exists $c > 0$ such that

$$\|y - x\| + |h(y) - h(x)| \leq c \Longrightarrow \operatorname{dist}(0, \partial h(y)) \geq c.$$

Then (3) is fulfilled for $\varphi(s) = \frac{1}{c}s$ and every

$$y \in B(x, c/2) \cap \{z \in \mathcal{H} : h(x) - c/2 < h(z) < h(x) + c/2\}.$$

If h is continously differentiable around x, then (3) becomes

$$\varphi'(h(y) - h(x))\|\nabla h(y)\| = \|\nabla(\varphi \circ (h - h(x)))(y)\| \geq 1 \tag{4}$$

for every

$$y \in U \cap \{z \in \mathcal{H} : h(x) < h(z) < h(x) + \eta\}.$$

Łojasiewicz (1963)

If $h : \mathcal{H} \to \mathbb{R}$ is a real-analytic function and $x \in \mathcal{H}$ a critical point, then there exist $\theta \in [1/2, 1)$ and $C, \varepsilon > 0$ such that (Łojasiewicz property)

$$|h(y) - h(x)|^{\theta} \leq C\|\nabla h(y)\| \text{ for every } y \in \mathcal{H} \text{ with } \|y - x\| < \varepsilon.$$

Thus, (4) is fulfilled for $\varphi(s) = \frac{1}{1-\theta} C s^{1-\theta}$ and every

$$y \in B(x, \varepsilon) \cap \{z \in H : h(x) < h(z) < +\infty\}.$$

If h is continously differentiable around x, then (3) becomes

$$\varphi'(h(y) - h(x))\|\nabla h(y)\| = \|\nabla(\varphi \circ (h - h(x)))(y)\| \geq 1 \tag{4}$$

for every

$$y \in U \cap \{z \in \mathcal{H} : h(x) < h(z) < h(x) + \eta\}.$$

Łojasiewicz (1963)

If $h : \mathcal{H} \to \mathbb{R}$ is a real-analytic function and $x \in \mathcal{H}$ a critical point, then there exist $\theta \in [1/2, 1)$ and $C, \varepsilon > 0$ such that (Łojasiewicz property)

$$|h(y) - h(x)|^\theta \leq C\|\nabla h(y)\| \text{ for every } y \in \mathcal{H} \text{ with } \|y - x\| < \varepsilon.$$

Thus, (4) is fulfilled for $\varphi(s) = \frac{1}{1-\theta}Cs^{1-\theta}$ and every

$$y \in B(x, \varepsilon) \cap \{z \in H : h(x) < h(z) < +\infty\}.$$

If h is continously differentiable around x, then (3) becomes

$$\varphi'(h(y) - h(x))\|\nabla h(y)\| = \|\nabla(\varphi \circ (h - h(x)))(y)\| \geq 1 \tag{4}$$

for every

$$y \in U \cap \{z \in \mathcal{H} : h(x) < h(z) < h(x) + \eta\}.$$

Łojasiewicz (1963)

If $h : \mathcal{H} \to \mathbb{R}$ is a real-analytic function and $x \in \mathcal{H}$ a critical point, then there exist $\theta \in [1/2, 1)$ and $C, \varepsilon > 0$ such that (Łojasiewicz property)

$$|h(y) - h(x)|^\theta \leq C\|\nabla h(y)\| \text{ for every } y \in \mathcal{H} \text{ with } \|y - x\| < \varepsilon.$$

Thus, (4) is fulfilled for $\varphi(s) = \frac{1}{1-\theta}Cs^{1-\theta}$ and every

$$y \in B(x, \varepsilon) \cap \{z \in H : h(x) < h(z) < +\infty\}.$$

Examples of KL functions

- semi-algebraic functions, i.e., functions having as graph semi-algebraic sets, namely, sets of the form

$$\bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{u \in \mathbb{R}^m : g_{ij}(u) = 0 \text{ and } h_{ij}(u) < 0\},$$

  where $g_{ij}, h_{ij} : \mathbb{R}^m \to \mathbb{R}$ are polynomial functions;
- real polynomial functions;
- indicator functions of semi-algebraic sets;
- finite sums and product of semi-algebraic functions;
- compositions of semi-algebraic functions;
- $\|\cdot\|_p$ for $p \in \mathbb{Q}$ (including the case $p = 0$);
- convex functions fulfilling a certain growth condition;
- uniformly convex functions.

Theorem

If $f+g$ is coercive and $H : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}}$,

$$H(x,y) = (f+g)(x) + M_2\|x-y\|^2$$

is a KL function, then there exists $\overline{x} \in \text{crit}(f+g)$ such that $\lim_{n\to+\infty} x_n = \overline{x}$.

- Step 1 (decrease property):

$$H(x_{n+1},x_n) + (M_1 - M_2)\|x_{n+1}-x_n\|^2 \leq H(x_n, x_{n-1}) \; \forall n \geq 1.$$

- Step 2 (subgradient lower bound for the iterates gap):
For every $n \geq 1$ there exists

$$w_{n+1} = (y_{n+1} + 2M_2(x_{n+1}-x_n), 2M_2(x_n - x_{n+1})) \in \partial H(x_{n+1}, x_n),$$

where

$$y_{n+1} = \frac{x_n - x_{n+1}}{\alpha_n} + \nabla g(x_{n+1}) - \nabla g(x_n) + \frac{\beta_n}{\alpha_n}(x_n - x_{n-1}),$$

such that

$$\|w_{n+1}\| \leq N(\|x_{n+1}-x_n\| + \|x_n - x_{n-1}\|).$$

Here,

$$0 < N = \sup_{n\geq 1}\left\{\frac{1}{\alpha_n} + L_{\nabla g} + 4M_2, \frac{\beta_n}{\alpha_n}\right\} < +\infty.$$

Theorem

If $f+g$ is coercive and $H : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}}$,

$$H(x,y) = (f+g)(x) + M_2\|x-y\|^2$$

is a KL function, then there exists $\overline{x} \in \operatorname{crit}(f+g)$ such that $\lim_{n\to+\infty} x_n = \overline{x}$.

- Step 1 (decrease property):

$$H(x_{n+1},x_n) + (M_1 - M_2)\|x_{n+1}-x_n\|^2 \leq H(x_n,x_{n-1}) \ \forall n \geq 1.$$

- Step 2 (subgradient lower bound for the iterates gap):

For every $n \geq 1$ there exists

$$w_{n+1} = (y_{n+1} + 2M_2(x_{n+1}-x_n), 2M_2(x_n - x_{n+1})) \in \partial H(x_{n+1},x_n),$$

where

$$y_{n+1} = \frac{x_n - x_{n+1}}{\alpha_n} + \nabla g(x_{n+1}) - \nabla g(x_n) + \frac{\beta_n}{\alpha_n}(x_n - x_{n-1}),$$

such that

$$\|w_{n+1}\| \leq N(\|x_{n+1}-x_n\| + \|x_n - x_{n-1}\|).$$

Here,

$$0 < N = \sup_{n\geq 1}\left\{\frac{1}{\alpha_n} + L_{\nabla g} + 4M_2, \frac{\beta_n}{\alpha_n}\right\} < +\infty.$$

Theorem

If $f+g$ is coercive and $H : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}}$,

$$H(x,y) = (f+g)(x) + M_2\|x-y\|^2$$

is a KL function, then there exists $\overline{x} \in \operatorname{crit}(f+g)$ such that $\lim_{n\to+\infty} x_n = \overline{x}$.

- Step 1 (decrease property):

$$H(x_{n+1},x_n) + (M_1 - M_2)\|x_{n+1}-x_n\|^2 \leq H(x_n,x_{n-1}) \ \forall n \geq 1.$$

- Step 2 (subgradient lower bound for the iterates gap):

For every $n \geq 1$ there exists

$$w_{n+1} = (y_{n+1} + 2M_2(x_{n+1}-x_n), 2M_2(x_n - x_{n+1})) \in \partial H(x_{n+1},x_n),$$

where

$$y_{n+1} = \frac{x_n - x_{n+1}}{\alpha_n} + \nabla g(x_{n+1}) - \nabla g(x_n) + \frac{\beta_n}{\alpha_n}(x_n - x_{n-1}),$$

such that

$$\|w_{n+1}\| \leq N(\|x_{n+1}-x_n\| + \|x_n - x_{n-1}\|).$$

Here,

$$0 < N = \sup_{n\geq 1}\left\{\frac{1}{\alpha_n} + L_{\nabla g} + 4M_2, \frac{\beta_n}{\alpha_n}\right\} < +\infty.$$

▶ Step 3 (applying the KL property):

Let $x \in \text{crit}(f+g)$ be a cluster point of $(x_n)_{n\geq 0}$ and $H(x_n, x_{n-1}) > H(x,x)$ for every $n \geq 1$. Then there exists $\overline{n} \geq 1$ such that for every $n \geq \overline{n}$

$$\frac{(M_1 - M_2)\|x_{n+1} - x_n\|^2}{N(\|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\|)} \leq \frac{H(x_n, x_{n-1}) - H(x_{n+1}, x_n)}{\|w_{n+1}\|} \leq$$
$$\frac{(H(x_n, x_{n-1}) - H(x_{n+1}, x_n))}{\text{dist}((0,0), \partial H(x_n, x_{n-1}))} \leq$$
$$\varphi'(H(x_n, x_{n-1}) - H(x,x)) \cdot (H(x_n, x_{n-1}) - H(x_{n+1}, x_n)) \leq$$
$$\varphi(H(x_n, x_{n-1}) - H(x,x)) - \varphi(H(x_{n+1}, x_n) - H(x,x)).$$

By denoting for every $n \geq 1$

$$\varepsilon_n = \frac{N}{M_1 - M_2}\left(\varphi(H(x_n, x_{n-1}) - H(x,x)) - \varphi(H(x_{n+1}, x_n) - H(x,x))\right)$$
$$a_n = \|x_n - x_{n-1}\|,$$

it holds

$$a_{n+1} \leq \sqrt{\varepsilon_n(a_n + a_{n-1})} \leq \frac{1}{4}(a_n + a_{n-1}) + \varepsilon_n \ \forall n \geq \overline{n}.$$

Since $\sum_{n\geq 1} \varepsilon_n < +\infty$, it follows that

$$\sum_{n\geq 1} a_n = \sum_{n\geq 1} \|x_n - x_{n-1}\| < +\infty.$$

Hence $(x_n)_{n\geq 0}$ is a Cauchy sequence and, consequently, convergent.

▶ Step 3 (applying the KL property):

Let $x \in \operatorname{crit}(f+g)$ be a cluster point of $(x_n)_{n\geq 0}$ and $H(x_n, x_{n-1}) > H(x,x)$ for every $n \geq 1$. Then there exists $\overline{n} \geq 1$ such that for every $n \geq \overline{n}$

$$\frac{(M_1 - M_2)\|x_{n+1} - x_n\|^2}{N(\|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\|)} \leq \frac{H(x_n, x_{n-1}) - H(x_{n+1}, x_n)}{\|w_{n+1}\|} \leq \frac{(H(x_n, x_{n-1}) - H(x_{n+1}, x_n))}{\operatorname{dist}((0,0), \partial H(x_n, x_{n-1}))} \leq \varphi'(H(x_n, x_{n-1}) - H(x,x)) \cdot (H(x_n, x_{n-1}) - H(x_{n+1}, x_n)) \leq \varphi(H(x_n, x_{n-1}) - H(x,x)) - \varphi(H(x_{n+1}, x_n) - H(x,x)).$$

By denoting for every $n \geq 1$

$$\varepsilon_n = \frac{N}{M_1 - M_2}\left(\varphi(H(x_n, x_{n-1}) - H(x,x)) - \varphi(H(x_{n+1}, x_n) - H(x,x))\right)$$
$$a_n = \|x_n - x_{n-1}\|,$$

it holds

$$a_{n+1} \leq \sqrt{\varepsilon_n(a_n + a_{n-1})} \leq \frac{1}{4}(a_n + a_{n-1}) + \varepsilon_n \ \forall n \geq \overline{n}.$$

Since $\sum_{n\geq 1} \varepsilon_n < +\infty$, it follows that

$$\sum_{n\geq 1} a_n = \sum_{n\geq 1} \|x_n - x_{n-1}\| < +\infty.$$

Hence $(x_n)_{n\geq 0}$ is a Cauchy sequence and, consequently, convergent.

▶ Step 3 (applying the KL property):

Let $x \in \text{crit}(f+g)$ be a cluster point of $(x_n)_{n\geq 0}$ and $H(x_n, x_{n-1}) > H(x,x)$ for every $n \geq 1$. Then there exists $\overline{n} \geq 1$ such that for every $n \geq \overline{n}$

$$\frac{(M_1 - M_2)\|x_{n+1} - x_n\|^2}{N(\|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\|)} \leq \frac{H(x_n, x_{n-1}) - H(x_{n+1}, x_n)}{\|w_{n+1}\|} \leq$$
$$\frac{(H(x_n, x_{n-1}) - H(x_{n+1}, x_n))}{\text{dist}((0,0), \partial H(x_n, x_{n-1}))} \leq$$
$$\varphi'(H(x_n, x_{n-1}) - H(x,x)) \cdot (H(x_n, x_{n-1}) - H(x_{n+1}, x_n)) \leq$$
$$\varphi(H(x_n, x_{n-1}) - H(x,x)) - \varphi(H(x_{n+1}, x_n) - H(x,x)).$$

By denoting for every $n \geq 1$

$$\varepsilon_n = \frac{N}{M_1 - M_2}\left(\varphi(H(x_n, x_{n-1}) - H(x,x)) - \varphi(H(x_{n+1}, x_n) - H(x,x))\right)$$
$$a_n = \|x_n - x_{n-1}\|,$$

it holds

$$a_{n+1} \leq \sqrt{\varepsilon_n(a_n + a_{n-1})} \leq \frac{1}{4}(a_n + a_{n-1}) + \varepsilon_n \ \forall n \geq \overline{n}.$$

Since $\sum_{n\geq 1} \varepsilon_n < +\infty$, it follows that

$$\sum_{n\geq 1} a_n = \sum_{n\geq 1} \|x_n - x_{n-1}\| < +\infty.$$

Hence $(x_n)_{n\geq 0}$ is a Cauchy sequence and, consequently, convergent.

Corollary

If $f + g$ is coercive and semi-algebraic, then

(a) $\sum_{n \geq 0} \|x_{n+1} - x_n\| < +\infty$;

(b) there existsthen there exists $\overline{x} \in \mathrm{crit}(f + g)$ such that $\lim_{n \to +\infty} x_n = \overline{x}$.

Numerical experiment I

Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1+x_1^2) + x_2^2$$

- $f:\mathbb{R}^2\to\mathbb{R}, f(x_1,x_2) = |x_1| - |x_2|$ is nonconvex and continuous;
- For $\gamma > 0$ and $x = (x_1,x_2)\in\mathbb{R}^2$ it holds:

$$\operatorname{prox}_{\gamma f}(x) = \operatorname{prox}_{\gamma|\cdot|}(x_1) \times \operatorname{prox}_{\gamma(-|\cdot|)}(x_2),$$

where

$$\operatorname{prox}_{\gamma|\cdot|}(x_1) = x_1 - \operatorname{sgn}(x_1)\cdot\min\{|x_1|,\gamma\}$$

and

$$\operatorname{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2+\gamma, & \text{if } x_2>0 \\ x_2-\gamma, & \text{if } x_2<0 \\ \{-\gamma,\gamma\}, & \text{if } x_2=0; \end{cases}$$

- $g:\mathbb{R}^2\to\mathbb{R}, g(x_1,x_2) = x_1^2 - \log(1+x_1^2) + x_2^2$, is continuously differentiable, while ∇g is $\frac{9}{4}$-Lipschitz continuous;
- $f+g$ is coercive;
- $(0,\frac{1}{2})$ and $(0,-\frac{1}{2})$ are the only optimal solutions.

Numerical experiment I

Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1+x_1^2) + x_2^2$$

- $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1,x_2) = |x_1| - |x_2|$ is nonconvex and continuous;
- For $\gamma > 0$ and $x = (x_1, x_2) \in \mathbb{R}^2$ it holds:

$$\text{prox}_{\gamma f}(x) = \text{prox}_{\gamma|\cdot|}(x_1) \times \text{prox}_{\gamma(-|\cdot|)}(x_2),$$

where

$$\text{prox}_{\gamma|\cdot|}(x_1) = x_1 - \text{sgn}(x_1) \cdot \min\{|x_1|, \gamma\}$$

and

$$\text{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2 + \gamma, & \text{if } x_2 > 0 \\ x_2 - \gamma, & \text{if } x_2 < 0 \\ \{-\gamma, \gamma\}, & \text{if } x_2 = 0; \end{cases}$$

- $g : \mathbb{R}^2 \to \mathbb{R}, g(x_1,x_2) = x_1^2 - \log(1+x_1^2) + x_2^2$, is continuously differentiable, while ∇g is $\frac{9}{4}$-Lipschitz continuous;
- $f + g$ is coercive;
- $(0, \frac{1}{2})$ and $(0, -\frac{1}{2})$ are the only optimal solutions.

Numerical experiment I

Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1+x_1^2) + x_2^2$$

- $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1, x_2) = |x_1| - |x_2|$ is nonconvex and continuous;
- For $\gamma > 0$ and $x = (x_1, x_2) \in \mathbb{R}^2$ it holds:

$$\text{prox}_{\gamma f}(x) = \text{prox}_{\gamma|\cdot|}(x_1) \times \text{prox}_{\gamma(-|\cdot|)}(x_2),$$

where

$$\text{prox}_{\gamma|\cdot|}(x_1) = x_1 - \text{sgn}(x_1) \cdot \min\{|x_1|, \gamma\}$$

and

$$\text{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2 + \gamma, & \text{if } x_2 > 0 \\ x_2 - \gamma, & \text{if } x_2 < 0 \\ \{-\gamma, \gamma\}, & \text{if } x_2 = 0; \end{cases}$$

- $g : \mathbb{R}^2 \to \mathbb{R}, g(x_1, x_2) = x_1^2 - \log(1+x_1^2) + x_2^2$, is continuously differentiable, while ∇g is $\frac{9}{4}$-Lipschitz continuous;
- $f + g$ is coercive;
- $(0, \frac{1}{2})$ and $(0, -\frac{1}{2})$ are the only optimal solutions.

Numerical experiment I

Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1+x_1^2) + x_2^2$$

- $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1,x_2) = |x_1| - |x_2|$ is nonconvex and continuous;
- For $\gamma > 0$ and $x = (x_1,x_2) \in \mathbb{R}^2$ it holds:

$$\text{prox}_{\gamma f}(x) = \text{prox}_{\gamma|\cdot|}(x_1) \times \text{prox}_{\gamma(-|\cdot|)}(x_2),$$

where

$$\text{prox}_{\gamma|\cdot|}(x_1) = x_1 - \text{sgn}(x_1)\cdot\min\{|x_1|,\gamma\}$$

and

$$\text{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2+\gamma, & \text{if } x_2 > 0 \\ x_2-\gamma, & \text{if } x_2 < 0 \\ \{-\gamma,\gamma\}, & \text{if } x_2 = 0; \end{cases}$$

- $g : \mathbb{R}^2 \to \mathbb{R}, g(x_1,x_2) = x_1^2 - \log(1+x_1^2) + x_2^2$, is continuously differentiable, while ∇g is $\frac{9}{4}$-Lipschitz continuous;
- $f+g$ is coercive;
- $(0,\frac{1}{2})$ and $(0,-\frac{1}{2})$ are the only optimal solutions.

Numerical experiment I

Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1+x_1^2) + x_2^2$$

- $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1,x_2) = |x_1| - |x_2|$ is nonconvex and continuous;
- For $\gamma > 0$ and $x = (x_1,x_2) \in \mathbb{R}^2$ it holds:

$$\text{prox}_{\gamma f}(x) = \text{prox}_{\gamma|\cdot|}(x_1) \times \text{prox}_{\gamma(-|\cdot|)}(x_2),$$

where

$$\text{prox}_{\gamma|\cdot|}(x_1) = x_1 - \text{sgn}(x_1)\cdot\min\{|x_1|,\gamma\}$$

and

$$\text{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2+\gamma, & \text{if } x_2 > 0 \\ x_2-\gamma, & \text{if } x_2 < 0 \\ \{-\gamma,\gamma\}, & \text{if } x_2 = 0; \end{cases}$$

- $g : \mathbb{R}^2 \to \mathbb{R}, g(x_1,x_2) = x_1^2 - \log(1+x_1^2) + x_2^2$, is continuously differentiable, while ∇g is $\frac{9}{4}$-Lipschitz continuous;
- $f + g$ is coercive;
- $(0, \frac{1}{2})$ and $(0, -\frac{1}{2})$ are the only optimal solutions.

Numerical experiment I

Consider the optimization problem

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} |x_1| - |x_2| + x_1^2 - \log(1+x_1^2) + x_2^2$$

- $f : \mathbb{R}^2 \to \mathbb{R}, f(x_1,x_2) = |x_1| - |x_2|$ is nonconvex and continuous;
- For $\gamma > 0$ and $x = (x_1,x_2) \in \mathbb{R}^2$ it holds:

$$\text{prox}_{\gamma f}(x) = \text{prox}_{\gamma|\cdot|}(x_1) \times \text{prox}_{\gamma(-|\cdot|)}(x_2),$$

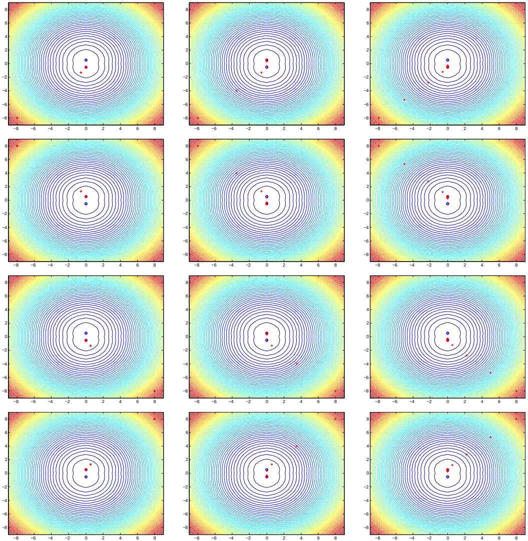
where

$$\text{prox}_{\gamma|\cdot|}(x_1) = x_1 - \text{sgn}(x_1)\cdot \min\{|x_1|,\gamma\}$$

and

$$\text{prox}_{\gamma(-|\cdot|)}(x_2) = \begin{cases} x_2+\gamma, & \text{if } x_2 > 0 \\ x_2-\gamma, & \text{if } x_2 < 0 \\ \{-\gamma,\gamma\}, & \text{if } x_2 = 0; \end{cases}$$

- $g : \mathbb{R}^2 \to \mathbb{R}, g(x_1,x_2) = x_1^2 - \log(1+x_1^2) + x_2^2$, is continuously differentiable, while ∇g is $\frac{9}{4}$-Lipschitz continuous;
- $f+g$ is coercive;
- $(0,\frac{1}{2})$ and $(0,-\frac{1}{2})$ are the only optimal solutions.



Iterations: 100; Starting points: $(-8,-8), (-8,8), (8,-8)$ and $(8,8)$, respectively;
First column: the non-inertial version ($\beta_n = \beta = 0\ \forall n \geq 1$); Second column: $\beta_n = \beta = 0.199\ \forall n \geq 1$; Third column: $\beta_n = \beta = 0.299\ \forall n \geq 1$.

Numerical experiment II (restoration of noisy blurred images)

For a given matrix $A \in \mathbb{R}^{m\times m}$ describing a blur operator and a given vector $b \in \mathbb{R}^m$ representing the blurred and noisy image, the task is to estimate the unknown original image $\overline{x} \in \mathbb{R}^m$ fulfilling

$$A\overline{x} = b.$$

We solve the regularized nonconvex minimization problem

$$\inf_{x\in\mathbb{R}^m} \left\{ \sum_{k=1}^{M} \sum_{l=1}^{N} \varphi\big((Ax-b)_{kl}\big) + \lambda \|Wx\|_0 \right\},$$

where

- $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(t) = \log(1+t^2)$, is derived form the Student t distribution;
- $\lambda > 0$ is a regularization parameter;
- $W : \mathbb{R}^m \to \mathbb{R}^m$ is a discrete Haar wavelet transform with four levels;
- $\|y\|_0 = \sum_{i=1}^{m} |\operatorname{sgn}(y_i)|$, for $y = (y_1, ..., y_m)$.

In this context, $x \in \mathbb{R}^m$ represents the vectorized image $X \in \mathbb{R}^{M\times N}$, where $m = M \cdot N$.

Numerical experiment II (restoration of noisy blurred images)

For a given matrix $A \in \mathbb{R}^{m \times m}$ describing a blur operator and a given vector $b \in \mathbb{R}^m$ representing the blurred and noisy image, the task is to estimate the unknown original image $\overline{x} \in \mathbb{R}^m$ fulfilling

$$A\overline{x} = b.$$

We solve the regularized nonconvex minimization problem

$$\inf_{x \in \mathbb{R}^m} \left\{ \sum_{k=1}^{M} \sum_{l=1}^{N} \varphi\Big((Ax - b)_{kl}\Big) + \lambda \|Wx\|_0 \right\},$$

where

- $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(t) = \log(1 + t^2)$, is derived form the Student t distribution;
- $\lambda > 0$ is a regularization parameter;
- $W : \mathbb{R}^m \to \mathbb{R}^m$ is a discrete Haar wavelet transform with four levels;
- $\|y\|_0 = \sum_{i=1}^m |\operatorname{sgn}(y_i)|$, for $y = (y_1, ..., y_m)$.

In this context, $x \in \mathbb{R}^m$ represents the vectorized image $X \in \mathbb{R}^{M \times N}$, where $m = M \cdot N$.

- Since $WW^T = W^T W = I_m$,

$$\mathrm{prox}_{\gamma\|W(\cdot)\|_0}(x) = W^T \, \mathrm{prox}_{\lambda\gamma\|\cdot\|_0}(Wx) \ \forall x \in \mathbb{R}^m \ \forall \gamma > 0,$$

where for all $u = (u_1, ..., u_m)$ we have

$$\mathrm{prox}_{\lambda\gamma\|\cdot\|_0}(u) = (\mathrm{prox}_{\lambda\gamma|\cdot|_0}(u_1), ..., \mathrm{prox}_{\lambda\gamma|\cdot|_0}(u_m))$$

and for all $t \in \mathbb{R}$

$$\mathrm{prox}_{\lambda\gamma|\cdot|_0}(t) = \begin{cases} t, & \text{if } |t| > \sqrt{2\lambda\gamma}, \\ \{0, t\}, & \text{if } |t| = \sqrt{2\lambda\gamma}, \\ 0, & \text{otherwise.} \end{cases}$$

- For the experiments we used the 256×256 boat test image which we first blurred by using a Gaussian blur operator of size 9×9 and standard deviation 4 and to which we afterward added a zero-mean white Gaussian noise with standard deviation 10^{-6}.
- We took as Lipschitz constant of the gradient of the smooth misfit function $L_{\nabla g} = 2$.

- Since $WW^T = W^TW = I_m$,

$$\operatorname{prox}_{\gamma\|W(\cdot)\|_0}(x) = W^T \operatorname{prox}_{\lambda\gamma\|\cdot\|_0}(Wx) \ \forall x \in \mathbb{R}^m \ \forall \gamma > 0,$$

where for all $u = (u_1, ..., u_m)$ we have

$$\operatorname{prox}_{\lambda\gamma\|\cdot\|_0}(u) = (\operatorname{prox}_{\lambda\gamma|\cdot|_0}(u_1), ..., \operatorname{prox}_{\lambda\gamma|\cdot|_0}(u_m))$$

and for all $t \in \mathbb{R}$

$$\operatorname{prox}_{\lambda\gamma|\cdot|_0}(t) = \begin{cases} t, & \text{if } |t| > \sqrt{2\lambda\gamma}, \\ \{0, t\}, & \text{if } |t| = \sqrt{2\lambda\gamma}, \\ 0, & \text{otherwise.} \end{cases}$$

- For the experiments we used the 256×256 boat test image which we first blurred by using a Gaussian blur operator of size 9×9 and standard deviation 4 and to which we afterward added a zero-mean white Gaussian noise with standard deviation 10^{-6}.

- We took as Lipschitz constant of the gradient of the smooth misfit function $L_{\nabla g} = 2$.

- Since $WW^T = W^TW = I_m$,

$$\mathrm{prox}_{\gamma\|W(\cdot)\|_0}(x) = W^T\,\mathrm{prox}_{\lambda\gamma\|\cdot\|_0}(Wx)\ \forall x \in \mathbb{R}^m\ \forall \gamma > 0,$$

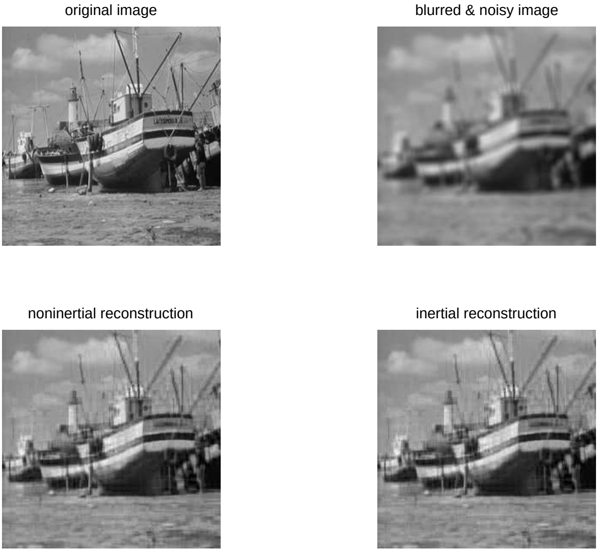
where for all $u = (u_1, ..., u_m)$ we have

$$\mathrm{prox}_{\lambda\gamma\|\cdot\|_0}(u) = (\mathrm{prox}_{\lambda\gamma|\cdot|_0}(u_1), ..., \mathrm{prox}_{\lambda\gamma|\cdot|_0}(u_m))$$

and for all $t \in \mathbb{R}$

$$\mathrm{prox}_{\lambda\gamma|\cdot|_0}(t) = \begin{cases} t, & \text{if } |t| > \sqrt{2\lambda\gamma}, \\ \{0, t\}, & \text{if } |t| = \sqrt{2\lambda\gamma}, \\ 0, & \text{otherwise.} \end{cases}$$

- For the experiments we used the 256×256 boat test image which we first blurred by using a Gaussian blur operator of size 9×9 and standard deviation 4 and to which we afterward added a zero-mean white Gaussian noise with standard deviation 10^{-6}.
- We took as Lipschitz constant of the gradient of the smooth misfit function $L_{\nabla g} = 2$.



The first row shows the original 256×256 boat test image and the blurred and noisy one and the second row the reconstructed images after 300 iterations.

D.C. programming

Consider the optimization problem

$$\min \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} \tag{5}$$

- $\mathcal{G}$ and $\mathcal{H}$ are finite-dimensional real Hilbert spaces;
- $g : \mathcal{H} \to \overline{\mathbb{R}}$ and $h : \mathcal{G} \to \overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions;
- $K : \mathcal{H} \to \mathcal{G}$ is a linear mapping;
- $\varphi : \mathcal{H} \to \mathbb{R}$ is convex, Fréchet differentiable and $\nabla\varphi$ is $L_{\nabla\varphi}$-Lipschitz continuous.

Toland dual problem

$$\min \{h^*(y) - (g + \varphi)^*(K^*y) \mid y \in \mathcal{G}\}. \tag{6}$$

Primal-dual formulation

$$\min \{\Phi(x, y) \mid x \in \mathcal{H}, y \in \mathcal{G}\}, \tag{7}$$

$$\Phi : \mathcal{H} \times \mathcal{G} \to \overline{\mathbb{R}}, \quad \Phi(x, y) := g(x) + \varphi(x) + h^*(y) - \langle y, Kx \rangle.$$

Φ is proper and lower semicontinuous.

D.C. programming

Consider the optimization problem

$$\min\{g(x)+\varphi(x)-h(Kx) \mid x\in\mathcal{H}\} \tag{5}$$

- $\mathcal{G}$ and $\mathcal{H}$ are finite-dimensional real Hilbert spaces;
- $g:\mathcal{H}\to\overline{\mathbb{R}}$ and $h:\mathcal{G}\to\overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions;
- $K:\mathcal{H}\to\mathcal{G}$ is a linear mapping;
- $\varphi:\mathcal{H}\to\mathbb{R}$ is convex, Fréchet differentiable and $\nabla\varphi$ is $L_{\nabla\varphi}$-Lipschitz continuous.

Toland dual problem

$$\min\{h^*(y)-(g+\varphi)^*(K^*y) \mid y\in\mathcal{G}\}. \tag{6}$$

Primal-dual formulation

$$\min\{\Phi(x,y) \mid x\in\mathcal{H}, y\in\mathcal{G}\}, \tag{7}$$

$$\Phi:\mathcal{H}\times\mathcal{G}\to\overline{\mathbb{R}},\quad \Phi(x,y):=g(x)+\varphi(x)+h^*(y)-\langle y,Kx\rangle.$$

Φ is proper and lower semicontinuous.

D.C. programming

Consider the optimization problem

$$\min \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} \tag{5}$$

- $\mathcal{G}$ and $\mathcal{H}$ are finite-dimensional real Hilbert spaces;
- $g : \mathcal{H} \to \overline{\mathbb{R}}$ and $h : \mathcal{G} \to \overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions;
- $K : \mathcal{H} \to \mathcal{G}$ is a linear mapping;
- $\varphi : \mathcal{H} \to \mathbb{R}$ is convex, Fréchet differentiable and $\nabla\varphi$ is $L_{\nabla\varphi}$-Lipschitz continuous.

Toland dual problem

$$\min \{h^*(y) - (g + \varphi)^*(K^* y) \mid y \in \mathcal{G}\}. \tag{6}$$

Primal-dual formulation

$$\min \{\Phi(x, y) \mid x \in \mathcal{H}, y \in \mathcal{G}\}, \tag{7}$$

$$\Phi : \mathcal{H} \times \mathcal{G} \to \overline{\mathbb{R}}, \quad \Phi(x, y) := g(x) + \varphi(x) + h^*(y) - \langle y, Kx\rangle.$$

Φ is proper and lower semicontinuous.

Proposition

1. The optimal values of (5), (6) and (7) are equal.
2. For all $x \in \mathcal{H}$ and $y \in \mathcal{G}$,

$$\Phi(x, y) \geq g(x) + \varphi(x) - h(Kx) \qquad \text{and}$$
$$\Phi(x, y) \geq h^*(y) - (g + \varphi)^*(K^* y).$$

3. Let $\bar{x} \in \mathcal{H}$ be an optimal solution of (5). Then

$$K^* \partial h(K\bar{x}) \subseteq \partial(h \circ K)(\bar{x}) \subseteq \partial g(\bar{x}) + \nabla \varphi(\bar{x}).$$

4. Let $\bar{y} \in \mathcal{G}$ be an optimal solution of (6). Then

$$K \partial (g + \varphi)^*(K\bar{y}) \subseteq \partial((g + \varphi)^* \circ K^*)(\bar{y}) \subseteq \partial h^*(\bar{y}).$$

5. Let $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ be an optimal solution of (7). Then $\bar{x}$ is an optimal solution of (5), and $\bar{y}$ is an optimal solution of (6). Furthermore,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}), \tag{8}$$
$$K\bar{x} \in \partial h^*(\bar{y}). \tag{9}$$

Proposition

1. The optimal values of (5), (6) and (7) are equal.
2. For all $x \in \mathcal{H}$ and $y \in \mathcal{G}$,

$$\Phi(x,y) \geq g(x) + \varphi(x) - h(Kx) \qquad \text{and}$$
$$\Phi(x,y) \geq h^*(y) - (g+\varphi)^*(K^*y).$$

3. Let $\bar{x} \in \mathcal{H}$ be an optimal solution of (5). Then

$$K^*\partial h(K\bar{x}) \subseteq \partial(h \circ K)(\bar{x}) \subseteq \partial g(\bar{x}) + \nabla\varphi(\bar{x}).$$

4. Let $\bar{y} \in \mathcal{G}$ be an optimal solution of (6). Then

$$K\partial(g+\varphi)^*(K\bar{y}) \subseteq \partial((g+\varphi)^* \circ K^*)(\bar{y}) \subseteq \partial h^*(\bar{y}).$$

5. Let $(\bar{x},\bar{y}) \in \mathcal{H} \times \mathcal{G}$ be an optimal solution of (7). Then $\bar{x}$ is an optimal solution of (5), and $\bar{y}$ is an optimal solution of (6). Furthermore,

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}), \tag{8}$$
$$K\bar{x} \in \partial h^*(\bar{y}). \tag{9}$$

Proposition

1. The optimal values of (5), (6) and (7) are equal.
2. For all $x \in \mathcal{H}$ and $y \in \mathcal{G}$,

$$\Phi(x,y) \geq g(x) + \varphi(x) - h(Kx) \qquad \text{and}$$
$$\Phi(x,y) \geq h^*(y) - (g+\varphi)^*(K^*y).$$

3. Let $\bar{x} \in \mathcal{H}$ be an optimal solution of (5). Then

$$K^*\partial h(K\bar{x}) \subseteq \partial(h \circ K)(\bar{x}) \subseteq \partial g(\bar{x}) + \nabla\varphi(\bar{x}).$$

4. Let $\bar{y} \in \mathcal{G}$ be an optimal solution of (6). Then

$$K\partial(g+\varphi)^*(K\bar{y}) \subseteq \partial((g+\varphi)^* \circ K^*)(\bar{y}) \subseteq \partial h^*(\bar{y}).$$

5. Let $(\bar{x},\bar{y}) \in \mathcal{H} \times \mathcal{G}$ be an optimal solution of (7). Then $\bar{x}$ is an optimal solution of (5), and $\bar{y}$ is an optimal solution of (6). Furthermore,

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}), \tag{8}$$
$$K\bar{x} \in \partial h^*(\bar{y}). \tag{9}$$

Proposition

1. The optimal values of (5), (6) and (7) are equal.
2. For all $x \in \mathcal{H}$ and $y \in \mathcal{G}$,

$$\Phi(x,y) \geq g(x) + \varphi(x) - h(Kx) \quad \text{and}$$
$$\Phi(x,y) \geq h^*(y) - (g+\varphi)^*(K^*y).$$

3. Let $\bar{x} \in \mathcal{H}$ be an optimal solution of (5). Then

$$K^*\partial h(K\bar{x}) \subseteq \partial(h \circ K)(\bar{x}) \subseteq \partial g(\bar{x}) + \nabla\varphi(\bar{x}).$$

4. Let $\bar{y} \in \mathcal{G}$ be an optimal solution of (6). Then

$$K\partial(g+\varphi)^*(K\bar{y}) \subseteq \partial((g+\varphi)^* \circ K^*)(\bar{y}) \subseteq \partial h^*(\bar{y}).$$

5. Let $(\bar{x},\bar{y}) \in \mathcal{H} \times \mathcal{G}$ be an optimal solution of (7). Then $\bar{x}$ is an optimal solution of (5), and $\bar{y}$ is an optimal solution of (6). Furthermore,

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}), \tag{8}$$
$$K\bar{x} \in \partial h^*(\bar{y}). \tag{9}$$

Proposition

1. The optimal values of (5), (6) and (7) are equal.
2. For all $x \in \mathcal{H}$ and $y \in \mathcal{G}$,

$$\Phi(x, y) \geq g(x) + \varphi(x) - h(Kx) \qquad \text{and}$$
$$\Phi(x, y) \geq h^*(y) - (g + \varphi)^*(K^* y).$$

3. Let $\bar{x} \in \mathcal{H}$ be an optimal solution of (5). Then

$$K^* \partial h(K\bar{x}) \subseteq \partial(h \circ K)(\bar{x}) \subseteq \partial g(\bar{x}) + \nabla\varphi(\bar{x}).$$

4. Let $\bar{y} \in \mathcal{G}$ be an optimal solution of (6). Then

$$K\partial(g + \varphi)^*(K\bar{y}) \subseteq \partial((g + \varphi)^* \circ K^*)(\bar{y}) \subseteq \partial h^*(\bar{y}).$$

5. Let $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ be an optimal solution of (7). Then $\bar{x}$ is an optimal solution of (5), and $\bar{y}$ is an optimal solution of (6). Furthermore,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}), \tag{8}$$
$$K\bar{x} \in \partial h^*(\bar{y}). \tag{9}$$

Critical points of Φ

We say that $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of the objective function Φ of (7) if

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}),$$
$$K\bar{x} \in \partial h^*(\bar{y}).$$

We denote by $\mathrm{crit}\Phi$ the set of critical points of the function Φ.

Critical points of $g + \varphi - h \circ K$

$$\mathrm{crit}(g + \varphi - h \circ K) := \{x \in \mathcal{H} : K^*\partial h(Kx) \cap (\partial g(x) + \nabla\varphi(x)) \neq \emptyset\}$$

Critical points of $h^* - (g + \varphi)^* \circ K^*$

$$\mathrm{crit}(h^* - (g + \varphi)^* \circ K^*) := \{y \in \mathcal{G} : K\partial(g + \varphi)^*(K^*y) \cap \partial h^*(y) \neq \emptyset\}$$

If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of Φ, then

$$K^*\bar{y} \in K^*\partial h(K\bar{x}) \cap (\partial g(\bar{x}) + \nabla\varphi(\bar{x})),$$
$$K\bar{x} \in K\partial(g + \varphi)^*(K^*\bar{y}) \cap \partial h^*(\bar{y}).$$

Thus, $\bar{x}$ is a critical point of $g + \varphi - h \circ K$ and $\bar{y}$ is a critical point of $h^* - (g + \varphi)^* \circ K^*$

Critical points of Φ

We say that $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of the objective function Φ of (7) if

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}),$$
$$K\bar{x} \in \partial h^*(\bar{y}).$$

We denote by $\text{crit}\Phi$ the set of critical points of the function Φ.

Critical points of $g + \varphi - h \circ K$

$$\text{crit}(g + \varphi - h \circ K) := \{x \in \mathcal{H} : K^*\partial h(Kx) \cap (\partial g(x) + \nabla\varphi(x)) \neq \emptyset\}$$

Critical points of $h^* - (g + \varphi)^* \circ K^*$

$$\text{crit}(h^* - (g + \varphi)^* \circ K^*) := \{y \in \mathcal{G} : K\partial(g + \varphi)^*(K^*y) \cap \partial h^*(y) \neq \emptyset\}$$

If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of Φ, then

$$K^*\bar{y} \in K^*\partial h(K\bar{x}) \cap (\partial g(\bar{x}) + \nabla\varphi(\bar{x})),$$
$$K\bar{x} \in K\partial(g + \varphi)^*(K^*\bar{y}) \cap \partial h^*(\bar{y}).$$

Thus, $\bar{x}$ is a critical point of $g + \varphi - h \circ K$ and $\bar{y}$ is a critical point of $h^* - (g + \varphi)^* \circ K^*$

Critical points of Φ

We say that $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of the objective function Φ of (7) if

$$K^*\bar{y} \in \partial g(\bar{x}) + \nabla\varphi(\bar{x}),$$
$$K\bar{x} \in \partial h^*(\bar{y}).$$

We denote by $\text{crit}\Phi$ the set of critical points of the function Φ.

Critical points of $g + \varphi - h \circ K$

$$\text{crit}(g + \varphi - h \circ K) := \{x \in \mathcal{H} : K^*\partial h(Kx) \cap (\partial g(x) + \nabla\varphi(x)) \neq \emptyset\}$$

Critical points of $h^* - (g + \varphi)^* \circ K^*$

$$\text{crit}(h^* - (g + \varphi)^* \circ K^*) := \{y \in \mathcal{G} : K\partial(g + \varphi)^*(K^*y) \cap \partial h^*(y) \neq \emptyset\}$$

If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of Φ, then

$$K^*\bar{y} \in K^*\partial h(K\bar{x}) \cap (\partial g(\bar{x}) + \nabla\varphi(\bar{x})),$$
$$K\bar{x} \in K\partial(g + \varphi)^*(K^*\bar{y}) \cap \partial h^*(\bar{y}).$$

Thus, $\bar{x}$ is a critical point of $g + \varphi - h \circ K$ and $\bar{y}$ is a critical point of $h^* - (g + \varphi)^* \circ K^*$

Critical points of Φ

We say that $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of the objective function Φ of (7) if

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}),$$
$$K\bar{x} \in \partial h^*(\bar{y}).$$

We denote by $\mathrm{crit}\Phi$ the set of critical points of the function Φ.

Critical points of $g + \varphi - h \circ K$

$$\mathrm{crit}(g + \varphi - h \circ K) := \{x \in \mathcal{H} : K^* \partial h(Kx) \cap (\partial g(x) + \nabla \varphi(x)) \neq \emptyset\}$$

Critical points of $h^* - (g + \varphi)^* \circ K^*$

$$\mathrm{crit}(h^* - (g + \varphi)^* \circ K^*) := \{y \in \mathcal{G} : K \partial (g + \varphi)^*(K^* y) \cap \partial h^*(y) \neq \emptyset\}$$

If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is a critical point of Φ, then

$$K^* \bar{y} \in K^* \partial h(K\bar{x}) \cap (\partial g(\bar{x}) + \nabla \varphi(\bar{x})),$$
$$K\bar{x} \in K \partial (g + \varphi)^*(K^* \bar{y}) \cap \partial h^*(\bar{y}).$$

Thus, $\bar{x}$ is a critical point of $g + \varphi - h \circ K$ and $\bar{y}$ is a critical point of $h^* - (g + \varphi)^* \circ K^*$

A double-proximal gradient algorithm

Let $(x_0, y_0) \in \mathcal{H} \times \mathcal{G}$, and let $(\gamma_n)_{n\geq 0}$ and $(\mu_n)_{n\geq 0}$ be sequences of positive numbers. For all $n \geq 0$ set

$$x_{n+1} := \operatorname{prox}_{\gamma_n g}\left(x_n + \gamma_n K^* y_n - \gamma_n \nabla\varphi(x_n)\right),$$
$$y_{n+1} := \operatorname{prox}_{\mu_n h^*}\left(y_n + \mu_n K x_{n+1}\right).$$

Important inequalities

For all $n \geq 0$

$$\Phi(x_{n+1}, y_n) - \Phi(x_n, y_n) \leq \left(\frac{L_{\nabla\varphi}}{2} - \frac{1}{\gamma_n}\right) \|x_n - x_{n+1}\|^2,$$
$$\Phi(x_{n+1}, y_{n+1}) - \Phi(x_{n+1}, y_n) \leq -\frac{1}{\mu_n} \|y_n - y_{n+1}\|^2.$$

A double-proximal gradient algorithm

Let $(x_0, y_0) \in \mathcal{H} \times \mathcal{G}$, and let $(\gamma_n)_{n\geq 0}$ and $(\mu_n)_{n\geq 0}$ be sequences of positive numbers. For all $n \geq 0$ set

$$\begin{aligned} x_{n+1} &:= \operatorname{prox}_{\gamma_n g}\left(x_n + \gamma_n K^* y_n - \gamma_n \nabla\varphi(x_n)\right), \\ y_{n+1} &:= \operatorname{prox}_{\mu_n h^*}\left(y_n + \mu_n K x_{n+1}\right). \end{aligned}$$

Important inequalities

For all $n \geq 0$

$$\Phi(x_{n+1}, y_n) - \Phi(x_n, y_n) \leq \left(\frac{L_{\nabla\varphi}}{2} - \frac{1}{\gamma_n}\right) \|x_n - x_{n+1}\|^2,$$

$$\Phi(x_{n+1}, y_{n+1}) - \Phi(x_{n+1}, y_n) \leq -\frac{1}{\mu_n} \|y_n - y_{n+1}\|^2.$$

Proposition

- For all $n \geq 0$, if $0 < \gamma_n \leq \frac{2}{L_{\nabla\varphi}}$, then

$$\Phi(x_{n+1}, y_{n+1}) \leq \Phi(x_{n+1}, y_n) \leq \Phi(x_n, y_n).$$

- Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and

$$0 < \inf_{n \geq 0} \gamma_n \leq \sup_{n \geq 0} \gamma_n < \frac{2}{L_{\nabla\varphi}} \quad \text{and} \quad 0 < \inf_{n \geq 0} \mu_n \leq \sup_{n \geq 0} \mu_n < +\infty. \tag{10}$$

Then,

$$\sum_{n \geq 0} \|x_n - x_{n+1}\|^2 < +\infty \quad \text{and} \quad \sum_{n \geq 0} \|y_n - y_{n+1}\|^2 < +\infty.$$

Proposition

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (10) be satisfied. If $(x_n)_{n \geq 0}$ and $(y_n)_{n \geq 0}$ are bounded, then

1. every cluster point of $(x_n)_{n \geq 0}$ is a critical point of (5),
2. every cluster point of $(y_n)_{n \geq 0}$ is a critical point of (6),
3. every cluster point of $(x_n, y_n)_{n \geq 0}$ is a critical point of (7).

Proposition

- For all $n \geq 0$, if $0 < \gamma_n \leq \frac{2}{L_{\nabla\varphi}}$, then

$$\Phi(x_{n+1}, y_{n+1}) \leq \Phi(x_{n+1}, y_n) \leq \Phi(x_n, y_n).$$

- Let $\inf\{g(x) + \varphi(x) - h(Kx) \,|\, x \in \mathcal{H}\} > -\infty$ and

$$0 < \inf_{n\geq 0} \gamma_n \leq \sup_{n\geq 0} \gamma_n < \frac{2}{L_{\nabla\varphi}} \quad \text{and} \quad 0 < \inf_{n\geq 0} \mu_n \leq \sup_{n\geq 0} \mu_n < +\infty. \tag{10}$$

Then,

$$\sum_{n\geq 0} \|x_n - x_{n+1}\|^2 < +\infty \quad \text{and} \quad \sum_{n\geq 0} \|y_n - y_{n+1}\|^2 < +\infty.$$

Proposition

Let $\inf\{g(x) + \varphi(x) - h(Kx) \,|\, x \in \mathcal{H}\} > -\infty$ and (10) be satisfied. If $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ are bounded, then

1. every cluster point of $(x_n)_{n\geq 0}$ is a critical point of (5),
2. every cluster point of $(y_n)_{n\geq 0}$ is a critical point of (6),
3. every cluster point of $(x_n, y_n)_{n\geq 0}$ is a critical point of (7).

Proposition

- For all $n \geq 0$, if $0 < \gamma_n \leq \frac{2}{L_{\nabla\varphi}}$, then

$$\Phi(x_{n+1}, y_{n+1}) \leq \Phi(x_{n+1}, y_n) \leq \Phi(x_n, y_n).$$

- Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and

$$0 < \inf_{n \geq 0} \gamma_n \leq \sup_{n \geq 0} \gamma_n < \frac{2}{L_{\nabla\varphi}} \quad \text{and} \quad 0 < \inf_{n \geq 0} \mu_n \leq \sup_{n \geq 0} \mu_n < +\infty. \tag{10}$$

Then,

$$\sum_{n \geq 0} \|x_n - x_{n+1}\|^2 < +\infty \quad \text{and} \quad \sum_{n \geq 0} \|y_n - y_{n+1}\|^2 < +\infty.$$

Proposition

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (10) be satisfied. If $(x_n)_{n \geq 0}$ and $(y_n)_{n \geq 0}$ are bounded, then

1. every cluster point of $(x_n)_{n \geq 0}$ is a critical point of (5),
2. every cluster point of $(y_n)_{n \geq 0}$ is a critical point of (6)
3. every cluster point of $(x_n, y_n)_{n \geq 0}$ is a critical point of (7).

Proposition

Let (10) be satisfied. For any $n \geq 0$, the following statements are equivalent:

1. (x_n, y_n) is a critical point of Φ;
2. $(x_{n+1}, y_{n+1}) = (x_n, y_n)$;
3. $\Phi(x_{n+1}, y_{n+1}) = \Phi(x_n, y_n)$.

Let

$$\omega(x_0, y_0) := \{\text{set of cluster points of } (x_n, y_n)_{n \geq 0} \text{ when } x_0 \text{ and } y_0 \text{ are the initial points}\}.$$

Theorem (Convergence result)

Let (10) be satisfied and assume that the sequence $(x_n, y_n)_{n \geq 0}$ is bounded. Then the following assertions hold:

1. $\emptyset \neq \omega(x_0, y_0) \subseteq \text{crit}\Phi \subseteq \text{crit}(g + \varphi - h \circ K) \times \text{crit}(h^* - (g + \varphi)^* \circ K^*)$,
2. $\lim_{n\to\infty} \text{dist}((x_n, y_n), \omega(x_0, y_0)) = 0$,
3. if the common optimal value of the problems (5), (6) and (7) is finite, then $\omega(x_0, y_0)$ is a compact and connected set, and so are the sets of cluster points of the sequences $(x_n)_{n \geq 0}$ and $(y_n)_{n \geq 0}$,
4. the objective function Φ is finite and constant on $\omega(x_0, y_0)$ provided that the optimal value is finite.

Proposition

Let (10) be satisfied. For any $n \geq 0$, the following statements are equivalent:

1. (x_n, y_n) is a critical point of Φ;
2. $(x_{n+1}, y_{n+1}) = (x_n, y_n)$;
3. $\Phi(x_{n+1}, y_{n+1}) = \Phi(x_n, y_n)$.

Let

$\omega(x_0, y_0) := \{$set of cluster points of $(x_n, y_n)_{n\geq 0}$ when x_0 and y_0 are the initial points$\}$.

Theorem (Convergence result)

Let (10) be satisfied and assume that the sequence $(x_n, y_n)_{n\geq 0}$ is bounded. Then the following assertions hold:

1. $\emptyset \neq \omega(x_0, y_0) \subseteq \text{crit}\Phi \subseteq \text{crit}(g + \varphi - h \circ K) \times \text{crit}(h^* - (g + \varphi)^* \circ K^*)$,
2. $\lim_{n\to\infty} \text{dist}((x_n, y_n), \omega(x_0, y_0)) = 0$,
3. if the common optimal value of the problems (5), (6) and (7) is finite, then $\omega(x_0, y_0)$ is a compact and connected set, and so are the sets of cluster points of the sequences $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$,
4. the objective function Φ is finite and constant on $\omega(x_0, y_0)$ provided that the optimal value is finite.

Lemma (subgradient estimation)

For each $n \geq 1$ with $\gamma_{n-1} < \frac{2}{L_{\nabla \varphi}}$, there exist

$$\begin{pmatrix} x_n^* \\ y_n^* \end{pmatrix} = \begin{pmatrix} \frac{x_{n-1}-x_n}{\gamma_{n-1}} + K^*(y_{n-1} - y_n) + \nabla\varphi(x_n) - \nabla\varphi(x_{n-1}) \\ \frac{y_{n-1}-y_n}{\mu_{n-1}} \end{pmatrix} \in \partial\Phi(x_n, y_n),$$

thus,

$$\begin{aligned} \|x_n^*\| &\leq \|K\| \, \|y_{n-1} - y_n\| + \frac{1}{\gamma_{n-1}} \|x_{n-1} - x_n\|, \\ \|y_n^*\| &\leq \frac{1}{\mu_{n-1}} \|y_{n-1} - y_n\|. \end{aligned} \tag{11}$$

Theorem (convergence result when Φ is a KL function)

Let

$$0 < \underline{\gamma} := \inf_{n\geq 0} \gamma_n \leq \overline{\gamma} := \sup_{n\geq 0} \gamma_n < \frac{2}{L_{\nabla\varphi}},$$

$$0 < \underline{\mu} := \inf_{n\geq 0} \mu_n \leq \overline{\mu} := \sup_{n\geq 0} \mu_n < +\infty.$$

Suppose that Φ is in addition a KL function and that the sequences $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ are bounded. Then $(x_n, y_n)_{n\geq 0}$ is a Cauchy sequence, thus convergent to a critical point of Φ.

Lemma (subgradient estimation)

For each $n \geq 1$ with $\gamma_{n-1} < \frac{2}{L_{\nabla\varphi}}$, there exist

$$\begin{pmatrix} x_n^* \\ y_n^* \end{pmatrix} = \begin{pmatrix} \frac{x_{n-1}-x_n}{\gamma_{n-1}} + K^*(y_{n-1} - y_n) + \nabla\varphi(x_n) - \nabla\varphi(x_{n-1}) \\ \frac{y_{n-1}-y_n}{\mu_{n-1}} \end{pmatrix} \in \partial\Phi(x_n, y_n),$$

thus,

$$\begin{aligned} \|x_n^*\| &\leq \|K\| \, \|y_{n-1} - y_n\| + \frac{1}{\gamma_{n-1}} \|x_{n-1} - x_n\|, \\ \|y_n^*\| &\leq \frac{1}{\mu_{n-1}} \|y_{n-1} - y_n\|. \end{aligned} \tag{11}$$

Theorem (convergence result when Φ is a KL function)

Let

$$\begin{aligned} 0 < \underline{\gamma} &:= \inf_{n\geq 0} \gamma_n \leq \overline{\gamma} := \sup_{n\geq 0} \gamma_n < \frac{2}{L_{\nabla\varphi}}, \\ 0 < \underline{\mu} &:= \inf_{n\geq 0} \mu_n \leq \overline{\mu} := \sup_{n\geq 0} \mu_n < +\infty. \end{aligned}$$

Suppose that Φ is in addition a KL function and that the sequences $(x_n)_{n\geq 0}$ and $(y_n)_{n\geq 0}$ are bounded. Then $(x_n, y_n)_{n\geq 0}$ is a Cauchy sequence, thus convergent to a critical point of Φ.

Theorem (convergence rates)

In the hypotheses of the previous theorem, assume that Φ is a KL function with desingularization function $s \mapsto \frac{1}{1-\theta} C s^{1-\theta}$ for some $C > 0$ and $0 \leq \theta < 1$. Let $\bar{x}$ and $\bar{y}$ be the limit points of the sequences $(x_n)_{n \geq 0}$ and $(y_n)_{n \geq 0}$, respectively. Then the following convergence rates are guaranteed:

1. if $\theta = 0$, then there exists $n_0 \geq 0$, such that $x_n = x_{n_0}$ and $y_n = y_{n_0}$ for $n \geq n_0$;
2. if $0 < \theta \leq \frac{1}{2}$, then there exist $c > 0$ and $0 \leq q < 1$ such that

   $$\|x_n - \bar{x}\| \leq c q^n \qquad \text{and} \qquad \|y_n - \bar{y}\| \leq c q^n$$

   for all $n \geq 0$;
3. if $\frac{1}{2} < \theta < 1$, then there exists $c > 0$ such that

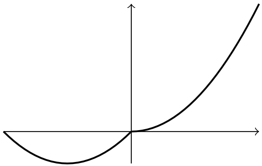
   $$\|x_n - \bar{x}\| \leq c n^{-\frac{1-\theta}{2\theta - 1}} \qquad \text{and} \qquad \|y_n - \bar{y}\| \leq c n^{-\frac{1-\theta}{2\theta - 1}}$$

   for all $n \geq 0$.

An example

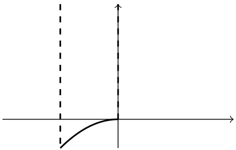
- Primal program

$$\min_{x \in \mathbb{R}} \left\{ \frac{1}{2}x^2 - \max\{-x, 0\} \right\}$$

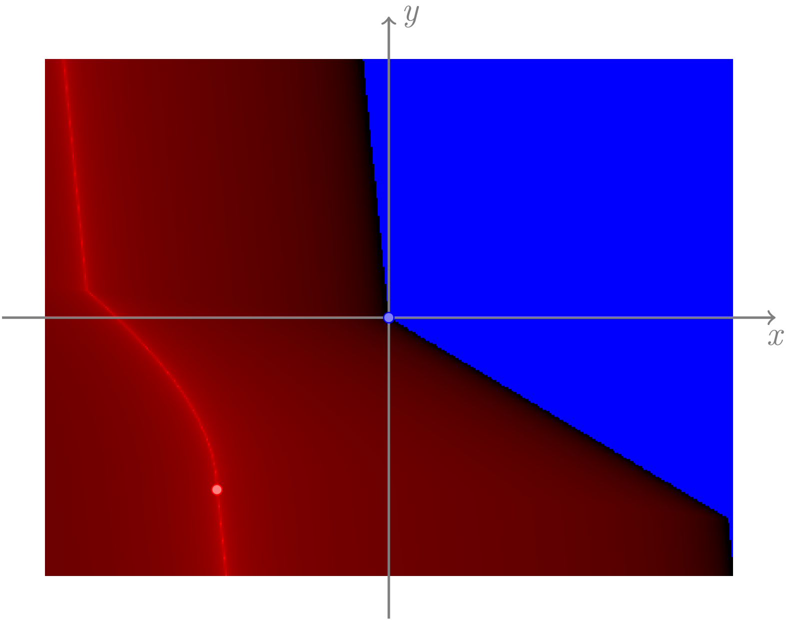


- Dual program

$$\min_{y \in [-1,0]} \left\{ -\frac{1}{2}y^2 \right\}$$



- Primal-dual critical points: $(-1,-1)$ and $(0,0)$.



y
x

Application to image processing

- We represent an image of the size $m \times n$ pixels by a vector $x \in \mathbb{R}^{mn}$ with entries in $[0,1]$ (where 0 represents pure black and 1 represents pure white).
- The original image $x \in \mathbb{R}^{mn}$ is assumed to be blurred by a linear operator $A : \mathbb{R}^{mn} \to \mathbb{R}^{mn}$ and corrupted with noise ν. Knowing $b = Ax + \nu$, we want to reconstruct the original image x by considering the minimization problem

$$\min_{x \in \mathbb{R}^{mn}} \left(\frac{\mu}{2} \|Ax - b\|^2 + J(Dx) \right),$$

where $\mu > 0$ is a regularization parameter, $D : \mathbb{R}^{mn} \to \mathbb{R}^{2mn}$ is the discrete gradient operator given by $Dx = (D_1 x, D_2 x)$,

$$D_1 : \mathbb{R}^{mn} \to \mathbb{R}^{mn}, (D_1 x)_{i,j} := \begin{cases} x_{i+1,j} - x_{i,j}, & i = 1, \dots, m-1; j = 1, \dots, n; \\ 0, & i = m; j = 1, \dots, n \end{cases}$$

$$D_2 : \mathbb{R}^{mn} \to \mathbb{R}^{mn}, (D_2 x)_{i,j} := \begin{cases} x_{i,j+1} - x_{i,j}, & i = 1, \dots, m; j = 1, \dots, n-1; \\ 0, & i = 1, \dots, m; j = n, \end{cases}$$

and $J : \mathbb{R}^{mn} \to \mathbb{R}$ is a regularizing functional penalizing noisy images.

Choices for the functional J:

- Zhang penalty (Zhang, 2009): $\mathrm{Zhang}_a(z) = \sum_{j=1}^{2mn} g_a(z_j)$, where $a > 0$ and

$$g_a(z_j) = \begin{cases} \frac{1}{a}\left|z_j\right| & \text{if } \left|z_j\right| < a, \\ 1 & \text{if } \left|z_j\right| \geq a \end{cases} = \frac{1}{a}\left|z_j\right| - \begin{cases} 0 & \text{if } \left|z_j\right| < a, \\ \frac{1}{a}(\left|z_j\right| - a) & \text{if } \left|z_j\right| \geq a. \end{cases}$$

Denoting the part after the curly brace as $h_a(z_j)$ and $h_a(z) := \sum_{j=1}^{2mn} h_a(z_j)$, we have

$$\mathrm{prox}_{\gamma h_a^*}(z) = \begin{cases} -\frac{1}{a} & \text{if } z \leq -\frac{1}{a} - \gamma a, \\ z + \gamma a & \text{if } -\frac{1}{a} - \gamma a \leq z \leq -\gamma a, \\ 0 & \text{if } -\gamma a \leq z \leq \gamma a, \\ z - \gamma a & \text{if } \gamma a \leq z \leq \frac{1}{a} + \gamma a, \\ \frac{1}{a} & \text{if } z \geq \frac{1}{a} + \gamma a. \end{cases}$$

- LZOX penalty (Lou, Zeng, Osher, Xin, 2009): $\mathrm{LZOX}_a(z) = \|z\|_{\ell^1} - a\|z\|_\times$, where

$$\|(u,v)\|_\times := \sum_{i=1}^{m}\sum_{j=1}^{n}\sqrt{u_{i,j}^2 + v_{i,j}^2}.$$

Choices for the functional J:

- Zhang penalty (Zhang, 2009): $\mathrm{Zhang}_a(z) = \sum_{j=1}^{2mn} g_a(z_j)$, where $a > 0$ and

$$g_a(z_j) = \begin{cases} \frac{1}{a}|z_j| & \text{if } |z_j| < a, \\ 1 & \text{if } |z_j| \geq a \end{cases} = \frac{1}{a}|z_j| - \begin{cases} 0 & \text{if } |z_j| < a, \\ \frac{1}{a}(|z_j| - a) & \text{if } |z_j| \geq a. \end{cases}$$

Denoting the part after the curly brace as $h_a(z_j)$ and $h_a(z) := \sum_{j=1}^{2mn} h_a(z_j)$, we have

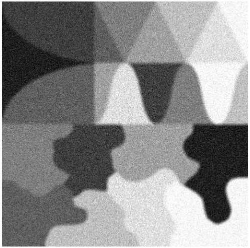
$$\mathrm{prox}_{\gamma h_a^*}(z) = \begin{cases} -\frac{1}{a} & \text{if } z \leq -\frac{1}{a} - \gamma a, \\ z + \gamma a & \text{if } -\frac{1}{a} - \gamma a \leq z \leq -\gamma a, \\ 0 & \text{if } -\gamma a \leq z \leq \gamma a, \\ z - \gamma a & \text{if } \gamma a \leq z \leq \frac{1}{a} + \gamma a, \\ \frac{1}{a} & \text{if } z \geq \frac{1}{a} + \gamma a. \end{cases}$$

- LZOX penalty (Lou, Zeng, Osher, Xin, 2009): $\mathrm{LZOX}_a(z) = \|z\|_{\ell^1} - a\|z\|_\times$, where

$$\|(u, v)\|_\times := \sum_{i=1}^{m} \sum_{j=1}^{n} \sqrt{u_{i,j}^2 + v_{i,j}^2}.$$

- We tested the MATLAB code on a PC with Intel Core i5 4670S ($4\times$ 3.10GHz) and 8GB DDR3 RAM (1600MHz);
- Stopping criterion: $\|(x_{n+1}, y_{n+1}) - (x_n, y_n)\|_\infty \leq 10^{-4}$;
- Stepsizes: $\mu_n = \gamma_n = \frac{1}{8\mu}$ for all $n \geq 0$;
- Initial values: $x_0 = b$, $y_0 \in \partial h(Kx_0)$.

(b) Original image



(c) Blurry image

- $\text{ISNR}(x_k) = 10 \log_{10} \left(\frac{\|x-b\|^2}{\|x-x_k\|^2} \right)$

	$a = 0.01$	$a = 0.03$	$a = 0.1$	$a = 0.3$	$a = 1.0$	$a = 3.0$
$\mu = 1.0$	−43.708	−33.711	−23.148	−13.846	−3.0288	2.4922
$\mu = 10.0$	−18.781	−9.9406	−3.2070	2.5442	5.9227	**6.97777**
$\mu = 20.0$	−11.270	−4.8428	0.43533	4.7768	6.76613	6.57299
$\mu = 50.0$	−4.8333	−1.05553	2.63959	6.46109	6.81752	3.952101
$\mu = 100.0$	−1.7546	−0.14560	3.16532	6.90202	5.29597	2.129705
$\mu = 200.0$	−0.41418	0.0619477	2.98543	6.38513	3.088196	1.110186
$\mu = 500.0$	0.0077144	0.121807	2.101321	3.816813	1.317390	0.482406
$\mu = 1000.0$	0.0528014	0.127592	1.423684	2.070959	0.692487	0.271777

ISNR values for Zhang after 50 iterations

	$a = 0.00$	$a = 0.2$	$a = 0.4$	$a = 0.5$	$a = 0.6$	$a = 0.8$	$a = 1.0$
$\mu = 1.0$	−3.0288	−4.2266	−3.7637	−3.6569	−3.5150	−4.3590	−13.701
$\mu = 10.0$	5.9227	6.26615	6.414791	6.44871	6.45780	6.28863	4.301090
$\mu = 20.0$	6.76613	6.90005	**6.93064**	6.917926	6.88018	6.61521	5.305623
$\mu = 50.0$	6.81752	6.78308	6.65411	6.4923	6.36250	5.780558	4.741993
$\mu = 100.0$	5.29597	5.23264	5.05189	4.91247	4.739717	4.287092	3.696120
$\mu = 200.0$	3.088196	3.060511	2.985871	2.930448	2.863122	2.693096	2.477708
$\mu = 500.0$	1.317390	1.312168	1.298834	1.288983	1.277010	1.246724	1.208036
$\mu = 1000.0$	0.692487	0.691049	0.687585	0.685057	0.682000	0.674272	0.664401

ISNR values for LZOX after 50 iterations

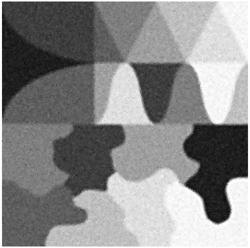
(d) LZOX, $\mu = 20$, $a = 0.4$

(e) LZOX, $\mu = 20$, $a = 1$

(f) LZOX, $\mu = 50$, $a = 0$

(g) Zhang, $\mu = 10$, $a = 3$

(h) Zhang, $\mu = 20$, $a = 1$



(i) Zhang, $\mu = 100$, $a = 0.1$

Reconstructions

References

H. Attouch, J. Bolte, B.F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*, Mathematical Programming 137(1-2) Series A, 91–129, 2013

S. Banert, R.I. Boţ, *A general double-proximal gradient algorithm for d.c. programming*, Mathematical Programming, DOI: 10.1007/s10107-018-1292-2

J. Bolte, A. Daniilidis, A. Lewis, M. Shota, *Clarke subgradients of stratifiable functions*, SIAM Journal on Optimization 18(2), 556–572, 2007

J. Bolte, A. Daniilidis, O. Ley, L. Mazet, *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity*, Transactions of the American Mathematical Society 362(6), 3319–3363, 2010

J. Bolte, S. Sabach, M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Mathematical Programming Series A (146)(1–2), 459–494, 2014

R.I. Boţ, E.R. Csetnek, *A forward-backward dynamical approach to the minimization of the sum of a nonsmooth convex with a smooth nonconvex function*, ESAIM: Control, Optimization and Calculus of Variations 24(2), 463-477, 2018

R.I. Boţ, E.R. Csetnek, S. László, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO Journal on Computational Optimization 4, 3–25, 2016

R.I. Boţ, E.R. Csetnek, D.-K. Nguyen, *A proximal minimization algorithm for structured nonconvex and nonsmooth problems*, to appear in SIAM Journal on Optimization

R.I. Boţ, D.-K. Nguyen, *The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates*, to appear in Mathematics of Operations Research