

Four (plus one) open questions about the minimization of structured nonsmooth and nonconvex functions

Radu Ioan Boţ

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

Operator Splitting Methods in Data Analysis
Flatiron Institute, New York
March 20-22, 2019

Contents

Proximal algorithms for structured nonconvex problems

- Problem formulation and algorithm

- Particular instances

Convergence analysis

- A descent inequality

- General conditions for $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ bounded

- The cluster points of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ are KKT points

- Global convergence under Kurdyka-Łojasiewicz assumptions

Bonus material: D.C. programming

- A double-proximal gradient algorithm

- Application to image processing

Proximal algorithms for structured nonconvex problems

Problem formulation and algorithm

We address the solving of optimization problems of the form

$$\min_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} \{f(Ax) + g(y) + h(x,y)\}, \quad (1)$$

where

- ▶ $h: \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a **Fréchet differentiable function**;
- ▶ $f: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{R}^q \rightarrow \mathbb{R} \cup \{+\infty\}$ are **proper and lower semicontinuous functions**;
- ▶ $A: \mathbb{R}^m \rightarrow \mathbb{R}^p$ is a linear operator.

If $m = p$ and A is the identity operator, then (1) becomes

$$\min_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} \{f(x) + g(y) + h(x,y)\}.$$

PALM: Proximal Alternating Linearized Minimization (Bolte, Sabach, Teboulle, Mathematical Programming, 2014)

Let $\gamma_1 > 1$ and $\gamma_2 > 1$. For a given starting point $(x^0, y^0) \in \mathbb{R}^m \times \mathbb{R}^q$, let be for any $n \geq 0$

$$\begin{aligned} x^{n+1} &\in \operatorname{prox}_{\frac{1}{\gamma_1 \ell(y^n)} f} \left\{ x^n - \frac{1}{\gamma_1 \ell(y^n)} \nabla_x h(x^n, y^n) \right\} \\ y^{n+1} &\in \operatorname{prox}_{\frac{1}{\gamma_2 \ell(x^{n+1})} g} \left\{ y^n - \frac{1}{\gamma_2 \ell(x^{n+1})} \nabla_y h(x^{n+1}, y^n) \right\}. \end{aligned}$$

Assumptions for PALM

- ▶ (i) f, g and h are bounded from below;
- ▶ (ii) for any fixed $y \in \mathbb{R}^q$ the function $x \mapsto h(x, y)$ is $C_{\ell_1(y)}^{1,1}$;
- ▶ (iii) for any fixed $x \in \mathbb{R}^m$ the function $y \mapsto h(x, y)$ is $C_{\ell_1(x)}^{1,1}$;
- ▶ (iv) $\lambda_1^+ \geq \sup_{n \geq 0} \ell_1(y^n) \geq \inf_{n \geq 0} \ell_1(y^n) \geq \lambda_1^- > 0$, $\lambda_2^+ \geq \sup_{n \geq 0} \ell_2(x^n) \geq \inf_{n \geq 0} \ell_2(x^n) \geq \lambda_2^- > 0$;
- ▶ (v) ∇h is Lipschitz continuous in bounded sets of $\mathbb{R}^m \times \mathbb{R}^q$.

If $m = p$ and A is the identity operator, then (1) becomes

$$\min_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} \{f(x) + g(y) + h(x,y)\}.$$

PALM: Proximal Alternating Linearized Minimization (Bolte, Sabach, Teboulle, Mathematical Programming, 2014)

Let $\gamma_1 > 1$ and $\gamma_2 > 1$. For a given starting point $(x^0, y^0) \in \mathbb{R}^m \times \mathbb{R}^q$, let be for any $n \geq 0$

$$\begin{aligned} x^{n+1} &\in \text{prox}_{\frac{1}{\gamma_1 \ell(y^n)} f} \left\{ x^n - \frac{1}{\gamma_1 \ell(y^n)} \nabla_x h(x^n, y^n) \right\} \\ y^{n+1} &\in \text{prox}_{\frac{1}{\gamma_2 \ell(x^{n+1})} g} \left\{ y^n - \frac{1}{\gamma_2 \ell(x^{n+1})} \nabla_y h(x^{n+1}, y^n) \right\}. \end{aligned}$$

Assumptions for PALM

- ▶ (i) f, g and h are bounded from below;
- ▶ (ii) for any fixed $y \in \mathbb{R}^q$ the function $x \mapsto h(x, y)$ is $C_{\ell_1(y)}^{1,1}$;
- ▶ (iii) for any fixed $x \in \mathbb{R}^m$ the function $y \mapsto h(x, y)$ is $C_{\ell_1(x)}^{1,1}$;
- ▶ (iv) $\lambda_1^+ \geq \sup_{n \geq 0} \ell_1(y^n) \geq \inf_{n \geq 0} \ell_1(y^n) \geq \lambda_1^- > 0$, $\lambda_2^+ \geq \sup_{n \geq 0} \ell_2(x^n) \geq \inf_{n \geq 0} \ell_2(x^n) \geq \lambda_2^- > 0$;
- ▶ (v) ∇h is Lipschitz continuous in bounded sets of $\mathbb{R}^m \times \mathbb{R}^q$.

Algorithm

Let $\mu, \beta, \tau > 0$ and $0 < \sigma \leq 1$. For a given starting point $(x_0, y_0, z_0, u_0) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p$, let be for any $n \geq 0$

$$y_{n+1} \in \arg \min_{y \in \mathbb{R}^q} \left\{ g(y) + \langle \nabla_y h(x_n, y_n), y \rangle + \frac{\mu}{2} \|y - y_n\|^2 \right\}$$

$$z_{n+1} \in \arg \min_{z \in \mathbb{R}^p} \left\{ f(z) + \langle u_n, Ax_n - z \rangle + \frac{\beta}{2} \|Ax_n - z\|^2 \right\}$$

$$x_{n+1} := x_n - \tau^{-1} (\nabla_x h(x_n, y_{n+1}) + A^T u_n + \beta A^T (Ax_n - z_{n+1}))$$

$$u_{n+1} := u_n + \sigma \beta (Ax_{n+1} - z_{n+1}).$$

Equivalent formulation

For any $n \geq 0$ let be

$$y_{n+1} \in \text{prox}_{\mu^{-1}g} (y_n - \mu^{-1} \nabla_y h(x_n, y_n))$$

$$z_{n+1} \in \text{prox}_{\beta^{-1}f} (Ax_n + \beta^{-1} u_n)$$

$$x_{n+1} := x_n - \tau^{-1} (\nabla_x h(x_n, y_{n+1}) + A^T u_n + \beta A^T (Ax_n - z_{n+1}))$$

$$u_{n+1} := u_n + \sigma \beta (Ax_{n+1} - z_{n+1}).$$

Algorithm

Let $\mu, \beta, \tau > 0$ and $0 < \sigma \leq 1$. For a given starting point $(x_0, y_0, z_0, u_0) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p$, let be for any $n \geq 0$

$$y_{n+1} \in \arg \min_{y \in \mathbb{R}^q} \left\{ g(y) + \langle \nabla_y h(x_n, y_n), y \rangle + \frac{\mu}{2} \|y - y_n\|^2 \right\}$$

$$z_{n+1} \in \arg \min_{z \in \mathbb{R}^p} \left\{ f(z) + \langle u_n, Ax_n - z \rangle + \frac{\beta}{2} \|Ax_n - z\|^2 \right\}$$

$$x_{n+1} := x_n - \tau^{-1} \left(\nabla_x h(x_n, y_{n+1}) + A^T u_n + \beta A^T (Ax_n - z_{n+1}) \right)$$

$$u_{n+1} := u_n + \sigma \beta (Ax_{n+1} - z_{n+1}).$$

Equivalent formulation

For any $n \geq 0$ let be

$$y_{n+1} \in \text{prox}_{\mu^{-1}g} \left(y_n - \mu^{-1} \nabla_y h(x_n, y_n) \right)$$

$$z_{n+1} \in \text{prox}_{\beta^{-1}f} \left(Ax_n + \beta^{-1} u_n \right)$$

$$x_{n+1} := x_n - \tau^{-1} \left(\nabla_x h(x_n, y_{n+1}) + A^T u_n + \beta A^T (Ax_n - z_{n+1}) \right)$$

$$u_{n+1} := u_n + \sigma \beta (Ax_{n+1} - z_{n+1}).$$

Particular instances

The case $g(y) = 0$ and $h(x, y) = h(x)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$

The optimization problem (1) becomes

$$\min_{x \in \mathbb{R}^m} \{f(Ax) + h(x)\},$$

while the iterative scheme reads for any $n \geq 0$

$$z_{n+1} \in \text{prox}_{\beta^{-1}f}(Ax_n + \beta^{-1}u_n)$$

$$x_{n+1} := x_n - \tau^{-1}(\nabla h(x_n) + A^T u_n + \beta A^T(Ax_n - z_{n+1}))$$

$$u_{n+1} := u_n + \sigma\beta(Ax_{n+1} - z_{n+1}).$$

The case $m = p$ and $A = \text{Id}$

The optimization problem (1) becomes

$$\min_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} \{f(x) + g(y) + h(x, y)\},$$

while the iterative scheme reads for any $n \geq 0$

$$y_{n+1} \in \text{prox}_{\mu^{-1}g}(y_n - \mu^{-1}\nabla_y h(x_n, y_n))$$

$$z_{n+1} \in \text{prox}_{\beta^{-1}f}(x_n + \beta^{-1}u_n)$$

$$x_{n+1} := x_n - \tau^{-1}(\nabla_x h(x_n, y_{n+1}) + u_n + \beta(x_n - z_{n+1}))$$

$$u_{n+1} := u_n + \sigma\beta(x_{n+1} - z_{n+1}).$$

Particular instances

The case $g(y) = 0$ and $h(x, y) = h(x)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$

The optimization problem (1) becomes

$$\min_{x \in \mathbb{R}^m} \{f(Ax) + h(x)\},$$

while the iterative scheme reads for any $n \geq 0$

$$\begin{aligned} z_{n+1} &\in \text{prox}_{\beta^{-1}f}(Ax_n + \beta^{-1}u_n) \\ x_{n+1} &:= x_n - \tau^{-1}(\nabla h(x_n) + A^T u_n + \beta A^T(Ax_n - z_{n+1})) \\ u_{n+1} &:= u_n + \sigma\beta(Ax_{n+1} - z_{n+1}). \end{aligned}$$

The case $m = p$ and $A = \text{Id}$

The optimization problem (1) becomes

$$\min_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} \{f(x) + g(y) + h(x, y)\},$$

while the iterative scheme reads for any $n \geq 0$

$$\begin{aligned} y_{n+1} &\in \text{prox}_{\mu^{-1}g}(y_n - \mu^{-1}\nabla_y h(x_n, y_n)) \\ z_{n+1} &\in \text{prox}_{\beta^{-1}f}(x_n + \beta^{-1}u_n) \\ x_{n+1} &:= x_n - \tau^{-1}(\nabla_x h(x_n, y_{n+1}) + u_n + \beta(x_n - z_{n+1})) \\ u_{n+1} &:= u_n + \sigma\beta(x_{n+1} - z_{n+1}). \end{aligned}$$

The case $m = p$, $A = \text{Id}$, $f(x) = 0$ and $h(x, y) = h(y)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$

The optimization problem (1) becomes

$$\min_{y \in \mathbb{R}^q} \{g(y) + h(y)\},$$

while the iterative scheme reads for any $n \geq 0$

$$y_{n+1} \in \text{prox}_{\mu^{-1}g} \left(y_n - \mu^{-1} \nabla h(y_n) \right).$$

A descent inequality

Assumption 1

1. the functions f, g and h are bounded from below;
2. the linear operator A is **surjective**;
3. for any fixed $y \in \mathbb{R}^q$ there exists $\ell_1(y) \geq 0$ such that

$$\|\nabla_x h(x, y) - \nabla_x h(x', y)\| \leq \ell_1(y) \|x - x'\| \quad \forall x, x' \in \mathbb{R}^m,$$

and for any fixed $x \in \mathbb{R}^m$ there exist $\ell_2(x), \ell_3(x) \geq 0$ such that

$$\|\nabla_y h(x, y) - \nabla_y h(x, y')\| \leq \ell_2(x) \|y - y'\| \quad \forall y, y' \in \mathbb{R}^q,$$

$$\|\nabla_x h(x, y) - \nabla_x h(x, y')\| \leq \ell_3(x) \|y - y'\| \quad \forall y, y' \in \mathbb{R}^q;$$

4. there exist $\ell_{i,+} > 0, i = 1, 2, 3$, such that

$$\sup_{n \geq 0} \ell_1(y_n) \leq \ell_{1,+}, \quad \sup_{n \geq 0} \ell_2(x_n) \leq \ell_{2,+}, \quad \sup_{n \geq 0} \ell_3(x_n) \leq \ell_{3,+}.$$

Regularized augmented Lagrangian

Define

$$\Psi: \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$$

as

$$\begin{aligned} (x, y, z, u, x', u') \mapsto & f(z) + g(y) + h(x, y) + \langle u, Ax - z \rangle + \frac{\beta}{2} \|Ax - z\|^2 \\ & + C_0 \|A^T(u - u') + \sigma B(x - x')\|^2 + C_1 \|x - x'\|^2, \end{aligned}$$

where

$$B := \tau \text{Id} - \beta A^T A, \quad C_0 := \frac{4(1-\sigma)}{\sigma^2 \beta \lambda_{\min}(AA^T)} \geq 0 \quad \text{and} \quad C_1 := \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma \beta \lambda_{\min}(AA^T)} > 0.$$

► If $2\tau \geq \beta \|A\|^2$, then

$$\|B\| \leq \tau.$$

Notations

$$\mathbf{R} := \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p$$

$$\mathbf{X} := (x, y, z, u, x', u')$$

$$\mathbf{X}_n := (x_n, y_n, z_n, u_n, x_{n-1}, u_{n-1}) \quad \forall n \geq 1$$

$$\Psi_n := \Psi(\mathbf{X}_n) \quad \forall n \geq 1.$$

Regularized augmented Lagrangian

Define

$$\Psi: \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$$

as

$$\begin{aligned} (x, y, z, u, x', u') \mapsto & f(z) + g(y) + h(x, y) + \langle u, Ax - z \rangle + \frac{\beta}{2} \|Ax - z\|^2 \\ & + C_0 \|A^T(u - u') + \sigma B(x - x')\|^2 + C_1 \|x - x'\|^2, \end{aligned}$$

where

$$B := \tau \text{Id} - \beta A^T A, \quad C_0 := \frac{4(1 - \sigma)}{\sigma^2 \beta \lambda_{\min}(AA^T)} \geq 0 \quad \text{and} \quad C_1 := \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma\beta \lambda_{\min}(AA^T)} > 0.$$

► If $2\tau \geq \beta \|A\|^2$, then

$$\|B\| \leq \tau.$$

Notations

$$\mathbf{R} := \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p$$

$$\mathbf{X} := (x, y, z, u, x', u')$$

$$\mathbf{X}_n := (x_n, y_n, z_n, u_n, x_{n-1}, u_{n-1}) \quad \forall n \geq 1$$

$$\Psi_n := \Psi(\mathbf{X}_n) \quad \forall n \geq 1.$$

A descent inequality

Let Assumption 1 be satisfied and $2\tau \geq \beta \|A\|^2$. Then for any $n \geq 1$ it holds

$$\Psi_{n+1} + C_2 \|x_{n+1} - x_n\|^2 + C_3 \|y_{n+1} - y_n\|^2 + C_4 \|u_{n+1} - u_n\|^2 \leq \Psi_n,$$

where

$$C_2 := \tau - \frac{\ell_{1,+} + \beta \|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma\beta\lambda_{\min}(AA^T)},$$

$$C_3 := \frac{\mu - \ell_{2,+}}{2} - \frac{8\ell_{3,+}^2}{\sigma\beta\lambda_{\min}(AA^T)},$$

$$C_4 := \frac{1}{\sigma\beta}.$$

Convergence results I

Let Assumption 1 be satisfied and the algorithm parameters be such that $2\tau \geq \beta \|A\|^2$ and $\min\{C_2, C_3, C_4\} > 0$. The following statements are true:

1. the sequence $\{\Psi_n\}_{n \geq 1}$ is bounded from below and convergent;
- 2.

$$x_{n+1} - x_n \rightarrow 0, \quad y_{n+1} - y_n \rightarrow 0, \quad z_{n+1} - z_n \rightarrow 0 \quad \text{and} \quad u_{n+1} - u_n \rightarrow 0 \quad \text{as} \quad n \rightarrow +\infty.$$

A descent inequality

Let Assumption 1 be satisfied and $2\tau \geq \beta \|A\|^2$. Then for any $n \geq 1$ it holds

$$\Psi_{n+1} + C_2 \|x_{n+1} - x_n\|^2 + C_3 \|y_{n+1} - y_n\|^2 + C_4 \|u_{n+1} - u_n\|^2 \leq \Psi_n,$$

where

$$C_2 := \tau - \frac{\ell_{1,+} + \beta \|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma\beta\lambda_{\min}(AA^T)},$$

$$C_3 := \frac{\mu - \ell_{2,+}}{2} - \frac{8\ell_{3,+}^2}{\sigma\beta\lambda_{\min}(AA^T)},$$

$$C_4 := \frac{1}{\sigma\beta}.$$

Convergence results I

Let Assumption 1 be satisfied and the algorithm parameters be such that $2\tau \geq \beta \|A\|^2$ and $\min\{C_2, C_3, C_4\} > 0$. The following statements are true:

1. the sequence $\{\Psi_n\}_{n \geq 1}$ is bounded from below and convergent;
- 2.

$$x_{n+1} - x_n \rightarrow 0, \quad y_{n+1} - y_n \rightarrow 0, \quad z_{n+1} - z_n \rightarrow 0 \quad \text{and} \quad u_{n+1} - u_n \rightarrow 0 \quad \text{as} \quad n \rightarrow +\infty.$$

- **Open question 1:** Is it possible to obtain a useful descent inequality (for an appropriate Lyapunov function) **without** assuming that the linear operator is surjective?

A possible choice of the algorithm parameters

Let

$$\begin{aligned}
 0 < \sigma < \frac{1}{24\kappa(AA^T)}, \quad \mu > \ell_{2,+} + \frac{16\ell_{3,+}^2}{\sigma\beta\lambda_{\min}(AA^T)} > 0, \\
 \beta > \frac{4\nu}{1 - 24\sigma\kappa(AA^T)} \left(4 + 3\sigma + \sqrt{24 + 24\sigma + 9\sigma^2 - 192\sigma\kappa(AA^T)} \right) > 0 \\
 \max \left\{ \frac{\beta\|A\|^2}{2}, \frac{\beta\lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} - \sqrt{\Delta'_\tau} \right) \right\} < \tau < \\
 \frac{\beta\lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} + \sqrt{\Delta'_\tau} \right),
 \end{aligned}$$

where

$$\nu := \frac{\ell_{1,+}}{\lambda_{\min}(AA^T)} > 0 \text{ and } \Delta'_\tau := 1 - \frac{32\nu}{\beta} - \frac{128\nu^2}{\beta^2} - \frac{24\nu\sigma}{\beta} - 24\sigma\kappa(AA^T) > 0.$$

Then we have

$$\min\{C_2, C_3, C_4\} > 0.$$

furthermore, there exist $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ such that

$$\frac{1}{\gamma_1} - \frac{\ell_{1,+}}{2\gamma_1^2} = \frac{1}{\beta\lambda_{\min}(AA^T)} \quad \text{and} \quad \frac{1}{\gamma_2} - \frac{\ell_{1,+}}{2\gamma_2^2} = \frac{2}{\beta\lambda_{\min}(AA^T)}. \quad (2)$$

General conditions for $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ bounded

Let Assumption 1 be satisfied and the algorithm parameters i be such that $2\tau \geq \beta \|A\|^2$, $\min\{C_2, C_3, C_4\} > 0$ and there exist $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ such that (2) holds. Suppose that **one** of the following conditions hold:

1. the function h is coercive;
2. the operator A is invertible, and f and g are coercive.

Then every sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by the algorithm is **bounded**.

► Denote

$$s_* := \sup_{n \geq 0} \{\|x_{n+1} - x_n\|, \|y_{n+1} - y_n\|, \|z_{n+1} - z_n\|, \|u_{n+1} - u_n\|\} < +\infty.$$

► Let $n \geq 1$ be fixed. We have

$$\Psi_1 \geq \dots \geq \Psi_n \geq \Psi_{n+1} \geq f(z_{n+1}) + g(y_{n+1}) + h(x_{n+1}, y_{n+1}) - \frac{1}{2\beta} \|u_{n+1}\|^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2$$

and, from here,

$$\begin{aligned} \Psi_1 \geq & f(z_{n+1}) + g(y_{n+1}) + h(x_{n+1}, y_{n+1}) - \frac{1}{\beta \lambda_{\min}(AA^T)} \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \\ & - \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2. \end{aligned}$$

General conditions for $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ bounded

Let Assumption 1 be satisfied and the algorithm parameters i be such that $2\tau \geq \beta \|A\|^2$, $\min\{C_2, C_3, C_4\} > 0$ and there exist $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ such that (2) holds. Suppose that **one** of the following conditions hold:

1. the function h is coercive;
2. the operator A is invertible, and f and g are coercive.

Then every sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by the algorithm is **bounded**.

► Denote

$$s_* := \sup_{n \geq 0} \{\|x_{n+1} - x_n\|, \|y_{n+1} - y_n\|, \|z_{n+1} - z_n\|, \|u_{n+1} - u_n\|\} < +\infty.$$

► Let $n \geq 1$ be fixed. We have

$$\Psi_1 \geq \dots \geq \Psi_n \geq \Psi_{n+1} \geq f(z_{n+1}) + g(y_{n+1}) + h(x_{n+1}, y_{n+1}) - \frac{1}{2\beta} \|u_{n+1}\|^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2$$

and, from here,

$$\begin{aligned} \Psi_1 \geq f(z_{n+1}) + g(y_{n+1}) + h(x_{n+1}, y_{n+1}) - \frac{1}{\beta \lambda_{\min}(AA^T)} \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \\ - \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2. \end{aligned}$$

General conditions for $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ bounded

Let Assumption 1 be satisfied and the algorithm parameters i be such that $2\tau \geq \beta \|A\|^2$, $\min\{C_2, C_3, C_4\} > 0$ and there exist $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ such that (2) holds. Suppose that **one** of the following conditions hold:

1. the function h is coercive;
2. the operator A is invertible, and f and g are coercive.

Then every sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by the algorithm is **bounded**.

► Denote

$$s_* := \sup_{n \geq 0} \{\|x_{n+1} - x_n\|, \|y_{n+1} - y_n\|, \|z_{n+1} - z_n\|, \|u_{n+1} - u_n\|\} < +\infty.$$

► Let $n \geq 1$ be fixed. We have

$$\Psi_1 \geq \dots \geq \Psi_n \geq \Psi_{n+1} \geq$$

$$f(z_{n+1}) + g(y_{n+1}) + h(x_{n+1}, y_{n+1}) - \frac{1}{2\beta} \|u_{n+1}\|^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2$$

and, from here,

$$\begin{aligned} \Psi_1 \geq & f(z_{n+1}) + g(y_{n+1}) + h(x_{n+1}, y_{n+1}) - \frac{1}{\beta \lambda_{\min}(AA^T)} \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \\ & - \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2. \end{aligned}$$

► 1. The conclusion follows from

$$\begin{aligned}
& \frac{1}{2} h(x_{n+1}, y_{n+1}) + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2 \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} f(z) - \inf_{y \in \mathbb{R}^m} g(y) \\
& \quad - \frac{1}{2} \inf_{n \geq 1} \left\{ h(x_{n+1}, y_{n+1}) - \left(\frac{1}{\gamma_1} - \frac{\ell_{1,+}}{2\gamma_1^2} \right) \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \right\} \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} f(z) - \inf_{y \in \mathbb{R}^q} g(y) \\
& \quad - \frac{1}{2} \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} h(x, y) < +\infty.
\end{aligned}$$

► 2. The conclusion follows from

$$\begin{aligned}
& f(z_{n+1}) + g(y_{n+1}) + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2 \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 \\
& \quad - \inf_{n \geq 1} \left\{ h(x_{n+1}, y_{n+1}) - \left(\frac{1}{\gamma_2} - \frac{\ell_{1,+}}{2\gamma_2^2} \right) \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \right\} \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} h(x, y) < +\infty.
\end{aligned}$$

► 1. The conclusion follows from

$$\begin{aligned}
& \frac{1}{2} h(x_{n+1}, y_{n+1}) + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2 \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} f(z) - \inf_{y \in \mathbb{R}^m} g(y) \\
& \quad - \frac{1}{2} \inf_{n \geq 1} \left\{ h(x_{n+1}, y_{n+1}) - \left(\frac{1}{\gamma_1} - \frac{\ell_{1,+}}{2\gamma_1^2} \right) \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \right\} \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} f(z) - \inf_{y \in \mathbb{R}^q} g(y) \\
& \quad - \frac{1}{2} \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} h(x, y) < +\infty.
\end{aligned}$$

► 2. The conclusion follows from

$$\begin{aligned}
& f(z_{n+1}) + g(y_{n+1}) + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2 \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 \\
& \quad - \inf_{n \geq 1} \left\{ h(x_{n+1}, y_{n+1}) - \left(\frac{1}{\gamma_2} - \frac{\ell_{1,+}}{2\gamma_2^2} \right) \|\nabla_x h(x_{n+1}, y_{n+1})\|^2 \right\} \\
& \leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} h(x, y) < +\infty.
\end{aligned}$$

Descent Lemma

Let $\psi: \mathbb{R}^d \rightarrow \mathbb{R}$ be Fréchet differentiable such that its gradient is Lipschitz continuous with constant $\ell > 0$. Then the following statements are true:

1. For every $x, y \in \mathbb{R}^d$ and every $z \in [x, y] = \{(1-t)x + ty: t \in [0, 1]\}$ it holds

$$\psi(y) \leq \psi(x) + \langle \nabla \psi(z), y - x \rangle + \frac{\ell}{2} \|y - x\|^2;$$

2. for any $\gamma \in \mathbb{R} \setminus \{0\}$ it holds

$$\inf_{x \in \mathbb{R}^d} \left\{ \psi(x) - \left(\frac{1}{\gamma} - \frac{\ell}{2\gamma^2} \right) \|\nabla \psi(x)\|^2 \right\} \geq \inf_{x \in \mathbb{R}^d} \psi(x).$$

- **Open question 2:** Is there another approach which can be used to guarantee boundedness for the sequence of generated iterates and which, in particular, works when each summand in the objective function is **nonsmooth**?

The cluster points of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ are KKT points

Assumption 2

1. the functions f, g and h are bounded from below;
2. the linear operator A is surjective;
3. every generated sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ is bounded:
4. ∇h is Lipschitz continuous with constant $L > 0$ on a convex bounded subset $B_1 \times B_2 \subseteq \mathbb{R}^m \times \mathbb{R}^q$ containing $\{(x_n, y_n)\}_{n \geq 0}$. In other words, for any $(x, y), (x', y') \in B_1 \times B_2$ it holds

$$\begin{aligned} & ||| (\nabla_x h(x, y) - \nabla_x h(x', y'), \nabla_y h(x, y) - \nabla_y h(x', y')) ||| \leq \\ & L ||| (x, y) - (x', y') |||; \end{aligned}$$

5. the parameters $\mu, \beta, \tau > 0$ and $0 < \sigma \leq 1$ are such that $2\tau \geq \beta \|A\|^2$ and $\min\{C_2, C_3, C_4\} > 0$, where

$$\begin{aligned} C_2 &:= \tau - \frac{L\sqrt{2} + \beta \|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - \frac{8(\sigma\tau + L\sqrt{2})^2}{\sigma\beta\lambda_{\min}(AA^T)}, \\ C_3 &:= \frac{\mu - L\sqrt{2}}{2} - \frac{16L^2}{\sigma\beta\lambda_{\min}(AA^T)}, C_4 := \frac{1}{\sigma\beta}. \end{aligned}$$

The hypotheses 3. and 4. in Assumption 1 are fulfilled for

$$\ell_{1,+} = \ell_{2,+} = \ell_{3,+} := L\sqrt{2}.$$

Upper estimates for the limiting subgradients

Let Assumption 2 be satisfied. Then for any $n \geq 1$ it holds

$$D_n := (d_x^n, d_y^n, d_z^n, d_u^n, d_{x'}^n, d_{u'}^n) \in \partial\Psi(\mathbf{X}_n),$$

where

$$\begin{aligned}d_x^n &:= \nabla_x h(x_n, y_n) + A^T u_n + \beta A^T (Ax_n - z_n) + 2C_1 (x_n - x_{n-1}) \\&\quad + 2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})), \\d_y^n &:= \nabla_y h(x_n, y_n) - \nabla_y h(x_{n-1}, y_{n-1}) + \mu (y_{n-1} - y_n), \\d_z^n &:= u_{n-1} - u_n + \beta A (x_{n-1} - x_n), \\d_u^n &:= Ax_n - z_n + 2C_0 A (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})), \\d_{x'}^n &:= -2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})) - 2C_1 (x_n - x_{n-1}), \\d_{u'}^n &:= -2C_0 A (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})).\end{aligned}$$

In addition, for any $n \geq 1$ it holds

$$\|D_n\| \leq C_5 \|x_n - x_{n-1}\| + C_6 \|y_n - y_{n-1}\| + C_7 \|u_n - u_{n-1}\|,$$

where

$$\begin{aligned}C_5 &:= 2\sqrt{2} \cdot L + \tau + \beta \|A\| + 4(\sigma\tau + \|A\|)\sigma\tau C_0 + 4C_1, \quad C_6 := L\sqrt{2} + \mu, \\C_7 &:= 1 + \frac{1}{\sigma\beta} + \left(\frac{2}{\sigma} - 1\right) \|A\| + 4(\sigma\tau + \|A\|) C_0 \|A\|.\end{aligned}$$

Upper estimates for the limiting subgradients

Let Assumption 2 be satisfied. Then for any $n \geq 1$ it holds

$$D_n := (d_x^n, d_y^n, d_z^n, d_u^n, d_{x'}^n, d_{u'}^n) \in \partial\Psi(\mathbf{X}_n),$$

where

$$\begin{aligned} d_x^n &:= \nabla_x h(x_n, y_n) + A^T u_n + \beta A^T (Ax_n - z_n) + 2C_1 (x_n - x_{n-1}) \\ &\quad + 2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})), \\ d_y^n &:= \nabla_y h(x_n, y_n) - \nabla_y h(x_{n-1}, y_{n-1}) + \mu (y_{n-1} - y_n), \\ d_z^n &:= u_{n-1} - u_n + \beta A (x_{n-1} - x_n), \\ d_u^n &:= Ax_n - z_n + 2C_0 A (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})), \\ d_{x'}^n &:= -2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})) - 2C_1 (x_n - x_{n-1}), \\ d_{u'}^n &:= -2C_0 A (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})). \end{aligned}$$

In addition, for any $n \geq 1$ it holds

$$\|D_n\| \leq C_5 \|x_n - x_{n-1}\| + C_6 \|y_n - y_{n-1}\| + C_7 \|u_n - u_{n-1}\|,$$

where

$$\begin{aligned} C_5 &:= 2\sqrt{2} \cdot L + \tau + \beta \|A\| + 4(\sigma\tau + \|A\|)\sigma\tau C_0 + 4C_1, \quad C_6 := L\sqrt{2} + \mu, \\ C_7 &:= 1 + \frac{1}{\sigma\beta} + \left(\frac{2}{\sigma} - 1\right) \|A\| + 4(\sigma\tau + \|A\|) C_0 \|A\|. \end{aligned}$$

- Denote by $\Omega := \Omega(\{\mathbf{X}_n\}_{n \geq 1})$ the **set of cluster points** of the sequence $\{\mathbf{X}_n\}_{n \geq 1}$.

Convergence results II

Let Assumption 2 be satisfied. The following statements are true:

1. if $\{(x_{n_k}, y_{n_k}, z_{n_k}, u_{n_k})\}_{k \geq 0}$ is a subsequence of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ which converges to (x_*, y_*, z_*, u_*) as $k \rightarrow +\infty$, then

$$\lim_{k \rightarrow +\infty} \Psi_{n_k} = \Psi(x_*, y_*, z_*, u_*, x_*, u_*);$$

2. it holds

$$\begin{aligned} \Omega \subseteq \text{crit}(\Psi) \subseteq \{ \mathbf{X}_* \in \mathbf{R} : -A^T u_* = \nabla_x h(x_*, y_*), u_* \in \partial f(z_*), \\ 0 \in \partial g(y_*) + \nabla_y h(x_*, y_*), z_* = Ax_* \}; \end{aligned}$$

3. it holds $\lim_{n \rightarrow +\infty} \text{dist}(\mathbf{X}_n, \Omega) = 0$;
4. the set Ω is nonempty, connected and compact;
5. the function Ψ takes on Ω the value

$$\Psi_* = \lim_{n \rightarrow +\infty} \Psi_n = \lim_{n \rightarrow +\infty} \{f(z_n) + g(y_n) + h(x_n, y_n)\}.$$

KKT point

- An element (x_*, y_*, z_*, u_*) fulfilling

$$-A^T u_* = \nabla_x h(x_*, y_*), u_* \in \partial f(z_*), 0 \in \partial g(y_*) + \nabla_y h(x_*, y_*), z_* = Ax_*.$$

- It holds

$$0 \in A^T \partial f(Ax_*) + \nabla_x h(x_*, y_*), \quad 0 \in \partial g(y_*) + \nabla_y h(x_*, y_*)$$

and for A injective this is equivalent to

$$\begin{aligned} 0 &\in \partial(f \circ A)(x_*) + \nabla_x h(x_*, y_*) = \partial_x(f \circ A + h)(x_*, y_*), \\ 0 &\in \partial g(y_*) + \nabla_y h(x_*, y_*) = \partial_y(g + h)(x_*, y_*), \end{aligned}$$

in other words, (x_*, y_*) is a **critical point** of the optimization problem (1).

KKT point

- An element (x_*, y_*, z_*, u_*) fulfilling

$$-A^T u_* = \nabla_x h(x_*, y_*), u_* \in \partial f(z_*), 0 \in \partial g(y_*) + \nabla_y h(x_*, y_*), z_* = Ax_*.$$

- It holds

$$0 \in A^T \partial f(Ax_*) + \nabla_x h(x_*, y_*), 0 \in \partial g(y_*) + \nabla_y h(x_*, y_*)$$

and for A injective this is equivalent to

$$\begin{aligned} 0 &\in \partial(f \circ A)(x_*) + \nabla_x h(x_*, y_*) = \partial_x(f \circ A + h)(x_*, y_*), \\ 0 &\in \partial g(y_*) + \nabla_y h(x_*, y_*) = \partial_y(g + h)(x_*, y_*), \end{aligned}$$

in other words, (x_*, y_*) is a **critical point** of the optimization problem (1).

Global convergence under Kurdyka-Łojasiewicz assumptions

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

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

such that

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

for every

$$y \in U \cap \{z \in \mathbb{R}^m : h(x) < h(z) < h(x) + \eta\}.$$

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

Łojasiewicz (1963)

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

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

Thus, (3) is fulfilled for $\varphi(s) = \frac{1}{1-\theta} C s^{1-\theta}$ and every

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

Global convergence under Kurdyka-Łojasiewicz assumptions

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

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

such that

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

for every

$$y \in U \cap \{z \in \mathbb{R}^m : h(x) < h(z) < h(x) + \eta\}.$$

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

Łojasiewicz (1963)

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

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

Thus, (3) is fulfilled for $\varphi(s) = \frac{1}{1-\theta} C s^{1-\theta}$ and every

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

Global convergence of the iterates

If Assumption 2 is satisfied and Ψ is a **KL function**, then:

1. the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ has finite length, namely,

$$\sum_{n \geq 0} \|x_{n+1} - x_n\| < +\infty, \quad \sum_{n \geq 0} \|y_{n+1} - y_n\| < +\infty,$$

$$\sum_{n \geq 0} \|z_{n+1} - z_n\| < +\infty, \quad \sum_{n \geq 0} \|u_{n+1} - u_n\| < +\infty;$$

2. the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ **converges to a KKT point** of the optimization problem (1).

Convergence rates for the regularized augmented Lagrangian

Let Assumption 2 be satisfied. If Ψ satisfies the **Łojasiewicz property** with Łojasiewicz exponent $\theta \in [0, 1)$, then the following statements are true:

1. if $\theta = 0$, then $\{\Psi_n\}_{n \geq 1}$ converges in finite time;
2. if $\theta \in (0, 1/2]$, then there exist $n_0 \geq 1$, $\widehat{C}_0 > 0$ and $Q \in [0, 1)$ such that for any $n \geq n_0$

$$0 \leq \Psi_n - \Psi_* \leq \widehat{C}_0 Q^n;$$

3. if $\theta \in (1/2, 1)$, then there exist $n_0 \geq 1$ and $\widehat{C}_1 > 0$ such that for any $n \geq n_0 + 1$

$$0 \leq \Psi_n - \Psi_* \leq \widehat{C}_1 n^{-\frac{1}{2\theta-1}}.$$

Global convergence of the iterates

If Assumption 2 is satisfied and Ψ is a **KL function**, then:

1. the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ has finite length, namely,

$$\sum_{n \geq 0} \|x_{n+1} - x_n\| < +\infty, \sum_{n \geq 0} \|y_{n+1} - y_n\| < +\infty,$$
$$\sum_{n \geq 0} \|z_{n+1} - z_n\| < +\infty, \sum_{n \geq 0} \|u_{n+1} - u_n\| < +\infty;$$

2. the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ **converges to a KKT point** of the optimization problem (1).

Convergence rates for the regularized augmented Lagrangian

Let Assumption 2 be satisfied. If Ψ satisfies the **Łojasiewicz property** with Łojasiewicz exponent $\theta \in [0, 1)$, then the following statements are true:

1. if $\theta = 0$, then $\{\Psi_n\}_{n \geq 1}$ converges in finite time;
2. if $\theta \in (0, 1/2]$, then there exist $n_0 \geq 1$, $\widehat{C}_0 > 0$ and $Q \in [0, 1)$ such that for any $n \geq n_0$

$$0 \leq \Psi_n - \Psi_* \leq \widehat{C}_0 Q^n;$$

3. if $\theta \in (1/2, 1)$, then there exist $n_0 \geq 1$ and $\widehat{C}_1 > 0$ such that for any $n \geq n_0 + 1$

$$0 \leq \Psi_n - \Psi_* \leq \widehat{C}_1 n^{-\frac{1}{2\theta-1}}.$$

Convergence rates for the iterates

Let Assumption 2 be satisfied. Suppose further that Ψ satisfies the **Łojasiewicz property** with Łojasiewicz exponent $\theta \in [0, 1)$. Let (x_*, y_*, z_*, u_*) be the KKT point of the optimization problem (1) to which $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ converges as $n \rightarrow +\infty$. Then the following statements are true:

1. if $\theta = 0$, then the algorithm converges in finite time;
2. if $\theta \in (0, 1/2]$, then there exist $n_0 \geq 1$, $\widehat{C}_{0,1}, \widehat{C}_{0,2}, \widehat{C}_{0,3}, \widehat{C}_{0,4} > 0$ and $\widehat{Q} \in [0, 1)$ such that for any $n \geq n_0$

$$\begin{aligned}\|x_n - x_*\| &\leq \widehat{C}_{0,1} \widehat{Q}^k, & \|y_n - y_*\| &\leq \widehat{C}_{0,2} \widehat{Q}^k, \\ \|z_n - z_*\| &\leq \widehat{C}_{0,3} \widehat{Q}^k, & \|u_n - u_*\| &\leq \widehat{C}_{0,4} \widehat{Q}^k;\end{aligned}$$

3. if $\theta \in (1/2, 1)$, then there exist $n_0 \geq 1$ and $\widehat{C}_{1,1}, \widehat{C}_{1,2}, \widehat{C}_{1,3}, \widehat{C}_{1,4} > 0$ such that for any $n \geq n_0 + 1$

$$\begin{aligned}\|x_n - x_*\| &\leq \widehat{C}_{1,1} n^{-\frac{1-\theta}{2\theta-1}}, & \|y_n - y_*\| &\leq \widehat{C}_{1,2} n^{-\frac{1-\theta}{2\theta-1}}, \\ \|z_n - z_*\| &\leq \widehat{C}_{1,3} n^{-\frac{1-\theta}{2\theta-1}}, & \|u_n - u_*\| &\leq \widehat{C}_{1,4} n^{-\frac{1-\theta}{2\theta-1}}.\end{aligned}$$

- Open question 3: Is there a generic class of nonconvex functions which fulfills the Kurdyka-Łojasiewicz property in infinite-dimensional Hilbert spaces?

Here is another source of nonconvexity ...

Composition with a nonlinear operator

Let be

$$\min_{x \in \mathbb{R}^m} \{f(Kx) + g(x) + h(x)\},$$

where

- ▶ $h: \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a **convex and Fréchet differentiable function with Lipschitz-continuous gradient**;
- ▶ $f: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are **proper, convex and lower semicontinuous functions**;
- ▶ $K: \mathbb{R}^m \rightarrow \mathbb{R}^p$ is a (possibly nonlinear) continuously differentiable operator.

Algorithm

Let $\beta, \tau > 0$. For a given starting point $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^p$, let be for any $n \geq 0$

$$\bar{y}_{n+1} = \text{prox}_{\beta g^*}(y_n + \beta K(x_n))$$

$$x_{n+1} = \text{prox}_{\tau^{-1}f}\left(x_n - \frac{1}{\tau}(\nabla h(x_n) + (\nabla K(x_n))^* \bar{y}_{n+1})\right)$$

$$y_{n+1} := \bar{y}_{n+1} + \beta \nabla K(x_n)(x_{n+1} - x_n).$$

▶ **Open question 4:** Does anybody have an idea on how to prove that this algorithm converges? ☺

Here is another source of nonconvexity ...

Composition with a nonlinear operator

Let be

$$\min_{x \in \mathbb{R}^m} \{f(Kx) + g(x) + h(x)\},$$

where

- ▶ $h: \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a **convex and Fréchet differentiable function with Lipschitz-continuous gradient**;
- ▶ $f: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are **proper, convex and lower semicontinuous functions**;
- ▶ $K: \mathbb{R}^m \rightarrow \mathbb{R}^p$ is a (possibly nonlinear) continuously differentiable operator.

Algorithm

Let $\beta, \tau > 0$. For a given starting point $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^p$, let be for any $n \geq 0$

$$\bar{y}_{n+1} = \text{prox}_{\beta g^*}(y_n + \beta K(x_n))$$

$$x_{n+1} = \text{prox}_{\tau^{-1}f}\left(x_n - \frac{1}{\tau}(\nabla h(x_n) + (\nabla K(x_n))^* \bar{y}_{n+1})\right)$$

$$y_{n+1} := \bar{y}_{n+1} + \beta \nabla K(x_n)(x_{n+1} - x_n).$$

▶ **Open question 4:** Does anybody have an idea on how to prove that this algorithm converges? ☺

Here is another source of nonconvexity ...

Composition with a nonlinear operator

Let be

$$\min_{x \in \mathbb{R}^m} \{f(Kx) + g(x) + h(x)\},$$

where

- ▶ $h: \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a **convex and Fréchet differentiable function with Lipschitz-continuous gradient**;
- ▶ $f: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are **proper, convex and lower semicontinuous functions**;
- ▶ $K: \mathbb{R}^m \rightarrow \mathbb{R}^p$ is a (possibly nonlinear) continuously differentiable operator.

Algorithm

Let $\beta, \tau > 0$. For a given starting point $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^p$, let be for any $n \geq 0$

$$\bar{y}_{n+1} = \text{prox}_{\beta g^*}(y_n + \beta K(x_n))$$

$$x_{n+1} = \text{prox}_{\tau^{-1}f}\left(x_n - \frac{1}{\tau}(\nabla h(x_n) + (\nabla K(x_n))^* \bar{y}_{n+1})\right)$$

$$y_{n+1} := \bar{y}_{n+1} + \beta \nabla K(x_n)(x_{n+1} - x_n).$$

▶ **Open question 4:** Does anybody have an idea on how to prove that this algorithm converges? 😊

Bonus material: D. C. programming

Consider the optimization problem

$$\min \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} \quad (4)$$

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

Toland dual problem

$$\min \{h^*(y) - (g + \varphi)^*(K^*y) \mid y \in \mathcal{G}\}. \quad (5)$$

Primal-dual formulation

$$\min \{\Phi(x, y) \mid x \in \mathcal{H}, y \in \mathcal{G}\}, \quad (6)$$

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

Φ is **proper** and **lower semicontinuous**.

Bonus material: D. C. programming

Consider the optimization problem

$$\min \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} \quad (4)$$

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

Toland dual problem

$$\min \{h^*(y) - (g + \varphi)^*(K^*y) \mid y \in \mathcal{G}\}. \quad (5)$$

Primal-dual formulation

$$\min \{\Phi(x, y) \mid x \in \mathcal{H}, y \in \mathcal{G}\}, \quad (6)$$

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

Φ is **proper** and **lower semicontinuous**.

Bonus material: D. C. programming

Consider the optimization problem

$$\min \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} \quad (4)$$

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

Toland dual problem

$$\min \{h^*(y) - (g + \varphi)^*(K^*y) \mid y \in \mathcal{G}\}. \quad (5)$$

Primal-dual formulation

$$\min \{\Phi(x, y) \mid x \in \mathcal{H}, y \in \mathcal{G}\}, \quad (6)$$

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

Φ is **proper** and **lower semicontinuous**.

Proposition

1. The optimal values of (4), (5) and (6) 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. If $\bar{x} \in \mathcal{H}$ be an **optimal solution** of (4), then $\bar{x} \in \mathcal{H}$ is a **strong critical point** of (4):

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

4. If $\bar{y} \in \mathcal{G}$ is an **optimal solution** of (5), then $\bar{y}$ is a **strong critical point** of (5):

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

5. If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is **optimal solution** of (6), then $\bar{x}$ is an **optimal solution** of (4), and $\bar{y}$ is an **optimal solution** of (5). On the other hand,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}), \tag{7}$$

$$K\bar{x} \in \partial h^*(\bar{y}). \tag{8}$$

Proposition

1. The optimal values of (4), (5) and (6) 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. If $\bar{x} \in \mathcal{H}$ be an **optimal solution** of (4), then $\bar{x} \in \mathcal{H}$ is a **strong critical point** of (4):

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

4. If $\bar{y} \in \mathcal{G}$ is an **optimal solution** of (5), then $\bar{y}$ is a **strong critical point** of (5):

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

5. If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is **optimal solution** of (6), then $\bar{x}$ is an **optimal solution** of (4), and $\bar{y}$ is an **optimal solution** of (5). On the other hand,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}), \tag{7}$$

$$K\bar{x} \in \partial h^*(\bar{y}). \tag{8}$$

Proposition

1. The optimal values of (4), (5) and (6) 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. If $\bar{x} \in \mathcal{H}$ be an **optimal solution** of (4), then $\bar{x} \in \mathcal{H}$ is a **strong critical point** of (4):

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

4. If $\bar{y} \in \mathcal{G}$ is an **optimal solution** of (5), then $\bar{y}$ is a **strong critical point** of (5):

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

5. If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is **optimal solution** of (6), then $\bar{x}$ is an **optimal solution** of (4), and $\bar{y}$ is an **optimal solution** of (5). On the other hand,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}), \tag{7}$$

$$K\bar{x} \in \partial h^*(\bar{y}). \tag{8}$$

Proposition

1. The optimal values of (4), (5) and (6) 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. If $\bar{x} \in \mathcal{H}$ be an **optimal solution** of (4), then $\bar{x} \in \mathcal{H}$ is a **strong critical point** of (4):

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

4. If $\bar{y} \in \mathcal{G}$ is an **optimal solution** of (5), then $\bar{y}$ is a **strong critical point** of (5):

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

5. If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is **optimal solution** of (6), then $\bar{x}$ is an **optimal solution** of (4), and $\bar{y}$ is an **optimal solution** of (5). On the other hand,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}), \tag{7}$$

$$K\bar{x} \in \partial h^*(\bar{y}). \tag{8}$$

Proposition

1. The optimal values of (4), (5) and (6) 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. If $\bar{x} \in \mathcal{H}$ be an **optimal solution** of (4), then $\bar{x} \in \mathcal{H}$ is a **strong critical point** of (4):

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

4. If $\bar{y} \in \mathcal{G}$ is an **optimal solution** of (5), then $\bar{y}$ is a **strong critical point** of (5):

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

5. If $(\bar{x}, \bar{y}) \in \mathcal{H} \times \mathcal{G}$ is **optimal solution** of (6), then $\bar{x}$ is an **optimal solution** of (4), and $\bar{y}$ is an **optimal solution** of (5). On the other hand,

$$K^* \bar{y} \in \partial g(\bar{x}) + \nabla \varphi(\bar{x}), \tag{7}$$

$$K\bar{x} \in \partial h^*(\bar{y}). \tag{8}$$

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 (6) if

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

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

Weak critical points of $g + \varphi - h \circ K$

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

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

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

It holds

$$\text{crit}\Phi \subseteq \text{crit}(g + \varphi - h \circ K) \times \text{crit}(h^* - (g + \varphi)^* \circ K^*).$$

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 (6) if

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

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

Weak critical points of $g + \varphi - h \circ K$

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

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

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

It holds

$$\text{crit}\Phi \subseteq \text{crit}(g + \varphi - h \circ K) \times \text{crit}(h^* - (g + \varphi)^* \circ K^*).$$

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 (6) if

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

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

Weak critical points of $g + \varphi - h \circ K$

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

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

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

It holds

$$\text{crit}\Phi \subseteq \text{crit}(g + \varphi - h \circ K) \times \text{crit}(h^* - (g + \varphi)^* \circ K^*).$$

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 (6) if

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

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

Weak critical points of $g + \varphi - h \circ K$

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

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

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

It holds

$$\text{crit}\Phi \subseteq \text{crit}(g + \varphi - h \circ K) \times \text{crit}(h^* - (g + \varphi)^* \circ K^*).$$

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

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

Two descent inequalities and consequences

► For all $n \geq 0$

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

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

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

$$0 < \inf_{n \geq 0} \gamma_n \leq \sup_{n \geq 0} \gamma_n < \frac{2}{L \nabla \varphi} \quad \text{and} \quad 0 < \inf_{n \geq 0} \mu_n \leq \sup_{n \geq 0} \mu_n < +\infty. \quad (9)$$

Then,

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

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

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

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

Two descent inequalities and consequences

► For all $n \geq 0$

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

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

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

$$0 < \inf_{n \geq 0} \gamma_n \leq \sup_{n \geq 0} \gamma_n < \frac{2}{L_{\nabla \varphi}} \quad \text{and} \quad 0 < \inf_{n \geq 0} \mu_n \leq \sup_{n \geq 0} \mu_n < +\infty. \quad (9)$$

Then,

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

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

Proposition

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (9) 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 **weak critical point** of (4),
2. every cluster point of $(y_n)_{n \geq 0}$ is a **weak critical point** of (5)
3. every cluster point of $(x_n, y_n)_{n \geq 0}$ is a **critical point** of (6).

Proposition

Let (9) 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 (9) 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. the objective function Φ is finite and constant on $\omega(x_0, y_0)$ provided that the optimal value is finite.

Proposition

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (9) 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 **weak critical point** of (4),
2. every cluster point of $(y_n)_{n \geq 0}$ is a **weak critical point** of (5)
3. every cluster point of $(x_n, y_n)_{n \geq 0}$ is a **critical point** of (6).

Proposition

Let (9) 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 (9) 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. the objective function Φ is finite and constant on $\omega(x_0, y_0)$ provided that the optimal value is finite.

Proposition

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (9) 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 **weak critical point** of (4),
2. every cluster point of $(y_n)_{n \geq 0}$ is a **weak critical point** of (5)
3. every cluster point of $(x_n, y_n)_{n \geq 0}$ is a **critical point** of (6).

Proposition

Let (9) 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 (9) 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. the objective function Φ is finite and constant on $\omega(x_0, y_0)$ provided that the optimal value is finite.

Theorem (convergence result when Φ is a KL function)

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (9) be satisfied. 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** $s \mapsto \frac{1}{1-\theta}Cs^{1-\theta}$ 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

$$\|x_n - \bar{x}\| \leq cq^n \quad \text{and} \quad \|y_n - \bar{y}\| \leq cq^n$$

for all $n \geq 0$;

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

$$\|x_n - \bar{x}\| \leq cn^{-\frac{1-\theta}{2\theta-1}} \quad \text{and} \quad \|y_n - \bar{y}\| \leq cn^{-\frac{1-\theta}{2\theta-1}}$$

for all $n \geq 0$.

Theorem (convergence result when Φ is a KL function)

Let $\inf \{g(x) + \varphi(x) - h(Kx) \mid x \in \mathcal{H}\} > -\infty$ and (9) be satisfied. 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

$$\|x_n - \bar{x}\| \leq cq^n \quad \text{and} \quad \|y_n - \bar{y}\| \leq cq^n$$

for all $n \geq 0$;

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

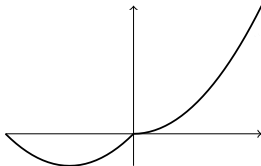
$$\|x_n - \bar{x}\| \leq cn^{-\frac{1-\theta}{2\theta-1}} \quad \text{and} \quad \|y_n - \bar{y}\| \leq cn^{-\frac{1-\theta}{2\theta-1}}$$

for all $n \geq 0$.

An example

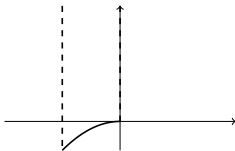
► Primal program

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

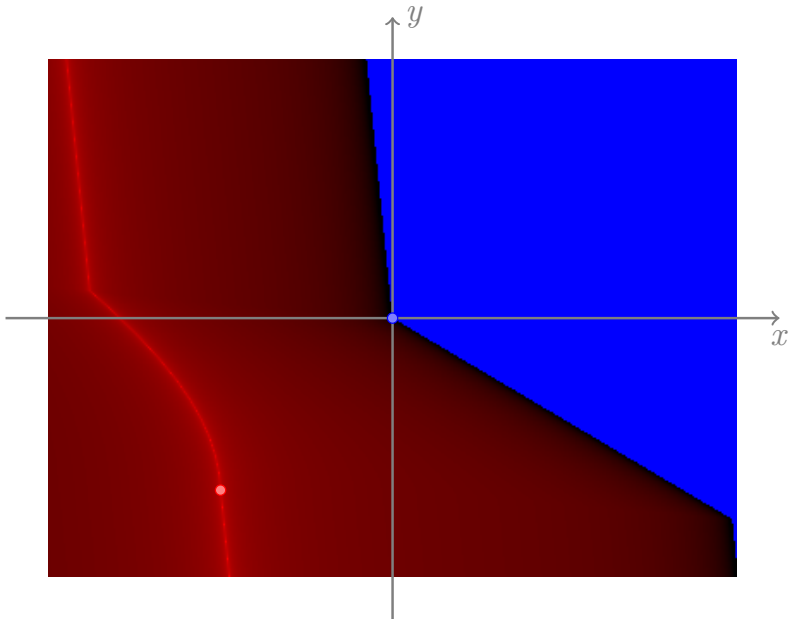


► Dual program

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



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



- Open question 5: Is there a modification of the double proximal-gradient algorithm which converges to a primal-dual pair of strong critical points?

- Open question 5: Is there a modification of the double proximal-gradient algorithm which converges to a primal-dual pair of strong critical points?

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 $A : \mathbb{R}^{mn} \rightarrow \mathbb{R}^{mn}$ and corrupted with **noise** ν . Knowing $b = Ax + \nu$, we want to reconstruct the original image x by considering the minimization problem (Lou, Zeng, Osher, Xin, SIIMAG 2009)

$$\min_{x \in \mathbb{R}^{mn}} \left(\frac{\mu}{2} \|Ax - b\|^2 + \|Dx\|_{\ell^1} - a \|Dx\|_{\times} \right),$$

where $\mu > 0$ is a **regularization parameter**, $a > 0$, $D : \mathbb{R}^{mn} \rightarrow \mathbb{R}^{2mn}$ is the **discrete gradient operator** given by $Dx = (D_1x, D_2x)$,

$$D_1 : \mathbb{R}^{mn} \rightarrow \mathbb{R}^{mn}, (D_1x)_{i,j} := \begin{cases} x_{i+1,j} - x_{i,j}, & i = 1, \dots, m-1; j = 1, \dots, n; \\ 0, & i = m; j = 1, \dots, n \end{cases}$$
$$D_2 : \mathbb{R}^{mn} \rightarrow \mathbb{R}^{mn}, (D_2x)_{i,j} := \begin{cases} x_{i,j+1} - x_{i,j}, & i = 1, \dots, m; j = 1, \dots, n-1; \\ 0, & i = 1, \dots, m; j = n, \end{cases}$$

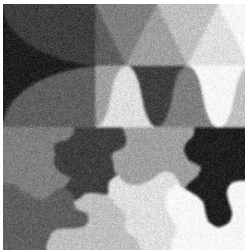
and

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

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



(b) Original image



(c) Blurry image

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

	$a = 0.00$	$a = 0.2$	$a = 0.4$	$a = 0.5$	$a = 0.6$	$a = 0.8$	$a = 1.0$
$\mu = 1.0$	-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 after 50 iterations



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



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



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

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*, SIAM Journal on Optimization, 2019



R.I. Boţ, D.-K. Nguyen, *The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates*, arXiv:1801.01994, 2018