A nonlinear primal–dual extragradient method for nonsmooth PDE-constrained optimization

Christian Clason[1] Tuomo Valkonen[2]

[1]Faculty of Mathematics, Universität Duisburg-Essen
[2]Department of Mathematical Sciences, University of Liverpool

International Conference on Continuous Optimization
Tokyo, August 10, 2016

Motivation

Primal-dual extragradient method:

- first-order algorithm for nonsmooth convex problems with linear operators [Chambolle/Pock 2011]
- *very* popular in imaging (TV denoising, deblurring, ...)
- acceleration (Nesterov, $\mathcal{O}(1/k^2)$ convergence)
- version for nonlinear operators [Valkonen 2014]

Here:

- application to parameter identification for PDEs
- $\rightsquigarrow$ function space algorithm

Difficulty:

- convergence proof requires set-valued analysis in infinite-dimensional spaces

Model problems

L^1-fitting

$$\min_u \frac{1}{\alpha}\|S(u) - y^\delta\|_{L^1} + \frac{1}{2}\|u\|_{L^2}^2$$

State constraints

$$\min_{u\in L^2(\Omega)} \frac{1}{2}\|S(u) - y^d\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L^2}^2 \quad \text{s. t.} \quad S(u)(x) \leqslant M \quad \text{a. e. in } \Omega$$

$S : U \subset L^2(\Omega) \to L^2(\Omega), \quad S(u) =: y$ satisfies

$$\begin{cases} -\Delta y + uy = f \\ \partial_\nu y = 0 \end{cases}$$

Problem statement

$$\min_{u \in X} F(K(u)) + G(u)$$

- $F : Y \to \overline{\mathbb{R}}, G : X \to \overline{\mathbb{R}}$ convex, lower semicontinuous
- X, Y Hilbert spaces
- $K \in C^2(X, Y)$ (here: $K(u) = S(u) - y^\delta$ or $K(u) = S(u)$)
- saddle point formulation:

$$\min_{u \in X} \sup_{v \in Y^*} G(u) + \langle K(u), v \rangle - F^*(v)$$

- $F^* : Y^* \to \overline{\mathbb{R}}$ Fenchel conjugate

Algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

K linear:

$$\begin{cases} u^{k+1} = \text{prox}_{\tau G}(u^k - \tau K^* v^k) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k \\ v^{k+1} = \text{prox}_{\sigma F^*}(v^k + \sigma K \bar{u}^{k+1}) \end{cases}$$

- σ, τ step sizes, $\quad \sigma\tau < \|K\|^{-2}$
- $\text{prox}_{\sigma F}(v) = \arg\min_u \frac{1}{2\sigma}\|u - v\|^2 + F(u)$ proximal mapping

Algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

K nonlinear:

$$\begin{cases} u^{k+1} = \mathrm{prox}_{\tau G}(u^k - \tau K'(u^k)^* v^k) \\ \bar{u}^{k+1} = 2u^{k+1} - u^k \\ v^{k+1} = \mathrm{prox}_{\sigma F^*}(v^k + \sigma K(\bar{u}^{k+1})) \end{cases}$$

- σ, τ step sizes, $\sigma\tau < \sup_{u \in B_R} \|K'(u)\|^{-2}$
- $\mathrm{prox}_{\sigma F}(v) = \arg\min_u \frac{1}{2\sigma}\|u - v\|^2 + F(u)$ proximal mapping
- $K'(u)$ Fréchet derivative, $K'(u)^*$ adjoint

Algorithm

K nonlinear, accelerated:

$$\begin{cases} u^{k+1} = \text{prox}_{\tau_k G}(u^k - \tau^k K'(u^k)^* v^k) \\ \omega_k = 1/\sqrt{1+2c\tau^k} \quad \tau^{k+1} = \omega_k \tau^k \quad \sigma_{k+1} = \sigma_k/\omega_k \\ \bar{u}^{k+1} = u^{k+1} + \omega_k(u^{k+1} - u^k) \\ v^{k+1} = \text{prox}_{\sigma_{k+1}F^*}(v^k + \sigma_{k+1}K(\bar{u}^{k+1})) \end{cases}$$

- σ, τ step sizes, $\sigma_0\tau_0 < \sup_{u\in B_R} \|K'(u)\|^{-2}$
- $\text{prox}_{\sigma F}(v) = \arg\min_u \frac{1}{2\sigma}\|u-v\|^2 + F(u)$ proximal mapping
- $K'(u)$ Fréchet derivative, $K'(u)^*$ adjoint
- $c \geqslant 0$ acceleration parameter

Convergence

Theorem

Iterates converge locally to saddle point $(\bar{u}, \bar{v})$ if

1. *G is c_G-strongly convex* (here: $c_G = 1$)
2. $c = c_n \in [0, c_G)$, $c_n = 0$ *for* $n > N \in \mathbb{N}$ (finite acceleration)
3. *metric regularity around saddle point*

(cf. [Valkonen 2014])

Difficulty:

- metric regularity in function spaces
- $\rightsquigarrow$ requires infinite-dimensional set-valued analysis
- here: only rough outline, no details

Metric regularity

Saddle-point problem

$$\min_{u\in X} \sup_{v\in Y^*} G(u) + \langle K(u), v\rangle - F^*(v)$$

Primal-dual optimality conditions

$$\begin{cases} K(\bar u) \in \partial F^*(\bar v) \\ -K'(\bar u)^*\bar v \in \partial G(\bar u) \end{cases}$$

Metric regularity

Primal-dual optimality conditions

$$\begin{cases} K(\bar{u}) \in \partial F^*(\bar{v}) \\ -K'(\bar{u})^*\bar{v} \in \partial G(\bar{u}) \end{cases}$$

Set inclusion for $H : L^2(\Omega)^2 \rightrightarrows L^2(\Omega)^2$

$$0 \in H_{\bar{u}}(\bar{u}, \bar{v}) := \begin{pmatrix} \partial G(\bar{u}) + K'(\bar{u})^*\bar{v} \\ \partial F^*(\bar{v}) - K(\bar{u}) \end{pmatrix}$$

Metric regularity

Set inclusion for $H : L^2(\Omega)^2 \rightrightarrows L^2(\Omega)^2$

$$0 \in H_{\bar u}(\bar u, \bar v) := \begin{pmatrix} \partial G(\bar u) + K'(\bar u)^* \bar v \\ \partial F^*(\bar v) - K(\bar u) \end{pmatrix}$$

Metric regularity at $(\bar u, \bar v)$

$$\inf_{q:\, w \in H_{\bar u}(q)} \|q - (\bar u, \bar v)\| \leqslant L\|w\| \quad \text{for all } \|w\| \leqslant \rho$$

- interpretation: small perturbation w of 0
  $\Rightarrow$ small perturbation q of saddle point $(\bar u, \bar v)$
- Lipschitz property for set-valued $H_{\bar u}^{-1}$ at $((\bar u, \bar v), 0)$

Metric regularity

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Metric regularity at $(\bar u, \bar v)$

$$\inf_{q:\, w\in H_{\bar u}(q)} \|q - (\bar u, \bar v)\| \leqslant L\|w\| \quad \text{for all } \|w\| \leqslant \rho$$

Mordukhovich criterion

$$L_H = \inf_{t>0} \sup \left\{ \|\widehat{D}^* H(q'|w')\| \,\middle|\, q' \in B((\bar u, \bar v), t),\ w' \in H(q') \cap B(w,t) \right\}$$

- Aubin constant L_H is minimal choice of L
- $\widehat{D}^* H$ regular coderivative of H (cf. $L = \|\nabla f\|$ for $f \in C^1$)
- $\rightsquigarrow$ set-valued analysis in function spaces

Set-valued analysis in function space

Difficulties:

- multiple non-equivalent concepts (regular, limiting)
- calculus not tight

Here:

- set-valued mappings from subdifferentials of pointwise functionals
- $\rightsquigarrow$ infinite-dimensional (regular) derivatives pointwise via nice finite-dimensional (regular, graphical) derivatives
- cf. pointwise Fenchel conjugates, subdifferentials [Ekeland]

Application to saddle point inclusion

$$0 \in H_{\bar{u}}(\bar{u},\bar{v}) = \begin{pmatrix} \partial G(\bar{u}) + K'(\bar{u})^* \bar{v} \\ \partial F^*(\bar{v}) - K(\bar{u}) \end{pmatrix}$$

$$DH_{\bar{u}}(q|w)(\Delta q) = \begin{pmatrix} D[\partial G](u|\xi - K'(\bar{u})^* v)(\Delta u) + K'(\bar{u})^* \Delta v \\ D[\partial F^*](v|\eta + K'(\bar{u})u + c_{\bar{u}})(\Delta v) - K'(\bar{u})\Delta u \end{pmatrix}$$

$$\widetilde{DH_{\bar{u}}}(q|w)(\Delta q) = \begin{cases} T_q \Delta q + V(q|w)^\circ & \Delta q \in V(q|w) \\ \emptyset & \text{otherwise} \end{cases}$$

- $q = (u,v), \quad w = (\xi,\eta), \quad c_{\bar{u}} = K(\bar{u}) - K'(\bar{u})\bar{u}$
- T_q linear Operator (independent of w), $\quad V(q|w)$ cone
- $\widehat{D}^* H_{\bar{u}}(q|w) = [\widetilde{DH_{\bar{u}}(q|w)}]^{*+}$ upper adjoint of convexification

Mordukhovich criterion

$$\widetilde{DH_{\bar{u}}}(q|w)(\Delta q) = \begin{cases} T_q \Delta q + V(q|w)^\circ & \Delta q \in V(q|w) \\ \emptyset & \text{otherwise} \end{cases}$$

Then: Aubin constant $L_H \leqslant c < \infty$ **iff**

$$\sup_{t>0} \inf_{\substack{(\Delta w, z) \in W^t(q|w), \\ \|\Delta w\| > 0}} \frac{\|T_q^* \Delta w - z\|}{\|\Delta w\|} \geqslant c^{-1} > 0$$

$$W^t(q|w) = \bigcup \left\{ V(q'|w') \times V(q'|w')^\circ \;\middle|\; \begin{array}{l} w' \in H_{\bar{u}}(q'), \\ \|q' - q\| < t, \; \|w' - w\| < t \end{array} \right\}$$

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Primal-dual optimality conditions

$$\begin{cases} S(\bar{u}) - y^\delta \in \partial F^*(\bar{v}) \\ -S'(\bar{u})^*\bar{v} = \bar{u} \end{cases}$$

Metric regularity around $(\bar{u}, \bar{v})$ if either

1. $\sup_{t>0} \inf \left\{ \frac{\|S'(\bar{u})S'(\bar{u})^*z - v\|}{\|z\|} \,\middle|\, (z, v) \in V^t_{\partial F^*}(\bar{v}|y^\delta - S(\bar{u})),\ z \neq 0 \right\} > 0$
2. Moreau–Yosida regularization: $F^* \mapsto F^*_\gamma := F^* + \frac{\gamma}{2}\|\cdot\|^2$
3. finite-dimensional data: $Y \rightsquigarrow Y_h$

In case 1: $\|S'(\bar{u})^*z\| \geqslant c\|z\|$ for $z \in V^t_{\partial F^*}(\bar{v}|y^\delta - S(\bar{u}))$ necessary!

L^1 fitting

$$\min_u \frac{1}{\alpha} \|S(u) - y^\delta\|_{L^1} + \frac{1}{2}\|u\|_{L^2}^2$$

- $F(y) = \int_\Omega \alpha^{-1}|y(x)|\,dx \quad \rightsquigarrow \quad f^* = \delta_{[-\alpha^{-1},\alpha^{-1}]}(z)$
- $S : U \subset L^2(\Omega) \to L^2(\Omega), \quad S(u) =: y$ satisfies

$$\begin{cases} -\Delta y + uy = f \\ \partial_\nu y = 0 \end{cases}$$

L^1 fitting: metric regularity

Here: $z \in V^t_{\partial F^*}(v|\eta)$ if

$$z(x) \in \begin{cases} \{0\} & |v'(x)| = \alpha^{-1} \text{ and } \eta'(x) \neq 0 \\ -\operatorname{sign} v'(x)[0,\infty) & |v'(x)| = \alpha^{-1} \text{ and } \eta'(x) = 0 \\ \mathbb{R} & |v'(x)| < \alpha^{-1} \text{ and } \eta'(x) = 0 \end{cases}$$

for some $\|v' - \bar v\| \leqslant t$, $\|\eta' - \bar\eta\| \leqslant t$

- S compact operator: $\|S'(\bar u)^* z\| \geqslant c\|z\|$ only holds for $z = 0$
- $\bar\eta = S(\bar u) - y^\delta, \quad \alpha \bar v \in \operatorname{sign} \bar\eta$
- $\rightsquigarrow$ in general not satisfied!

L^1 fitting: algorithm

$$\begin{cases} z^{k+1} = S'(u^k)^* v^k \\ u^{k+1} = \dfrac{1}{1+\tau_k}(u^k - \tau_k z^{k+1}) \\ \omega_k = 1/\sqrt{1+2c\tau_k} \quad \tau_{k+1} = \omega_k \tau_k \quad \sigma_{k+1} = \sigma_k/\omega_k \\ \bar{u}^{k+1} = u^{k+1} + \omega_k(u^{k+1} - u^k) \\ v^{k+1} = \operatorname{proj}_{[-\alpha^{-1},\alpha^{-1}]}\left(\frac{1}{1+\sigma_{k+1}\gamma}(v^k + \sigma_{k+1}(S(\bar{u}^{k+1}) - y^\delta))\right) \end{cases}$$

- $S'(u^k)^* v^k$ solution of adjoint equation
- proj_C pointwise projection on convex set $C \subset \mathbb{R}$
- Moreau–Yosida parameter $\gamma \geqslant 0$
- local convergence if $\gamma > 0$ or finite-dimensional

State constraints

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

$$\min_{u\in L^2(\Omega)} \frac{1}{2\alpha}\|S(u)-y^d\|_{L^2}^2 + \frac{1}{2}\|u\|_{L^2}^2 \quad \text{s. t.} \quad S(u)(x) \leqslant M \quad \text{a. e. in } \Omega$$

- $F(y) = \frac{1}{2\alpha}\|y-y^d\|_{L^2}^2 + \delta_{\{|y(x)|\leqslant M\}}(y)$
- $\rightsquigarrow f^*(x,z) = \begin{cases} Mz - \frac{1}{2\alpha}|M-y^d(x)|^2 & z > \alpha^{-1}(M-y^d(x)), \\ \frac{\alpha}{2}|z|^2 + zy^d(x) & z \leqslant \alpha^{-1}(M-y^d(x)). \end{cases}$
- $S : U \subset L^2(\Omega) \to L^2(\Omega)$ as before
- if strict complementarity ($\alpha v(x) \neq M - y^d(x)$): $\quad z(x) \neq 0$
  $\rightsquigarrow$ estimate not satisfied

State constraints: algorithm

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

$$
\begin{cases}
u^{k+1} = \frac{1}{1+\tau_k}(u^k - \tau_k S'(u^k)^* v^k) \\
\omega_k = 1/\sqrt{1+2\bar{\gamma}\tau^k} \quad \tau^{k+1} = \omega_k \tau^k \quad \sigma_{k+1} = \sigma_k/\omega_k \\
\bar{u}^{k+1} = u^{k+1} + \omega_k(u^{k+1} - u^k) \\
r^{k+1} = v^k + \sigma_{k+1} S(\bar{u}^{k+1}) \\
\chi^{k+1} = \left[\!\left[r^{k+1} > \frac{1+\sigma_{k+1}\gamma}{\alpha}(M - y^d) + \sigma_{k+1} M \right]\!\right] \\
v^{k+1} = \frac{1}{1+\sigma_{k+1}\gamma}\chi^{k+1}\left(r^{k+1} - \sigma_{k+1} c\right) + \frac{1}{1+\sigma_{k+1}(\alpha+\gamma)}(1-\chi^{k+1})\left(r^{k+1} - \sigma_{k+1} y^d\right)
\end{cases}
$$

- $[\![P]\!](x) = 1$ if $P(x)$ true, 0 else (*Iverson bracket*)
- local convergence if $\gamma > 0$ or finite-dimensional

L^1 fitting

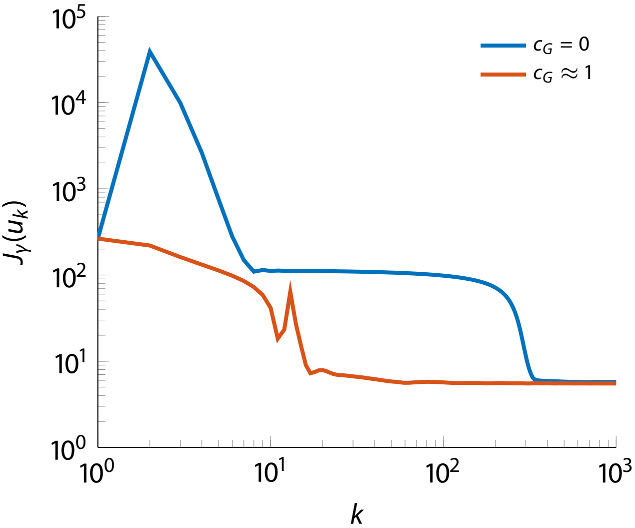
- $\Omega = [-1, 1]$, FE (P_1–P_0) discretization of (y, u)
- random impulsive noise:

$$y^\delta(x) = \begin{cases} y^\dagger(x) + \|y^\dagger\|_\infty \xi(x) & \text{with probability } 0.3 \\ y^\dagger(x) & \text{else} \end{cases}$$

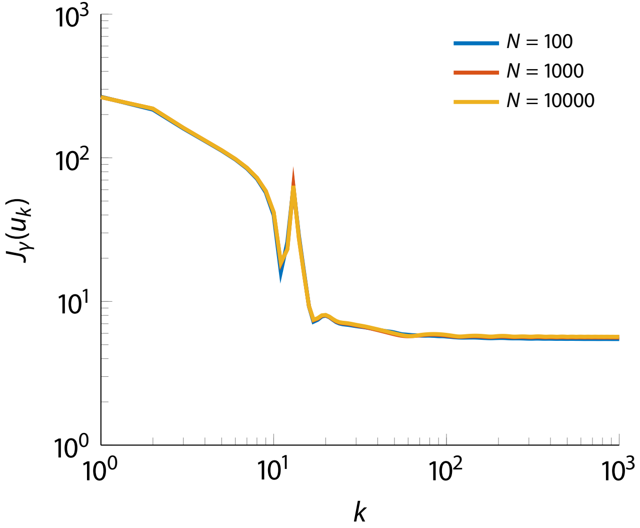
$y^\dagger = S(u^\dagger)$, $\xi(x) \in \mathcal{N}(0, 0.1)$, $\rightsquigarrow \alpha = 10^{-2}$

- $\sigma_0 = \tilde{L}^{-1}, \tau_0 = 0.99\tilde{L}^{-1}$, $\tilde{L} = \|S(u^0)\|/\|u^0\|$
- $\gamma = 10^{-12}$, $u^0 \equiv 1, v^0 = 0$ (no warmstart!)
- compare $c \in \{0, 1 - 10^{-16}\}$, $N \in \{100, 1000, 10000\}$

L^1 fitting: acceleration (same data)



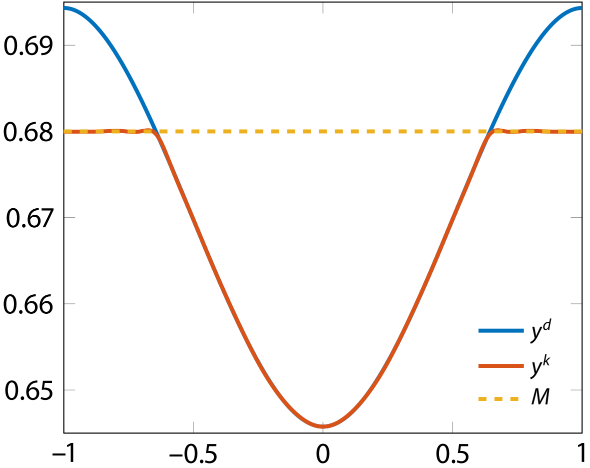
L^1 fitting: discretization (avg. of 10)



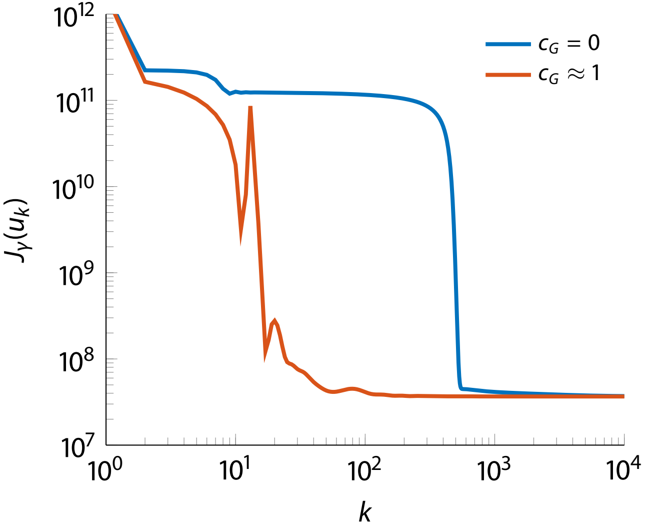
UNIVERSITÄT DUISBURG ESSEN
Open-Minded

- $\Omega = [-1, 1]$, FE (P_1–P_0) discretization of (y, u)
- $y^d = S(u^\dagger)$, $\alpha = 10^{-12}$, $M = 0.68 < y^d(x)$ for some $x \in \Omega$
- $\gamma = 10^{-12}$, $\sigma_0 = \tilde{L}^{-1}, \tau_0 = 0.99\tilde{L}^{-1}$, $\tilde{L} = \|S(u^0)\|/\|u^0\|$
- $u^0 \equiv 1, v^0 = 0$ (no warmstart!)
- compare $c \in \{0, 1 - 10^{-16}\}$, $N \in \{100, 1000, 10000\}$

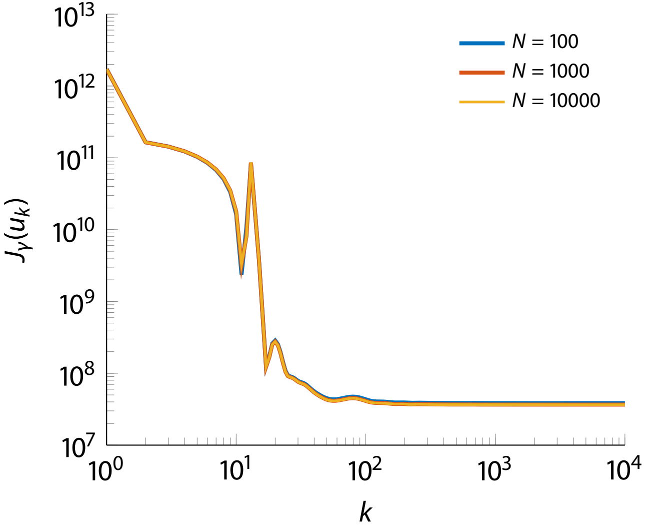
State constraints: target



State constraints: acceleration



State constraints: discretization



Conclusion

Primal-dual extragradient methods in function space:

- can be accelerated
- analyzed using set-valued analysis in function space
- requires Moreau–Yosida regularization
- $\rightsquigarrow$ no norm gap, continuation needed; mesh-independence

Outlook:

- full acceleration
- partial stability (w.r.t. primal variable only)
- other PDE-constrained optimization problems

Preprints/Code:

`http://www.uni-due.de/mathematik/agclason/clason_pub.php`