A primal-dual extragradient method for nonsmooth nonlinear inverse problems for PDEs

Christian Clason[1] Tuomo Valkonen[2]

[1]Faculty of Mathematics, Universität Duisburg-Essen
[2]Department of Applied Mathematics and Theoretical Physics, University of Cambridge

International Conference on Inverse Problems, Imaging, and Applications
Zhejiang University, Hangzhou, August 7, 2015

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:

- in function space
- $\rightsquigarrow$ parameter identification for PDEs

Difficulty:

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

Model problems

L^1-fitting

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

L^∞-fitting/Morozov

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

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

K linear:

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

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

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:

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

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

Convergence

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Theorem

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

1. *G is c_G-strongly convex (here: $c_G = \alpha$)*
2. *$c \in [0, c_G)$, $c = c_n = 0$ 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

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

$$\partial G(u) = \left\{ x^* \in X : \langle x^*, \tilde{x} - x \rangle \leqslant G(\tilde{x}) - G(x) \right\}$$ convex subdifferential

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 $((\bar{u}, \bar{v}), 0)$

Set-valued analysis

Metric regularity:

- related to graphical derivative of $H : L^2(\Omega)^2 \rightrightarrows L^2(\Omega)^2$
- $\rightsquigarrow$ set-valued analysis in infinite dimensions
- difficulties compared to finite dimensions:
    1. multiple non-equivalent concepts (Fréchet, Mordukhovich)
    2. calculus not tight

Here:

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

Examples: L^2 norm

$$F(u) = \int_\Omega f(u(x))\,dx, \qquad f(z) = \tfrac{1}{2}z^2, \qquad \partial f(z) = \{z\}$$

- Fréchet derivative of ∂f

$$D(\partial f)(z|\zeta)(\Delta z) = \begin{cases} \Delta z & \zeta = z \\ \emptyset & \text{otherwise} \end{cases}$$

- $\rightsquigarrow$ Fréchet derivative of ∂F

$$\left[D(\partial F)(u|\xi)(\Delta u)\right](x) = D\left(\partial f\left(u(x)|\xi(x)\right)\right)(\Delta u(x)) = \Delta u(x)$$

- $\rightsquigarrow$ Fréchet coderivative of ∂F

$$\widehat{D}^*(\partial F)(u|\xi)(\Delta\xi) = [D(\partial F)(u|\xi)]^{*+}(\Delta\xi) = \Delta\xi$$

Examples: indicator function

$$F(u) = \int_\Omega f(u(x))\,dx, \qquad f(z) = \iota_{[-1,1]}(z) = \begin{cases} 0 & |z| \leqslant 1 \\ \infty & |z| > 1 \end{cases}$$

- Fréchet derivative of ∂f

$$D(\partial f)(z|\zeta)(\Delta z) = \begin{cases} \{0\} & |z| < 1,\ \zeta = 0 \\ [0,\infty)z & |z| = 1,\ \zeta = 0,\ z\Delta z \leqslant 0 \\ \mathbb{R} & |z| = 1,\ \zeta \in (0,\infty),\ z\Delta z = 0 \\ \emptyset & \text{otherwise} \end{cases}$$

- $\rightsquigarrow$ Fréchet derivative of ∂F

$$D[\partial F](v|\eta)(\Delta v) = \begin{cases} V_{\partial F}(v|\eta)^\perp & \Delta v \in V_{\partial F}(v|\eta) \text{ and } \eta \in \partial F(v) \\ \emptyset & \text{otherwise} \end{cases}$$

$$V_{\partial F}(v|\eta) = \left\{ z \in L^2(\Omega) \mid z(x) = 0 \text{ if } |v(x)| = 1,\ x \in \Omega \right\}$$

Examples: L^1 norm

$$F(u) = \int_\Omega f(u(x))\,dx, \qquad f(z) = |z|, \qquad \partial f(z) = \operatorname{sign}(z)$$

- Fréchet derivative of ∂f

$$D(\partial f)(z|\zeta)(\Delta z) = \begin{cases} \{0\} & z \neq 0,\ \zeta = \operatorname{sign} z \\ \{0\} & z = 0,\ \Delta z \neq 0,\ \zeta|\Delta z| = \Delta z \\ (-\infty, 0]\zeta & z = 0,\ \Delta z = 0,\ |\zeta| = 1 \\ \mathbb{R} & z = 0,\ \Delta z = 0,\ |\zeta| < 1 \\ \emptyset & \text{otherwise} \end{cases}$$

- $\rightsquigarrow$ Fréchet derivative of ∂F

$$D[\partial F](v|\eta)(\Delta v) = \begin{cases} V_{\partial F}(v|\eta)^\perp & \Delta v \in V_{\partial F}(v|\eta) \text{ and } \eta \in \partial F(v) \\ \emptyset & \text{otherwise} \end{cases}$$

$$V_{\partial F}(v|\eta) = \left\{ z \in L^2(\Omega) \mid z(x) = 0 \text{ if } v(x) = 0,\ x \in \Omega \right\}$$

Application to PDEs

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Primal-dual optimality conditions

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

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: $Y \mapsto Y_h$

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

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

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

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

L^1 fitting: algorithm

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

- $S'(u^k)^* v^k$ 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

L^1 fitting: metric regularity

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

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

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 $z = 0$
- $\bar{\eta} = S(\bar{u}) - y^\delta, \quad \bar{v} \in \operatorname{sign} \bar{\eta}$
- $\rightsquigarrow$ in general not satisfied!

L^∞ fitting

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

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

- $F(y) = \delta_{\{|y(x)|\leqslant\delta\}}(y) \quad \rightsquigarrow \quad f^* = \delta|z|$
- $S : U \subset L^2(\Omega) \to L^2(\Omega)$, $S(u) =: y$ satisfies

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

L^∞ fitting: algorithm

$$
\begin{cases}
z^{k+1} = S'(u^k)^* v^k \\
u^{k+1} = \dfrac{1}{1+\tau_k \alpha}(u^k - \tau_k z^{k+1}) \\
\omega_k = 1/\sqrt{1+2c\tau_k} \quad \tau_{k+1} = \omega_k \tau_k \quad \sigma_{k+1} = \sigma_k/\omega_k \\
\bar{u}^{k+1} = u^{k+1} + \omega_k(u^{k+1} - u^k) \\
\bar{v}^{k+1} = v^k + \sigma_{k+1}(S(\bar{u}^{k+1}) - y^\delta) \\
v^{k+1} = \frac{1}{1+\sigma_{k+1}\gamma}(|\bar{v}^{k+1}| - \delta\sigma)^+ \operatorname{sign}(\bar{v}^{k+1})
\end{cases}
$$

- local convergence if $\gamma > 0$ or finite-dimensional
- for $t = 0$: $z(x) = 0$ if $|S(\bar{u})(x) - y^\delta(x)| < \delta$ $\rightsquigarrow$ estimate unlikely

L^1 fitting

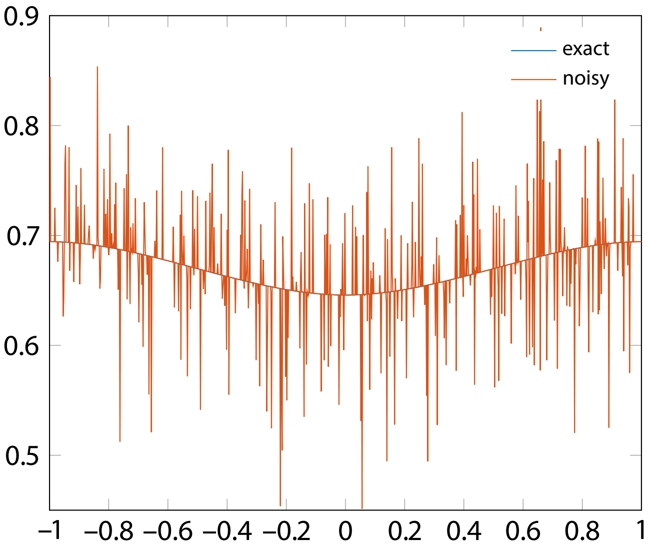
- $\Omega = [-1, 1]$, FE (P_1–P_0) discretization, $N = 1000$ nodes
- random impulsive noise:

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

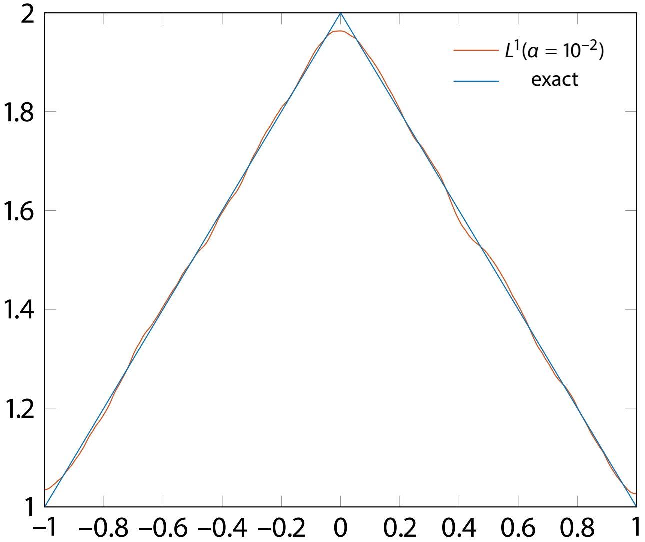
$$y^\dagger = S(u^\dagger), \quad \xi(x) \in \mathcal{N}(0, 0.1)$$

- $\gamma = 10^{-12}, c = \alpha = 10^{-2}, \sigma = 2, \tau = 3.5$
- $u^0 \equiv 1, v^0 = 0$ (no warmstart!)
- 1000 iterations

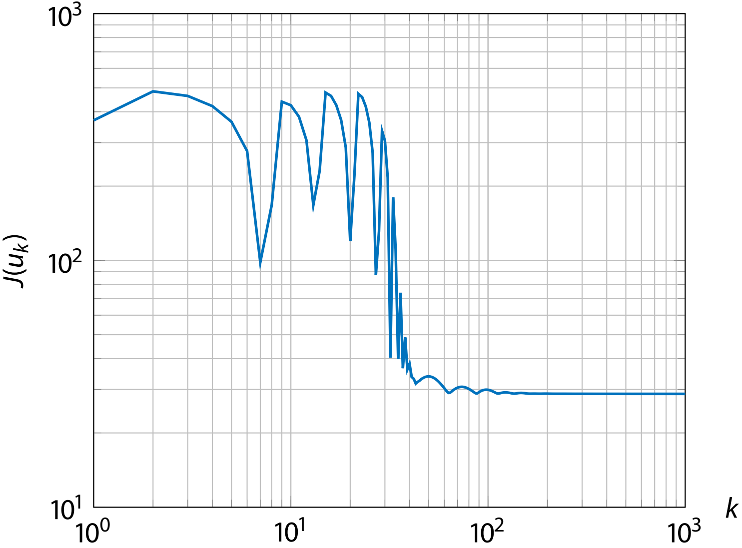
L^1 fitting: data



L^1 fitting: reconstruction



L^1 fitting: functional value



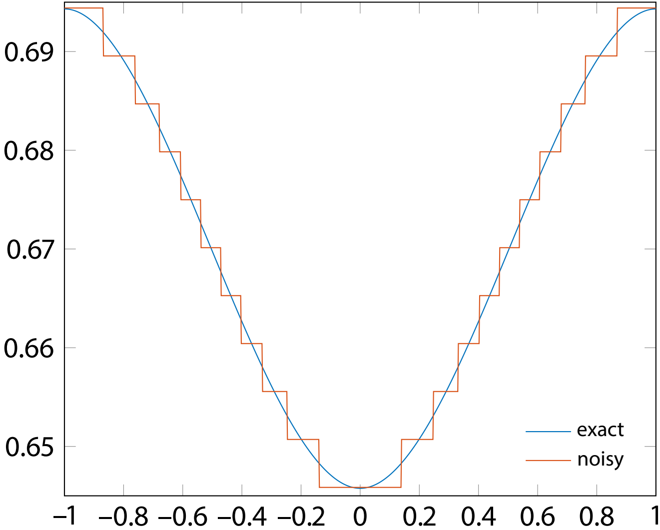
L^∞ fitting

- $\Omega = [-1, 1]$, FE (P_1–P_0) discretization, $N = 1000$ nodes
- quantization noise:

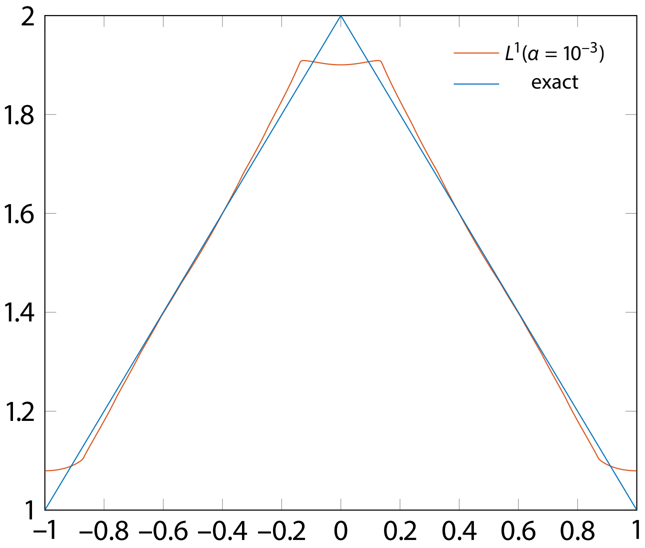
  $$\text{round } y^\dagger \text{ to } n_b = 11 \text{ equidistant values}$$

- $\gamma = 10^{-12}$, $c = \alpha = 10^{-3}$, $\sigma = 2$, $\tau = 3$
- $u^0 \equiv 1$, $v^0 = 0$ (no warmstart!)
- 1000 iterations

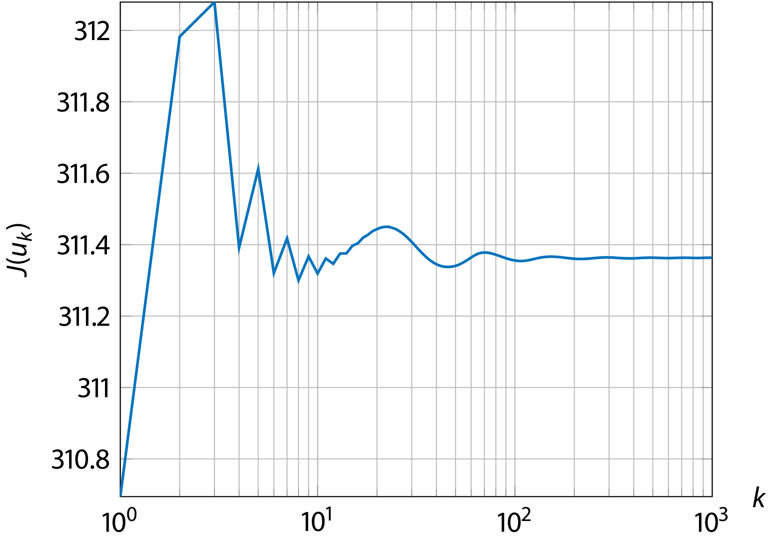
L^∞ fitting: data



L^∞ fitting: reconstruction



L^∞ fitting: functional value



Conclusion

Primal-dual extragradient methods in function space:

- can be accelerated
- analyzed using pointwise set-valued analysis
- requires Moreau–Yosida regularization
- does not require norm gap, continuation

Outlook:

- full acceleration
- partial stability (w.r.t. primal variable only)
- other PDE-constrained optimization problems
- pointwise set-valued analysis for bilevel problems