Accelerated primal–dual methods for PDE-constrained optimization

Christian Clason[1] Stanislav Mazurenko[2] Tuomo Valkonen[3]

[1]Faculty of Mathematics, Universität Duisburg-Essen
[2]Department of Mathematical Sciences, University of Liverpool
[3]MODEMAT, Escuela Politécnica Nacional, Quito, Ecuador

International Workshop on
PDE-Constrained Optimization, Optimal Controls and Application

Tsinghua Sanya International Mathematics Forum, Hainan
December 10, 2018

Motivation

Primal-dual proximal splitting:

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

Here:

- application to PDE-constrained optimization
- convergence rates for acceleration
- $\rightsquigarrow$ function space algorithm

Model problems

Parameter identification with L¹-fitting

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

Optimal control with state constraints

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

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

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

Derivation of the algorithm

Optimization problem

$$\min_{u \in X} F(u) + G(Ku) =: J(u)$$

- $F : X \to \overline{\mathbb{R}},\ G : Y \to \overline{\mathbb{R}}$ convex, lower semicontinuous
- X, Y Hilbert spaces
- $K \in L(X, Y)$

Derivation of the algorithm

Optimization problem

$$\min_{u \in X} F(u) + G(Ku) =: J(u)$$

Optimality condition

$$0 \in \partial J(\bar{u})$$

Derivation of the algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Optimization problem

$$\min_{u \in X} F(u) + G(Ku) =: J(u)$$

Optimality condition

$$0 \in \partial J(\bar{u})$$

Optimality condition (any $\gamma > 0$)

$$\bar{u} \in (\mathrm{Id} + \gamma \partial J)^{-1}(\bar{u}) =: \mathcal{R}_{\gamma \partial J}(\bar{u})$$

(resolvent: single-valued, Lipschitz; coincides w. proximal mapping)

Derivation of the algorithm

UNIVERSITÄT DUISBURG ESSEN Open-Minded

Optimality condition (any $\gamma > 0$)

$$\bar{u} \in (\mathrm{Id} + \gamma\partial J)^{-1}(\bar{u}) =: \mathcal{R}_{\gamma\partial J}(\bar{u})$$

(resolvent: single-valued, Lipschitz; coincides w. proximal mapping)

Fixed point iteration: proximal point method

$$u^{k+1} = \mathcal{R}_{\gamma_k\partial J}(u^k)$$

Pro: converges for any $\gamma_k \geqslant \gamma_0 > 0$

Con: each step equivalent to original problem

Optimization problem

$$\min_{u \in X} F(u) + G(Ku) =: J(u)$$

Optimality condition (sum, chain rule)

$$0 \in \partial J(\bar{u}) = \partial F(\bar{u}) + K^* \partial G(K\bar{u})$$

Derivation of the algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Optimization problem

$$\min_{u \in X} F(u) + G(Ku) =: J(u)$$

Optimality condition (sum, chain rule)

$$0 \in \partial J(\bar{u}) = \partial F(\bar{u}) + K^* \partial G(K\bar{u})$$

Optimality condition (sum, chain rule, Fenchel conjugate)

$$\begin{aligned} -K^* \bar{p} &\in \partial F(\bar{u}) \\ K\bar{u} &\in \partial G^*(\bar{p}) \end{aligned}$$

Derivation of the algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Optimality condition (sum, chain rule, Fenchel conjugate)

$$
\begin{aligned}
-K^*\bar{p} &\in \partial F(\bar{u}) \\
K\bar{u} &\in \partial G^*(\bar{p})
\end{aligned}
$$

Optimality condition (any $\sigma, \tau > 0$)

$$
\begin{aligned}
\bar{u} &= \mathcal{R}_{\tau\partial F}(\bar{u} - \tau K^*\bar{p}) \\
\bar{p} &= \mathcal{R}_{\sigma\partial G^*}(\bar{p} + \sigma K\bar{u})
\end{aligned}
$$

Derivation of the algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Optimality condition (any $\sigma, \tau > 0$)

$$\bar{u} = \mathcal{R}_{\tau\partial F}(\bar{u} - \tau K^* \bar{p})$$
$$\bar{p} = \mathcal{R}_{\sigma\partial G^*}(\bar{p} + \sigma K \bar{u})$$

Fixed point iteration

$$u^{k+1} = \mathcal{R}_{\tau_k\partial F}(u^k - \tau K^* p^k)$$
$$p^{k+1} = \mathcal{R}_{\sigma_k\partial G^*}(p^k + \sigma K u^{k+1})$$

Pro: each step easy to evaluate (for “nice” F, G, K)

Con: convergence?

Algorithm

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Idea: write as preconditioned primal-dual proximal point method

$$z^{k+1} = \mathcal{R}_{M^{-1}H}(z^k) \quad \Leftrightarrow \quad 0 \in H(z^{k+1}) + M_k(z^{k+1} - z^k)$$

- $z^k = (u^k, p^k)$
- $H(z) = \begin{pmatrix} \partial F(u) + K^* y \\ \partial G^*(p) - Ku \end{pmatrix}, \quad M_k = \begin{pmatrix} \tau_k^{-1}\,\mathrm{Id} & -K^* \\ 0 & \sigma_k^{-1}\,\mathrm{Id} \end{pmatrix}$ preconditioner

Algorithm

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Idea: write as preconditioned primal-dual proximal point method

$$z^{k+1} = \mathcal{R}_{M^{-1}H}(z^k) \quad \Leftrightarrow \quad 0 \in H(z^{k+1}) + M_k(z^{k+1} - z^k)$$

- $z^k = (u^k, p^k)$
- $H(z) = \begin{pmatrix} \partial F(u) + K^* y \\ \partial G^*(p) - Ku \end{pmatrix}, \quad M_k = \begin{pmatrix} \tau_k^{-1}\,\mathrm{Id} & -K^* \\ -K & \sigma_k^{-1}\,\mathrm{Id} \end{pmatrix}$ preconditioner

Idea: write as preconditioned primal-dual proximal point method

$$z^{k+1} = \mathcal{R}_{M^{-1}H}(z^k) \quad \Leftrightarrow \quad 0 \in H(z^{k+1}) + M_k(z^{k+1} - z^k)$$

- $z^k = (u^k, p^k)$
- $H(z) = \begin{pmatrix} \partial F(u) + K^* y \\ \partial G^*(p) - Ku \end{pmatrix}, \quad M_k = \begin{pmatrix} \tau_k^{-1}\,\mathrm{Id} & -K^* \\ -K & \sigma_k^{-1}\,\mathrm{Id} \end{pmatrix}$ preconditioner

Primal-dual proximal splitting

$$\begin{aligned} u^{k+1} &= \mathcal{R}_{\tau_k \partial F}(u^k - \tau_k K^* p^k) \\ \hat{u}^{k+1} &= 2u^{k+1} - u^k \\ p^{k+1} &= \mathcal{R}_{\sigma_k \partial G^*}(p^k + \sigma_k K \hat{u}^{k+1}) \end{aligned}$$

Convergence

$$0 \in H(z^{k+1}) + M_k(z^{k+1} - z^k) \quad \Leftrightarrow \quad z^{k+1} = \mathcal{R}_{M^{-1}H}(z^k)$$

- $H = \begin{pmatrix} \partial F & K^* \\ -K & \partial G^* \end{pmatrix}$ maximally monotone + skew symmetric
- $M_k = \begin{pmatrix} \tau_k^{-1}\,\mathrm{Id} & -K^* \\ -K & \sigma_k^{-1}\,\mathrm{Id} \end{pmatrix}$ self-adjoint
- M_k positive definite if $1 - \sqrt{\sigma_k \tau_k}\|K\| > 0$

$\leadsto M^{-1}H$ maximally monotone $\leadsto$ weak convergence to $\bar{z} \in H^{-1}(0)$

[Chambolle/Pock '09, He/Yuan '12]

Convergence rates by testing

Nesterov: strong convergence with rate under strong convexity

- F γ-strongly convex if $F - \frac{\gamma}{2}\|\cdot\|^2$ convex for $\gamma \geqslant 0$:

$$\langle \partial F(u) - \partial F(\tilde{u}), u - \tilde{u} \rangle \geqslant \gamma \|u - \tilde{u}\|^2 \quad \text{for all } u, \tilde{u}$$

- G^* ρ-strongly convex if $G^* - \frac{\rho}{2}\|\cdot\|^2$ convex for $\rho \geqslant 0$:

$$\langle \partial G^*(p) - \partial G^*(\tilde{p}), p - \tilde{p} \rangle \geqslant \rho \|p - \tilde{p}\|^2 \quad \text{for all } p, \tilde{p}$$

(always holds for Moreau–Yosida regularization G_ρ, $\rho > 0$!)

Testing: consider weighted norm with weight $\to \infty$ at rate
[Valkonen/Pock '17, Valkonen '18]

Convergence rates by testing

Abstract iteration

$$0 \in W_{k+1}H(z^{k+1}) + M_{k+1}(z^{k+1} - z^k)$$

- W_{k+1} step size operator
- H as before with $0 \in H(\bar{z})$
- M_{k+1} preconditioner (multiplied by step sizes)

Abstract iteration

$$0 \in Z_{k+1} W_{k+1} H(z^{k+1}) + Z_{k+1} M_{k+1}(z^{k+1} - z^k)$$

- W_{k+1} step size operator
- H as before with $0 \in H(\bar{z})$
- M_{k+1} preconditioner (multiplied by step sizes)
- Z_{k+1} testing operator

Convergence rates by testing

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Abstract iteration

$$0 \in Z_{k+1} W_{k+1} H(z^{k+1}) + Z_{k+1} M_{k+1} (z^{k+1} - z^k)$$

Weighted Γ-strong monotonicity for $\Gamma = \mathrm{diag}(\gamma, \rho) \geqslant 0$:

$$\langle H(z^{k+1}) - H(\bar{z}), z^{k+1} - \bar{z} \rangle_{W_{k+1} Z_{k+1}} \geqslant \|z^{k+1} - \bar{z}\|^2_{W_{k+1} Z_{k+1} \Gamma}$$

Convergence rates by testing

Abstract iteration

$$0 \in Z_{k+1} W_{k+1} H(z^{k+1}) + Z_{k+1} M_{k+1}(z^{k+1} - z^k)$$

Weighted Γ-strong monotonicity for $\Gamma = \operatorname{diag}(\gamma, \rho) \geqslant 0$:

$$\langle H(z^{k+1}) - H(\bar{z}), z^{k+1} - \bar{z} \rangle_{W_{k+1} Z_{k+1}} \geqslant \|z^{k+1} - \bar{z}\|^2_{W_{k+1} Z_{k+1} \Gamma}$$

Use iteration, $0 \in H(\bar{z})$, Pythagoras identity in Hilbert space:

$$\frac{1}{2}\|z^{k+1} - \bar{z}\|^2_{Z_{k+1}(M_{k+1} + 2W_{k+1}\Gamma)} + \frac{1}{2}\|z^{k+1} - z^k\|^2_{Z_{k+1} M_{k+1}} \leqslant \frac{1}{2}\|z^k - \bar{z}\|^2_{Z_{k+1} M_{k+1}}$$

Convergence rates by testing

$$\frac{1}{2}\|z^{k+1}-\bar{z}\|^2_{Z_{k+1}(M_{k+1}+2W_{k+1}\Gamma)}+\frac{1}{2}\|z^{k+1}-z^k\|^2_{Z_{k+1}M_{k+1}} \leqslant \frac{1}{2}\|z^k-\bar{z}\|^2_{Z_{k+1}M_{k+1}}$$

If:

1. $Z_{k+1}M_{k+1}$ self-adjoint, positive semi-definite
2. $Z_{k+1}(M_{k+1}+2W_{k+1}\Gamma) \geqslant Z_{k+2}M_{k+2}$

$$\frac{1}{2}\|z^{k+1}-\bar{z}\|^2_{Z_{k+2}M_{k+2}} \leqslant \frac{1}{2}\|z^k-\bar{z}\|^2_{Z_{k+1}M_{k+1}}$$

(quantitative Féjer monotonicity)

Convergence rates by testing

$$\frac{1}{2}\|z^{k+1}-\bar{z}\|^2_{Z_{k+1}(M_{k+1}+2W_{k+1}\Gamma)}+\frac{1}{2}\|z^{k+1}-z^k\|^2_{Z_{k+1}M_{k+1}}\leqslant\frac{1}{2}\|z^k-\bar{z}\|^2_{Z_{k+1}M_{k+1}}$$

If:

1. $Z_{k+1}M_{k+1}$ self-adjoint, positive semi-definite
2. $Z_{k+1}(M_{k+1}+2W_{k+1}\Gamma)\geqslant Z_{k+2}M_{k+2}$

$$\frac{1}{2}\|z^N-\bar{z}\|^2_{Z_{N+1}M_{N+1}}\leqslant\frac{1}{2}\|z^0-\bar{z}\|^2_{Z_1M_1}$$

$$\frac{1}{2}\|z^N - \bar{z}\|^2_{Z_{N+1}M_{N+1}} \leqslant \frac{1}{2}\|z^0 - \bar{z}\|^2_{Z_1 M_1}$$

3 $Z_{N+1}M_{N+1} \geqslant \mu(N)\,\mathrm{Id}$ for $\mu(N) \to \infty$:

Convergence rate

$$\|z^N - \bar{z}\|^2 = \mathcal{O}(\mu(N)^{-1})$$

UNIVERSITÄT DUISBURG ESSEN Open-Minded

Abstract iteration

$$0 \in Z_{k+1} W_{k+1} H(z^{k+1}) + Z_{k+1} M_{k+1}(z^{k+1} - z^k)$$

Primal-dual proximal splitting:

- $H = \begin{pmatrix} \partial F & K^* \\ -K & \partial G^* \end{pmatrix}$ $\qquad M_{k+1} = \begin{pmatrix} \mathrm{Id} & -\tau_k K^* \\ -\omega_k \sigma_{k+1} K & \mathrm{Id} \end{pmatrix}$
- $W_{k+1} = \begin{pmatrix} \tau_k \,\mathrm{Id} & 0 \\ 0 & \sigma_{k+1}\,\mathrm{Id} \end{pmatrix}$ $\qquad Z_{k+1} = \begin{pmatrix} \varphi_k \,\mathrm{Id} & 0 \\ 0 & \psi_{k+1}\,\mathrm{Id} \end{pmatrix}$
- $\rightsquigarrow$ H weighted strongly monotone iff strongly monotone

$\rightsquigarrow$ choose $\tau_k, \sigma_{k+1}, \omega_k, \varphi_k, \psi_{k+1}$ such that 1 – 3

UNIVERSITÄT DUISBURG ESSEN Open-Minded

1. $Z_{k+1}M_{k+1}$ self-adjoint iff

$$\omega_k = \sigma_k^{-1}\psi_{k+1}^{-1}\varphi_k\tau_k$$

2. $Z_{k+1}(M_{k+1} + 2W_{k+1}\Gamma) - Z_{k+2}M_{k+2} \geqslant 0$ if

$$\varphi_{k+1} \leqslant \varphi_k(1 + 2\gamma\tau_k), \quad \psi_{k+1} \leqslant \psi_k(1 + 2\rho\sigma_{k+1}), \quad \varphi_k\tau_k = \psi_k\sigma_k$$

$\rightsquigarrow$ $Z_{k+1}M_{k+1}$ positive definite if

$$1 - \sigma_k\tau_k\|K\|^2 > 0$$

$\rightsquigarrow$ choose $\sigma_k\tau_k \equiv \sigma_0\tau_0$

3. growth of $Z_{k+1}M_{k+1}$?

Application to primal-dual proximal splitting

F strongly monotone for $\gamma > 0$:

$$\varphi_{k+1} = \varphi_k(1 + 2\gamma\tau_k), \qquad \psi_{k+1} = \psi_k$$

$\rightsquigarrow$ choose

- $\tau_0\sigma_0 < \|K\|^{-2}$
- $\omega_k = (1 + 2\gamma\tau_k)^{-1/2}$
- $\tau_{k+1} = \tau_k\omega_k$
- $\sigma_{k+1} = \sigma_k/\omega_k$

$$\varphi_k = \mathcal{O}(k^2) \quad \Rightarrow \quad \|u^k - \bar{u}\|^2 = \mathcal{O}(k^{-2})$$

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

F strongly monotone for $\gamma > 0$:

Accelerated primal-dual proximal splitting

$$
\begin{aligned}
u^{k+1} &= \mathcal{R}_{\tau_k \partial F}(u^k - \tau_k K^* p^k) \\
\omega_k &= (1 + 2\gamma\tau_k)^{-1/2}, \quad \tau_{k+1} = \tau_k \omega_k, \quad \sigma_{k+1} = \sigma_k / \omega_k \\
\hat{u}^{k+1} &= \omega_k (u^{k+1} - u^k) + u^{k+1} \\
p^{k+1} &= \mathcal{R}_{\sigma_{k+1} \partial G^*}(p^k + \sigma_{k+1} K \hat{u}^{k+1})
\end{aligned}
$$

Application to primal-dual proximal splitting

F and G^* strongly monotone for $\gamma > 0, \rho > 0$:

$$\varphi_{k+1} = \varphi_k(1 + 2\gamma\tau_k), \qquad \psi_{k+1} = \psi_k(1 + 2\gamma\sigma_k)$$

$\rightsquigarrow$ choose

- $\tau_k \equiv \tau_0, \sigma_k \equiv \sigma_0$
- $\tau_0\sigma_0 < \|K\|^{-2}$
- $\omega_k = (1 + 2\min\{\gamma, \rho\})^{-1}$

$$\varphi_k, \psi_k = \mathcal{O}(C(\gamma, \rho)^k) \quad \Rightarrow \quad \|u^k - \bar{u}\|^2 + \|p^k - \bar{p}\|^2 = \mathcal{O}(C(\gamma, \rho)^{-k})$$

Application to primal-dual proximal splitting

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

F and G^* strongly monotone for $\gamma > 0$, $\rho > 0$:

Accelerated primal-dual proximal splitting

$$
\begin{aligned}
u^{k+1} &= \mathcal{R}_{\tau\partial F}(u^k - \tau K^* p^k) \\
\omega &= (1 + 2\min\{\gamma, \rho\})^{-1} \\
\hat{u}^{k+1} &= \omega(u^{k+1} - u^k) + u^{k+1} \\
p^{k+1} &= \mathcal{R}_{\sigma\partial G^*}(p^k + \sigma K \hat{u}^{k+1})
\end{aligned}
$$

Nonlinear primal-dual proximal splitting

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

If $K : X \to Y$ nonlinear, differentiable

Nonlinear primal-dual proximal splitting [Valkonen '14]

$$u^{k+1} = \mathcal{R}_{\tau_k \partial F}(u^k - \tau_k K'(u^k)^* p^k)$$
$$\hat{u}^{k+1} = 2u^{k+1} - u^k$$
$$p^{k+1} = \mathcal{R}_{\sigma_k \partial G^*}(p^k + \sigma_k K(\hat{u}^{k+1}))$$

- $K'(u)$ Fréchet derivative of K
- $K'(u)^*$ adjoint
- analogously for acceleration under strong convexity
- convergence using same approach (with "linearized" H_{k+1})

Convergence

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

General assumptions:

1. K' locally Lipschitz
2. K' locally uniformly bounded
3. F, G^* locally strongly convex for $\gamma, \rho \geqslant 0$
4. for some $\lambda, \theta, \tilde{\gamma} > 0$ and all $u, \hat{u}$ near $\bar{u}$:

Three-point condition

$$\langle [K'(\hat{u}) - K'(u)]^* \bar{p}, u - \bar{u} \rangle + (\gamma - \tilde{\gamma}) \|u - \bar{u}\|^2 \geqslant \theta \|K(\bar{u}) - K(u) - K'(u)(\bar{u} - u)\| - \frac{\lambda}{2} \|u - \hat{u}\|^2$$

(If 1 – 4 globally $\rightsquigarrow$ convergence results global)

Three-point condition

Three-point condition

$$\langle [K'(\hat{u}) - K'(u)]^* \bar{p}, u - \bar{u} \rangle + (\gamma - \tilde{\gamma}) \|u - \bar{u}\|^2 \geqslant \theta \|K(\bar{u}) - K(u) - K'(u)(\bar{u} - u)\| - \frac{\lambda}{2} \|u - \hat{u}\|^2$$

- holds if

$$(\gamma - \tilde{\gamma}) \|u - \bar{u}\|^2 + \left\langle \left(K'(u) - K'(\bar{u}) \right) (u - \bar{u}), \bar{p} \right\rangle \geqslant \delta \|u - \bar{u}\|^2$$

- quadratic growth condition at solution w.r.t. primal variable
- cf. sufficient second order condition

Weak convergence

Assume:

- F, G^* convex (not strongly)
- conditions 1 – 4 hold for any weak limit
- $u \mapsto K'(u)$ weak-to-strong continuous
- $(u, p) \mapsto (K'(u)^* p, K(u))$ weak-to-weak continuous
- σ, τ sufficiently small

Then $(u^k, p^k) \rightharpoonup (\bar{u}, \bar{p})$ with $0 \in H(\bar{u}, \bar{p})$ locally

Convergence with rates

Assume:

- F strongly convex for $\gamma > 0$
- conditions 1 – 4 hold
- σ, τ and $\tilde{\gamma} < \gamma$ sufficiently small
- $\omega_k = (1 + 2\tilde{\gamma}\tau_k)^{-1/2}, \quad \tau_{k+1} = \tau_k\omega_k, \quad \sigma_{k+1} = \sigma_k/\omega_k$

Then $\|u^k - \bar{u}\|^2 = \mathcal{O}(k^{-2})$ locally

Convergence with rates

Assume:

- F and G^* strongly convex for $\gamma, \rho > 0$
- conditions 1 – 4 hold
- τ and $\tilde{\gamma} < \gamma$ and $\tilde{\rho} < \rho$ sufficiently small
- $\omega_k = (1 + 2\tilde{\gamma}\tau)^{-1}, \quad \sigma = \frac{\tilde{\gamma}}{\tilde{\rho}}\tau$

Then $\|u^k - \bar{u}\|^2 + \|p^k - \bar{p}\|^2 = \mathcal{O}(C(\tilde{\rho})^{-k})$ locally

Numerical example: L^1 fitting

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

- $F(u) = \frac{1}{2}\|u\|_{L^2}^2 \quad \rightsquigarrow \gamma = 1$
- $G(y) = \alpha^{-1}\|y(x)\|_{L^1} \quad \rightsquigarrow \quad G^*(y) = \delta_{[-\alpha^{-1},\alpha^{-1}]}(y)$ a.e.
- $S : U \subset L^2(\Omega) \to L^2(\Omega), \quad S(u) =: y$ satisfies

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

- $K(u) = S(u) - y^\delta$
- τ, σ using estimated Lipschitz constant, $u^0 \equiv 1, p^0 \equiv 0$
- $(\bar{u}, \bar{p})$ estimated by iterate $(u^{2\max}, p^{2\max})$

Numerical example: L^1 fitting

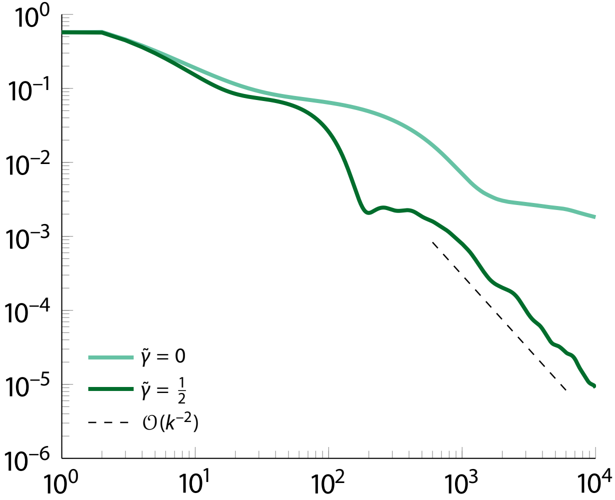


Figure: comparison of standard ($\tilde{\gamma} = 0$) and accelerated ($\tilde{\gamma} = 1/2 < \gamma$)

Numerical example: L^1 fitting

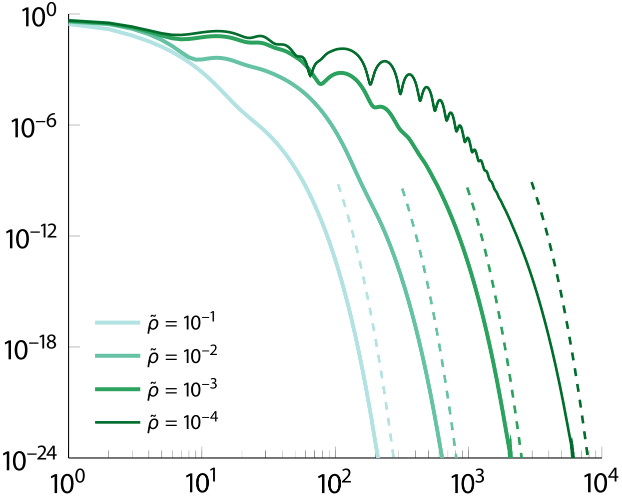


Figure: Moreau–Yosida regularization G_ρ: rate $\mathcal{O}(C^k)$ (dashed)

Numerical example: state constraints

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

$$\min_u \frac{1}{2\alpha}\|S(u)-y^d\|_{L^2} + \frac{1}{2}\|u\|_{L^2}^2 \quad \text{s.t.} \quad S(u) \leqslant c$$

- $F(u) = \frac{1}{2}\|u\|_{L^2}^2 \quad \rightsquigarrow \gamma = 1$
- $G(y) = (2\alpha)^{-1}\|y(x) - y^d\|_{L^2}^2 + \delta_{(-\infty,c]}(y)$.
- $S : U \subset L^2(\Omega) \to L^2(\Omega), \quad S(u) =: y$ satisfies

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

- $K(u) = S(u)$
- τ, σ using estimated Lipschitz constant, $u^0 \equiv 1, p^0 \equiv 0$
- $(\bar{u}, \bar{p})$ estimated by iterate $(u^{2\max}, p^{2\max})$

Numerical example: state constraints

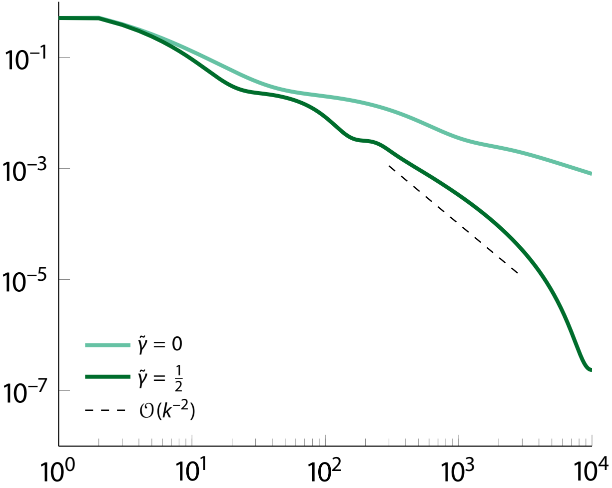


Figure: comparison of standard ($\tilde{\gamma} = 0$) and accelerated ($\tilde{\gamma} = 1/2 < \gamma$)

Numerical example: state constraints

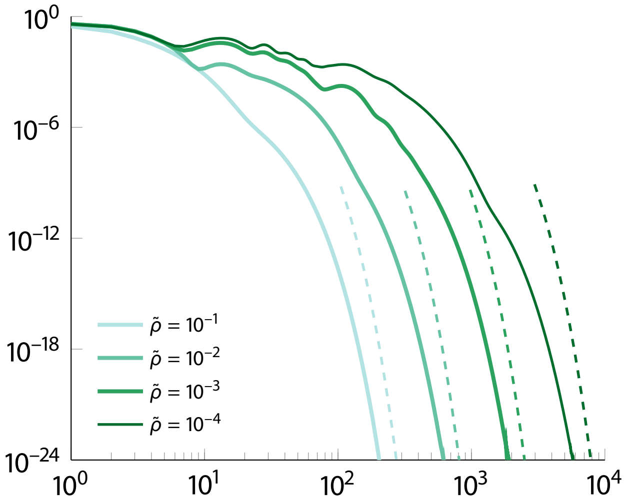


Figure: Moreau–Yosida regularization G_ρ: rate $\mathcal{O}(C(\tilde{\rho})^k)$ (dashed)

Generalized saddle-point problems

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Saddle-point problem

$$\min_{u \in X} \max_{p \in Y} F(u) + \langle p, Ku \rangle - G^*(p)$$

Primal-dual proximal splitting

$$\begin{aligned} u^{k+1} &= \mathcal{R}_{\tau_k \partial F}(u^k - \tau_k K^* p^k) \\ \hat{u}^{k+1} &= 2u^{k+1} - u^k \\ p^{k+1} &= \mathcal{R}_{\sigma_k \partial G^*}(p^k + \sigma_k K \hat{u}^{k+1}) \end{aligned}$$

Generalized saddle-point problems

Generalized saddle-point problem

$$\min_{u \in X} \max_{p \in Y} F(u) + K(u,p) - G^*(p)$$

Generalized primal-dual proximal splitting

$$u^{k+1} = \mathcal{R}_{\tau_k \partial F}(u^k - \tau_k K_u(u^k, p^k))$$
$$\hat{u}^{k+1} = 2u^{k+1} - u^k$$
$$p^{k+1} = \mathcal{R}_{\sigma_k \partial G^*}(p^k + \sigma_k K_p(\bar{u}^{k+1}, p^k))$$

- $K : X \times Y \to \mathbb{R}$ nonlinear, partial derivatives K_u, K_p

Application to Nash equilibrium problems

Two player (for simplicity) non-cooperative game:

- strategies $x_i \in X, i = 1, 2$
- $X_i \subset \mathbb{R}$ set of valid strategies, $i = 1, 2$ $\quad (X = X_1 \times X_2)$
- payout functions $\varphi_i(x_1, x_2), i = 1, 2$

Nash equilibrium

$$\varphi_1(x_1^*, x_2) = \min_{x_1 \in X_1} \varphi_1(x_1, x_2) \quad \text{f.a. } x_2 \in X_2$$
$$\varphi_2(x_1, x_2^*) = \min_{x_2 \in X_2} \varphi_1(x_1, x_2) \quad \text{f.a. } x_1 \in X_1$$

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Nash equilibrium

$$\varphi_1(x_1^*, x_2) = \min_{x_1 \in X_1} \varphi_1(x_1, x_2) \quad \text{f.a. } x_2 \in X_2$$
$$\varphi_2(x_1, x_2^*) = \min_{x_2 \in X_2} \varphi_1(x_1, x_2) \quad \text{f.a. } x_1 \in X_1$$

Nikaido–Isoda function

$$K(x, y) = \big(\varphi_1(x_1, y_2) - \varphi_1(x_1, x_2)\big) + \big(\varphi_2(y_1, x_2) - \varphi_2(x_1, x_2)\big)$$

UNIVERSITÄT DUISBURG ESSEN

Open-Minded

Nikaido–Isoda function

$$K(x,y) = \big(\varphi_1(x_1,y_2) - \varphi_1(x_1,x_2)\big) + \big(\varphi_2(y_1,x_2) - \varphi_2(x_1,x_2)\big)$$

Nash equilibrium

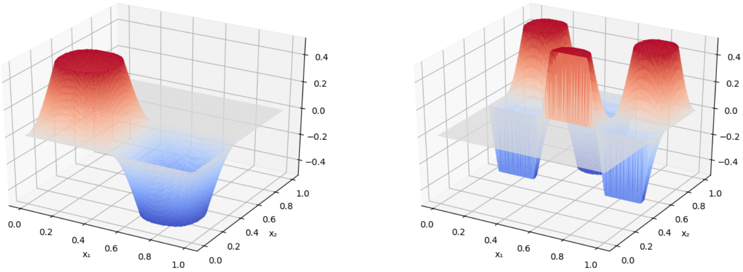
$$\min_{x\in X}\max_{y\in Y} K(x,y) = \min_{x\in \mathbb{R}^2}\max_{x\in \mathbb{R}^2} \delta_X(x) + K(x,y) - \delta_X(y)$$

Elliptic Nash equilibrium problem

- $X_1 = X_2 = \{u \in L^2(\Omega) : u(x) \in [a, b] \text{ a.e. }\}$
- $\varphi_i(u_1, u_2) = \frac{1}{2}\|S(u_1, u_2) - z_i\|^2 + \frac{\alpha_i}{2}\|u_i\|^2$
- $y = S(u_1, u_2)$ solves $\quad -\Delta y = u_1|_{\omega_1} + u_2|_{\omega_2} + f$
- $\omega_1, \omega_2 \subset \Omega$ disjoint

$$K_u(u, v) = \begin{pmatrix} p_1(u, v) + \alpha_1 u_1 \\ p_2(u, v) + \alpha_2 u_2 \end{pmatrix} \qquad K_v(u, v) = \begin{pmatrix} q_1(u, v) - \alpha_1 v_1 \\ q_2(u, v) - \alpha_2 v_2 \end{pmatrix}$$

$$\begin{aligned} -\Delta p_1 &= 2S(u_1, u_2) - S(u_1, v_2) - z_1 \qquad & -\Delta q_1 &= -S(v_1, u_2) + z_1 \\ -\Delta p_2 &= 2S(u_1, u_2) - S(v_1, u_2) - z_2 \qquad & -\Delta q_2 &= -S(u_1, v_2) + z_2 \end{aligned}$$



(a) u_1^* (b) u_2^*

Figure: constructed solution [Borzì/Kanzow '13]

Numerical example: convergence

k	$N = 64$	$N = 128$	$N = 256$	$N = 512$	$N = 1024$
1	$1.298 \cdot 10^{-1}$	$1.319 \cdot 10^{-1}$	$1.330 \cdot 10^{-1}$	$1.335 \cdot 10^{-1}$	$1.338 \cdot 10^{-1}$
2	$3.889 \cdot 10^{-6}$	$4.048 \cdot 10^{-6}$	$4.074 \cdot 10^{-6}$	$4.088 \cdot 10^{-6}$	$4.097 \cdot 10^{-6}$
3	$3.835 \cdot 10^{-10}$	$3.977 \cdot 10^{-10}$	$4.010 \cdot 10^{-10}$	$4.026 \cdot 10^{-10}$	$4.032 \cdot 10^{-10}$
4	$3.811 \cdot 10^{-14}$	$3.952 \cdot 10^{-14}$	$3.986 \cdot 10^{-14}$	$4.001 \cdot 10^{-14}$	$4.008 \cdot 10^{-14}$
5	$3.787 \cdot 10^{-18}$	$3.928 \cdot 10^{-18}$	$3.963 \cdot 10^{-18}$	$3.977 \cdot 10^{-18}$	$3.985 \cdot 10^{-18}$

Table: error $\|u^k - u^*\|^2$ for different mesh sizes $h = N^{-1}$

Conclusion

Primal-dual methods for PDE-constrained optimization:

- convergence analysis in function space
- accelerated convergence rates
- no C^2, Moreau–Yosida regularization needed
- robust w.r.t. initialization

Outlook:

- generalized saddle-point problems (non-convex functionals)
- risk-averse optimization (CVaR $\rightsquigarrow$ projection)
- application to non-smooth PDEs/variational inequalities

Preprints/Code:

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