UNIVERSITY OF GRAZ
Department of Mathematics and
Scientific Computing

Primal-dual proximal splitting and generalized conjugation in nonsmooth nonconvex optimization

Christian Clason[1] Stanislav Mazurenko[2] Tuomo Valkonen[3,4]

[1]Department of Mathematics and Scientific Computing, University of Graz
[2]Loschmidt Laboratories, Masaryk University, Brno, Czechia
[3]MODEMAT, Escuela Politécnica Nacional, Quito, Ecuador
[4]Department of Mathematics and Statistics, University of Helsinki, Finland

Workshop Recent Advances in Scientific Computing and Inverse Problems
Hong Kong, March 11, 2024

Motivation: convex optimization in imaging

TV-denoising: Rudin–Osher–Fatemi (ROF) model

$$\min_x \frac{1}{2}\|x - f\|_2^2 + \alpha \||Dx|_p\|_1$$

- f noisy image
- $\alpha > 0$ regularization parameter
- $D : \mathbb{R}^{n\times m} \longrightarrow \mathbb{R}^{n\times m\times 2}$ discrete gradient

Total variation penalty

$$F(Dx) = \||Dx|_p\|_1$$

- penalizes jump length and height
- $\rightsquigarrow$ piecewise constant (but staircasing)
- convex

Motivation: convex optimization in imaging

TV-denoising: Rudin–Osher–Fatemi (ROF) model

$$\min_x \frac{1}{2}\|x - f\|_2^2 + \alpha \||Dx|_p\|_1$$

Total variation penalty

$$F(Dx) = \||Dx|_p\|_1 = \sup_y \langle Dx, y\rangle - F^*(y)$$

- F^* Fenchel conjugate, always convex; here

$$F^*(y) = \delta_{B_\infty^q}(y) = \begin{cases} 0 & |z_{ij}|_q \leq 1 \\ \infty & \text{else} \end{cases}$$

Motivation: convex optimization in imaging

Saddle point problem

$$\min_x \max_y G(x) + \langle Dx, y \rangle - F^*(y)$$

Primal-dual proximal splitting (Chambolle–Pock)

$$\begin{aligned}
x^{i+1} &= \operatorname{prox}_{\tau_i G}(x^i - \tau_i D^* y^i) \\
\bar{x}^{i+1} &= x^{i+1} + \omega_i (x^{i+1} - x^i) \\
y^{i+1} &= \operatorname{prox}_{\sigma_{i+1} F^*}(y^i + \sigma_{i+1} D \bar{x}^{i+1})
\end{aligned}$$

- prox_G, $\operatorname{prox}_{F^*}$ proximal point mappings (projection)
- τ_i, σ_i step lengths
- ω_i overrelaxation

Motivation: nonconvex optimization in imaging

ℓ^0-TV-denoising: Potts model

$$\min_x \frac{1}{2}\|x - f\|_2^2 + \alpha\||Dx|_p\|_0$$

ℓ^0 "norm"

$$\|z\|_0 = \sum_{i,j} |z_{ij}|_0 \qquad |z|_0 = \begin{cases} 1 & z \neq 0 \\ 0 & z = 0 \end{cases}$$

- penalizes jump length, not height
- $\rightsquigarrow$ piecewise constant, no staircasing
- nonconvex

Goals:

- write ℓ^0 norm as generalized conjugate of convex F^*
- $\rightsquigarrow$ generalized primal-dual proximal splitting

1 Overview

2 Generalized conjugation

3 Generalized primal-dual proximal splitting

4 Potts model denoising

5 Nash equilibrium problems

6 Conclusion

1 Overview

2 Generalized conjugation

3 Generalized primal-dual proximal splitting

4 Potts model denoising

5 Nash equilibrium problems

6 Conclusion

Generalized conjugation: scalar motivation

$$f(t) = |t|_0 = \begin{cases} 1 & t \neq 0 \\ 0 & t = 0 \end{cases}$$

Consider $\rho : \mathbb{R} \longrightarrow \mathbb{R}$ with

1. $\rho(0) = 0$
2. $\sup_{t \leq 0} \rho(t) = 0$
3. $\sup_{t > 0} \rho(t) = 1$

Case distinction:

$$f(t) = \sup_s \rho(st) - 0$$

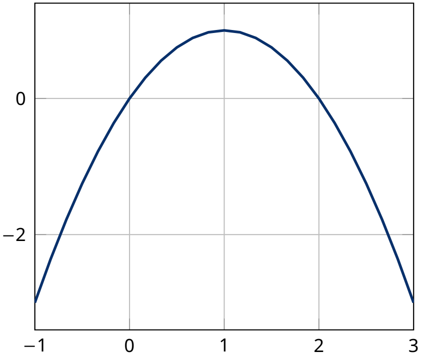
Example (smooth):

$$\rho(t) = 2t - t^2$$

Generalized conjugation: scalar motivation

Example (smooth):

$$\rho(t) = 2t - t^2$$

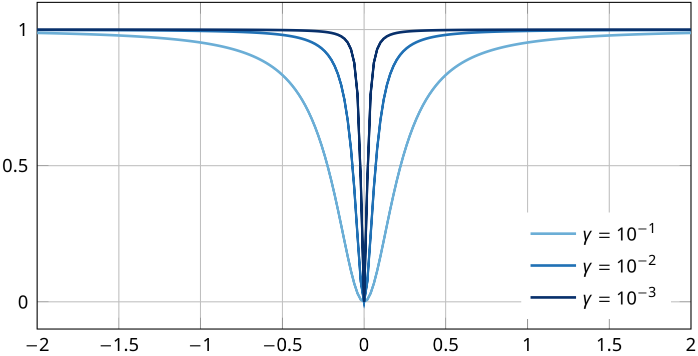


Generalized conjugation: scalar motivation

But: generalized conjugate $f^\rho = 0$ not strongly convex

$\rightsquigarrow$ Huber regularization for $\gamma > 0$

$$f_\gamma(t) = \sup_s \rho(st) - \frac{\gamma}{2}|s|^2 = \frac{2t^2}{2t^2+\gamma}$$



Generalized conjugation: ℓ^0-TV

$$F(z) = \sum_{i,j} |z_{ij,1}|_0 + |z_{ij,2}|_0$$

- anisotropic Potts model
- counts jump per pixel per direction
- separable

$$F_\gamma(z) = \sup_y \kappa_1(z, y) - \frac{\gamma}{2}\|y\|_2^2, \qquad \gamma \geq 0$$

$$\kappa_1(z, y) = \sum_{i,j} \rho(z_{ij,1} y_{ij,1}) + \rho(z_{ij,2} y_{ij,2})$$

Generalized conjugation: ℓ^0-TV

$$F(z) = \| |z|_\infty \|_0 = \sum_{ij} \big| |z_{ij}|_\infty \big|_0 = \sum_{ij} \max\{|z_{ij,1}|_0, |z_{ij_2}|_0\}$$

- isotropic Potts model
- counts jump per pixel
- not separable

$$F_\gamma(z) = \sup_y \kappa_\infty(z, y) - \frac{\gamma}{2}\|y\|_2^2, \qquad \gamma \geq 0$$

$$\kappa_\infty(z, y) = \sum_{i,j} \rho(z_{ij,1} y_{ij,1} + z_{ij,2} y_{ij,2})$$

1 Overview

2 Generalized conjugation

3 Generalized primal-dual proximal splitting

4 Potts model denoising

5 Nash equilibrium problems

6 Conclusion

Generalized primal-dual proximal splitting

Saddle point problem

$$\min_x \max_y G(x) + \langle Dx, y \rangle - F^*(y)$$

Primal-dual proximal splitting

$$\begin{aligned} x^{i+1} &= \operatorname{prox}_{\tau_i G}(x^i - \tau_i D^* y^i) \\ \bar{x}^{i+1} &= x^{i+1} + \omega_i (x^{i+1} - x^i) \\ y^{i+1} &= \operatorname{prox}_{\sigma_{i+1} F^*}(y^i + \sigma_{i+1} D \bar{x}^{i+1}) \end{aligned}$$

- prox_G, $\operatorname{prox}_{F^*}$ proximal point mappings (projection)
- τ_i, σ_i step lengths
- ω_i overrelaxation

Generalized primal-dual proximal splitting

Generalized saddle point problem

$$\min_x \max_y G(x) + K(x, y) - F^*(y)$$

Generalized Primal-dual proximal splitting

$$\begin{aligned} x^{i+1} &= \text{prox}_{\tau_i G}(x^i - \tau_i K_x(x^i, y^i)) \\ \bar{x}^{i+1} &= x^{i+1} + \omega_i (x^{i+1} - x^i) \\ y^{i+1} &= \text{prox}_{\sigma_{i+1} F^*}(y^i + \sigma_{i+1} K_y(\bar{x}^{i+1}, y^i)) \end{aligned}$$

- prox_G, prox_{F^*} proximal point mappings (projection)
- τ_i, σ_i step lengths
- ω_i overrelaxation

GPDPS: weak convergence

Assume that

1. F^*, G convex
2. K twice Lipschitz continuously differentiable (can be weakened)
3. second-order growth condition at saddle point
4. bounds on the step lengths and overrelaxation (technical!)

Then GPDPS locally converges weakly to saddle point.

GPDPS: strong convergence

Assume that

1. F^* convex, G strongly convex
2. K twice Lipschitz continuously differentiable (can be weakened)
3. second-order growth condition at saddle point
4. bounds on the constant step lengths and overrelaxation $\omega_i = 1$ (technical!)

Then GPDPS locally converges strongly to saddle point with rate $\mathcal{O}(1/N)$.

(Satisfied for Potts model)

GPDPS: linear convergence

Assume that

1. F^*, G strongly convex
2. K twice Lipschitz continuously differentiable (can be weakened)
3. second-order growth condition at saddle point
4. usual choice of constant step lengths and overrelaxation $\omega_i < 1$ (technical!)

Then GPDPS locally converges strongly to saddle point with rate $\mathcal{O}(c^N)$.

(Satisfied for Potts model with Huber regularization)

1 Overview

2 Generalized conjugation

3 Generalized primal-dual proximal splitting

4 Potts model denoising

5 Nash equilibrium problems

6 Conclusion

Potts model denoising

Generalized saddle point problem

$$\min_x \max_y G(x) + K(x, y) - F^*(y)$$

Here:

$$G(x) = \frac{1}{2\alpha}\|x - f\|_2^2,$$
$$F^*(y) = \frac{\gamma}{2}\|y\|_2^2,$$
$$K(x, y) = \kappa_p(D_h x, y),$$

Example:

- $\alpha = 1$, $\gamma = 10^{-3}$, $p \in \{1, \infty\}$
- F^* strongly convex $\rightsquigarrow$ constant step lengths, $\omega_i < 1$

Potts model denoising

Generalized Primal-dual proximal splitting

$$\begin{aligned} x^{i+1} &= \mathrm{prox}_{\tau_i G}(x^i - \tau_i K_x(x^i, y^i)) \\ \bar{x}^{i+1} &= x^{i+1} + \omega_i (x^{i+1} - x^i) \\ y^{i+1} &= \mathrm{prox}_{\sigma_{i+1} F^*}(y^i + \sigma_{i+1} K_y(\bar{x}^{i+1}, y^i)) \end{aligned}$$

$$\mathrm{prox}_{\tau G}(x) = \frac{1}{1+\frac{\tau}{\alpha}}\left(x + \frac{\tau}{\alpha} f\right), \qquad \mathrm{prox}_{\sigma F_\gamma^*}(y) = \frac{1}{1+\gamma\sigma} y$$

$$K_x(x,y) = D^* \kappa_{p,z}(Dx, y) \qquad K_y(x,y) = \kappa_{p,y}(Dx, y),$$

$p = 1$:

$$\begin{aligned} [\kappa_{1,z}(z,y)]_{ij,k} &= 2(1 - z_{ij,k} y_{ij,k}) y_{ij,k} \\ [\kappa_{1,y}(z,y)]_{ij,k} &= 2(1 - z_{ij,k} y_{ij,k}) z_{ij,k} \end{aligned}$$

Potts model denoising

Generalized Primal-dual proximal splitting

$$
\begin{aligned}
x^{i+1} &= \operatorname{prox}_{\tau_i G}(x^i - \tau_i K_x(x^i, y^i)) \\
\bar{x}^{i+1} &= x^{i+1} + \omega_i (x^{i+1} - x^i) \\
y^{i+1} &= \operatorname{prox}_{\sigma_{i+1} F^*}(y^i + \sigma_{i+1} K_y(\bar{x}^{i+1}, y^i))
\end{aligned}
$$

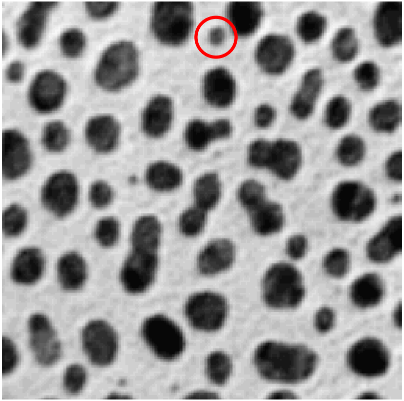
$$
\operatorname{prox}_{\tau G}(x) = \frac{1}{1+\frac{\tau}{\alpha}}\left(x + \frac{\tau}{\alpha} f\right), \qquad \operatorname{prox}_{\sigma F_\gamma^*}(y) = \frac{1}{1+\gamma\sigma} y
$$

$$
K_x(x,y) = D^* \kappa_{p,z}(Dx, y) \qquad K_y(x,y) = \kappa_{p,y}(Dx, y),
$$

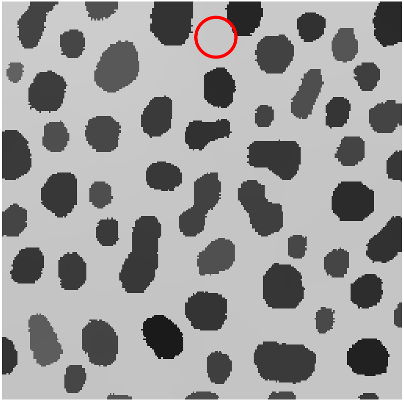
$p = \infty$:

$$
\begin{aligned}
[\kappa_{\infty,z}(z,y)]_{ij,k} &= 2(1 - z_{ij,1} y_{ij,1} - z_{ij,2} y_{ij,2}) y_{ij,k} \\
[\kappa_{\infty,y}(z,y)]_{ij,k} &= 2(1 - z_{ij,1} y_{ij,1} - z_{ij,2} y_{ij,2}) z_{ij,k}
\end{aligned}
$$

Potts model denoising: example

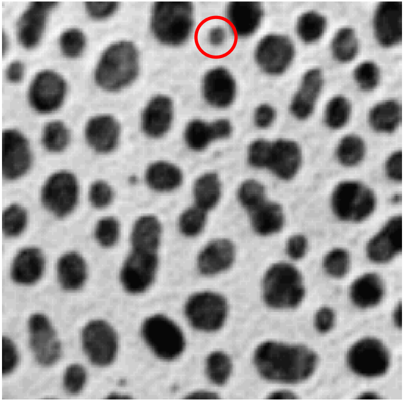


(a) original image f

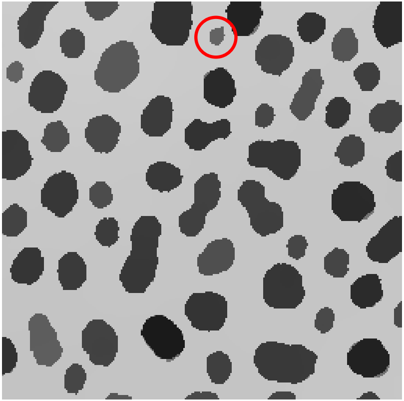


(b) x^N for $p = 1$

Potts model denoising: example



(a) original image f



(b) x^N for $p = \infty$

Potts model denoising: results

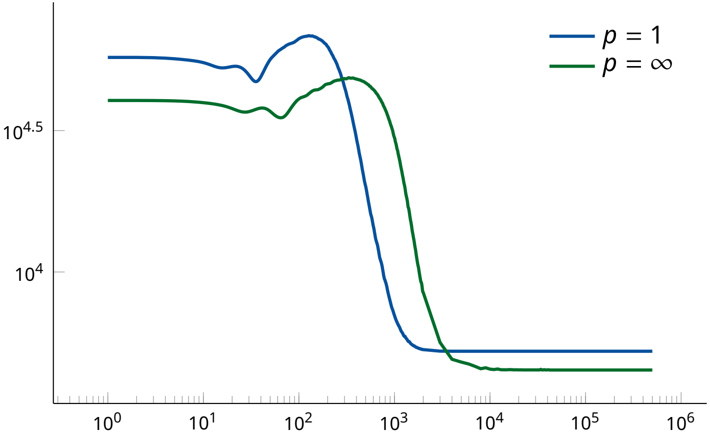


Figure: function values

Potts model denoising: results

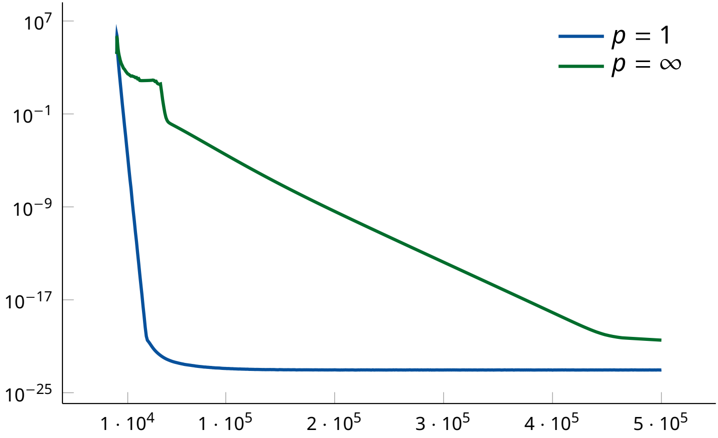


Figure: primal-dual errors

1 Overview

2 Generalized conjugation

3 Generalized primal-dual proximal splitting

4 Potts model denoising

5 Nash equilibrium problems

6 Conclusion

Nash equilibrium problems

Application: non-cooperative n-player games (in Hilbert spaces)

Player k:

- strategy $x_k \in X_k$
- payout function $\varphi_k : X := X_1 \times \cdots \times X_n \to \mathbb{R}$

Strategy vector $x := (x_1, \ldots, x_n)$; after player k changes $x_k \mapsto z$:

$$(x_{-k}|z) := (x_1, \ldots, x_{k-1}, z, x_{k+1}, \ldots x_n) \qquad (1 \leq k \leq n,\ z \in X_k)$$

Strategy vector $x^* \in X$ Nash equilibrium if

$$\varphi_k(x^*) = \varphi_k(x^*_{-k}|x^*_k) = \min_{z \in X_K} \varphi_k(x^*_{-k}|z) \qquad (1 \leq k \leq n).$$

(no player can change only their own strategy without being worse off)

Nash equilibrium problems: reformulation

Nikaido–Isoda function (Ky Fan function)

$$\Psi(x, y) = \sum_{k=1}^{n} \left(\varphi_k(x_{-k}|x_k) - \varphi_k(x_{-k}|y_k)\right) \qquad (x, y \in X)$$

Optimum response function

$$V(x) = \max_{y \in X} \Psi(x, y) \qquad (x \in X).$$

$\rightsquigarrow$ $x^* \in X$ Nash equilibrium iff solution to

$$\min_x \max_y \delta_X(x) + \Psi(x, y) - \delta_X(y).$$

$\rightsquigarrow$ generalized primal-dual proximal splitting

Application: elliptic Nash equilibrium problems

Two-player elliptic Nash equilibrium problem (ENEP):

- strategies $u_k \in L^2(\omega_k)$, $\omega_k \subsetneq \Omega$

$$X_k = \left\{ w \in L^2(\omega_k) : w(x) \in [a,b] \text{ a.e. } x \in \omega_k \right\}$$

- payout function

$$\varphi_k(u_1,u_2) = \frac{1}{2}\|S(u_1,u_2) - z_k\|^2_{L^2(\Omega)} + \frac{\alpha_k}{2}\|u_k\|^2_{L^2(\omega_k)}$$

- $S : L^2(\omega_1) \times L^2(\omega_2) \longrightarrow L^2(\Omega), \quad u \mapsto y$ solving

$$\begin{cases} -\Delta y = u_1 + u_2 + f & \text{on } \Omega, \\ \quad y = 0 & \text{on } \partial\Omega, \end{cases}$$

ENEP: Algorithm

$$u^{i+1} = \text{proj}_{[a,b]}(x^i - \tau_i K_u(u^i, v^i))$$
$$\bar{u}^{i+1} = 2u^{i+1} - u^i$$
$$v^{i+1} = \text{prox}_{[a,b]}(v^i + \sigma_{i+1} K_v(\bar{u}^{i+1}, v^i))$$

adjoint calculus, linearity of adjoint equation:

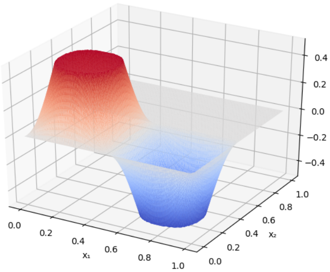
$$K_u(u,v) = \begin{pmatrix} p_1(u,v) + \alpha_1 u_1 \\ p_2(u,v) + \alpha_2 u_2 \end{pmatrix} \qquad K_v(u,v) = \begin{pmatrix} q_1(u,v) - \alpha_1 v_1 \\ q_2(u,v) - \alpha_2 v_2 \end{pmatrix}$$

with

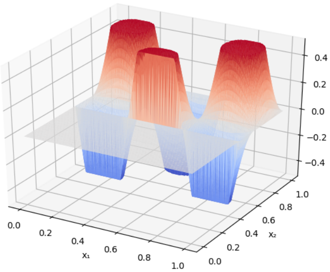
$$-\Delta p_1 = 2S(u_1,u_2) - S(u_1,v_2) - z_1 \qquad -\Delta q_1 = -S(v_1,u_2) + z_1$$
$$-\Delta p_2 = 2S(u_1,u_2) - S(v_1,u_2) - z_2 \qquad -\Delta q_2 = -S(u_1,v_2) + z_2$$

$\rightsquigarrow$ 9 PDE solves per iteration

ENEP: Results



(a) u_1^*



(b) u_2^*

Figure: Constructed solution for elliptic NEP example ($N = 128$)

ENEP: Results

Table: Results for elliptic NEP example for different N

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

Conclusion

Generalized conjugation:

- convex reformulation of nonconvex problems …
- …using nonlinear coupling term
- $\rightsquigarrow$ generalized primal-dual proximal splitting
- applicable to Potts model denoising
- $\rightsquigarrow$ linear convergence with Huber regularization

Outlook:

- convergence in Hilbert space (stronger conditions on K)
- application to (generalized) Nash equilibrium problems
- application to other imaging or inverse problems?

Preprints, Julia code:

`http://homepage.uni-graz.at/c.clason/publikationen`