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

10th International Congress on Industrial and Applied Mathematics
Tokyo, August 21, 2023

# 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)

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

- $\text{prox}_G$, $\text{prox}_{F^*}$ proximal point mappings (projection)
- $\tau_i, \sigma_i$ step lengths
- $\omega_i$ overrelaxation

# Motivation: nonconvex optimization in imaging

## $\ell^0$-TV-denoising: Potts model

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

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

# 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:

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

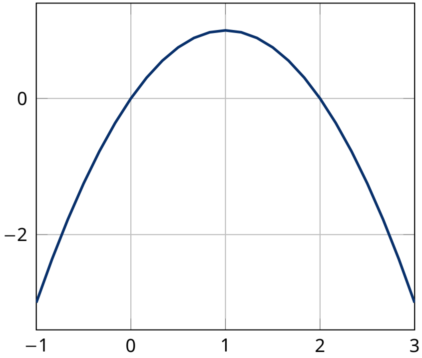
Example (smooth):

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

# Generalized conjugation: scalar motivation

Example (smooth):

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

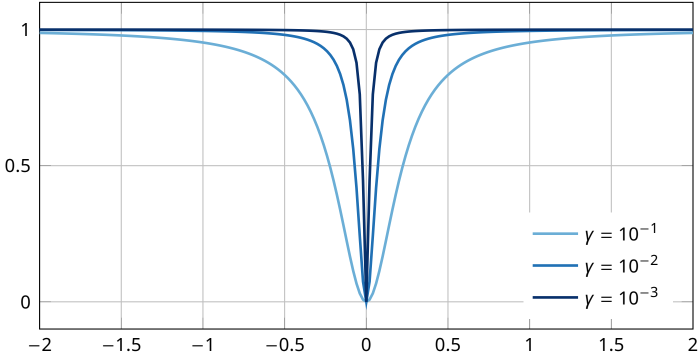


# Generalized conjugation: scalar motivation

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: $\ell^0$-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: $\ell^0$-TV

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

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

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

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

# Generalized primal-dual proximal splitting

## Saddle point problem

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

Primal-dual proximal splitting

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

- $\mathrm{prox}_G$, $\mathrm{prox}_{F^*}$ proximal point mappings (projection)
- $\tau_i, \sigma_i$ step lengths
- $\omega_i$ 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

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

- $\operatorname{prox}_G$, $\operatorname{prox}_{F^*}$ proximal point mappings (projection)
- $\tau_i, \sigma_i$ step lengths
- $\omega_i$ 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)

# 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

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

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

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

$p = 1$:

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

# Potts model denoising

## Generalized Primal-dual proximal splitting

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

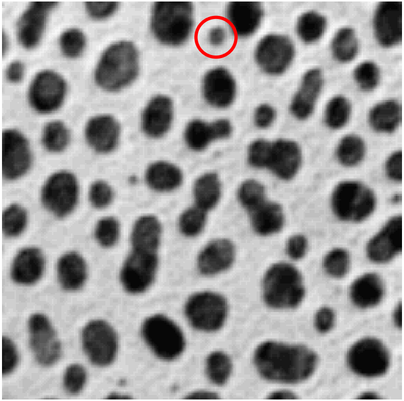
$$
\mathrm{prox}_{\tau G}(x) = \frac{1}{1 + \frac{\tau}{\alpha}} \left( x + \frac{\tau}{\alpha} f \right), \qquad \mathrm{prox}_{\sigma F_\gamma^*}(y) = \frac{1}{1 + \gamma\sigma} y
$$

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

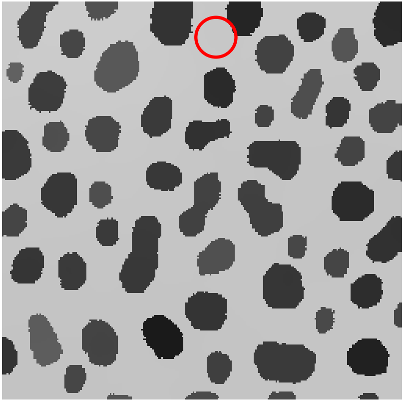
$p = \infty$:

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

# Potts model denoising: example

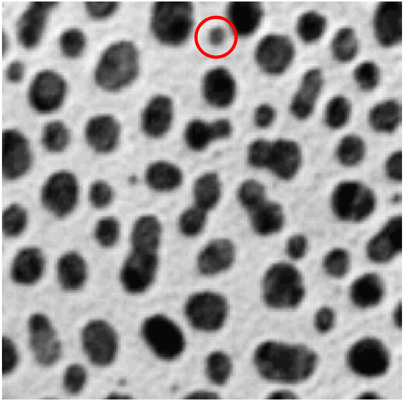


(a) original image $f$

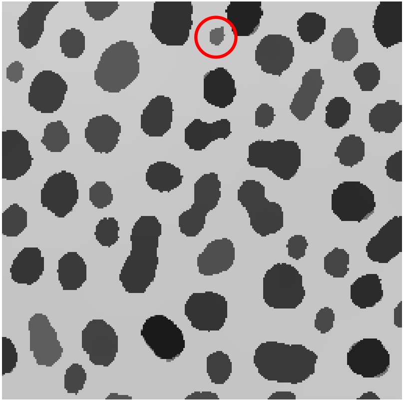


(b) $x^N$ for $p = 1$

# Potts model denoising: example



(a) original image $f$



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

# Potts model denoising: results

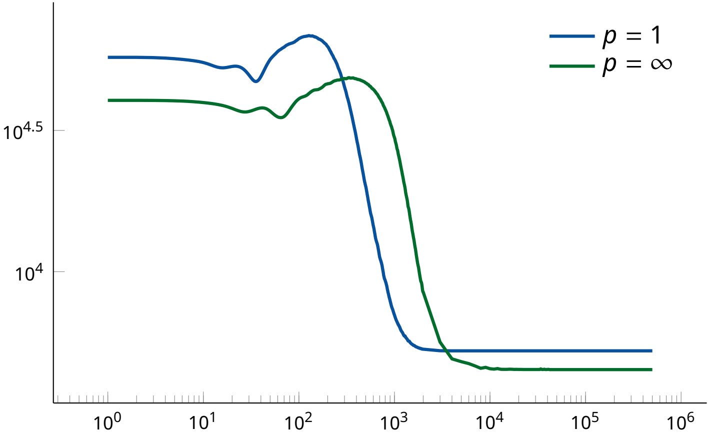



Figure: function values

# Potts model denoising: results

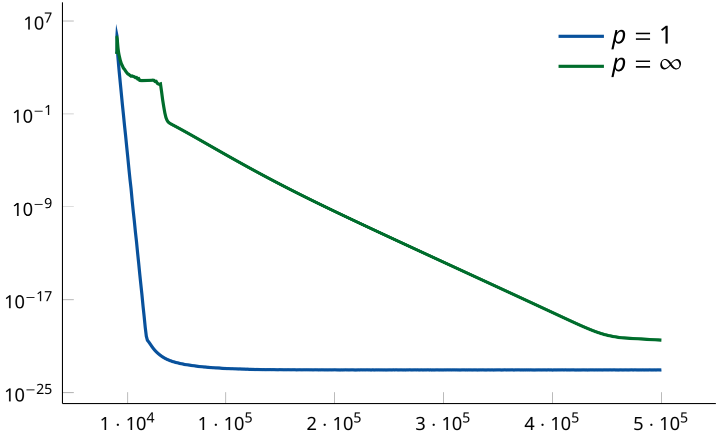



Figure: primal-dual errors

# 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 Nash equilibrium problems
- application to other imaging or inverse problems?

Preprints, codes:

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