The Primal-Dual Hybrid Gradient Method for Semiconvex Splittings

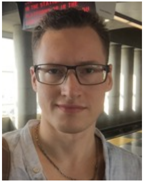
SIAM Conference on Imaging Science, Albuquerque
Minisymposium on Non-Convex Regularization Methods in Image Restoration

May 26th, 2016

Thomas Möllenhoff

Computer Vision Group
Department of Computer Science
Technical University of Munich

Joint work with:



Evgeny Strekalovskiy

Michael Moeller

Daniel Cremers

Motivation: First-Order Splitting Methods

- Many relevant optimization problems in image processing, computer vision, machine learning have structured form

$$\min_{x \in \mathbb{R}^n} \; G(x) + F(Kx)$$

Motivation: First-Order Splitting Methods

- Many relevant optimization problems in image processing, computer vision, machine learning have structured form

$$\min_{x \in \mathbb{R}^n} \; G(x) + F(Kx)$$

- Usually $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a data fidelity term and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ a regularization term encouraging spatial smoothness, e.g. by some linear gradient operator $K : \mathbb{R}^n \to \mathbb{R}^m$

Motivation: First-Order Splitting Methods

- Many relevant optimization problems in image processing, computer vision, machine learning have structured form

$$\min_{x \in \mathbb{R}^n} \; G(x) + F(Kx)$$

- Usually $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a data fidelity term and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ a regularization term encouraging spatial smoothness, e.g. by some linear gradient operator $K : \mathbb{R}^n \to \mathbb{R}^m$
- Trend: employ first-order splitting methods to quickly solve the above up to modest accuracy

Motivation: First-Order Splitting Methods

- Many relevant optimization problems in image processing, computer vision, machine learning have structured form

$$\min_{x \in \mathbb{R}^n} \; G(x) + F(Kx)$$

- Usually $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a data fidelity term and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ a regularization term encouraging spatial smoothness, e.g. by some linear gradient operator $K : \mathbb{R}^n \to \mathbb{R}^m$
- Trend: employ first-order splitting methods to quickly solve the above up to modest accuracy
- A popular and versatile algorithm is the primal-dual hybrid-gradient (PDHG) method [Pock, Cremers, Bischof, Chambolle '09], [Esser, Zhang, Chan '10], [Pock, Chambolle '11]

Motivation: First-Order Splitting Methods

- Many relevant optimization problems in image processing, computer vision, machine learning have structured form

$$\min_{x \in \mathbb{R}^n} \; G(x) + F(Kx)$$

- Usually $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a data fidelity term and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ a regularization term encouraging spatial smoothness, e.g. by some linear gradient operator $K : \mathbb{R}^n \to \mathbb{R}^m$
- Trend: employ first-order splitting methods to quickly solve the above up to modest accuracy
- A popular and versatile algorithm is the primal-dual hybrid-gradient (PDHG) method [Pock, Cremers, Bischof, Chambolle '09], [Esser, Zhang, Chan '10], [Pock, Chambolle '11]
- Well-established theory in the convex setting

Motivation: Nonconvex Regularization

- Popular choice of regularizer are discrete TV-type energies, $\varphi : \mathbb{R} \to \mathbb{R}$

$$F(\nabla x) = \sum_{i \in \Omega} \varphi \left(\|(\nabla x)_i\| \right)$$

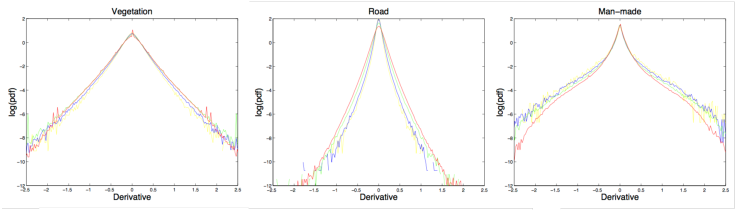
Motivation: Nonconvex Regularization

- Popular choice of regularizer are discrete TV-type energies, $\varphi : \mathbb{R} \to \mathbb{R}$

$$F(\nabla x) = \sum_{i \in \Omega} \varphi\left(\|(\nabla x)_i\|\right)$$

- Derivative statistics of natural images suggest nonconvex φ [Huang, Mumford '09]

$$\varphi(t) = t^q, \ q < 1$$



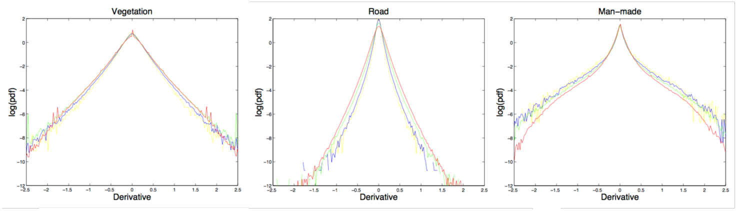
Motivation: Nonconvex Regularization

- Popular choice of regularizer are discrete TV-type energies, $\varphi : \mathbb{R} \to \mathbb{R}$

$$F(\nabla x) = \sum_{i \in \Omega} \varphi\left(\|(\nabla x)_i\|\right)$$

- Derivative statistics of natural images suggest nonconvex φ [Huang, Mumford '09]

$$\varphi(t) = t^q, \ q < 1$$



- Important limit case is the Potts model ($q = 0$) [Potts '52] (unsupervised segmentation, image cartooning, blind deblurring, ...)

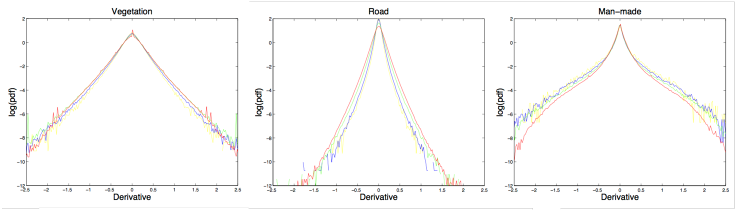
Motivation: Nonconvex Regularization

- Popular choice of regularizer are discrete TV-type energies, $\varphi : \mathbb{R} \to \mathbb{R}$

$$F(\nabla x) = \sum_{i \in \Omega} \varphi\left(\|(\nabla x)_i\|\right)$$

- Derivative statistics of natural images suggest nonconvex φ [Huang, Mumford '09]

$$\varphi(t) = t^q, \; q < 1$$



- Important limit case is the Potts model ($q = 0$) [Potts '52] (unsupervised segmentation, image cartooning, blind deblurring, ...)
- Piecewise smooth approximations [Blake, Zisserman '87], [Geman, Geman '84]

$$\varphi(t) = \min\{\lambda, \alpha t^2\}$$

Motivation: Nonconvex Proximal Mappings

- Key computation in most splitting methods is evaluation of *proximal mapping*

$$\mathrm{prox}_{\tau,f}(z) := \arg\min_{x \in \mathbb{R}^n} \; f(x) + \frac{\|x - z\|^2}{2\tau}$$

Motivation: Nonconvex Proximal Mappings

- Key computation in most splitting methods is evaluation of *proximal mapping*

$$\mathsf{prox}_{\tau,f}(z) := \arg\min_{x\in\mathbb{R}^n} \; f(x) + \frac{\|x-z\|^2}{2\tau}$$

- Often can be efficiently (and globally optimally) evaluated even if f is nonconvex

Motivation: Nonconvex Proximal Mappings

- Key computation in most splitting methods is evaluation of *proximal mapping*

$$\mathsf{prox}_{\tau,f}(z) := \arg\min_{x \in \mathbb{R}^n} \ f(x) + \frac{\|x - z\|^2}{2\tau}$$

- Often can be efficiently (and globally optimally) evaluated even if f is nonconvex
- Can apply most proximal splitting algorithms “as is” in the nonconvex setting

Motivation: Nonconvex Proximal Mappings

- Key computation in most splitting methods is evaluation of *proximal mapping*

$$\mathsf{prox}_{\tau,f}(z) := \arg\min_{x \in \mathbb{R}^n} \; f(x) + \frac{\|x - z\|^2}{2\tau}$$

- Often can be efficiently (and globally optimally) evaluated even if f is nonconvex
- Can apply most proximal splitting algorithms "as is" in the nonconvex setting
- The use of nonconvex proximal mappings emerged around ≈ 2009, amongst many others [Chartrand '09], [Blumensath, Davies '09], [Fornasier, Ward '09], ...

Motivation: Nonconvex Proximal Mappings

- Key computation in most splitting methods is evaluation of *proximal mapping*

$$\text{prox}_{\tau,f}(z) := \arg\min_{x \in \mathbb{R}^n} \; f(x) + \frac{\|x - z\|^2}{2\tau}$$

- Often can be efficiently (and globally optimally) evaluated even if f is nonconvex
- Can apply most proximal splitting algorithms "as is" in the nonconvex setting
- The use of nonconvex proximal mappings emerged around ≈ 2009, amongst many others [Chartrand '09], [Blumensath, Davies '09], [Fornasier, Ward '09], ...
- Remark: many other possibilities exist for nonsmooth nonconvex optimization

The Primal-Dual Hybrid Gradient Method (PDHG)

- Idea: introduce variable splitting for constraint $z = Kx$

$$\min_{x,z} \; G(x) + F(z), \quad \text{s.t. } z = Kx$$

The Primal-Dual Hybrid Gradient Method (PDHG)

- Idea: introduce variable splitting for constraint $z = Kx$

$$\min_{x,z} \; G(x) + F(z), \quad \text{s.t. } z = Kx$$

- Add Lagrange multiplier y for linear constraint

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

The Primal-Dual Hybrid Gradient Method (PDHG)

- Idea: introduce variable splitting for constraint $z = Kx$

$$\min_{x,z} \; G(x) + F(z), \quad \text{s.t. } z = Kx$$

- Add Lagrange multiplier y for linear constraint

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Under convexity assumptions equivalent to

$$\min_{x} \max_{y} \; G(x) - F^*(y) + \langle y, Kx \rangle$$

The Primal-Dual Hybrid Gradient Method (PDHG)

- Idea: introduce variable splitting for constraint $z = Kx$

$$\min_{x,z} \; G(x) + F(z), \quad \text{s.t. } z = Kx$$

- Add Lagrange multiplier y for linear constraint

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Under convexity assumptions equivalent to

$$\min_{x} \max_{y} \; G(x) - F^*(y) + \langle y, Kx \rangle$$

- PDHG algorithm performs alternating ascent/descent in y and x, followed by a subsequent overrelaxation

$$\begin{aligned} y^{k+1} &= \mathrm{prox}_{\sigma, F^*}\left(y^k + \sigma K \bar{x}^k\right), \\ x^{k+1} &= \mathrm{prox}_{\tau, G}\left(x^k - \tau K^T y^{k+1}\right), \\ \bar{x}^{k+1} &= x^{k+1} + \theta\left(x^{k+1} - x^k\right). \end{aligned}$$

The Primal-Dual Hybrid Gradient Method (PDHG)

- Idea: introduce variable splitting for constraint $z = Kx$

$$\min_{x,z} \; G(x) + F(z), \quad \text{s.t. } z = Kx$$

- Add Lagrange multiplier y for linear constraint

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Under convexity assumptions equivalent to

$$\min_{x} \max_{y} \; G(x) - F^*(y) + \langle y, Kx \rangle$$

- PDHG algorithm performs alternating ascent/descent in y and x, followed by a subsequent overrelaxation

$$\begin{aligned} y^{k+1} &= \text{prox}_{\sigma, F^*}\left(y^k + \sigma K \bar{x}^k\right), \\ x^{k+1} &= \text{prox}_{\tau, G}\left(x^k - \tau K^T y^{k+1}\right), \\ \bar{x}^{k+1} &= x^{k+1} + \theta\left(x^{k+1} - x^k\right). \end{aligned}$$

- Convergence to saddle-point $(\,\widehat{x},\ \widehat{y}\,)$ for $\tau\sigma\|K\|^2 \leq 1$, $\theta = 1$

The Primal-Dual Hybrid Gradient Method (PDHG)

- Idea: introduce variable splitting for constraint $z = Kx$

$$\min_{x,z} \; G(x) + F(z), \quad \text{s.t. } z = Kx$$

- Add Lagrange multiplier y for linear constraint

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Under convexity assumptions equivalent to

$$\min_{x} \max_{y} \; G(x) - F^*(y) + \langle y, Kx \rangle$$

- PDHG algorithm performs alternating ascent/descent in y and x, followed by a subsequent overrelaxation

$$\begin{aligned}
y^{k+1} &= \text{prox}_{\sigma,F^*}\left(y^k + \sigma K \bar{x}^k\right), \\
x^{k+1} &= \text{prox}_{\tau,G}\left(x^k - \tau K^T y^{k+1}\right), \\
\bar{x}^{k+1} &= x^{k+1} + \theta\left(x^{k+1} - x^k\right).
\end{aligned}$$

- Convergence to saddle-point $(\ \widehat{x},\ \widehat{y}\ )$ for $\tau\sigma\|K\|^2 \leq 1$, $\theta = 1$

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

- For the previous penalty functions φ, we have $F^{**} \equiv 0$

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

- For the previous penalty functions φ, we have $F^{**} \equiv 0$
- The nonconvex aspects of the regularizer do not enter the algorithm

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

- For the previous penalty functions φ, we have $F^{**} \equiv 0$
- The nonconvex aspects of the regularizer do not enter the algorithm
- Idea: formulate the algorithm in terms of the primal variable z

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

- For the previous penalty functions φ, we have $F^{**} \equiv 0$
- The nonconvex aspects of the regularizer do not enter the algorithm
- Idea: formulate the algorithm in terms of the primal variable z

$$\max_y \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Leads to the following [Strekalovskiy, Cremers '14], [M., et al. '15]:

$$\begin{aligned} z^{k+1} &= \text{prox}_{1/\sigma, F} \left(y^k/\sigma + K\bar{x}^k \right), \\ y^{k+1} &= y^k + \sigma(K\bar{x}^k - z^{k+1}), \\ x^{k+1} &= \text{prox}_{\tau, G} \left(x^k - \tau K^T y^{k+1} \right), \\ \bar{x}^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k). \end{aligned}$$

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

- For the previous penalty functions φ, we have $F^{**} \equiv 0$
- The nonconvex aspects of the regularizer do not enter the algorithm
- Idea: formulate the algorithm in terms of the primal variable z

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Leads to the following [Strekalovskiy, Cremers '14], [M., et al. '15]:

$$\begin{aligned} z^{k+1} &= \text{prox}_{1/\sigma,F}\left(y^k/\sigma + K\bar{x}^k\right), \\ y^{k+1} &= y^k + \sigma(K\bar{x}^k - z^{k+1}), \\ x^{k+1} &= \text{prox}_{\tau,G}\left(x^k - \tau K^T y^{k+1}\right), \\ \bar{x}^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k). \end{aligned}$$

- **Proposition:** Equivalent to original algorithm for convex F

Equivalent Reformulation of the PDHG

- Algorithm solves convex relaxation due to the occurance of F^* in y^{k+1}-step

$$\min_{x \in \mathbb{R}^n} G(x) + F^{**}(Kx)$$

- For the previous penalty functions φ, we have $F^{**} \equiv 0$
- The nonconvex aspects of the regularizer do not enter the algorithm
- Idea: formulate the algorithm in terms of the primal variable z

$$\max_{y} \min_{x,z} \; G(x) + F(z) + \langle y, Kx - z \rangle$$

- Leads to the following [Strekalovskiy, Cremers '14], [M., et al. '15]:

$$\begin{aligned} z^{k+1} &= \text{prox}_{1/\sigma, F}\left(y^k/\sigma + K\bar{x}^k\right), \\ y^{k+1} &= y^k + \sigma(K\bar{x}^k - z^{k+1}), \\ x^{k+1} &= \text{prox}_{\tau, G}\left(x^k - \tau K^T y^{k+1}\right), \\ \bar{x}^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k). \end{aligned}$$

- **Proposition:** Equivalent to original algorithm for convex F
- Can be applied to nonconvex F in a meaningful way

Application: Minimization of the Mumford-Shah functional

- Reformulated PDHG applied to Mumford-Shah [Strekalovskiy, Cremers '14]

$$\min_{u:\Omega\to\mathbb{R}^k} \underbrace{\frac{1}{2}\|u-f\|^2}_{=:G(u)} + \underbrace{\sum_{i\in\Omega} \min\{\lambda, \alpha\|(\nabla u)_i\|^2\}}_{=:F(\nabla u)}$$

Application: Minimization of the Mumford-Shah functional

- Reformulated PDHG applied to Mumford-Shah [Strekalovskiy, Cremers '14]

$$\min_{u:\Omega\to\mathbb{R}^k} \underbrace{\frac{1}{2}\|u-f\|^2}_{=:G(u)} + \underbrace{\sum_{i\in\Omega} \min\{\lambda, \alpha\|(\nabla u)_i\|^2\}}_{=:F(\nabla u)}$$

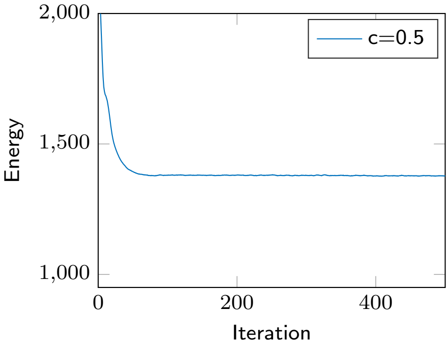
- Behaviour of algorithm with constant steps is similar to subgradient method

Application: Minimization of the Mumford-Shah functional

- Reformulated PDHG applied to Mumford-Shah [Strekalovskiy, Cremers '14]

$$\min_{u:\Omega\to\mathbb{R}^k} \underbrace{\frac{1}{2}\|u-f\|^2}_{=:G(u)} + \underbrace{\sum_{i\in\Omega} \min\{\lambda, \alpha\|(\nabla u)_i\|^2\}}_{=:F(\nabla u)}$$

- Behaviour of algorithm with constant steps is similar to subgradient method
- Set for some $c > 0$, $\tau \propto c$, $\sigma \propto \frac{1}{c}$, $\tau\sigma\|\nabla\|^2 \leq 1$, fixed $\theta = 1$

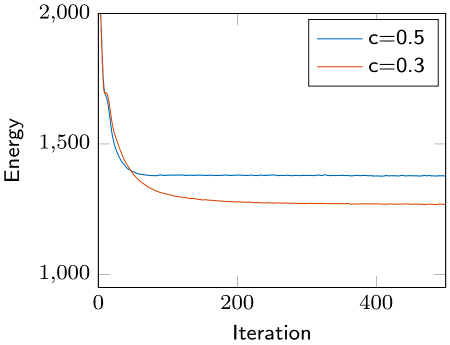


Application: Minimization of the Mumford-Shah functional

- Reformulated PDHG applied to Mumford-Shah [Strekalovskiy, Cremers '14]

$$\min_{u:\Omega\to\mathbb{R}^k} \underbrace{\frac{1}{2}\|u-f\|^2}_{=:G(u)} + \underbrace{\sum_{i\in\Omega} \min\{\lambda, \alpha\|(\nabla u)_i\|^2\}}_{=:F(\nabla u)}$$

- Behaviour of algorithm with constant steps is similar to subgradient method
- Set for some $c>0$, $\tau \propto c$, $\sigma \propto \frac{1}{c}$, $\tau\sigma\|\nabla\|^2 \leq 1$, fixed $\theta = 1$

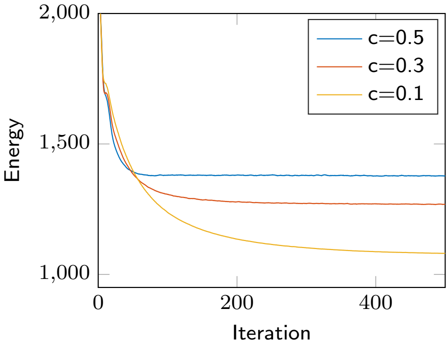


Application: Minimization of the Mumford-Shah functional

- Reformulated PDHG applied to Mumford-Shah [Strekalovskiy, Cremers '14]

$$\min_{u:\Omega\to\mathbb{R}^k} \underbrace{\frac{1}{2}\|u-f\|^2}_{=:G(u)} + \underbrace{\sum_{i\in\Omega} \min\{\lambda, \alpha\|(\nabla u)_i\|^2\}}_{=:F(\nabla u)}$$

- Behaviour of algorithm with constant steps is similar to subgradient method
- Set for some $c > 0$, $\tau \propto c$, $\sigma \propto \frac{1}{c}$, $\tau\sigma\|\nabla\|^2 \leq 1$, fixed $\theta = 1$



Application: Minimization of the Mumford-Shah functional

- Motivates adaptive choice of steps with $\tau_k \to 0$, $\sigma_k \to +\infty$

Application: Minimization of the Mumford-Shah functional

- Motivates adaptive choice of steps with $\tau_k \to 0$, $\sigma_k \to +\infty$
- Possible choice which works extremely well in practice, for $\gamma > 0$:

$$\theta_k = \frac{1}{\sqrt{1 + 2\gamma\tau_k}},\ \sigma_{k+1} = \sigma_k/\theta_k,\ \tau_{k+1} = \tau_k\theta_k,$$

Application: Minimization of the Mumford-Shah functional

- Motivates adaptive choice of steps with $\tau_k \to 0$, $\sigma_k \to +\infty$
- Possible choice which works extremely well in practice, for $\gamma > 0$:

$$\theta_k = \frac{1}{\sqrt{1 + 2\gamma\tau_k}},\ \sigma_{k+1} = \sigma_k / \theta_k,\ \tau_{k+1} = \tau_k \theta_k,$$

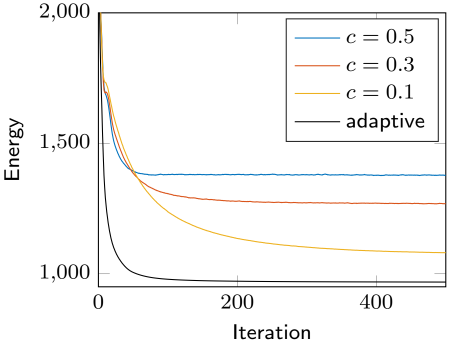
- In convex setting gives $\mathcal{O}(1/k^2)$ convergence if either F^* or G is strongly convex with modulus γ [Chambolle, Pock '11]

Application: Minimization of the Mumford-Shah functional

- Motivates adaptive choice of steps with $\tau_k \to 0$, $\sigma_k \to +\infty$
- Possible choice which works extremely well in practice, for $\gamma > 0$:

$$\theta_k = \frac{1}{\sqrt{1 + 2\gamma\tau_k}}, \ \sigma_{k+1} = \sigma_k / \theta_k, \ \tau_{k+1} = \tau_k \theta_k,$$

- In convex setting gives $\mathcal{O}(1/k^2)$ convergence if either F^* or G is strongly convex with modulus γ [Chambolle, Pock '11]

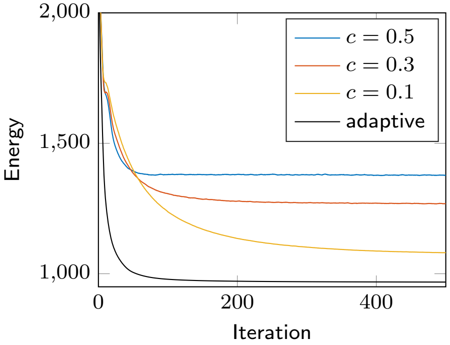


Application: Minimization of the Mumford-Shah functional

- Motivates adaptive choice of steps with $\tau_k \to 0$, $\sigma_k \to +\infty$
- Possible choice which works extremely well in practice, for $\gamma > 0$:

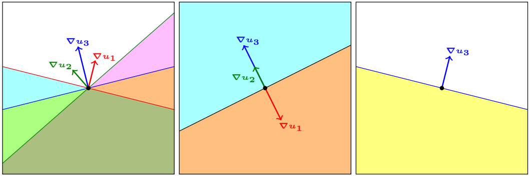
$$\theta_k = \frac{1}{\sqrt{1 + 2\gamma\tau_k}}, \; \sigma_{k+1} = \sigma_k/\theta_k, \; \tau_{k+1} = \tau_k\theta_k,$$

- In convex setting gives $\mathcal{O}(1/k^2)$ convergence if either F^* or G is strongly convex with modulus γ [Chambolle, Pock '11]



- On recent GPU: ≈ 30ms for 640×480 color image [Strekalovskiy, Cremers '14] ($\to$ application: real-time video cartooning!)

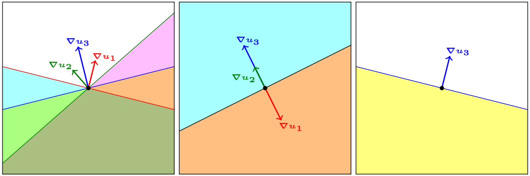
TV^q and TGV^q-like Regularization for Color Images



- For color images $u : \Omega \to \mathbb{R}^3$, consider at every pixel $i \in \Omega$ the Jacobian matrix

$$(\nabla u)_i = \begin{pmatrix} (\partial_x u_1)_i & (\partial_x u_2)_i & (\partial_x u_3)_i \\ (\partial_y u_1)_i & (\partial_y u_2)_i & (\partial_y u_3)_i \end{pmatrix}$$

TV^q and TGV^q-like Regularization for Color Images

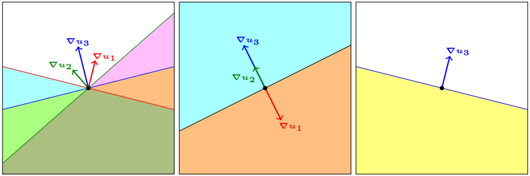


- For color images $u : \Omega \to \mathbb{R}^3$, consider at every pixel $i \in \Omega$ the Jacobian matrix

$$(\nabla u)_i = \begin{pmatrix} (\partial_x u_1)_i & (\partial_x u_2)_i & (\partial_x u_3)_i \\ (\partial_y u_1)_i & (\partial_y u_2)_i & (\partial_y u_3)_i \end{pmatrix}$$

- Possible extensions of total variation to color images by considering various matrix norms of Jacobian [Sapiro, Ringach '96], [Bresson, Chan '08]

TV^q and TGV^q-like Regularization for Color Images



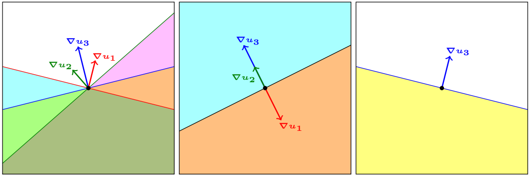
- For color images $u : \Omega \to \mathbb{R}^3$, consider at every pixel $i \in \Omega$ the Jacobian matrix

$$(\nabla u)_i = \begin{pmatrix} (\partial_x u_1)_i & (\partial_x u_2)_i & (\partial_x u_3)_i \\ (\partial_y u_1)_i & (\partial_y u_2)_i & (\partial_y u_3)_i \end{pmatrix}$$

- Possible extensions of total variation to color images by considering various matrix norms of Jacobian [Sapiro, Ringach '96], [Bresson, Chan '08]
- Nonconvex generalizations of above [M., Strekalovskiy, Moeller, Cremers '15]

$$TV_F^q(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_F^q, \; TV_{S^q}^q(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_{S^q}^q, \; q < 1$$

TV^q and TGV^q-like Regularization for Color Images



- For color images $u : \Omega \to \mathbb{R}^3$, consider at every pixel $i \in \Omega$ the Jacobian matrix

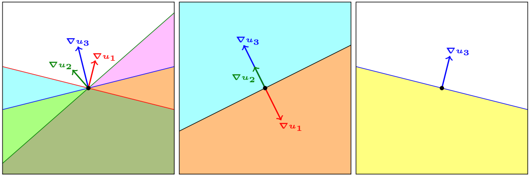
$$(\nabla u)_i = \begin{pmatrix} (\partial_x u_1)_i & (\partial_x u_2)_i & (\partial_x u_3)_i \\ (\partial_y u_1)_i & (\partial_y u_2)_i & (\partial_y u_3)_i \end{pmatrix}$$

- Possible extensions of total variation to color images by considering various matrix norms of Jacobian [Sapiro, Ringach '96], [Bresson, Chan '08]
- Nonconvex generalizations of above [M., Strekalovskiy, Moeller, Cremers '15]

$$TV_F^q(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_F^q, \; TV_{S^q}^q(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_{S^q}^q, \; q < 1$$

- Furthermore, we propose similar nonconvex generalizations for the total generalized variation (TGV) [Bredies '10], [Bredies '14]

TV^q and TGV^q-like Regularization for Color Images



- For color images $u : \Omega \to \mathbb{R}^3$, consider at every pixel $i \in \Omega$ the Jacobian matrix

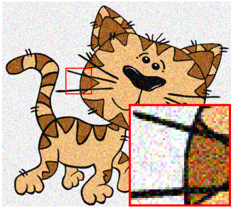
$$(\nabla u)_i = \begin{pmatrix} (\partial_x u_1)_i & (\partial_x u_2)_i & (\partial_x u_3)_i \\ (\partial_y u_1)_i & (\partial_y u_2)_i & (\partial_y u_3)_i \end{pmatrix}$$

- Possible extensions of total variation to color images by considering various matrix norms of Jacobian [Sapiro, Ringach '96], [Bresson, Chan '08]
- Nonconvex generalizations of above [M., Strekalovskiy, Moeller, Cremers '15]

$$TV_F^q(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_F^q, \; TV_{S^q}^q(u) = \sum_{i \in \Omega} \|(\nabla u)_i\|_{S^q}^q, \; q < 1$$

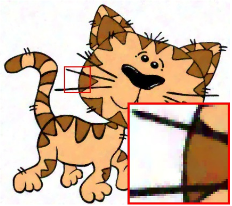
- Furthermore, we propose similar nonconvex generalizations for the total generalized variation (TGV) [Bredies '10], [Bredies '14]
- Can be efficiently solved using the nonconvex PDHG

TV^q and TGV^q-like Regularization for Color Images, $q = 1/2$



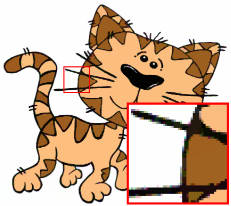
Noisy

$\sigma = 0.15$



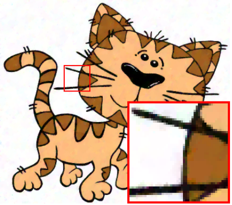
TV_F

PSNR=26.9



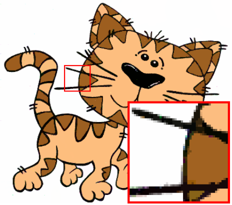
TV_F^q

PSNR=28.4



TV_{S^1}

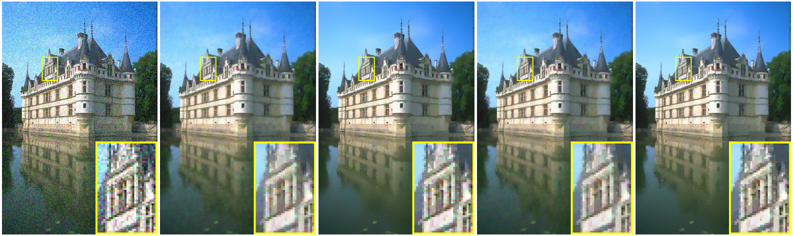
PSNR=27.9



$TV_{S^q}^q$

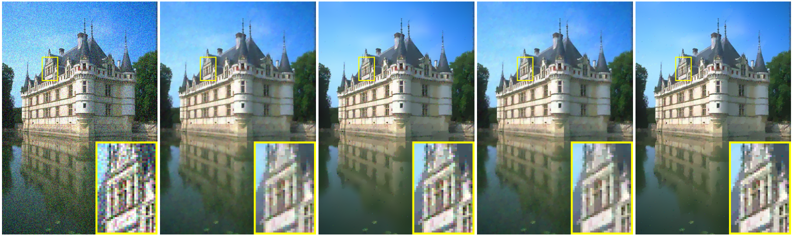
PSNR=**29.8**

TV^q and TGV^q-like Regularization for Color Images, $q = 3/4$



Noisy	TGV_F	TGV_F^q	TGV_{S^1}	$TGV_{S^q}^q$
$\sigma = 0.1$	PSNR=28.5	PSNR=28.9	PSNR=29.0	PSNR=**29.4**

TV^q and TGV^q-like Regularization for Color Images, $q = 3/4$



Noisy	TGV_F	TGV_F^q	TGV_{S^1}	$TGV_{S^q}^q$
$\sigma = 0.1$	PSNR=28.5	PSNR=28.9	PSNR=29.0	PSNR=**29.4**

Algorithm works well in practice. Theoretical convergence properties?

Convergence Analysis of Nonconvex PDHG

Theorem ([M., Strekalovskiy, Moeller, Cremers '15])

Let $G - \frac{c}{2}\|\cdot\|^2$ *and* $F + \frac{\omega}{2}\|\cdot\|^2$ *be convex with* $c > \omega\|K\|^2$*. Then the (ergodic) iterates* (X^k) *produced by the PDHG converge to the (unique) global minimizer*

$$\widehat{x} = \arg\min_x \; G(x) + F(Kx),$$

with $\|X^k - \widehat{x}\|^2 = \mathcal{O}(1/k)$ *for* $0 < \sigma = 2\omega, \; \tau\sigma\|K\|^2 \leq 1$*, and any* $\theta \in [0,1]$*.*

Convergence Analysis of Nonconvex PDHG

Theorem ([M., Strekalovskiy, Moeller, Cremers '15])

Let $G - \frac{c}{2}\|\cdot\|^2$ *and* $F + \frac{\omega}{2}\|\cdot\|^2$ *be convex with* $c > \omega\|K\|^2$. *Then the (ergodic) iterates* (X^k) *produced by the PDHG converge to the (unique) global minimizer*

$$\widehat{x} = \arg\min_{x} \; G(x) + F(Kx),$$

with $\|X^k - \widehat{x}\|^2 = \mathcal{O}(1/k)$ *for* $0 < \sigma = 2\omega$, $\tau\sigma\|K\|^2 \leq 1$, *and any* $\theta \in [0, 1]$.

- Notice that for our proof, σ has to be twice as big as to make the proximal minimization subproblem in z convex:

$$\text{prox}_{1/\sigma, F}(\tilde{z}) = \arg\min_{z} \; F(z) + \frac{\sigma}{2}\|z - \tilde{z}\|^2$$

Convergence Analysis of Nonconvex PDHG

Theorem ([M., Strekalovskiy, Moeller, Cremers '15])

Let $G - \frac{c}{2}\|\cdot\|^2$ and $F + \frac{\omega}{2}\|\cdot\|^2$ be convex with $c > \omega\|K\|^2$. Then the (ergodic) iterates (X^k) produced by the PDHG converge to the (unique) global minimizer

$$\widehat{x} = \arg\min_x \; G(x) + F(Kx),$$

with $\|X^k - \widehat{x}\|^2 = \mathcal{O}(1/k)$ for $0 < \sigma = 2\omega$, $\tau\sigma\|K\|^2 \leq 1$, and any $\theta \in [0,1]$.

- Notice that for our proof, σ has to be twice as big as to make the proximal minimization subproblem in z convex:

$$\text{prox}_{1/\sigma,F}(\tilde{z}) = \arg\min_z \; F(z) + \frac{\sigma}{2}\|z - \tilde{z}\|^2$$

- **A posteriori convergence** (similar to [Esser, Zhang '14]): if $\|x^{n+1} - x^n\| \to 0$ and $\|y^{n+1} - y^n\| \to 0$ and additionally (x^n), (y^n) and (z^n) are bounded then the iteration converges to critical points along subsequences

Convergence Analysis of Nonconvex PDHG

Theorem ([M., Strekalovskiy, Moeller, Cremers '15])

Let $G - \frac{c}{2}\|\cdot\|^2$ *and* $F + \frac{\omega}{2}\|\cdot\|^2$ *be convex with* $c > \omega\|K\|^2$. *Then the (ergodic) iterates* (X^k) *produced by the PDHG converge to the (unique) global minimizer*

$$\widehat{x} = \arg\min_{x} \; G(x) + F(Kx),$$

with $\|X^k - \widehat{x}\|^2 = \mathcal{O}(1/k)$ *for* $0 < \sigma = 2\omega$, $\tau\sigma\|K\|^2 \leq 1$, *and any* $\theta \in [0, 1]$.

- Notice that for our proof, σ has to be twice as big as to make the proximal minimization subproblem in z convex:

$$\text{prox}_{1/\sigma, F}(\tilde{z}) = \arg\min_{z} \; F(z) + \frac{\sigma}{2}\|z - \tilde{z}\|^2$$

- **A posteriori convergence** (similar to [Esser, Zhang '14]): if $\|x^{n+1} - x^n\| \to 0$ and $\|y^{n+1} - y^n\| \to 0$ and additionally (x^n), (y^n) and (z^n) are bounded then the iteration converges to critical points along subsequences
- **Theory-practice gap I:** convergence proof if overall energy is semiconvex (experiments indicate that in the overall nonconvex setting an additional requirement is differentiability of F)

Convergence Analysis of Nonconvex PDHG

Theorem ([M., Strekalovskiy, Moeller, Cremers '15])

Let $G - \frac{c}{2}\|\cdot\|^2$ and $F + \frac{\omega}{2}\|\cdot\|^2$ be convex with $c > \omega\|K\|^2$. Then the (ergodic) iterates (X^k) produced by the PDHG converge to the (unique) global minimizer

$$\widehat{x} = \arg\min_x \; G(x) + F(Kx),$$

with $\|X^k - \widehat{x}\|^2 = \mathcal{O}(1/k)$ *for* $0 < \sigma = 2\omega, \;\; \tau\sigma\|K\|^2 \leq 1$, *and any* $\theta \in [0, 1]$.

- Notice that for our proof, σ has to be twice as big as to make the proximal minimization subproblem in z convex:

$$\text{prox}_{1/\sigma, F}(\tilde{z}) = \arg\min_z \; F(z) + \frac{\sigma}{2}\|z - \tilde{z}\|^2$$

- **A posteriori convergence** (similar to [Esser, Zhang '14]): if $\|x^{n+1} - x^n\| \to 0$ and $\|y^{n+1} - y^n\| \to 0$ and additionally (x^n), (y^n) and (z^n) are bounded then the iteration converges to critical points along subsequences
- **Theory-practice gap I:** convergence proof if overall energy is semiconvex (experiments indicate that in the overall nonconvex setting an additional requirement is differentiability of F)
- **Theory-practice gap II:** for adaptive step sizes, experiments indicate that the algorithm converges for general nonconvex energies

Sharpness of the Step-Size Restriction and Consequences

- Consider the minimization of $\frac{\lambda-1}{2}x^2$ for some $\lambda > 1$:

$$\min_{x\in\mathbb{R}} \quad \underbrace{\frac{\lambda}{2}x^2}_{G(x)} \underbrace{-\frac{1}{2}x^2}_{F(x)} \tag{*}$$

Sharpness of the Step-Size Restriction and Consequences

- Consider the minimization of $\frac{\lambda-1}{2}x^2$ for some $\lambda > 1$:

$$\min_{x\in\mathbb{R}} \ \underbrace{\frac{\lambda}{2}x^2}_{G(x)} \ \underbrace{-\frac{1}{2}x^2}_{F(x)} \qquad (*)$$

- **Proposition:** the step-size restriction $\sigma \geq 2\omega$ is sharp. For any $\sigma < 2$ there exists a $\lambda > 1$ such that PDHG applied to (*) diverges

Sharpness of the Step-Size Restriction and Consequences

- Consider the minimization of $\frac{\lambda-1}{2}x^2$ for some $\lambda > 1$:

$$\min_{x\in\mathbb{R}} \underbrace{\frac{\lambda}{2}x^2}_{G(x)} \underbrace{-\frac{1}{2}x^2}_{F(x)} \tag{*}$$

- **Proposition:** the step-size restriction $\sigma \geq 2\omega$ is sharp. For any $\sigma < 2$ there exists a $\lambda > 1$ such that PDHG applied to (*) diverges
- **Consequence of our work:** for strongly convex energies, one has some additional modeling freedom by introducing (slightly) non-convex terms

Sharpness of the Step-Size Restriction and Consequences

- Consider the minimization of $\frac{\lambda-1}{2}x^2$ for some $\lambda > 1$:

$$\min_{x \in \mathbb{R}} \underbrace{\frac{\lambda}{2}x^2}_{G(x)} \underbrace{-\frac{1}{2}x^2}_{F(x)} \qquad (*)$$

- **Proposition:** the step-size restriction $\sigma \geq 2\omega$ is sharp. For any $\sigma < 2$ there exists a $\lambda > 1$ such that PDHG applied to (*) diverges
- **Consequence of our work:** for strongly convex energies, one has some additional modeling freedom by introducing (slightly) non-convex terms
- Example: enhanced ROF model, $\alpha > 0$ sufficiently small

$$\min_{u:\Omega \to \mathbb{R}^k} \frac{\lambda}{2}\|u - f\|^2 + \sum_{i \in \Omega} \|(\nabla u)_i\| + \alpha R_{\mathsf{ncvx}}(u)$$

Sharpness of the Step-Size Restriction and Consequences

- Consider the minimization of $\frac{\lambda-1}{2}x^2$ for some $\lambda > 1$:

$$\min_{x \in \mathbb{R}} \underbrace{\frac{\lambda}{2}x^2}_{G(x)} \underbrace{-\frac{1}{2}x^2}_{F(x)} \qquad (*)$$

- **Proposition:** the step-size restriction $\sigma \geq 2\omega$ is sharp. For any $\sigma < 2$ there exists a $\lambda > 1$ such that PDHG applied to (*) diverges
- **Consequence of our work:** for strongly convex energies, one has some additional modeling freedom by introducing (slightly) non-convex terms
- Example: enhanced ROF model, $\alpha > 0$ sufficiently small

$$\min_{u:\Omega \to \mathbb{R}^k} \frac{\lambda}{2}\|u - f\|^2 + \sum_{i \in \Omega} \|(\nabla u)_i\| + \alpha R_{\text{ncvx}}(u)$$

- Application: convex non-convex (CNC) models [Parekh, Selesnick '15], [Lanza, Morigi, Sgallari '16]

Recent Convergence Results for Nonconvex Splittings

- Recently, general convergence results have been obtained for nonconvex ADMM [Li, Pong '15], [Wang, Yin, Zeng '16] and Douglas-Rachford [Li, Pong '15]

Recent Convergence Results for Nonconvex Splittings

- Recently, general convergence results have been obtained for nonconvex ADMM [Li, Pong '15], [Wang, Yin, Zeng '16] and Douglas-Rachford [Li, Pong '15]
- For $K = I$, $\theta = 1$, PDHG is equivalent to Douglas-Rachford splitting and ADMM (convergence results carry over)

Recent Convergence Results for Nonconvex Splittings

- Recently, general convergence results have been obtained for nonconvex ADMM [Li, Pong '15], [Wang, Yin, Zeng '16] and Douglas-Rachford [Li, Pong '15]
- For $K = I$, $\theta = 1$, PDHG is equivalent to Douglas-Rachford splitting and ADMM (convergence results carry over)
- For $K \neq I$, $\theta = 1$, PDHG can be seen as inexact ADMM on the *dual problem* (hasn't been studied for nonconvex F to best of our knowledge)

Recent Convergence Results for Nonconvex Splittings

- Recently, general convergence results have been obtained for nonconvex ADMM [Li, Pong '15], [Wang, Yin, Zeng '16] and Douglas-Rachford [Li, Pong '15]
- For $K = I$, $\theta = 1$, PDHG is equivalent to Douglas-Rachford splitting and ADMM (convergence results carry over)
- For $K \neq I$, $\theta = 1$, PDHG can be seen as inexact ADMM on the *dual problem* (hasn't been studied for nonconvex F to best of our knowledge)
- Local convergence result for nonlinear K [Valkonen '13], can also be used to do nonconvex regularization [Shekhovtsov, Reinbacher, Graber, Pock '16]

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization
- Main theoretical result: PDHG converges for the splitting of a convex energy into a strongly convex and semiconvex part

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization
- Main theoretical result: PDHG converges for the splitting of a convex energy into a strongly convex and semiconvex part
- Step-sizes have to be chosen with care, $\sigma \geq 2\omega$

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization
- Main theoretical result: PDHG converges for the splitting of a convex energy into a strongly convex and semiconvex part
- Step-sizes have to be chosen with care, $\sigma \geq 2\omega$
- Practical implication: introduce nonconvexity in the presence of strong convexity

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization
- Main theoretical result: PDHG converges for the splitting of a convex energy into a strongly convex and semiconvex part
- Step-sizes have to be chosen with care, $\sigma \geq 2\omega$
- Practical implication: introduce nonconvexity in the presence of strong convexity
- Gaps between theory and practice: convergence analysis in fully nonconvex setting and convergence with adaptive step sizes are open problems

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization
- Main theoretical result: PDHG converges for the splitting of a convex energy into a strongly convex and semiconvex part
- Step-sizes have to be chosen with care, $\sigma \geq 2\omega$
- Practical implication: introduce nonconvexity in the presence of strong convexity
- Gaps between theory and practice: convergence analysis in fully nonconvex setting and convergence with adaptive step sizes are open problems
- Future work: can the recent tools developed for nonconvex ADMM/Douglas-Rachford be applied to extend our results for PDHG?

Conclusions / Future work

- Reformulated PDHG to be applicable to nonconvex F
- PDHG (with adaptive steps) is an efficient algorithm for nonconvex regularization
- Main theoretical result: PDHG converges for the splitting of a convex energy into a strongly convex and semiconvex part
- Step-sizes have to be chosen with care, $\sigma \geq 2\omega$
- Practical implication: introduce nonconvexity in the presence of strong convexity
- Gaps between theory and practice: convergence analysis in fully nonconvex setting and convergence with adaptive step sizes are open problems
- Future work: can the recent tools developed for nonconvex ADMM/Douglas-Rachford be applied to extend our results for PDHG?

Thank you for your attention!

thomas.moellenhoff@in.tum.de

References

1. T. Möllenhoff, E. Strekalovskiy, M. Moeller, D. Cremers. The primal-dual hybrid-gradient method for semiconvex splittings, SIAM Journal on Imaging Science, 2015.
2. T. Möllenhoff, E. Strekalovskiy, M. Moeller, D. Cremers. Low rank priors for color image regularization, Energy Minimization Methods in Computer Vision and Pattern Recognition, 2015.
3. E. Strekalovskiy, D. Cremers. Real-time minimization of the piecewise smooth Mumford-Shah functional, European Conference on Computer Vision, 2014.
4. T. Pock, D. Cremers, H. Bischof, A. Chambolle. An algorithm for minimizing the Mumford-Shah functional, International Conference on Computer Vision, 2009.