Variational denoising for manifold-valued data

Andreas Weinmann

Helmholtz Center Munich & TU München

Paris,
le 21 novembre 2014

Overview

An algorithm for TV minimization for manifold-valued data (joint work with L. Demaret and M.Storath)

Second order TV type functionals for $\mathbb{S}^1$-valued data (joint work with R. Bergmann, F. Laus, G. Steidl)

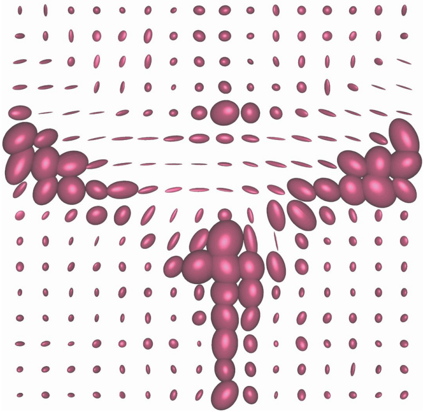
Potts and Blake-Zisserman functionals for manifold-valued signals with a few jumps (joint work with L. Demaret and M.Storath)

Manifold-valued data in DTI

- In diffusion tensor imaging (DTI) (Basser et al. '94) the data are positive(-definite) matrices.
- It is reasonable (cf. Pennec et al. '2004) to equip Pos_n with the Riemannian metric

$$g_P(A, B) = \mathrm{trace}(P^{-\frac{1}{2}} A P^{-1} B P^{-\frac{1}{2}}),$$

P positive and A, B symmetric.



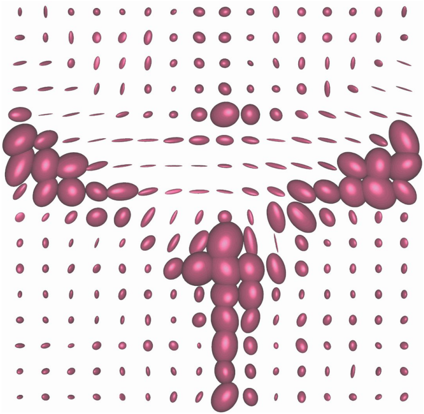
Positive Matrices visualized as ellipsoids.

Manifold-valued data in DTI

- In diffusion tensor imaging (DTI) (Basser et al. '94) the data are positive(-definite) matrices.
- It is reasonable (cf. Pennec et al. '2004) to equip Pos_n with the Riemannian metric

$$g_P(A,B) = \mathrm{trace}(P^{-\frac{1}{2}} A P^{-1} B P^{-\frac{1}{2}}),$$

  P positive and A, B symmetric.
- Pos_n with the metric g_P is a Cartan Hadamard manifold (complete, nonpositive sectional curvature, simply connected).



Positive Matrices visualized as ellipsoids.

- log and exp can be computed explicitly by

$$\log_P Q = P^{\frac{1}{2}} \log(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}}) P^{\frac{1}{2}}, \qquad \exp_P A = P^{\frac{1}{2}} \exp(P^{-\frac{1}{2}} A P^{-\frac{1}{2}}) P^{\frac{1}{2}}.$$

TV functionals for manifold-valued data

We consider the variational denoising problem given by the (discrete, anisotropic, bivariate) functionals

$$F_\alpha(u) = \sum_{i,j} \mathrm{dist}(u_i, f_i)^p + \alpha \sum_{i,j} \mathrm{dist}(u_{ij}, u_{i-1,j})^q + \alpha \sum_{i,j} \mathrm{dist}(u_{ij}, u_{i,j-1})^q,$$

with data f and $p, q \geq 1$.

- Choosing q=1 corresponds to (anisotropic) TV minimization/ROF model in Lagrange form (Rudin, Osher, Fatemi '90).

TV functionals for manifold-valued data

We consider the variational denoising problem given by the (discrete, anisotropic, bivariate) functionals

$$F_\alpha(u) = \sum_{i,j} \mathrm{dist}(u_i, f_i)^p + \alpha \sum_{i,j} \mathrm{dist}(u_{ij}, u_{i-1,j})^q + \alpha \sum_{i,j} \mathrm{dist}(u_{ij}, u_{i,j-1})^q,$$

with data f and $p, q \geq 1$.

- Choosing q=1 corresponds to (anisotropic) TV minimization/ROF model in Lagrange form (Rudin, Osher, Fatemi ’90).
- Choose the Riemannian distance dist to obtain the corresponding functionals for manifold-valued data.

TV functionals for manifold-valued data

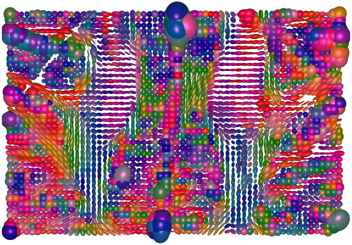
We consider the variational denoising problem given by the (discrete, anisotropic, bivariate) functionals

$$F_\alpha(u) = \sum_{i,j} \mathrm{dist}(u_i, f_i)^p + \alpha \sum_{i,j} \mathrm{dist}(u_{ij}, u_{i-1,j})^q + \alpha \sum_{i,j} \mathrm{dist}(u_{ij}, u_{i,j-1})^q,$$

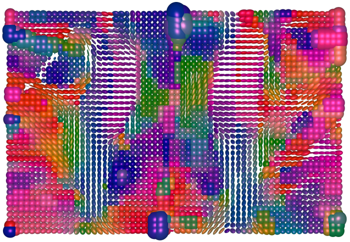
with data f and $p, q \geq 1$.

- Choosing q=1 corresponds to (anisotropic) TV minimization/ROF model in Lagrange form (Rudin, Osher, Fatemi '90).
- Choose the Riemannian distance dist to obtain the corresponding functionals for manifold-valued data.
- Increase anisotropy by additionally considering diagonals, knight moves, ...

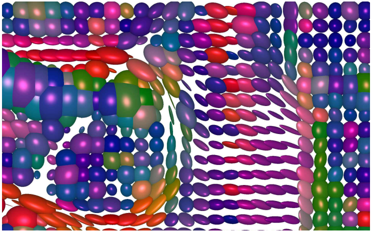
TV denoising on real DTI data (Camino project, Cook et. al. '06)

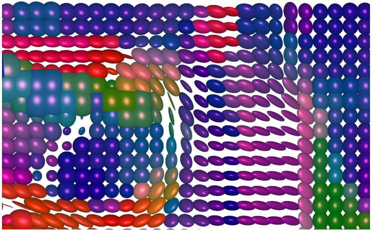


Real data



Our method for $\ell^2 - TV$
$\alpha = 0.11$.





Minimization algorithms - TV problem

- **Idea:** Write (for simplicity univariate, multivarite analogous):

$$F(u) = \gamma \sum_i \mathrm{dist}(u_i, u_{i-1})^q + \sum_j \mathrm{dist}(u_j, f_j)^p = \sum_i F_i(u) + G(u),$$

$$\text{where} \quad F_i(u) = \gamma\, \mathrm{dist}(u_i, u_{i-1})^q, \quad G = \sum_j \mathrm{dist}(u_j, f_j)^p.$$

Minimization algorithms - TV problem

- **Idea:** Write (for simplicity univariate, multivarite analogous):

$$F(u) = \gamma \sum_i \mathrm{dist}(u_i, u_{i-1})^q + \sum_j \mathrm{dist}(u_j, f_j)^p = \sum_i F_i(u) + G(u),$$

$$\text{where} \quad F_i(u) = \gamma \, \mathrm{dist}(u_i, u_{i-1})^q, \quad G = \sum_j \mathrm{dist}(u_j, f_j)^p.$$

- Apply the cyclic proximal point algorithm (Bacak, Bertsekas) : Iterate the proximal mappings (Moreau) of G and F_i, $i = l, \ldots, r$,

$$\mathrm{prox}_{\lambda F_i}(u) = \arg\min_v \tfrac{1}{2}\mathrm{dist}(u, v)^2 + \lambda F_i(v).$$

- **Central Point:** The proximal mappings of F_i, G can be computed explicitely (next slide).

Minimization algorithms - TV problem

Minimize $F(u) = \sum_i F_i(u) + G(u)$,

$$F_i(u) = \gamma\,\mathrm{dist}(u_i, u_{i-1})^q, \quad G = \sum_j \mathrm{dist}(u_j, f_j)^p.$$

- The proximal mapping of G is explicitly given by

$$\mathrm{prox}_{\lambda G}(u)_i = [u_i, f_i]_t, \quad t = \begin{cases} \frac{2\lambda}{(1+2\lambda)}\mathrm{dist}(u_i, f_i) & \text{for p= 2,} \\ \min(\lambda, \mathrm{dist}(u_i, f_i)) & \text{for p= 1.} \end{cases}$$

("Soft thresholding" for $p = 1$.)

- The proximal mapping of F_i is explicitly given by (Demaret, Storath, W.)

$$\mathrm{prox}_{\lambda F_i}(u)_j = \begin{cases} u_j & \text{if } j \neq i, i-1, \\ [u_i, u_{i-1}]_t & \text{if } j = i, \\ [u_{i-1}, u_i]_t & \text{if } j = i-1, \end{cases}$$

$t = \frac{\gamma\lambda}{(2+2\gamma\lambda)}\mathrm{dist}(u_i, u_{i-1})$ for q=2, $t = \min(\lambda\gamma, \frac{1}{2}\mathrm{dist}(u_i, u_{i-1}))$ for q=1.

Minimization algorithms - TV problem

Minimize $F(u) = \sum_i F_i(u) + \sum_i G_i(u)$,

$$F_i(u) = \gamma\,\mathrm{dist}(u_i, u_{i-1})^q, \quad G_i = \mathrm{dist}(u_i, f_i)^p.$$

- **A parallel proximal point algorithm**: Calculate the proximal mappings of F_i, G_i at $u^{(k)}$

$$u_i^{(k+1)} = \mathrm{prox}_{\lambda F_i}(u^{(k)}), \quad u_{n+i}^{(k+1)} = \mathrm{prox}_{\lambda G_i}(u^{(k)}),$$

and then average them using **intrinsic means** (Cartan, Frechet, Karcher, ...)

$$u^{(k+1)} = \arg\min_u \sum_i \mathrm{dist}(u, u_i^{(k+1)})^2.$$

Minimization algorithms - TV problem

Minimize $F(u) = \sum_i F_i(u) + \sum_i G_i(u)$,

$$F_i(u) = \gamma \,\mathrm{dist}(u_i, u_{i-1})^q, \quad G_i = \mathrm{dist}(u_i, f_i)^p.$$

- **A parallel proximal point algorithm**: Calculate the proximal mappings of F_i, G_i at $u^{(k)}$

$$u_i^{(k+1)} = \mathrm{prox}_{\lambda F_i}(u^{(k)}), \quad u_{n+i}^{(k+1)} = \mathrm{prox}_{\lambda G_i}(u^{(k)}),$$

  and then average them using **intrinsic means** (Cartan, Frechet, Karcher, ...)

$$v = u^{(k+1)} = \arg\min_u \sum\nolimits_i \mathrm{dist}(u, u_i^{(k+1)})^2.$$

- To compute the minimizer, we use the gradient descent (Karcher)

$$v_{\mathrm{new}} = \exp_{v_{\mathrm{old}}}\left(\tfrac{1}{N} \sum\nolimits_{i=1}^{N} \log_{v_{\mathrm{old}}} u_i^{(k+1)}\right).$$

Minimization algorithms - TV problem

Minimize $F(u) = \sum_i F_i(u) + \sum_i G_i(u)$,

$$F_i(u) = \gamma \,\mathrm{dist}(u_i, u_{i-1})^q, \quad G_i = \mathrm{dist}(u_i, f_i)^p.$$

- **A parallel proximal point algorithm**: Calculate the proximal mappings of F_i, G_i at $u^{(k)}$

$$u_i^{(k+1)} = \mathrm{prox}_{\lambda F_i}(u^{(k)}), \quad u_{n+i}^{(k+1)} = \mathrm{prox}_{\lambda G_i}(u^{(k)}),$$

  and then average them using **intrinsic means** (Cartan, Frechet, Karcher, ...)

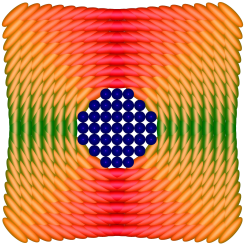
$$v = u^{(k+1)} = \arg\min_u \sum_i \mathrm{dist}(u, u_i^{(k+1)})^2.$$

- To compute the minimizer, we use the gradient descent (Karcher)

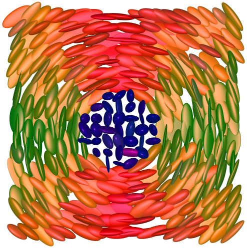
$$v_{\mathrm{new}} = \exp_{v_{\mathrm{old}}}\left(\frac{1}{N} \sum_{i=1}^N \log_{v_{\mathrm{old}}} u_i^{(k+1)}\right).$$

- **Fast Variant:** Approximate the mean by iterated geodesic averages.

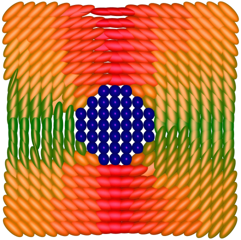
Synthetic DTI example



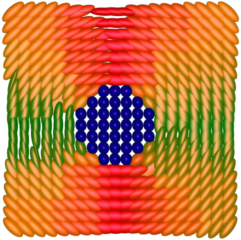
Synthetic DT image



Rician noise, $\sigma = 90$.



ℓ^2-TV (our cyclic PPA)



ℓ^2-TV (our parallel PPA)

Analytic Results

Theorem (Demaret, Storath, W.)

In a Cartan-Hadamard manifold (complete, simply connected, nonpositive sectional curvature) the proposed algorithms (cyclic, parallel and the parallel variant with approximative mean computation) for L^p-TV minimization converge towards a global minimizer.

Analytic Results

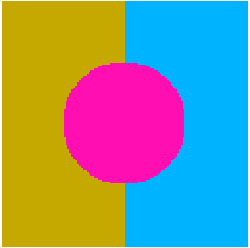
Theorem (Demaret, Storath, W.)

In a Cartan-Hadamard manifold (complete, simply connected, nonpositive sectional curvature) the proposed algorithms (cyclic, parallel and the parallel variant with approximative mean computation) for L^p-TV minimization converge towards a global minimizer.

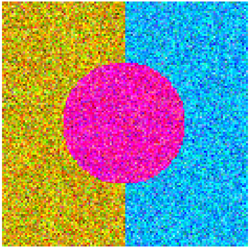
Sceleton of proof:

- Proof that in a connected, complete Riemannian manifold, the proximal mappings of the first differences and the distances are given by the formulas derived above.
- For the cyclic PPA apply the convergence result of Bacak (Bacak '14).
- For the parallel PPAs base on techniques used in (Bacak '14) and find suitable modifications.

Denoising on the LCh color model ($\mathbb{S}^1 \times \mathbb{R}^2$).

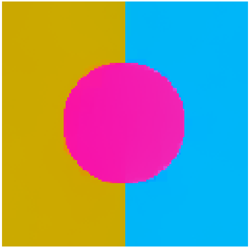


Synthetic image



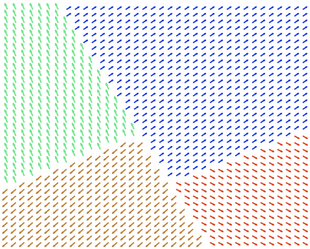
Gaussian noise (PSNR: 15.64).

ℓ^2-TV on RGB (PSNR:23.92)

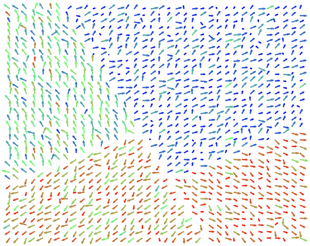


ℓ^2-TV on LCh (PSNR:32.19)

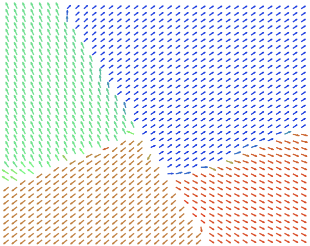
Denoising $\mathbb{S}^2$ data.



Original

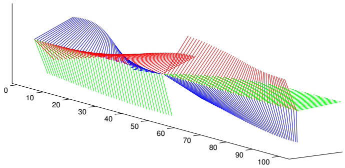


Von Mises-Fisher noise ($\kappa = 12.7$)

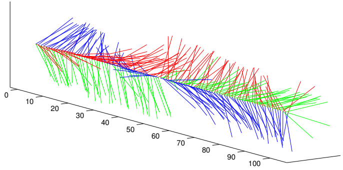


ℓ^1-TV regularization

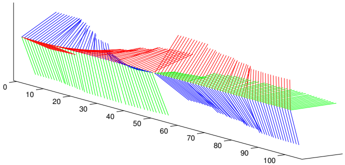
Denoising SO_3 data.



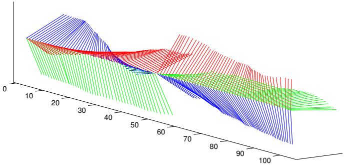
Synthetic signal



Fisher noise ($\kappa = 75.$)

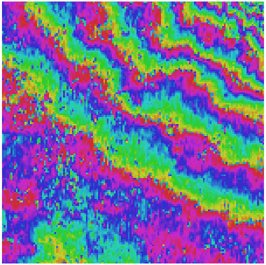


ℓ^2-TV

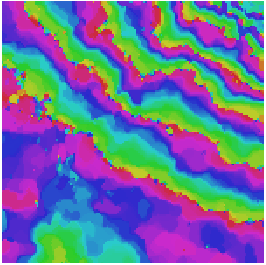


ℓ^2-Huber

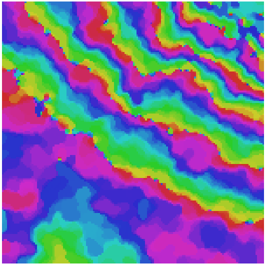
Denoising inSAR data



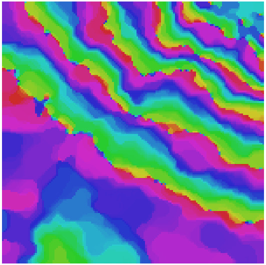
Real data



L^2-TV denoising



L^1-TV denoising



TV with Huber data term

Part II:

Second order TV type functionals for $\mathbb{S}^1$-valued data

(joint work with R. Bergmann, F. Laus, G. Steidl)

Second order TV for S^1 valued data

- Second order TV type functional for real-valued data:

$$F(u) = \|u - f\|_2^2 + \alpha\|\nabla_1 u\|_1 + \beta\|\nabla_2 u\|_1.$$

  Here,

$$\nabla_2 u(i) = u(i-1) - 2u(i) + u(i+1).$$

- **Question:** What are second differences for S^1 valued data?

Second order TV for S^1 valued data

- Second order TV type functional for real-valued data:

$$F(u) = \|u - f\|_2^2 + \alpha\|\nabla_1 u\|_1 + \beta\|\nabla_2 u\|_1.$$

  Here,

$$\nabla_2 u(i) = u(i-1) - 2u(i) + u(i+1).$$

- **Question:** What are second differences for S^1 valued data?
- **Idea:** Translate

$$\nabla_2 u(i) = (u(i-1) - u(i)) + (u(i-1) - u(i))$$

  to the manifold setting:

$$\nabla_2 u(i) = \exp^{-1}_{u(i)} u(i-1) + \exp^{-1}_{u(i)} u(i+1).$$

Second order TV for S^1 valued data

- Second order TV type functional for real-valued data:

$$F(u) = \|u - f\|_2^2 + \alpha\|\nabla_1 u\|_1 + \beta\|\nabla_2 u\|_1.$$

Here,

$$\nabla_2 u(i) = u(i-1) - 2u(i) + u(i+1).$$

- **Question:** What are second differences for S^1 valued data?
- **Idea:** Translate

$$\nabla_2 u(i) = (u(i-1) - u(i)) + (u(i-1) - u(i))$$

to the manifold setting:

$$\nabla_2 u(i) = \exp_{u(i)}^{-1} u(i-1) + \exp_{u(i)}^{-1} u(i+1).$$

- **Problem:** These second differences are not continuous in u_{i-1}, u_i, u_{i+1}.

Second order TV for S^1 valued data

- **Alternative**: View $u_i \in]-\pi, \pi]$ as real-valued data and define the absolute cyclic difference

$$d_2(f_{i-1}, f_i, f_{i+1}) = \min_{k,l,m=-1,0,1} |\nabla_2(f_{i-1} + k2\pi, f_i + l2\pi, f_{i+1} + m2\pi)|$$

  These differences are continuous in f_{i-1}, f_i, f_{i+1}.
- Equivalent: Consider all liftings and take the minimal difference on the lifted $\mathbb{R}$-valued data.
- For nearby f_{i-1}, f_i, f_{i+1} the manifold and the lifting definition agree.

Second order TV for S^1 valued data

- **Alternative**: View $u_i \in]-\pi, \pi]$ as real-valued data and define the absolute cyclic difference

$$d_2(f_{i-1}, f_i, f_{i+1}) = \min_{k,l,m=-1,0,1} |\nabla_2(f_{i-1} + k2\pi, f_i + l2\pi, f_{i+1} + m2\pi)|$$

  These differences are continuous in f_{i-1}, f_i, f_{i+1}.
- Equivalent: Consider all liftings and take the minimal difference on the lifted $\mathbb{R}$-valued data.
- For nearby f_{i-1}, f_i, f_{i+1} the manifold and the lifting definition agree.
- The proximal mappings for d_2 can be computed explicitely (Bergmann, Laus, Steidl, W. '14): for $w = (1, -2, 1)^T$, and $|\langle f, w\rangle| < \pi$,

$$\mathrm{prox}_{\lambda d_2}(f) = (f - swm)_{2\pi}, \quad m = \min\left(\lambda, \frac{\langle f,w\rangle}{\|w\|_2^2}\right), \quad s = \mathrm{sign}\langle f, w\rangle.$$

- All ingredients for the cyclic proximal point algorithm are available.

Convergence of the cyclic proximal point algorithms

Theorem (Bergmann, Laus, Steidl, W., 2014)

For data f with nearby data items and small enough parameters α, β, the cyclic proximal point algorithm for second order TV type minimization converges to a minimizer.

- What nearby means and α, β can be quantified.

Convergence of the cyclic proximal point algorithms

Theorem (Bergmann, Laus, Steidl, W., 2014)

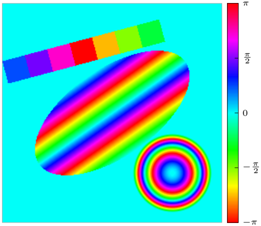
For data f with nearby data items and small enough parameters α, β, the cyclic proximal point algorithm for second order TV type minimization converges to a minimizer.

- What nearby means and α, β can be quantified.

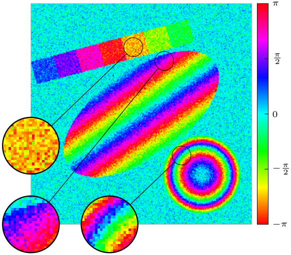
Idea of proof:

- Lift the setting to the covering space $\mathbb{R}$.
- For $\mathbb{R}$-valued data we have convergence and the distance of the iterates can be estimated basing on (Bacak, Bertsekas).
- Lifting commutes with the proximal mappings and all other relevant operations for the considered data.
- Conclude nearness for $\mathbb{S}^1$ data and derive convergence.

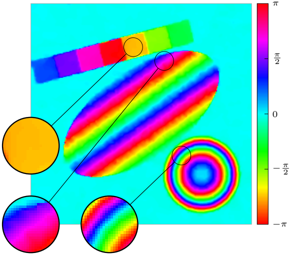
Second order TV minimization - synthetic example.



Original



Noisy



Second order TV.

Denoising the H channel in HSV space.

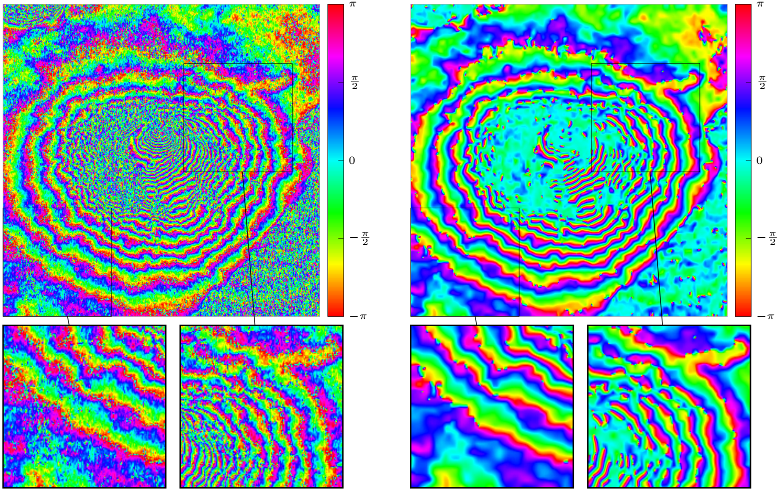
Image

Noisy hue

hue denoising on $\mathbb{R}$

hue denoising on $\mathbb{S}^1$

Second order TV for real SAR data of Mt. Vesuvius.



Original

Second order TV denoising.

Part III:

Potts and Blake-Zisserman functionals for manifold-valued signals with a few jumps

(joint work with L. Demaret and M.Storath)

Potts and Blake-Zisserman functionals for manifold-valued data

Define (univariate) Potts functionals P_γ for manifold-valued data,

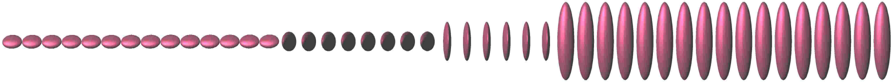
$$P_\gamma(u) = \gamma \, \#\{i : u_i \neq u_{i-1}\} + \sum_i \mathrm{dist}(u_i, f_i)^p,$$

and Blake-Zisserman functionals B_γ for manifold-valued data,

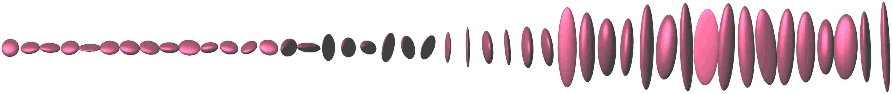
$$B_\gamma(u) = \gamma \, \sum_i \min(s^q, \mathrm{dist}(u_i, u_{i-1})^q) + \sum_i \mathrm{dist}(u_i, f_i)^p.$$

Example: L^2-Potts minimization for DTI data.

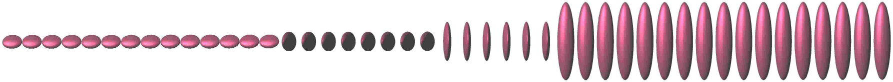
Original (synthetic) signal:



Noisy data (Rician noise with $\sigma = 60$):

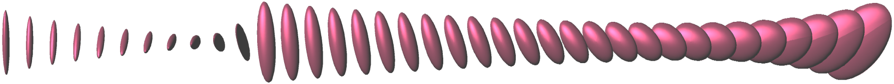


L^2-Potts reconstruction:

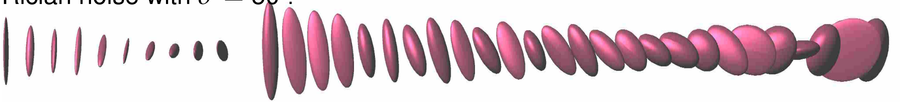


Example: Blake-Zisserman vs. Potts.

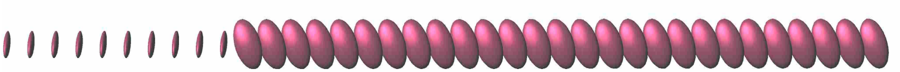
Original (synthetic) signal:



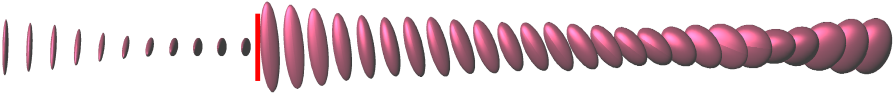
Rician noise with $\sigma = 50$:



L^2-Potts reconstruction:



L^2-Blake-Zisserman reconstruction:



Minimization algorithm - Potts problem

$$P_\gamma(u) = \gamma \, \#\{i : u_i \neq u_{i-1}\} + \sum_{i=1}^{n} \mathrm{dist}(u_i, f_i)^p \to \min,$$

- Algorithm based on dynamic programming.
- Most time consuming: For each subinterval $[l, r]$ calculate

$$u = \arg\min \sum_{i=l}^{r} \mathrm{dist}(u, f_i)^p,$$

(p=2: Riemannian center of mass; p=1: Riemannian median.)

We use (sub-)gradient descent; e.g., for $p = 1$,

$$u^{k+1} = \exp_{u^k}\left(\tau_k \sum_{i=l}^{r} \frac{\log_{u^k} f_i}{\|\log_{u^k} f_i\|} \right).$$

Converges for $p = 1$ when $\tau \in \ell^2 \setminus \ell^1$ (Arnaudon et al. '11).

Minimization algorithm - Blake-Zisserman problem

$$B_\gamma(u) = \gamma \sum_i \min(s^q, \operatorname{dist}(u_i, u_{i-1})^q) + \sum_j \operatorname{dist}(u_j, f_j)^p \to \min .$$

- Algorithm based on dynamic programming.
- For each subinterval $[l, r]$ calculate the minimizer of

$$F(u) = \gamma \sum_i \operatorname{dist}(u_i, u_{i-1})^q + \sum_j \operatorname{dist}(u_j, f_j)^p$$

- This is a *TV* minimization problem (or, more general, ℓ^q variation minimization) for manifold data which can be solved by the developed methods.

Minimization algorithms

Theorem (Demaret, Storath, W.)

Let $p, q \geq 1$. In a Cartan-Hadamard manifold, our algorithm for the minimization of the (univariate) Potts functionals P_γ produces a minimizer.

Theorem (Demaret, Storath, W.)

Let $p, q \geq 1$. In a Hadamard space, our algorithm for the minimization of the (inivariate) Blake-Zisserman functionals B_γ produces a minimizer.

Minimization algorithms

Theorem (Demaret, Storath, W.)

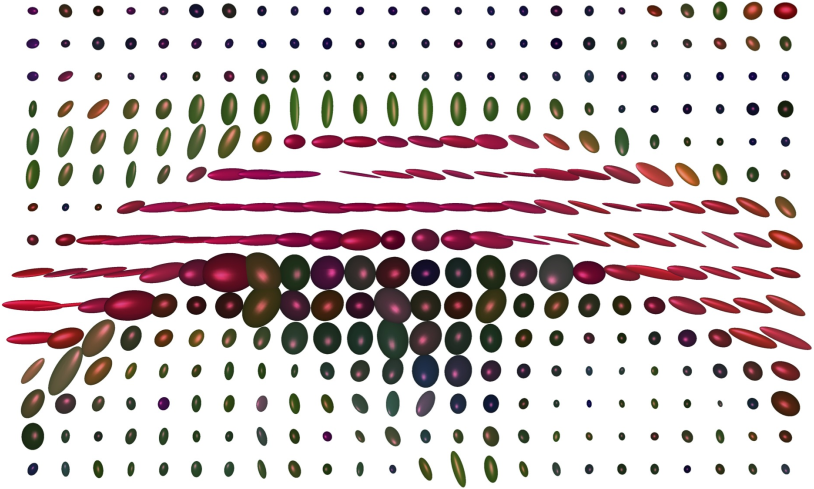
Let $p, q \geq 1$. In a Cartan-Hadamard manifold, our algorithm for the minimization of the (univariate) Potts functionals P_γ produces a minimizer.

Theorem (Demaret, Storath, W.)

Let $p, q \geq 1$. In a Hadamard space, our algorithm for the minimization of the (inivariate) Blake-Zisserman functionals B_γ produces a minimizer.

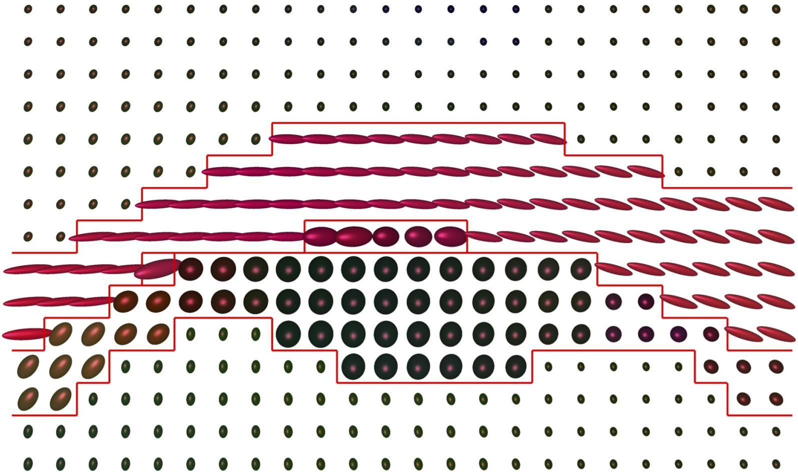
- For multivariate data, the Potts and the Blake-Zisserman problem are NP hard.
- In this case, we use a splitting approach (cf. W., Demaret, Storath '14).

Segmentation: real data from the Camino project (Cook et al. ’06)



Segmentation: real data from the Camino project (Cook et al. '06)

Edge between neighbouring $P, Q \iff \operatorname{dist}(P, Q) \geq s$ (s B.-Z. parameter).



Blake-Zisserman regularization ($p, q = 1$) with $s = 0.67$, $\gamma = 4.3$.

Summary

- We have derived algorithms for TV minimization for manifolds.
- We have shown convergence to a minimizer for Hadamard manifolds.
- We have seen the potential in various applications.
- We have derived an algorithm for second order TV type functionals for S^1 data.
- We have obtained convergence for nearby neighboring data and shown applications.
- We have obtained algorithms for Potts and Blake-Zisserman problems for manifold valued data.
- We have seen a segmentation of a real corpus callosum.

Some References

M. Bačák.
Computing medians and means in Hadamard spaces.
SIAM J. Optim. (to appear), 2014.

J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers.
Total variation regularization for functions with values in a manifold.
In *IEEE ICCV 2013*, pages 2944–2951, 2013.

A. Weinmann, L. Demaret, and M. Storath.
Total variation regularization for manifold-valued data.
SIAM J. Imaging Sci., 7(4):2226–2257, 2014.

R. Bergmann, F. Laus, G. Steidl, and A. Weinmann.
Second order differences of cyclic data and applications in variational denoising.
SIAM J. Imaging Sci. (to appear), 2014.

A. Weinmann, L. Demaret, and M. Storath.
Mumford-Shah and Potts regularization for manifold-valued data with applications to DTI and Q-ball imaging.
Preprint, 2014.