

Variational Methods: Setup

Forward operator between Banach spaces

$$K : \mathcal{U}(\Omega) \rightarrow \mathcal{V}(\Sigma)$$

with finite dimensional approximation (sampling, averaging)

$$K^m : \mathcal{U}(\Omega) \rightarrow \mathcal{V}^m$$

Inverse Problems and Ill-posedness

Consider

$$Ku = f$$

with **compact operator** K acting between Banach spaces

This problem is **ill-posed**:

- Potential non-existence and non-uniqueness
- **Instability**, discontinuous dependence on data f

Regularization

The problem needs to be **approximated by well-posed ones**

Topologists answer: restrict domain to a **compact** set

[Tikhonov 1943]

Hilbert space theory: approximate least-squares

$$\|Ku - f\|^2 \rightarrow \min_u$$

by

$$\|Ku - f\|^2 + \alpha\|u\|^2 \rightarrow \min_u$$

[Tikhonov 43/63, Phillips 62, Ridge 70] [Tikhonov, Glasko 64, Morozov 66]



A. Tikhonov

Maximum Likelihood / Bayes

Reconstruct **maximum-likelihood estimate**

$$\hat{u} = \arg \max_u p(u|f)$$

Model of **posterior probability (Bayes)**

$$p(u|f) = \frac{p(f|u)p(u)}{p(f)}$$

Yields regularized variational problem for finite m

Use prior model: Bayesian estimation

Likelihood

$$p(f|u) \sim e^{-D(Ku, f)}$$

Prior related to some regularization functional R)

$$p_0(u) \sim F(R(u))$$

Negative log posterior

$$-\log p(u|f) = D(Ku, f) - \log F(R(u)) + \text{const}$$

MAP estimate

Minimize negative log posterior with respect to u

$$-\log p(u|f) = D(Ku, f) - \log F(R(u)) + \text{const}$$

Simple Lagrange multiplier argument shows equivalence to a minimization problem of the form

$$J(u) = D(Ku, f) + \alpha R(u)$$

(Note rigorous derivation in Banach spaces based on distributions, logarithmic derivative of measures

Stuart-Dashti 14, Helin-mb 16, Agapios-mb-Dashti-Helin 17, Sullivan 17

Example: data terms

Additive Gaussian noise leads to weighted least squares (Hilbert space norm determined by covariance matrix)

$$D(Ku, f) = \frac{1}{2} \|Ku - f\|_{\mathcal{H}}^2$$

Poisson noise (photon counting) leads to Kullback-Leibler

$$D(Ku, f) = \int (Ku - f - f \log \frac{Ku}{f}) d\mu$$

Example Gauss:

Additive noise, i.i.d. on each pixel, mean zero, variance σ

$$p(f|u) \sim e^{-\frac{\|K^m u - f\|_2^2}{2\sigma^2}}$$

Minimization of negative posterior log-likelihood yields

$$\hat{u} = \arg \min_{u \in \mathcal{W}(\Omega)} \left\{ \frac{1}{2} \|K^m u - f\|_2^2 + \alpha J(u) \right\}$$

Asymptotic variational model

$$\hat{u} = \arg \min_{u \in \mathcal{W}(\Omega)} \left\{ \frac{1}{2} \|Ku - f\|_{\mathcal{V}(\Sigma)}^2 + \alpha J(u) \right\}$$

Non-Gaussian Models

Analogous treatment of Non-Gaussian Noise models

Laplace:

$$p(f|u) = (2\sigma)^{-m} \exp \left(- \sum_{i=1}^m \frac{|f_i - (K^m u)_i|}{\sigma} \right)$$

Asymptotic of MAP model

$$\hat{u} = \arg \min_{u \in \mathcal{W}(\Omega)} \left\{ \int_{\Sigma} |(Ku)(y) - f(y)| \, d\mu(y) + \alpha J(u) \right\}$$

Non-Gaussian Models

Poisson (*Shepp-Vardi 83*):

$$\hat{u} = \arg \min_{u \in \mathcal{W}(\Sigma)} \left\{ \int_{\Sigma} \left[f(y) \ln \left(\frac{f(y)}{(Ku)(y)} \right) - f(y) + (Ku)(y) \right] d\mu(y) + \alpha J(u) \right\}$$

Multiplicative (various models, here *Aubert-Aujol 08*):

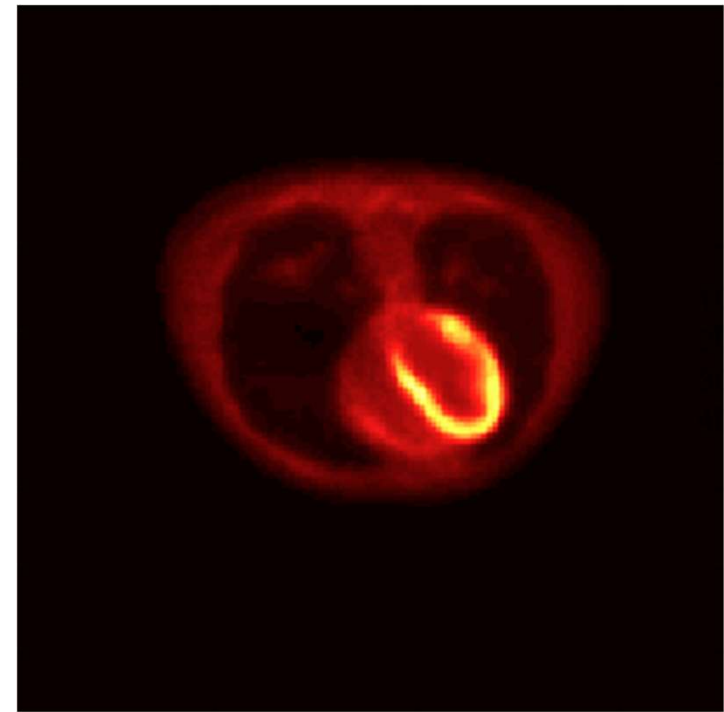
$$\hat{u} \in \arg \min_{u \in \mathcal{W}(\Omega)} \left\{ \int_{\Sigma} \left[\ln \left(\frac{(Ku)(y)}{f(y)} \right) + \frac{f(y)}{(Ku)(y)} - 1 \right] d\mu(y) + \alpha J(u) \right\}$$

Standard Methods

Operator / Matrix K full and not much structured
Standard solution by simple iterative methods
„early stopping“ (iterative regularization)s

Poisson Statistics: *EM-Algorithm with appropriate termination (Shepp-Vardi 83):*

$$u_{k+1} = \frac{u_k}{K^*1} K^* \left(\frac{f}{K u_k} \right)$$



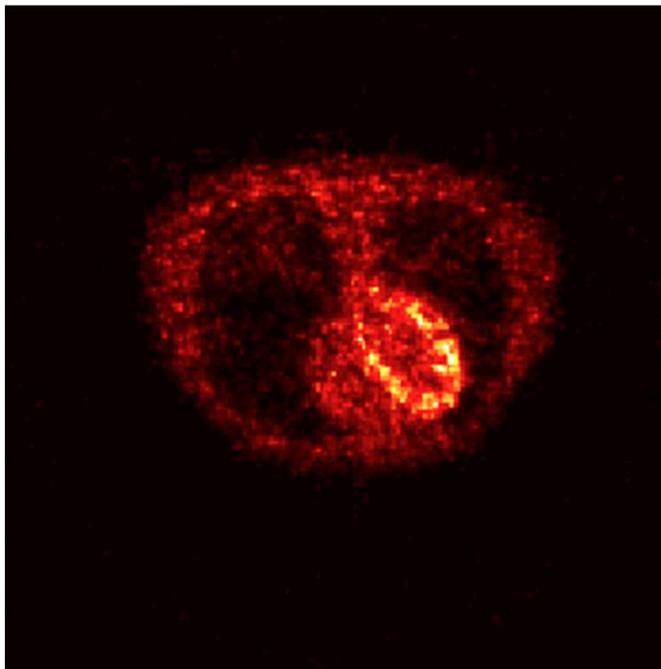
20 min ^{18}F -FDG Scan

Small Time Intervals: Low Photon Counts

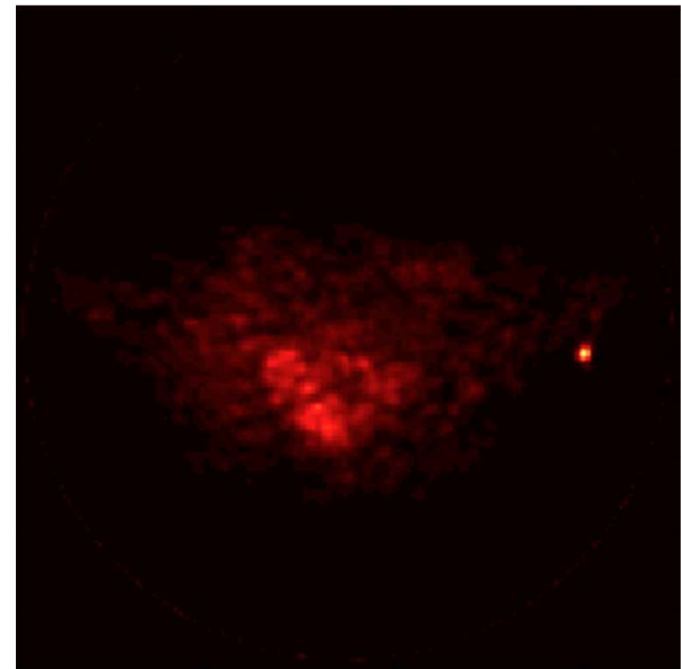
Signal-dependent undersampling /noise

Low quality reconstruction with standard methods

^{15}F -FDG
5s interval



H_2^{15}O
 $\sim 60\text{s}$



Variational Model: Regularization

Likelihood for Poisson data to be used instead of simple least-square data fittings

Additional regularization (=Bayesian MAP estimation)

$$\hat{u} = \arg \min_{u \in \mathcal{W}(\Sigma)} \left\{ \int_{\Sigma} \left[f(y) \ln \left(\frac{f(y)}{(Ku)(y)} \right) - f(y) + (Ku)(y) \right] d\mu(y) + \alpha J(u) \right\}$$

Example: regularization terms

Use total variation regularization

$$R(u) = \int |Du|$$

(Pre)Dual definition

$$R(u) = \sup_{\varphi \in C_0^\infty(\Omega)^d, \|\varphi\|_\infty \leq 1} \int_\Omega u \nabla \cdot \varphi \, dx$$

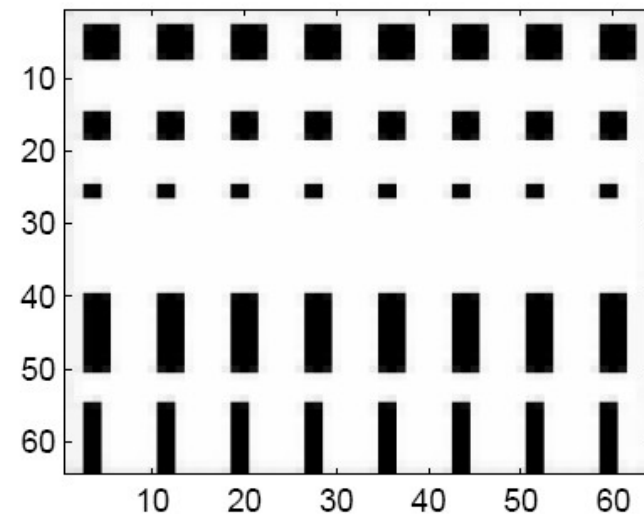
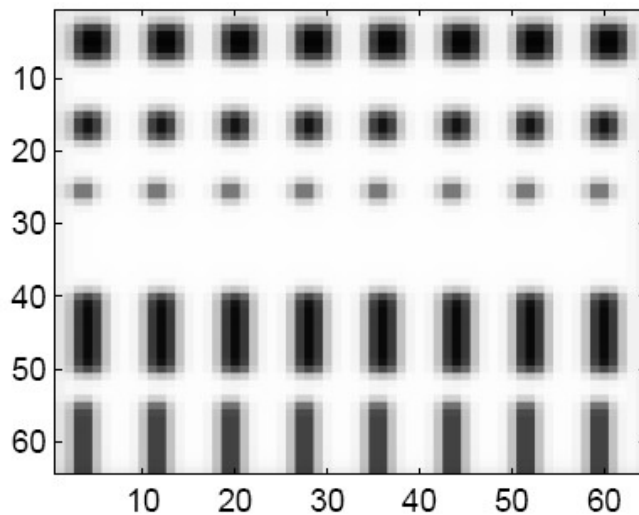
Rudin-Osher-Fatemi 1989/1992

50 years after Tikhonov

Beyond Hilbert space

$$\frac{1}{2} \|Ku - f\|^2 + \alpha J(u) \rightarrow \min_u$$

Detailed structure of reconstructed solutions for positive regularization parameter matters



TV-Method: Structural Prior

Formal

$$J(u) := \int |\nabla u| \, dx$$

Exact

$$J(u) := \sup_{\mathbf{g} \in C_0^\infty} \int u \operatorname{div} \mathbf{g} \, dx$$

ROF-Model for denosing g : J minimized subject to

$$\int (u - g)^2 \, dx \leq \sigma^2$$

Rudin-Osher-Fatemi 89,92

TV-Methods and Geometry

Relation to minimal surface problems via coarea-formula

$$J(u) = \int \int_{\partial\{u < \alpha\}} ds \, d\alpha$$

Fitting terms decomposes into volume integrals with weight $u-g$

Solution of isoperimetric problems on sublevel sets !

$$\frac{\lambda}{2} \int (u - g)^2 \, dx + J(u) \rightarrow \min_{u \in BV}$$



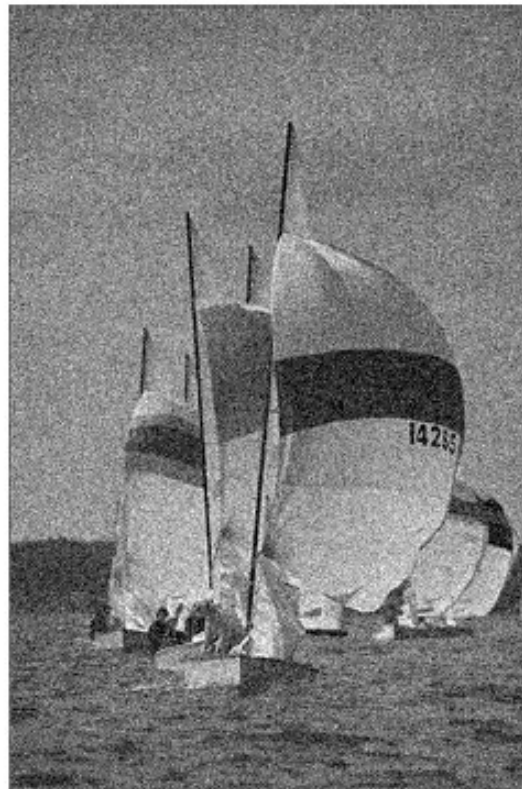
WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

ROF Model

clean



noisy



ROF



Why TV-Methods ?

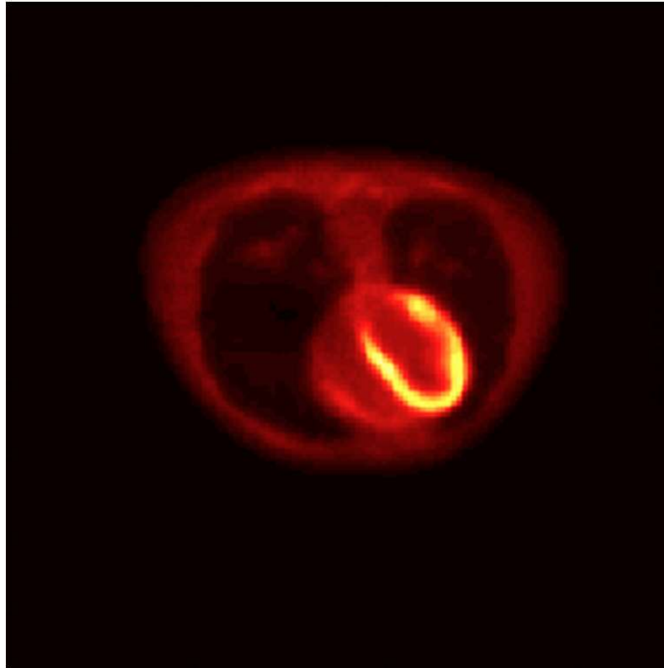
Cartooning



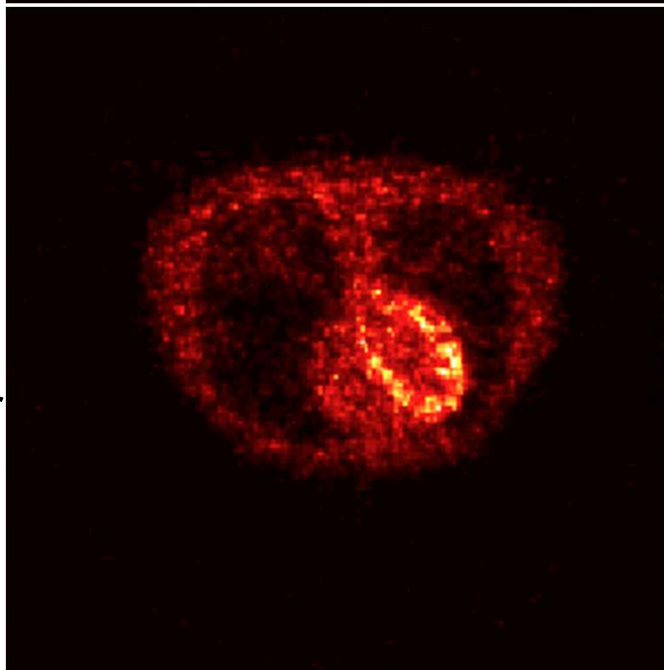
ROF Model with increasing allowed variance

^{18}F -FDG

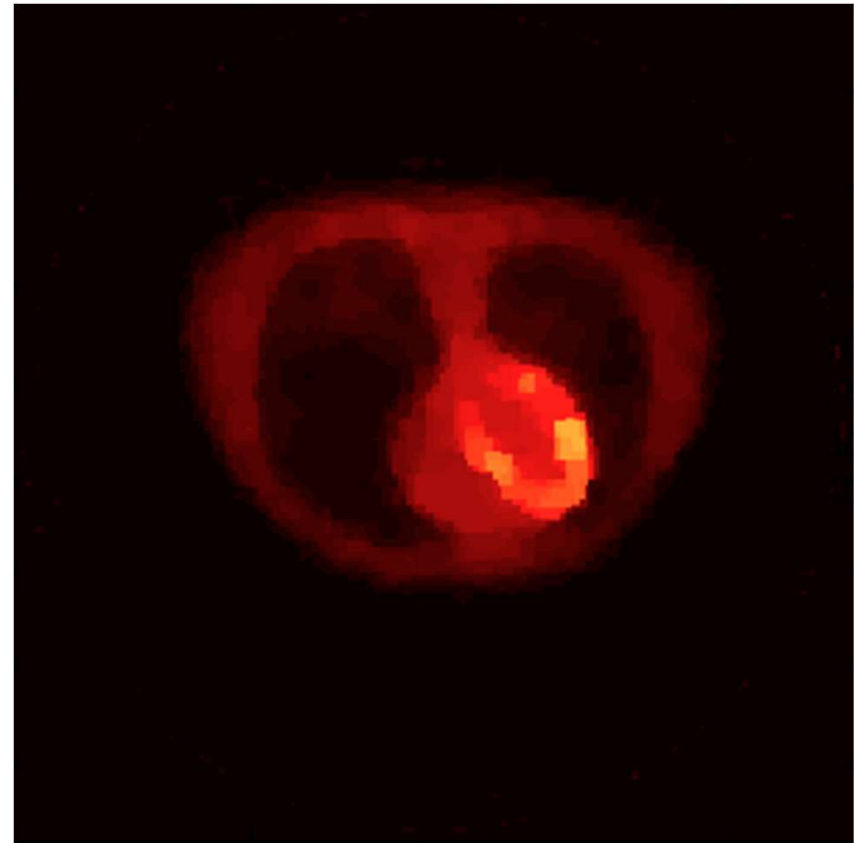
EM, 20 min



EM, 5s



EM-TV, 5s



*Jahn Müller, 2011
Data from Nuclear
Medicine
Department, UKM*

Beyond Staircasing: Higher Order TV

Improved version of infimal convolution between first and second order total variation (stronger dual constraint)

$$\text{GTV}_\beta(u) := \sup_{\substack{q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^d)) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1}} \int_{\Omega} u \, \text{div}^2(q) \, dx$$

Primal version

$$\text{GTV}_\beta(u) = \inf_{\substack{u=v+w \\ s \in \mathcal{N}(\text{div})}} \left\{ \int_{\Omega} |\nabla v - s| + \beta \int_{\Omega} |\nabla^2 w + \nabla s| \right\}$$

Bredies et al 2011

Sparsity

What is a good model

General rule: simple explanations are more likely !

Sparsity: Simplest Explanations = multiples of basis vectors

Inverse solutions u related to expansions in basis

Combinations of few basis functions likely: Sparsity

$$J(u) = ||u||_0 \text{ minimized subject to } Ku = f$$

$$\text{More general } J(u) = ||Tu||_0$$

L1 and L1-type

0-norm far from convex, NP-hard in nD

Not even defined in infinite dimension, no limit

Relaxation from 0-norm to 1-norm (Rudin-Osher-Fatemi, Donoho, Candes-Tao,) :

$$J(u) = ||u||_1 \text{ minimized subject to } Ku = f$$

$$\text{More general } J(u) = ||Tu||_1$$

Biomedical Imaging: 2000 vs 2010

<i>Modality</i>	<i>State of the art 2000</i>	<i>State of the art 2010</i>
Full CT	Filtered Backprojection	Exact Reconstruction
PET/SPECT	Filtered Backprojection/EM	EM-TV / Dynamic Sparse
PET-CT	-	EM-AnatomicalTV
Photoacoustic	-	Wavelet Sparse / TV
EEG/MEG	Least Norm, LORETA	Sparsity / Bayesian
ECG-BSPM	Least Norm	L1 of normal derivative
Microscopy	None, linear Filter	Poisson-TV / Shearlet-L1

Why is sparse / L1-type better ?

Start with standard quadratic functionals

$$J(u) = \frac{1}{2} \|Lu\|^2$$

Advantages:

Differentiability, strict convexity

Standard tools from Analysis (Hilbert spaces) and standard computational methods (linear systems, Newton methods, ...)

Why is sparse / L1-type better ?

Standard quadratic functionals

$$J(u) = \frac{1}{2} \|Lu\|^2$$

Disadvantage:

Oversmoothing of the solution

$$\alpha L^* Lu = K^*(f - Ku)$$

Hence solution is in the range of $(L^* L)^{-1} K^*$

Why is sparse / L1-type better ?

Typical inverse problem: K is integral operator

$$(Ku)(x) = \int k(x, y)u(y) \, dy$$

Even for L being the identity in L^2 , we have

$$u(x) = \frac{1}{\alpha} \int k(y, x) \left(f(y) - \int k(y, z)u(z) \right) \, dz$$

Smoothness of u determined by the integral kernel !

Why is sparse / L1-type better ?

Image denoising: K is the identity in L^2 , L equals gradient

$$\int (u - f)^2 + \alpha |\nabla u|^2 \, dx \rightarrow \min_u$$

Optimality yields Euler equation

Regularity yields

$$-\alpha \Delta u + u = f, \text{ no edges}$$

$$u \in H^2$$

Why is sparse / L1-type better ?

Nonlinear denoising

$$\int (u - f)^2 + \alpha G(\nabla u) \, dx \rightarrow \min_u$$

Smooth convex function G yields

$$-\alpha \nabla \cdot (G'(\nabla u) \nabla u) + u = f$$

Still full elliptic regularity !

Nonsmooth G needed to keep edges !

Why is sparse / L1-type better ?

Example: sparsity with respect to basis / frame (e.g. wavelet)

Let K be an **operator defined on the coefficients**, i.e. from $\ell^2(\mathbb{Z})$ to some Hilbert space

Tikhonov solution satisfies

$$u_k = \frac{1}{\alpha} (K^* (f - Ku))_k$$

Decay properties of coefficients are strongly determined by the adjoint operator, in general **not sparse**

Why is sparse / L1-type better ?

Example: sparsity with respect to basis / frame (e.g. wavelet)

Nonlinear regularization $J(u) = \sum_k r(u_k)$

yields

$$r'(u_k) = \frac{1}{\alpha} (K^*(f - Ku))_k$$

If r is strictly convex with minimizer zero, **as few nonzero entries as in the quadratic case !**

Why is sparse / L1-type better ?

Example: sparsity with respect to basis / frame (e.g. wavelet)

Simple not strictly convex choice

$$J(u) = \sum_k |u_k|$$

yields

$$\text{sign}(u_k) \ni \frac{1}{\alpha} (K^*(f - Ku))_k$$

All small entri

Numerical Differentiation with TV

From Master Thesis of
Markus Bachmayr, 2007

$$u_{H^1}^\alpha = \arg \min_u \left\{ H(u, f^\delta) + \alpha \int_0^1 |u'|^2 dt \right\},$$

$$u_{L^2}^\alpha = \arg \min_u \left\{ H(u, f^\delta) + \alpha \int_0^1 |u|^2 dt \right\},$$

$$u_{BV}^\alpha = \arg \min_u \left\{ H(u, f^\delta) + \alpha |u|_{BV[0,1]} \right\}$$

where the fitting functional H is given by

$$H(u, f^\delta) = \int_0^1 |Tu - f^\delta|^2 dt.$$

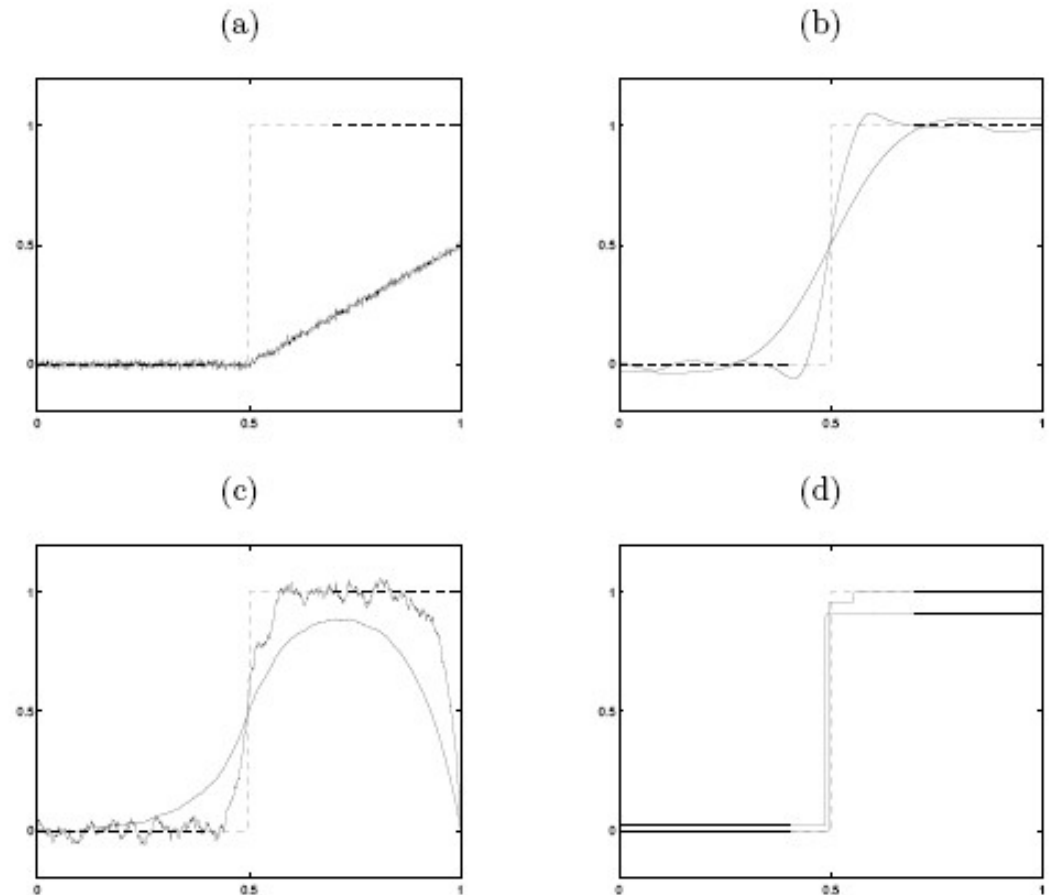


Figure 1.1: Comparison of results for numerical differentiation: (a) exact solution u and given primitive data f^δ with 5% noise, (b) $u_{H^1}^\alpha$ for $\alpha = 10^{-4}$, 10^{-6} , (c) $u_{L^2}^\alpha$ for $\alpha = 10^{-2}$, 10^{-3} , (d) u_{BV}^α for $\alpha = 10^{-3}$, 10^{-5} .

Challenges for the Theory

Need to setup analysis in Banach spaces, usually non-reflexive ones like BV

Existence of solutions for positive regularization parameter by standard techniques of calculus of variation

Quantifying errors ? No norm estimates

Understanding fine structures

Basic Properties and Structure

Can be understood from optimality condition

$$K^*(Ku - f) + \alpha p = 0, \quad p \in \partial J(u)$$

and structure of subdifferential.

Solutions reselectively their subgradients satisfy source condition

$$p = K^*w$$

Basic Properties and Structure

E.g. in the l_1 -case

$$p_i \in \text{sign}(u_i) \cap (K^*w)_i$$

For smoothing operator with adjoint mapping into some l_p -space, solutions immediately have finite number of nonzero entries.

Example: EEG/ MEG model, depth bias

Difficulty in EEG/MEG is huge nullspace and structure of the forward operator

For understanding use scalar forward model

$$-\Delta v = u \quad \text{in } \Omega \subset \mathbb{R}^3$$

$$\frac{\partial v}{\partial n} = 0 \text{ on } \partial\Omega \quad \int_{\partial\Omega} v \, d\sigma = 0$$

Simplified EEG model

Forward operator

$$K : L^p_{\diamond}(\Omega) \rightarrow L^2_{\diamond}(\partial\Omega)$$
$$u \mapsto v$$

Optimality : subgradients always in the range of the adjoint operator K^* !!!

Simplified EEG model

Source condition is effectively existence of Lagrange multiplier for

$$R(u) \rightarrow \min_u \quad \text{subject to } Ku = f.$$

Formal computation based on Lagrangian

$$\mathcal{L}(u, p, q) = R(u) + \int_{\partial\Omega} (v - f)p \, d\sigma + \int_{\Omega} (\nabla v \cdot \nabla q - fu) \, dx$$

Simplified EEG model

Saddle-point of the Lagrangian satisfies (for some p)

$$\begin{aligned} q &\in \partial R(u) \\ -\Delta q &= 0 \quad \text{in } \Omega \\ p + \frac{\partial q}{\partial n} &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

Simplified EEG model

Saddle-point of the Lagrangian satisfies (for some p)

$$\begin{aligned} q &\in \partial R(u) \\ -\Delta q &= 0 \quad \text{in } \Omega \\ p + \frac{\partial q}{\partial n} &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

Does that explain the depth bias ?

Minimum Norm Solutions

Regularization

$$R(u) = \frac{1}{2} \int_{\Omega} u^2 \, dx$$
$$\partial R(u) = \{u\}$$

Hence, solution u is harmonic !

In particular, u attains its maximum at the boundary
(maximum principle) - depth bias is always present !

Weighted Minimum Norm Solutions

Regularization

$$R(u) = \frac{1}{2} \int_{\Omega} \frac{u^2}{w} dx$$

Hence, solution u is $\partial R(u) = \left\{ \frac{u}{w} \right\}$ function and weight
 Depth bias can be reduced (in particular if w is highly localized)
 Good choice of weight is crucial

Sparse Solutions

Regularization

$$R(u) = \int_{\Omega} |u| \, dx$$

$$\partial R(u) = \{s\} \quad s(x) \begin{cases} = 1 & \text{if } u(x) > 0 \\ = -1 & \text{if } u(x) < 0 \\ \in [0, 1] & \text{if } u(x) = 0 \end{cases}$$

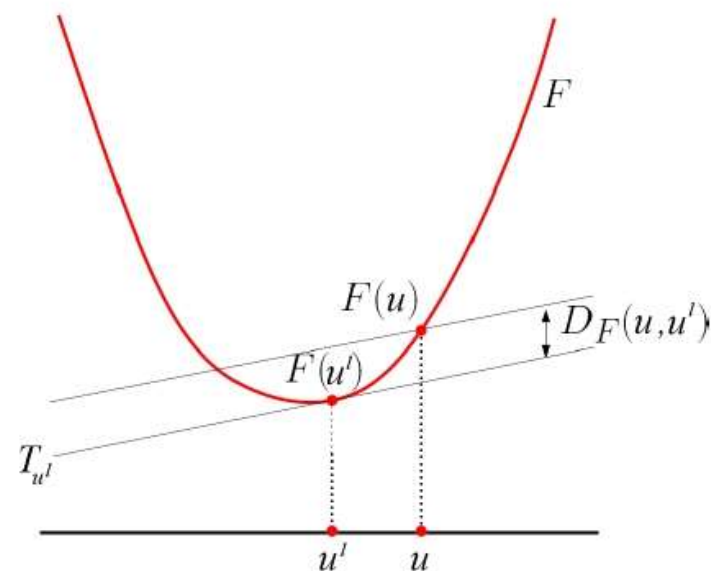
Hence, sign s is harmonic !

s attains its maximum only at the boundary (+1 / -1)

u is supported only on the boundary !!

Measuring Errors

For degenerate regularizations norms do not tell much (too weak or too strong)



Error estimates need appropriate distance measure, generalized

Bregman-distance

$$D_J^p(v, u) = J(v) - J(u) - \langle p, v - u \rangle, \quad p \in \partial J(u)$$

mb-Osher 04, Resmerita 05, mb-Resmerita-He 07, Benning-mb 10

Error Estimation

Define **generalized Bregman distance** for each subgradient

$$p \in \partial J(v)$$

$$D_J^p(u, v) = J(u) - J(v) - \langle p, u - v \rangle$$

Symmetric version

$$d_J(u, v) = \langle p_u - p_v, u - v \rangle$$

Kiwiel 97, Chen-Teboulle 97

Error Estimation

Again under source condition (on exact solution)

$$\tilde{p} = K^* \tilde{w}$$

Starting point is again optimality condition

$$K^*(Ku - K\tilde{u}) + \alpha(p - \tilde{p}) = K^*(f - K\tilde{u}) - \alpha\tilde{p}$$

$$Ku - K\tilde{u} + \alpha(w - \tilde{w}) = f - K\tilde{u} - \alpha\tilde{w}$$

Error Estimation

Take squared norm

$$\begin{aligned} \|Ku - K\tilde{u}\|^2 + 2\alpha\langle u - \tilde{u}, p - \tilde{p} \rangle + \alpha^2\|w - \tilde{w}\|^2 \\ = \|f - K\tilde{u}\|^2 - 2\alpha\langle f - K\tilde{u}, \tilde{w} \rangle + \alpha^2\|\tilde{w}\|^2 \end{aligned}$$

All terms on the left are positive, including Bregman distance

Terms on the right are **noise and bias**

Error Estimation

Statistical model: additive Gaussian noise

$$E[\|f - K\tilde{u}\|^2 - 2\alpha\langle f - K\tilde{u}, \tilde{w} \rangle + \alpha^2\|\tilde{w}\|^2] = \sigma^2 + \alpha^2\|\tilde{w}\|^2$$

Error estimate

$$D_J^{\text{symm}} \leq \frac{\sigma^2}{\alpha} + \alpha\|\tilde{w}\|^2$$

Error Estimation

For squared Hilbert space norm the Bregman distance is squared norm distance – consistency with Hilbert space theory

For energies homogeneous of degree one, we have

$$J(v) = \langle p, v \rangle$$

Bregman distance becomes

$$D_{TV}^p(u, v) = J(u) - \langle p, u \rangle$$

Error Estimation

Bregman distance for singular energies is not a strict distance, can be zero for $u \neq v$

In particular d_{TV} is **zero for contrast change**

$$d_{TV}(u, f(u)) = 0$$

Resmerita-Scherzer 06

Bregman distance is still **not negative** (convexity)

Bregman distance can provide **information about edges**

Error Estimation

Let v be piecewise constant with white background and grey values c_i on regions Ω_i

Then we obtain subgradients of the form

$$p^\epsilon = \nabla \cdot (f^\epsilon(b_i) \nabla b_i)$$

with signed distance function b_i^{and}

$$0 \leq f_i^\epsilon \leq 1, \quad f_i^\epsilon(0) = 1$$

$$\text{supp } f_i^\epsilon = -(\epsilon, \epsilon)$$

Error Estimation

Bregman distances given by

$$D_{TV}^{p^\epsilon}(u, v) = \sup_{g, \|g\|_\infty \leq 1} \int u \nabla \cdot (g - f^\epsilon(b_i) \nabla b_i)$$

In the limit we obtain for being piecewise continuous

$$\liminf D_{TV}^{p^\epsilon}(u, v) \geq d^0(u, v) = |u|_{TV(\Omega \setminus \cup \partial \Omega_i)}$$

Local Loss of Contrast

TV minimization **loses contrast locally**, total variation of the reconstruction is smaller than total variation of clean image.

Image features left in residual $f-u$

a clean



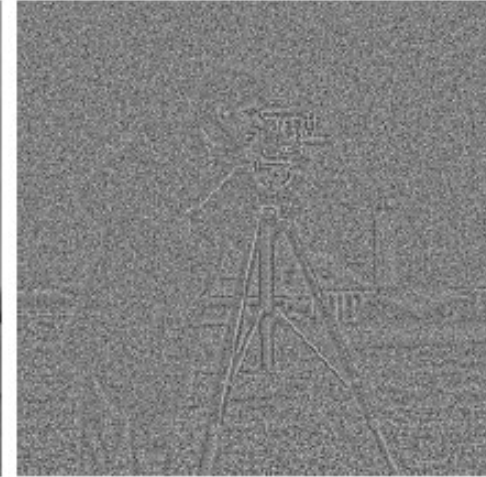
f noisy



u ROF



$f-u$



Loss of Contrast = Systematic Bias of TV

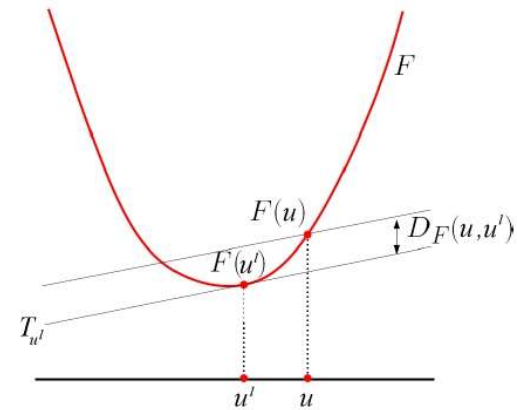
Becomes more severe in ill-posed problems with operator K

Not just simple vision effect to be corrected, but loss of information

Simple idea for Least-Squares:

add back the noise to amplify = Augmented Lagrangian

$$\begin{cases} u^{l+1} &= \arg \min_u \left\{ \frac{1}{2} \|Ku - f^l\|_{L^2(\Sigma)}^2 + \alpha J(u) \right\} \\ f^{l+1} &= f^l + f - Ku^{l+1}, \end{cases}$$



Bregman Iteration

Can be shown to be equivalent to Bregman iteration

$$u^{l+1} = \arg \min_u \left\{ \frac{1}{2} \|Ku - f\|_2^2 + \alpha D_J^{p^l}(u, u^l) \right\}$$

$$D_J^p(v, u) = J(v) - J(u) - \langle p, v - u \rangle, \quad p \in \partial J(u)$$

Immediate generalization to convex fidelities and regularizers

$$\min_{u \in \mathcal{W}(\Omega)} \{H_f(Ku - f) + \alpha J(u)\}$$

Primal Bregman Iteration

$$u^{l+1} = \arg \min_{u \in \mathcal{W}(\Omega)} \left\{ H_f(Ku - f) + \alpha D_J^{p^l}(u, u^l) \right\}$$

Bregman Iteration

Note: Primal Bregman Iteration not equivalent to ALM for non-quadratic fidelities !

ALM

$$u^{l+1} = \arg \min_{u \in \mathcal{W}(\Omega)} \{ H_f(Ku + r^l - f) + \alpha J(u) \}$$

$$r^{l+1} = r^l + Ku^{l+1} - f$$

is equivalent

$$q^{l+1} = \arg \min_{q \in \mathcal{V}(\Sigma)^*} \left\{ \alpha J^*\left(\frac{1}{\alpha} K^*(q)\right) - \langle f, q \rangle + D_{H_f^*}^{r^l}(-q, -q^l) \right\}$$

Bregman Iteration

Properties like iterative regularization method

Regularizing effect from **appropriate termination of the iteration**

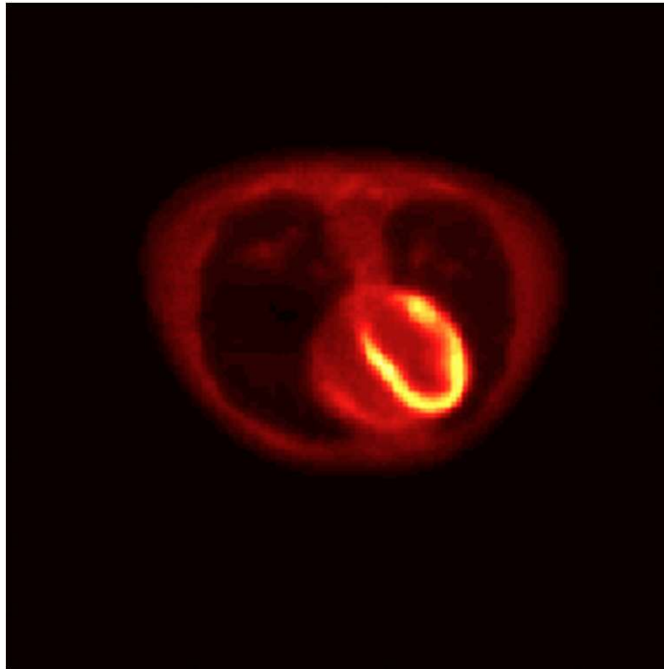
Better performance for oversmoothing single steps, i.e. **regularization parameter α very large**

Limit: Inverse Scale Space Method

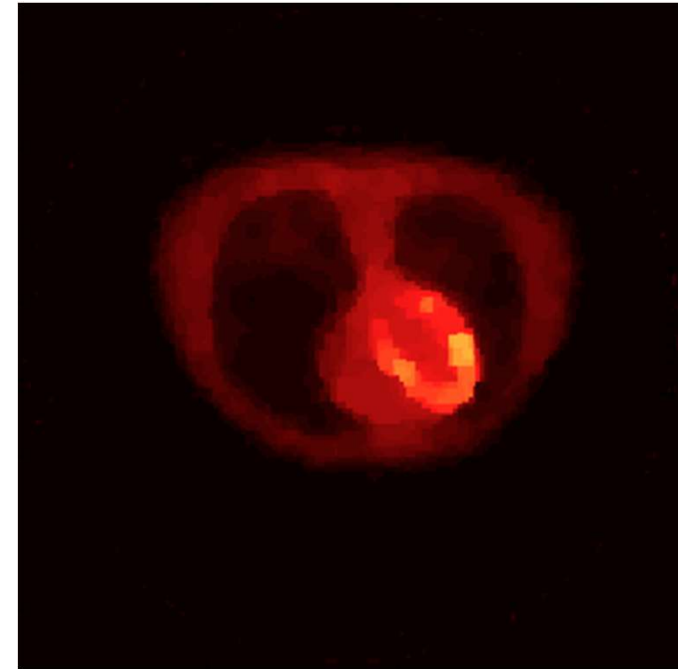
$$ml \quad \frac{d}{dt} p(t) = -K^*(\partial H_f(Ku(t) - f))$$

^{18}F -FDG

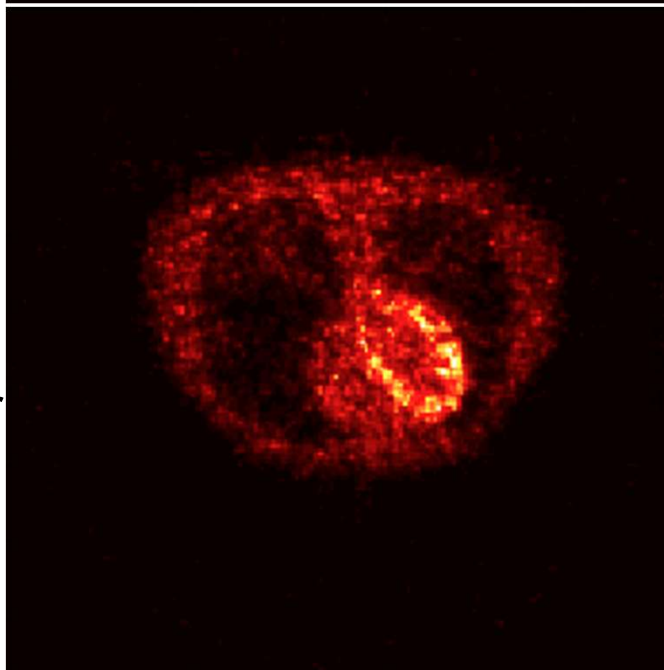
EM, 20 min



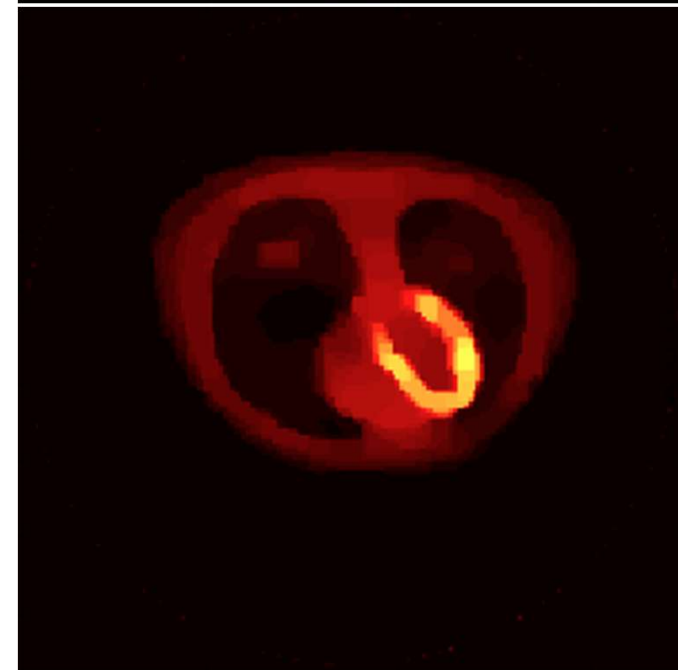
EM-TV, 5s



EM, 5s



BRG/ISS, 5s



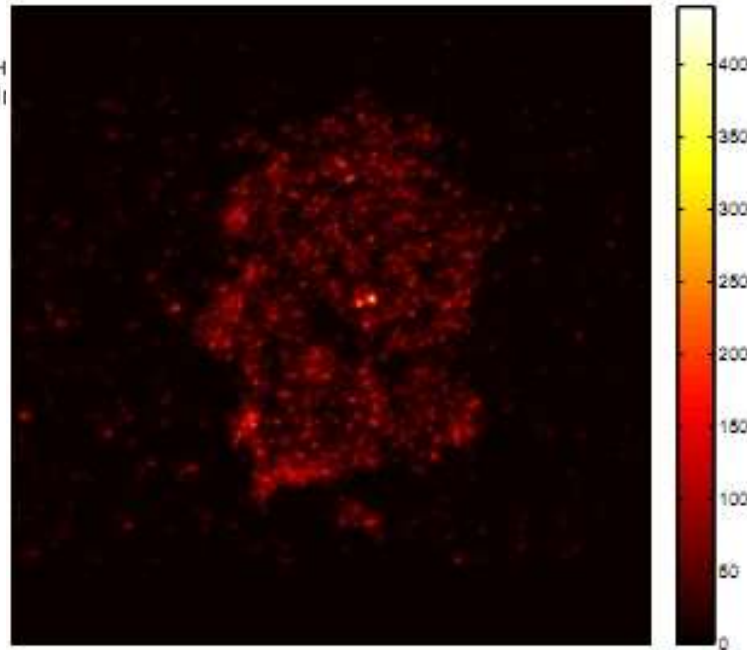
*Jahn Müller, 2011
Data from Nuclear
Medicine
Department, UKM*



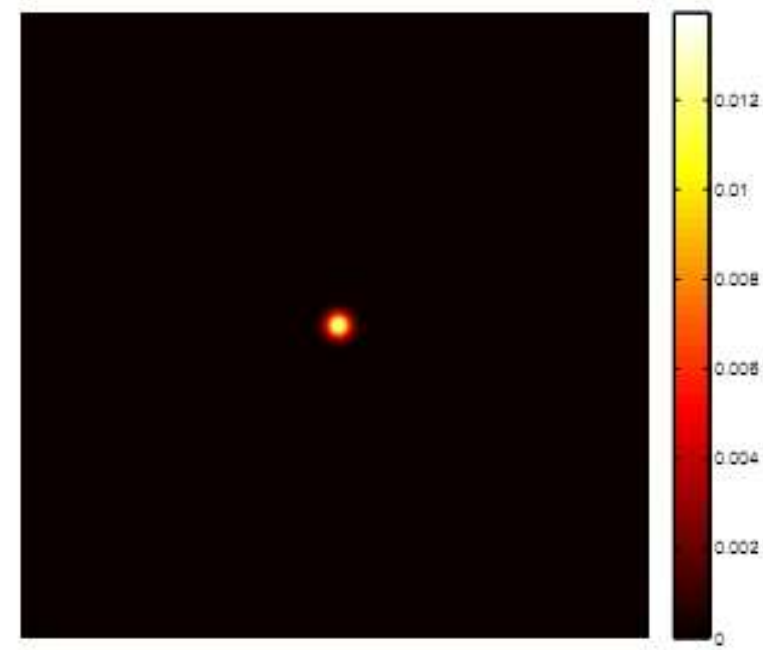
WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

STED Microscopy

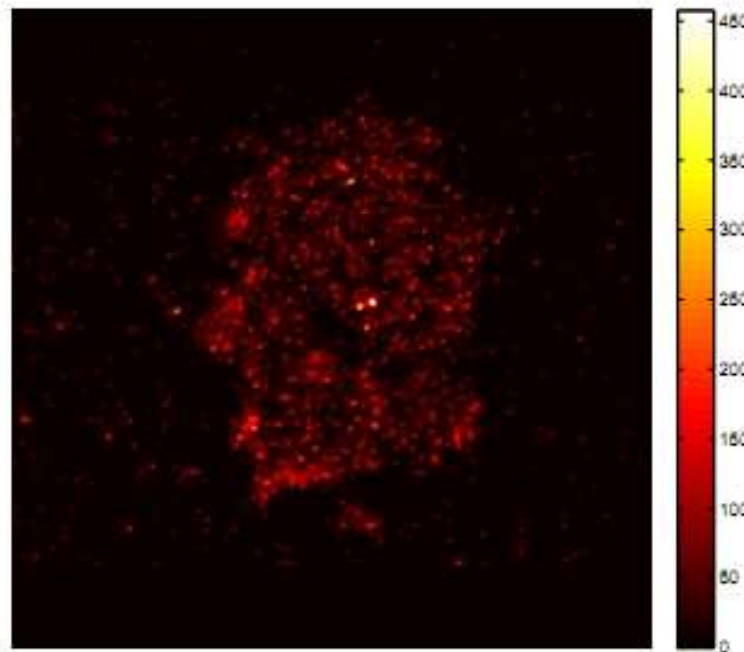
*Christoph Brune,
2009
Data from MPI for
Biophys. Chem.
Göttingen
(K.Willig,
A.Schönle,
Hell)*



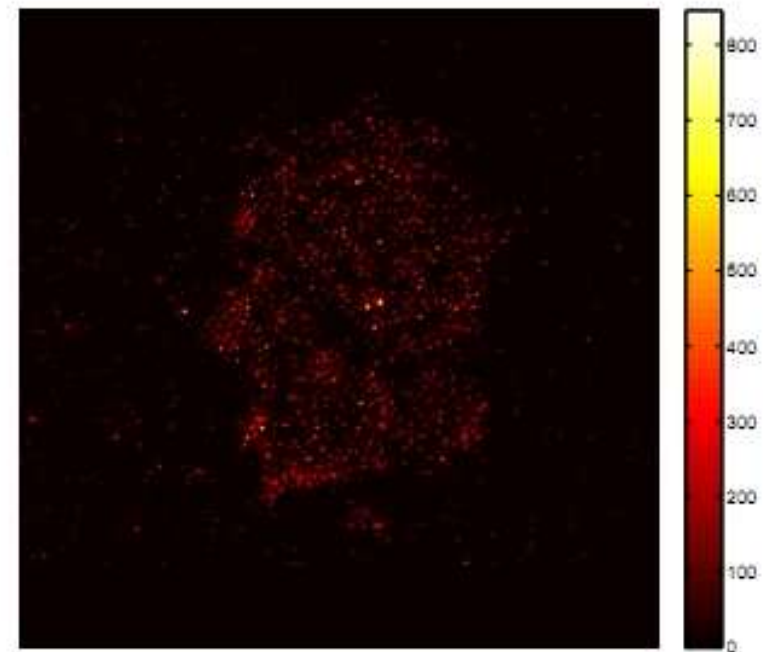
(a) Syntaxin



(b) psf 3d visualization



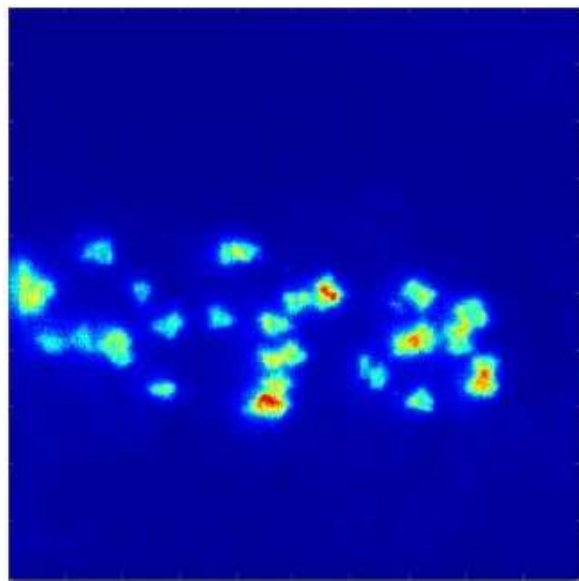
(c) EM-TV



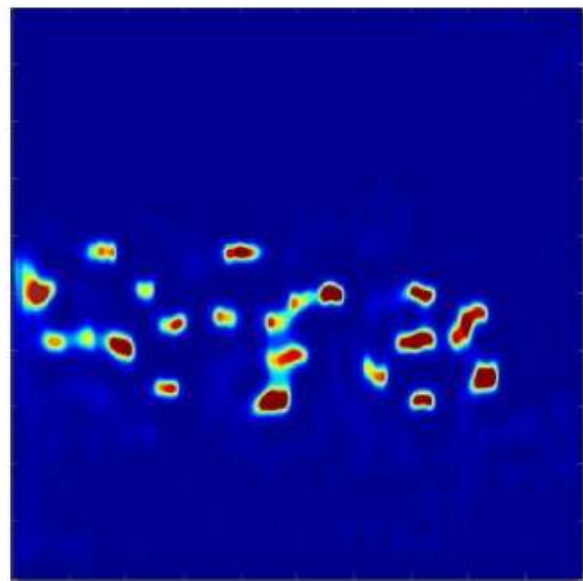
(d) Bregman-EM-TV

Protein Bruchpilot (4pi)

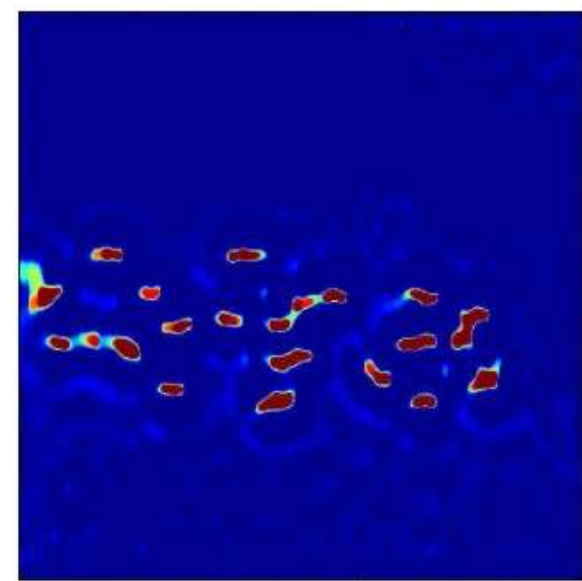
Protein in neuromuscular junction of larval *Drosophila*



Experimental data

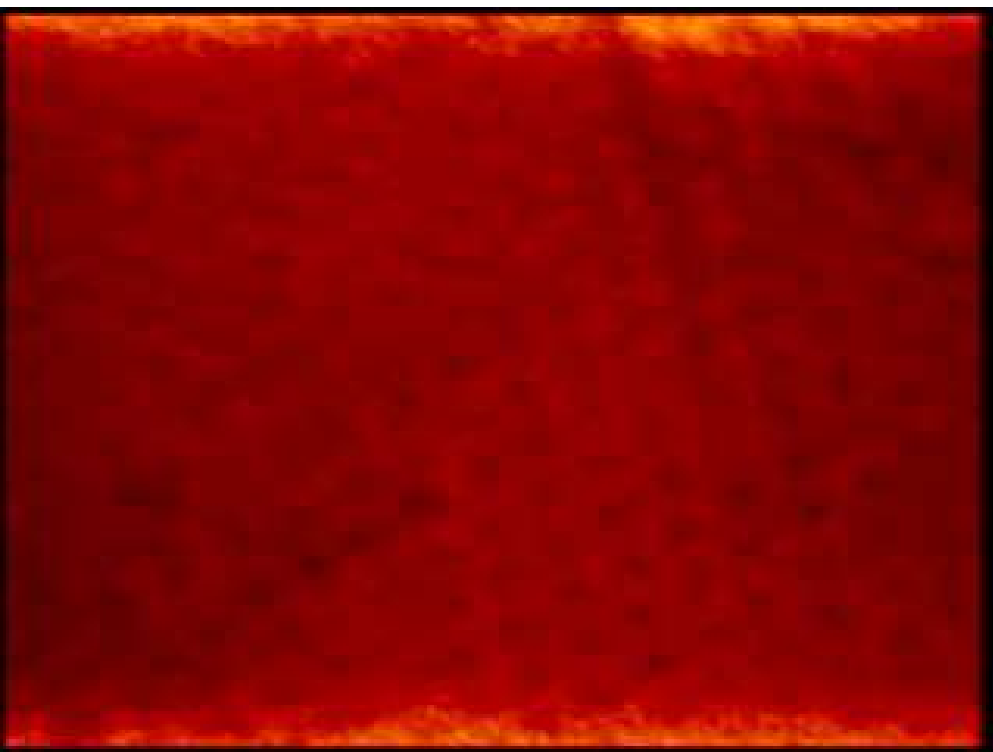


EM-TV

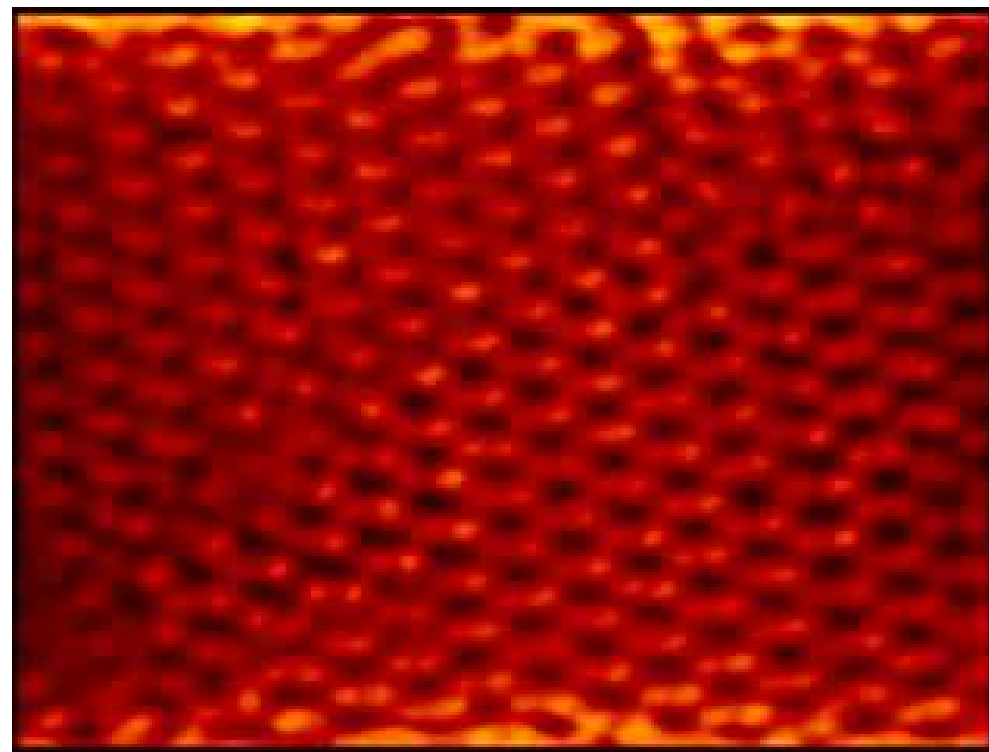


Bregman EM-TV

3D Beads Crystal Structure



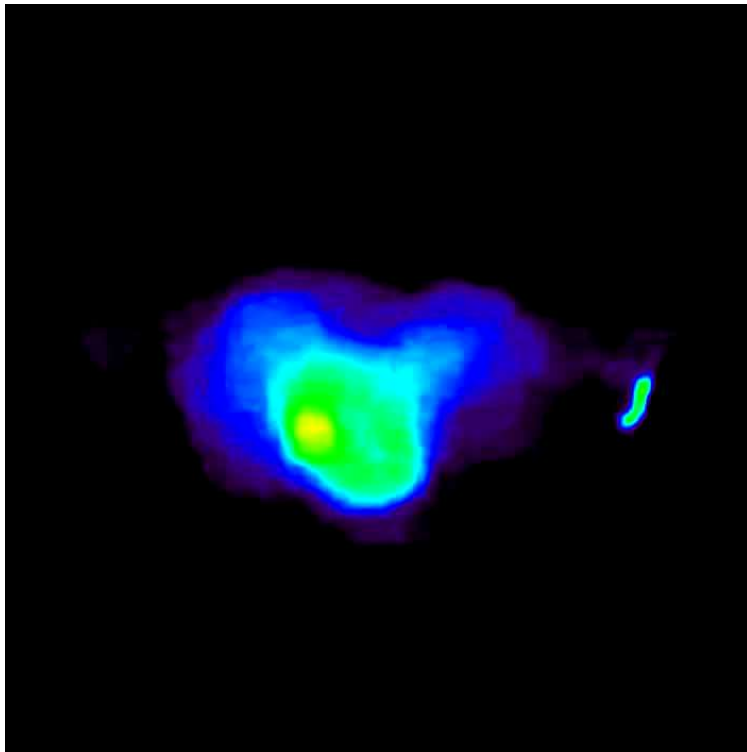
Original STED



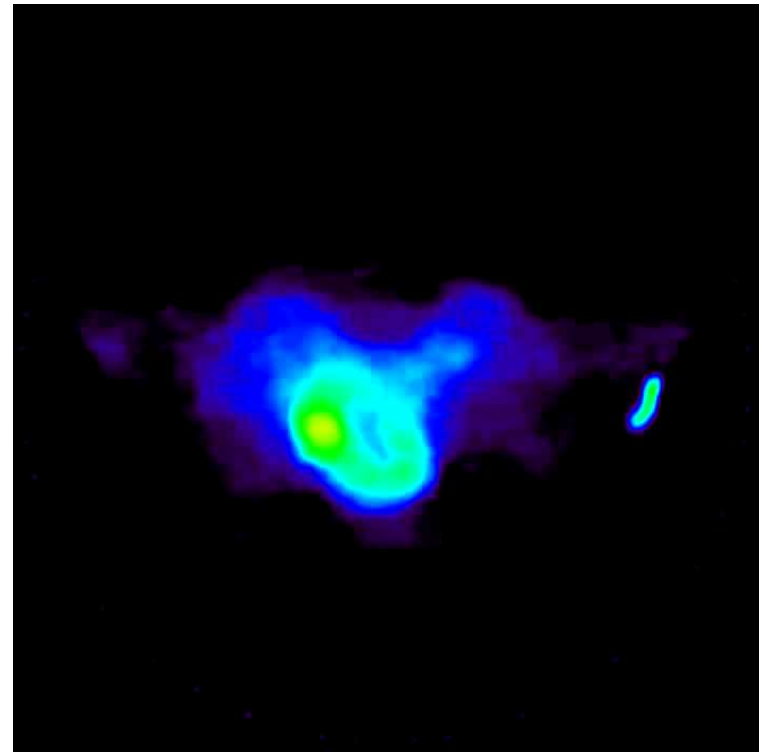
EM-TV

Cardiac H_2^{15}O -Scan

EM-TV



EM-TV Bregman



Jahn Müller, 2011

Data from Nuclear Medicine Department, UKM

Why does Inverse Scale Space work ?

Singular value decomposition in fully quadratic case

$$J(u) = \frac{1}{2}\|u\|^2, \quad H_f(Ku - f) = \frac{1}{2}\|Ku - f\|^2$$

Eigenfunctions: $\sigma_n^2 u_n = K^* K u_n$

$$f = c K u_n \text{ s.t. } u = c_n(t) u_n$$

$$\frac{dc_n(t)}{dt} = \sigma_n^2 (c - c_n(t))$$

Convergence faster in small frequencies (large eigenvalues)

Why does Inverse Scale Space work ?

Convex **one-homogeneous** regularization J (TV, l_1 , ...)

Eigenfunctions: $\lambda K^* K \hat{u} \in \partial J(\hat{u})$

$$f = cK\hat{u} \text{ yields } u = \begin{cases} 0 & t < \frac{1}{\lambda c} \\ c\hat{u} & t \geq \frac{1}{\lambda c} \end{cases}$$

Again large frequencies appear later. Not at all for small t !

Eigenvalues in TV indeed related to jump measures

mb-Benning, 2013

Why does Inverse Scale Space work ?

Multiple frequencies not simple for nonlinear case

However, various theoretical and computational results confirming **exact scale decomposition**

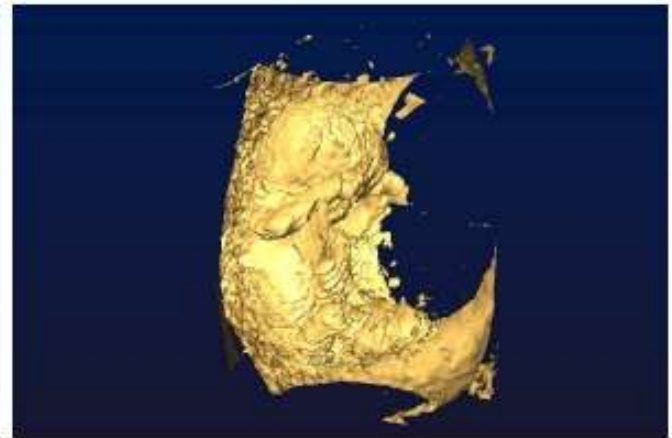
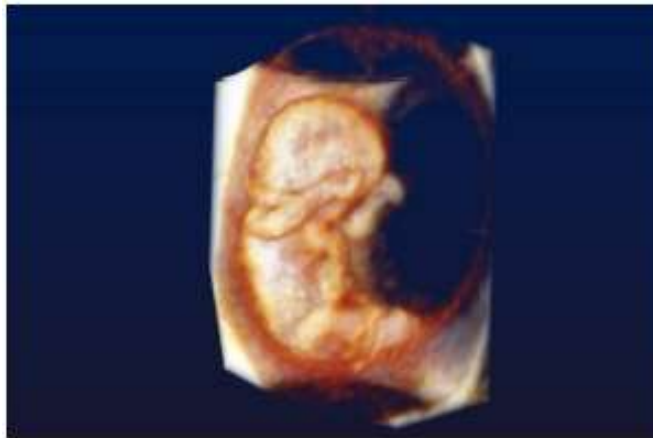
PhD-Thesis Benning, 2011 / mb-Frick-Scherzer-Osher 2007

Complete characterization of inverse scale space **for discrete l1-functionals**, yields jump dynamics in time, **adaptive basis pursuit method** with guaranteed convergence

mb-Möller-Benning-Osher, 2011

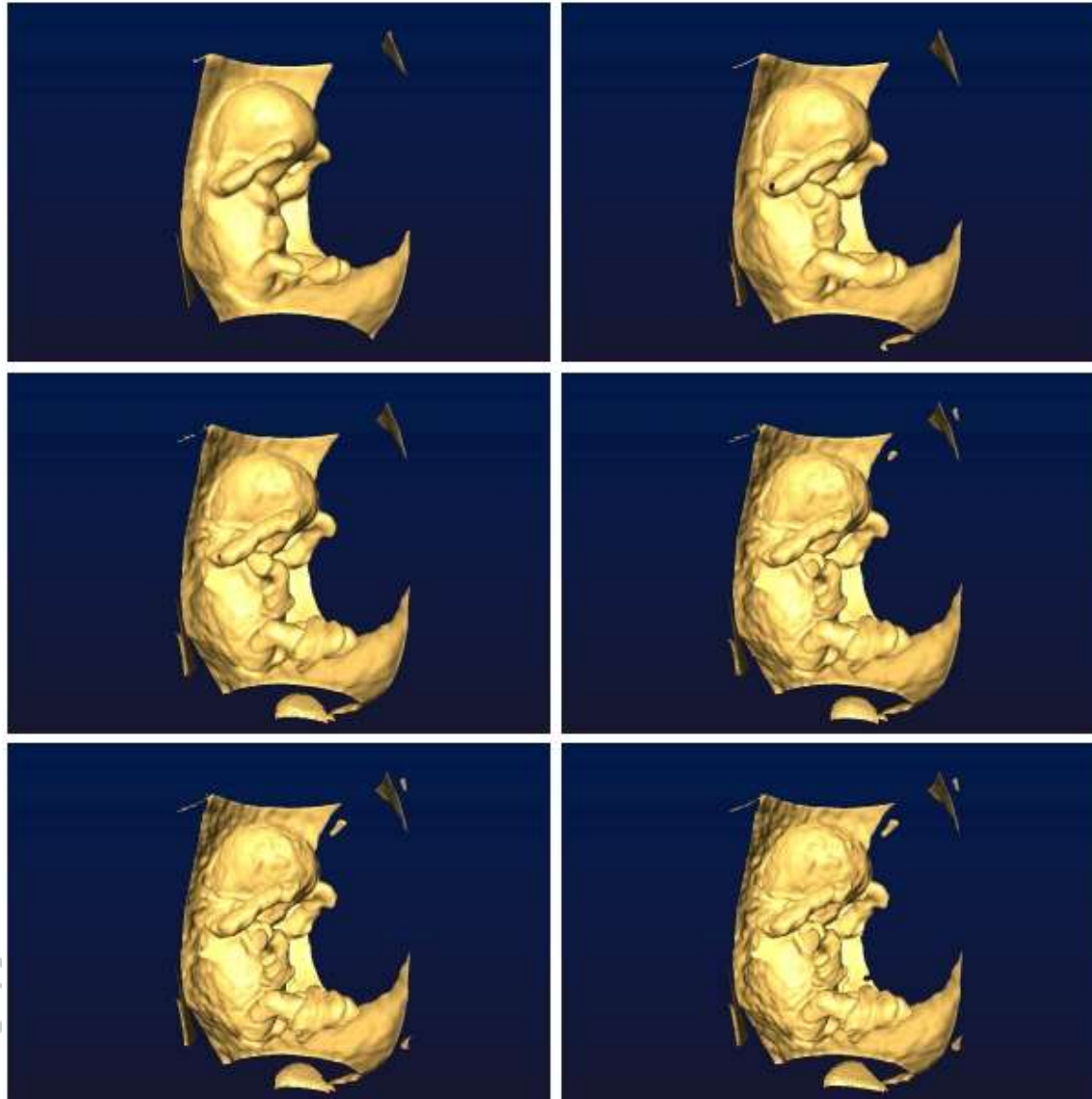
Surface Smoothing

Smoothing of surfaces in level set representation



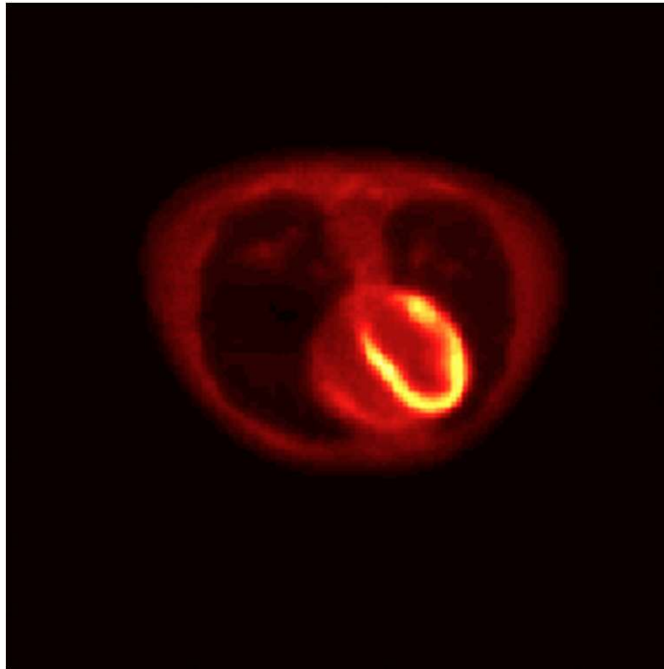
3D Ultrasound, *level set, surface*

Bregman Iteration / Inverse Scale Space

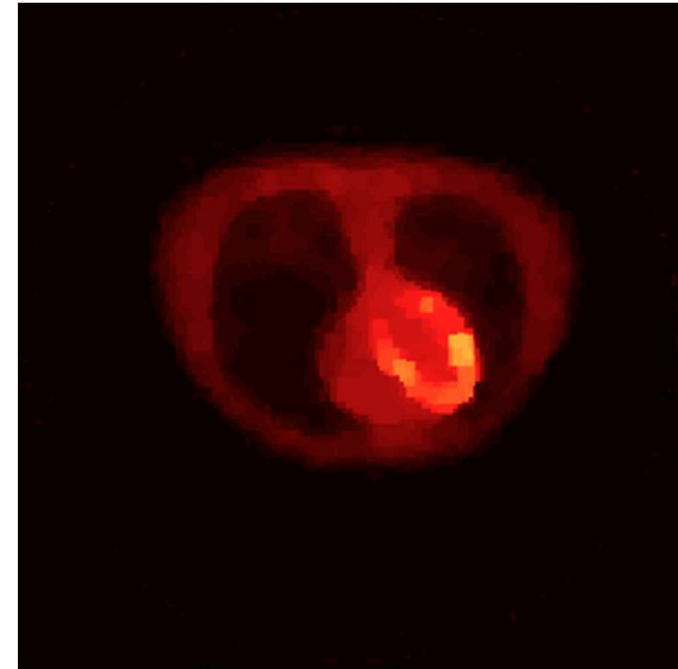


^{18}F -FDG

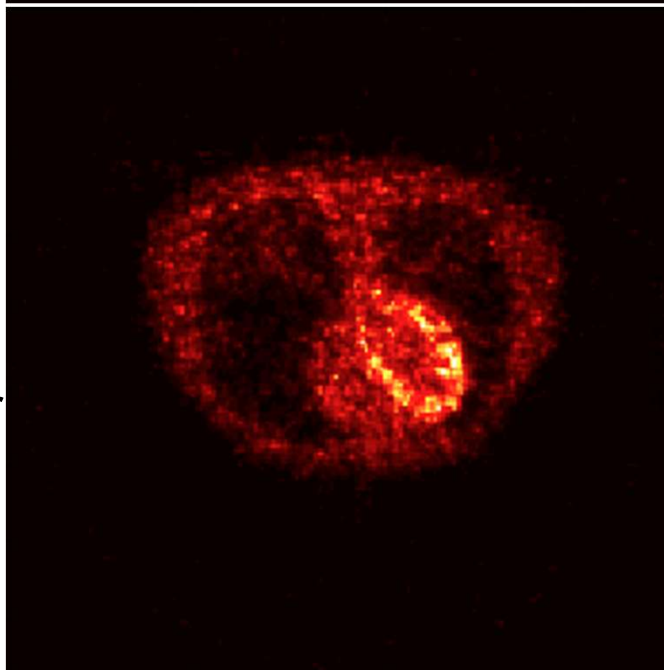
EM, 20 min



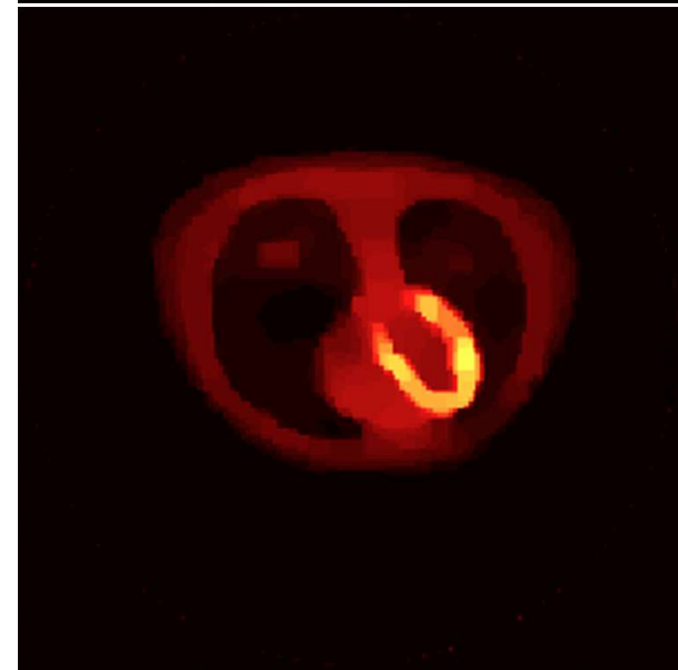
EM-TV, 5s



EM, 5s



BREGMAN
EM-TV 5s



*Jahn Müller, 2011
Data from Nuclear
Medicine
Department, UKM*

Beyond staircasing

Major idea: combine TV with higher order TV

Infimal convolution

$$\text{ICTV}_\beta(u) := (\text{TV} \square \beta \text{TV}^2)(u) = \inf_{u=v+w} (\text{TV}(v) + \beta \text{TV}^2(w))$$

Dual version

$$\text{ICTV}_\beta(u) = \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^d) \\ q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^d)) \\ \|p\|_\infty \leq 1, \|q\|_\infty \leq 1 \\ \beta \text{div}^2(q) = \text{div}(p)}} \int_{\Omega} u \text{div}(p) \, dx$$

Chambolle-Lions 97

GTV

Decomposition by inf-convolution not optimal, improvement by stronger dual constraint

$$\text{GTV}_\beta(u) := \sup_{\substack{q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^d)) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1}} \int_{\Omega} u \, \text{div}^2(q) \, dx$$

Primal

$$\text{GTV}_\beta(u) = \inf_{\substack{u=v+w \\ s \in \mathcal{N}(\text{div})}} \left\{ \int_{\Omega} |\nabla v - s| + \beta \int_{\Omega} |\nabla^2 w + \nabla s| \right\}$$

Bredies et al 2011

GTV

Single tensor: **generalized TV** (Bredies et al 2010, Setzer et al 2010)

$$\text{GTV}_\beta(u) := \sup_{\substack{q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1}} \int_{\Omega} u \, \text{div}^2(q) \, dx$$

Note

$$\text{GTV}_\beta(u) \leq \text{ICTV}_\beta(u) \leq \text{TV}(u)$$

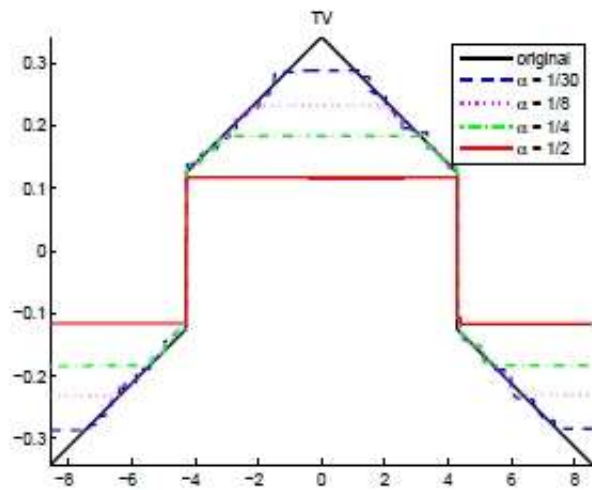
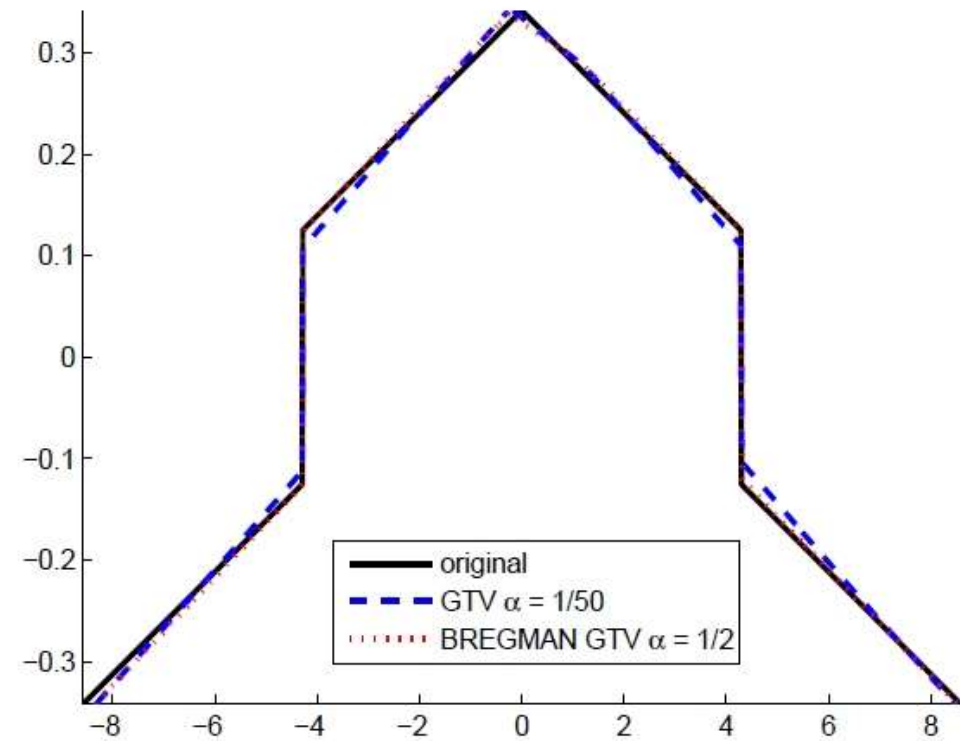
All functionals induce equivalent norms on functions with vanishing zero and first moment

But different reconstructions !

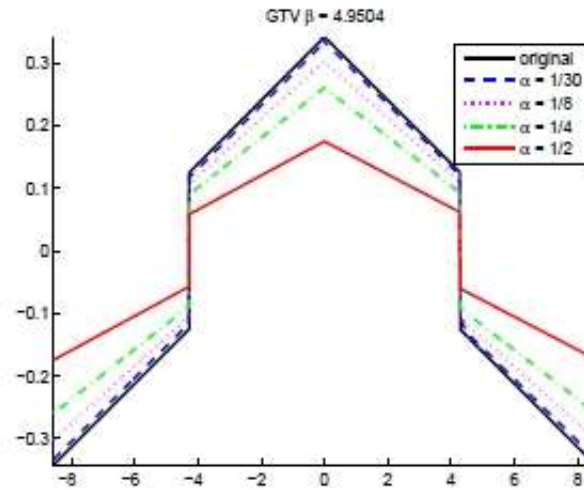


WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

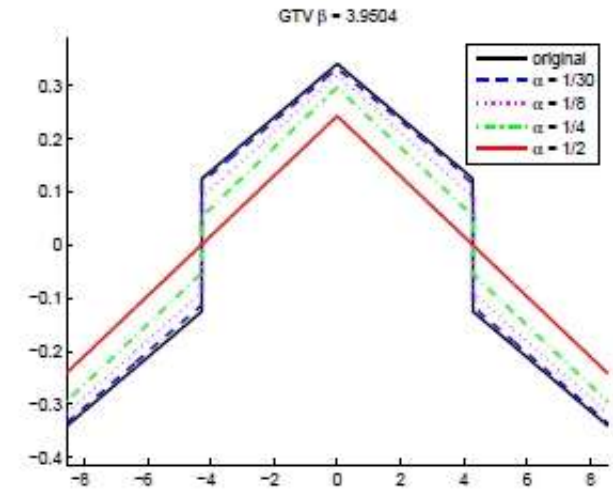
Staircasing



(a) TV

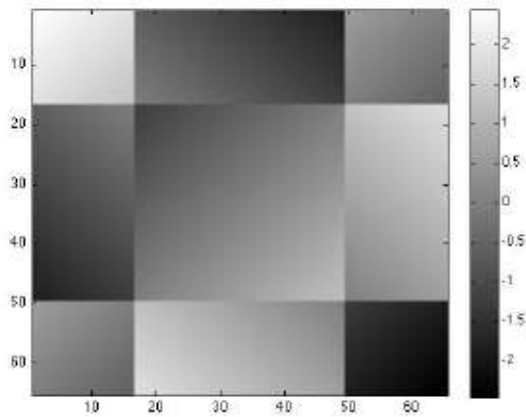


(b) GTV $\frac{32}{2\sqrt{3}+3}$

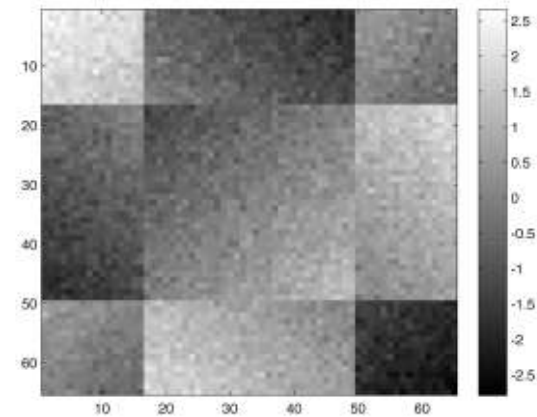


(c) GTV $\frac{32}{2\sqrt{3}+3} - 1$

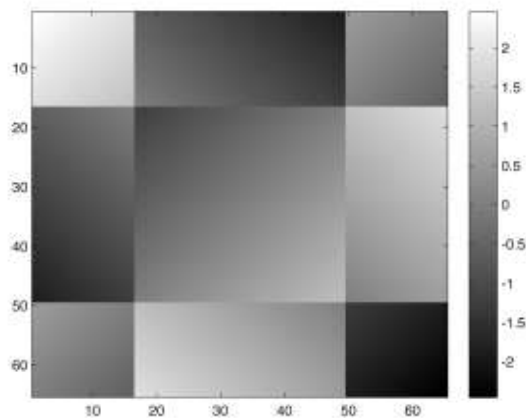
Bregman-GTV Denoising



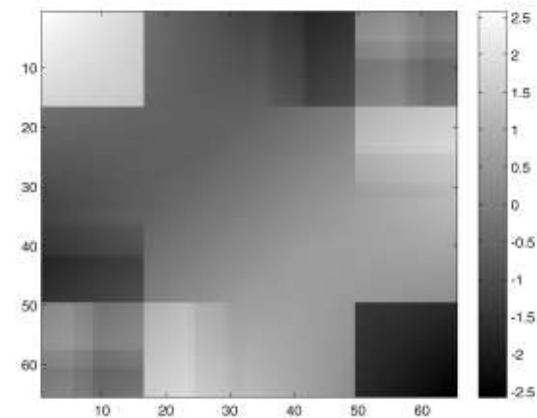
(a) original image



(b) noisy image



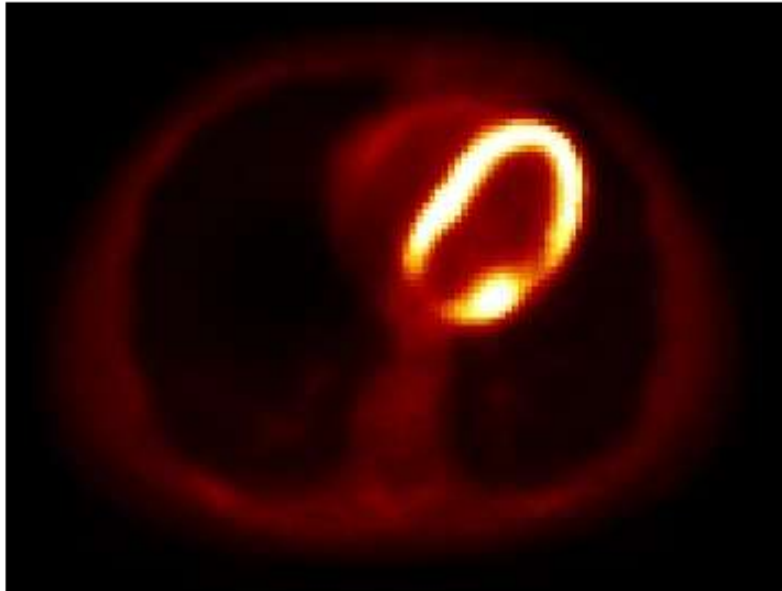
(c) Bregman GTV



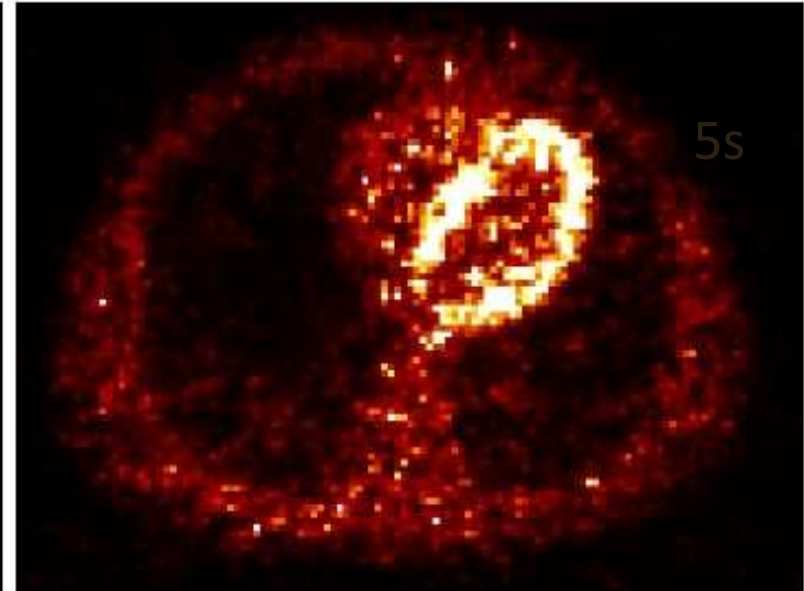
(d) Bregman ICTV

Bregman EM-GTV in PET

EM,
20 min



Bregman
EM-TV,
5s



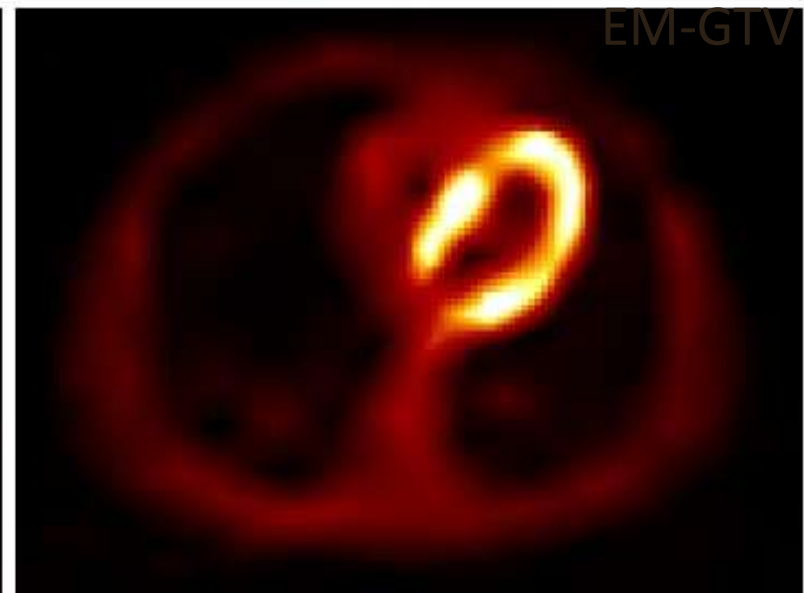
EM,

5s

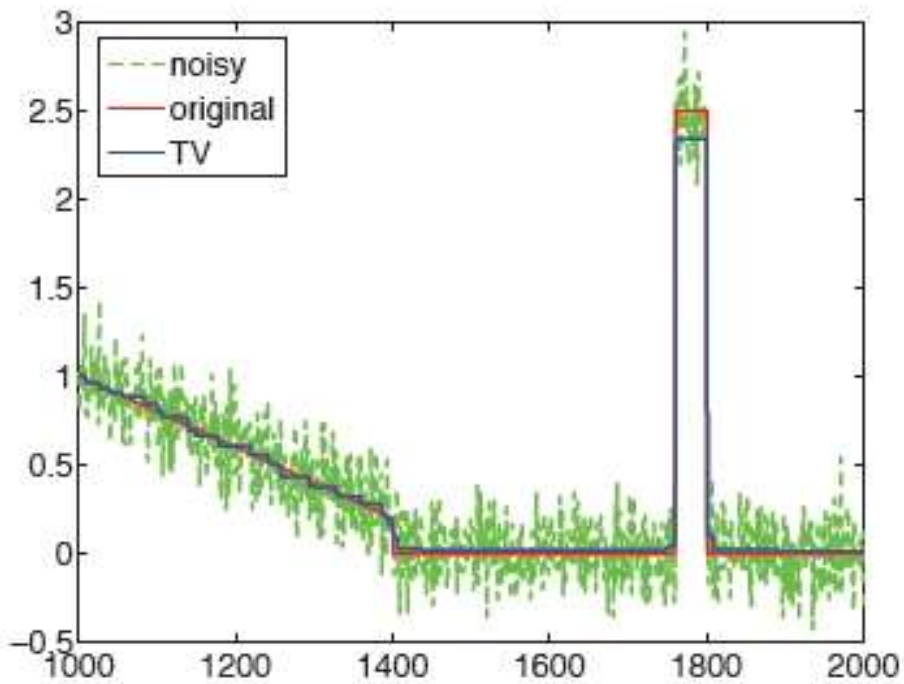
Bregman

EM-GTV

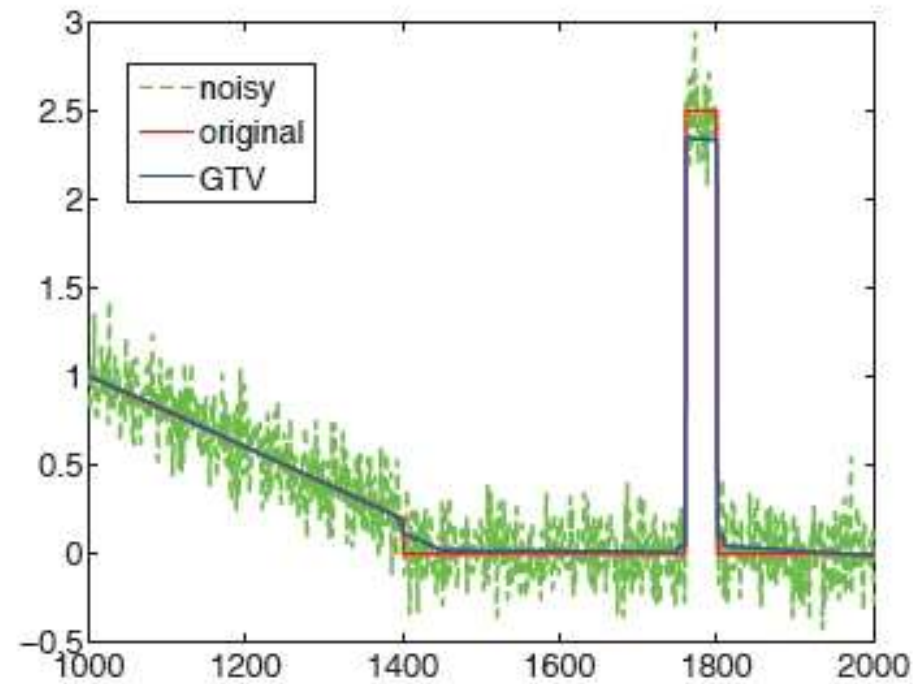
5s



ICTV / GTV denoising



(a) TV

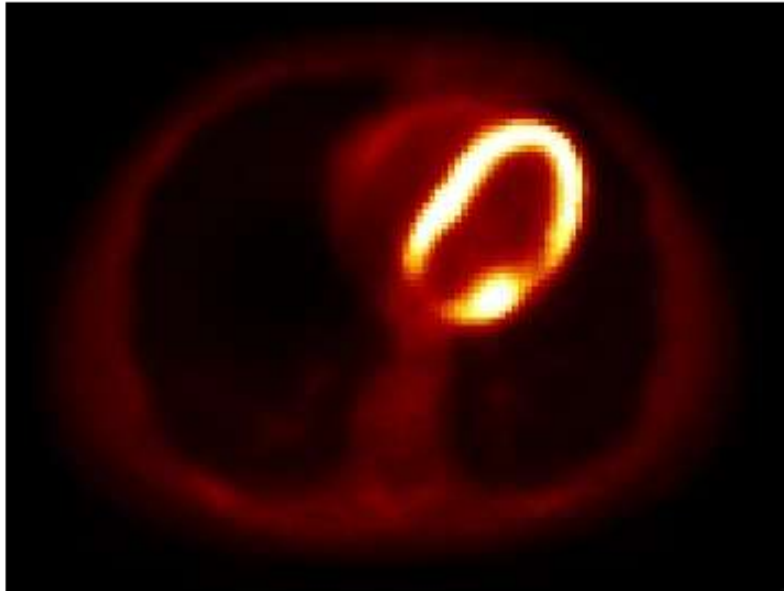


(b) GTV

Benning-Brune-mb-Müller 2013

Bregman EM-GTV in PET

EM,
20 min

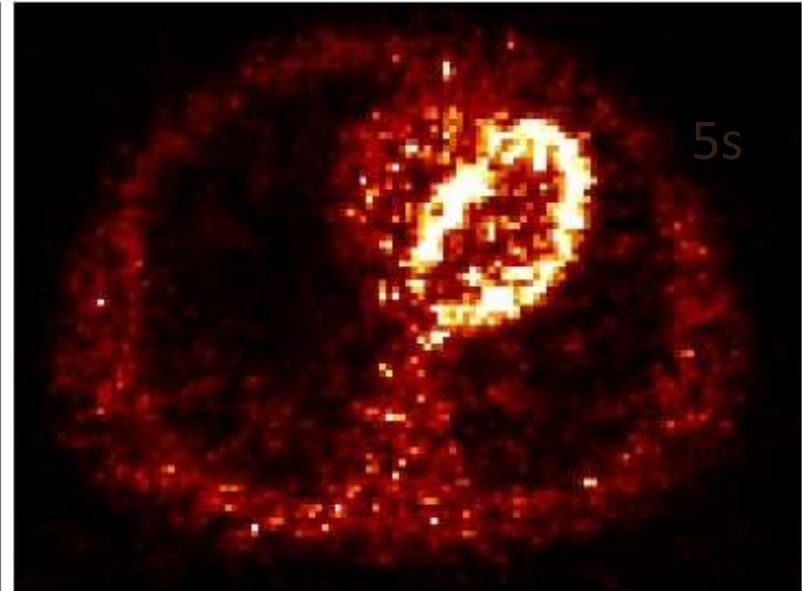


Bregman
EM-TV,
5s



Martin Burge

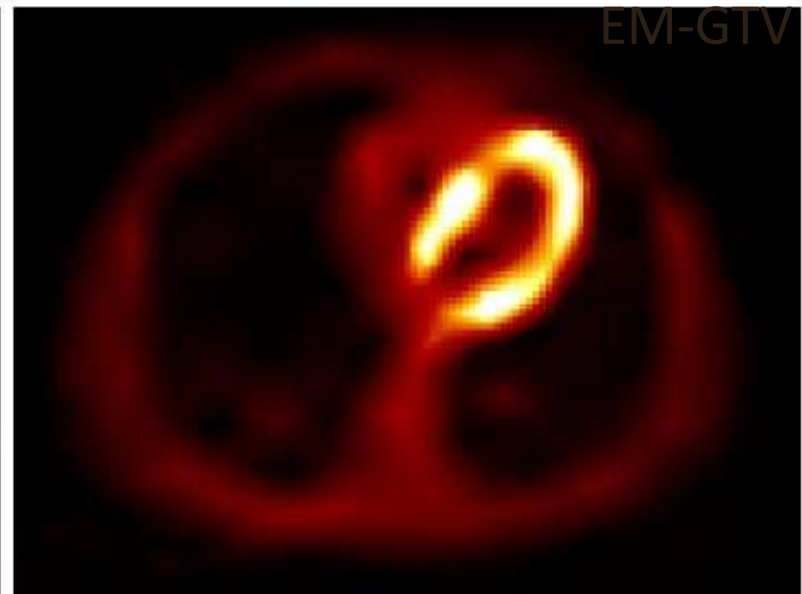
EM,



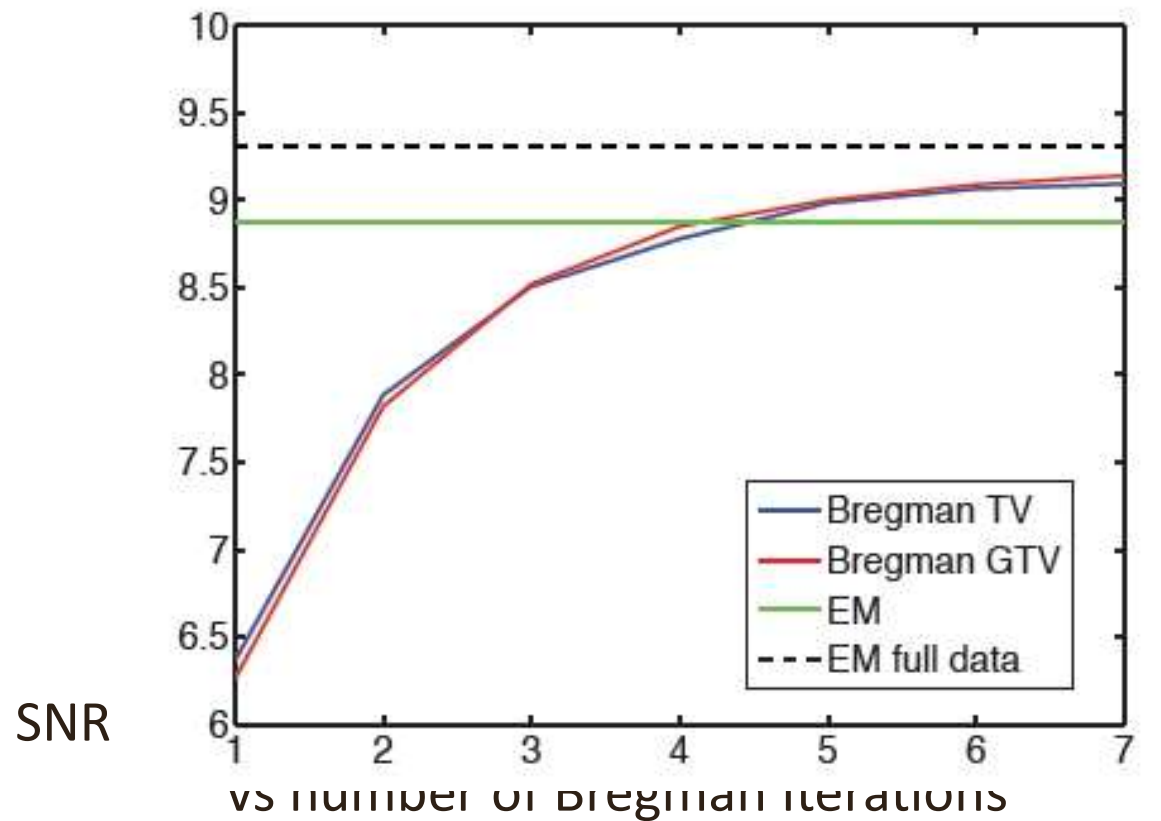
Bregman

EM-GTV

5s

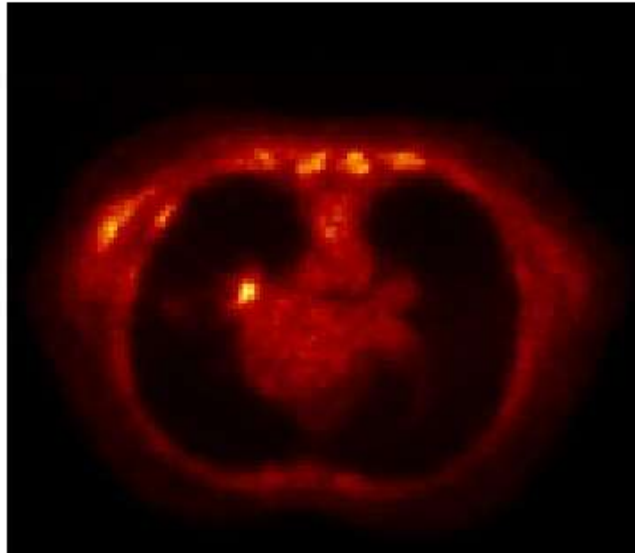


Realistic XCAT-GATE Software Phantom

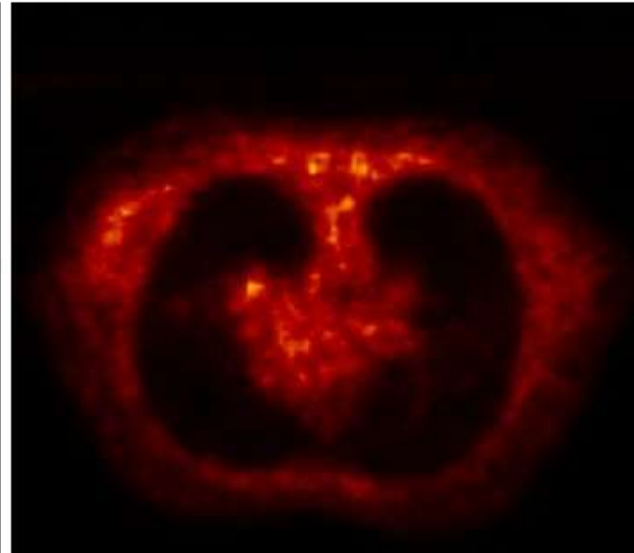


Oncology

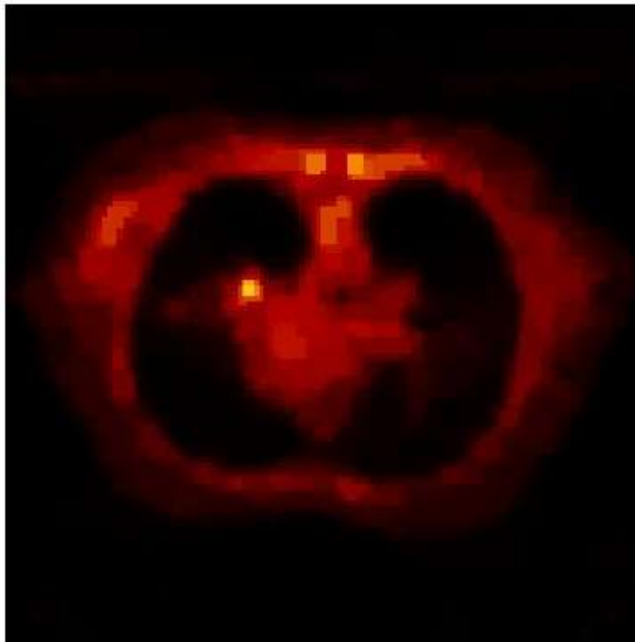
EM,
7 min



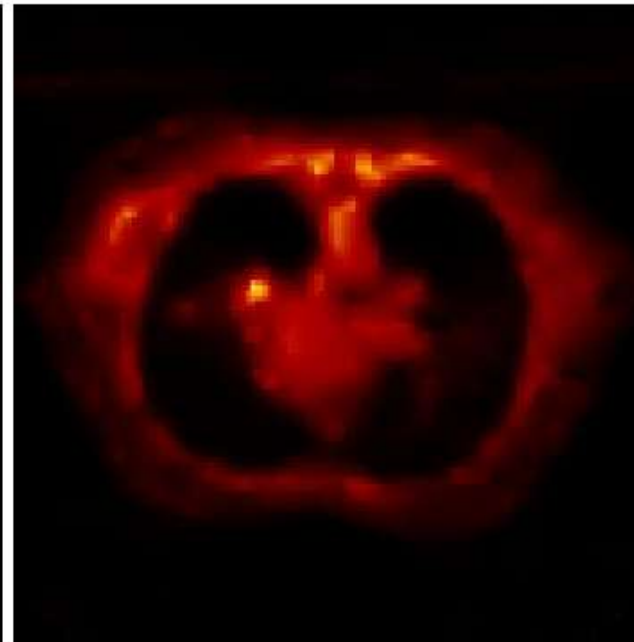
EM,



Bregman
EM-TV,
30s



Bregman
EM-GTV
30s



Bregman iteration, Inverse Scale Space Flow

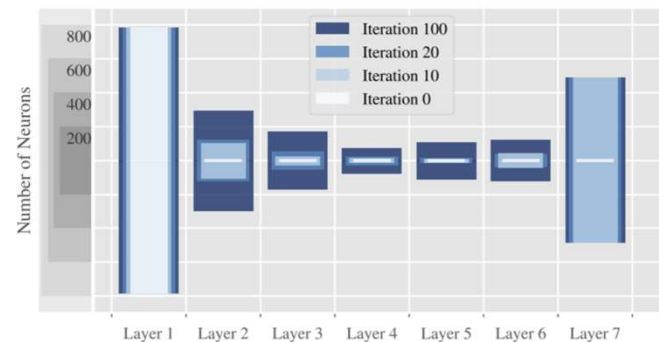
Bregman Iteration

$$p^{k+1} = p^k + \tau K^* \partial F(Ku^{k+1}, f)$$

interpreted as implicit Euler discretization with time step t of *inverse scale space*

$$\partial_t p = -K^* \partial F(Ku, f), \quad p \in \partial J(u)$$

[mb-Gilboa-Osher-Xu 2006, mb-Frick-Osher-Scherzer 2007, Brune-Sawatzky-mb 2011, mb-Möller-Benning-Osher 2012]



Recent development: stochastic linearized Bregman methods for training sparse deep neural networks [Bungert-Roith-Tenbrinck-mb 2021]

Robust Optimization Algorithms

Solve Problem of the form

$$\int_{\Sigma} D(Ku, f) \, dx + \alpha J(u) \rightarrow \min_u$$

Basic paradigms:

- **Never solve** linear systems **with** full matrices (derived from **K**)
- **Never** use **linearization** or crude approximations **of TV or similar**
- **Modularity in regularization** (if other prior knowledge)

Obvious choice: forward-backward splitting (*Lions-Mercier 79*)

Forward Backward Splitting

Explicit half step

$$u^{k+1/2} = u^k - \tau K^* \partial_u D(Ku, f)$$

Combined with implicit half step

$$u^{k+1} + \alpha \tau \partial J(u^{k+1}) \ni u^{k+1/2}$$

Forward Backward Splitting

First half step only needs efficient application of operator K and its adjoint

Second half-step is a standard denoising problem

$$\frac{1}{2} \int (u - f)^2 + \alpha \tau J(u) \rightarrow \min_u$$

Well-established techniques for denoising

Robust Algorithms for Poisson: FB-EM-TV

Example: Poisson Noise with TV regularization of density image

$$\min_{\substack{u \in BV(\Omega) \\ u \geq 0 \text{ a.e.}}} \int_{\Sigma} (Ku - f \log Ku) \, d\mu + \alpha |u|_{BV(\Omega)}$$

Forward step can destroy positivity and well-definedness in forward splitting

Advanced Issues: Novel Inverse Problems

- **Dynamic** Reconstructions (4D or even 5D with spectral info)
- **Fast** measurements (leading to low SNR)
- **Quantitative** analysis needed (coupling to mathematical models)
- **Underdetermination** by data (prior knowledge, models)
- **Uncertainty** to be quantified (stochastics, sensitivity)

Model-based Imaging

In 4D imaging (space + time), prior information about the dynamics can often be used to constrain set of possible images

Dynamics obtained from mathematical models:

- Fluid dynamics
- Electrical activity in the heart (bidomain models)
- Biochemical reactions
- Models for cell dynamics, intracellular flow
- **Models for the dynamics of tracer uptake in PET**

Regularization and Constraints

Dynamical priors as constraints or as regularization:

constraints: model equations to be satisfied, only parameters in the model to be determined – new (lower-dimensional) set of unknowns

- regularization: appropriate basis sets extracted from analytical or numerical solutions of the models, regularization favours sparse representations (explanation by few effects)