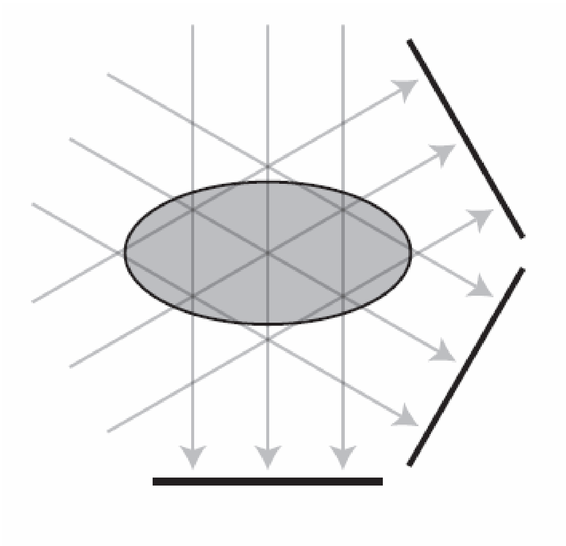
Total Variation Regularization and Large Scale Volume Reconstructions in Tomography

Sami S. Brandt

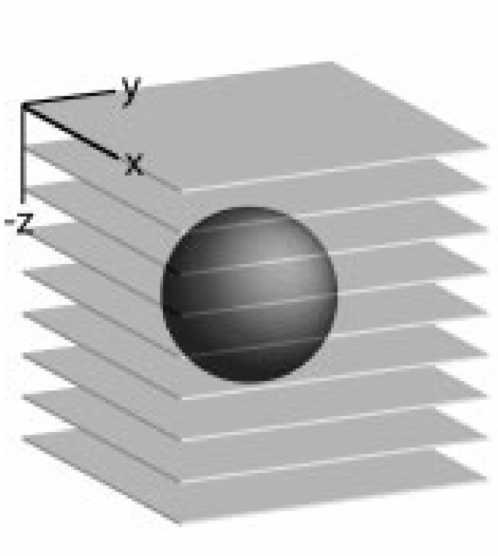
CCBR - SYNARC
WHERE SCIENCE MEETS SERVICE

UNIVERSITY of OULU
OULUN YLIOPISTO

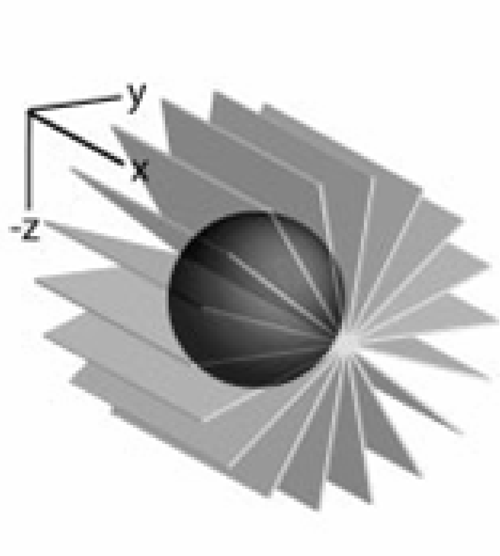
Linear Volumetric Imaging



(a) Tomography



(b) Confocal z-stack imaging

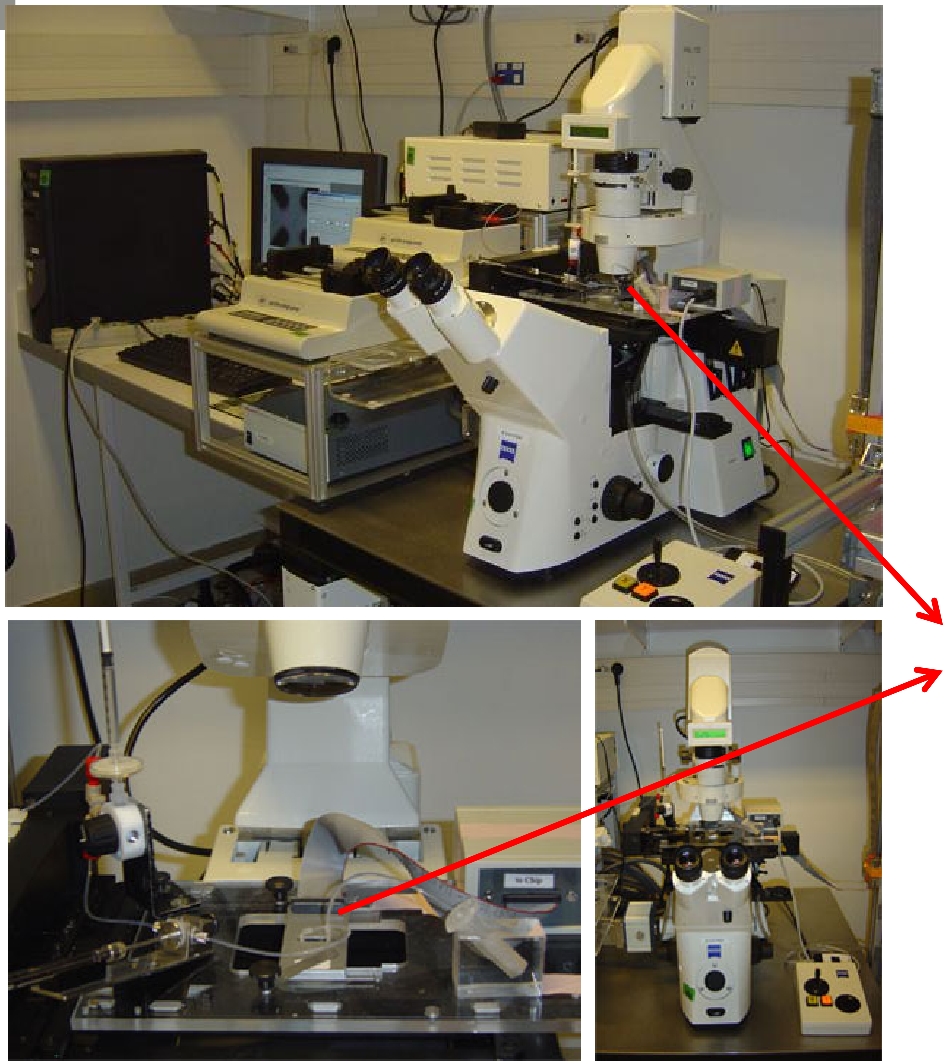


(c) Microrotation confocal imaging

Contents

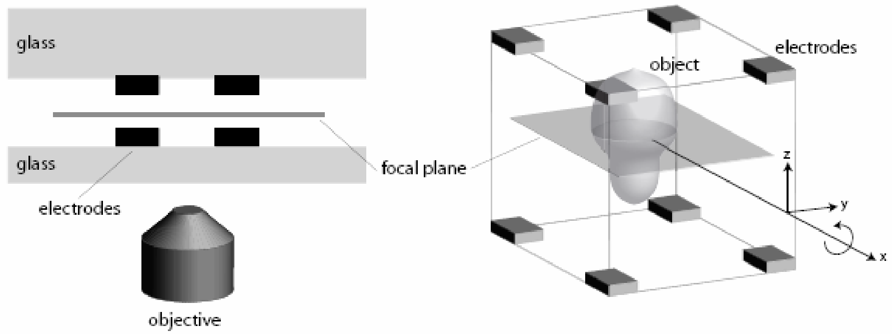
- Bayesian reconstruction problem in volumetric imaging
- Sparsity regularisation by total variation and spatial derivative priors
- High Dimensional Optimization
  - EM
  - Skilling-Bryan
- Application example
  - Microrotation reconstruction

Micro-rotation fluorescence imaging



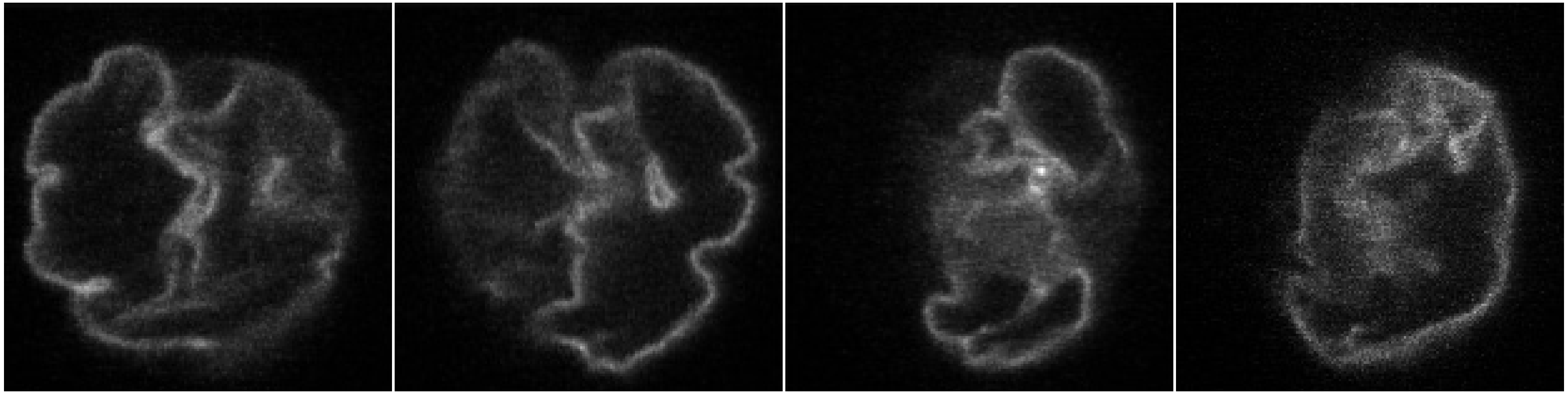
Micro-rotation imaging aims at

1) To image living cells in natural environment
2) To improve image resolution in 3D



http://www.pfid.org/AUTOMATION/

Examples of micro-rotation images



A human living cell, expressing fluorescence at nuclear envelope

Laksameethanasan et al. 2008

Step 0: Image Registration

- Imaging geometry need to be solved prior to the reconstruction problem
- Beyond the scope of this talk, but we have studied that too
  - TEM: Brandt et al. (2001a,2001b), Brandt and Ziese (2005), Brandt (2006);
  - X-ray tomography: Brandt and Kolehmainen (2007);
  - Microrotation confocal microscopy: Brandt and Mevorah (2006), Palander (2007)

Projection Model

- Assume a linear projection model

$$m_i(x,y) = A_i f(x,y,z),$$

where $A_i: C_3 \rightarrow C_2$

- Assuming a linear and shift invariant system

$$m_i(x,y) = h(x,y,z) * f_i(x,y,z)\Big|_{z=d},$$

Bayesian Reconstruction

- Consider the discretised model

$$\hat{\mathbf{m}} = \mathbf{A}\mathbf{f}$$

- The complete solution is the posterior

$$p(\mathbf{f}|\mathbf{m}) = \frac{p(\mathbf{f})p(\mathbf{m}|\mathbf{f})}{p(\mathbf{m})} \propto p(\mathbf{f})p(\mathbf{m}|\mathbf{f}),$$

where $p(\mathbf{m}|\mathbf{f})$ is the likelihood and $p(\mathbf{f})$ is the prior

Likelihood

- Obtained from the noise model
- Gaussian noise

$$p(\mathbf{m}|\mathbf{f}) \propto \exp\left(-\frac{1}{2\sigma}\|\mathbf{m}-\mathbf{Af}\|^2\right)$$

- Poisson noise (photon counting)

$$p(\mathbf{m}|\mathbf{f}) = \prod_{j=1}^{KM}\left(\frac{1}{m_j!}\right)\exp\left(\mathbf{m}^{\mathrm{T}}\log(\mathbf{Af}) - \mathbf{1}^{\mathrm{T}}\mathbf{Af}\right)$$

Sparsity Prior

- We may take the desired sparsity of the solution into account in the prior $p(\mathbf{f})$
- What choices do we have?

For instance:

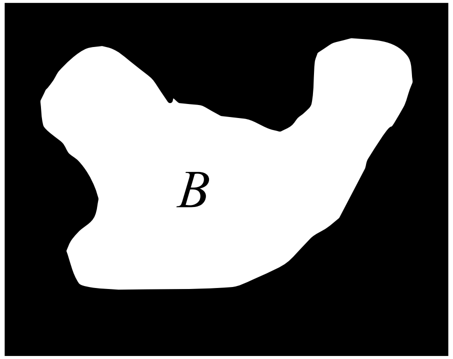
- Pseudo norm
- One norm (Lasso)
- Total variation
- Spatial derivative priors

Total variation Prior

- In the continuous case

$$p_f \propto \exp\left(-\lambda \int \left|\nabla f\right| \mathrm{d}V\right)$$

- If f is the characteristic function of the set B

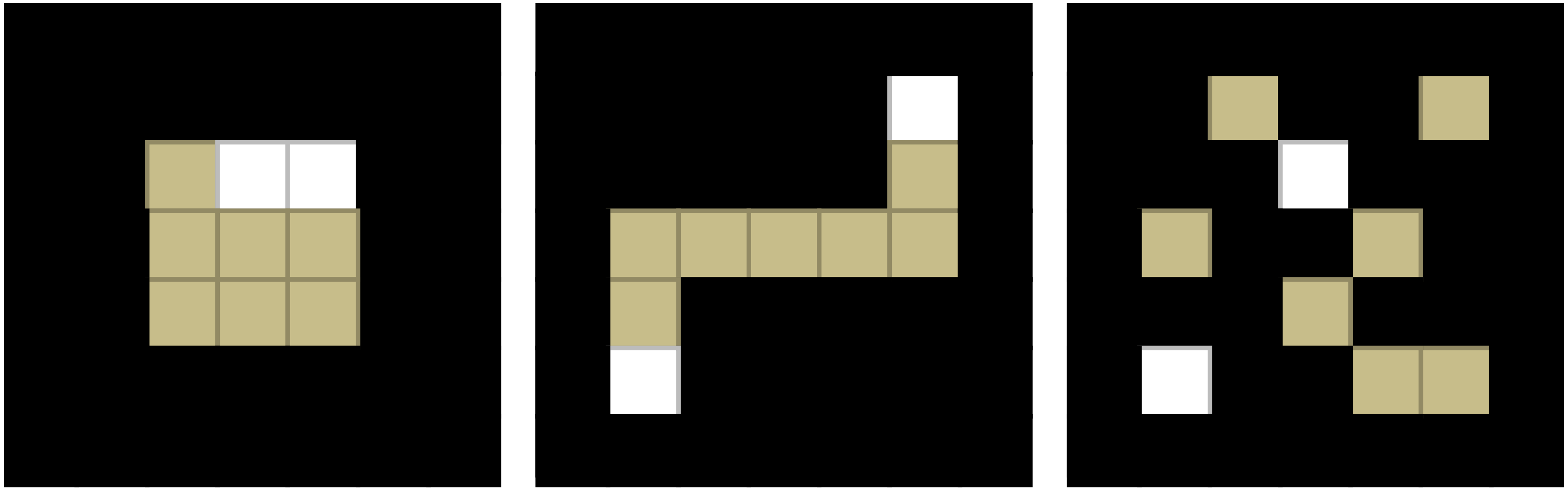


$$\mathrm{TV}(f) \equiv \int \left|\nabla f\right| \mathrm{d}V = \mathrm{length}(\partial B)$$

- Discrete definition (four neighbourhood)

$$p(\mathbf{f}) = \exp\left(-\frac{\lambda}{2} \sum_{i \in N_j} \left|f_i - f_j\right|\right)$$

Total Variation Example



TV = 18 TV = 26 TV = 42

Images with the same energy but increasing total variation

Spatial Derivative Priors

- We may use a more general class of priors

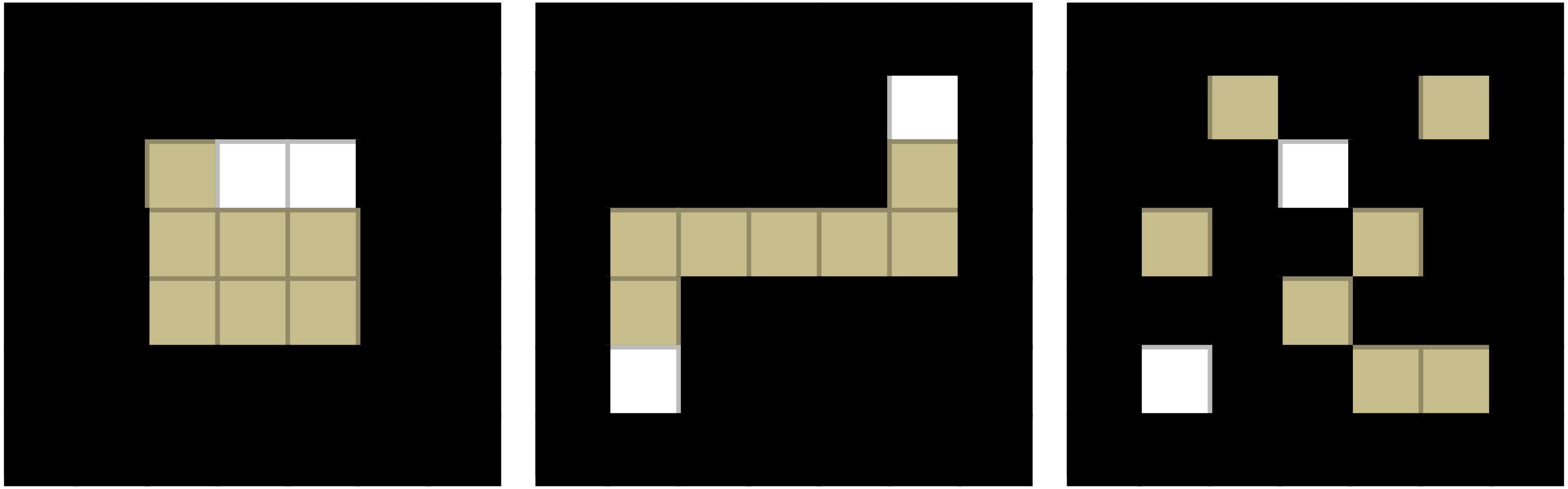
$$p_f \propto \exp\left(-\lambda \int \left| G\, f \right| \mathrm{d}V \right)$$

where $G\!: C_3 \rightarrow C_3$ is a linear operator

- We have used the Laplacian instead of gradient

$$p_f \propto \exp\left(-\lambda \int \left| \Delta f \right| \mathrm{d}V \right)$$

Discrete Laplacian Example



SD = 32   SD = 42   SD = 52

	1	
1	-4	1
	1	

Images with the same energy but increasing total absolute (discrete) Laplacian

How does this relate to sparsity of the solution?

- Consider negative log posterior

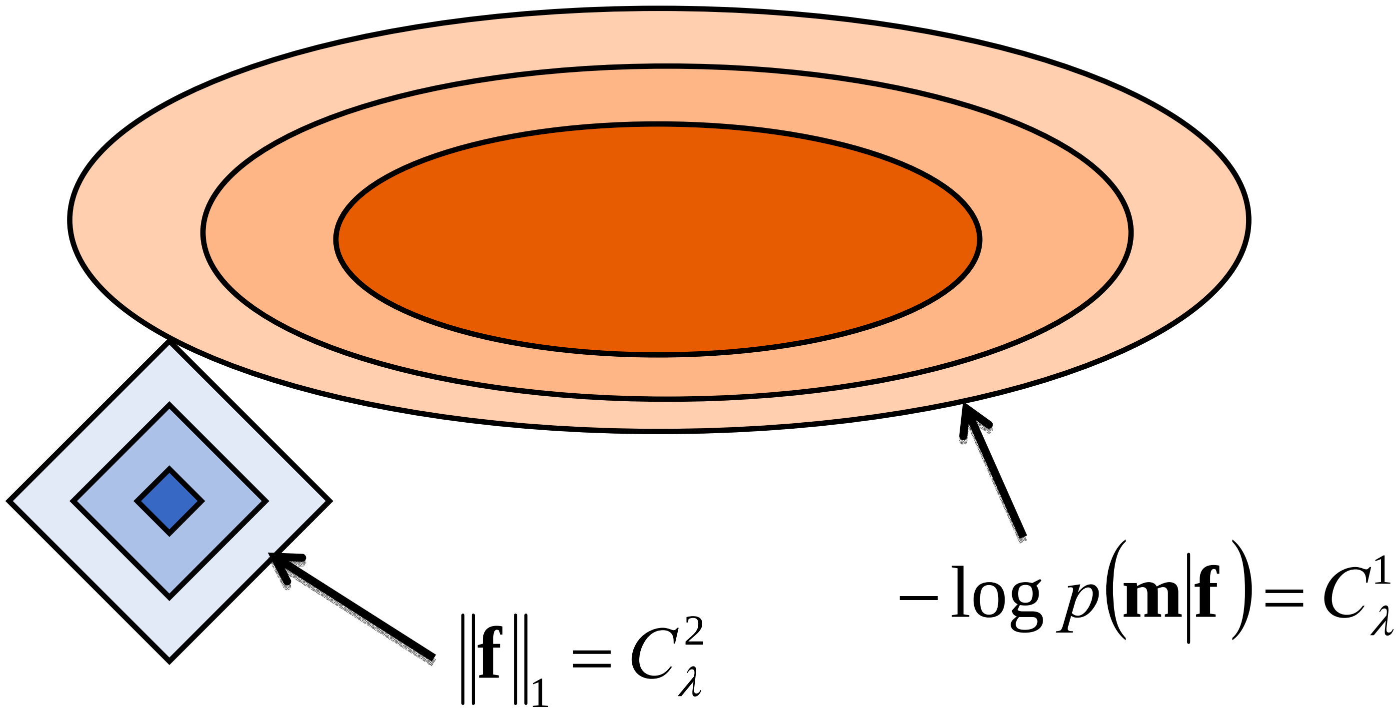
$$E(\mathbf{f}) = -\log p(\mathbf{m}|\mathbf{f}) - \lambda \log p(\mathbf{f})$$

- Computing the MAP estimate is multiobjective optimization
- The regularization parameter is chosen so that the fit (likelihood) is at desired level (Morozov discrepancy principle)
- From that subset the prior is maximized

How does this relate to sparsity of the solution?

- Consider first the one norm prior

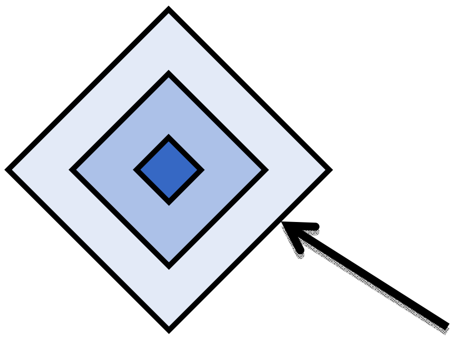
$$E(\mathbf{f}) = -\log p(\mathbf{m}|\mathbf{f}) + \lambda \|\mathbf{f}\|_1$$



How does this relate to sparsity of the solution?

- For the spatial derivative priors

$$E(\mathbf{f}) = -\log p(\mathbf{m}|\mathbf{f}) + \lambda \|\mathbf{G}\mathbf{f}\|_1$$

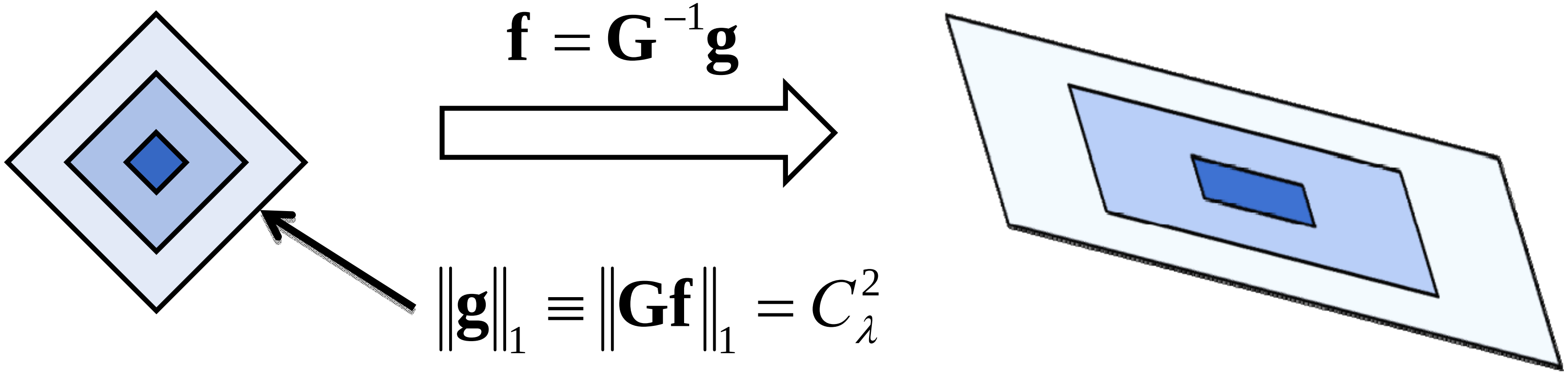


$$\|\mathbf{g}\|_1 \equiv \|\mathbf{G}\mathbf{f}\|_1 = C_\lambda^2$$

How does this relate to sparsity of the solution?

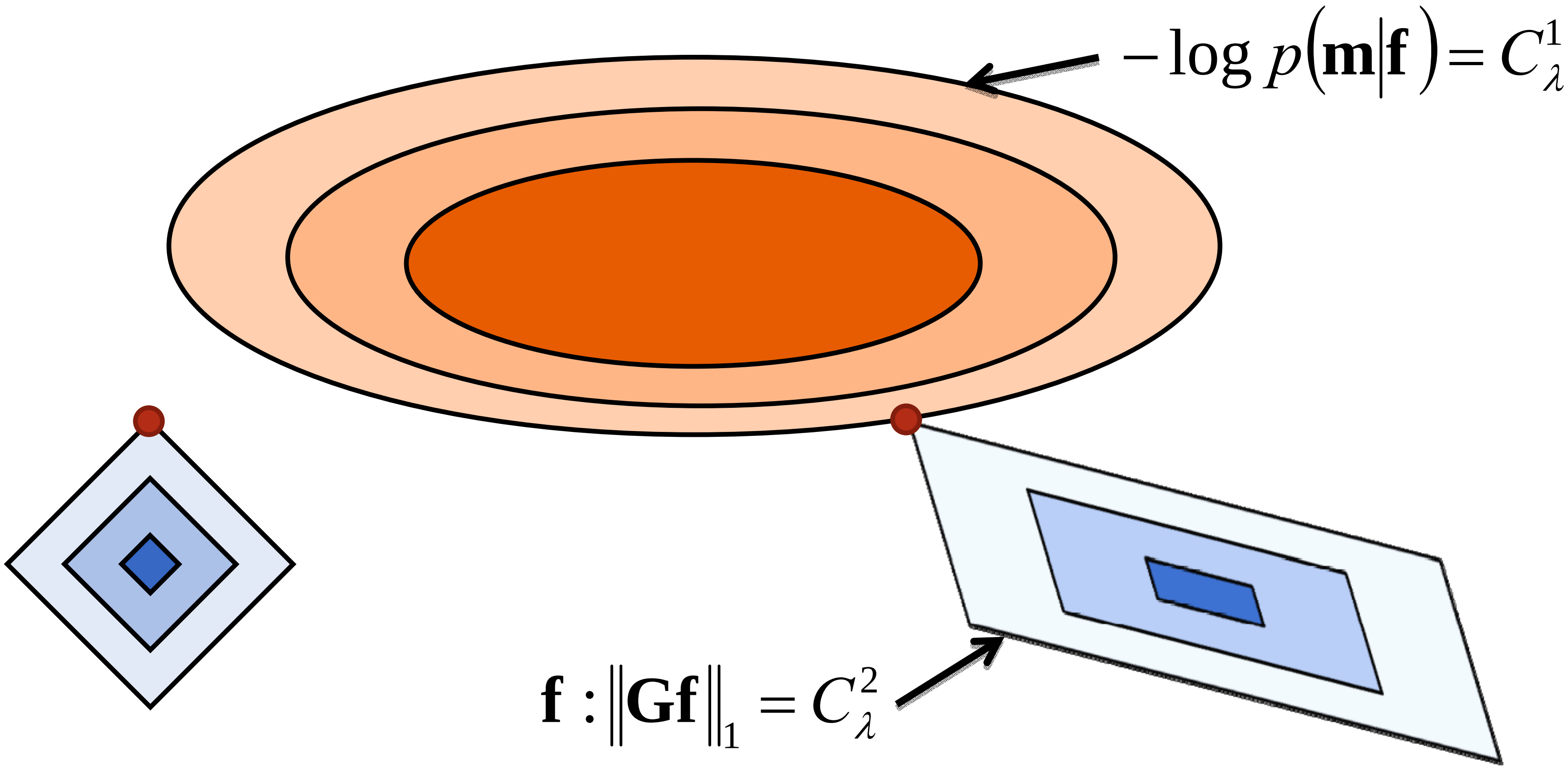
- For the spatial derivative priors

$$E(\mathbf{f}) = -\log p\left(\mathbf{m}|\mathbf{f}\right) + \lambda \|\mathbf{G}\mathbf{f}\|_1$$



How does this relate to sparsity of the solution?

The linearity preserves the vertex



How does this relate to sparsity of the solution?

- Total variation favours sparse solutions in the first derivative (edge preservation)
- Our Spatial derivative prior favours sparsity in the second derivative (edge and smoothness preservation)
- The 1-norm imposes the sparsity for a large class of linear operators

Implementation of the Prior

- The laplacian computed by convolution with the LoG filter (Gaussian interpolation)

$$\mathrm{TV}(f) = \int \left| G\, f \right| \mathrm{d}V$$

$$\approx \sum_{l} \left| \sum_{k} f_l \Delta g(\mathbf{r}_l - \mathbf{r}_k) \right| = \left\| \mathbf{Gf} \right\|_1$$

where G is the Toepliz matrix corresponding to the 3D convolution with the LoG kernel

Implementation of the Prior

- To make the energy function differentiable at zero, we approximate

$$|t| \approx \beta^{-1}\cosh(\beta t)$$

in $\|\mathbf{Gf}\|_1 \approx \|\mathbf{Gf}\|_{\tilde{1}}$

- The prior finally takes the form

$$p(\mathbf{f}) \propto \exp\left(-\lambda \|\mathbf{Gf}\|_{\tilde{1}}\right)$$

Computation of the MAP Estimate

- Poisson noise and the spatial derivative prior yields the optimization problem

$$\min_{\mathbf{f}} \left\{ \mathbf{1}^{\mathrm{T}} \left(\mathbf{Af} \right) - \mathbf{m}^{\mathrm{T}} \log \left(\mathbf{Af} \right) + \lambda \left\| \mathbf{Gf} \right\|_{\tilde{1}} \right\}$$

with subject to $f_i \geq 0 \forall i$

- Here we consider two algorithms:
  - Expectation Maximization (EM)
  - Non-linear optimization by Skilling-Bryan

EM algorithm

- Solution by the iteration (Green 1990, Dey 2006, Laksameethanasan et al. 2008)

$$\mathbf{f}_{k+1} = \frac{\mathbf{f}_k}{\mathbf{A}^{\mathrm{T}}\mathbf{1} + \lambda \nabla \left\| \mathbf{G}\mathbf{f}_k \right\|_{\tilde{1}}} \left(\mathbf{A}^{\mathrm{T}} \frac{\mathbf{m}}{\mathbf{A}\mathbf{f}_k} \right)$$

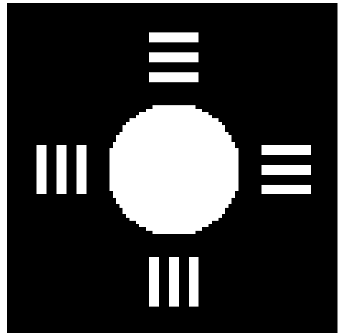
where

$$\nabla \left\| \mathbf{G}\mathbf{f}_k \right\|_{\tilde{1}} = \mathbf{G}^{\mathrm{T}} \tanh\left(\beta \mathbf{G}\mathbf{f}_k \right)$$

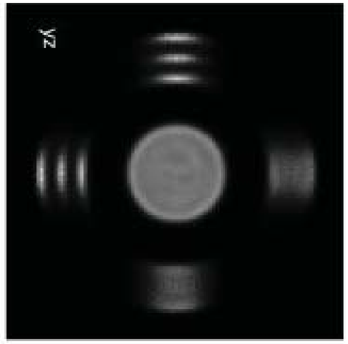
- Note: the matrices for forward projection

$\mathbf{A}$ and its adjoint $\mathbf{A}^{\mathrm{T}}$ are *not* computed

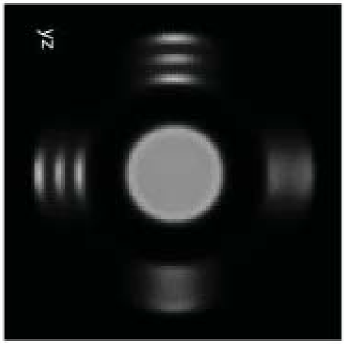
Toy Reconstruction Example



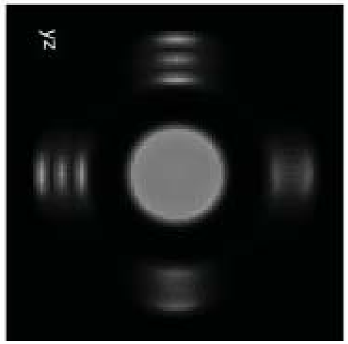
ML

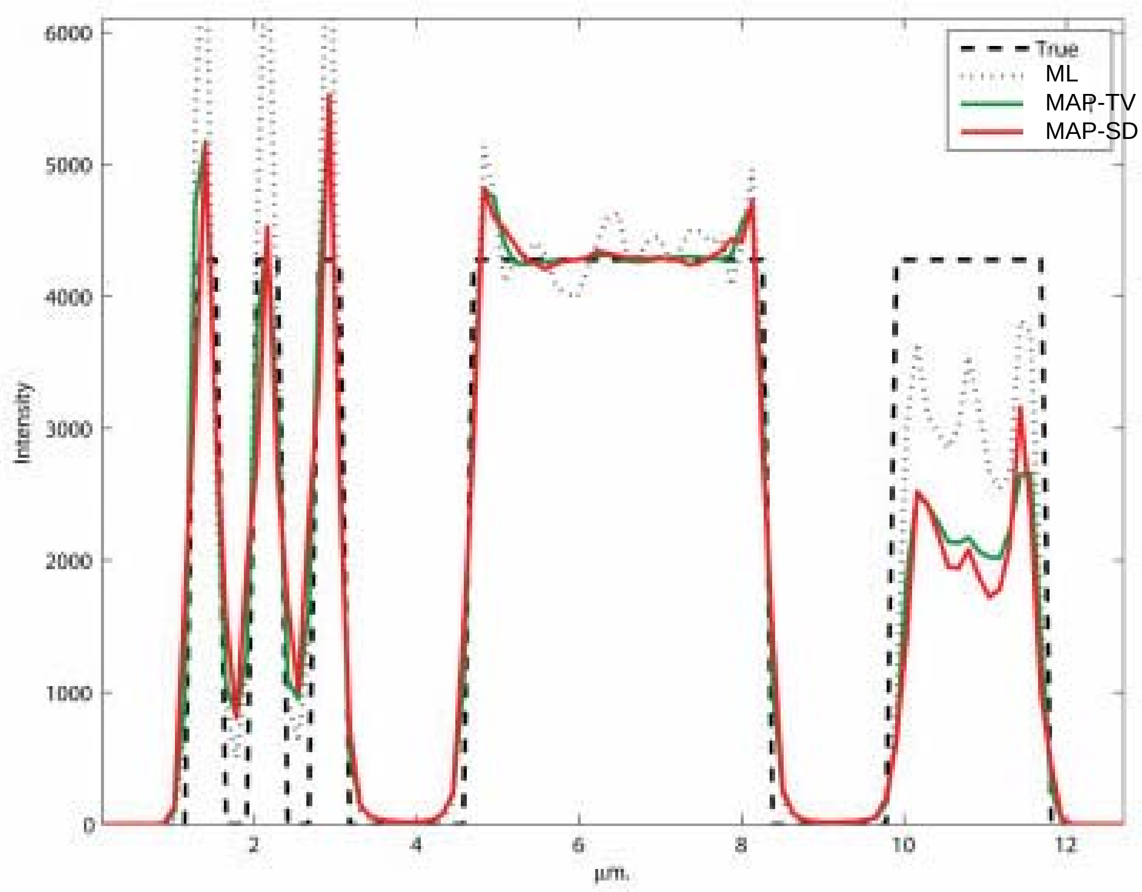


MAP-TV



MAP-SD





(d) Intensity profile

Generalized Skilling-Bryan method

- A 2nd order, non-linear optimization algorithm to compute the MAP estimate

$$\arg\max_{\mathbf{f}} p(\mathbf{f}|\mathbf{m}), \qquad f_i \geq 0 \forall i$$

which is equivalent to maximizing

$$q(\mathbf{f}) = s(\mathbf{f}) - \lambda c(\mathbf{f}), \quad f_i \geq 0 \forall i$$

where $s(\mathbf{f}) = \log p(\mathbf{f}), \; c(\mathbf{f}) = -\log p(\mathbf{m}|\mathbf{f})$

and λ is the regularisation parameter

TABLE 1. Likelihood and prior terms with their gradients and Hessians used in the extended Skilling–Bryan minimisation method. The operations between two vectors are performed element by element.

	Functions	Gradient $\mathbf{g}$	Hessian $\mathbf{H}$
Gaussian noise c_G	$\frac{1}{2}\|\mathbf{m}-\mathbf{Af}\|^2$	$-\mathbf{A}^T(\mathbf{m}-\mathbf{Af})$	$\mathbf{A}^T\mathbf{A}$
Poisson noise c_P	$\mathbf{1}^T(\mathbf{Af})-\mathbf{m}^T\log(\mathbf{Af})$	$-\mathbf{A}^T(\frac{\mathbf{m}}{\mathbf{Af}}-\mathbf{1})$	$\mathbf{A}^T\text{diag}\left(\frac{\mathbf{m}}{(\mathbf{Af})(\mathbf{Af})}\right)\mathbf{A}$
Guassian prior s_G	$-\frac{1}{2}\|\mathbf{f}\|^2$	$-\mathbf{f}$	$-\mathbf{I}$
Entropy prior s_E	$\mathbf{1}^T\mathbf{f}-\mathbf{f}^T\log(\mathbf{f}/\mathbf{f}_0)$	$-\log(\mathbf{f}/\mathbf{f}_0)$	$-\text{diag}(\mathbf{1}/\mathbf{f})$
TV prior s_T	$-\mathbf{1}^T\beta^{-1}\log(\cosh(\beta\mathbf{Gf}))$	$-\mathbf{G}^T\tanh(\beta\mathbf{Gf})$	$-\mathbf{G}^T\text{diag}\left(\beta\,\text{sech}^2(\beta\mathbf{Gf})\right)\mathbf{G}$

Skilling-Bryan Algorithm

- The objective function is approximated by the 2nd order Taylor expansion

$$q(\mathbf{f}+\mathbf{p}) \approx q(\mathbf{f}) + \mathbf{g}_q^{\mathrm{T}}\mathbf{p} + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}_q\mathbf{p}$$

- To allow high-dimensional search, the maximization is done in a subspace.

Subspace Selection

- The step is solved up the 2nd order:

$$
\begin{aligned}
\mathbf{p} &= -(\mathbf{H}_q + \gamma \mathbf{I})^{-1}\mathbf{g}_q, \\
&\approx -(\mathbf{I} + \gamma^{-1}\mathbf{H}_q)\mathbf{g}_q, \\
&\approx -\mathbf{g}_s + \lambda \mathbf{g}_c + \gamma^{-1}[(\mathbf{H}_s - \lambda \mathbf{H}_c)(\mathbf{g}_s - \lambda \mathbf{g}_c)]
\end{aligned}
$$

- The step lies in the subspace spanned by

$$\mathbf{g}_c, \mathbf{g}_s, \mathbf{H}_c\mathbf{g}_c, \mathbf{H}_s\mathbf{g}_s, \mathbf{H}_s\mathbf{g}_c \text{ and } \mathbf{H}_c\mathbf{g}_s$$

Subspace Selection

- The basis is summarised by

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{f}\mathbf{g}_s, \\ \mathbf{e}_2 &= \mathbf{f}\mathbf{g}_c, \\ \mathbf{e}_3 &= \mathbf{f}\mathbf{H}_c\left(\frac{\mathbf{e}_1}{||\mathbf{g}_s||} - \frac{\mathbf{e}_2}{||\mathbf{g}_c||}\right) + \mathbf{f}\mathbf{H}_s\left(\frac{\mathbf{e}_1}{||\mathbf{g}_s||} - \frac{\mathbf{e}_2}{||\mathbf{g}_c||}\right) \end{aligned}$$

- The gradient directions are weighted by $\mathbf{f}$ to increase the weight for high values (positivity constraint)

Some final details

- The step is solved in the subspace

$$\mathbf{p} = \mathbf{E}\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3,$$

under the constraint $c < c_{\text{aim}}$

- The new iterate is obtained as

$$\mathbf{f}_{\text{new}} = \mathbf{f} + \mathbf{E}\mathbf{x},$$

whilst it needs to be protected against negative values (optional rescaling)

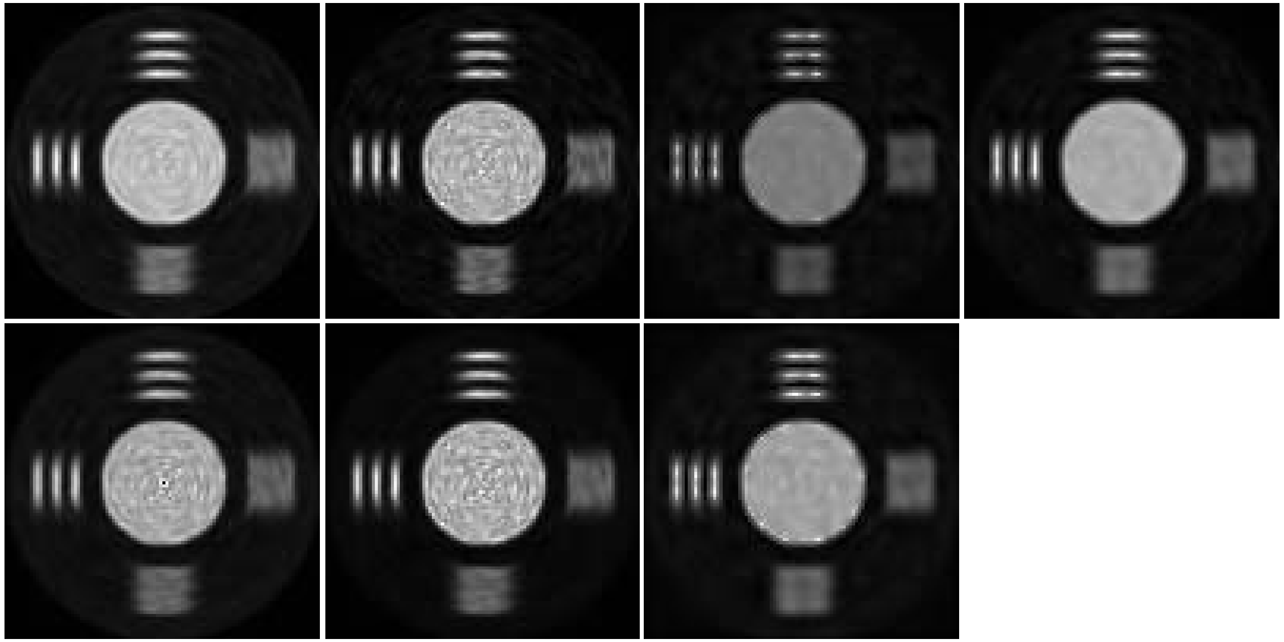
Properties of the Skilling-Bryan Algorithm

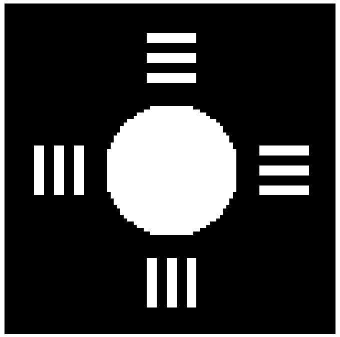
- Resembles Levenberg-Marguardt Method
- Inherent positivity constraint – thus natural for tomographic reconstruction
- Converges faster than 1st order methods based on line-search
- Each iteration evaluates six projection-backprojection pairs (computational bottleneck)

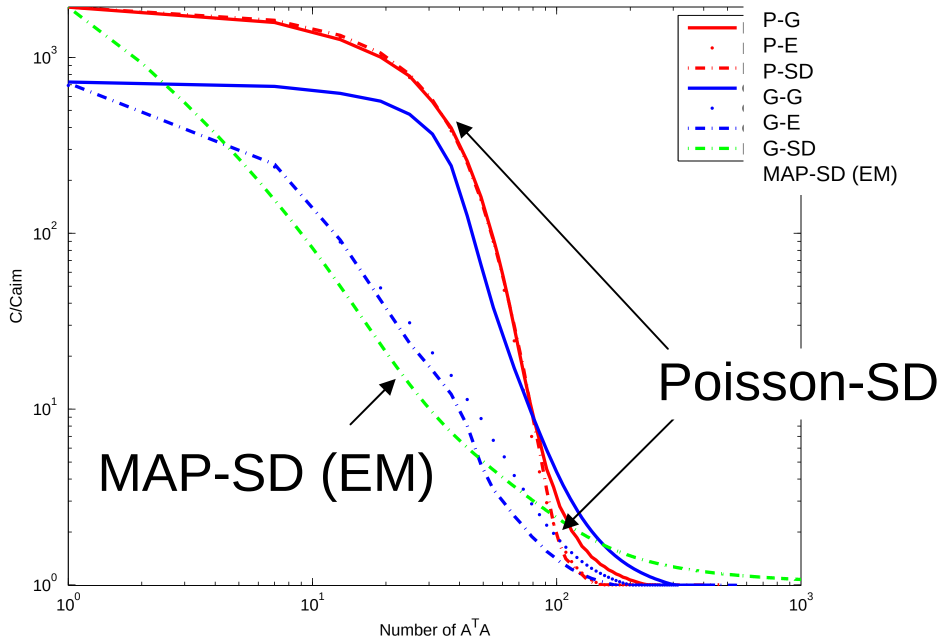
G E SD MAP-SD (EM)

Poisson

Gaussian







Conclusions

- We have studied the Bayesian approach for volumetric reconstruction problems
- TV and Spatial Derivative Priors impose sparse solutions in the derivatives
- Extended Skilling-Bryan Optimization
  - Especially for convex objective functions with the positivity constraint
  - Suitable for very high dimensional problems
  - Fast convergence (≈ 20 iterations)

Acknowledgements

- Danai Laksameethanasan, Olivier Renaud, Spencer Shorte
- AUTOMATION project (EU sixth framework programme)

http://www.pfid.org/AUTOMATION/

References

Brandt, S. et al. (2001a), J. Struct. Biol. **133**: 10-12

Brandt, S. et al. (2001b), J. Struct. Biol. **136**: 201-213.

Brandt, S. and Ziese U. (2006), J. Microsc. **222**: 1-14

Brandt, S.S. (2006), in Frank J. (editor), Electron Tomography (2nd ed.)

Brandt, S.S. and Kolehmainen, V. (2007), IEEE Trans. Med. Imaging **26**:238-248;

Brandt, S.S. and Mevorah, M. (2006), Proc. SMVP

Dey, N. et al. (2006), Microsc. Res. Tech. **69**: 260-266

Green, P.J. 1990, IEEE Trans. Med. Imaging **9**:84-93

Laksameethanasan, D. and Brandt, S.S. (2007). Proc. AIP Conf. (MaxEnt) 954, 354-361

Laksameethanasan, D., et al (2008). Microscopy Research Technique, 72(2) 158-167

Palander, K. (2007), Master's Thesis, Helsinki Univ. Tech.