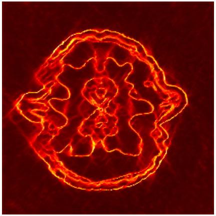
Sample-based Sparse Bayesian Inversion in Biomedical Imaging

Felix Lucka
University College London, UK
f.lucka@ucl.ac.uk



Centre for Medical Image Computing

ICIAM
Beijing, August 11, 2015

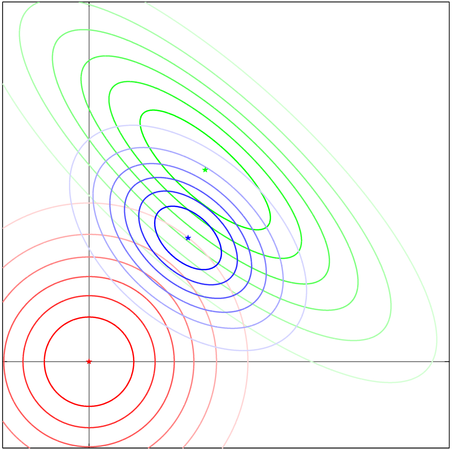
Ill-posed inverse problems with additive Gaussian noise:

$$f = \mathcal{A}(u) + \varepsilon$$

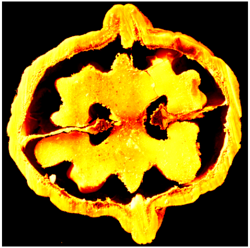
$p_{like}(f|u) \propto$
$\exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}}\right)$

$p_{prior}(u) \propto$
$\exp\left(-\lambda\,\|D^T u\|_2^2\right)$

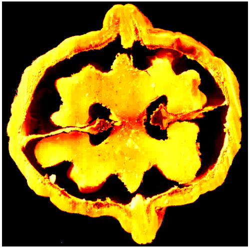
$p_{post}(u|f) \propto$
$\exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_2^2\right)$



Probabilistic representation allows for a rigorous quantification of the solution's uncertainties.



(a) 100%



(b) 10%

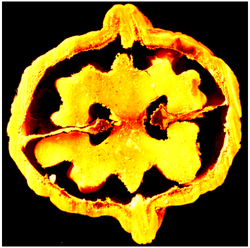
(c) 1%

Sparsity a-priori constraints are used in variational regularization, compressed sensing and ridge regression:

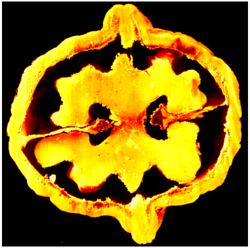
$$\hat{u}_\lambda \quad = \underset{u}{\operatorname{argmin}} \left\{ \tfrac{1}{2} \| f - A\, u \|_2^2 + \lambda \| D^T u \|_1 \right\}$$

(e.g. *total variation*, *wavelet shrinkage*, *LASSO*,...)

Sparsity / Compressible Representation



(a) 100%



(b) 10%

(c) 1%

Sparsity a-priori constraints are used in variational regularization, compressed sensing and ridge regression:

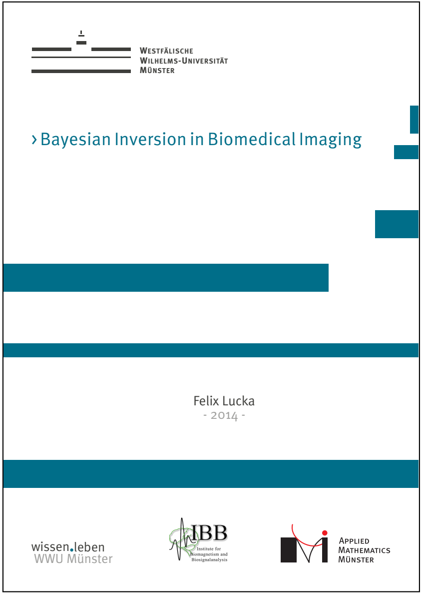
$$\hat{u}_\lambda \quad = \underset{u}{\operatorname{argmin}} \left\{ \tfrac{1}{2} \| f - A\, u \|_2^2 + \lambda \| D^T u \|_1 \right\}$$

(e.g. *total variation*, *wavelet shrinkage*, *LASSO*,...)

How about sparsity as a-priori information in the Bayesian approach?

PhD Thesis "Bayesian Inversion in Biomedical Imaging"

- Submitted 2014, supervised by Martin Burger and Carsten H. Wolters.
- Linear inverse problems in biomedical imaging applications.
- Simulated data scenarios and experimental CT and EEG/MEG data.
- Sparsity by means of
  - ℓ_p-norm based priors
  - Hierarchical prior modeling
- Focus on Bayesian computation and application.


Westfälische Wilhelms-Universität Münster

› Bayesian Inversion in Biomedical Imaging

Felix Lucka
- 2014 -

wissen.leben WWU Münster

IBB Institute for Biomagnetism and Biosignalanalysis

Applied Mathematics Münster

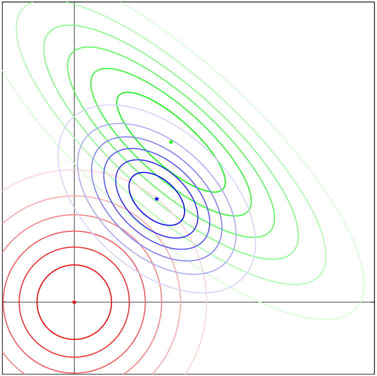
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda \, \|D^T u\|_p^p\right)$$

Decrease p from 2 to 0 and stop at $p = 1$ for convenience.

The ℓ_p Approach to Sparse Bayesian Inversion

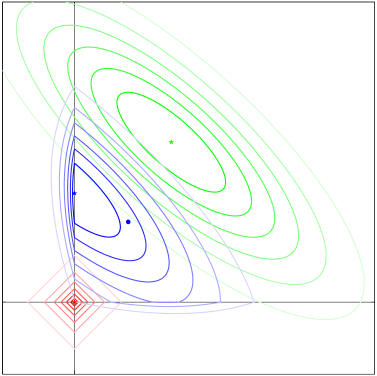
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_p^p\right)$$

Decrease p from 2 to 0 and stop at $p = 1$ for convenience.



$$\exp\left(-\lambda\,\|D^T u\|_2^2\right)$$

$$\exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_2^2\right)$$

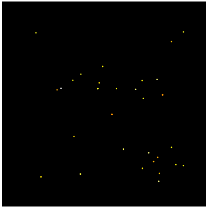


$$\exp\left(-\lambda\,\|D^T u\|_1\right)$$

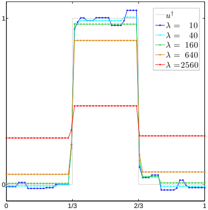
$$\exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_1\right)$$

Bayesian Inference with ℓ_1 Priors

$$p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_1\right)$$



Aims: Bayesian inversion in high dimensions ($n \to \infty$).

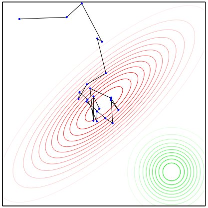


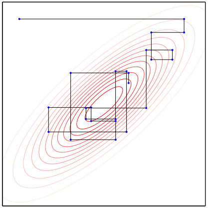
Priors: Simple ℓ_1, total variation (TV), Besov space priors.

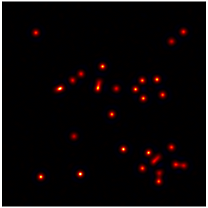
Starting points:

- **Lassas & Siltanen, 2004.** *Can one use total variation prior for edge-preserving Bayesian inversion?* *Inverse Problems*, 20.
- **Lassas, Saksman & Siltanen, 2009.** *Discretization invariant Bayesian inversion and Besov space priors.* *Inverse Problems and Imaging*, 3(1).
- **Kolehmainen, Lassas, Niinimäki & Siltanen, 2012.** *Sparsity-promoting Bayesian inversion.* *Inverse Problems*, 28(2).

Efficient MCMC Techniques for ℓ_1 Priors







Task: Monte Carlo integration by samples from

$$p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_1\right)$$

Problem: Standard Markov chain Monte Carlo (MCMC) sampler (Metropolis-Hastings) inefficient for large n or λ.

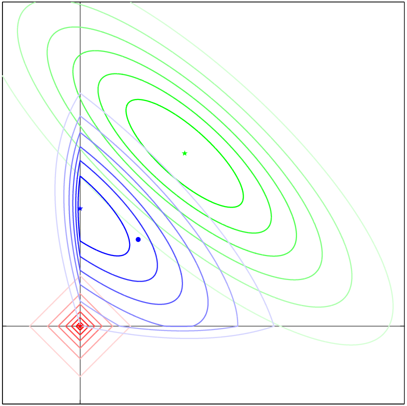
Contributions:

- Development of explicit single component Gibbs sampler.
- Tedious implementation for different scenarios.
- Still efficient in high dimensions ($n > 10^6$).
- Detailed evaluation and comparison to MH.

L, 2012. *Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.* Inverse Problems, 28(12):125012.

$$\hat{u}_{\mathrm{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ \, p_{post}(u|f) \} \quad \text{vs.} \quad \hat{u}_{\mathrm{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

- CM preferred in theory, dismissed in practice.
- MAP discredited by theory, chosen in practice.



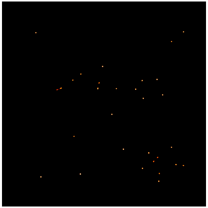
New Theoretical Ideas for an Old Bayesian Debate

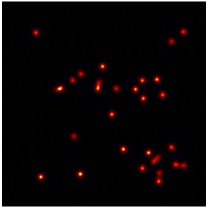
$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ p_{post}(u|f) \} \quad \text{vs.} \quad \hat{u}_{\text{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

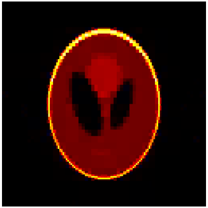
- CM preferred in theory, dismissed in practice.
- MAP discredited by theory, chosen in practice.

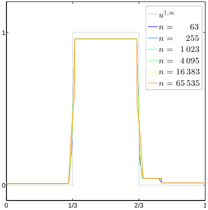
However:

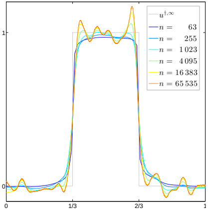
- MAP results looks/performs better or similar to CM.
- Gaussian priors: MAP = CM. Funny coincidence?
- Theoretical argument has a logical flaw.









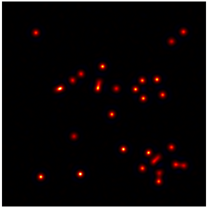


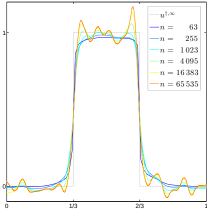
$$\hat{u}_{\mathrm{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ p_{post}(u|f) \} \quad \text{vs.} \quad \hat{u}_{\mathrm{CM}} := \int u \, p_{post}(u|f) \, \mathrm{d}u$$

- CM preferred in theory, dismissed in practice.
- MAP discredited by theory, chosen in practice.

Contributions:

- Theoretical rehabilitation of MAP.
- Key: Bayes cost functions based on Bregman distances.
- Gaussian case consistent in this framework.

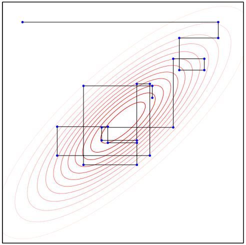




Burger & L, 2014. *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*, *Inverse Problems*, 30(11):114004.

Helin & Burger, 2015. *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems*, *Inverse Problems*, 31(8)

$$p_{prior}(u) \propto \exp\left(-\lambda \|D^T u\|_1\right)$$

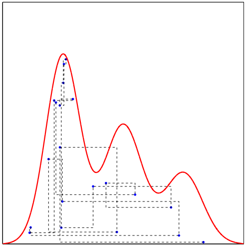


Limitations:

- D must be diagonalizable (synthesis priors):
- ℓ_p^q-prior: $\exp\left(-\lambda \|D^T u\|_p^q\right)$? TV in 2D/3D?
- Non-negativity or other hard-constraints?

Contributions:

- Replace explicit by generalized slice sampling.
- Implementation & evaluation for most common priors.

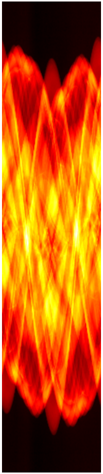


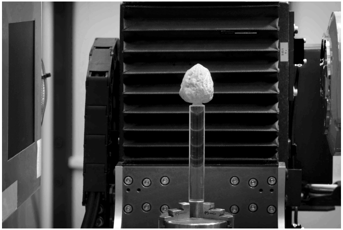
Neal, 2003. *Slice Sampling.* Annals of Statistics 31(3)

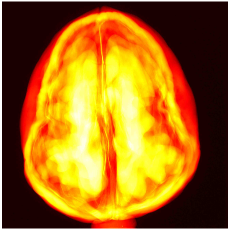
L, 2015. *Fast Gibbs sampling for high-dimensional Bayesian inversion.* (in preparation)

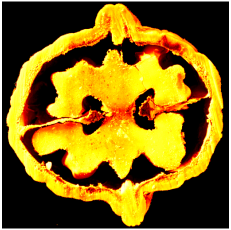
Application to Experimental Data: Walnut-CT

- Cooperation with Samuli Siltanen, Esa Niemi et al.
- Implementation of MCMC methods for Fanbeam-CT.
- Besov and TV prior; non-negativity constraints.
- Stochastic noise modeling.
- Bayesian perspective on limited angle CT.





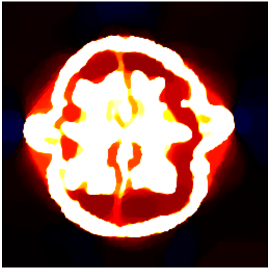




Use the data set for your own work:
http://www.fips.fi/dataset.php (documentation: arXiv:1502.04064)

Walnut-CT with TV Prior: Full Angle

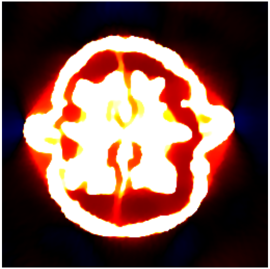
(a) MAP



(b) MAP, special color scale

(c) CStd

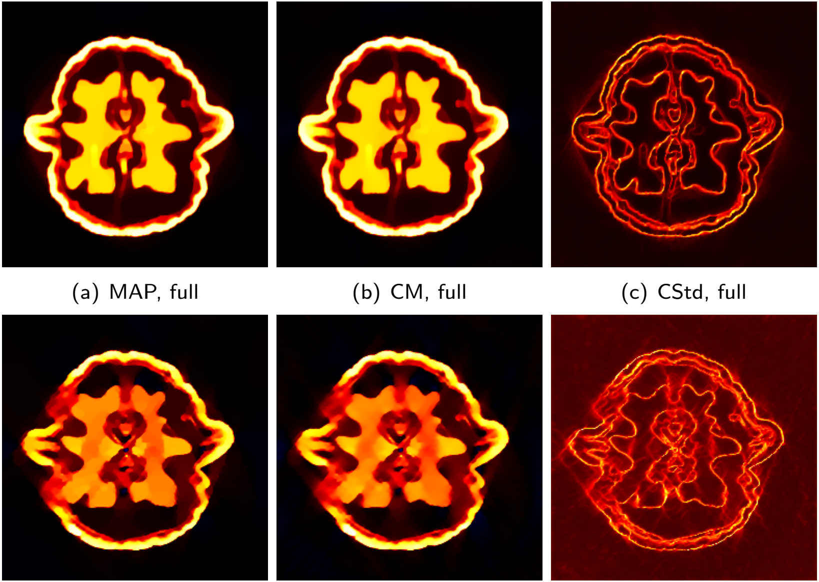
(d) CM



(e) CM, special color scale

(f) CM of $\|\nabla u\|_2$

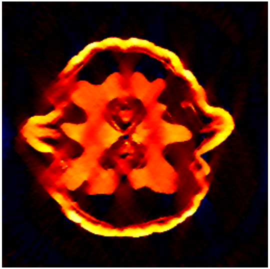
Walnut-CT with TV Prior: Full vs. Limited Angle



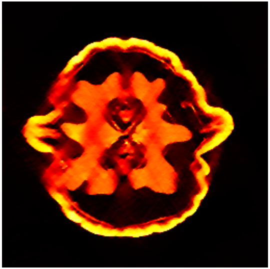
(a) MAP, full (b) CM, full (c) CStd, full

(d) MAP, limited (e) CM, limited (f) CStd, limited

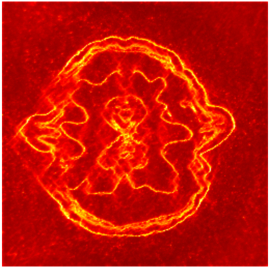
Walnut-CT with TV Prior: Non-Negativity Constraints, Limited Angle



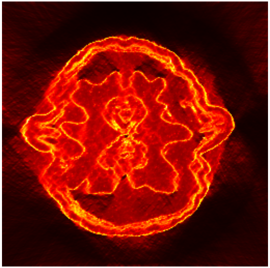
(a) CM, uncon



(b) CM, non-neg



(c) CStd, uncon



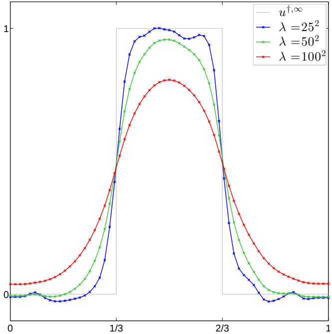
(d) CStd, non-neg

Outline

Gaussian increment prior:

$$p_{prior}(u) \propto \prod_i \exp\left(-\frac{(u_{i+1} - u_i)^2}{\gamma}\right)$$

- Gaussian variables take values on a characteristic scale, determined by γ.
- Similar amplitudes are likely, sparsity (= outliers) is unlikely.

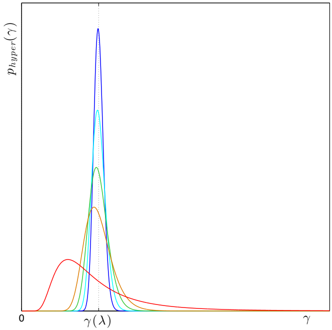


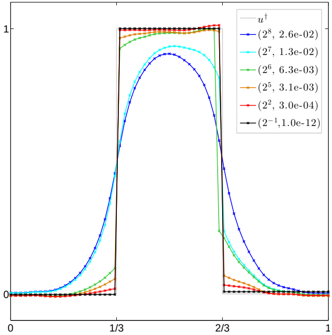
Conditionally Gaussian increment prior:

$$p_{prior}(u|\gamma) \propto \prod_i \exp\left(-\frac{(u_{i+1}-u_i)^2}{\gamma_i}\right)$$

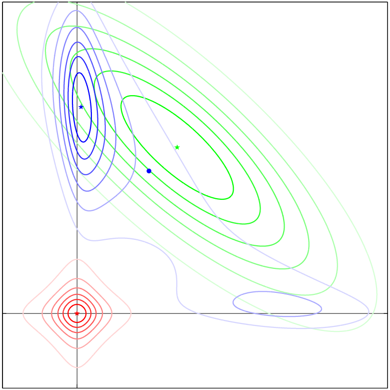
Scale-invariant hyperprior to approximate un-informative γ_i^{-1} prior:

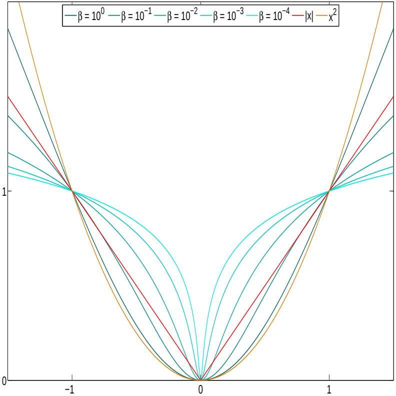
$$p_{hyper}(\gamma_i) \propto \gamma_i^{-(\alpha+1)} \exp\left(-\frac{\beta}{\gamma_i}\right), \quad \textit{inverse gamma distribution}$$





The Implicit Energy Functional behind HBM





Implicit prior is a Student's t-prior with $\nu = 2\alpha, \theta = \beta/(2\alpha)$:

$$p_{prior}(u) \propto \prod_i \left(1 + \frac{u_i^2}{\nu\theta}\right)^{-\frac{\nu-1}{2}}$$

$$p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \frac{\nu-1}{2}\sum_i \log\left(1 + \frac{u_i^2}{\nu\theta}\right)\right)$$

$$p_{post}(u, \gamma | f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \sum_i^n \left(\frac{u_i^2 + 2\beta}{2\gamma_i} + (\alpha + 1/2)\log(\gamma_i)\right)\right)$$

All computational approaches (optimization or sampling) exploit the conditional structure:

- Fix γ and update u by solving n-dim linear problem.
- Fix u and update γ by solving n 1-dim non-linear problems.

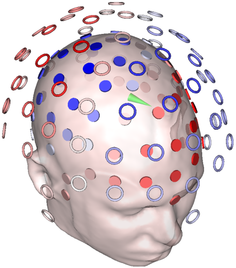
Major difficulty: Multimodality of posterior.

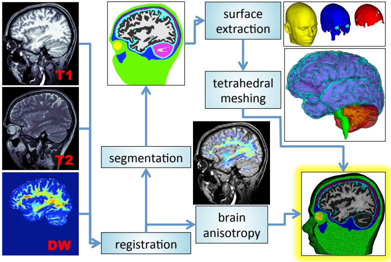
Heuristic Full-MAP computation:

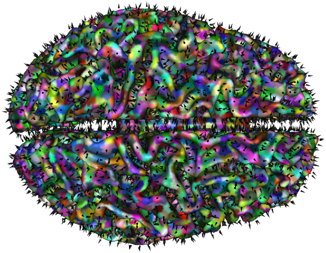
- Use MCMC to explore posterior (avoids very sub-optimal local modes).
- Initialize alternating optimization by local MCMC averages to compute local modes.

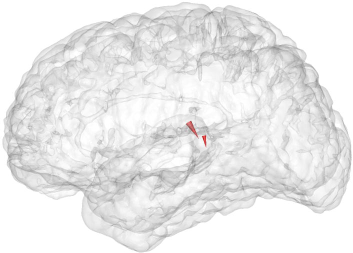
No guarantee for finding highest mode but usually an acceptable result.

Why HBM? EEG/MEG Source Reconstruction









Notoriously ill-posed problem!

Why HBM? EEG/MEG Source Reconstruction

- Inversion with log-concave priors suffers from systematic depth miss-localization, HBM does not.
- HBM shows promising results for focal brain networks with simulated and real data.

L., Aydin, Vorwerk, Burger, Wolters, 2013. *Hierarchical Fully-Bayesian Inference for Combined EEG/MEG Source Analysis of Evoked Responses: From Simulations to Real Data.*
BaCI 2013, Geneva.

L., Pursiainen, Burger, Wolters, 2012. *Hierarchical Fully-Bayesian Inference for EEG/MEG combination: Examination of Depth Localization and Source Separation using Realistic FE Head Models.*
Biomag 2012, Paris

L., Pursiainen, Burger, Wolters, 2012. *Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: Depth localization and source separation for focal primary currents.*
NeuroImage, 61(4):1364–1382.

Bayesian Modeling:

- Sparsity can be modeled in different ways.
- HBM is an interesting but challenging alternative to ℓ_p priors.
- Combine ℓ_p-type and hierarchical priors: ℓ_p-hypermodels.

Bayesian Computation:

- Elementary MCMC samplers may perform very differently.
- Contrary to common beliefs sample-based Bayesian inversion in high dimensions ($n > 10^6$) is feasible if tailored samplers are developed.
- Fast samplers can be used for simulated annealing.
- Reason for the efficiency of the Gibbs samplers is unclear.
- Adaptation, parallelization, multimodality, multi-grid.

Bayesian Estimation / Uncertainty Quantification

- MAP estimates are proper Bayes estimators.
- But: Everything beyond "MAP or CM?" is far more interesting and can really complement variational approaches.
- However: Extracting information from posterior samples (*data mining*) is a non-trivial (future research) topic.
- Application studies had proof-of-concept character up to now.
- Specific UQ task to explore full potential of the Bayesian approach.

L., 2014. *Bayesian Inversion in Biomedical Imaging*
PhD Thesis, University of Münster.

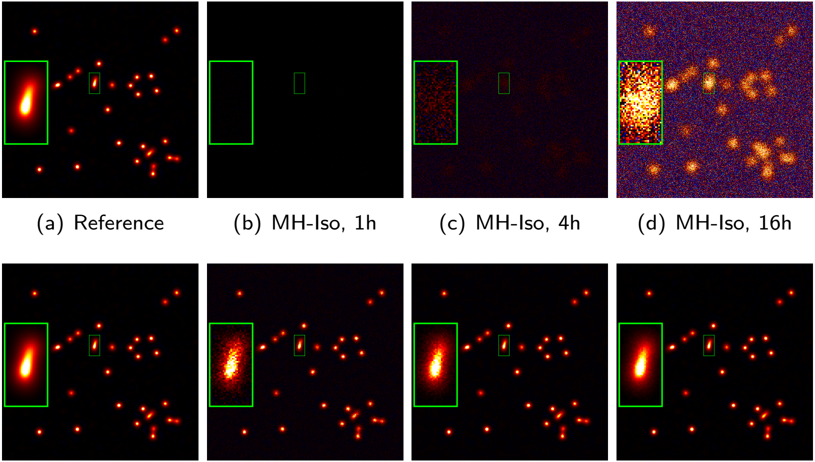
M. Burger, L., 2014. *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*
Inverse Problems, 30(11):114004.

L., 2012. *Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.*
Inverse Problems, 28(12):125012.

L., Pursiainen, Burger, Wolters, 2012. *Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: Depth localization and source separation for focal primary currents.*
NeuroImage, 61(4):1364–1382.

Thank you for your attention!

- **L., 2014.** *Bayesian Inversion in Biomedical Imaging*
  PhD Thesis, University of Münster.
- **M. Burger, L., 2014.** *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*
  Inverse Problems, 30(11):114004.
- **L., 2012.** *Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.*
  Inverse Problems, 28(12):125012.
- **L., Pursiainen, Burger, Wolters, 2012.** *Hierarchical Bayesian inference for the EEG inverse problem using realistic FE head models: Depth localization and source separation for focal primary currents.*
  NeuroImage, 61(4):1364–1382.



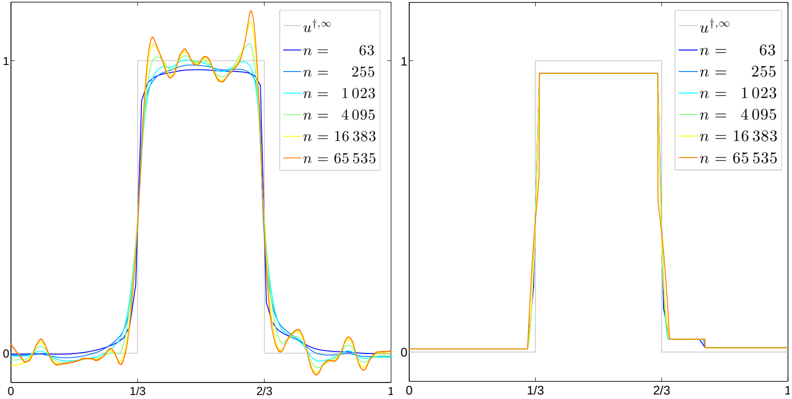
(a) Reference (b) MH-Iso, 1h (c) MH-Iso, 4h (d) MH-Iso, 16h

(e) Reference (f) SC Gibbs, 1h (g) SC Gibbs, 4h (h) SC Gibbs, 16h

Deconvolution, simple ℓ_1 prior, $n = 513 \times 513 = 263\,169$.

Numerical verification of the discretization dilemma of the TV prior (Lassas & Siltanen, 2004):

- For $\lambda_n = const.$, $n \longrightarrow \infty$ the TV prior diverges.
- CM diverges.
- MAP converges to edge-preserving limit.

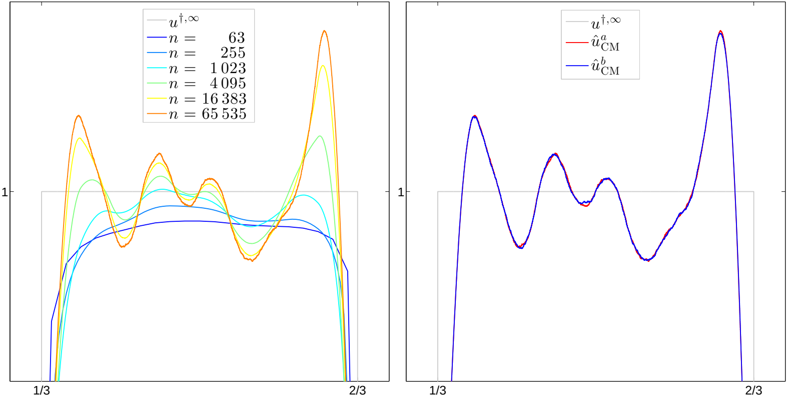


(a) CM by our Gibbs Sampler

(b) MAP by ADMM

Numerical verification of the discretization dilemma of the TV prior (Lassas & Siltanen, 2004):

- For $\lambda_n = const.$, $n \longrightarrow \infty$ the TV prior diverges.
- CM diverges.
- MAP converges to edge-preserving limit.

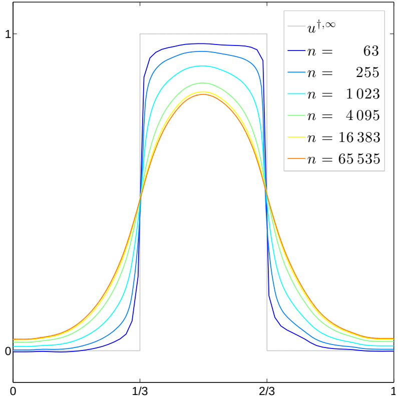


(a) Zoom into CM estimates (b) MCMC convergence check

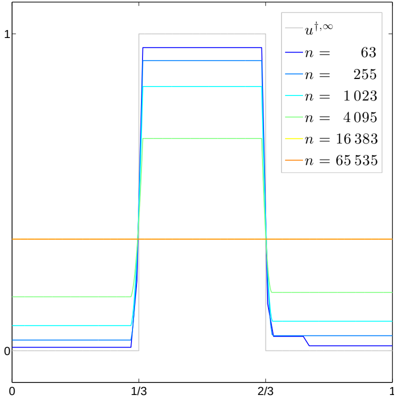
Verification of Theoretical Predictions by MCMC

Numerical verification of the discretization dilemma of the TV prior (Lassas & Siltanen, 2004):

- For $\lambda_n \propto \sqrt{n+1}$, $n \longrightarrow \infty$ the TV prior converges to a smoothness prior.
- CM converges to smooth limit.
- MAP converges to constant.



(a) CM by our Gibbs Sampler



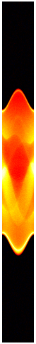
(b) MAP by ADMM

For images dimensions > 1: No theory yet...but we can compute it.

Test scenario:

- CT using only 45 projection angles and 500 measurement pixel.

real solution



data f

colormap

For images dimensions > 1: No theory yet...but we can compute it.

MAP, $n = 64^2$, $\lambda = 500$

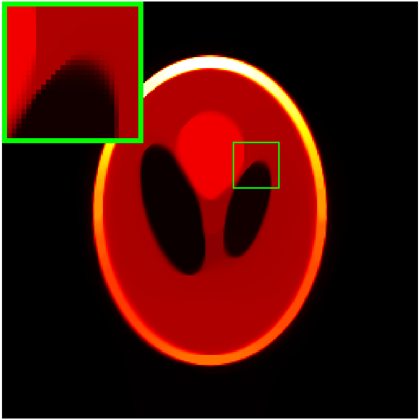
CM, $n = 64^2$, $\lambda = 500$

For images dimensions > 1: No theory yet...but we can compute it.

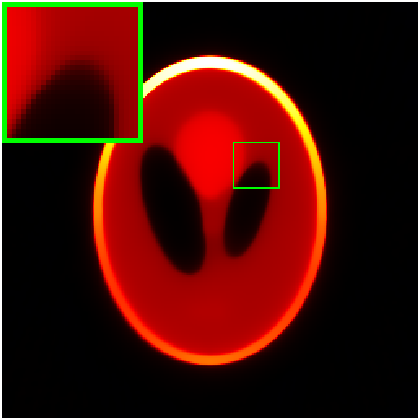
MAP, $n = 128^2$, $\lambda = 500$

CM, $n = 128^2$, $\lambda = 500$

For images dimensions > 1: No theory yet...but we can compute it.



MAP, $n = 256^2$, $\lambda = 500$

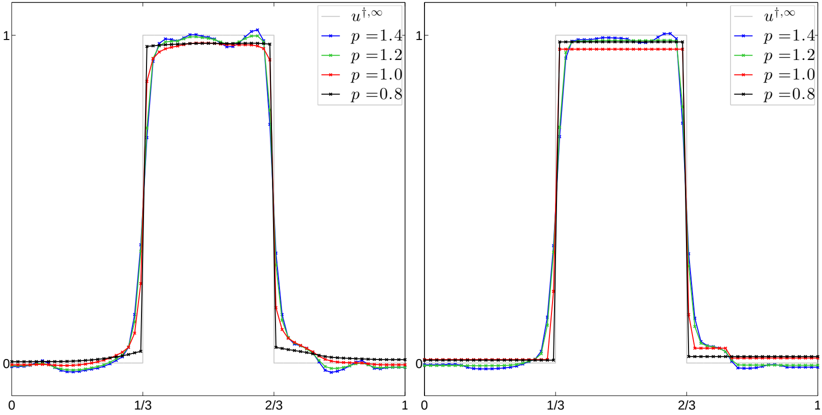


CM, $n = 256^2$, $\lambda = 500$

cf. Louchet, 2008, Louchet & Moisan, 2013 for the denoising case, $A = I$.

Examination of Alternative Priors by MCMC: TV-p

$$p_{post}(u) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_p^p\right)$$



(c) CM (Gibbs-MCMC) (d) MAP (Simulated Annealing)

Examination of Besov Space Priors by MCMC

An ℓ_1-type, wavelet-based prior: $$p_{prior}(u) \propto \exp\left(-\lambda \|WV^T u\|_1\right)$$

motivated by:

M. Lassas, E. Saksman, S. Siltanen, 2009. *Discretization invariant Bayesian inversion and Besov space priors.*, *Inverse Probl Imaging*, 3(1).

V. Kolehmainen, M. Lassas, K. Niinimäki, S. Siltanen, 2012. *Sparsity-promoting Bayesian inversion*, *Inverse Probl*, 28(2).

K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, S. Siltanen, 2013. *Sparse Tomography*, *SIAM J Sci Comput*, 35(3).