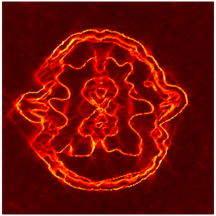
Recent Advances in Bayesian Inference for Inverse Problems

Felix Lucka

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



Centre for Medical Image Computing

Applied Inverse Problems
Helsinki, May 25, 2015

Noisy, ill-posed inverse problems:

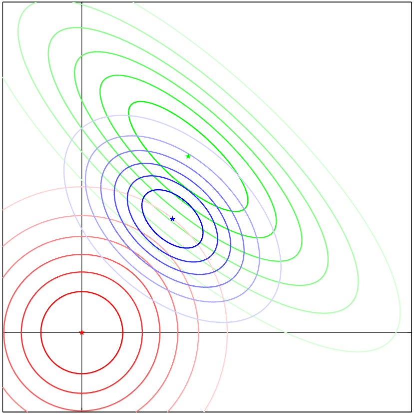
$$f = N(\mathcal{A}(u), \varepsilon)$$

Example: $f = Au + \varepsilon$

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

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

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



Probabilistic representation of solution allows for a rigorous quantification of its uncertainties.

Increased Interest in Bayesian Inversion

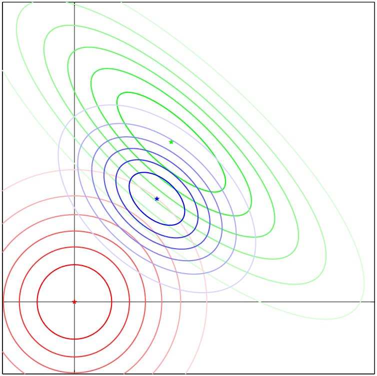
Inverse problems in the Bayesian framework
edited by Daniela Calvetti, Jari P Kaipio and Erkki Somersalo.
Special issue of *Inverse Problems*, November 2014.

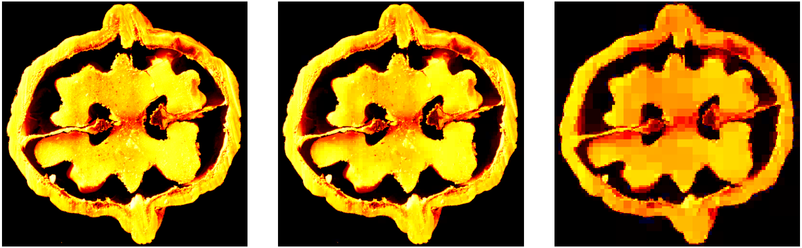
UQ and a Model Inverse Problem
Marco Iglesias and Andrew M. Stuart
SIAM News, July/August 2014.

Advantageous for high uncertainties:

- Strongly non-linear problems.
- Severely ill-posed problems.
- Little analytical structure
- Additional model uncertainties.

- Uncertainty quantification of inverse solutions.
- Dynamic Bayesian inversion for prediction or control of dynamical systems
- Infinite dimensional Bayesian inversion.
  ⇝ M26: "Theoretical perspectives in Bayesian inverse problems"
- Incorporating model uncertainties.
- New ways of encoding a-priori information.
  ⇝ "M29: Priors and SPDEs"
- Large-scale posterior sampling techniques.
  ⇝ M23: "Sampling methods for high dimensional Bayesian inverse problems"



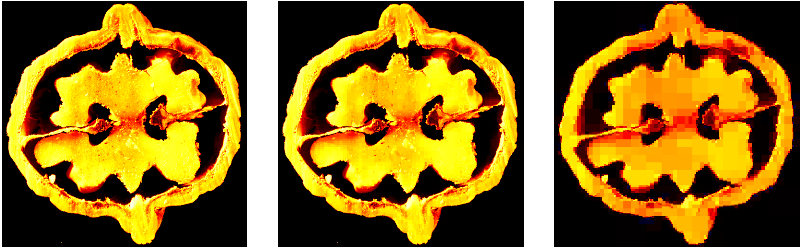


(a) 100% (b) 10% (c) 1%

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

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

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



(a) 100% (b) 10% (c) 1%

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

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

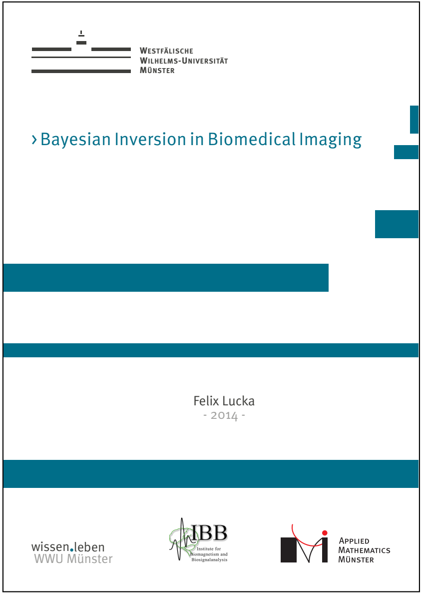
(e.g. *total variation*, *wavelet shrinkage*, *LASSO*,...)

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 computation and application.

Here: Results for ℓ_p-priors and CT.


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

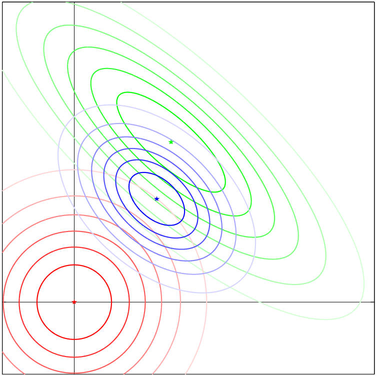
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda \, \|D^T u\|_p^p\right)$$

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

The ℓ_p Approach to Sparse Bayesian Inversion

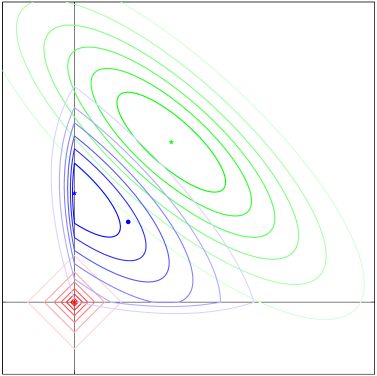
$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda \, \|D^T u\|_p^p\right)$

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



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

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



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

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

Bayesian Inference with ℓ_1 Priors

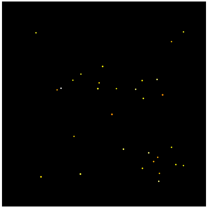
$$p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_1\right)$$

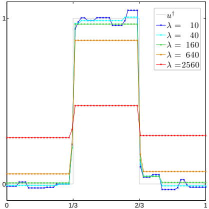
Aims: Bayesian inversion in high dimensions ($n \to \infty$).

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

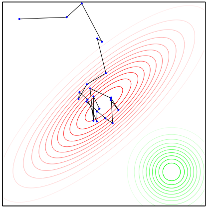
Starting points:

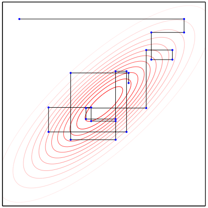
- M. Lassas, S. Siltanen, 2004. *Can one use total variation prior for edge-preserving Bayesian inversion?* *Inverse Problems*, 20.
- M. Lassas, E. Saksman, S. Siltanen, 2009. *Discretization invariant Bayesian inversion and Besov space priors.* *Inverse Problems and Imaging*, 3(1).
- V. Kolehmainen,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

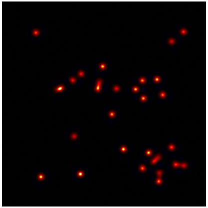




Efficient MCMC Techniques for ℓ_1 Priors







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.

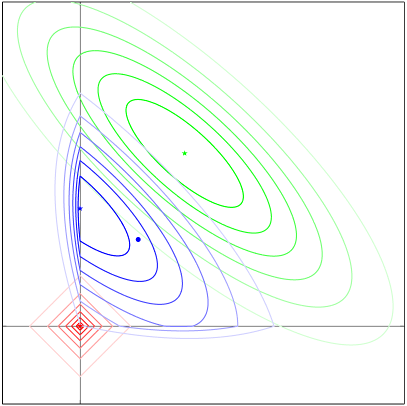
Contributions:

- 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.

F.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.

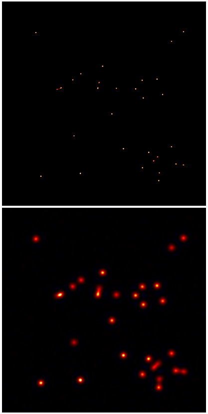
$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ \, p_{post}(u|f) \} \quad \text{vs.} \quad \hat{u}_{\text{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

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



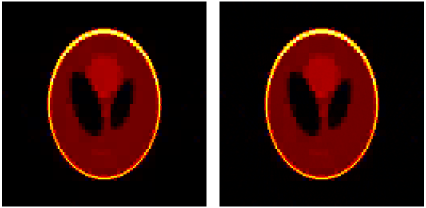
$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ \, p_{post}(u|f) \right\} \quad \text{vs.} \quad \hat{u}_{\text{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

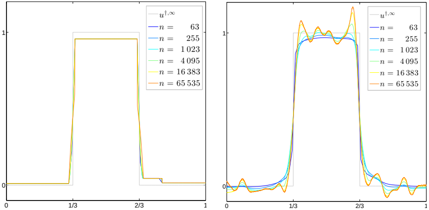
- CM preferred in theory, dismissed in practice.
- MAP discredited by theory, chosen in practice.



However:

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





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.

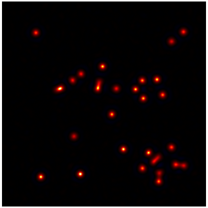
Contributions:

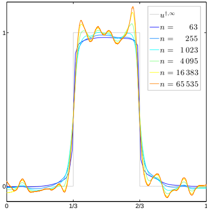
- Theoretical rehabilitation of MAP.
- Key: Bayes cost functions based on Bregman distances.
- Gaussian case consistent in this framework.

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

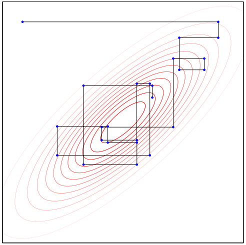
T. Helin, M. Burger, 2015. *Maximum a posteriori probability estimates in infinite-dimensional Bayesian inverse problems.*, arXiv:1412.5816v2

⇝ Talk by Martin Burger in M40-III





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

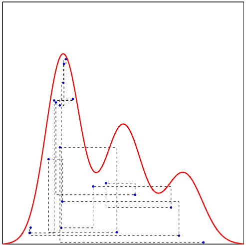


Limitations:

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

Contributions:

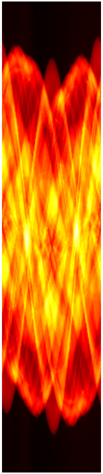
- Replace explicit by generalized slice sampling.
- Implementation & evaluation for most common priors.



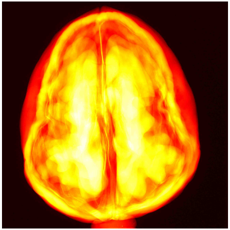
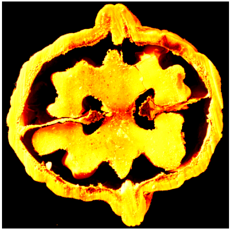
R.M. Neal, 2003. *Slice Sampling.* Annals of Statistics 31(3)

F.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.

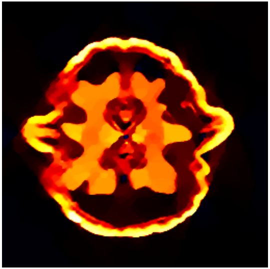




Walnut-CT with TV Prior: Full vs. Limited Angle

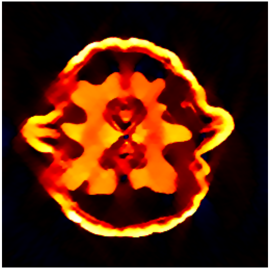
(a) MAP, full

(b) CM, full

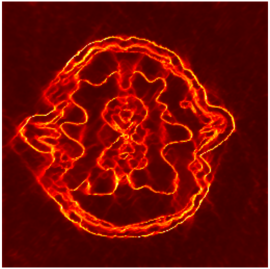
(c) CStd, full



(d) MAP, limited

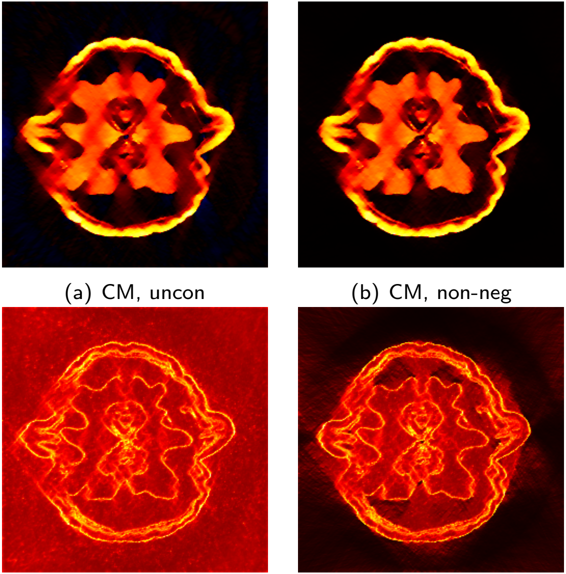


(e) CM, limited



(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

Summary & Conclusions

- Sparsity as a-priori information can be modeled in different ways.
- The elementary MCMC posterior samplers may show very different performance.
- Contrary to common beliefs they are not in general slow and scale bad with increasing dimension.
- Sample-based Bayesian inversion in high dimensions ($n > 10^6$) is feasible if tailored samplers are developed.
- MAP estimates are proper Bayes estimators.
- But "MAP or CM?" is NOT the key question in Bayesian inversion.
- Everything beyond point-estimates is far more interesting and can really complement variational approaches.

Current Related Work & Outlook

- Fast samplers can be used for simulated annealing.
- Reason for the efficiency of the Gibbs samplers is unclear.
- Adaptation, parallelization, multimodality, multi-grid.
- Combine ℓ_p-type and hierarchical priors: ℓ_p-hypermodels.
- Application studies had proof-of-concept character up to now.
- Specific UQ task to explore full potential of the Bayesian approach.

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

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

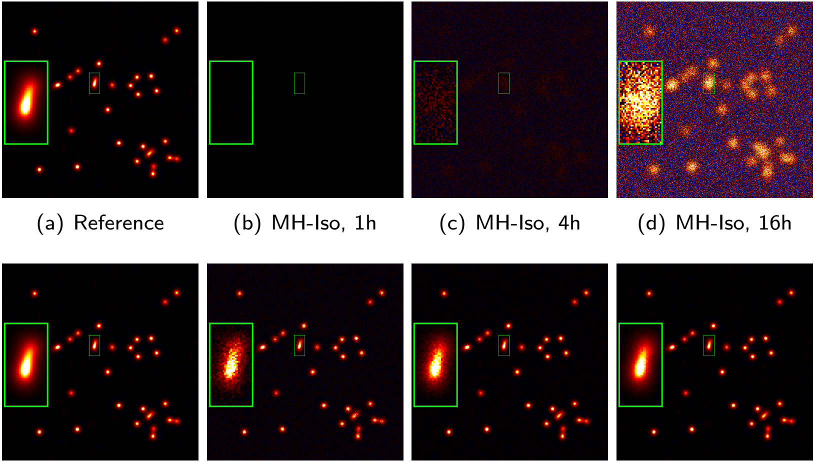
F.L., 2012. *Fast Markov chain Monte Carlo sampling for sparse Bayesian inference in high-dimensional inverse problems using L1-type priors.*
Inverse Problems, 28(12):125012.

EPSRC
Engineering and Physical Sciences Research Council

Thank you for your attention!

- F.L., 2014. *Bayesian Inversion in Biomedical Imaging*
  PhD Thesis, University of Münster.
- M. Burger, F.L., 2014. *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*
  Inverse Problems, 30(11):114004.
- F.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.

EPSRC
Engineering and Physical Sciences Research Council



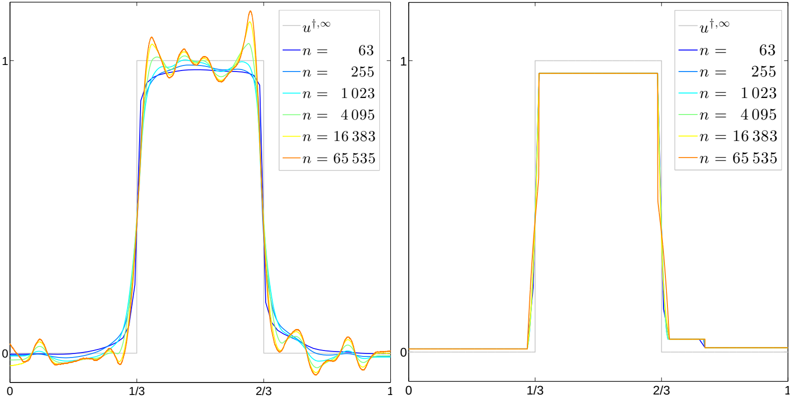
(a) Reference (b) MH-Iso, 1h (c) MH-Iso, 4h (d) MH-Iso, 16h

(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.

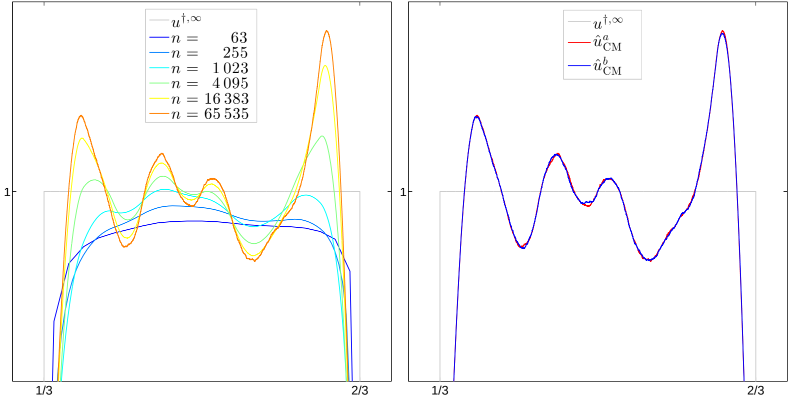


(a) CM by our Gibbs Sampler

(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.

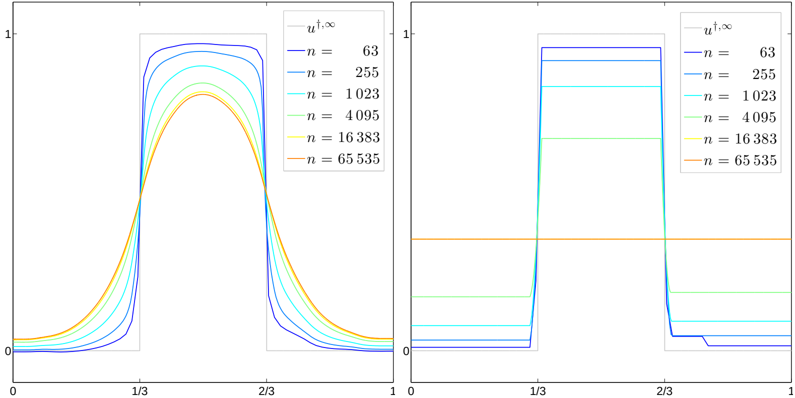


(a) Zoom into CM estimates (b) MCMC convergence check

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.

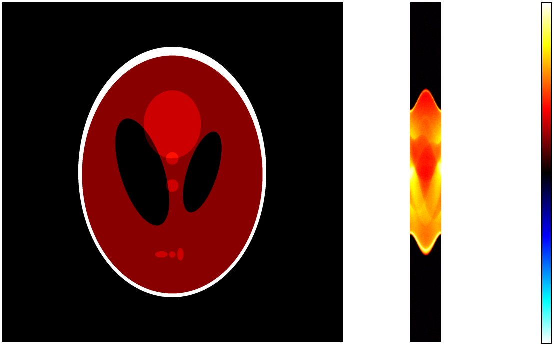


(a) CM by our Gibbs Sampler (b) MAP by ADMM

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

Need for New Theoretical Predictions

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

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

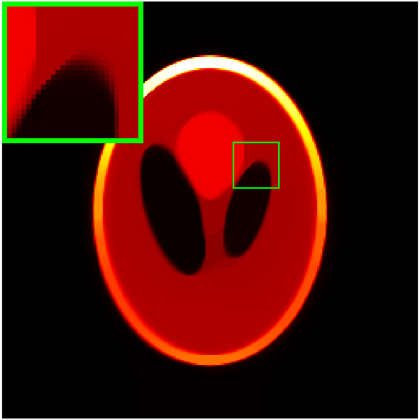
CM, $n = 64^2$, $\lambda = 500$

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

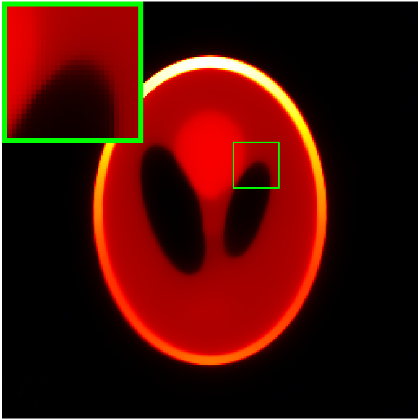
MAP, $n = 128^2$, $\lambda = 500$

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

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



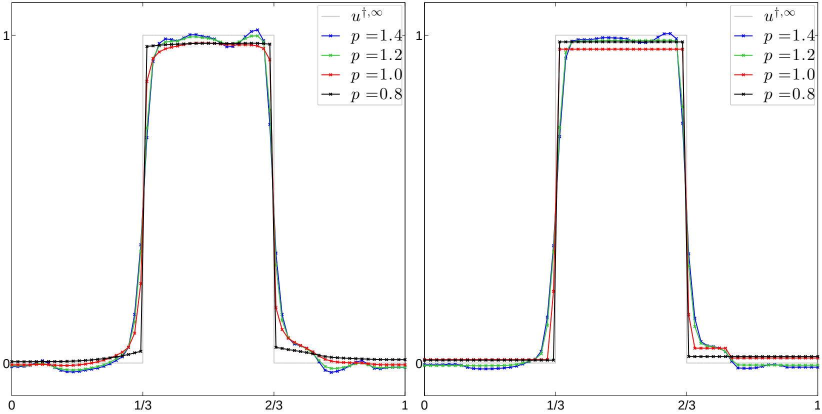
MAP, $n = 256^2$, $\lambda = 500$



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

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

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



(c) CM (Gibbs-MCMC)

(d) MAP (Simulated Annealing)

Examination of Besov Space Priors by MCMC

An ℓ_1-type, wavelet-based prior:

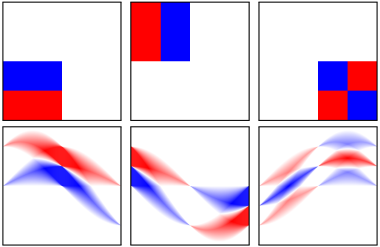
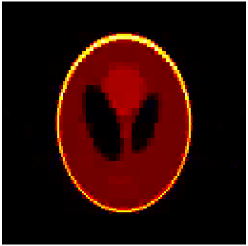
$$p_{prior}(u) \propto \exp\left(-\lambda \| W V^T u \|_1\right)$$

motivated by:

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

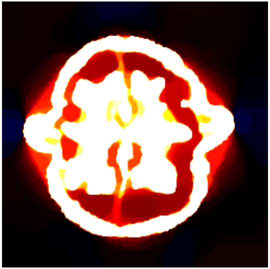
- V. Kolehmainen, M. Lassas, K. Niinimäki, S. Siltanen, 2012. *Sparsity-promoting Bayesian inversion*, *Inverse Probl*, 28(2).

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


Walnut-CT with TV Prior: Full Angle

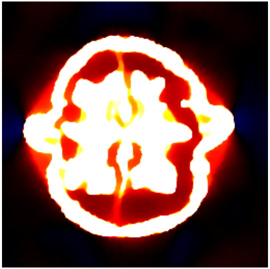
(a) MAP



(b) MAP, special color scale

(c) CStd

(d) CM



(e) CM, special color scale

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