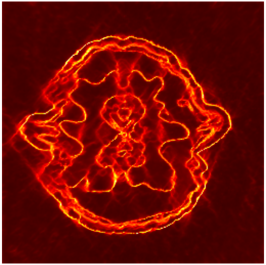
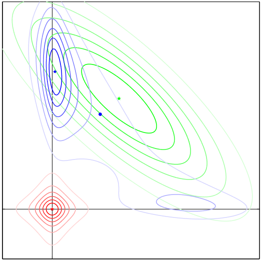
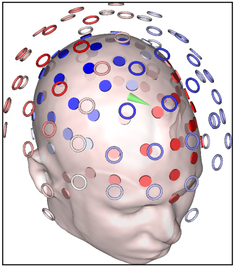
Recent Advances in Bayesian Inference for Biomedical Imaging







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

Centre for Medical Image Computing

Edinburgh, March 17, 2016

Noisy, ill-posed inverse problems:

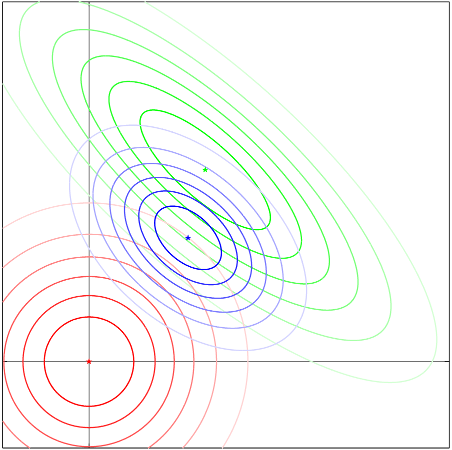
$$\boxed{f = N\left(\mathcal{A}(u), \varepsilon\right)}$$

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

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

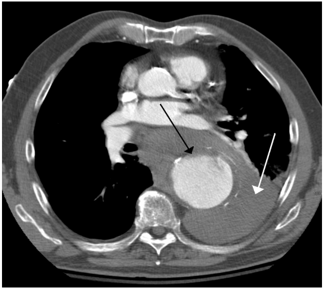
UQ and a Model Inverse Problem
Marco Iglesias and Andrew M. Stuart
SIAM News, July/August 2014.

Advantageous for high uncertainties:

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

Traditional task: Produce results to be interpreted by trained experts
$\Longrightarrow$ *Qualitative* usage of the reconstructed information.

Example: Conventional *computer tomography* (*CT*).

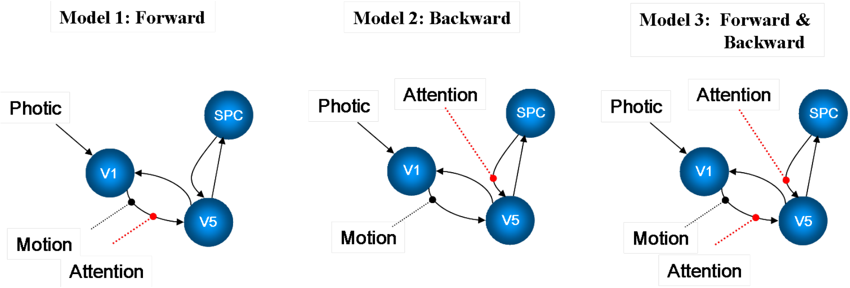


Source: Wikimedia Commons

Traditional task: Produce results to be interpreted by trained experts
$\Longrightarrow$ *Qualitative* usage of the reconstructed information.

New demand: Produce results for automatized analysis procedures / hypothesis testing; Multimodal imaging.
$\Longrightarrow$ *Quantitative* usage of the reconstructed information.

Example: *Dynamical causal modeling* (*DCM*).

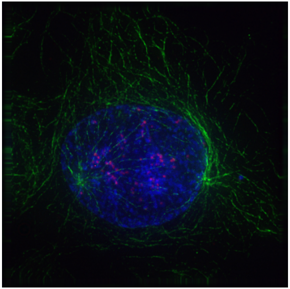


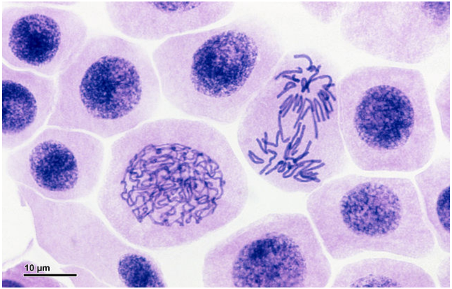
Source: Andre C. Marreiros et al. (2010), Scholarpedia, 5(7):9568.

Traditional task: Produce results to be interpreted by trained experts
$\Longrightarrow$ *Qualitative* usage of the reconstructed information.

New demand: Produce results for automatized analysis procedures / hypothesis testing; Multimodal imaging.
$\Longrightarrow$ *Quantitative* usage of the reconstructed information.

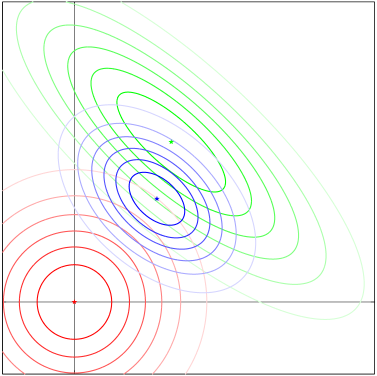
Example: Statistical analysis of microscopy images.



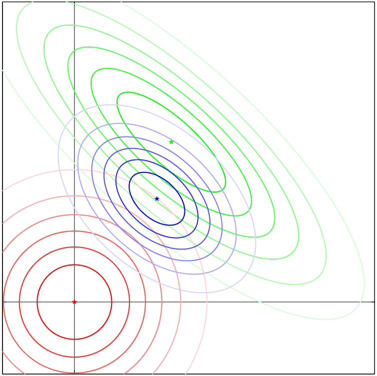


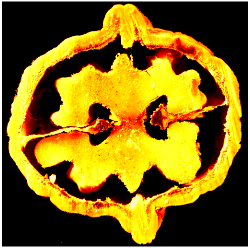
Source: Wikimedia Commons

- Uncertainty quantification of inverse solutions.
- Dynamic Bayesian inversion for prediction or control of dynamical systems
- Infinite dimensional Bayesian inversion.
- Incorporating model uncertainties.
- New ways of encoding a-priori information.
- Large-scale posterior sampling techniques.

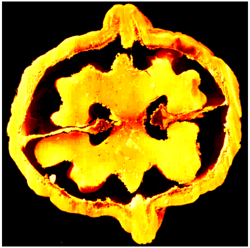


- Uncertainty quantification of inverse solutions.
- Dynamic Bayesian inversion for prediction or control of dynamical systems
- Infinite dimensional Bayesian inversion.
- Incorporating model uncertainties.
- New ways of encoding a-priori information.
- Large-scale posterior sampling techniques.





(a) 100%



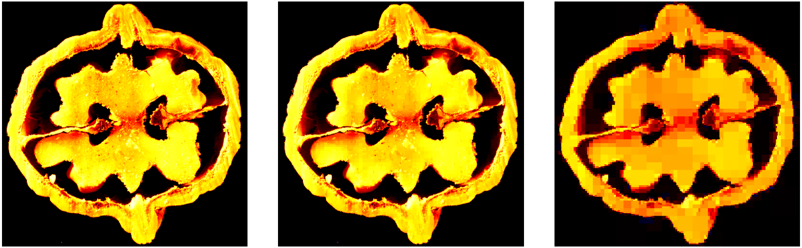
(b) 10%

(c) 1%

Sparsity a-priori constraints are used in variational regularization, compressed sensing and variable selection:

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

(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 variable selection:

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

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

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

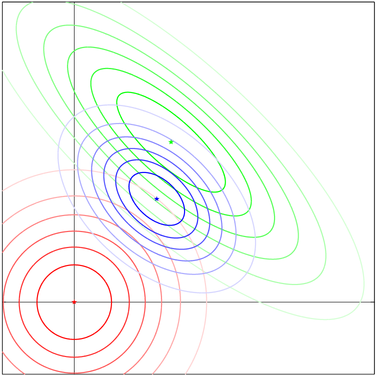
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|_{\Sigma_\varepsilon^{-1}}^2 - \lambda \, \|D^T u\|_p^p\right)$$

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

The ℓ_p Approach to Sparse Bayesian Inversion

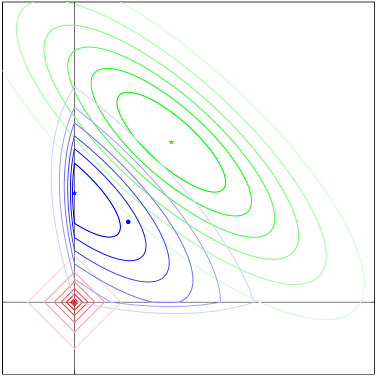
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_p^p\right)$$

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



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

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

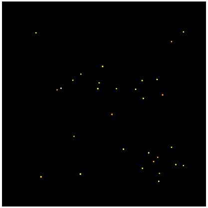


$\exp\left(-\lambda\,\|D^T u\|_1\right)$

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

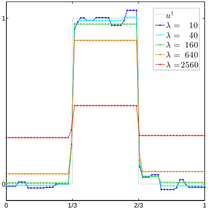
Bayesian Inference with ℓ_1 Priors

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



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

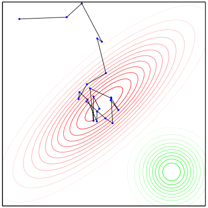
$$\boxed{p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_1\right)}$$

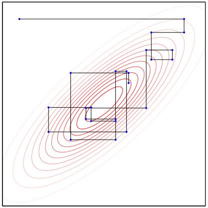
Problem: Standard Markov chain Monte Carlo (MCMC) sampler (Metropolis-Hastings) inefficient for large n or λ.

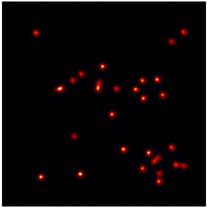
Contributions:

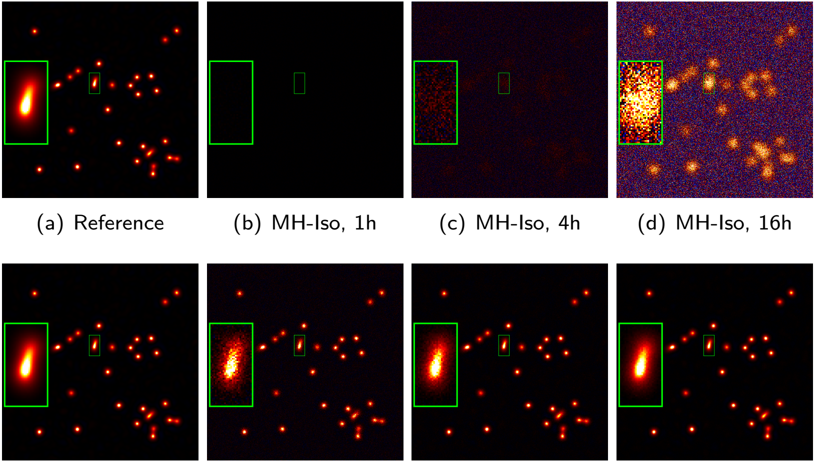
- Development of explicit single component Gibbs sampler.
- Tedious implementation for different scenarios.
- Still efficient in high dimensions ($n > 10^6$).
- Detailed evaluation and comparison to MH.

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.







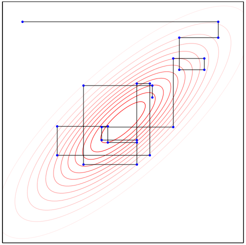


(a) Reference (b) MH-Iso, 1h (c) MH-Iso, 4h (d) MH-Iso, 16h

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

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

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

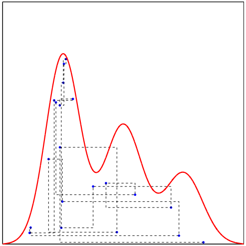


Limitations:

- 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.

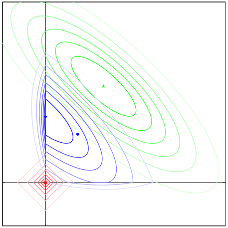


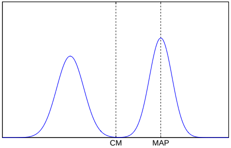
Neal, 2003. *Slice Sampling.* Annals of Statistics 31(3)

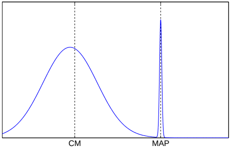
L, 2016. *Fast Gibbs sampling for high-dimensional Bayesian inversion.* submitted, arXiv:1602.08595

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

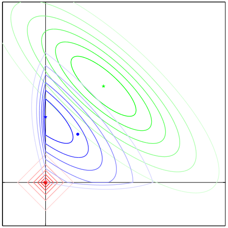
Classical Bayes cost formalism discriminates MAP (= variational regularization) and advocates CM.







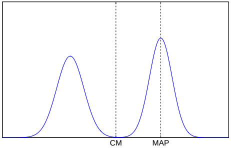
$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ p_{post}(u|f) \} \quad \text{OR} \quad \hat{u}_{\text{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

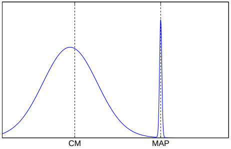


Classical Bayes cost formalism discriminates MAP (= variational regularization) and advocates CM.

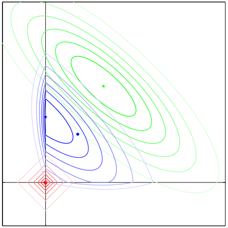
However...

- Theoretical argument has a logical flaw.
- Discrimination of MAP estimate is not intuitive.
- Gaussian priors: MAP = CM. Funny coincidence?
- Non-Gaussian priors: Poor computational validation!





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

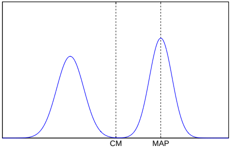


Classical Bayes cost formalism discriminates MAP (= variational regularization) and advocates CM.

However...

- Theoretical argument has a logical flaw.
- Discrimination of MAP estimate is not intuitive.
- Gaussian priors: MAP = CM. Funny coincidence?
- Non-Gaussian priors: Poor computational validation!

$\Longrightarrow$ Let's compute some examples!



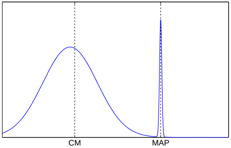
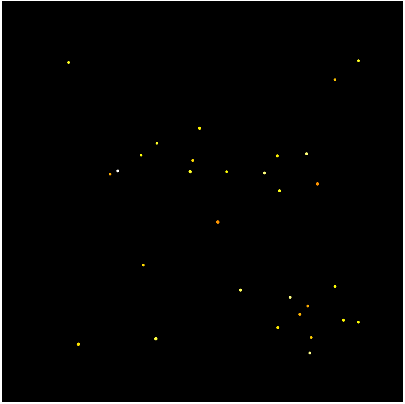
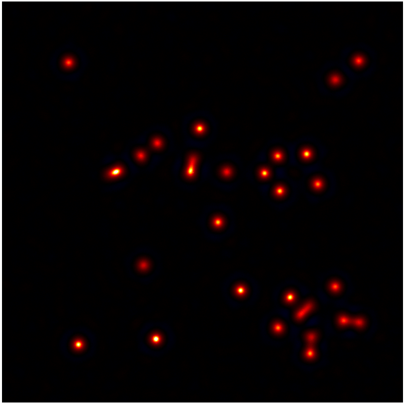


Image Deblurring Example in 2D



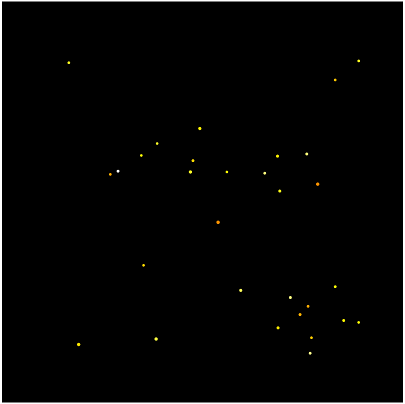
(a) Unknown function $\tilde{u}$



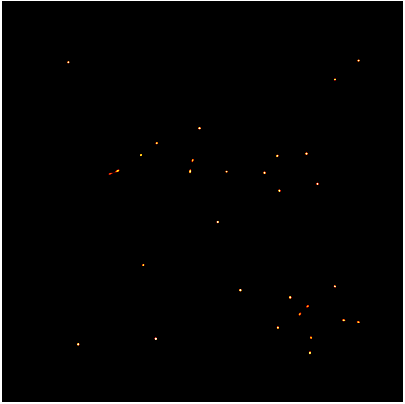
(b) CM estimate by our Gibbs sampler

Deconvolution, simple ℓ_1 prior, $n = 1023 \times 1023 = 1\,046\,529$.

Image Deblurring Example in 2D



(a) Unknown function $\tilde{u}$

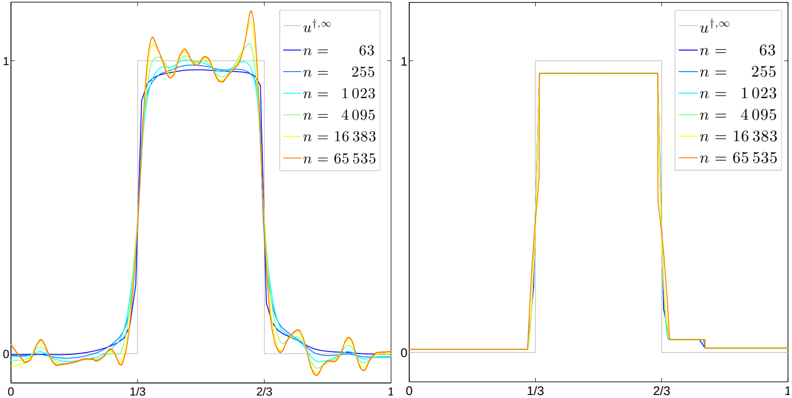


(b) MAP estimate by ADMM

Deconvolution, simple ℓ_1 prior, $n = 1023 \times 1023 = 1\,046\,529$.

"Can one use total variation prior for edge-preserving Bayesian inversion?"

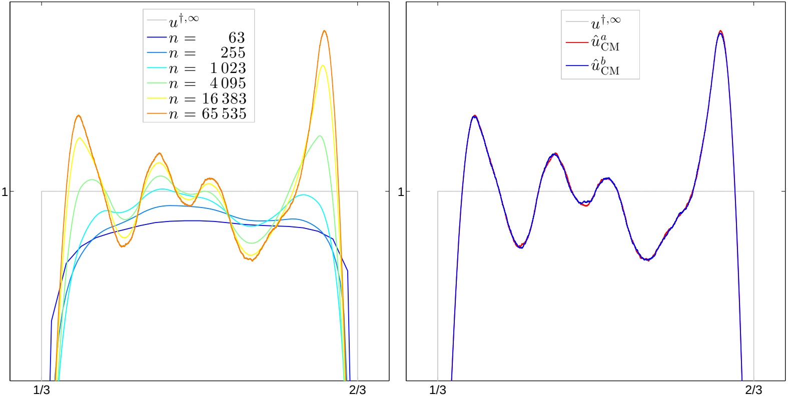
- For $\lambda_n = const.$ and $n \longrightarrow \infty$ the TV prior diverges.
- CM diverges.
- MAP converges to edge-preserving limit.



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

"Can one use total variation prior for edge-preserving Bayesian inversion?"

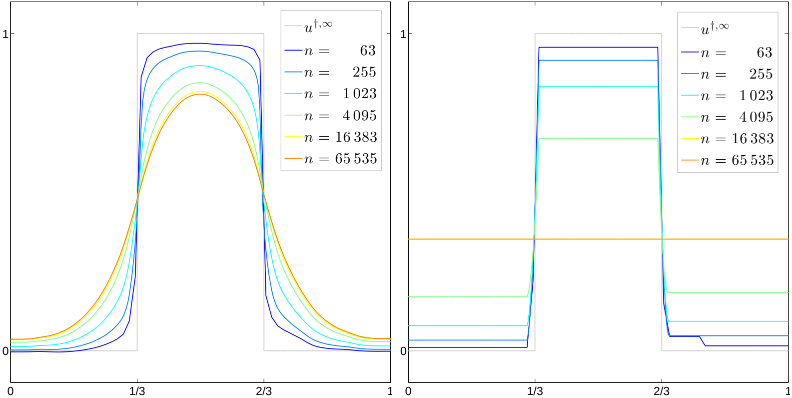
- For $\lambda_n = const.$ and $n \longrightarrow \infty$ the TV prior diverges.
- CM diverges.
- MAP converges to edge-preserving limit.



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

"Can one use total variation prior for edge-preserving Bayesian inversion?"

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



(a) CM by our Gibbs Sampler

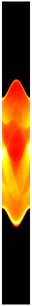
(b) MAP by ADMM

Sparse-Angle-CT with Discretization Invariant Besov Priors

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).

real solution $\tilde{u}$



data f

colormap

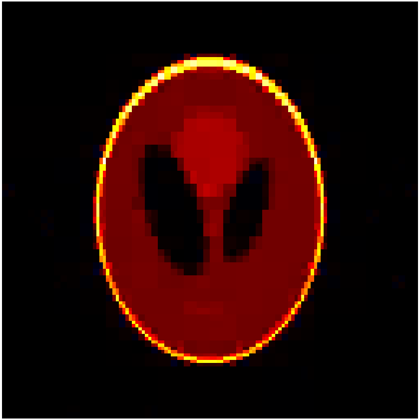
- CT using only 45 projection angles and 500 measurement pixel.
- Besov space priors using Haar wavelets.

Reconstructions for $\lambda = 2\mathrm{e}4$, $n = 64\times 64 = 4.096$

MAP estimate (by ADMM)

CM estimate (by our Gibbs sampler)

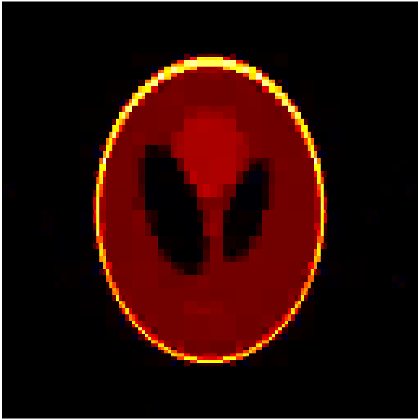
Reconstructions for $\lambda = 2e4$, $n = 128 \times 128 = 16.384$



MAP estimate (by ADMM)

CM estimate (by our Gibbs sampler)

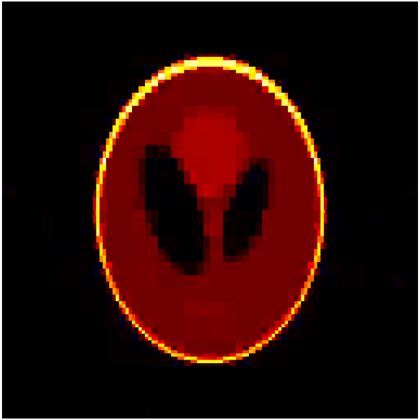
Reconstructions for $\lambda = 2e4$, $n = 256 \times 256 = 65.536$



MAP estimate (by ADMM)

CM estimate (by our Gibbs sampler)

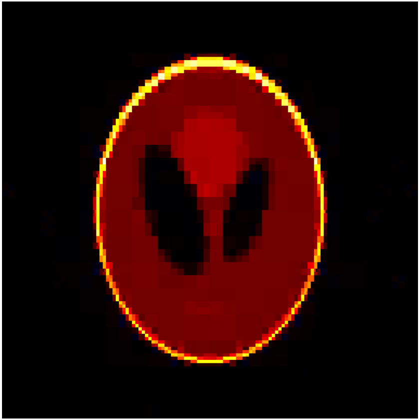
Reconstructions for $\lambda = 2e4$, $n = 512 \times 512 = 262.144$



MAP estimate (by ADMM)

CM estimate (by our Gibbs sampler)

Reconstructions for $\lambda = 2e4$, $n = 1024 \times 1024 = 1.048.576$

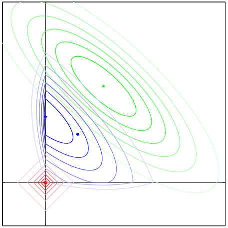


MAP estimate (by ADMM)

CM estimate (by our Gibbs sampler)

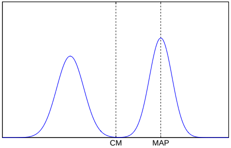
New Light on an Old Debate: MAP vs CM Estimates

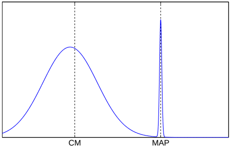
$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ \, p_{post}(u|f) \} \quad \text{OR} \quad \hat{u}_{\text{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$



Summary:

- Gaussian priors: MAP = CM. Funny coincidence?
- For reasonable priors, CM and MAP look quite similar. Fundamentally different?
- If the CM estimate looks good, it looks like the MAP.
- MAP estimates are sparser, sharper, look and perform better...
- Gribonval, Marchart, Louchet and Moisan, 2011-2013: CM are MAP estimates for different priors.





$\Longrightarrow$ Classical theory cannot be complete!

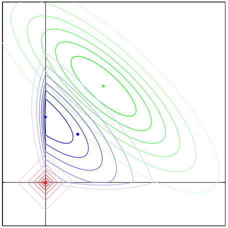
$$\hat{u}_{\mathrm{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ p_{post}(u|f) \} \quad \text{OR} \quad \hat{u}_{\mathrm{CM}} := \int u \, p_{post}(u|f) \, \mathrm{d}u$$

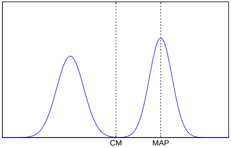
We developed new Bayes cost functions such that

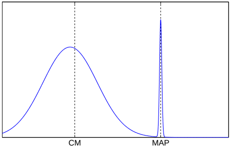
- Both MAP and CM are proper Bayes estimators for proper, convex cost functions.
- Key ingredient: Bregman distances.
- Gaussian case is no strange exception but consistent in this framework.

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*, *Inverse Problems, 31(8):085009.*

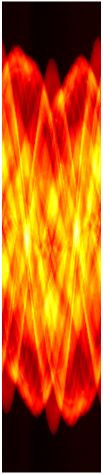


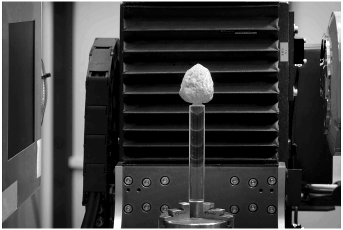


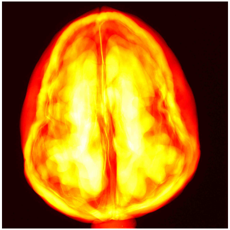


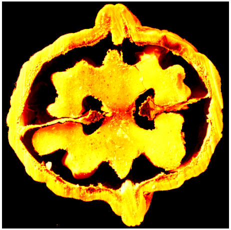
Application to Experimental Data: Walnut-CT

- 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.



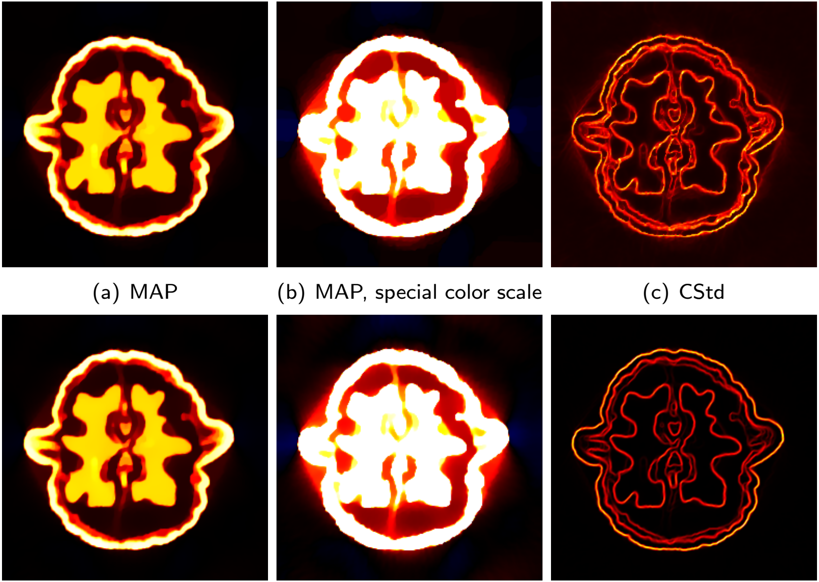






Use the data set for your own work:
http://www.fips.fi/dataset.php (documentation: arXiv:1502.04064)

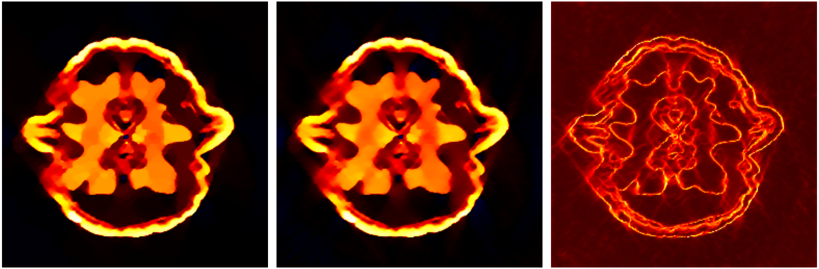
Walnut-CT with TV Prior: Full Angle



(a) MAP
(b) MAP, special color scale
(c) CStd
(d) CM
(e) CM, special color scale
(f) CM of $\|\nabla u\|_2$

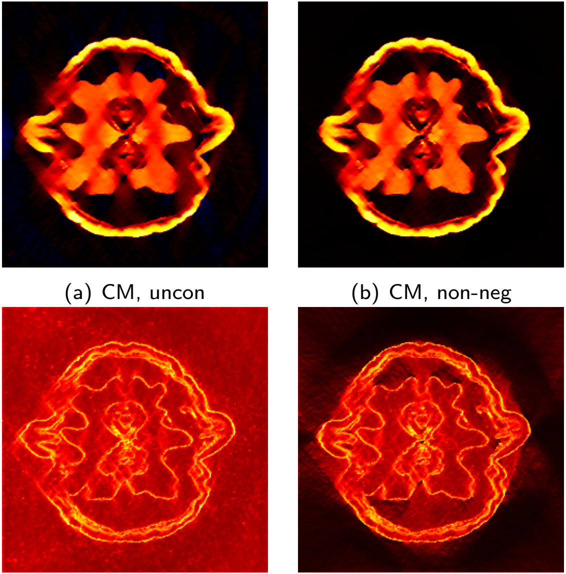
Walnut-CT with TV Prior: Full vs. Limited Angle

(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

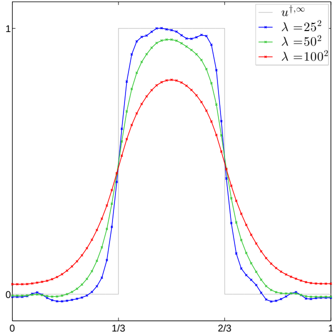
Outline

1. Introduction: Bayesian Inversion
2. Sparsity by ℓ_p Priors
3. Hierarchical Bayesian Modeling
4. Discussion, Conclusion and Outlook
5. Appendix

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.

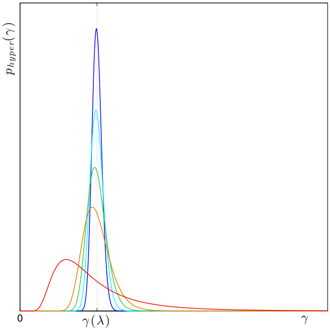


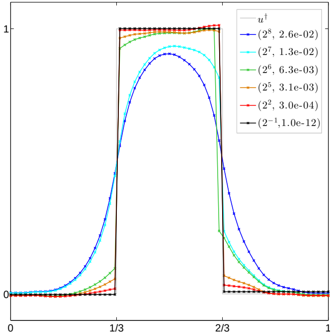
Conditionally Gaussian increment prior:

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

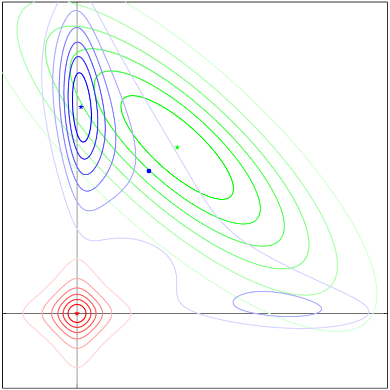
Scale-invariant hyperprior to approximate un-informative γ_i^{-1} prior:

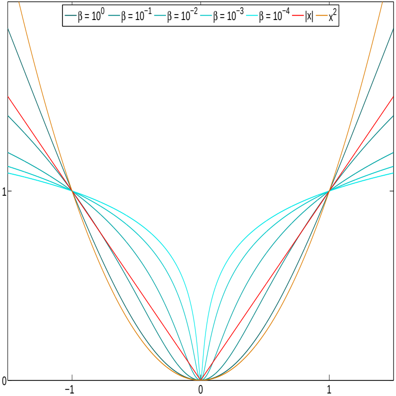
$$p_{hyper}(\gamma_i) \propto \gamma_i^{-(\alpha+1)} \exp\left(-\frac{\beta}{\gamma_i}\right), \quad \textit{inverse gamma distribution}$$





The Implicit Energy Functional behind HBM





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

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

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

Two Approaches to Sparsity

feature	ℓ_p prior	HBM
$\mathcal{J}(u)$	$\|u\|_p^p$	$\frac{\nu+1}{2}\sum \log\left(1+\frac{u^2}{\nu\theta}\right)$
sparsifying parameter	$p>0$	$\nu>0$
quadratic limit	$p=2$	$\nu\to\infty$
sparse limit	$p\to 0$	$\nu\to 0$
limit functional	$\|u\|_0$	$\sum_i^n \log(\|u_i\|)$ if all $u_i\neq 0$, $-\infty$ else
solutions	sparse	compressible
differentiable	$p>1$	always
convex	everywhere for $p\geqslant 1$	$\|u\|_\infty<\sqrt{\nu\theta}$
homogeneous	yes	no

Other stuff related to HBM: *Graphical models, general linear models, latent variable models, Variational Bayes, expectation maximization, scale mixture models, empirical priors, parametric empirical Bayes, automatic relevance determination...*

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

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

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

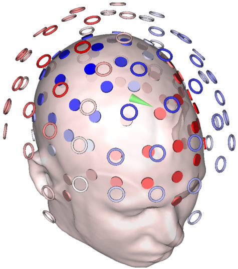
Major difficulty: Multimodality of posterior.

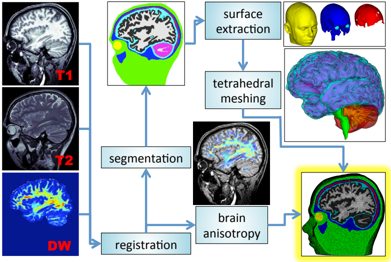
Heuristic Full-MAP computation:

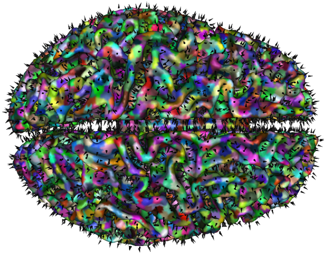
- Use MCMC to explore posterior (avoids very sub-optimal modes).
- Initialize alternating optimization by local MCMC averages to compute local modes.

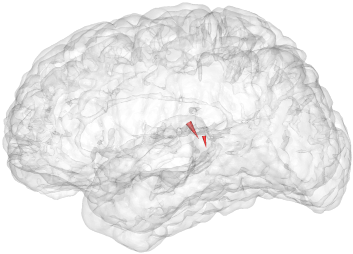
No guarantee for finding highest mode but usually an acceptable result.

Why HBM? EEG/MEG Source Reconstruction









Notoriously ill-posed problem!

- 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.
- Reason for the efficiency of the Gibbs samplers is unclear.
- Adaptation, parallelization, multimodality, multi-grid.
- Heuristic, fully Bayesian computation for HBM looks promising but needs more careful examination.

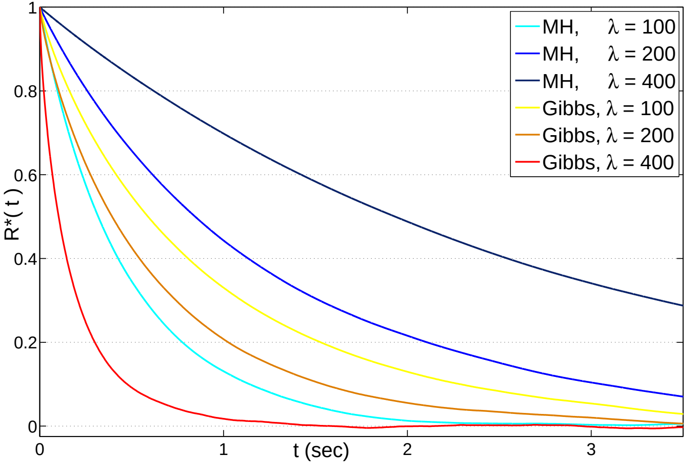
Bayesian Estimation / Uncertainty Quantification

- 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.

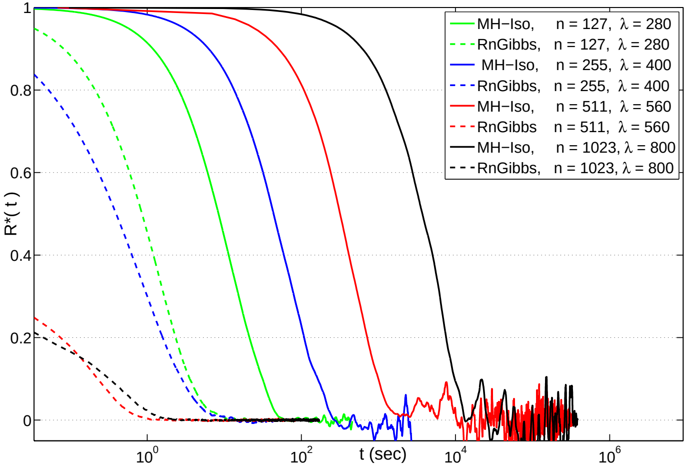
- **L, 2016.** *Fast Gibbs sampling for high-dimensional Bayesian inversion.* submitted, arXiv:1602.08595
- **L., 2014.** *Bayesian Inversion in Biomedical Imaging* PhD Thesis, University of Münster.
- **M. Burger, L., 2014.** *Maximum a posteriori estimates in linear inverse problems with log-concave 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, 2016.** *Fast Gibbs sampling for high-dimensional Bayesian inversion.* submitted, arXiv:1602.08595
- **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.



Temporal autocorrelation $R^*(t)$ for 1D TV-deblurring, $n = 63$.



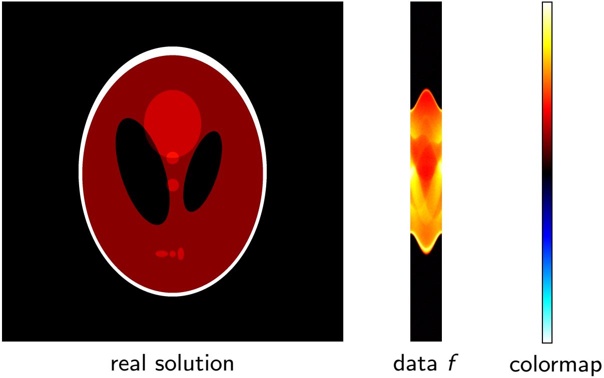
Temporal autocorrelation $R^*(t)$ for 1D TV-deblurring.

Need for New Theoretical Predictions

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

Test scenario:

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



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

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

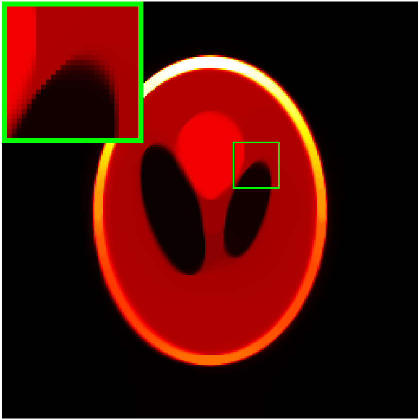
CM, $n = 64^2$, $\lambda = 500$

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

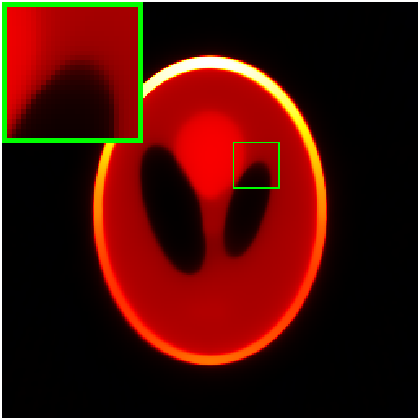
MAP, $n = 128^2$, $\lambda = 500$

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

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



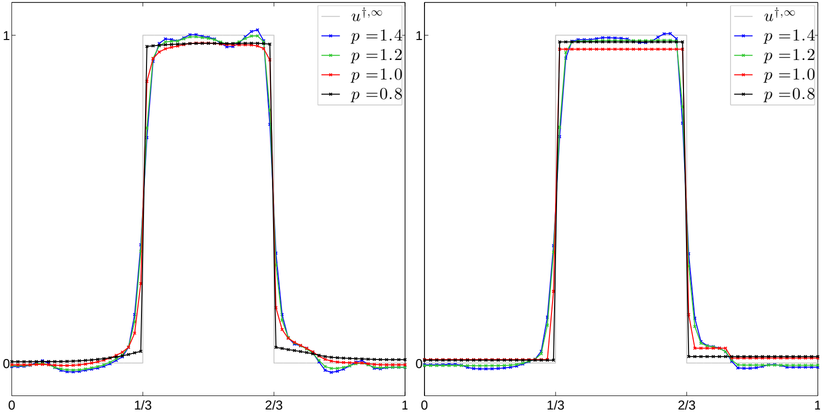
MAP, $n = 256^2$, $\lambda = 500$



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

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

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



(i) CM (Gibbs-MCMC)

(j) MAP (Simulated Annealing)

A theoretical argument "decides" the conflict: The Bayes cost formalism.

- An estimator is a random variable, as it relies on f and u.
- How does it perform on average? Which estimator is "best"?
- $\rightsquigarrow$ Define a cost function $\Psi(u, v)$.
- Bayes cost is the expected cost:

$$BC(\hat{u}) = \iint \Psi(u, \hat{u}(f))\, p_{like}(f|u)\,\mathrm{d}f\; p_{prior}(u)\,\mathrm{d}u$$

- Bayes estimator $\hat{u}_{BC}$ for given Ψ minimizes Bayes cost. Turns out:

$$\hat{u}_{BC}(f) = \operatorname*{argmin}_{\hat{u}} \left\{ \int \Psi(u, \hat{u}(f))\, p_{post}(u|f)\,\mathrm{d}u \right\}$$

MAP vs. CM Estimates: The Classical View

Main classical arguments pro CM and contra MAP estimates:

- CM is Bayes estimator for $\Psi(u, \hat{u}) = \|u - \hat{u}\|_2^2$ (MSE).
- Also the minimum variance estimator.
- The mean value is intuitive, it is the "center of mass", the known "average".
- MAP estimate can be seen as an asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 0, & \text{if} \quad \|u - \hat{u}\|_\infty \leqslant \epsilon \\ 1 & \text{otherwise}, \end{cases}$$

  for $\epsilon \to 0$ (uniform cost). $\Longrightarrow$ It is not a proper Bayes estimator.

- MAP and CM seem theoretically and computationally fundamentally different $\Longrightarrow$ one should decide.
- *"A real Bayesian would not use the MAP estimate"*
- People feel "ashamed" when they have to compute MAP estimates (even when their results are good).

A False Conclusion

"A real Bayesian would not use the MAP estimate as it is not a proper Bayes estimator".

"MAP estimate can be seen as an asymptotic Bayes estimator of

$$\Psi_\epsilon(u, \hat{u}) = \begin{cases} 0, & \text{if} \quad \|u - \hat{u}\|_\infty < \epsilon \\ 1 & \text{otherwise,} \end{cases}$$

for $\epsilon \to 0$.
???$\Longrightarrow$??? It is not a proper Bayes estimator."

"MAP estimator is asymptotic Bayes estimator for some degenerate Ψ"
$\not\Rightarrow$ "MAP can't be Bayes estimator for some proper Ψ" !!!!

Define

(a) $\Psi_{\text{LS}}(u, \hat{u}) := \|A(\hat{u} - u)\|^2_{\Sigma_\varepsilon^{-1}} + \beta \|L(\hat{u} - u)\|_2^2$

(b) $\Psi_{\text{Brg}}(u, \hat{u}) := \|A(\hat{u} - u)\|^2_{\Sigma_\varepsilon^{-1}} + \lambda D_{\mathcal{J}}(\hat{u}, u)$

for a regular L and $\beta > 0$.

Properties:

- Proper, convex cost functions
- For $\mathcal{J}(u) = \beta/\lambda \|Lu\|_2^2$ (Gaussian case!) we have $\lambda D_{\mathcal{J}}(\hat{u}, u) = \beta \|L(\hat{u} - u)\|_2^2$, and $\Psi_{\text{LS}}(u, \hat{u}) = \Psi_{\text{Brg}}(u, \hat{u})$!

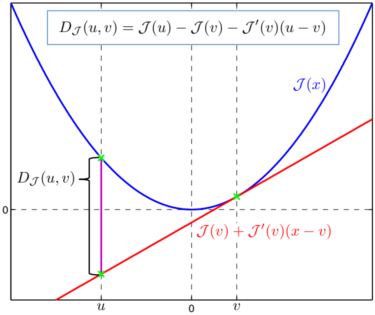
Theorems:

(I) The CM estimate is the Bayes estimator for $\Psi_{\text{LS}}(u, \hat{u})$

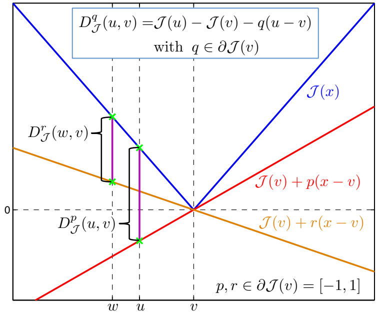
(II) The MAP estimate is the Bayes estimator for $\Psi_{\text{Brg}}(u, \hat{u})$

For a proper, convex functional $\Psi : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\infty\}$, the *Bregman distance* $D_\Psi^p(f,g)$ between $f,g \in \mathbb{R}^n$ for a subgradient $p \in \partial\Psi(g)$ is defined as

$$D_\Psi^p(f,g) = \Psi(f) - \Psi(g) - \langle p, f-g \rangle, \qquad p \in \partial\Psi(g)$$



(k) $\mathcal{J}(x) = x^2$



(l) $\mathcal{J}(x) = |x|$

Basically, $D_\Psi(f,g)$ measures the difference between Ψ and its linearization in f at another point g