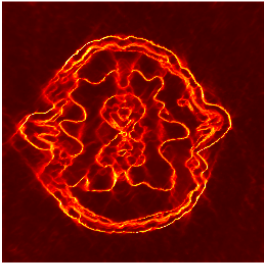
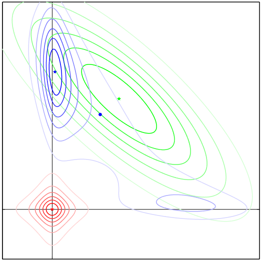
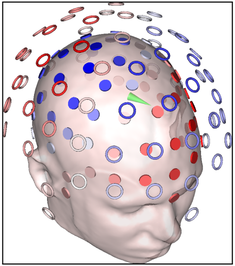
Sparse Bayesian Inversion in Biomedical Imaging

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

Centre for Medical Image Computing

Télécom ParisTech, March 3, 2016

Noisy, ill-posed inverse problems:

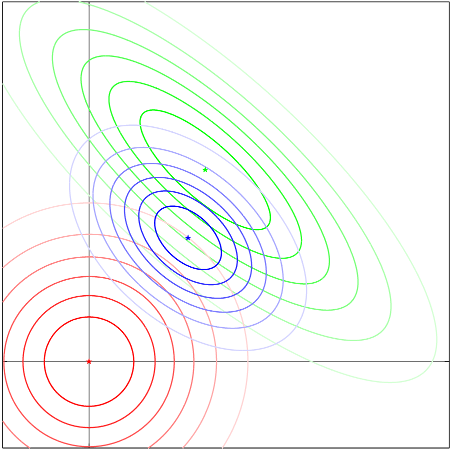
$$f = N\left(\mathcal{A}(u), \varepsilon\right)$$

Example: $f = Au + \varepsilon$

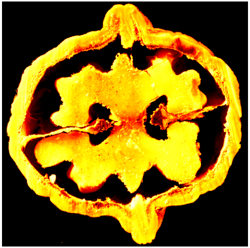
$p_{like}(f|u) \propto$
$\exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}}\right)$

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

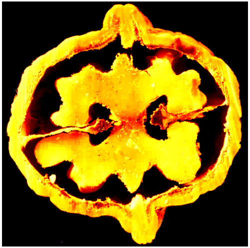
$p_{post}(u|f) \propto$
$\exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_2^2\right)$



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



(a) 100%



(b) 10%

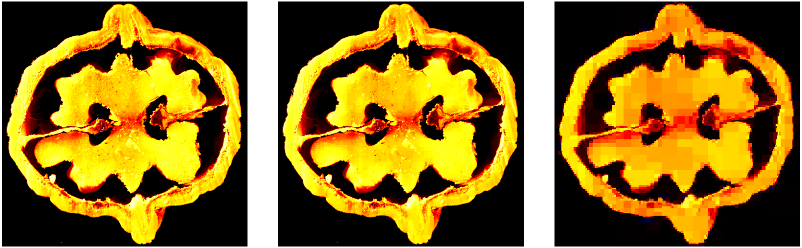
(c) 1%

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%

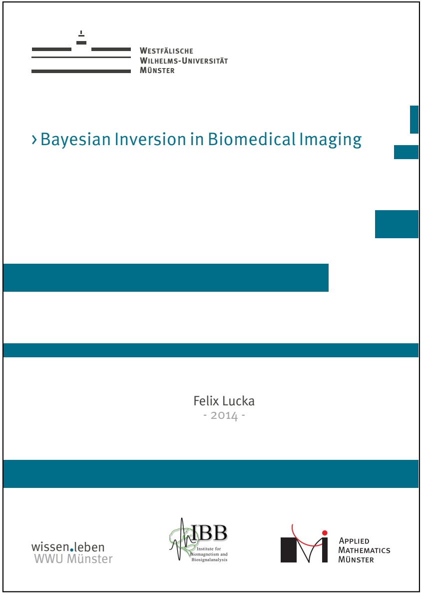
Sparsity a-priori constraints are used in variational regularization, compressed sensing and ridge regression:

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

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

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

- 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

Outline

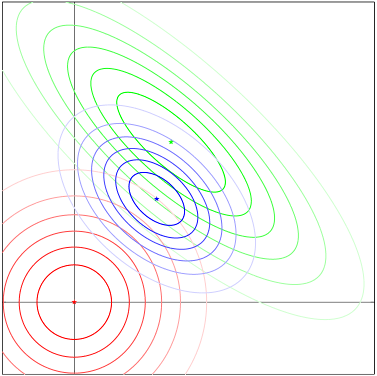
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_p^p\right)$$

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

The ℓ_p Approach to Sparse Bayesian Inversion

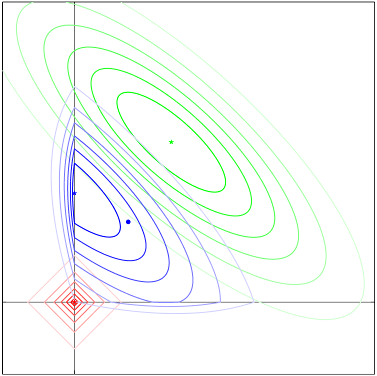
$$p_{prior}(u) \propto \exp\left(-\lambda \, \|D^T u\|_p^p\right), \quad p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda \, \|D^T u\|_p^p\right)$$

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



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

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

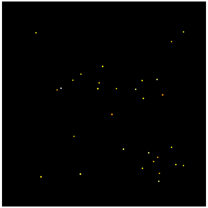


$$\exp\left(-\lambda \, \|D^T u\|_1\right)$$

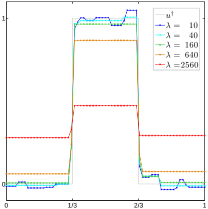
$$\exp\left(-\tfrac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda \, \|D^T u\|_1\right)$$

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 ($n \to \infty$).

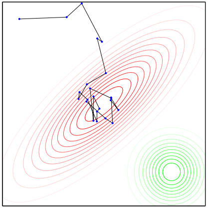


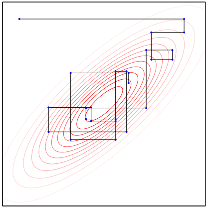
Priors: Simple ℓ_1, total variation (TV), Besov space priors.

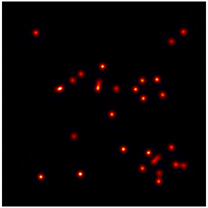
Starting points:

- **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

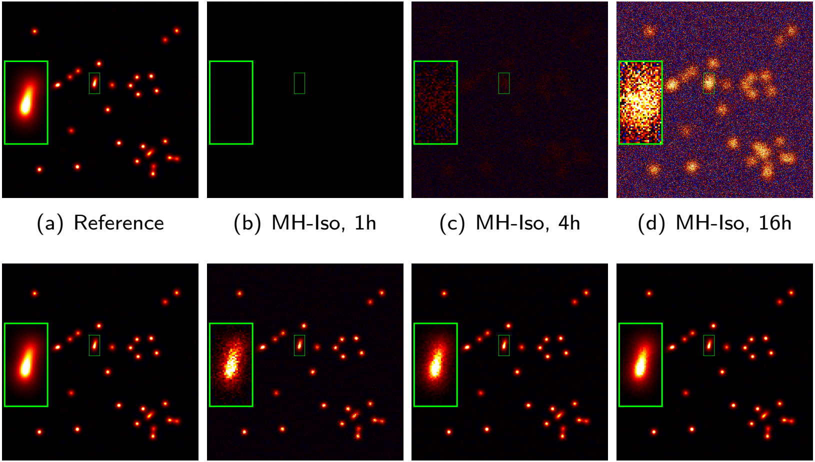
$$p_{post}(u|f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|^2_{\Sigma_\varepsilon^{-1}} - \lambda\,\|D^T u\|_1\right)$$

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.



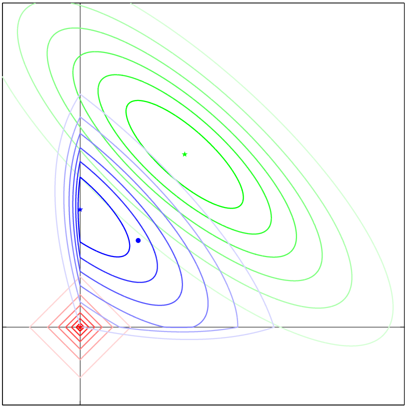
(a) Reference (b) MH-Iso, 1h (c) MH-Iso, 4h (d) MH-Iso, 16h

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

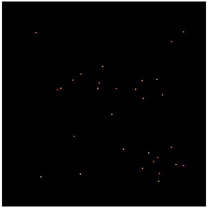
Deconvolution, simple ℓ_1 prior, $n = 513 \times 513 = 263\,169$.

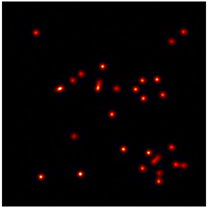
$$\hat{u}_{\text{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{\, p_{post}(u|f) \} \quad \text{vs.} \quad \hat{u}_{\text{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

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



New Theoretical Ideas for an Old Bayesian Debate



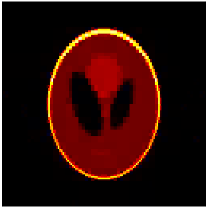


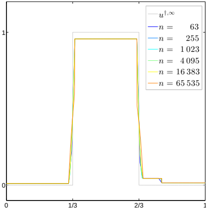
$$\hat{u}_{\mathrm{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \{ \, p_{post}(u|f) \} \quad \text{vs.} \quad \hat{u}_{\mathrm{CM}} := \int u \; p_{post}(u|f) \, \mathrm{d}u$$

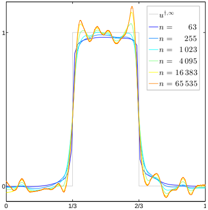
- CM preferred in theory, dismissed in practice.
- MAP discredited by theory, chosen in practice.

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

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

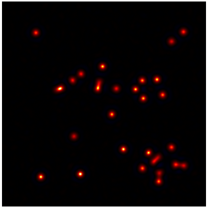
- CM preferred in theory, dismissed in practice.
- MAP discredited by theory, chosen in practice.

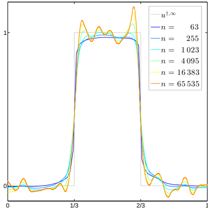
Contributions:

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

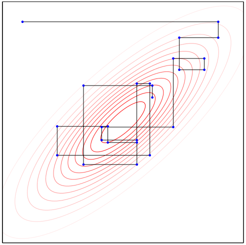
Burger & L, 2014. *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*, *Inverse Problems*, 30(11):114004.

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





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

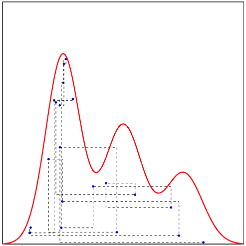


Limitations:

- 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.

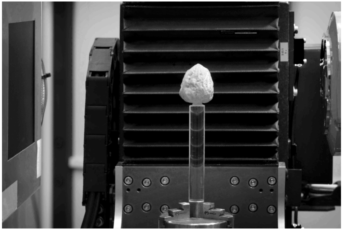
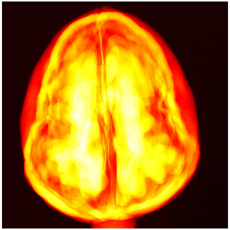
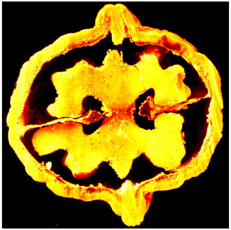
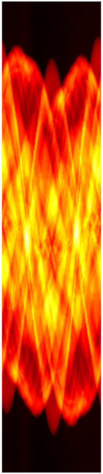


Neal, 2003. *Slice Sampling.* Annals of Statistics 31(3)

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

Application to Experimental Data: Walnut-CT

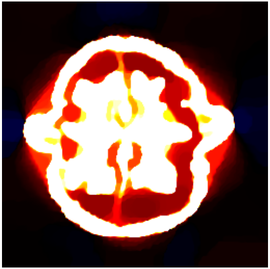
- Cooperation with Samuli Siltanen, Esa Niemi et al.
- Implementation of MCMC methods for Fanbeam-CT.
- Besov and TV prior; non-negativity constraints.
- Stochastic noise modeling.
- Bayesian perspective on limited angle CT.

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

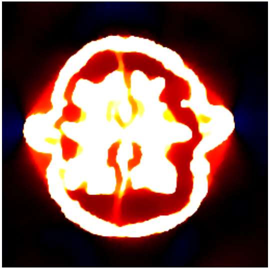
(a) MAP



(b) MAP, special color scale

(c) CStd

(d) CM

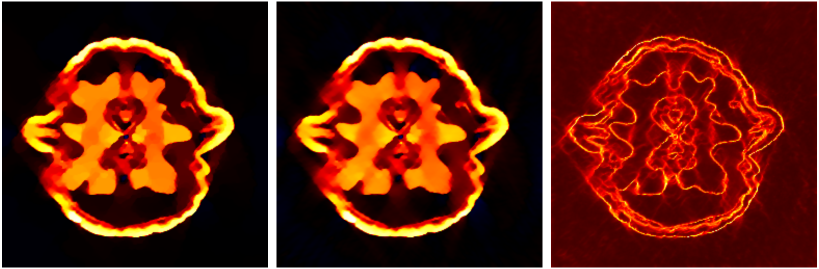


(e) CM, special color scale

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

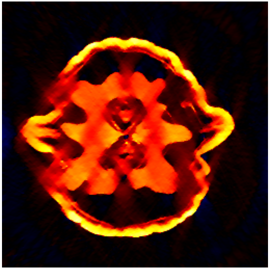
Walnut-CT with TV Prior: Full vs. Limited Angle

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

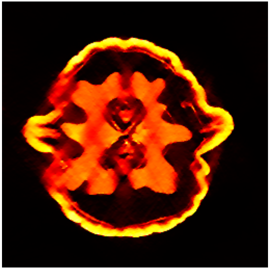


(d) MAP, limited (e) CM, limited (f) CStd, limited

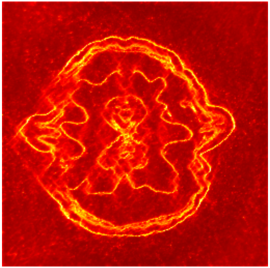
Walnut-CT with TV Prior: Non-Negativity Constraints, Limited Angle



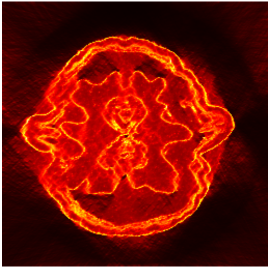
(a) CM, uncon



(b) CM, non-neg



(c) CStd, uncon

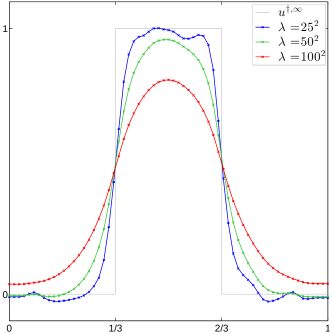


(d) CStd, non-neg

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.

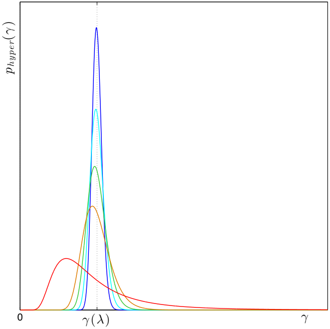


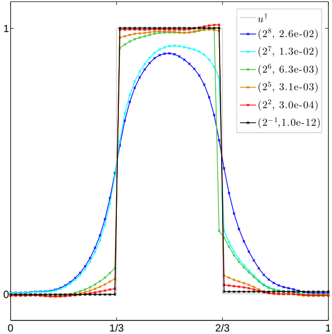
Conditionally Gaussian increment prior:

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

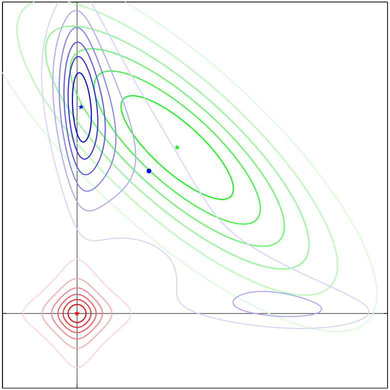
Scale-invariant hyperprior to approximate un-informative γ_i^{-1} prior:

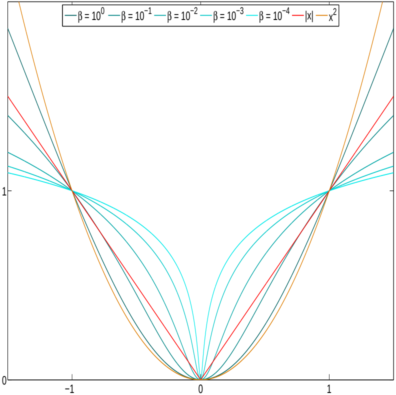
$$p_{hyper}(\gamma_i) \propto \gamma_i^{-(\alpha+1)} \exp\left(-\frac{\beta}{\gamma_i}\right), \quad \textit{inverse gamma distribution}$$





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...*

Hierarchical, Fully Bayesian Computation

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

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

Aim: Reconstruction of brain activity by non-invasive measurement of induced electromagnetic fields (bioelectromagnetism) outside of the skull.

source: Wikimedia Commons

source: Wikimedia Commons

Why HBM? EEG/MEG Source Reconstruction

Aim: Reconstruction of brain activity by non-invasive measurement of induced electromagnetic fields (bioelectromagnetism) outside of the skull.

source: Wikimedia Commons

source: Wikimedia Commons

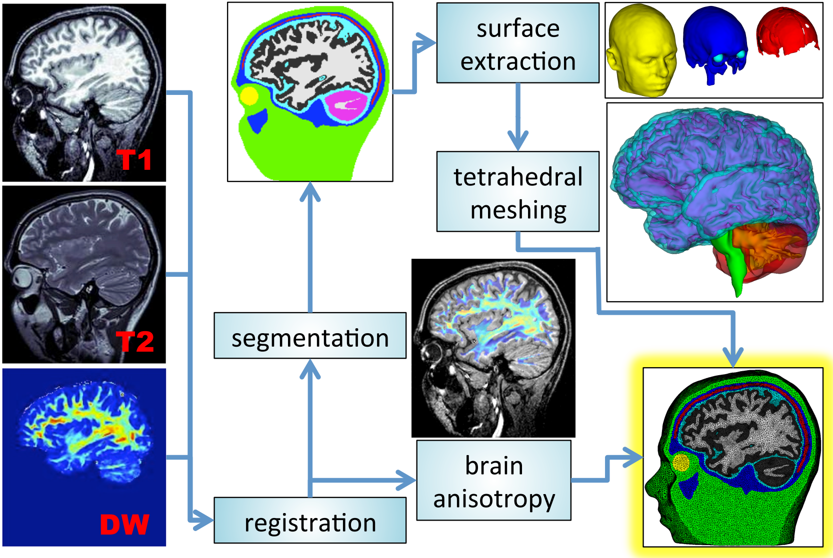
Notoriously ill-posed problem!

Workgroup "Methods in Bioelectromagnetism" in Münster

Aim: Improve quality, applicability and reliability of EEG/MEG based source reconstruction for the presurgical diagnosis of epilepsy patients.

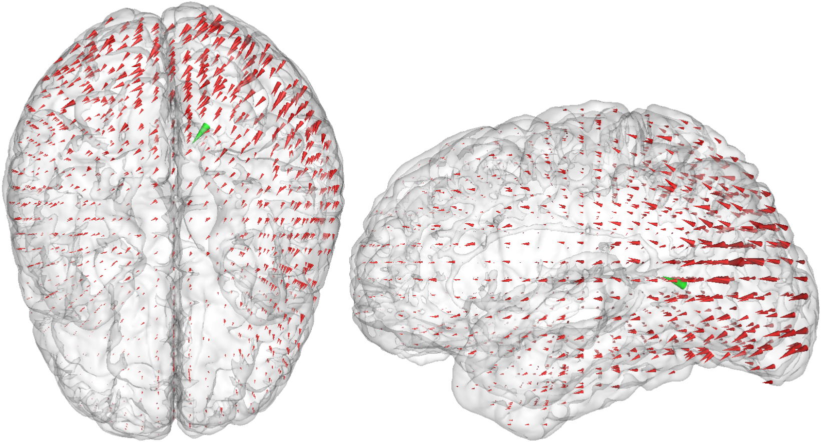
Challenges: Forward Modeling & Computation

Realistic and individual head models for simulating the forward equations.



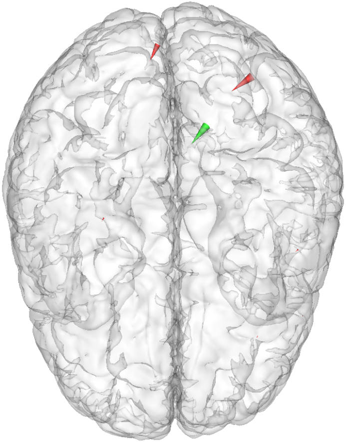
Reference (green cone) and MAP for Gaussian prior (red cones):

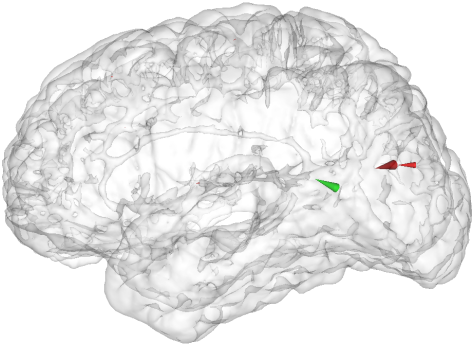
$$u_{\text{MAP}} = \underset{u}{\operatorname{argmin}} \left\{ \| f - A\, u \|_2^2 + \lambda \| u_{\text{amp}} \|_2^2 \right\}$$



Reference (green cone) and MAP for ℓ_1 prior (red cones):

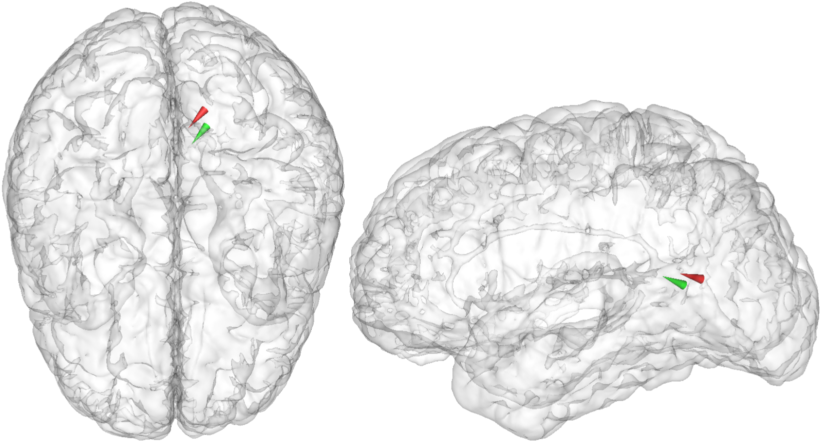
$$u_{\mathrm{MAP}} = \underset{u}{\operatorname{argmin}} \left\{ \| f - A\,u \|_2^2 + \lambda \| u_{\mathrm{amp}} \|_1 \right\}$$





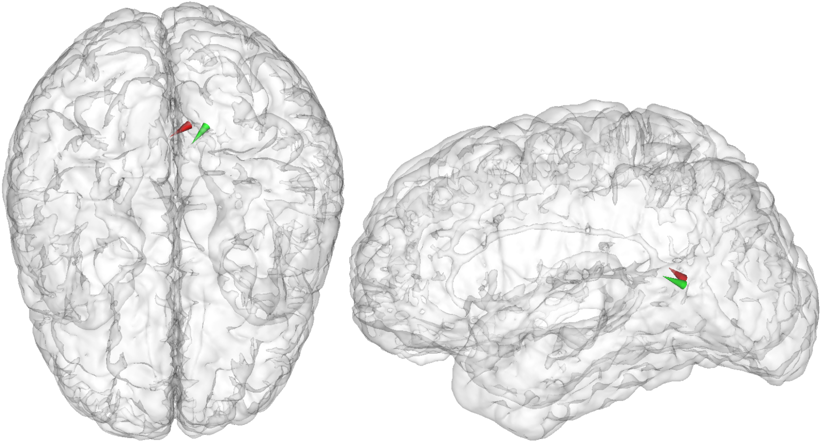
Reference (green cone) and single dipole scan (red cone):

$$u_{\mathrm{SDS}} = \operatorname*{argmin}_u \left\{ \| f - A\,u \|_2^2 + N_1(u) \right\}, \qquad N_1(u) = \begin{cases} 0 & \text{if } |u_{\mathrm{amp}}|_0 = 1 \\ \infty & \text{else} \end{cases}$$



Reference (green cone) and HBM-MAP estimate (red cone):

something like $u_{\text{MAP}} \simeq \underset{u}{\operatorname{argmin}} \left\{ \|f - A\,u\|_2^2 + \frac{\nu - 1}{2} \log\left(1 + \frac{u_{\text{amp}}^2}{\nu\theta}\right)\right\}$



The Curse of Convexity: Depth Bias

"Theorem": All MAP estimates for posteriors like

$$p_{post}(u|f) \propto \exp\left(-\tfrac{1}{2}\|f - A\,u\|_2^2 + \sum_i g(|u_i|)\right)$$

with priors that are uniform in i (no weighting) with convex g have depth bias:

- $|\hat{u}_i|$ has its maximum at the boundary of the gray matter.
- The proof combines properties of the adjoint problem of EEG/MEG with convex analysis (appendix).

Our (earlier) empirical results for EEG confirm this:

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.

- HBM does not suffer from systematic depth miss-localization.
- HBM shows promising results for focal brain networks with simulated and real data.
- Focus of my PhD work: HBM for EEG-MEG combination.

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!

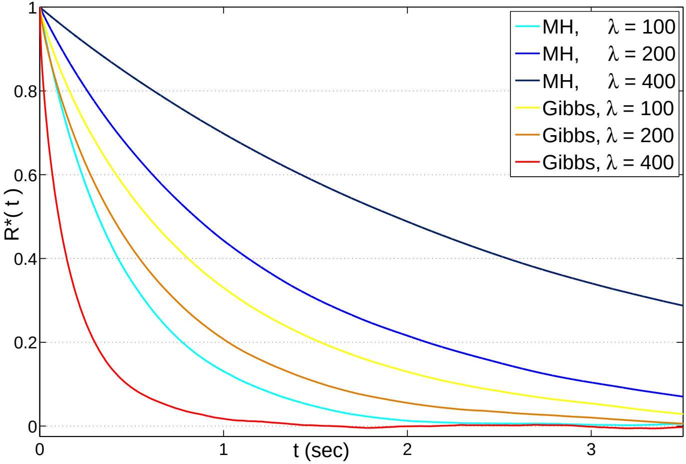
L, 2016. *Fast Gibbs sampling for high-dimensional Bayesian inversion.* submitted, arXiv:1602.08595

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

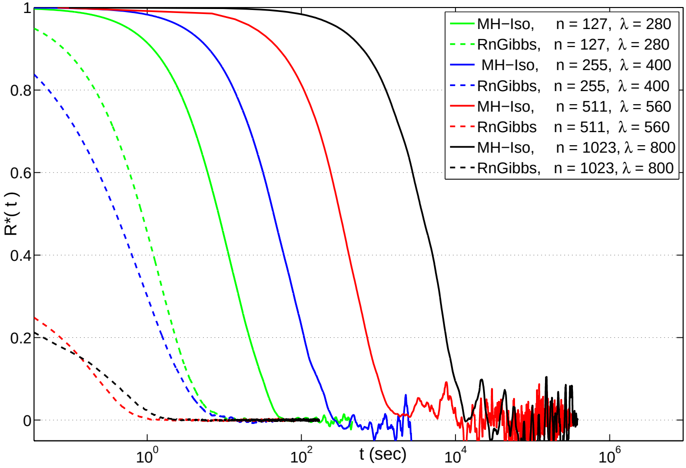
M. Burger, L., 2014. *Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators*
Inverse Problems, 30(11):114004.

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.



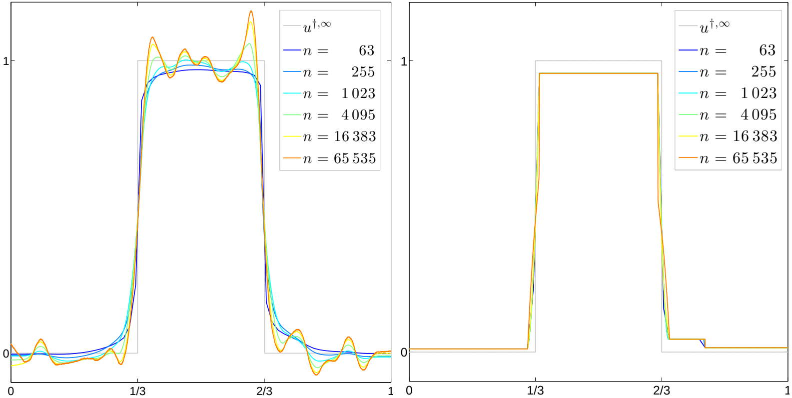
Temporal autocorrelation $R^*(t)$ for 1D TV-deblurring, $n = 63$.



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

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.

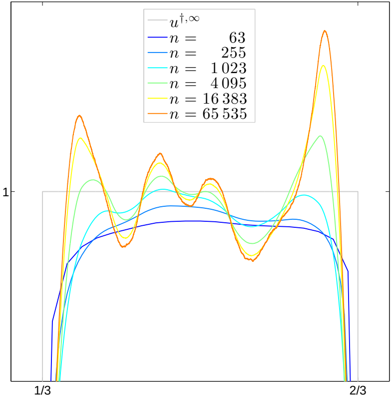


(a) CM by our Gibbs Sampler

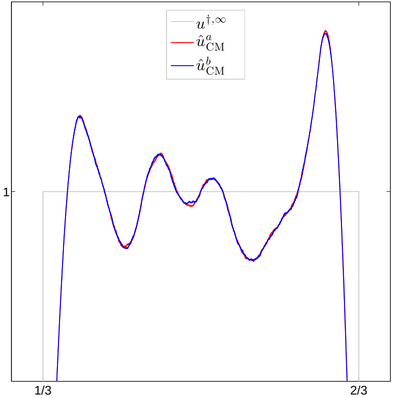
(b) MAP by ADMM

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.



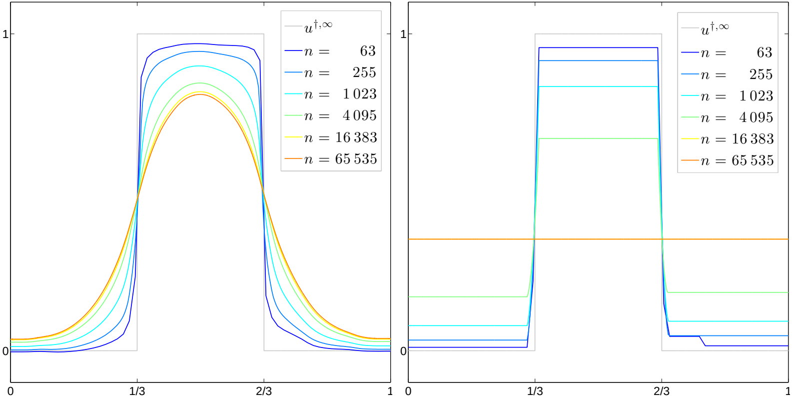
(a) Zoom into CM estimates



(b) MCMC convergence check

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

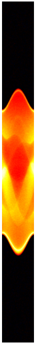
Need for New Theoretical Predictions

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

Test scenario:

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

real solution



data f

colormap

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

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

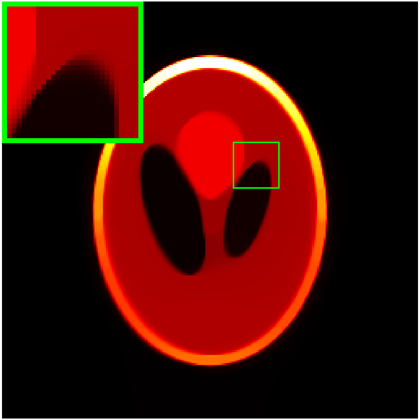
CM, $n = 64^2$, $\lambda = 500$

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

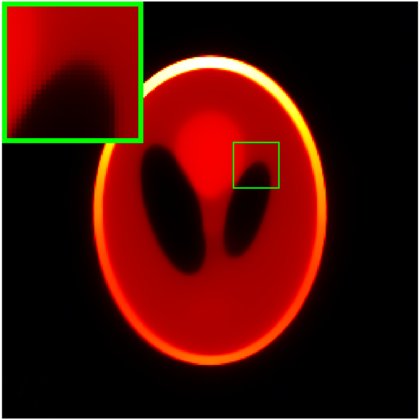
MAP, $n = 128^2$, $\lambda = 500$

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

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



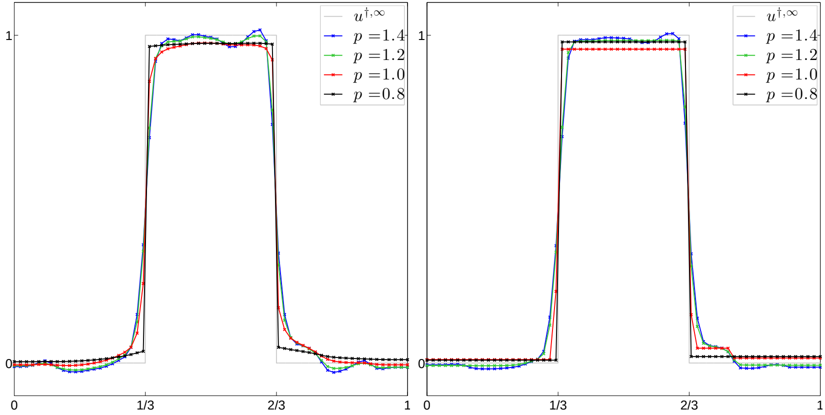
MAP, $n = 256^2$, $\lambda = 500$



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

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

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



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

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:

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

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 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 $\Psi_{\text{LS}}(u, \hat{u}) = \Psi_{\text{Brg}}(u, \hat{u})$!

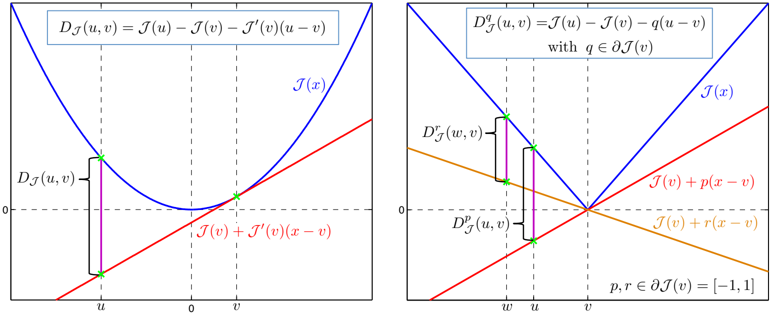
Theorems:

(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

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



(e) $\mathcal{J}(x) = x^2$ (f) $\mathcal{J}(x) = |x|$

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

Variational regularization:

$$\hat{u} = \underset{u}{\operatorname{argmin}} \left\{ \|f - A\,u\|_2^2 + \mathcal{J}(u) \right\}$$

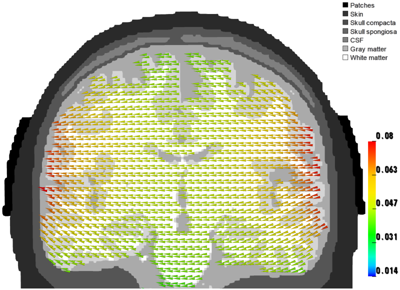
First order optimality condition:

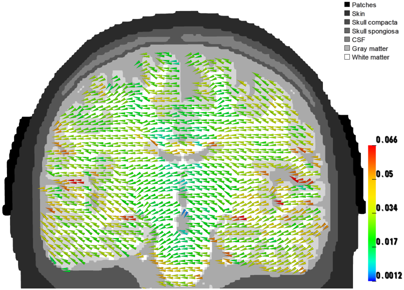
$$-A^T\,(f - A\,\hat{u}) + \mathcal{J}'(\hat{u}) \stackrel{!}{=} 0 \qquad \Longleftrightarrow \qquad \mathcal{J}'(\hat{u}) = A^T\,(f - A\,\hat{u})$$

That means: $\mathcal{J}'(\hat{u}) \in Range(A^T)$. How does $Range(A^T)$ look like?

- A^T is a discretization of the adjoint PDE to EEG / MEG.
- It maps electric potentials / magnetic fields to currents in the brain.
- Essentially solves the tCS / TMS brain stimulation problem.

Vallaghé, Papadopoulo, Clerc, 2009. *The adjoint method for general EEG and MEG sensor-based lead field equations* Phy. Med. Bio.





Wagner, 2015. *Optimizing tCS and TMS multi-sensor setups using realistic head models* PhD Thesis, University of Münster.

See his poster: *"Optimized stimulation protocols in transcranial direct current stimulation"*.

$\mathcal{J}'(\hat{u}) \in Range(A^T) \implies \mathcal{J}'(\hat{u})$ fulfills maximum principle (in continuous limit) and obtains its maximum at the gray matter boundary!

Assume

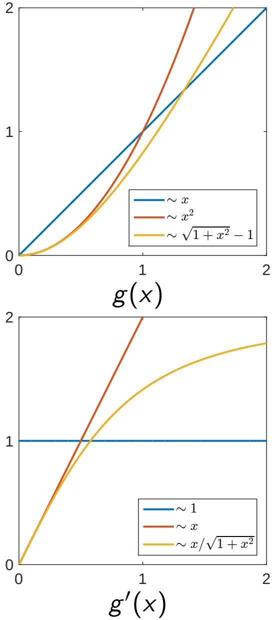
- $\mathcal{J}(u) \propto \sum_i g(|u_i|)$ (uniform in i).
- for simplicity, u is scalar.
- $g(x) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ non-decreasing: $g'(x) \geqslant 0$.

If g is convex, s "inherits" maximum principle:

- $g(x)$ is convex
  $\implies g''(x) \geqslant 0$.
- $g'(x) \geqslant 0$, $g''(x) \geqslant 0$
  $\implies g'(x)$ is positive, non-decreasing.
- $g'(|u_i|) \geqslant g'(|u_j|)$
  $\implies |u_i| \geqslant |u_j|$.
- $(\mathcal{J}'(\hat{u}))_i = g'(|\hat{u}_i|)$ has its maximum on boundary
  $\implies |\hat{u}_i|$ has its maximum at the boundary

$\implies$ Depth bias!

(nothing really changes in the vectorial case; for $g'(0) \neq 0$ or other non-smoothness, we need subdifferential calculus)



Depth Bias: The Blessings of Non-Convexity

Assume

- $\mathcal{J}(u) \propto \sum_i g(|u_i|)$, and that u is scalar.
- $g(x) : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ non-decreasing: $g(x)' \geqslant 0$.

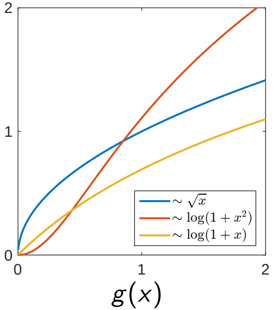
If g is non-convex, $g'(x)$ does not necessarily induce an order and $\hat{u}$ does not need to "inherit" maximum principle!

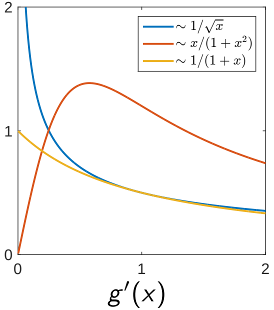
But caution:

- We need to analyze second order optimality condition as well!

Comments:

- Multiple-dipole scans are (extremely) non-convex.
- Heuristic justifies fully-Bayesian inference which preserves and explores the non-convexity.





What About Weightings?

Non-uniform convexity $\mathcal{J}(u) \propto \sum_i g\left(\frac{|u_i|}{w_i(A_i)}\right)$
such as WMNE, WMCE,...

Or post-processing by weighting (noise-normalization):

$$\tilde{u}_i = w_i(\hat{u}_i), \qquad \hat{u} = \underset{u}{\operatorname{argmin}} \left\{ \|f - A\,u\|_2^2 + \mathcal{J}(u) \right\}$$

such as sLORETA, DSPM, ...

Does that help?

- Static weights are often optimized to recover single sources.
- Empirically, sub-optimal for multiple sources (contrary to common misconception).
- Adaptive, iterative weighting often actually optimizes underlying non-convex model.