WESTFÄLISCHE
WILHELMS-UNIVERSITÄT
MÜNSTER

Sparsity Constraints in Bayesian Inversion

"Inverse Days" conference in Jyväskylä, Finland.

wissen.leben
WWU Münster

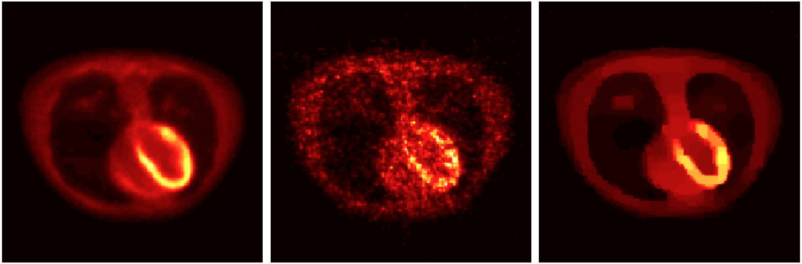
Felix Lucka

19.12.2012

Sparsity Constraints in Inverse Problems

Current trend in high dimensional inverse problems: Sparsity constraints.

- Compressed Sensing: High quality reconstructions from a small amount of data, if a sparse basis/dictionary is a-priori known (e.g., wavelets).
- Total Variation (TV) imaging: Sparsity constraints on the gradient of the unknowns.



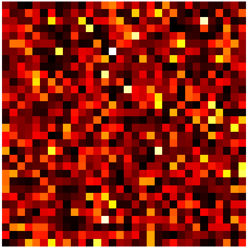
(a) 20 min, EM (b) 5 sec, EM (c) 5 sec, Bregman EM-TV

Thank's to Jahn Müller for these images!

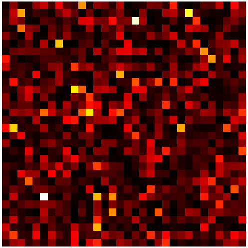
Sparsity Constraints in the Bayesian Approach

Sparsity as a-priori information are encoded into the prior distribution $p_{prior}(u)$:

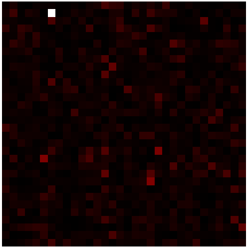
1. Turning the functionals used in variational regularization directly into priors, e.g., L1-type priors:
   - Convenient, as prior is log-concave.
   - MAP estimate is sparse, but the prior itself is not sparse.
2. Hierarchical Bayesian modeling (HBM): Sparsity is introduced at the hyperparameter level.
   - Relies on a slightly different concept of sparsity.
   - Resulting implicit priors over unknowns are usually not log-concave.



(a) $\exp\left(-\frac{1}{2}\|u\|_2^2\right)$

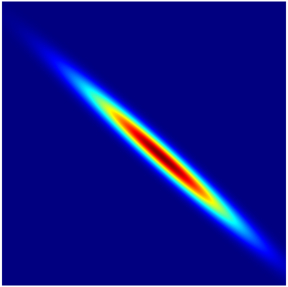


(b) $\exp\left(-|u|_1\right)$

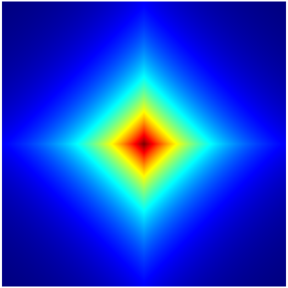


(c) $(1 + u^2/3)^{-2}$

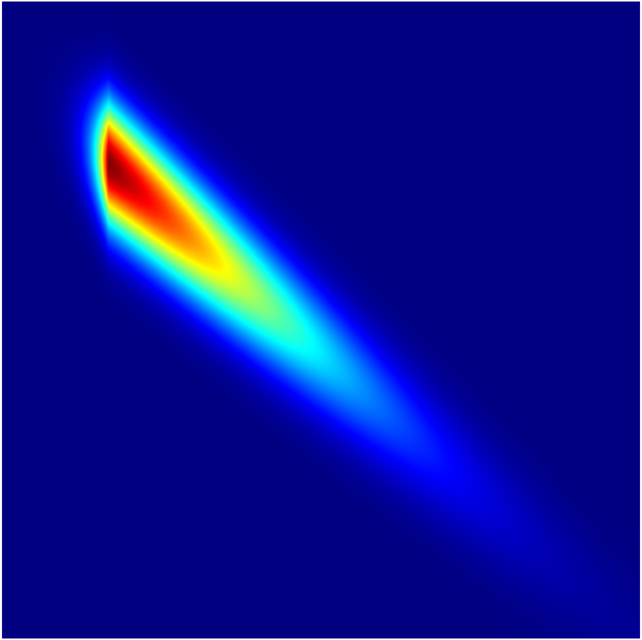
Outline



Likelihood:
$\exp\left(-\frac{1}{2\,\sigma^2}\|f - K\,u\|_2^2\right)$



Prior: $\exp\left(-\lambda\,|u|_1\right)$
(λ via discrepancy principle)



Posterior: $\exp\left(-\frac{1}{2\,\sigma^2}\|f - K\,u\|_2^2 - \lambda\,|u|_1\right)$

L1-type Priors in Bayesian Inversion

- Influence on MAP estimate is well understood (boring).
- Influence on other quantities (CM, credible regions, histograms, sophisticated stuff...): Interesting.
- Interesting scenario for the comparison between variational regularization theory and Bayesian inference!

M. Lassas and S. Siltanen, 2004.
Can one use total variation prior for edge-preserving Bayesian inversion?

M. Lassas, E. Saksman, and S. Siltanen, 2009.
Discretization invariant Bayesian inversion and Besov space priors.

V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, 2012.
Sparsity-promoting Bayesian inversion.

S. Comelli, 2011.
A Novel Class of Priors for Edge-Preserving Methods in Bayesian Inversion
Master thesis, supervision by V. Capasso and M. Burger

Sample and Integration-based Inference for Sparse Bayesian Inversion

Computation of many Bayesian quantities relies on drawing independent samples from posterior.

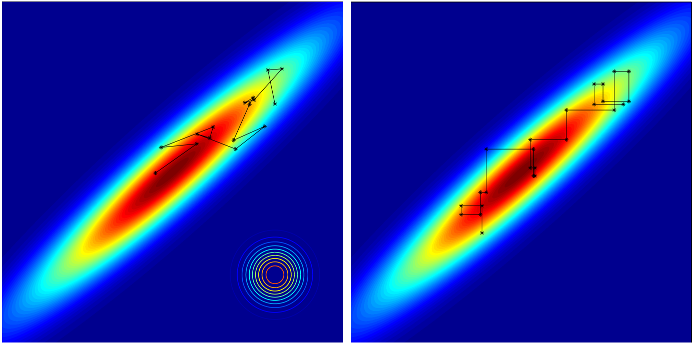
- Most widely used "black box" MCMC sampling algorithm: Metropolis-Hastings (MH) [Metropolis et al., 1953; Hastings, 1970].
- Breaks down for ill-posed inverse problems with sparsity constraints and $n \longrightarrow \infty$.
- Examining $n \longrightarrow \infty$ is interesting.

This talk summarizes partial results from:

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). arXiv:1206.0262v2.

Own Contributions

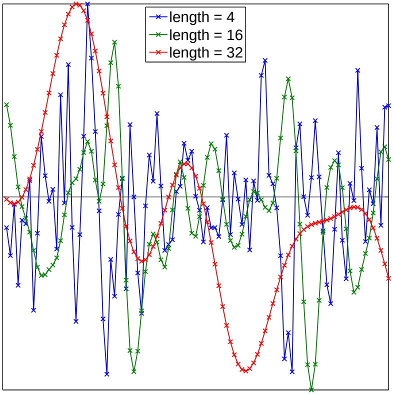
- Single component Gibbs sampling for posteriors from L1-type priors:
  - Fast and explicit computation of 1-dim. cond. densities.
  - Exact and robust sampling from 1-dim densities.
  - Further acceleration by stochastic over-relaxation techniques.
- Detailed comparison of the most basic variants of MH and Gibbs sampling.
- Scenarios with up to 261 121 unknowns.



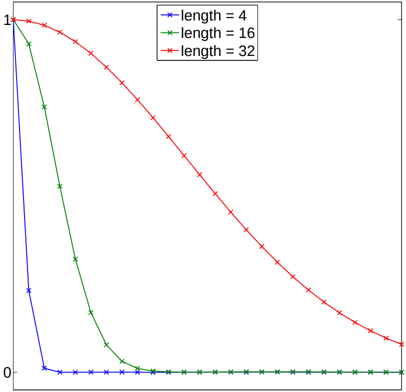
(a) Metropolis Hastings scheme (b) Gibbs scheme

Evaluating Sampling Efficiency by Autocorrelation Functions

- Sampling efficiency: Fast generation of independent samples.
- $R(\tau) \in [0, 1]$: Average correlation between samples x_i, $x_{i+\tau}$ w.r.t. to a test function (rescale by computation time per sample: $R(\tau) \longrightarrow R^*(t)$).
- Rapid decay of $R(\tau)/R^*(t) \Longrightarrow$ Samples get uncorrelated fast!



(c) Stochastic processes...

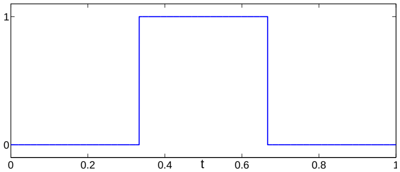


(d) ...and their autocorrelation functions

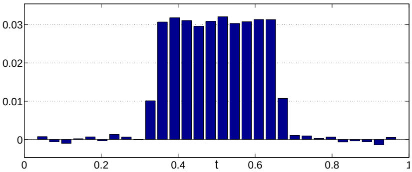
Total Variation Deblurring Example in 1D (from Lassas & Siltanen, 2004)

- Model of a charge coupled device (CCD) in 1D.
- Unknown light intensity $\tilde{u} : [0, 1] \rightarrow \mathbb{R}^+$, indicator on $[\frac{1}{3}, \frac{2}{3}]$.
- Integrated into $k = 30$ CCD pixels $[\frac{1}{k+2}, \frac{k+1}{k+2}] \subset [0, 1]$.
- Noise is added.
- $\tilde{u}$ is reconstructed on a regular, n-dim. grid.
- D is the forward finite difference operator with NB cond.

$$p_{post}(u|f) \propto \exp\left(-\frac{1}{2\,\sigma^2}\|f - K\,u\|_2^2 - \lambda\,|D\,u|_1\right)$$



(a) The unknown function $\tilde{u}(t)$



(b) The measurement data f

Total Variation Deblurring Example in 1D (from Lassas & Siltanen, 2004)

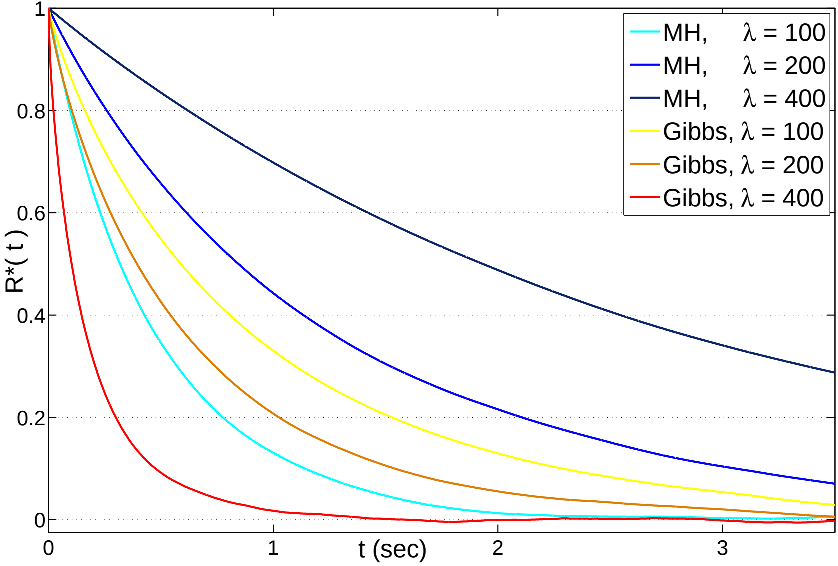


Figure: Temporal autocorrelation plots $R^*(t)$ for $n = 63$.

Total Variation Deblurring Example in 1D (from Lassas & Siltanen, 2004)

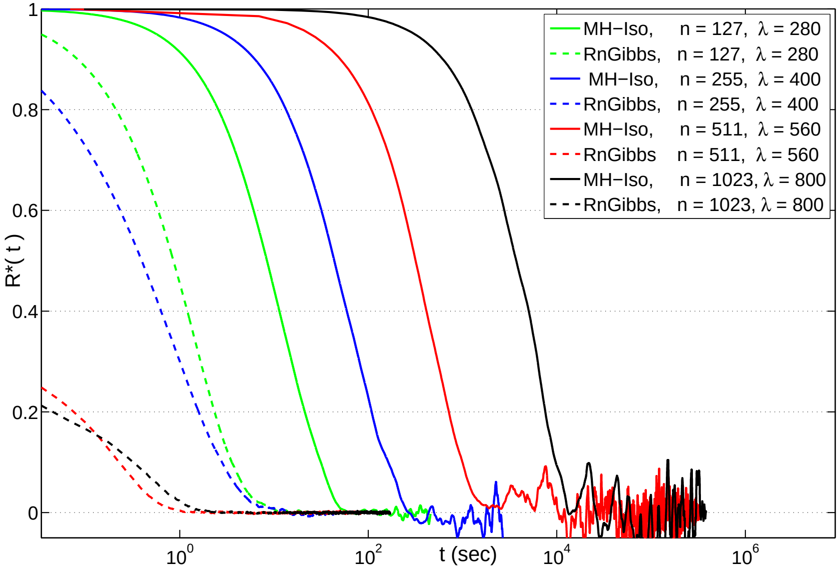


Figure: Temporal autocorrelation plots $R^*(t)$.

Total Variation Deblurring Example in 1D (from Lassas & Siltanen, 2004)

New sampler can be used to address theoretical questions:

- Lassas & Siltanen, 2004: For $\lambda_n \propto \sqrt{n+1}$, the TV prior converges to a smoothness prior in the limit $n \longrightarrow \infty$.
- MH sampling to compute CM estimate for $n = 63, 255, 1023, 4095$.
- Even after a month of computation time only partly satisfying results.

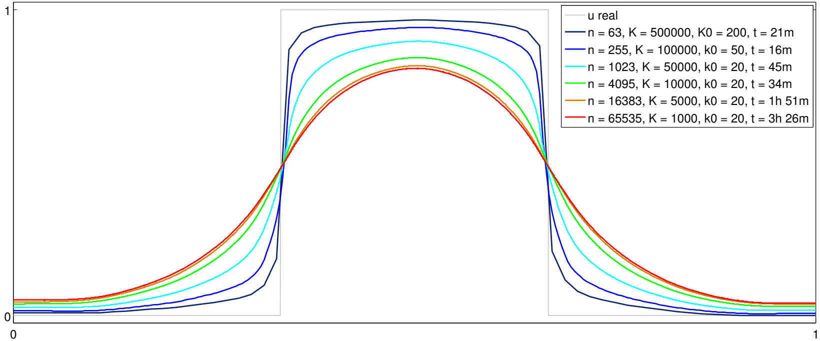
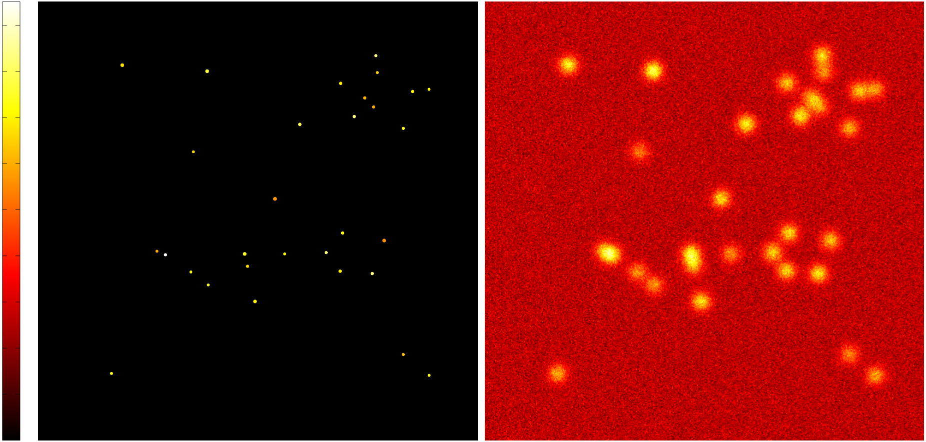


Figure: CM estimate computed for $n = 63, 255, 1023, 4095, 16383, 65535$ using Gibbs sampler on a comparable CPU.

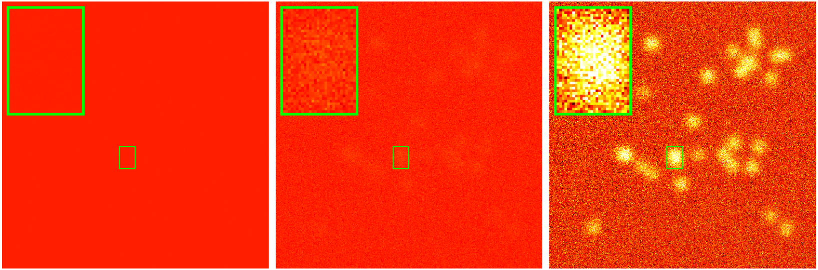
Image Deblurring Example in 2D



Unknown function $\tilde{u}$ Measurement data m

- Gaussian blurring kernel
- Relative noise level of 10%
- Reconstruction using $n = 511 \times 511 = 261\,121$.

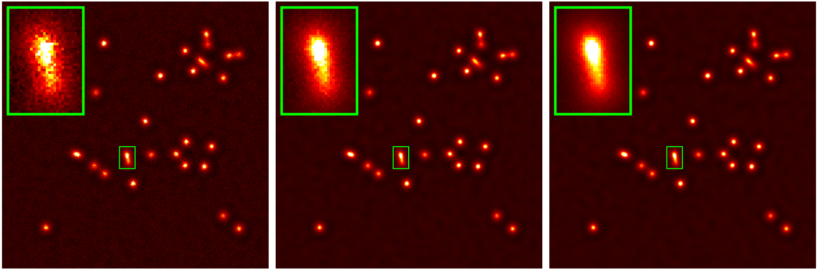
Image Deblurring Example in 2D



(a) 1h comp. time (b) 5h comp. time (c) 20h comp. time

Figure: CM estimates by MH sampler

Image Deblurring Example in 2D



(a) 1h comp. time (b) 5h comp. time (c) 20h comp. time

Figure: CM estimates by Gibbs sampler

Take Home Messages, Part I

- High dimensional inverse problems using sparsity constraints pose specific challenges for MCMC.
- MH and Gibbs may show very different performance.
- Contrary to common beliefs, MCMC is not in general slow and scales bad with increasing dimension!
- MCMC for inverse problems is far less elaborate compared to optimization up to now.
- Sample-based inference in high dimensions is feasible ($n > 10^6$ works).

CAUTION:

- These results do not generalize!
- MH and Gibbs sampling have both pro's and con's!

Outline

Cooperation with...

Aalto University
School of Science

Dr. Sampsa Pursiainen
Department of Mathematics and Systems Analysis,
Aalto University, Finland

Applied Mathematics Münster

Prof. Dr. Martin Burger
Institute for Applied Mathematics,
University of Münster, Germany

IBB
Institute for Biomagnetism and Biosignalanalysis

PD. Dr. Carsten Wolters
Institute for Biomagnetism and Biosignalanalysis,
University of Münster, Germany

Background of this Part of the Talk

Felix Lucka.
Hierarchical Bayesian Approaches to the Inverse Problem of EEG/MEG Current Density Reconstruction.
Diploma thesis in mathematics, University of Münster, March 2011

Felix Lucka., Sampsa Pursiainen, Martin Burger,Carsten H. Wolters.
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), 2012.

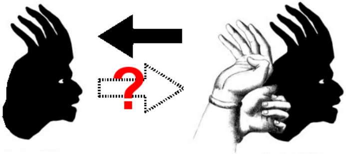
Source Reconstruction by Electroencephalography (EEG) and Magnetoencephalography (MEG)

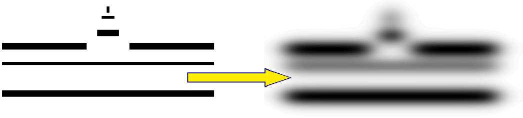
Aim: Reconstruction of brain activity by non-invasive measurement of induced electromagnetic fields (bioelectromagnetism) outside of the skull.

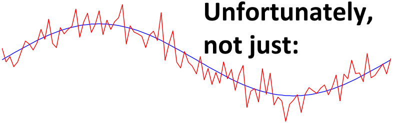
source: Wikimedia Commons

source: Wikimedia Commons

One Challenge of Source Reconstruction: The Inverse Problem







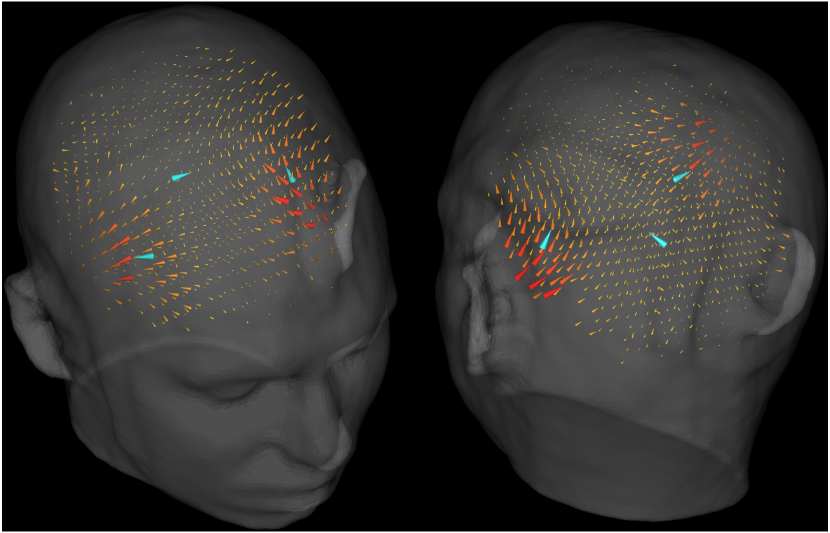
Unfortunately, not just:

- (Presumably) under-determined
- Severely ill-conditioned, special spatial characteristics.
- Signal is contaminated by a complex spatio-temporal mixture of external and internal noise and nuisance sources.

Discretization Approach: Current Density Reconstruction (CDR)

Continuous (ion current) vector field ≈ Spatial grid with 3 orthogonal elementary sources at each node (in general more sophisticated).

$$f = K\,u, \qquad \Longrightarrow \qquad p_{li}(f|u) \propto \exp(-\tfrac{1}{2}\|\Sigma_\varepsilon^{-1/2}(f - K\,u)\|_2^2)$$



Tasks and Problems for EEG/MEG in Presurgical Epilepsy Diagnosis

EEG/MEG in epileptic focus localization:

- *Focal epilepsy* is believed to originate from networks of focal sources.
- Active in inter-ictal spikes.
- Task 1: Determine number of focal sources (*multi focal epilepsy*?).
- Task 2: Determine location and extend of sources.

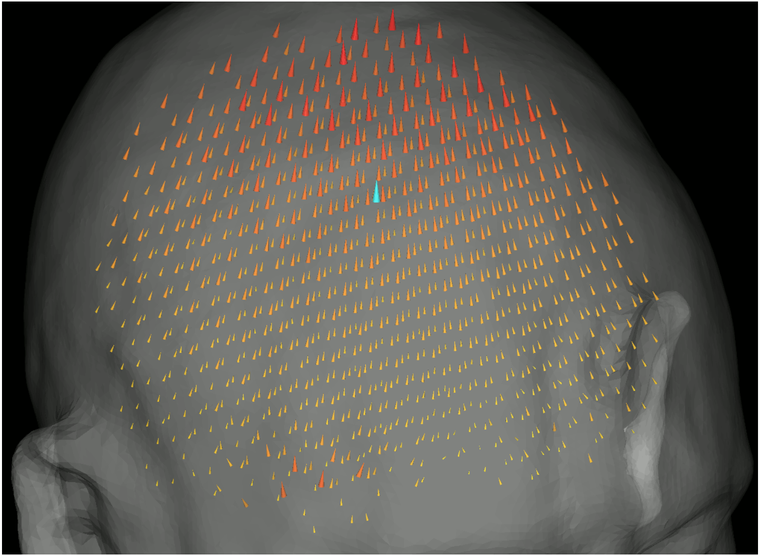
Problems of established CDR methods:

- Depth-Bias: Reconstruction of deeper sources too close to the surface.
- Masking: Near-surface sources “mask“ deep-lying ones.

Depth Bias: Illustration

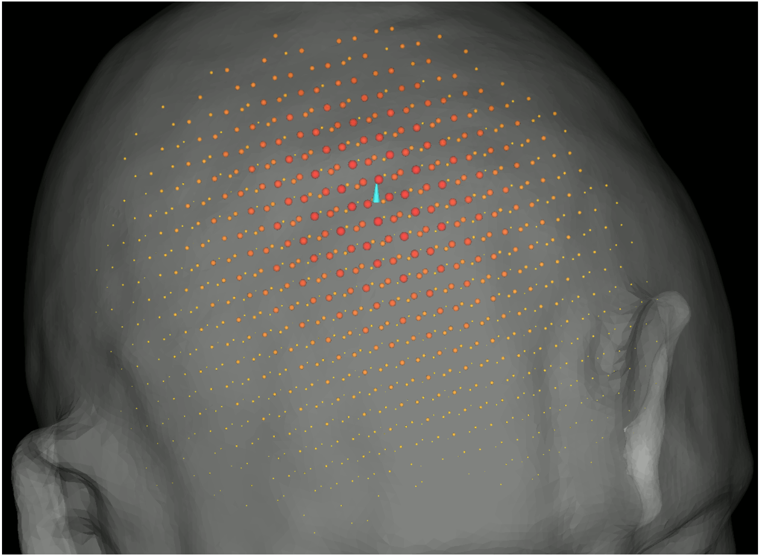
One deep-lying reference source (blue cone) and minimum norm estimate:

$$u_{\text{MNE}} = \operatorname{argmin}\{\|\Sigma_{\varepsilon}^{-1/2}(f - K\,u)\|_2^2 + \lambda\|u\|_2^2\}$$



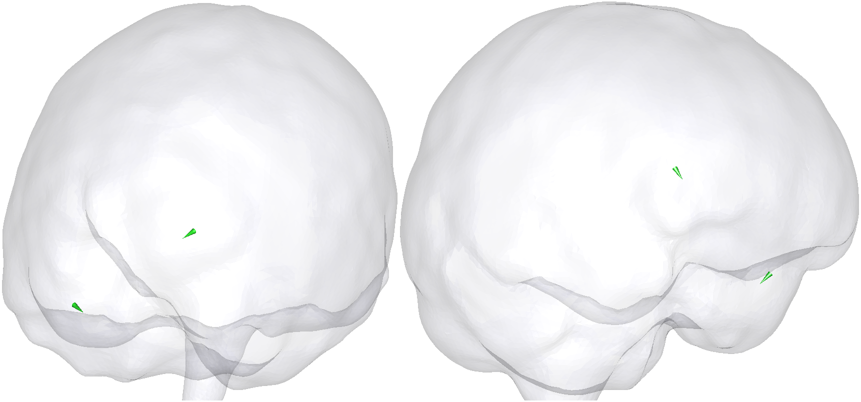
Depth Bias: Illustration

Reweighting minimum norm estimate (sLORETA, Pascual-Marqui, 2002).



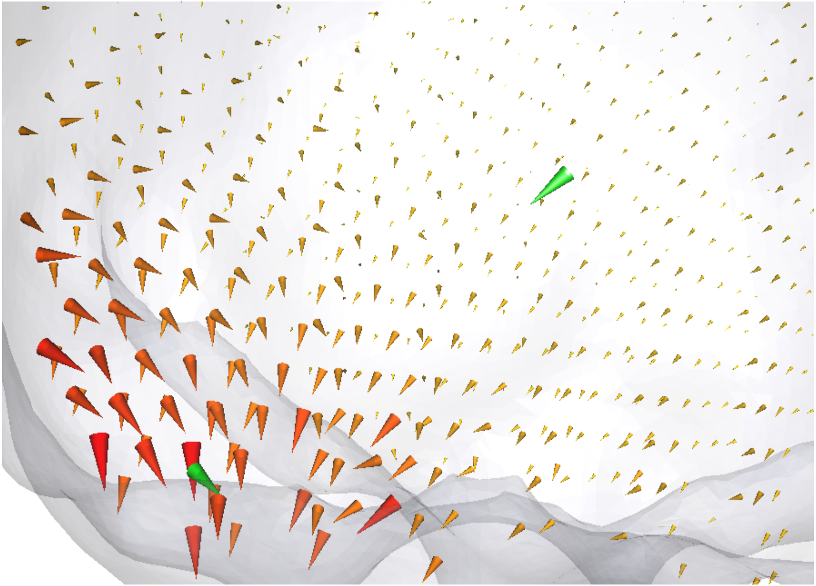
Masking: Illustration

Reference sources.



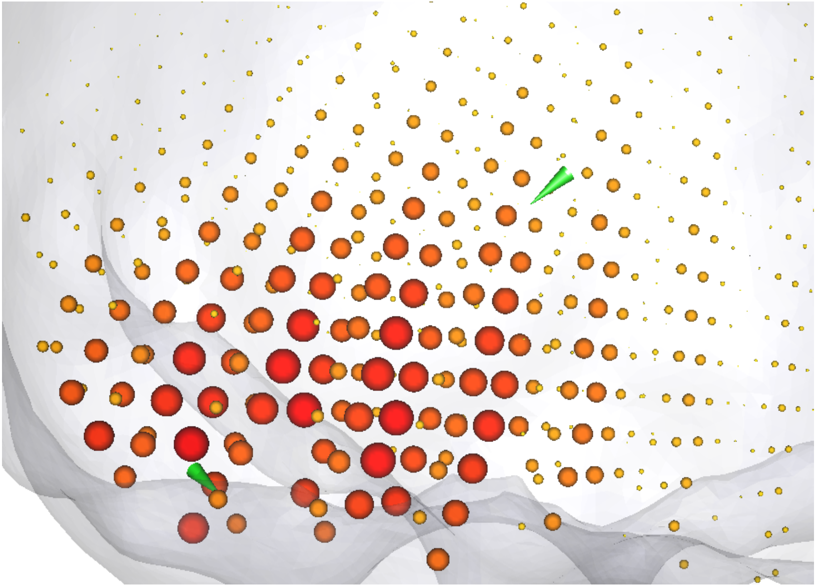
Masking: Illustration

MNE result and reference sources (green cones).



Masking: Illustration

sLORETA result and reference sources (green cones).



Basic Ideas of Sparsity Promoting HBM

Gaussian prior with fixed, uniform diagonal covariance matrix.

$$p_{pr}(u) \sim \mathcal{N}(0, \gamma \cdot \mathrm{Id})$$

Basic Ideas of Sparsity Promoting HBM

Gaussian prior with fixed, uniform diagonal covariance matrix.

$$p_{pr}(u) \sim \mathcal{N}(0, \gamma \cdot \mathrm{Id})$$

Advantages:

- Posterior is Gaussian.
- Simple, well understood analytical structure.

Problems:

- Gaussian variables = characteristic scale given by variance. (not *scale invariant*)
- All sources have variance γ

$\Longrightarrow$ Similar amplitudes are likely.

$\Longrightarrow$ Sparse solutions are very unlikely.

Basic Ideas of Sparsity Promoting HBM

Gaussian prior with flexible, individual diagonal covariance matrix:

$$p_{pr}(u|\gamma) \sim \mathcal{N}\left(0, \operatorname{diag}\left[\gamma_1, , \ldots, \gamma_n\right]\right)$$

- Let the data determine γ_i (hyperparameters).
- Bayesian inference: $\boldsymbol{\gamma}$ are random variables as well.
- Their prior distribution $p_{hypr}(\boldsymbol{\gamma})$ is called hyperprior.
- Encode sparsity constraints into hyperprior.

Basic Ideas of Sparsity Promoting HBM

Gaussian prior with flexible, individual diagonal covariance matrix:

$$p_{pr}(u|\gamma) \sim \mathcal{N}(0, \operatorname{diag}[\gamma_1, , \ldots, \gamma_n])$$

and heavy-tailed, non-informative, i.e., scale invariant prior on γ_i, e.g.,

$$p_{hypr}(\gamma_i) \propto \gamma_i^{-1}$$

to promote sparsity on the hyperparameter level.

Basic Ideas of Sparsity Promoting HBM

Gaussian prior with flexible, individual diagonal covariance matrix:

$$p_{pr}(u|\gamma) \sim \mathcal{N}\left(0, \operatorname{diag}\left[\gamma_1, , \ldots, \gamma_n\right]\right)$$

and heavy-tailed, non-informative, i.e., scale invariant prior on γ_i, e.g.,

$$p_{hypr}(\gamma_i) \propto \gamma_i^{-1}$$

to promote sparsity on the hyperparameter level.

Problem: Improper prior, improper posterior.

Basic Ideas of Sparsity Promoting HBM

Gaussian prior with flexible, individual diagonal covariance matrix:

$$p_{pr}(u|\gamma) \sim \mathcal{N}\left(0, \operatorname{diag}\left[\gamma_1, , \ldots, \gamma_n\right]\right)$$

and heavy-tailed, weakly-informative prior on γ_i, e.g., inverse gamma distribution:

$$\gamma_i^{-1} \longrightarrow \propto \gamma_i^{-(1+\alpha)} \exp\left(-\frac{\beta}{\gamma_i}\right)$$

to promote sparsity on the hyperparameter level.

Advantage: Conjugate hyperprior, computationally convenient.

Basic Ideas of Sparsity Promoting HBM

Gaussian prior with flexible, individual diagonal covariance matrix:

$$p_{pr}(u|\gamma) \sim \mathcal{N}\left(0, \operatorname{diag}\left[\gamma_1, , \ldots, \gamma_n\right]\right)$$

and heavy-tailed, weakly-informative prior on γ_i, e.g., inverse gamma distribution:

$$\gamma_i^{-1} \longrightarrow \propto \gamma_i^{-(1+\alpha)} \exp\left(-\frac{\beta}{\gamma_i}\right)$$

to promote sparsity on the hyperparameter level.

Advantage: Conjugate hyperprior, computationally convenient.

- By the direct correspondence, we get sparsity over the primary unknowns u as well.
- Generalization: $\operatorname{diag}\left[\gamma_1, , \ldots, \gamma_n\right] \longrightarrow \sum \gamma_i C_i$ (Gaussian scale mixture models)
- C_i to model complex spatio-temporal covariance structures.

Some Basic Features of Sparsity Promoting HBM

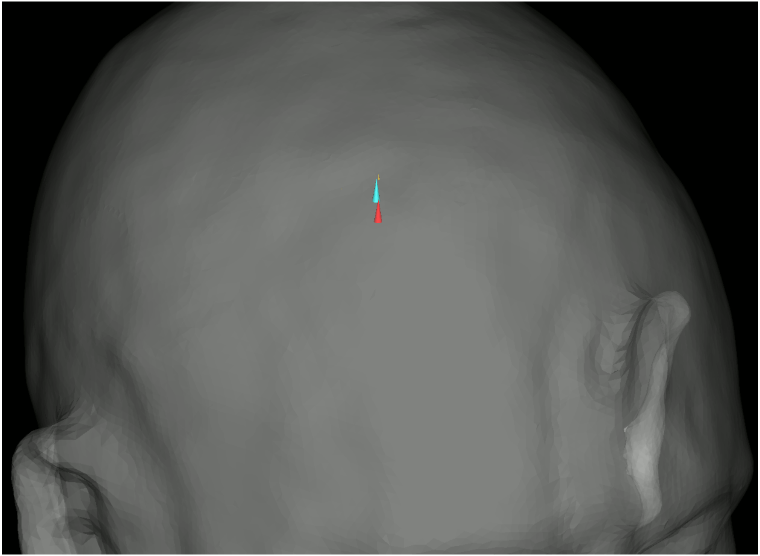
Posterior:

$$p_{post}(u, \gamma | f) \propto$$
$$\exp\left(-\frac{1}{2} \|\Sigma_{\varepsilon}^{-1/2}(f - K\,u)\|_2^2 - \sum_{i=1}^{k} \left(\frac{\frac{1}{2}\|u_{i*}\|^2 + \beta}{\gamma_i} + (\alpha + \tfrac{5}{2}) \ln \gamma_i \right) \right)$$

- Gaussian with respect to u.
- Factorizes over γ_i's.
- Energy is non-convex w.r.t. (u, γ) (posterior is multimodal).

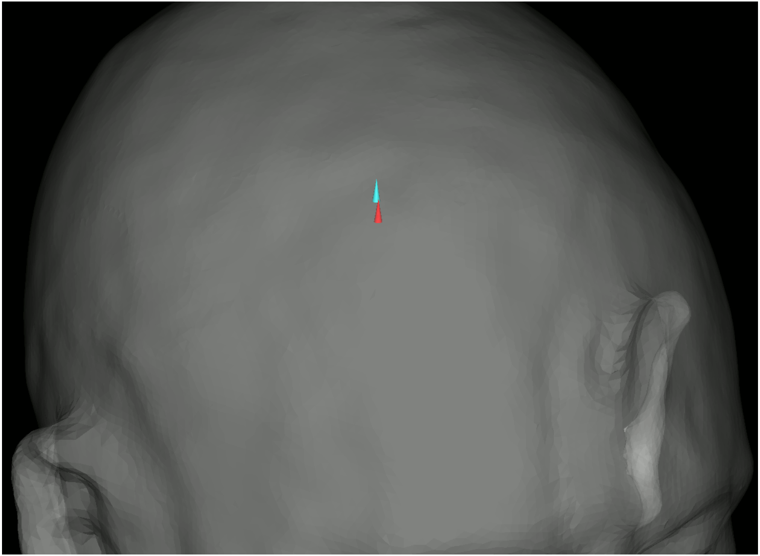
Depth Bias: Full-CM

Computed by blocked Gibbs sampler.



Depth Bias: Full-MAP

Computed by alternating optimization initialized at the CM estimate.

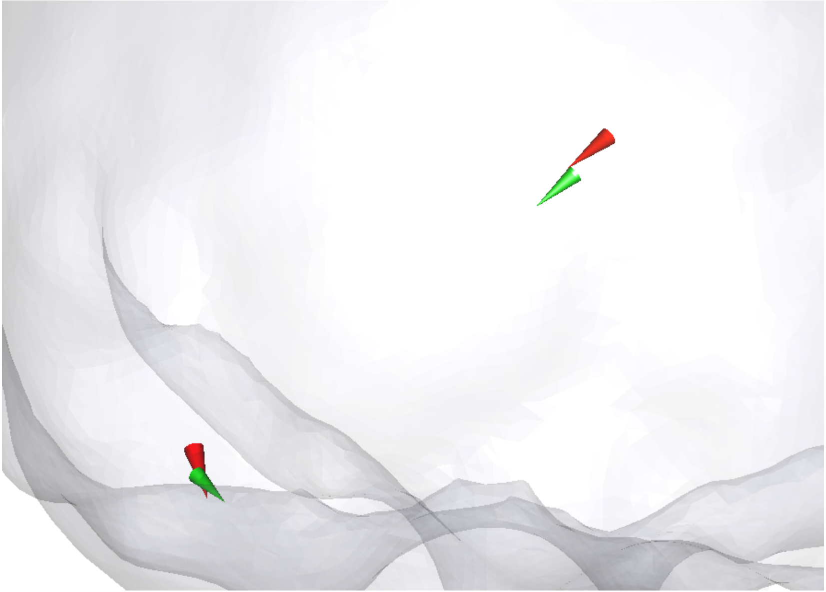


Masking: Result Full-CM

Computed by blocked Gibbs sampler.

Masking: Result Full-MAP

Computed by alternating optimization initialized at the CM estimate.



Contributions of our Studies

Elaborate up on:

Daniela Calvetti, Harri Hakula, Sampsa Pursiainen, Erkki Somersalo, 2009.
Conditionally Gaussian hypermodels for cerebral source localization

- Implementation of Full-MAP and Full-CM inference for HBM with realistic, high resolution Finite Element (FE) head models.
- Improve algorithms for Full-MAP estimation.
- Examination of general properties, parameter choices, etc.
- Systematic examination of performance concerning depth-bias and masking in simulation studies.
- Introduction of suitable performance measures for validation of simulation studies (Wasserstein distances).

Systematic Studies: Summary

Results for Full-MAP and Full-CM estimation:

- Good performance in all validation measures.
- No depth bias.
- Good results w.r.t. orientation, amplitude and spatial extend.
- Full-MAP estimate (by our algorithm): Best results in every aspect examined.

Full results:

Felix Lucka., Sampsa Pursiainen, Martin Burger,Carsten H. Wolters.
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), 2012.

Take Home Messages: Hierarchical Bayesian Modeling

- Current trend in all areas of Bayesian inference.
- Extension of the prior model by hyperparameters γ and hyperpriors.
- Gaussian w.r.t. u, factorization w.r.t. γ, sparse hyperprior.
- Alternative formulation of sparse Bayesian inversion that has promising, interesting features for EEG/MEG source reconstruction:
  - No depth-bias.
  - Good source separation
  - Non-convex energy (multimodal posterior) but the possibility to infer the support from CM estimate.

Thank you for your attention!

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), 2012; arXiv:1206.0262v2.

F. L., Sampsa Pursiainen, Martin Burger,Carsten H. 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), 2012.