Westfälische Wilhelms-Universität Münster

Hierarchical Bayesian Modeling and Another Type of Sparsity

Applied Math Colloquium, UCLA

wissen.leben
WWU Münster

Felix Lucka

05-29-2013

Outline

Motivation: Depth Bias in EEG/MEG Source Reconstruction.

A Sparsity-Promoting Hierarchical Bayesian Model for EEG/MEG

Two Roads to Sparsity: ℓ_p vs. Hierarchical Bayesian Modeling

Take Home Messages & Conclusions

wissen.leben
WWU Münster

Electroencephalography (EEG) and Magnetoencephalography (MEG)

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

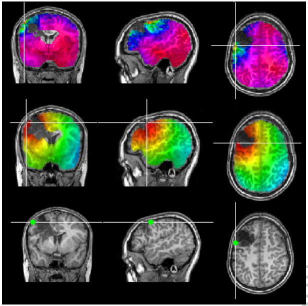
source: Wikimedia Commons

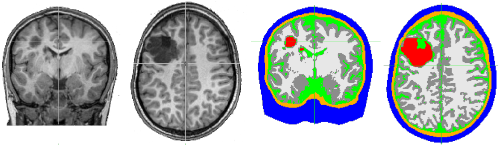
source: Wikimedia Commons

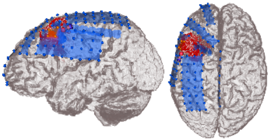
wissen.leben
WWU Münster

Applications of EEG/MEG

- Diagnostic tool in neurology, e.g., epilepsy.
- Examination tool in several fields of neuroscience.
- A broad separation can be made into
  - Sensor-level analysis
  - Source reconstruction



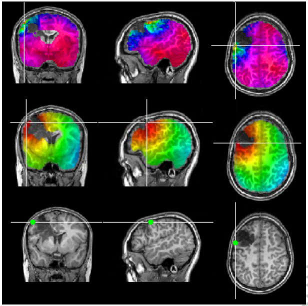


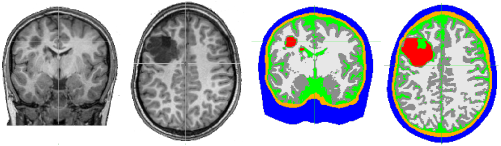


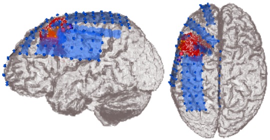
wissen.leben
WWU Münster

Applications of EEG/MEG

- Diagnostic tool in neurology, e.g., epilepsy.
- Examination tool in several fields of neuroscience.
- A broad separation can be made into
  - Sensor-level analysis
  - Source reconstruction







wissen.leben
WWU Münster

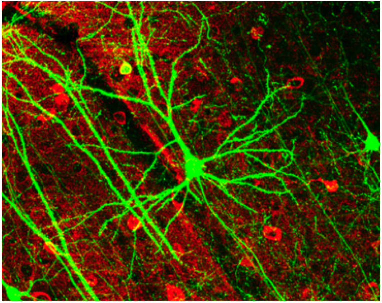
Challenges of Source Reconstruction: Mathematical Modeling

Mathematical modeling of bioelectromagnetism:

- Understand and model the transformation of the bio-chemical activity of the brain into ionic currents.
- Find reasonable simplifications to Maxwell's equations to formulate forward equations that relate ionic currents to measured signals:

$$\nabla \cdot (\sigma \nabla \phi) = \nabla \cdot \vec{j}^{pri} \quad + BC$$

- σ: volume conductor model

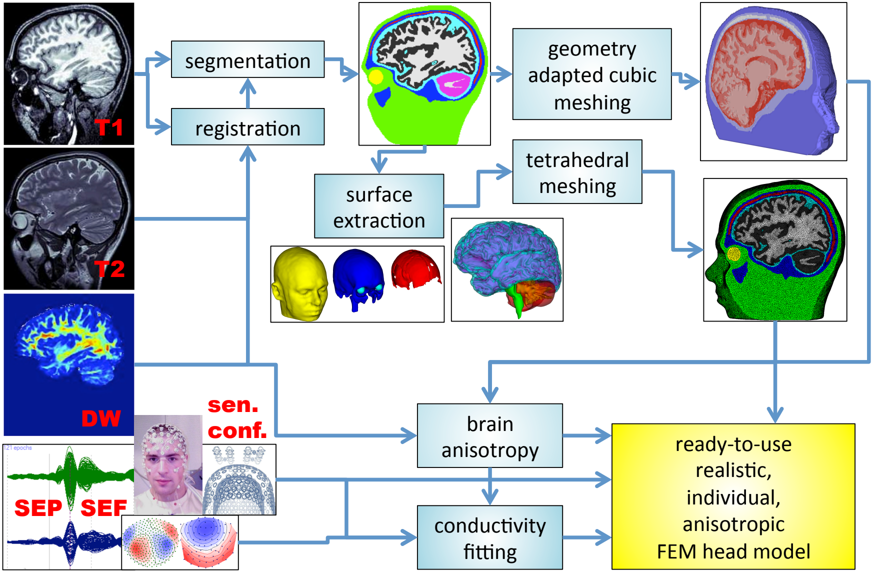


source: Wikimedia Commons

wissen.leben WWU Münster

Challenges of Source Reconstruction: Head Modeling

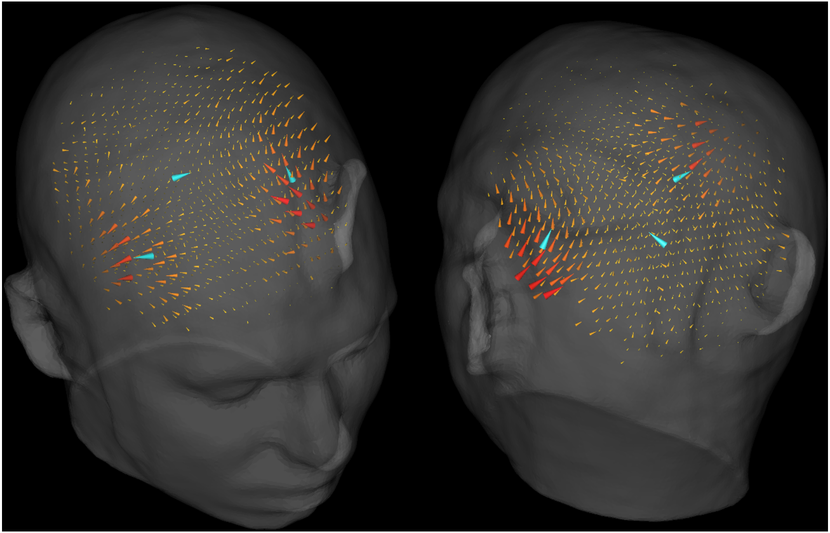
Realistic and individual head models for simulating the forward equations.



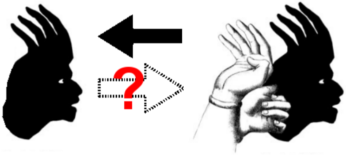
Discretization Approach: Current Density Reconstruction (CDR)

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

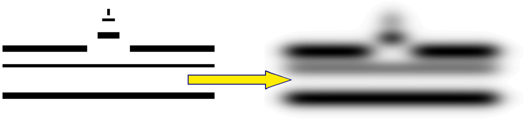
$$f = K\,u, \qquad \Longrightarrow \qquad p_{like}(f|u) \propto \exp(-\tfrac{1}{2}\|\Sigma_\varepsilon^{-1/2}(f - K\,u)\|_2^2)$$



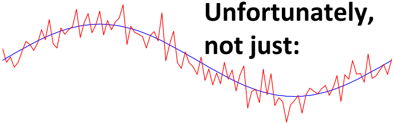
Challenges of Source Reconstruction: Inverse Problem



- (Presumably) under-determined



- Severely ill-conditioned



Unfortunately, not just:

- Signal is contaminated by a complex spatio-temporal mixture of external and internal noise and nuisance sources.

wissen.leben
WWU Münster

Specific Source Scenario: 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:

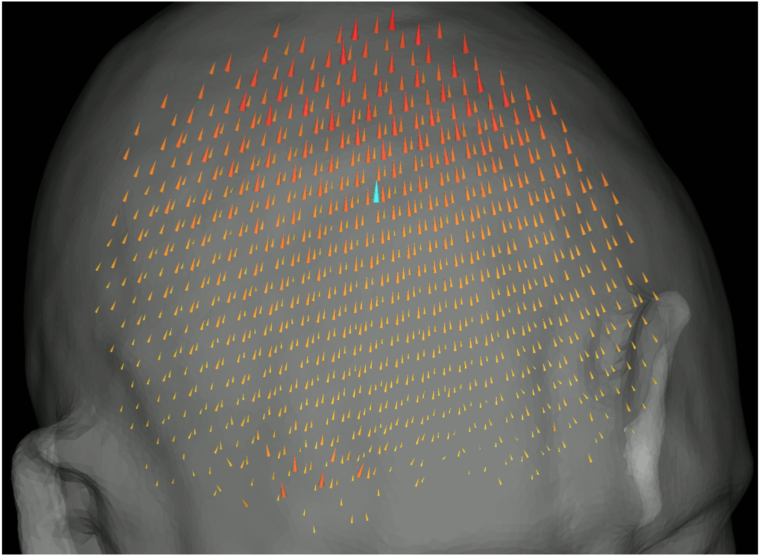
- Depth-Bias: Reconstruction of deeper sources too close to the surface.
- Masking: Near-surface sources "mask" deep-lying ones.

wissen.leben
WWU Münster

Depth Bias: Illustration

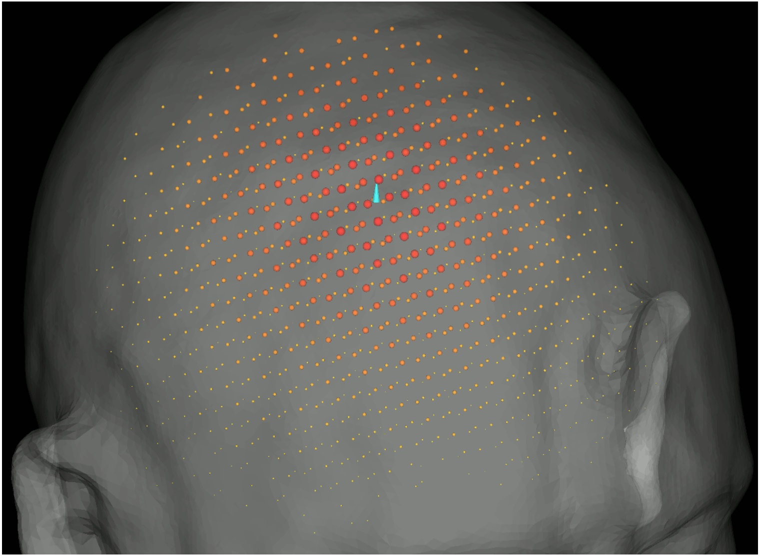
One deep-lying reference source (blue cone) and minimum norm estimate:

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



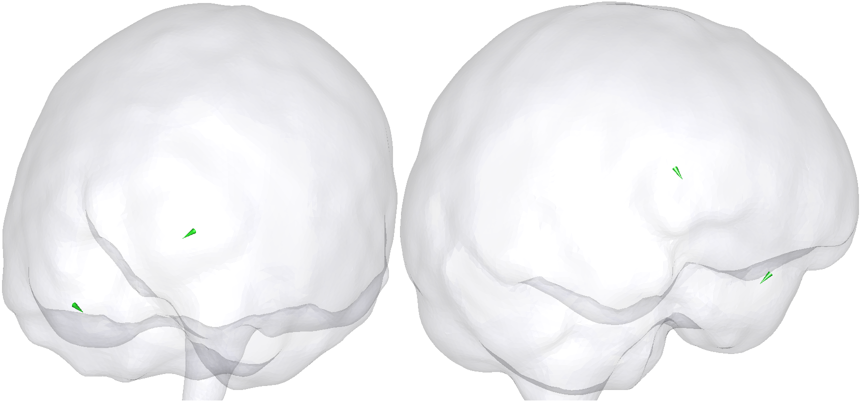
Depth Bias: Illustration

One deep-lying reference source (blue cone) and sLORETA result (Pascual-Marqui, 2002).



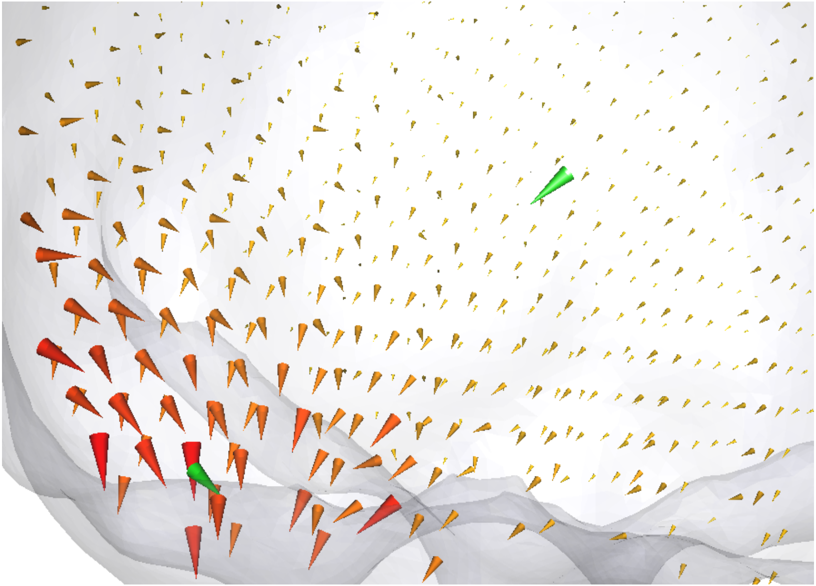
Masking: Illustration

Reference sources.



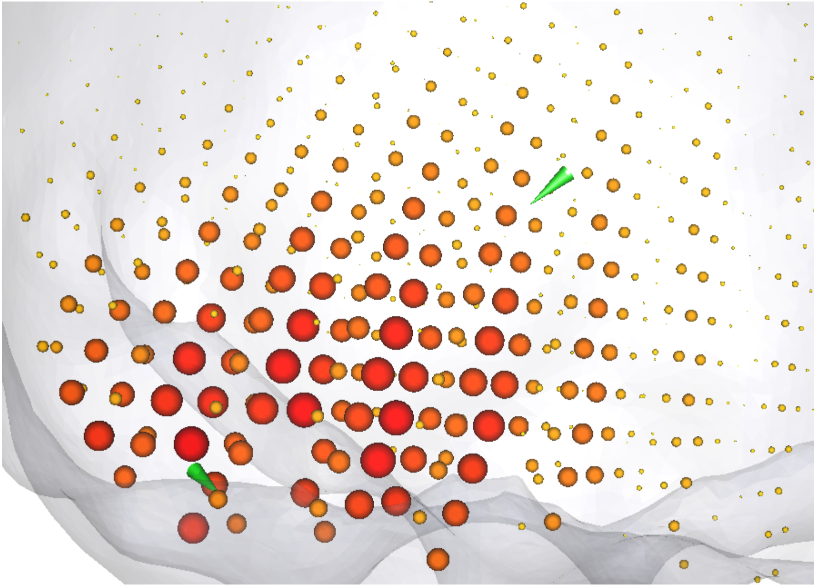
Masking: Illustration

MNE result and reference sources (green cones).



Masking: Illustration

sLORETA result and reference sources (green cones).



Problems of Classical Inverse Methods: Depth-Bias

- Using normal ℓ_2 and ℓ_1 type priors/regularizers: Depth-bias.
- Heuristic reason: Deep sources have weaker signal; Signal of single deep source can be generated by extended patch of near-surface sources.
- Theoretical reason in simplified EEG example:
  $q \in \partial \mathcal{J}(\hat{u})$ is a solution of a Laplace equation with Neumann BC
  $\Longrightarrow$ harmonic functions
  $\Longrightarrow$ maximum principle:
  - ℓ_2: $\hat{u}$ is harmonic $\Rightarrow$ maximum at boundary.
  - ℓ_1: $\text{sign}(\hat{u})$ is harmonic $\Rightarrow$ supported only at boundary.

wissen.leben
WWU Münster

Problems of Classical Inverse Methods: Depth-Bias

Introducing weighted norms ($\|u\|_2^2 \longrightarrow \|Wu\|_2^2$) to give deep sources an advantage.

- Partly solves depth-bias.
- Other draw-backs, e.g., larger spatial blurring $\Longrightarrow$ worse source separation.
- Critical from the Bayesian point of view: Would mean that deep sources usually have a stronger signal $\Longrightarrow$ unphysiological a-priori information.

Reweighting of the solution (e.g., sLORETA) also leads to problems w.r.t. source separation.

wissen.leben
WWU Münster

Outline

wissen.leben WWU Münster

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 the Talk

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.

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

wissen.leben
WWU Münster

Gentle Introduction to Sparsity Promoting HBMs

Wanted: A prior promoting sparse (focal) source activity.

First try:

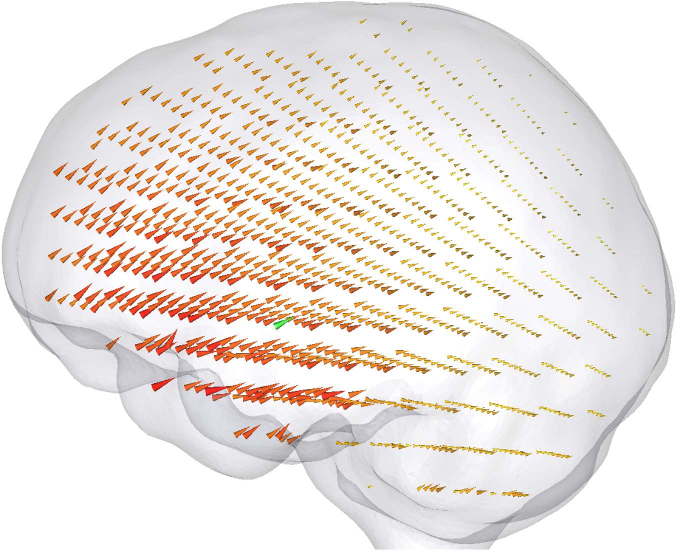
- Gaussian prior with fixed, uniform diagonal covariance $\Sigma_u = \gamma \cdot \mathrm{Id}$ (*Minimum norm estimation*).
- Compute MAP or CM estimate (equal)!

$$\hat{u}_{\mathrm{MAP}} := \underset{u \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ \exp\left(-\frac{1}{2\,\sigma^2} \|f - K\,u\|_2^2 - \frac{1}{2\gamma} \|u\|_2^2 \right) \right\}$$
$$= \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \|f - K\,u\|_2^2 + \frac{\sigma^2}{\gamma} \|u\|_2^2 \right\}$$

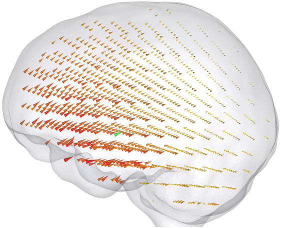
- Simple, well understood analytical structure.

wissen.leben
WWU Münster

First try: NOT a focal reconstruction.



Gentle Introduction to Sparsity Promoting HBMs



What went wrong?

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

$\Longrightarrow$ Similar amplitudes are likely.

$\Longrightarrow$ Focal activity is very unlikely.

wissen.leben
WWU Münster

Gentle Introduction to Sparsity Promoting HBMs

Idea: Gaussian prior with flexible, individual diagonal covariance:

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

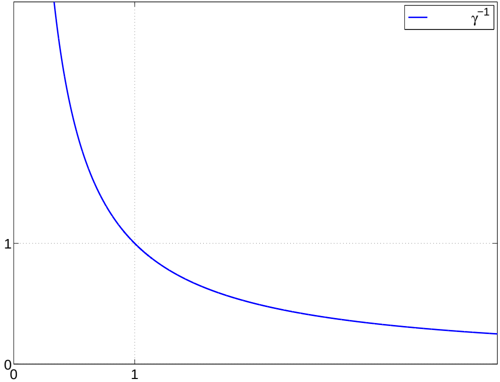
- Let the data determine γ_i (hyperparameters).
- Bayesian inference: γ are random variables as well.
- Their prior distribution $p_{hypr}(\gamma)$ is called hyperprior.
- Encode sparsity constraints into hyperprior $\rightsquigarrow$ by direct correspondence, we might get sparsity over the primary unknowns u as well.
- Generalization: $\mathrm{diag}[\gamma_1, , \ldots, \gamma_n] \longrightarrow \sum \gamma_i C_i$

wissen.leben WWU Münster

Gentle Introduction to Sparsity Promoting HBMs

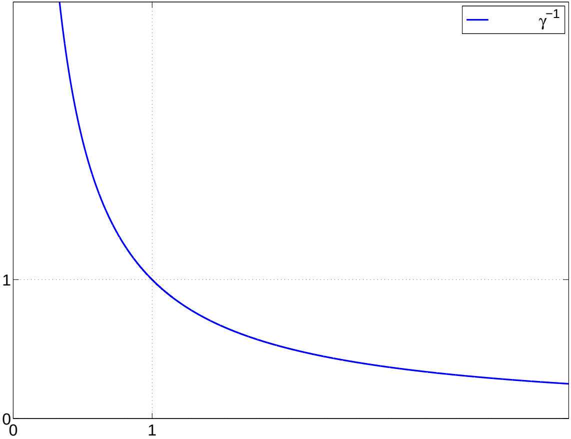
How to promote sparsity on hyperparameter (= scale variables) level?

$\longrightarrow$ Heavy-tailed, non-informative, i.e., scale invariant prior on γ_i!



wissen.leben WWU Münster

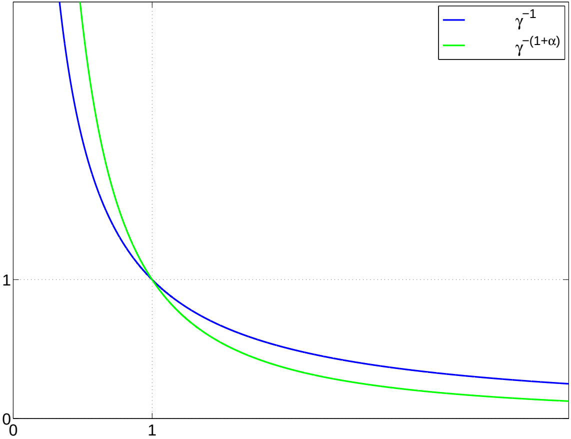
Gentle Introduction to Sparsity Promoting HBMs



Problem: Improper prior, improper posterior.

wissen.leben
WWU Münster

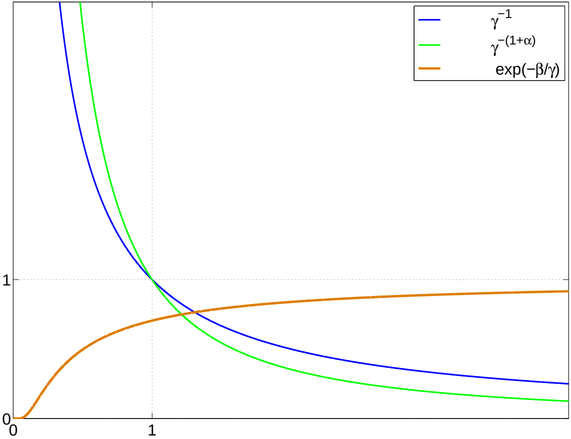
Gentle Introduction to Sparsity Promoting HBMs



make decay faster

wissen.leben
WWU Münster

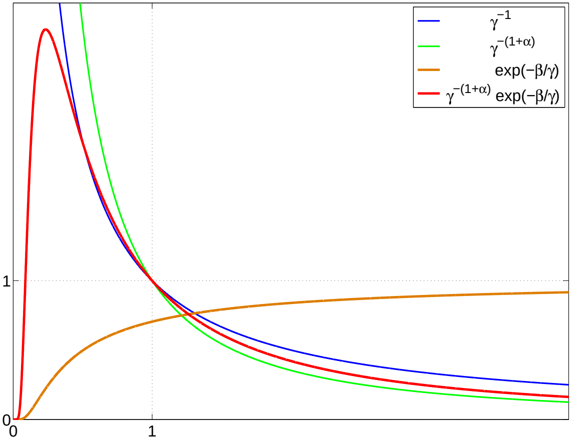
Gentle Introduction to Sparsity Promoting HBMs



multiply by mollifier

wissen.leben
WWU Münster

Gentle Introduction to Sparsity Promoting HBMs



Inverse gamma distribution: Conjugate hyperprior, computationally convenient.

wissen.leben
WWU Münster

Gentle Introduction to Sparsity Promoting HBMs

Posterior:

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

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

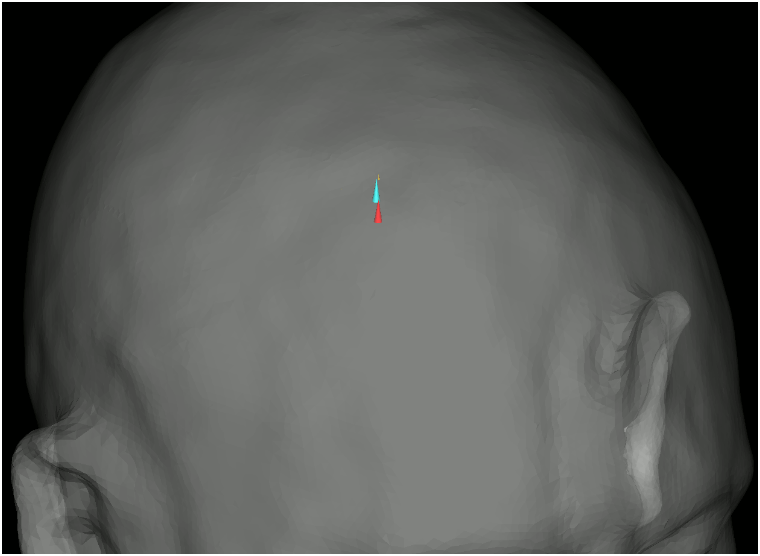
Bayesian Inference: Exploit information in posterior!

Focus of our work: Fully Bayesian inference (in constrast to, e.g. Variational Bayesian approaches).

wissen.leben WWU Münster

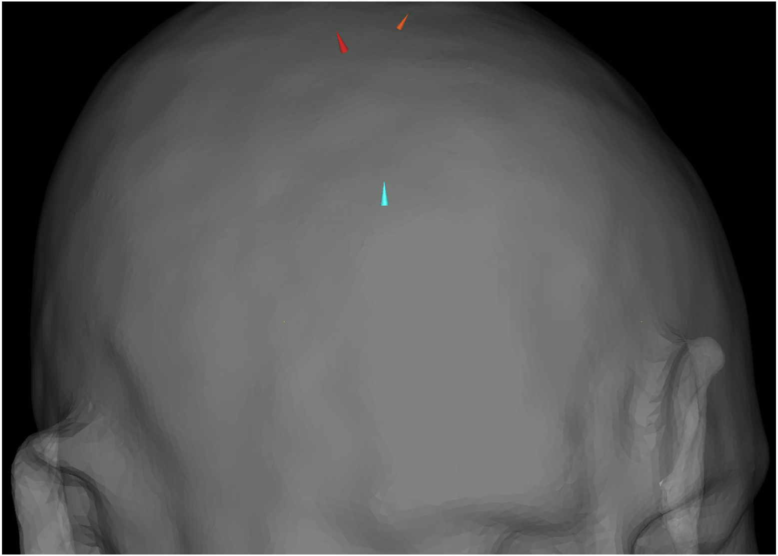
Depth Bias: Full-CM

Computed by blocked Gibbs MCMC sampler.



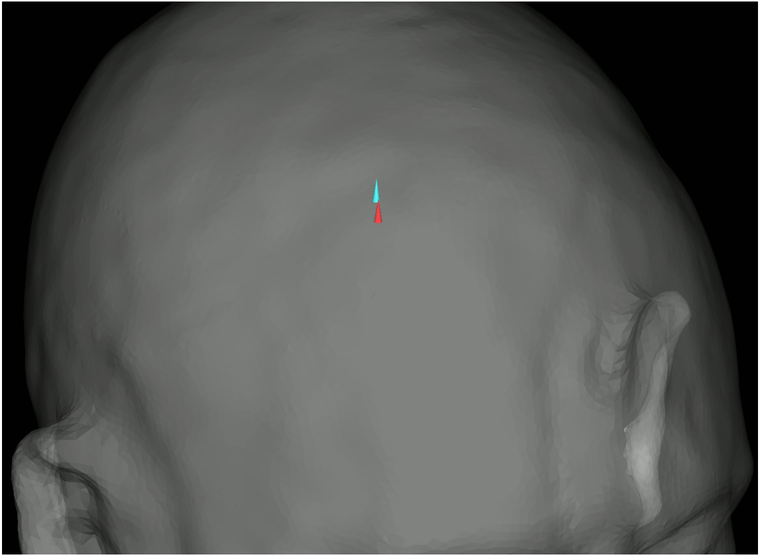
Depth Bias: Full-MAP, Algorithm I

Computed by alternating optimization, uniform initialization.



Depth Bias: Full-MAP, Algorithm II

Computed by alternating optimization initialized at the CM estimate.

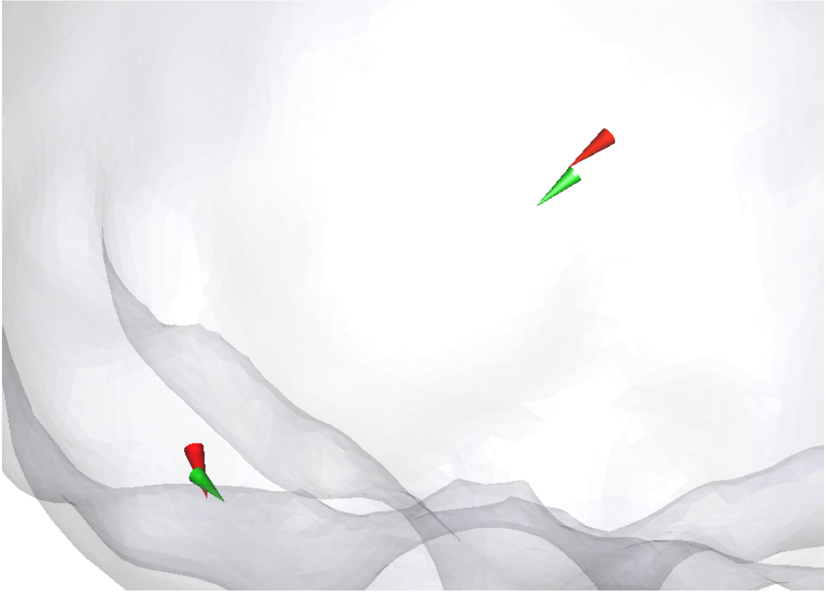


Masking: Result Full-CM

Computed by blocked Gibbs MCMC 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 by utilizing MCMC-based sampling.
- Systematic examination of performance concerning depth-bias and masking in extensive simulation studies.

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.

wissen.leben WWU Münster

A Complex Hierarchical Bayesian Model

HBM as a systematic approach to deal with

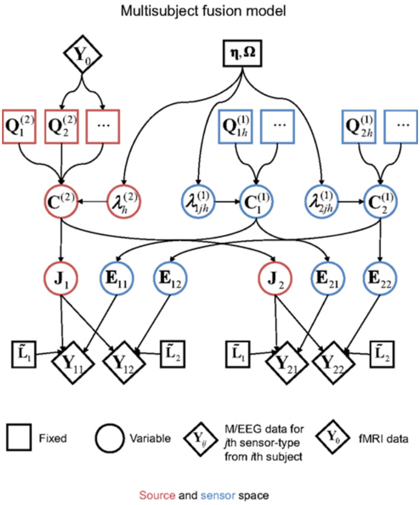
- Plenty of variables
- Various uncertainties
- Different a-priori information

<u>Example:</u>

HBM for

- Multisubject
- Multimodal (EEG/MEG/fMRI)

source reconstruction.



See: Henson RN, Wakeman DG, Litvak V and Friston KJ (2011). A parametric empirical Bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. in *Frontiers in Human Neuroscience*, 5:76.

Outline

Motivation: Depth Bias in EEG/MEG Source Reconstruction.

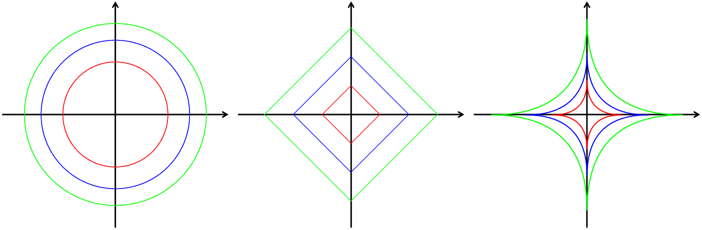
A Sparsity-Promoting Hierarchical Bayesian Model for EEG/MEG

Two Roads to Sparsity: ℓ_p vs. Hierarchical Bayesian Modeling

Take Home Messages & Conclusions

wissen.leben WWU Münster

The "ℓ_p Road to Sparsity"



(a) ℓ_2 (b) ℓ_1 (c) $\ell_{(1/2)}$

$$\min_u \sum |u_i|^p, \qquad s.t. \qquad Ku = f$$

- Start at $p = 2$ ("natural") and move to $p \to 0$.
- Stop at $p = 1$ for convenience.
- Non-differentiability leads to desired binary definition of sparsity:

$$|u|_0 := \#\{i : u_i \neq 0\}$$

- Positive homogeneity $\Longrightarrow$ uniform deformation of isocontours.

Sparsity-Promoting HBM: Implicit Prior

Implicit prior on u by integrating out hyperparameters:

$$
\begin{aligned}
p_{prior}(u) &= \int p_{prior}(u|\gamma)\, p_{hypr}(\gamma) d\gamma \\
&= \int \mathcal{N}(u; 0, \Sigma_u(\gamma))\, p_{hypr}(\gamma) d\gamma \qquad \rightsquigarrow \text{“Gaussian scale mixture”} \\
&\propto \prod_{i=1}^{k} \left(1 + \frac{(u_i^{\text{amp}})^2}{2\beta}\right)^{-(\alpha+3/2)} = \prod_{i=1}^{k} \left(1 + \frac{t_i^2}{\nu}\right)^{-\frac{1}{2}(\nu+1)}
\end{aligned}
$$

with $t_i = u_i^{\text{amp}}/\sqrt{\hat{\gamma}}$, $\hat{\gamma} = \beta/(\alpha+1)$, $\nu = 2(\alpha+1)$.

A Student's t-distribution on the (scaled) source amplitudes.

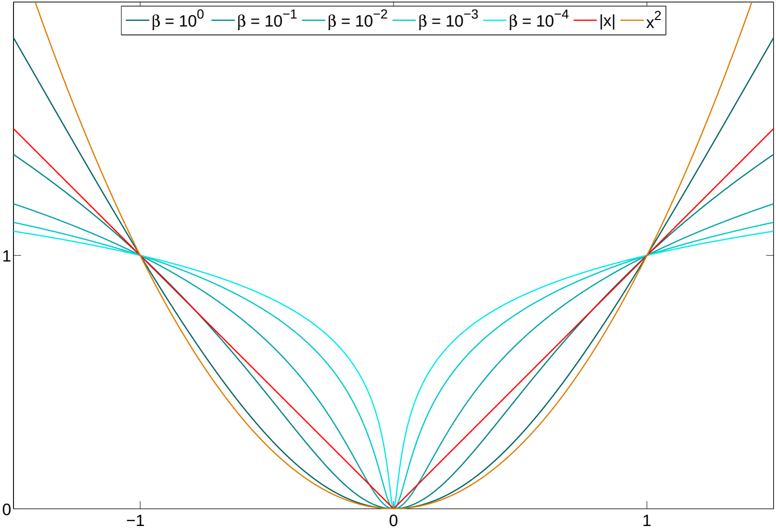
Caution: Only provides intuition, we always deal with the full prior!

wissen.leben
WWU Münster

Sparsity-Promoting HBM: Implicit Functional

Corresponds to the regularization functional:

$$\mathcal{J}(u) = (2\alpha + 3) \sum_{i=1}^{k} \log \left(1 + \frac{u_i^2}{2\beta} \right)$$



Sparsity-Promoting HBM: Implicit Functional

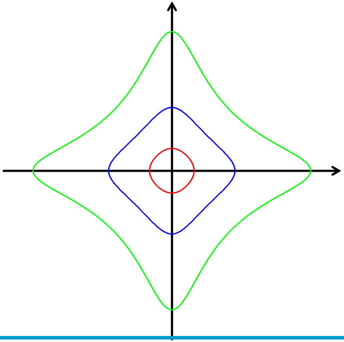
$$\mathcal{J}(u) = (2\alpha + 3) \sum_{i=1}^{k} \log\left(1 + \frac{u_i^2}{2\beta}\right)$$

- Convex for $|u_i| < \sqrt{2\beta}$.
- Concave for $|u_i| > \sqrt{2\beta}$.
- β defines "critical" scale.
- Always differentiable. $\Longrightarrow$ No zeros.
  In practice: thresholding.
  (*Automatic relevance determination*)
- Not homogeneous: Non uniform deformation of isocontours.

wissen.leben
WWU Münster

Sparsity-Promoting HBM: Implicit Functional

$$\mathcal{J}(u) = (2\alpha + 3) \sum_{i=1}^{k} \log\left(1 + \frac{u_i^2}{2\beta}\right)$$



- Convex for $|u_i| < \sqrt{2\beta}$.
- Concave for $|u_i| > \sqrt{2\beta}$.
- β defines "critical" scale.
- Always differentiable. $\Longrightarrow$ No zeros. In practice: thresholding. (*Automatic relevance determination*)
- Not homogeneous: Non uniform deformation of isocontours.

wissen.leben
WWU Münster

The "HBM Road to Sparsity"

$$\mathcal{J}(u) = (2\alpha + 3) \sum_{i=1}^{k} \log\left(1 + \frac{u_i^2}{2\beta}\right)$$

- Set $\alpha := \beta$
- Start at $\beta \to \infty$ and use $\log(1+y) \sim y$ for $y \ll 1$:

$$\mathcal{J}(u) = (2\beta + 3) \sum_{i=1}^{k} \log\left(1 + \frac{u_i^2}{2\beta}\right) \xrightarrow{\beta \to \infty} 2\beta \sum_{i=1}^{k} \frac{u_i^2}{2\beta} = \sum_{i=1}^{k} u_i^2$$

- Move to $\beta \to 0$:

$$\mathcal{J}(u) = (2\beta + 3) \sum_{i=1}^{k} \log\left(1 + \frac{u_i^2}{2\beta}\right) \xrightarrow{\beta \to 0} 3 \sum_{i=1}^{k} \log\left(\frac{u_i^2}{2\beta}\right) \propto \sum_{i=1}^{k} \log\left(|u_i|\right)$$

- Logarithm leads to a scale-based definition of sparsity:
  $\sum_{i=1}^{k} \log\left(|u_i|\right)$ measures the scale differences of the components of u.

wissen.leben
WWU Münster

Outline

wissen.leben WWU Münster

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.
- Interesting features for EEG/MEG source reconstruction:
  - No depth-bias, contrary to ℓ_p-based approaches.
  - Good source separation
  - Capture the various variables and their dependencies in EEG/MEG in a systematic way.
  - Promising for multimodal integration.

wissen.leben
WWU Münster

Hierarchical Bayesian Modeling Compared to ℓ_p-based Approaches

- Always non-convex, but differentiable energy (multimodal posterior).
- Stochastic framework and MCMC help to cope with that.
- Leads to scale-based interpretation of sparsity.

feature	$\ell_\mathbf{p}$	**HBM**
$\mathcal{J}(u)$	$\sum \lvert u_i \rvert^p$	$(2\alpha + 3) \sum \log\left(1 + \frac{u_i^2}{2\beta}\right)$
quadratic limit	$p = 2$	$\beta = \alpha \to \infty$
sparse limit	$p \to 0$	$\beta = \alpha \to 0$
limit functional	$\lvert u \rvert_0$	$\sum \log(\lvert u_i \rvert)$
interpretation	binary	scale based
differentiable	$p > 1$	always
convex	everywhere for $p \geqslant 1$	$\lvert u_i \rvert < \sqrt{2\beta}$
concave	everywhere for $p < 1$	$\lvert u_i \rvert > \sqrt{2\beta}$
homogeneous	yes	no

wissen.leben
WWU Münster

Thank you for your attention!

feature	$\ell_\mathbf{p}$	**HBM**
$\mathcal{J}(u)$	$\sum \lvert u_i \rvert^p$	$(2\alpha + 3) \sum \log \left(1 + \frac{u_i^2}{2\beta} \right)$
quadratic limit	$p = 2$	$\beta = \alpha \to \infty$
sparse limit	$p \to 0$	$\beta = \alpha \to 0$
limit functional	$\lvert u \rvert_0$	$\sum \log (\lvert u_i \rvert)$
interpretation	binary	scale based
differentiable	$p > 1$	always
convex	everywhere for $p \geqslant 1$	$\lvert u_i \rvert < \sqrt{2\beta}$
concave	everywhere for $p < 1$	$\lvert u_i \rvert > \sqrt{2\beta}$
homogeneous	yes	no

F. L., S. Pursiainen, M. Burger and C.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.