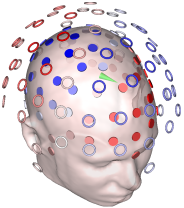
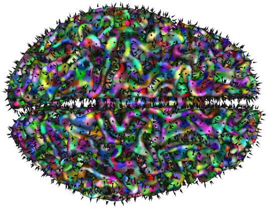
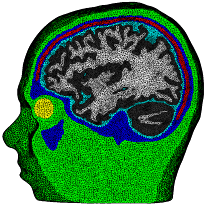
Hierarchical Bayesian Inference for Combined EEG/MEG Source Analysis







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

joint with:
Ümit Aydin, Johannes Vorwerk,
Martin Burger, Carsten H. Wolters.

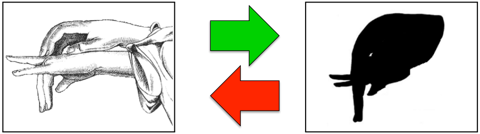
Centre for Medical Image Computing

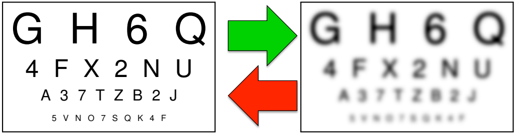
BaCI, Utrecht, September 2, 2015.

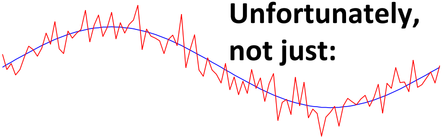
The Inverse Problem of EEG/MEG Source Reconstruction

$$f = Ls + \varepsilon$$

(current density reconstruction)







Unfortunately, not just:

- Under-determined:
  $\#$ sensors $\ll$ $\#$ sources
- Severely ill-conditioned, special spatial characteristics.
- Signal is contaminated by a complex spatio-temporal mixture of external and internal noise and nuisance sources.

Measurements alone are insufficient/unsuitable to determine solution!

Inverse modeling: Use a-priori information to solve the inverse problem.

Problems:

- No consensus, not even for "simple" brain activations.
- Very little research on reliable, physiological a-priori knowledge.
- Underestimation of the impact of prior information.

Consequences:

- Confusing zoo of inverse methods.
- A lot of folklore and funny explanations around.

However:

- Source reconstruction might (always) be a toolbox, but we can find the best tool for a given task / source scenario in a rigorous, objective way.

Networks of Focal Sources

Specific source scenario:

- Unknown number of focal sources.
- No a-priori information about location.
- May involve deep sources.

Challenges:

- Volume-based discretization of gray matter necessary.
- Deep sources are easily masked by superficial ones.

Examples:

- Presurgical epilepsy diagnosis.
- Functional mapping of the eloquent cortex.
- Early components of evoked potentials

More practical aspects in the upcoming talks!

Our Tool: Hierarchical Bayesian Modeling

Relies on Bayesian inversion:

- A priori information is encoded by probability distributions.
- Extend Gaussian prior by flexible, individual source variances γ_i.
- Let the data determine γ_i (hyperparameters).
- Incorporate focality constraints on hyperparameters.

$$p_{prior}(s|\gamma) \propto \prod \exp\left(-\frac{(s_{amp})_i^2}{\gamma_i}\right), \qquad p_{hyper}(\gamma_i) \propto \gamma_i^{-(\alpha+1)} \exp\left(-\frac{\beta}{\gamma_i}\right)$$

Our starting points:

Calvetti, Hakula, Pursiainen, Somersalo, 2009. *Conditionally Gaussian hypermodels for cerebral source localization.* SIAM J. Imaging Sci.

Wipf, Nagarajan, 2009. *A unified Bayesian framework for MEG/EEG source imaging.* Neuroimage

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

Non Convex Intuition Behind HBM

We use fully-Bayesian inference for the posterior:

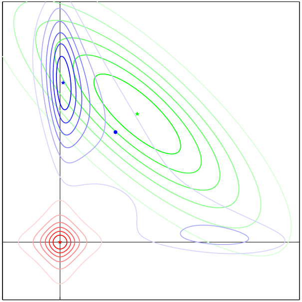
$$p_{post}(s, \gamma | f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|_2^2 - \sum_i^n \left(\frac{(s_{\text{amp}})_i^2 + 2\beta}{2\gamma_i} + (\alpha + 1/2)\log(\gamma_i) \right) \right)$$

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

$$p_{prior}(s) \propto \prod_i \left(1 + \frac{(s_{\text{amp}})_i^2}{\nu\theta} \right)^{-\frac{\nu-1}{2}}$$

$$-\log p_{post}(s|f) \propto$$

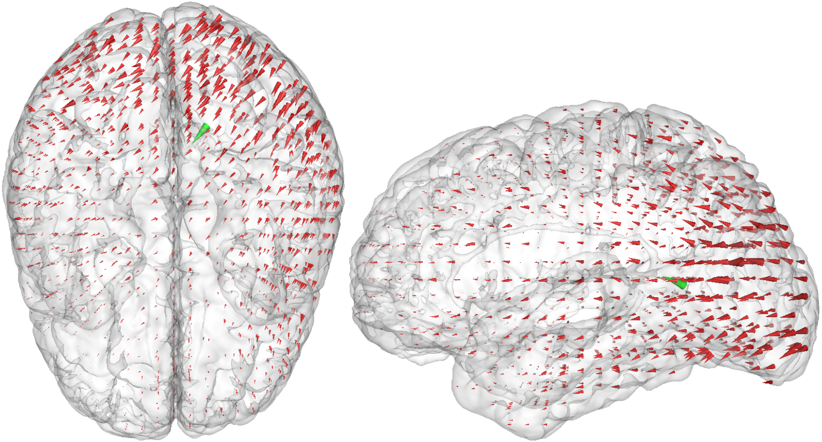
$$\frac{1}{2}\|f - A\,s\|_2^2 + \frac{\nu - 1}{2} \sum_i \log\left(1 + \frac{(s_{\text{amp}})_i^2}{\nu\theta} \right)$$



Non-convex regularization?! Why would anyone want to do that?

Reference (green cone) and minimum norm estimate (red cones):

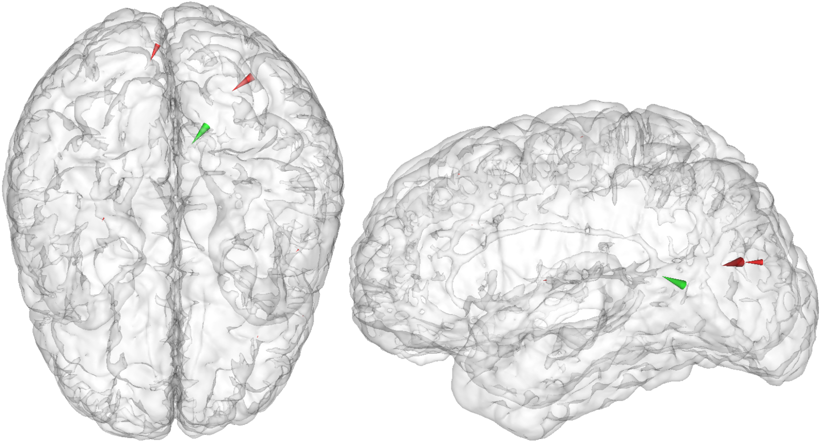
$$s_{\text{MNE}} = \underset{s}{\operatorname{argmin}} \left\{ \|f - L\,s\|_2^2 + \lambda \|s_{\text{amp}}\|_2^2 \right\}$$



Why Non-Convex Functionals?! An Illustration

Reference (green cone) and minimum current estimate (red cones):

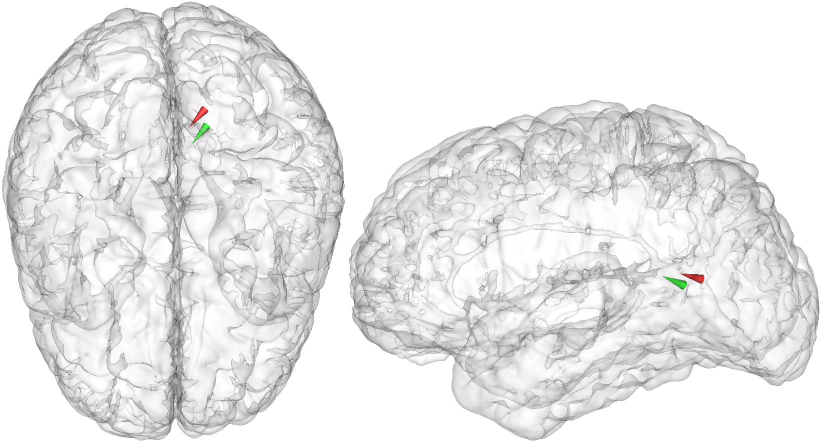
$$s_{\text{MCE}} = \underset{s}{\operatorname{argmin}} \left\{ \|f - L\,s\|_2^2 + \lambda \|s_{\text{amp}}\|_1 \right\}$$



Why Non-Convex Functionals?! An Illustration

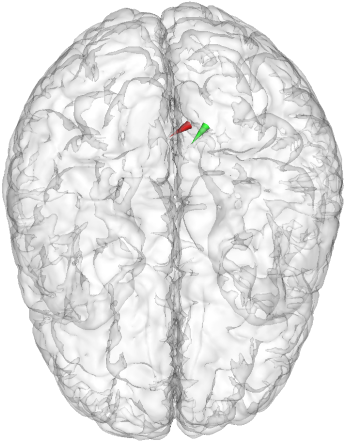
Reference (green cone) and single dipole scan (red cone):

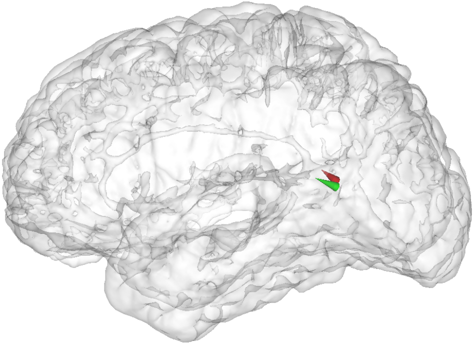
$$s_{\text{SDS}} = \underset{s}{\operatorname{argmin}} \left\{ \|f - L\,s\|_2^2 + N_1(s) \right\}, \qquad N_1(s) = \begin{cases} 0 & \text{if } \; |s_{\text{amp}}|_0 = 1 \\ \infty & \text{else} \end{cases}$$



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

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





The Curse of Convexity: Depth Bias

"Theorem": All variational regularization approaches

$$\hat{s} = \underset{s}{\operatorname{argmin}} \left\{ \|f - L\, s\|_2^2 + \sum_i g(|s_i|) \right\}$$

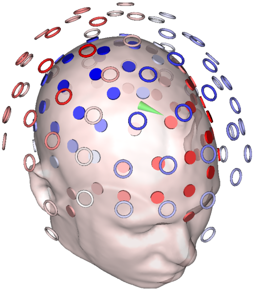
that are uniform in i (no weighting) with convex g have depth bias:

- $|\hat{s}_i|$ 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:

F.L., S. Pursiainen, M. Burger, 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*. *NeuroImage*, 61(4):1364–1382.

EEG vs. MEG and EMEG



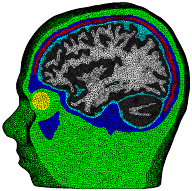
- Which modality is "better"?
- Does EMEG combine the deficits or strengths?

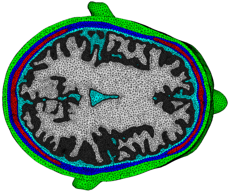
"EEG vs. MEG" has practical and theoretical aspects, don't mix them up!

Dassios, Fokas, 2013. *The definite non-uniqueness results for deterministic EEG and MEG data*, Inverse Problems.

Setting:

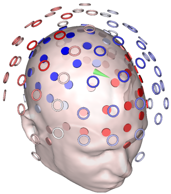
- Realistic head model.
- Equal number of EEG/MEG sensors.
- Sources in gray matter volume.
- One, two or three active sources.
- Evaluation using dipole localization error or earth mover's distance.





Inverse methods:

- Hierachical Bayesian Modeling (HBM)
- Minimum norm estimation (MNE)
- Different weighted MNE (WMNE) variants
- sLORETA



Results:

- Localization performance of HBM is equal for EEG and MEG.
- For WMNE variants and sLORETA, it is better for MEG.
- EMD (localization + extend) is better for EEG than MEG (all methods).
- HBM and sLORETA do not show any depth bias.
- Optimizing a-priori weights for WMNE is difficult: Most weights try to optimize single dipole recovery for one modality at the expense of source separation.

Conclusions:

- "Performance" of single modalities cannot be assessed independent of an inverse method used! This is a feature of the ill-posedness.
- MNE variants and sLORETA: Better localization of MEG comes at the costs of larger blurring.

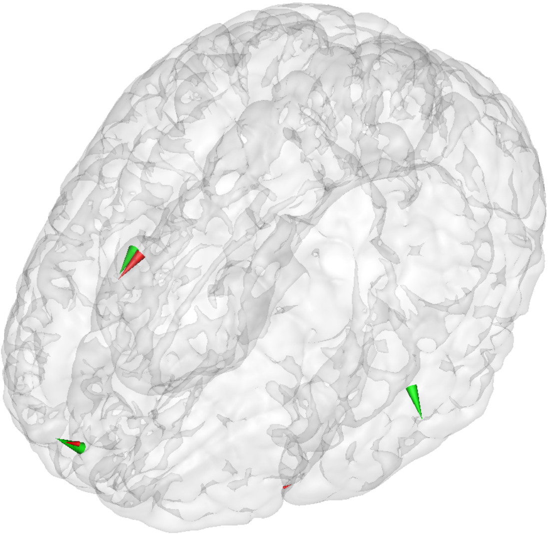
Results:

- EEG/MEG combination improves performance of all methods.
- Combination reduces variance and outliers in the error statistics.
- HBM source separation especially profits from combination.
- Depth localization does not always profit, especially if a single modality is very weak in that aspect.

Conclusions:

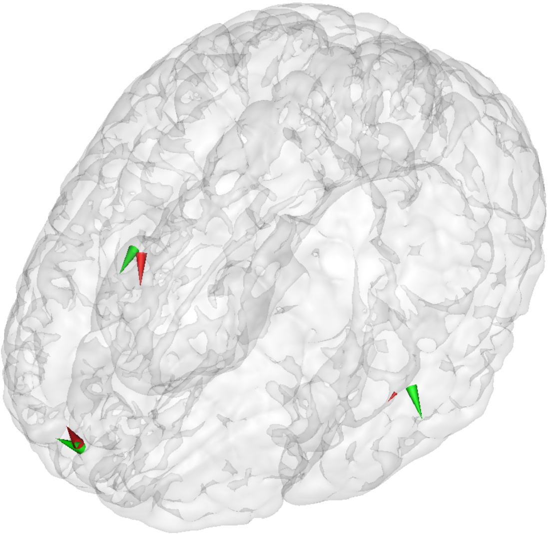
- EEG/MEG combination stabilizes and improves source reconstruction.
- No "combination of weaknesses"

HBM Simulation Studies, Specific Results I: EEG



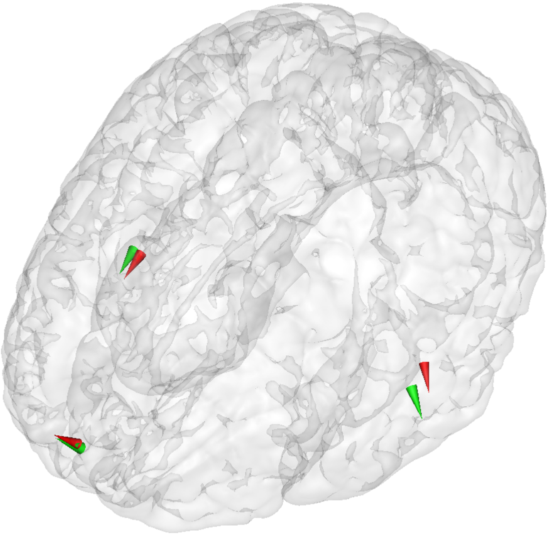
green cones: reference source red cones: HBM solution

HBM Simulation Studies, Specific Results I: MEG



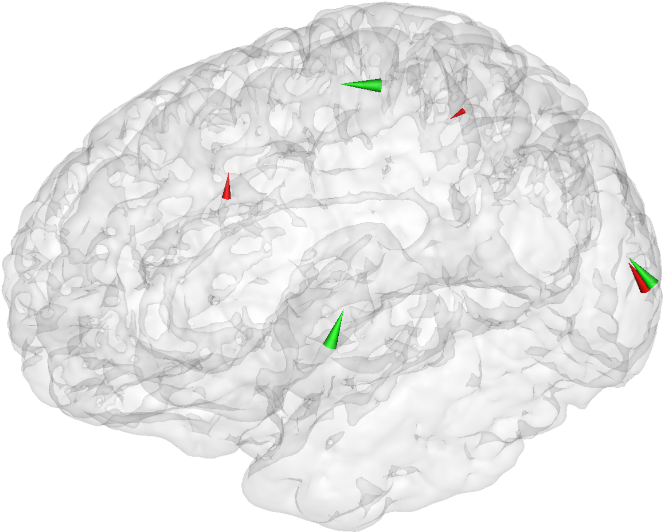
green cones: reference source   red cones: HBM solution

HBM Simulation Studies, Specific Results I: EMEG



green cones: reference source   red cones: HBM solution

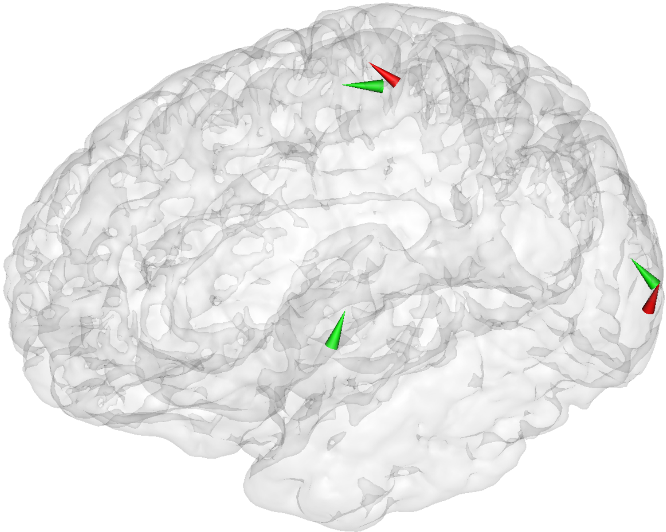
HBM Simulation Studies, Specific Results III: EEG



green cones: reference source

red cones: HBM solution

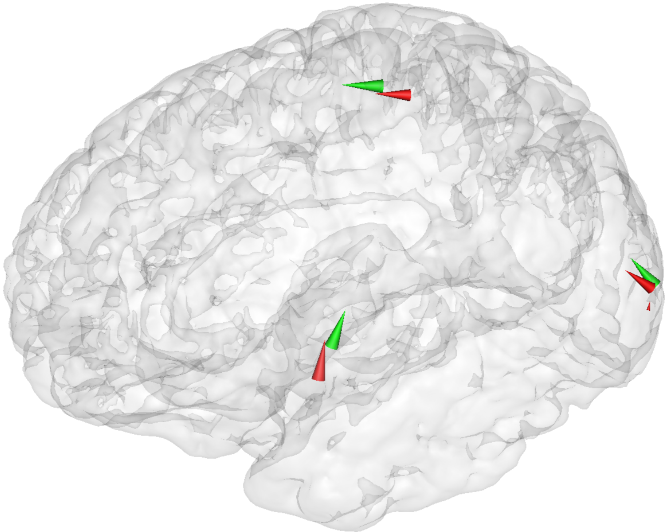
HBM Simulation Studies, Specific Results III: MEG



green cones: reference source

red cones: HBM solution

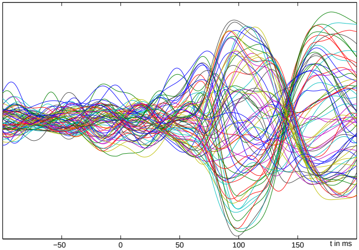
HBM Simulation Studies, Specific Results III: EMEG



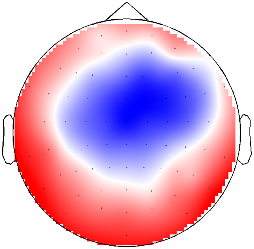
green cones: reference source

red cones: HBM solution

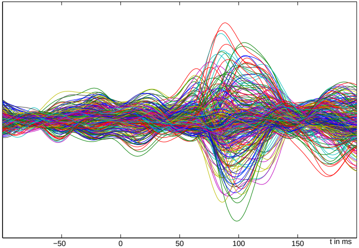
Validation with Auditory Evoked N100(m)



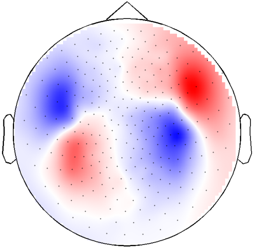
(a) AEP butterfly



(b) AEP topography

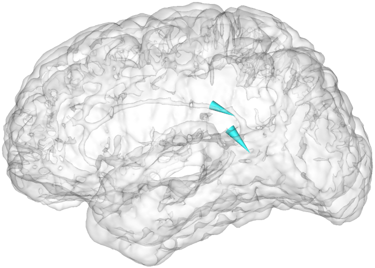


(c) AEF butterfly

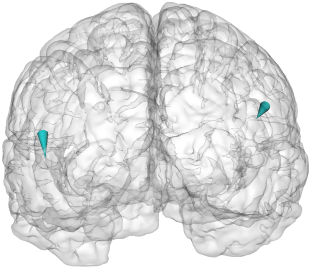


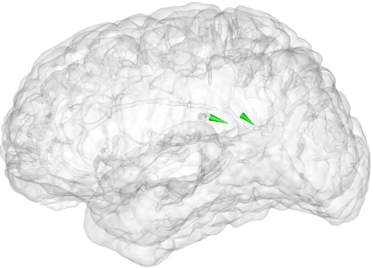
(d) AEF topography

Validation of HBM for AEP/AEF: EEG alone

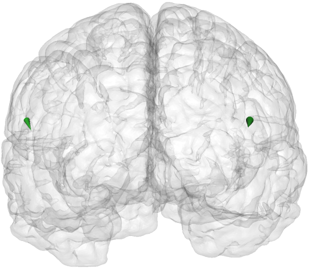


TDS

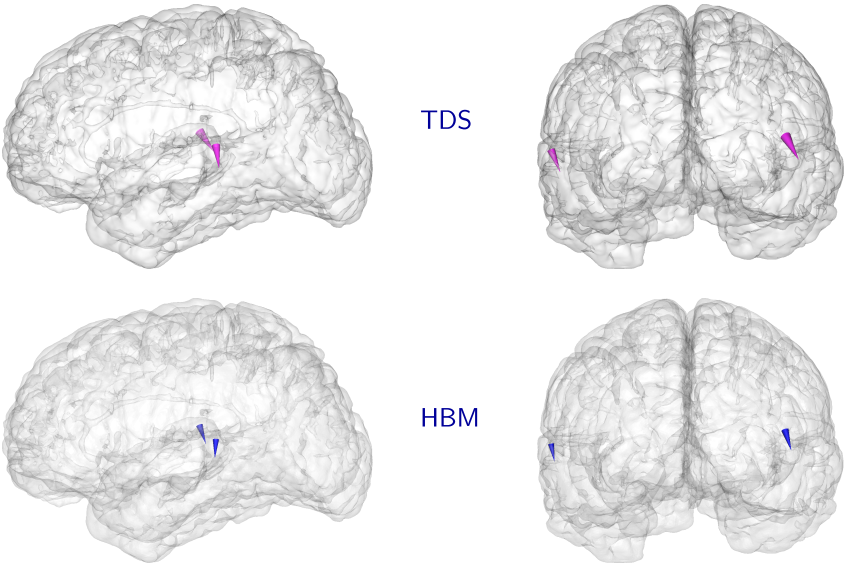




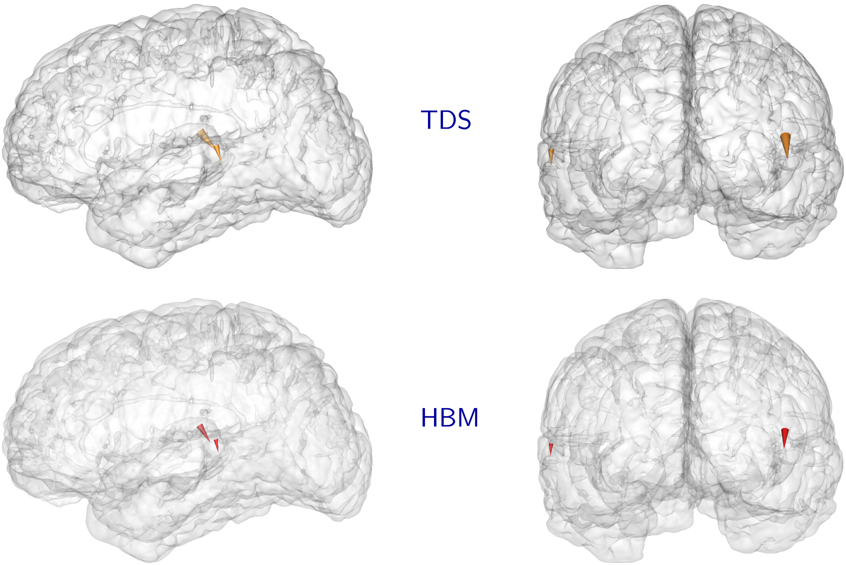
HBM



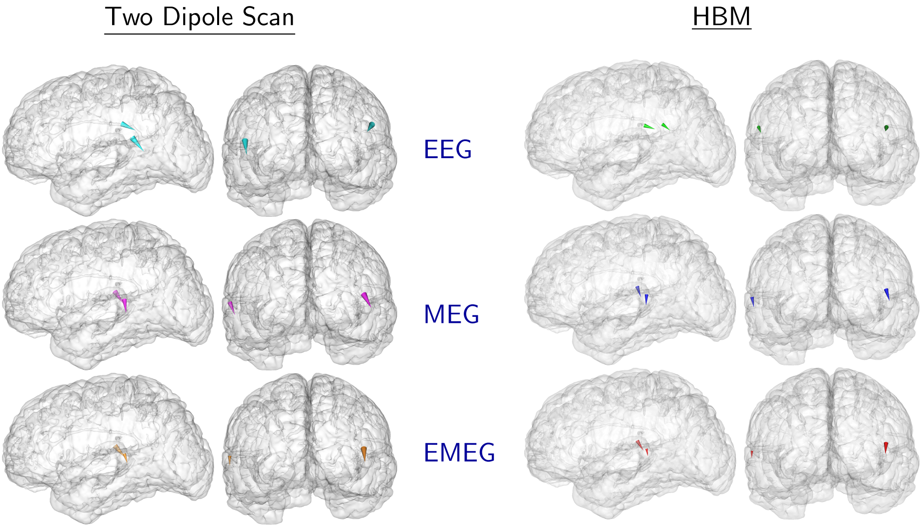
Validation of HBM for AEP/AEF: MEG alone



Validation of HBM for AEP/AEF: EMEG



Validation of HBM for AEP/AEF: All Results



HBM and Experimental Data: Sensitivity Studies

Disappointing first results (not shown here), also reported by others.

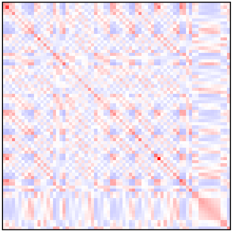
Non-linear, non-convex methods too sensitive to

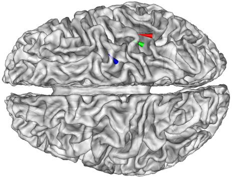
- Noise modeling errors?
- Source modeling errors / background activity?
- Forward modeling errors?

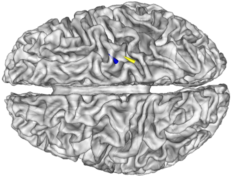
⇝ Examination through sensitivity studies.

Results:

- HBM estimates are surprisingly robust.





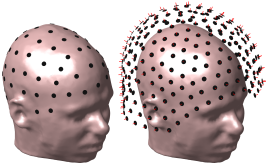


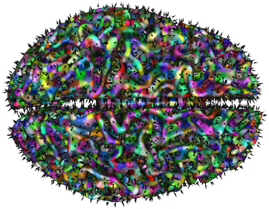
Another Perspective on Sparse EEG/MEG Combination

Aim: Interplay of realistic forward and ℓ_1-norm inverse modeling.

Methods:

- Compare exact recovery conditions developed in compressed sensing.
- Head model cascade, surface source spaces.

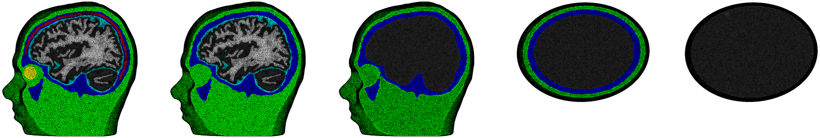




Results concerning EEG/MEG, EMEG:

- Combination boosts reconstruction performance.
- Strong conditions like coherence or RIP mislead.

L., Tellen, Wolters, Burger, 2013. *Sparse Recovery Conditions and Realistic Forward Modeling in EEG/MEG Source Reconstruction.* Compressed Sensing and its Applications, Berlin.



Source reconstruction:

- We need to accept the difficulty of source reconstruction.
- Toolbox of different, prior-dominated inverse methods.
- We need a rigorous, objective assessment of their pro's, con's and limitations for specific source scenarios.
- Example: Depth bias of uniform convex regularization.
- Pseudo-physiological motivations and folklore need to be overcome.
- Hope by multi-modal integration (EMEG, fMRI, NIRS, PET/SPECT,...), anatomical information (ROI, orientation), functional organization (atlas, DW-MRI), coupling to generative models?

Fully Bayesian inference for hierarchical Bayesian modeling:

- Promising results for focal source networks.
- Validated on simulated and experimental data.
- Non-convexity is challenging:
  - Heuristic optimization by multiple, MCMC-informed seeds.
  - Optimization community turn on such problems...

EEG vs. MEG, EMEG combination:

- Don't mix up practical and theoretical arguments.
- Theoretically, they provide complementary information of similar quality.
- "Performance" of single modalities cannot be assessed independent of an inverse method used!
- EMEG combines the strengths, not weaknesses of single modalities and stabilizes and improves source reconstruction.

F.L., Ü. Aydin, J. Vorwerk, M. Burger, C.H. Wolters. *Hierarchical Bayesian Inference for Combined EEG/MEG Source Analysis* (*in preparation*)

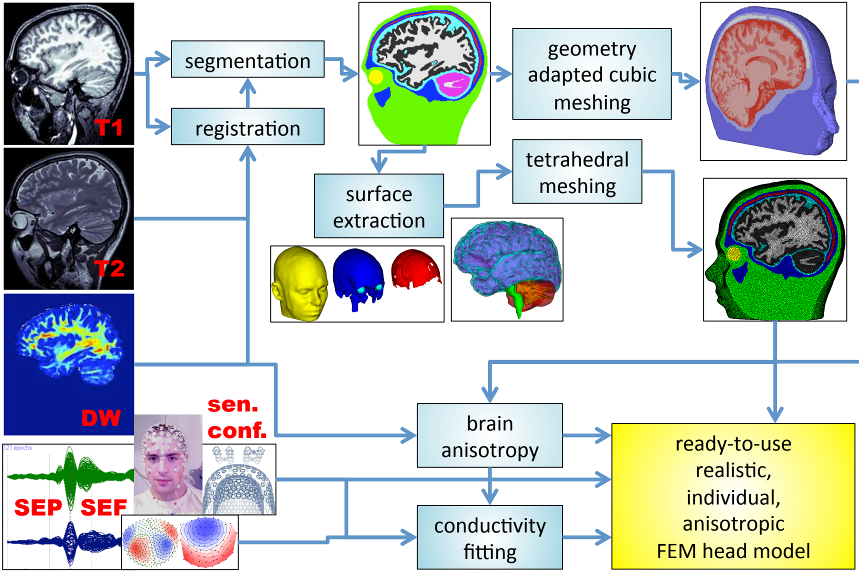
F.L., 2014. *Bayesian Inversion in Biomedical Imaging.* PhD Thesis, University of Münster.

F.L., S. Pursiainen, M. Burger, 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*. *NeuroImage*, 61(4):1364–1382.

Thank you for your attention!

- F.L., Ü. Aydin, J. Vorwerk, M. Burger, C.H. Wolters. *Hierarchical Bayesian Inference for Combined EEG/MEG Source Analysis* (*in preparation*)
- F.L., 2014. *Bayesian Inversion in Biomedical Imaging.* PhD Thesis, University of Münster.
- F.L., S. Pursiainen, M. Burger, 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*. *NeuroImage*, 61(4):1364–1382.

Realistic and individual head models for simulating the forward equations.



Fully-Bayesian Inference for HBM

$$p_{post}(s, \gamma | f) \propto \exp\left(-\frac{1}{2}\|f - A\,u\|_2^2 - \sum_i^n \left(\frac{(s_{\text{amp}})_i^2 + 2\beta}{2\gamma_i} + (\alpha + 1/2)\log(\gamma_i) \right) \right)$$

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

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

Major difficulty: Multimodality of posterior.

Heuristic Full-MAP computation:

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

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

Variational regularization:

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

First order optimality condition:

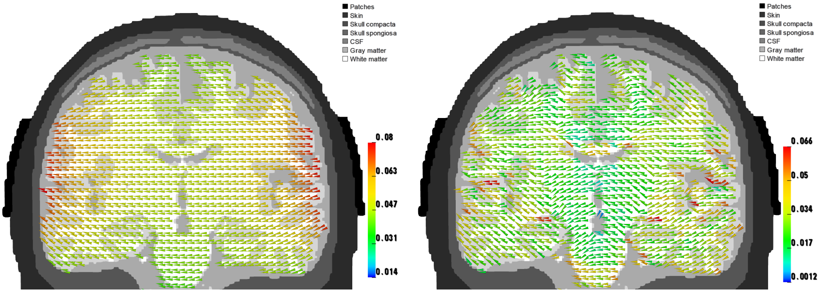
$$-L^T\,(f - L\,\hat{s}) + \mathcal{J}'(\hat{s}) \overset{!}{=} 0 \qquad \Longleftrightarrow \qquad \mathcal{J}'(\hat{s}) = L^T\,(f - L\,\hat{s})$$

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

- L^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.

Solutions to the tCS Problem



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"*.

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

Depth Bias: The Curse of (Uniform) Convexity

Assume

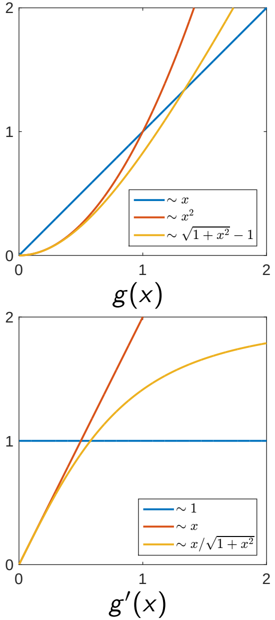
- $\mathcal{J}(s) \propto \sum_i g(|s_i|)$ (uniform in i).
- for simplicity, s is scalar.
- $g(x) : \mathbb{R}^+ \to \mathbb{R}^+$ non-decreasing: $g'(x) \geqslant 0$.

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

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

$\Longrightarrow$ Depth bias!

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

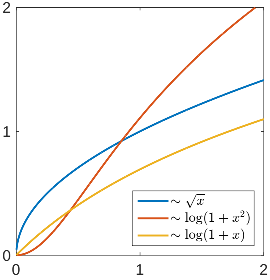


Depth Bias: The Blessings of Non-Convexity

Assume

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

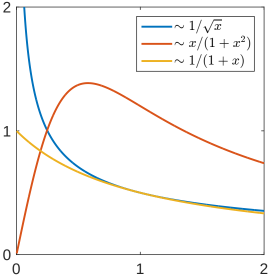
If g is non-convex, $g'(x)$ does not necessarily induce an order and $\hat{s}$ does not need to "inherit" maximum principle!



$g(x)$

But caution:

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



$g'(x)$

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 $\mathcal{J}(s) \propto \sum_i g\left(\frac{|s_i|}{w_i(L_i)}\right)$
such as WMNE, WMCE,...

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

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

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.

Wikipedia on MEG vs EEG: *"The decay of magnetic fields as a function of distance is more pronounced than for electric fields. Therefore, MEG is more sensitive to superficial cortical activity,..."*

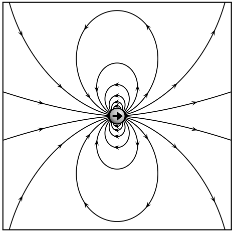
What are EM fields? (to my understanding)

- Charged particles experience an electromagnetic force.
- Everything else is mathematical description
- EM force can be described by EM field.
- Electric and magnetic fields are complementary appearances of the EM field (Maxwell).
- Electric and magnetic potentials often allow simpler description.

Current dipole:

- E and M fields decay like r^3.
- E and M potentials decay like r^2.
- Common to describe electric measurements by potentials and magnetic ones by fields.

But what do you actually measure, and how?



source: Wikimedia Commons

Work by Dassios, Fokas et al.:

- Electric and magnetic measurements carry different information about sources.
- In spherical geometry: Information is completely complementary.
- Even EMEG does not carry enough information for uniqueness...

Dassios, Fokas, 2013.
The definite non-uniqueness results for deterministic EEG and MEG data. Inverse Problems

Dassios, Fokas, Hadjiloizi, 2007.
On the complementarity of electroencephalography and magnetoencephalography. Inverse Problems