A hierarchical Krylov-Bayes iterative inverse solver for MEG with physiological preconditioning

D Calvetti[1] A Pascarella[3] F Pitolli[2] E Somersalo[1] B Vantaggi[2]

[1] Case Western Reserve University
Department of Mathematics, Applied Mathematics and Statistics
[2] University of Rome "La Sapienza"
Department of Basic and Applied Science for Engineering
[3] Istituto per le Applicazioni del Calcolo "Mario Picone"
CNR, Rome, Italy

MEG forward model

Impressed current and volume current:

$$\vec{J}_{\mathrm{tot}} = \vec{J} + \vec{J}_{\Omega} = \vec{J} + \sigma \vec{E} = \vec{J} - \sigma \nabla u,$$

where

$$\nabla \cdot (\sigma \nabla u) = \nabla \cdot \vec{J}, \quad \sigma \frac{\partial u}{\partial n}\bigg|_{\partial \Omega} = 0,$$

Biot-Savart law:

$$\vec{B}(\vec{p}) = \frac{\mu_0}{4\pi} \int_{\Omega} \vec{J}_{\mathrm{tot}}(\vec{r}) \times \frac{\vec{p} - \vec{r}}{|\vec{p} - \vec{r}|^3} d\vec{r}, \quad \vec{p} \in \mathbb{R}^3 \setminus \overline{\Omega},$$

Lead field

Magnetometer data: Position $\vec{p}_k$, orientation $\vec{\nu}_k$,

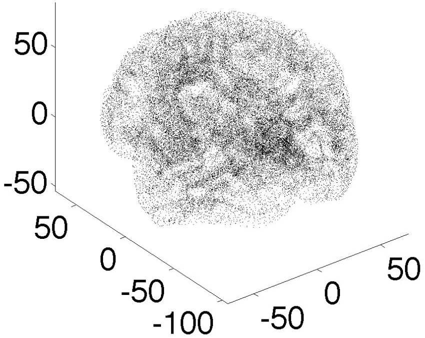
$$\begin{aligned}\beta_k &= \vec{\nu}_k \cdot \vec{B}(\vec{p}_k) = -\frac{\mu_0}{4\pi}\int_\Omega \frac{\vec{\nu}_k \times (\vec{p}_k - \vec{r})}{|\vec{p}_k - \vec{r}|^3} \cdot \vec{J}_{\mathrm{tot}}(\vec{r}) d\vec{r} \\ &= \int_\Omega \vec{\mathscr{M}}_{0,k}(\vec{r}) \cdot \vec{J}_{\mathrm{tot}}(\vec{r}) d\vec{r}.\end{aligned}$$

By linearity of Maxwell's equations, it is possible to write

$$\beta_k = \int_\Omega \vec{\mathscr{M}}_k(\vec{r}) \cdot \vec{J}(\vec{r}) d\vec{r},$$

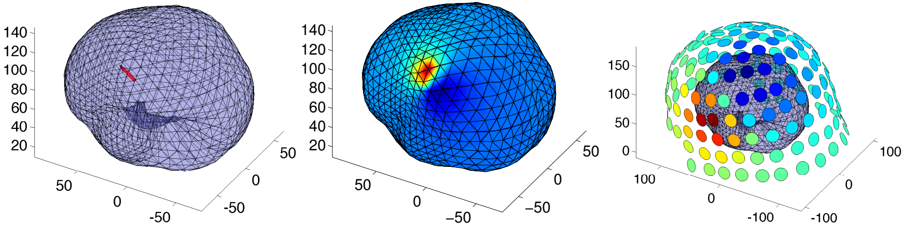
where $\vec{\mathscr{M}}_k$ depends on the geometry and conductivity of the head. assuming constant conductivity, $\vec{\mathscr{M}}_k$ can be approximated numerically by Geselowtz's formula using BEM.

Discretization: Dipole model



Brain model based on segmented MRI image. The grid points v_j represent the gray matter, and are the possible dipole locations.

Geometry



Current dipole (left), the surface electric potential computed by BEM (center), and the magnetic field at the magnetometers (right).

Discretization

$(\vec{r}_j, \vec{q}_j)$, $1 \leq j \leq n$, the positions and dipole moments of the dipoles.

$$\beta_k = \sum_{j=1}^{n} \vec{\mathscr{M}}_k(\vec{r}_j) \cdot \vec{q}_j, \quad 1 \leq k \leq m.$$

Furthermore, define

$$b_k = \beta_k + \varepsilon_k$$

with additive noise ε_k:

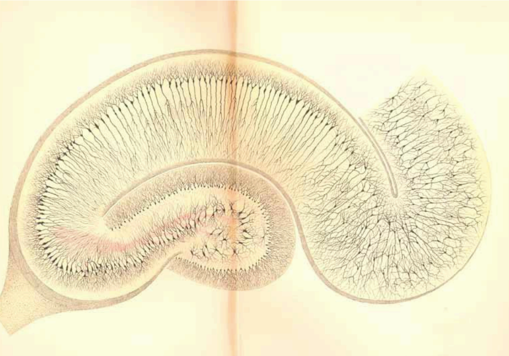
$$b = \sum_{j=1}^{n} \mathsf{M}_j \vec{q}_j + \varepsilon,$$

where $\mathsf{M}_j \in \mathbb{R}^{m \times 3}$ is a matrix with rows equal to $\vec{\mathscr{M}}_k(\vec{r}_j)^{\mathsf{T}}$,
ε = observation noise vector.

MEG inverse problem

- Given measurements of magnetic field at sensors positions, recover the location and strength of the activity in brain.
- Magnetometer data: ≈ 150 measurements
- Brain activity is represented by current dipoles at 25K possible fixed locations: the unknowns are the moments of the dipoles: 75K unknowns
- This is a linear inverse problem with a huge null space: the prior plays a major role.

What do we know about brain anatomy?



Anatomical prior

At each grid point $v_j \in \Omega$, $1 \leq j \leq n$,

- orthonormal triplet $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\} \in \mathbb{R}^3$ is given,
- $\vec{u}_3$ is in the preferential direction of the dipole (parallel to neuronal fascicle).

Define a SPD matrix

$$\mathsf{C}_j = \delta(\vec{u}_1 \vec{u}_1^{\mathsf{T}} + \vec{u}_2 \vec{u}_2^{\mathsf{T}}) + \vec{u}_3 \vec{u}_3^{\mathsf{T}} \in \mathbb{R}^{3\times 3},$$

where $0 < \delta < 1$ is a small parameter.

Anatomical prior

Define a conditionally Gaussian prior,

$$\pi^{j}_{\mathrm{prior}}(\vec{q}_j \mid \theta_j) \sim \mathcal{N}(0, \theta_j \mathsf{C}_j),$$

Explicitly:

$$\begin{aligned}\pi^{j}_{\mathrm{prior}}(\vec{q}_j \mid \theta_j) &= \frac{1}{\pi^{3/2}\sqrt{|\theta_j \mathsf{C}_j|}} \exp\left(-\frac{1}{2\theta_j} \vec{q}_j^{\mathsf{T}} \mathsf{C}_j^{-1} \vec{q}_j\right) \\ &\propto \exp\left(-\frac{1}{2}\frac{\|\vec{q}_j\|^2_{\mathsf{C}_j}}{\theta_j} - \frac{3}{2}\log\theta_j\right).\end{aligned}$$

where

$$\|\vec{q}_j\|^2_{\mathsf{C}_j} = \vec{q}_j^{\ \mathsf{T}} \mathsf{C}_j^{-1} \vec{q}_j.$$

Independency:

$$\pi_{\mathrm{prior}}(\vec{q}_1, \ldots, \vec{q}_n \mid \theta_1, \ldots, \theta_n) = \prod_{j=1}^{n} \pi^{j}_{\mathrm{prior}}(\vec{q}_j \mid \theta_j).$$

Hierarchical model

Hypermodel using gamma distribution,

$$\theta_j \sim \pi^j_{\mathrm{hyper}}(\theta_j \mid \theta^*_j, \beta_j) \propto \theta_j^{\beta_j - 1} \exp\left(-\frac{\theta_j}{\theta^*_j}\right),$$

where (β_j, θ^*_j) are hyperparameters.
Independency:

$$\pi_{\mathrm{hyper}}(\theta \mid \theta^*, \beta) = \prod_{j=1}^{n} \pi^j_{\mathrm{hyper}}(\theta_j \mid \theta^*_j, \beta_j).$$

Prior density

Notation:

$$Q = \begin{bmatrix} \vec{q}_1 \\ \vdots \\ \vec{q}_n \end{bmatrix} \in \mathbb{R}^{3n}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \in \mathbb{R}^{n}.$$

The joint prior model for (Q, θ) is written as

$$\pi_{\text{prior}}(Q, \theta \mid \theta^*, \beta) = \pi_{\text{prior}}(Q \mid \theta)\pi_{\text{hyper}}(\theta \mid \theta^*, \beta)$$

$$\propto \exp\left(-\frac{1}{2}\sum_{j=1}^{n} \frac{\|\vec{q}\|^2_{C_j}}{\theta_j} + \sum_{j=1}^{n}\left(\beta_j - \frac{5}{2}\right)\log\theta_j - \sum_{j=1}^{n}\frac{\theta_j}{\theta_j^*}\right).$$

Posterior density

Likelihood model:

$$b = \sum_{j=1}^{n} \mathsf{M}_j \vec{q}_j + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \Sigma),$$

where $\Sigma \in \mathbb{R}^{m \times m}$ is a SPD noise covariance matrix.
By Bayes' theorem,

$$\pi(Q, \theta \mid b, \theta^*, \beta) \propto \pi(b \mid Q)\pi(Q, \theta \mid \theta^*, \beta) \propto$$

$$\exp\left(-\frac{1}{2}\|b - \sum_{j=1}^{n} \mathsf{M}_j \vec{q}_j\|_{\Sigma}^2 - \frac{1}{2}\sum_{j=1}^{n} \frac{\|\vec{q}\|_{\mathsf{C}_j}^2}{\theta_j} + \sum_{j=1}^{n}\left(\beta_j - \frac{5}{2}\right)\log\theta_j - \sum_{j=1}^{n}\frac{\theta_j}{\theta_j^*}\right).$$

Gibbs energy

The negative of the log-posterior is

$$\begin{aligned}\mathscr{E}(Q,\theta) &= \frac{1}{2}\|b-\sum_{j=1}^{n}\mathsf{M}_j\vec{q}_j\|_{\Sigma}^2+\frac{1}{2}\sum_{j=1}^{n}\frac{\|\vec{q}_j\|_{\mathsf{C}_j}^2}{\theta_j}-\sum_{j=1}^{n}\eta_j\log\theta_j+\sum_{j=1}^{n}\frac{\theta_j}{\theta_j^*}\\ &= \frac{1}{2}\|b-\mathsf{M}Q\|_{\Sigma}^2+\frac{1}{2}\|Q\|_{\mathsf{D}_\theta}^2-\sum_{j=1}^{n}\eta_j\log\theta_j+\sum_{j=1}^{n}\frac{\theta_j}{\theta_j^*},\end{aligned}$$

where

$$\eta_j=\beta_j-\frac{5}{2}.$$

Maximum A Posteriori (MAP) estimate is the minimizer of $\mathscr{E}(Q,\theta)$.

MAP estimate and IAS algorithm

Iterative Alternating Scheme (IAS):

1. Initialize $\theta = \theta^0$, and set $k = 0$.
2. Update Q by defining

$$Q^{k+1} = \mathrm{argmin}\{\mathscr{E}(Q, \theta^k)\}; \tag{1}$$

3. Update θ by defininig

$$\theta^{k+1} = \mathrm{argmin}\{\mathscr{E}(Q^{k+1}, \theta)\}; \tag{2}$$

4. If a convergence criterion is met, stop, else increase k by one and continue from Step 2.

MAP estimate and IAS algorithm

Minimization by iterating the two steps:

1. Update Q:

$$Q^{k+1} = \operatorname{argmin}\{\mathscr{E}(Q \mid \theta^k)\} = \operatorname{argmin}\{\frac{1}{2}\|b - \mathsf{M}Q\|_{\Sigma}^2 + \frac{1}{2}\|Q\|_{\Gamma_{\theta^k}}^2\}.$$

   Approximate the solution using Krylov iterative method on

$$\mathsf{S}M\mathsf{C}_{\theta^k}^{-1}W = \mathsf{C}b, \quad \Sigma^{-1} = \mathsf{C}^{\mathsf{T}}\mathsf{C},\ \Gamma_{\theta^k}^{-1} = \mathsf{C}_{\theta^k}^{\mathsf{T}}\mathsf{C}_{\theta^k}$$

   stopping when the discrepancy falls below $\sqrt{m}$.

2. Update θ: Minimize component-wise

$$\mathscr{E}(\theta_j, \vec{q}_j^{\,k+1}) = \frac{1}{2}\frac{\|\vec{q}_j^{\,k+1}\|_{\mathsf{C}_j}^2}{\theta_j} - \eta_j \log\theta_j + \frac{\theta_j}{\theta_j^*}.$$

   by finding the unique positive critical point,

$$\theta_j^{k+1} = F_j(\vec{q}_j^{\,k+1}) = \theta_j^*\left(\frac{\eta_j}{2} + \sqrt{\frac{\eta_j^2}{4} + \frac{\|\vec{q}_j^{\,k+1}\|_{\mathsf{C}_j}^2}{2\theta_j^*}}\right), \quad \eta_j = \beta_j - 5/2.$$

Repeat the two steps when the stopping criterion is met.

MAP estimate and IAS algorithm

Theorem
The IAS algorithm converges to the unique minimizer $(\widehat{Q},\widehat{\theta})$ of the Gibbs energy $\mathscr{E}(Q,\theta)$. Moreover, the minimizer $(\widehat{Q},\widehat{\theta})$ satisfies the fixed point conditions,

$$\widehat{Q} = \mathrm{argmin}\{\mathscr{E}(Q, F(Q))\}, \quad \widehat{\theta} = F(\widehat{Q}),$$

where $F : \mathbb{R}^{3n} \to \mathbb{R}^n$ is the mapping with components $F_j : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$F_j(\vec{q}_j) = \theta_j^* \left(\frac{\eta_j}{2} + \sqrt{\frac{\eta_j^2}{4} + \frac{\|\vec{q}_j\|_{C_j}^2}{2\theta_j^*}} \right).$$

Hyperparameters and depth weighting

Convergence result implies that the global minimizer satisfies

$$\widehat{\theta}_j = \theta_j^* \left(\frac{\eta_j}{2} + \sqrt{\frac{\eta_j^2}{4} + \frac{\|\widehat{\vec{q}}_j\|_{\mathrm{C}_j}^2}{2\theta_j^*}} \right), \quad \eta_j = \beta_j - 5/2.$$

where $\widehat{Q}$ is the minimizer of the expression

$$\begin{aligned} \mathscr{E}(Q, F(Q)) \quad = \quad & \frac{1}{2}\|b - \sum_{j=1}^{n} \mathrm{M}_j \vec{q}_j\|_{\Sigma}^2 + \sum_{j=1}^{n} \left\{ \frac{1}{2} \frac{\|\vec{q}_j\|_{\mathrm{C}_j}^2}{\theta_j^*} \left(\frac{\eta_j}{2} + \sqrt{\frac{\eta_j^2}{4} + \frac{\|\widehat{\vec{q}}_j\|_{\mathrm{C}_j}^2}{2\theta_j^*}} \right)^{-1} \right. \\ & \left. + \left(\frac{\eta_j}{2} + \sqrt{\frac{\eta_j^2}{4} + \frac{\|\widehat{\vec{q}}_j\|_{\mathrm{C}_j}^2}{2\theta_j^*}} \right) - \eta_j \log \theta_j^* \left(\frac{\eta_j}{2} + \sqrt{\frac{\eta_j^2}{4} + \frac{\|\widehat{\vec{q}}_j\|_{\mathrm{C}_j}^2}{2\theta_j^*}} \right) \right\}. \end{aligned}$$

Hyperparameters and depth weighting

In the limit as $\eta_j \to 0+$,

$$\mathscr{E}(Q, F(Q)) \to \frac{1}{2}\|b - \sum_{j=1}^{n} \mathsf{M}_j \vec{q}_j\|_{\Sigma}^2 + \sqrt{2}\sum_{j=1}^{n} \frac{\|\vec{q}_j\|_{\mathsf{C}_j}}{\sqrt{\theta_j^*}},$$

which is a **weighted minimum current penalty** with weight $1/\sqrt{\theta_j^*}$.

Conclusion: The parameter β_j controls *sparsity promotion* of the prior.

Question: *Can* θ_j^* *be related to depth sensitivity?*

Hyperparameters and depth weighting

Hypermodel

$$\pi^j_{\mathrm{hyper}} \sim \mathrm{Gamma}(\beta_j, \theta^*_j)$$

implies that

$$\mathsf{E}\{\theta_j \mid \beta_j, \theta^*_j\} = \beta_j \theta^*_j.$$

Assume that the observed magnetic field is generated by a single dipole:

$$b = \mathsf{M}_j \vec{q}_j + \varepsilon, \quad \vec{q}_j \sim \mathcal{N}(0, \theta_j \mathsf{C}_j).$$

From the observation that

$$\mathsf{E}\{\mathsf{E}\{\vec{q}_j \vec{q}_j^{\mathsf{T}} \mid \theta_j\} \mid \theta^*_j\} = \mathsf{E}\{\theta_j C_j \mid \theta^*_j\} = \beta_j \theta^*_j C_j$$

it follows that the covariance of b conditioned on θ^*_j is

$$\Phi = \mathsf{E}\{bb^{\mathsf{T}} \mid \theta^*\} = \beta_j \theta^*_j \mathsf{M}_j \mathsf{C}_j \mathsf{M}_j^{\mathsf{T}} + \Sigma.$$

Hyperparameters and depth weighting

To derive an expression for θ_j^*, we take the trace of the first and last term and obtain

$$\begin{aligned} \theta_j^* &= \frac{\mathrm{trace}(\Phi) - \mathrm{trace}(\Sigma)}{\beta_j \|\mathsf{M}_j \mathsf{C}_j^{1/2}\|_{\mathrm{F}}^2} \\ &= \frac{\text{power of a noiseless signal}}{\text{sensitivity to } j\text{th dipole}}. \end{aligned}$$

To estimate Φ, use an empirical Bayesian approach: Given an observation time series,

$$\mathsf{B} = \begin{bmatrix} b^{(1)} & b^{(2)} & \cdots & b^{(T)} \end{bmatrix} \in \mathbb{R}^{m \times T},$$

approximate

$$\Phi \approx \frac{1}{T} \sum_{j=1}^{T} b^{(j)} \left(b^{(j)}\right)^{\mathsf{T}},$$

and further

$$\mathrm{trace}(\Phi) \approx \frac{1}{T} \sum_{j=1}^{T} \mathrm{trace}\left((b^{(j)}(b^{(j)})^{\mathsf{T}}\right) = \frac{1}{T} \sum_{j=1}^{T} \|b^{(j)}\|^2.$$

Hyperparameters and depth weighting

Observations:

- If the dipole orientation is known with certainty,

$$\mathsf{C}_j = \vec{e}\,\vec{e}^{\mathsf{T}},$$

  implying

$$\mathrm{trace}(\mathsf{M}_j\mathsf{C}_j\mathsf{M}_j^{\mathsf{T}}) = \mathrm{trace}(\mathsf{M}_j\vec{e}\,(\mathsf{M}_j\vec{e})^{\mathsf{T}}) = \|\mathsf{M}_j\vec{e}\|^2,$$

  which is the signal power of the unit dipole.
- The sensitivity of deep dipoles is smaller than superficial ones, hence

$$\theta^*_{\mathrm{deep}} > \theta^*_{\mathrm{superficial}},$$

  therefore deep dipoles are penalized less, as in sensitivity weighting.

Hyperparameters and depth weighting

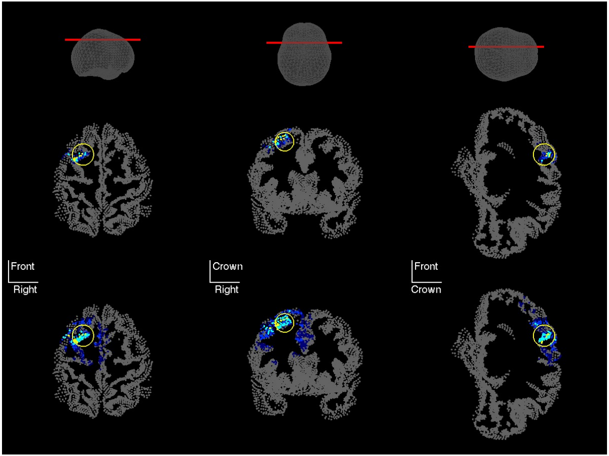
The sensitivity weighting gives deep dipoles a chance to explain the data, however:

- The expected variance of a deep (or radial) dipole with low sensitivity should not exceed a physiologically reasonable level.
- To avoid unrealistically large deep dipoles, introduce the truncation:

$$\theta_j^* = \min\left\{\frac{\mathrm{trace}(\Phi) - \mathrm{trace}(\Sigma)}{\beta_j \|\mathsf{M}_j \mathsf{C}_j^{1/2}\|_{\mathrm{F}}^2}, \theta_{\max}^*\right\},$$

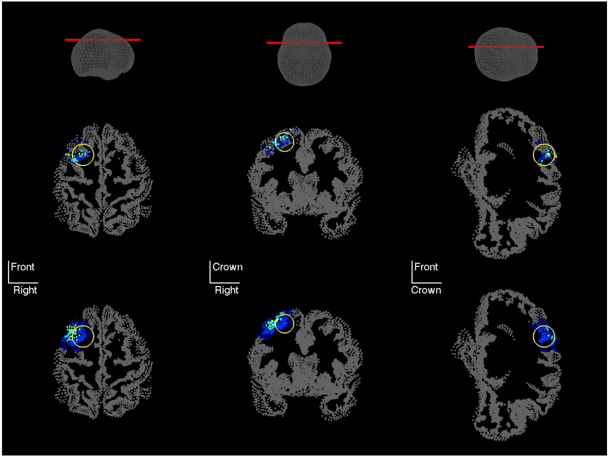
where $\theta_{\max}^*$ is a physiologically meaningful upper bound.

Computed examples: Effect of hyperparameter



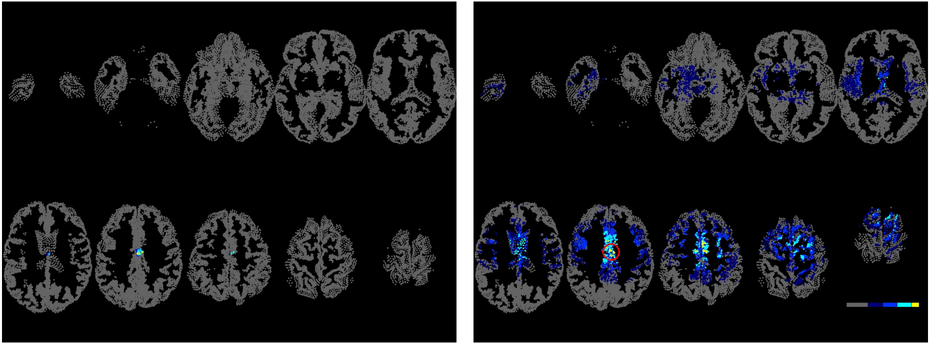
Estimated activity with two different hyperparameter values, $\eta = 0.005$ (upper row) and $\eta = 0.05$ (lower row)

Computed examples: Effect of anatomical prior



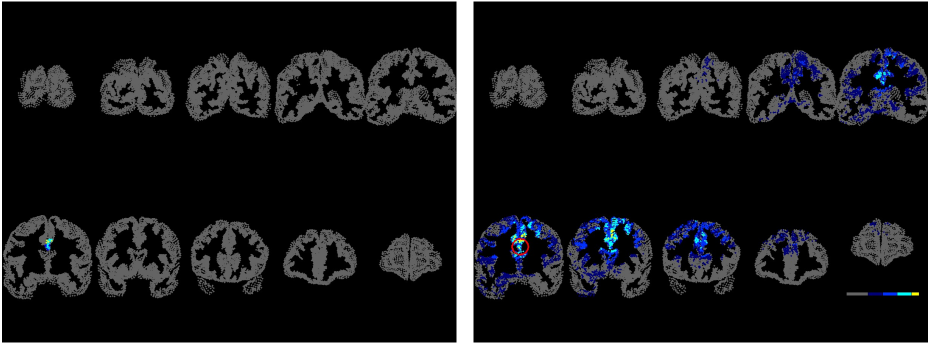
Estimates with the anatomical prior (top) and without (bottom). Here, $\eta = 0.005$ and $SNR = 15$.

Estimating deep sources: axial view



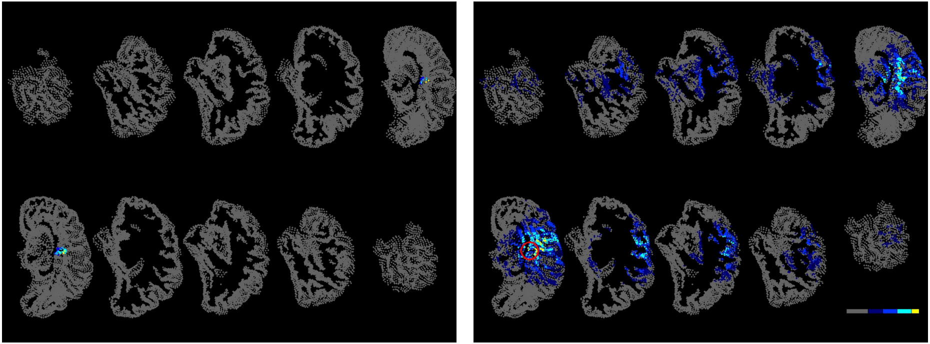
Original (left) and estimated (right) deep activity patch with Full Monty.

Estimating deep sources: coronal view



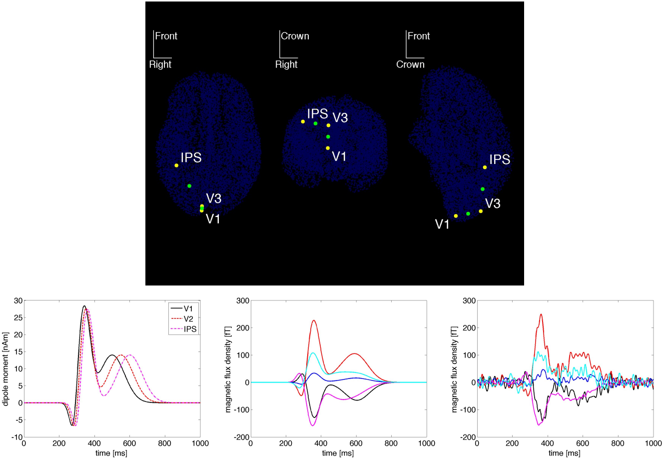
Original (left) and estimated (right) deep activity patch with Full Monty.

Estimating deep sources: sagittal view

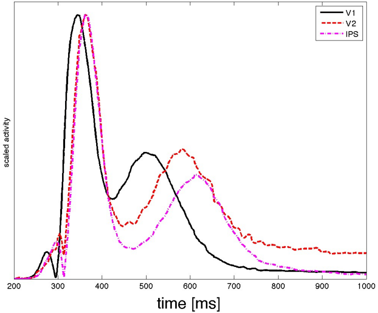


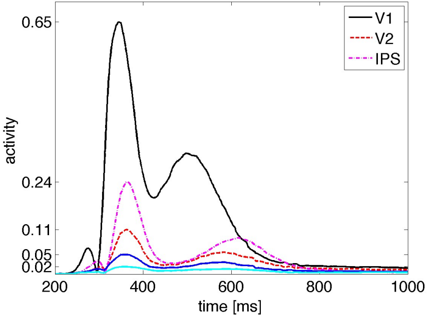
Original (left) and estimated (right) deep activity patch with Full Monty.

Computed examples: Time dependent sources



Computed examples: Time dependent sources





Conclusions

1. IAS combined with Bayes-Krylov solver leads to a fast and efficient numerical algorithm.
2. Hyperparameters control focality and depth sensitivity of the algorithm
3. No need for artificial depth weighting
4. Anatomical prior helps to discern physiologically meaningful source combinations without forcing the dipole orientations, accounting for uncertainties in the interpretation of segmented MRI images.
5. Numerical approximation of a globally convergent optimization method.

Reference

D Calvetti, A Pascarella, F Pitolli, E Somersalo, B Vantaggi: A hierarchical Krylov-Bayes iterative inverse solver for MEG with anatomical preconditioning. Inverse Problems 31(2015) 125005