Centrum Wiskunde & Informatica

Hierarchical Bayesian Uncertainty Quantification for EEG/MEG Source Reconstruction

Yousra Bekhti, Felix Lucka, Joseph Salmon, Alexandre Gramfort

International Congress on Industrial and Applied Mathematics
Valencia, Spain
17 July 2019

Introduction

Sparse Regression / Inverse Problems

Find sparse approximate solution to

$$\mathbf{m} = \mathbf{G}\mathbf{x} + \mathbf{e} \ , \qquad \mathbf{G} \in \mathbb{R}^{m \times n}, \qquad m \ll n$$

In general, combinatorial complexity!

Sparse Regression / Inverse Problems

Find sparse approximate solution to

$$\mathbf{m} = \mathbf{G}\mathbf{x} + \mathbf{e} \ , \qquad \mathbf{G} \in \mathbb{R}^{m \times n}, \qquad m \ll n$$

In general, combinatorial complexity!

Convexification:

$$\hat{\mathbf{x}}_\lambda \in \underset{x}{\operatorname{argmin}} \ \frac{1}{2} \left\| \mathbf{m} - \mathbf{G}\mathbf{x} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_1$$

- well understood for certain **G**, **x**, *e.g.*, **[Foucart & Rauhut '13]**
- otherwise, ad-hoc removal of ambiguity.
- bias: location and amplitude

Sparse Regression / Inverse Problems

Find sparse approximate solution to

$$\mathbf{m} = \mathbf{G}\mathbf{x} + \mathbf{e} \;, \qquad \mathbf{G} \in \mathbb{R}^{m \times n}, \qquad m \ll n$$

In general, combinatorial complexity!

Non-convex regression

$$\hat{\mathbf{x}}_\lambda \in \underset{x}{\operatorname{argmin}} \; \tfrac{1}{2} \left\| \mathbf{m} - \mathbf{G}\mathbf{x} \right\|_2^2 + \lambda \left\| \mathbf{x} \right\|_p^p \qquad 0 < p < 1$$

- ambiguity hidden in local-minima, quantification?
- empirically: less bias, superior performance
- computationally challenging

Majorization-Minimization (MM)

Idea: Minimize objective by successively minimizing upper bounds.

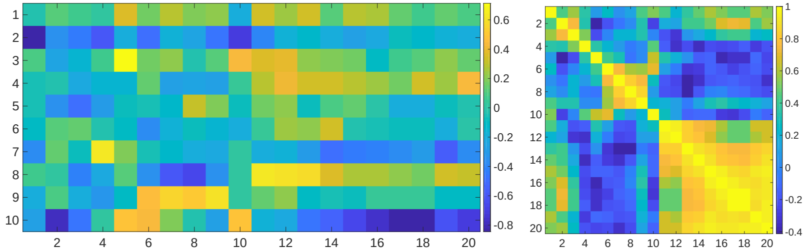
Use of ℓ_1-majorants for non-convex $\ell_{1/2}$-regression:

$$\hat{\mathbf{x}}^{(k)} \in \underset{\mathbf{x}}{\operatorname{argmin}} \, \frac{1}{2} \|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{n} \frac{|\mathbf{x}_i|}{2\sqrt{\left|\mathbf{x}_i^{(k-1)}\right|}}$$

- *Adaptive LASSO*, **[Zou 2006]**
- converges to local minimum $\Longrightarrow$ good initialization crucial!
- still no quantification of ambiguity

Strohmeier, Bekhti, Haueisen, Gramfort.
The Iterative Reweighted Mixed-Norm Estimate for Spatio-Temporal MEG/EEG Source Reconstruction. *IEEE-TMI, 2016*.

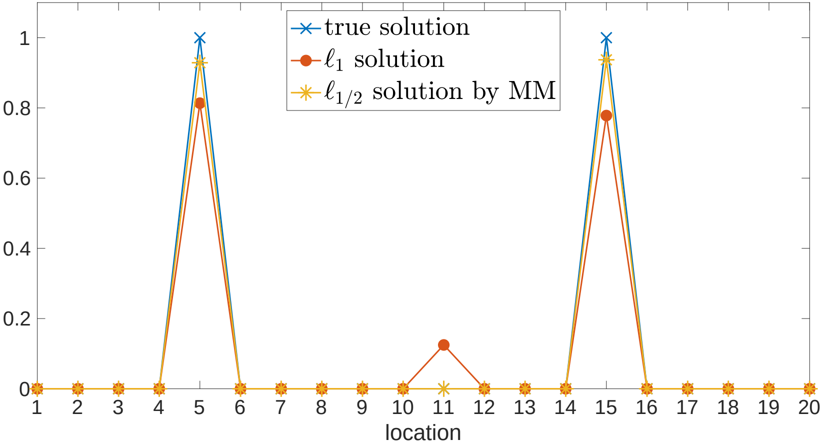
An Illustrative Example in 1D



(a) G **(b) $\mathbf{G}^\top\mathbf{G}$**

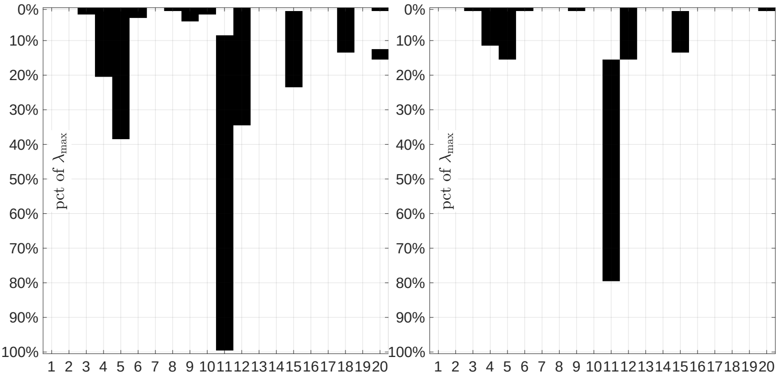
- $n = 20$
- **G**: low/ high coherence for locations $1 - 10$/ $11 - 20$
- $\mathbf{x}_5 = \mathbf{x}_{15} = 1$

Illustrative Example: ℓ_1 vs $\ell_{1/2}$



- 1% relative noise added
- higher sparsity, lower discrepancy $\|\mathbf{m} - \mathbf{Gx}\|_2$

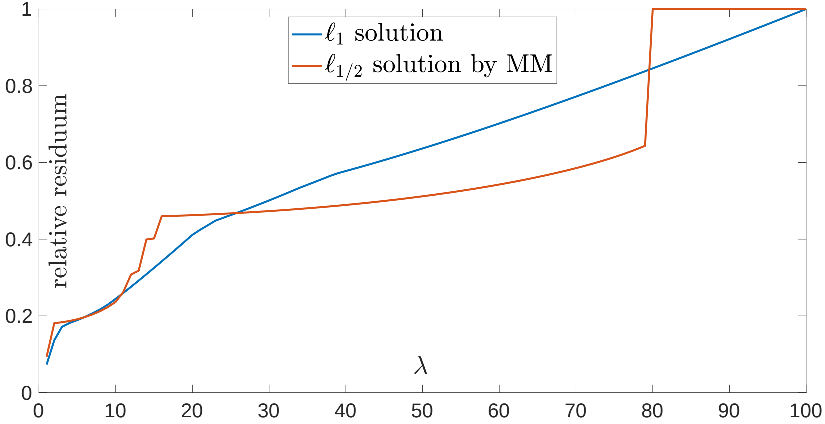
Illustrative Example: Support Recovery



(a) ℓ_1 **(b)** $\ell_{1/2}$

- 20% relative noise added
- $\lambda_{max} = \mathbf{G}^\top \mathbf{m}$
- No recovery of true support

Illustrative Example: Non-Convexity



Hierarchical Bayesian Models (HBM)

Uncertainty Quantification (UQ)? $\Longrightarrow$ Bayesian framework:

$$p_{post}(\mathbf{x}|\mathbf{m}) = \frac{p_{like}(\mathbf{m}|\mathbf{x})\, p_{prior}(\mathbf{x})}{p(\mathbf{m})} \;, \qquad p_{like}(\mathbf{m}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\left\|\mathbf{m} - \mathbf{G}\mathbf{x}\right\|_2^2\right)$$

Sparse priors?

- ℓ_1-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \left\|\mathbf{x}\right\|_1\right)$?
  ! dead-end **[Gribonval et al. '11,'12, Burger & FL '14, FL '14]**

Hierarchical Bayesian Models (HBM)

Uncertainty Quantification (UQ)? $\Longrightarrow$ Bayesian framework:

$$p_{post}(\mathbf{x}|\mathbf{m}) = \frac{p_{like}(\mathbf{m}|\mathbf{x})p_{prior}(\mathbf{x})}{p(\mathbf{m})} \ , \qquad p_{like}(\mathbf{m}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2\right)$$

Sparse priors?

- ℓ_1-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \|\mathbf{x}\|_1\right)$?
  ! dead-end **[Gribonval et al. '11,'12, Burger & FL '14, FL '14]**
- ℓ_p-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \|\mathbf{x}\|_p\right)$ with $0 < p < 1$?
  ! computationally challenging

Hierarchical Bayesian Models (HBM)

Uncertainty Quantification (UQ)? $\Longrightarrow$ Bayesian framework:

$$p_{post}(\mathbf{x}|\mathbf{m}) = \frac{p_{like}(\mathbf{m}|\mathbf{x}) p_{prior}(\mathbf{x})}{p(\mathbf{m})} \ , \qquad p_{like}(\mathbf{m}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\left\|\mathbf{m} - \mathbf{G}\mathbf{x}\right\|_2^2\right)$$

Sparse priors?

- ℓ_1-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \left\|\mathbf{x}\right\|_1\right)$?
  ! dead-end **[Gribonval et al. '11,'12, Burger & FL '14, FL '14]**
- ℓ_p-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \left\|\mathbf{x}\right\|_p\right)$ with $0 < p < 1$?
  ! computationally challenging
- Hierarchical ℓ_p-norm prior, $p \geqslant 1$

$$p_{prior}(\mathbf{x}|\gamma) \propto \exp\left(-\sum_{i=1}^{n} \frac{\left\|\mathbf{x}_i\right\|^p}{\gamma_i}\right),$$

Hierarchical Bayesian Models (HBM)

Uncertainty Quantification (UQ)? $\Longrightarrow$ Bayesian framework:

$$p_{post}(\mathbf{x}|\mathbf{m}) = \frac{p_{like}(\mathbf{m}|\mathbf{x}) p_{prior}(\mathbf{x})}{p(\mathbf{m})} \, , \qquad p_{like}(\mathbf{m}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}\left\|\mathbf{m} - \mathbf{G}\mathbf{x}\right\|_2^2\right)$$

Sparse priors?

- ℓ_1-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \left\|\mathbf{x}\right\|_1\right)$?
  ! dead-end **[Gribonval et al. '11,'12, Burger & FL '14, FL '14]**
- ℓ_p-norm priors, $p_{prior}(\mathbf{x}) \propto \exp\left(-\lambda \left\|\mathbf{x}\right\|_p\right)$ with $0 < p < 1$?
  ! computationally challenging
- Hierarchical ℓ_p-norm prior, $p \geqslant 1$, with generalized Gamma hyperprior

$$p_{prior}(\mathbf{x}|\gamma) \propto \exp\left(-\sum_{i=1}^{n} \frac{\left\|\mathbf{x}_i\right\|^p}{\gamma_i}\right), \quad p_{hyper}(\gamma) \propto \prod_i \gamma_i^{r\alpha-1} \exp\left(-\frac{\gamma_i^r}{\beta^r}\right)$$

HBM Inference

In the spirit of **[Calvetti, Somersalo et al. '07,'08,'09,'10]**,
we use full posterior

$$p_{post}(\mathbf{x}, \gamma | \mathbf{m}) \propto$$
$$\exp\left(-\frac{1}{2} \|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2 - \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|^p}{\gamma_i} + \frac{\gamma_i^r}{\beta^r} - (r\alpha - 1 - 1/p) \log \gamma_i \right)$$

Algorithms (optimization/sampling) exploit conditional structure and alternate

- Fix γ, update $\mathbf{x}$ by solving n-dim ℓ_2-ℓ_p problem.
- Fix $\mathbf{x}$, update γ by solving n 1-dim, smooth, non-linear problems.

Our aim:
Full *maximum-a-posteriori* estimates (full-MAP), posterior samples

Revisiting MM through HBM

Choose $p = 1$, $r = 1$ (Gamma distribution), $\alpha = 1 + 1/p$:

$$\exp\left(-\tfrac{1}{2}\|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2 - \sum_{i=1}^{n} \frac{\|\mathbf{x}_i\|}{\gamma_i} + \frac{\gamma_i}{\beta}\right)$$

Full-MAP alternation:

$$\mathbf{x}^{(k)} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \frac{1}{2}\|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2 + \sum_{i=1}^{n} \frac{|\mathbf{x}_i|}{\gamma_i^{(k-1)}} \right\}$$

$$\gamma_i^{(k)} = \sqrt{\beta}\sqrt{\left|\mathbf{x}_i^{(k)}\right|}, \quad \forall i$$

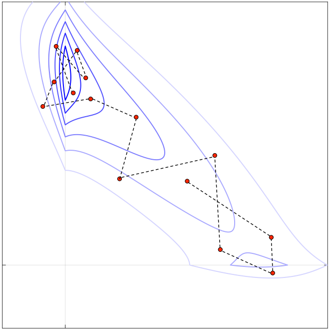
Choose $\beta = 4/\lambda^2$ and substitute γ_i to recover MM:

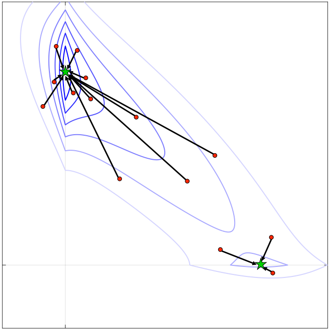
$$\hat{\mathbf{x}}^{(k)} \in \underset{\mathbf{x}}{\operatorname{argmin}}\, \tfrac{1}{2}\|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{n} \frac{|\mathbf{x}_i|}{2\sqrt{\left|\mathbf{x}_i^{(k-1)}\right|}}$$

$\longrightarrow$ stochastic framework behind MM! Does that help?

Mode Analysis with MCMC & Optimization

- MCMC chain of posterior samples.
- every sample initialization of local optimization
- chain of modes: sparse solutions with relative posterior mass





Efficient MCMC Sampling

$$\exp\left(-\frac{1}{2}\|\mathbf{m} - \mathbf{G}\mathbf{x}\|_2^2 - \sum_{i=1}^{n} \frac{|\mathbf{x}_i|^p}{\gamma_i} + \frac{\gamma_i}{\beta}\right)$$

Blocked Gibbs sampler:

$$\mathbf{x}^{(k)} \sim \; p_{post}(\mathbf{x} \quad , \boldsymbol{\gamma}^{(k-1)} | \mathbf{m}) \propto p_{post}(\mathbf{x}|\mathbf{m}, \gamma^{(k-1)}) \tag{1}$$

$$\gamma^{(k)} \sim \; p_{post}(\mathbf{x}^{(k)}, \boldsymbol{\gamma} \qquad | \mathbf{m}) \propto p_{post}(\boldsymbol{\gamma}|\mathbf{m}, \mathbf{x}^{(k)} \quad) \tag{2}$$

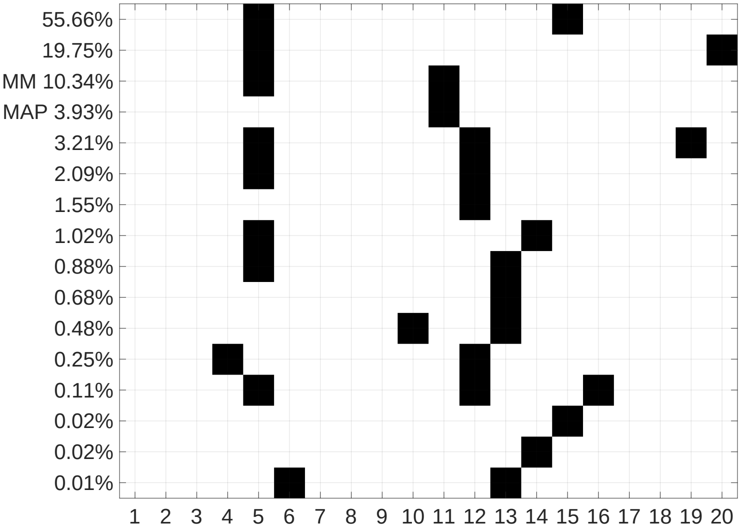
(1) via single component Gibbs sampling with nested slice sampling (*Slice-Within-Gibbs*) $\longrightarrow$ efficient due to fast burn-in

(2) via tailored accept-reject sampler

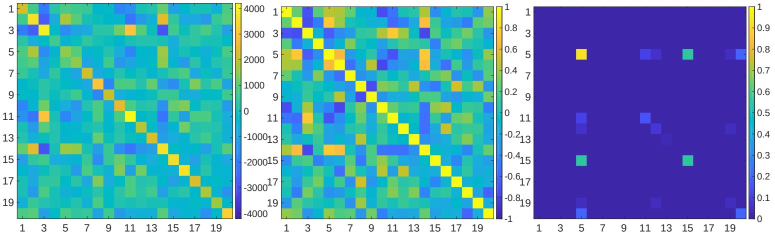
FL. Fast Gibbs sampling for high-dimensional Bayesian inversion *Inverse Problems, 2016*.

FL. Bayesian Inversion in Biomedical Imaging. *PhD Thesis, 2014*.

Illustrative Example: Support Frequency

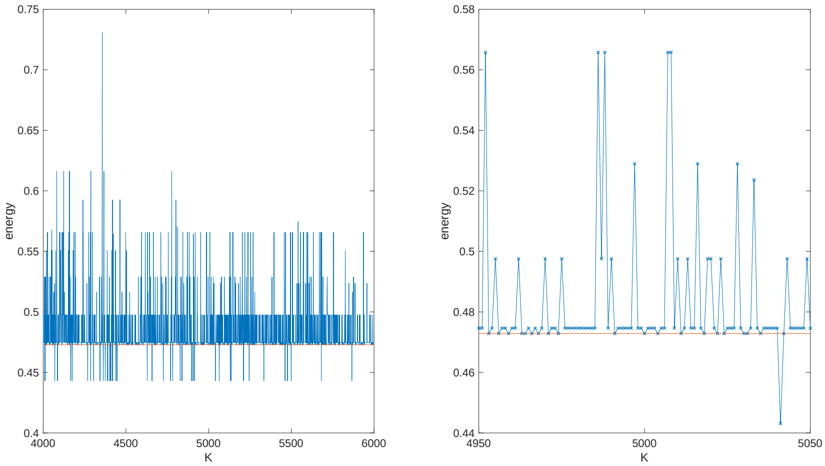


Comparison to Traditional UQ



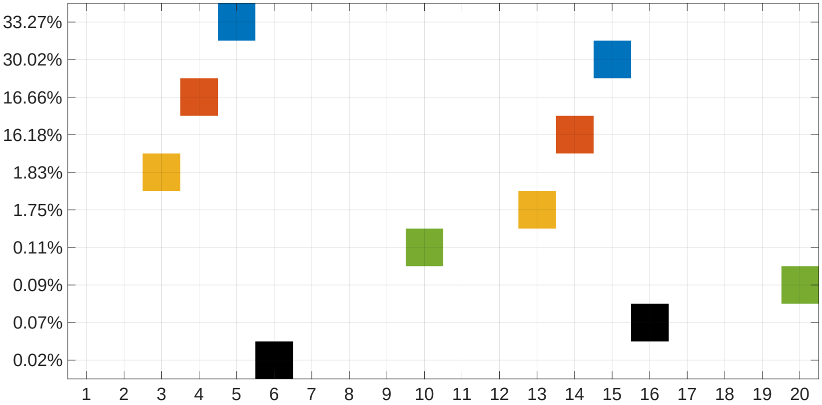
(a) covariance **(b)** correlation **(c)** simultaneous support activation

Illustrative Example: Sampler Mobility



- mode switch after 1.63 steps on average

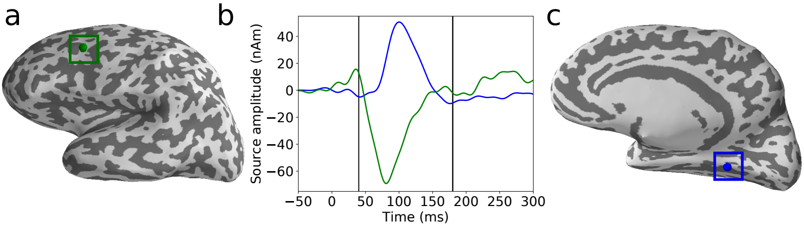
Symmetric Example: Support Frequency



- every mode has mirrored twin: $\mathbf{G}(:, 1{:}10) = \mathbf{G}(:, 11{:}20)$.

EEG/MEG Results

Simulated Data Example



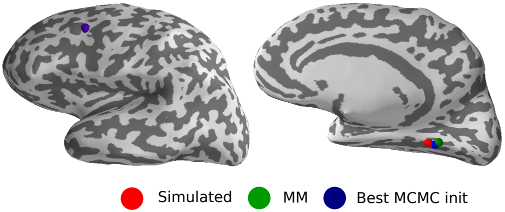
- superficial (motor cortex) and deep (inferior occipital gyrus) source
- $n = 7498$ cortical locations, $m = 306$ MEG sensors
- MNE open-source python toolbox, www.martinos.org/mne

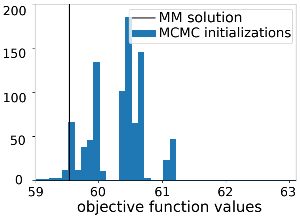
Multiple measurement vector (MMV) model:

$$\mathbf{M} = \mathbf{G}\mathbf{X} + \mathbf{E}$$

with $\mathbf{M}, \mathbf{E} \in \mathbb{R}^{m \times t}, \mathbf{G} \in \mathbb{R}^{m \times dn}, \mathbf{X} \in \mathbb{R}^{dn \times t}$.
$\longrightarrow$ all ℓ_1 terms become $\ell_{2,1}$ norms (*group sparsity*)

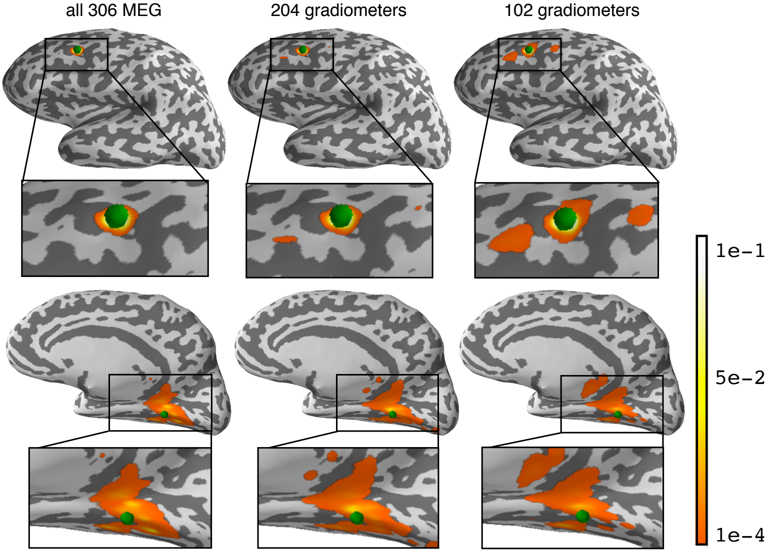
Simulated Data: Can We Improve Standard MM?



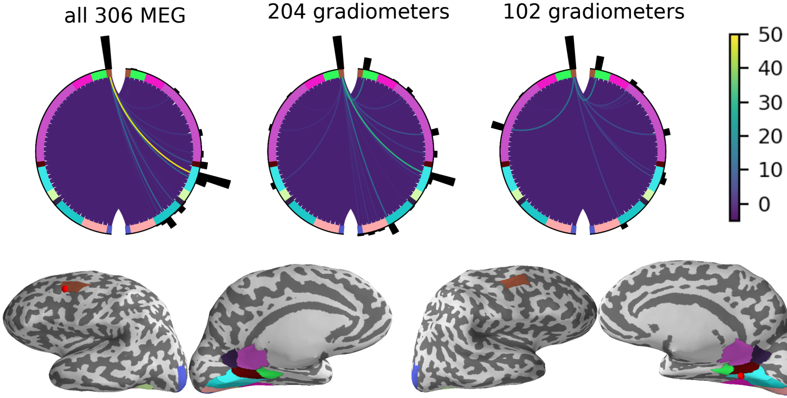


- ✓ better solution *w.r.t.* objective function
- ✓ better solution *w.r.t.* source distance

Simulated Data: Sparse Support Probability

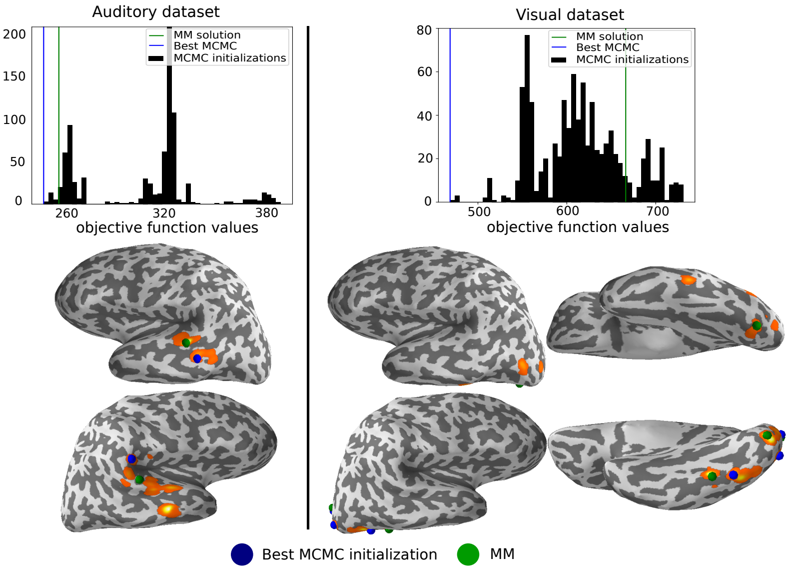


Simulated Data: Source Network Analysis



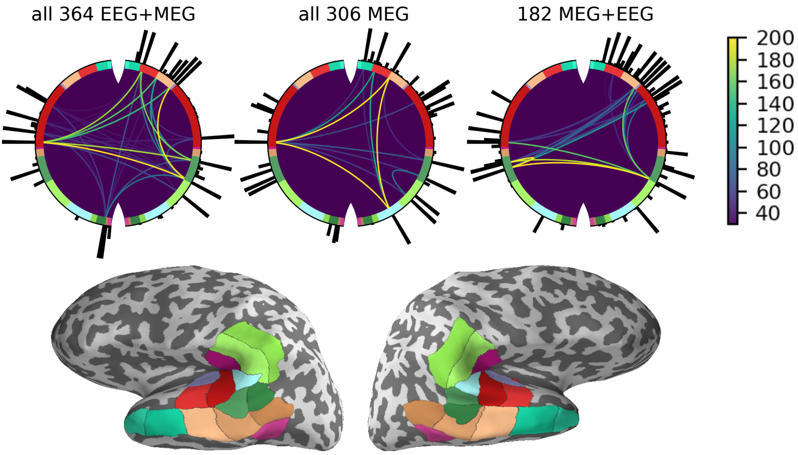
- increasing under-determinedness leads to higher ambiguity

Experimental Data: Can We Improve Standard MM?



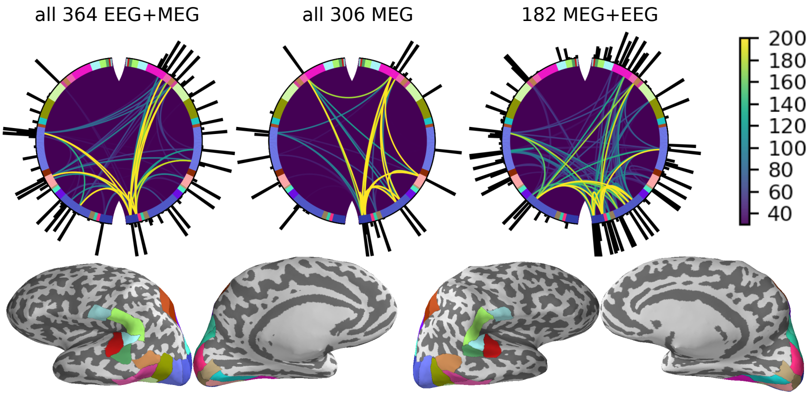
open MNE data sets [Gramfort et al., 2013]

Auditory Data: Source Network Analysis



- adding EEG produces more prominent modes $\longrightarrow$ less ambiguity
- more activity in the dark green area (auditory response)

Visual Data: Source Network Analysis



- higher complexity, posterior mass distributed widely
- majority of active sources in right hemisphere, consistent with checker board stimulation on left visual hemifield

Conclusion

Conclusions

Exploit link between MM for $\ell_{1/2}$-regression and ℓ_1-HBM:

- combine state-of-the-art optimization & sampling
- MCMC can jump between modes rapidly
- improve MM initializations
- reveal and explore different solutions to regression problem
- sparse Bayesian UQ vs. traditional Bayesian UQ

Sparse uncertainty quantification for EEG/MEG:

- result not single source configuration, but multiple hypotheses with quantification of their uncertainty
- helpful in clinical applications?

First steps here, need to develop towards consistent framework!

Bekhti, FL, Salmon, Gramfort. A hierarchical Bayesian perspective on majorization-minimization for non-convex sparse regression: application to M/EEG source imaging. *Inverse Problems, 2018*.

Strohmeier, Bekhti, Haueisen, Gramfort. The Iterative Reweighted Mixed-Norm Estimate for Spatio-Temporal MEG/EEG Source Reconstruction. *IEEE-TMI, 2016*.

FL. Fast Gibbs sampling for high-dimensional Bayesian inversion. *Inverse Problems, 2016*.

Thank you for your attention!
Questions?

- **Bekhti, FL, Salmon, Gramfort.**
  A hierarchical Bayesian perspective on majorization-minimization for non-convex sparse regression: application to M/EEG source imaging.
  Inverse Problems, 2018.
- **Strohmeier, Bekhti, Haueisen, Gramfort.** The Iterative Reweighted Mixed-Norm Estimate for Spatio-Temporal MEG/EEG Source Reconstruction. *IEEE-TMI, 2016*.
- **FL.** Fast Gibbs sampling for high-dimensional Bayesian inversion. *Inverse Problems, 2016*.