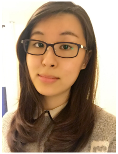
Hierarchical Bayesian Learning for Noise Estimation in Electromagnetic Brain Source Imaging

Ali Hashemi
Technische Universität Berlin

joint work with Chang Cai, Yijing Gao, Sanjay Ghosh,
Klaus-Robert Müller, Srikantan S. Nagarajan, and Stefan Haufe

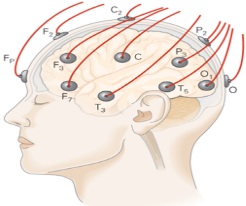
The 22nd International Conference on Biomagnetism (BIOMAG)
August 28th, 2022

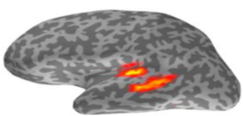
Electromagnetic Brain Source Imaging (BSI)

Electro-/Magnetoencephalography (E/MEG): A non-invasive brain imaging technique with high temporal resolution (order of ms).

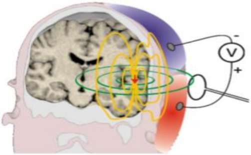
$\mathbf{Y} = \mathbf{L}\mathbf{X} + \mathbf{E}$	**Linear regression problem**
$\mathbf{Y} \in \mathbb{R}^{M \times T}$	M:#Sensors, T:#Samples,
$\mathbf{X} \in \mathbb{R}^{N \times T}$	N:#Sources, ($M \ll N$)
$\mathbf{L} \in \mathbb{R}^{M \times N}$	Lead Field Matrix (Known)



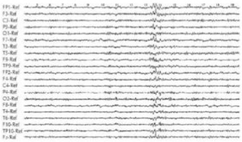
$\mathbf{X}$



$\mathbf{L}$

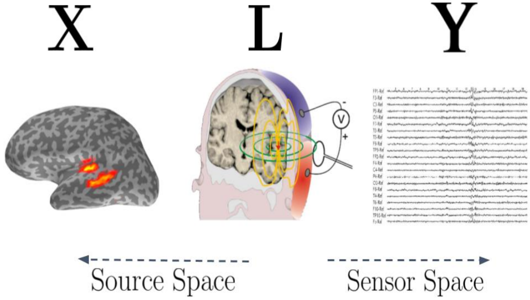


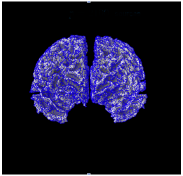
$\mathbf{Y}$



Source Space ⟵ - ⟶ Sensor Space

Electromagnetic Brain Source Imaging (BSI)





[Macsymenko '09][Massias '19]

Ill-posed inverse problem: (#Sensors$= 32 \sim 256$ vs #Sources$= 10^3 \sim 10^4$)

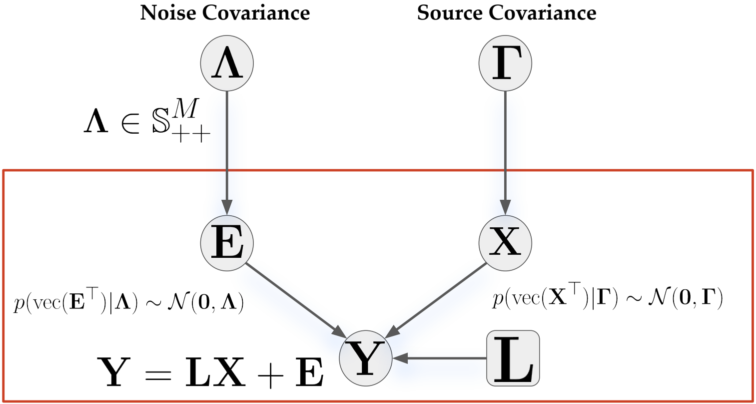
$$\mathbf{X}^* = \underset{\mathbf{X}}{\operatorname{argmin}} \; \underbrace{\|\mathbf{Y} - \mathbf{L}\mathbf{X}\|_F^2}_{\text{Likelihood:}p(\mathbf{Y}|\mathbf{X})} + \lambda \; \underbrace{\mathcal{R}(\mathbf{X})}_{\text{Prior:}p(\mathbf{X})}$$

1. Type-I MAP methods: ℓ_1, ℓ_2, $\ell_{1,2}$-norms, sparsity in transformed domains (Gabor).
   [Pascual-Marqui et al., '07][Haufe et al, '08, '11][Gramfort et al., '12, '13][Castaño-Candamil et al., '15]
2. Type-II ML approaches: different sparse Bayesian learning (SBL) variants ignoring the temporal dynamics. [Wipf et al., '09, '10, '11][Owen et al, '12][Cai et al., '17, '21]

Hierarchical Bayesian Learning

Probabilistic graphical model:

- Model the current activity of the brain sources as Gaussian scale mixtures
- $\mathbf{\Lambda}$: Noise covariance with full structure



Hierarchical Bayesian Inference and Type-II Loss

Posterior source distribution: $p(\mathbf{X}|\mathbf{Y}, \mathbf{\Gamma}, \mathbf{\Lambda}) = \prod_{t=1}^{T} \mathcal{N}(\bar{\mathbf{x}}(t), \mathbf{\Sigma}_{\mathbf{x}})$
where

$$\begin{aligned}
\bar{\mathbf{x}}(t) &= \mathbf{\Gamma}\mathbf{L}^{\top}(\mathbf{\Sigma}_{\mathbf{y}})^{-1}\mathbf{y}(t) \\
\mathbf{\Sigma}_{\mathbf{x}} &= \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{L}^{\top}(\mathbf{\Sigma}_{\mathbf{y}})^{-1}\mathbf{L}\mathbf{\Gamma} \\
\mathbf{\Sigma}_{\mathbf{y}} &= \mathbf{\Lambda} + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^{\top} .
\end{aligned}$$

$\mathbf{\Gamma}$, and $\mathbf{\Lambda}$ are learned by minimizing the negative log marginal likelihood (Type-II) loss, $-\log p(\mathbf{Y}|\mathbf{\Gamma}, \mathbf{\Lambda})$:

$$\text{Type} - \text{II Loss} : \mathcal{L}^{\text{II}}(\mathbf{\Gamma}, \mathbf{\Lambda}) = \log|\mathbf{\Lambda} + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^{\top}| + \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}(t)^{\top} \left(\mathbf{\Lambda} + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^{\top}\right)^{-1} \mathbf{y}(t) .$$

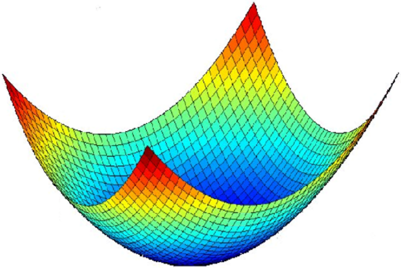
Challenges

$$\text{Type} - \text{II Loss}: \mathcal{L}^{\mathrm{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = \log|\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^{\top}| + \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}(t)^{\top}\left(\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^{\top}\right)^{-1}\mathbf{y}(t)\,.$$

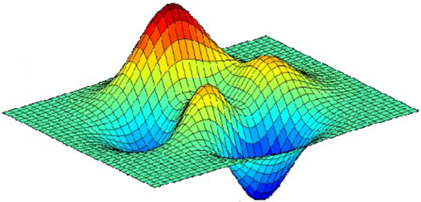
Challenges

$$\text{Type} - \text{II Loss}: \mathcal{L}^{\text{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = \log|\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top| + \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}(t)^\top \left(\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top\right)^{-1} \mathbf{y}(t) .$$

1. Non-convex Type-II ML loss function: Non-trivial to solve.



convex function

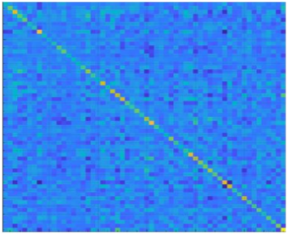


non-convex function

Challenges

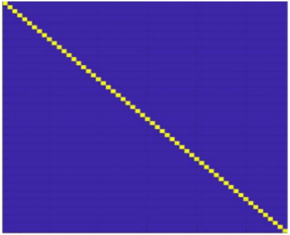
$$\text{Type} - \text{II Loss}: \mathcal{L}^{\text{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = \log|\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top| + \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}(t)^\top \left(\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top\right)^{-1}\mathbf{y}(t) .$$

1. Non-convex Type-II ML loss function: Non-trivial to solve.
2. Most contributions in the literature are limited to the estimation of only a diagonal noise covariance: Limiting assumption in practice.



Full structure

$\boldsymbol{\Lambda}$



Diagonal

Challenges

$$\text{Type} - \text{II Loss}: \mathcal{L}^{\text{II}}(\mathbf{\Gamma}, \mathbf{\Lambda}) = \log|\mathbf{\Lambda} + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top| + \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}(t)^\top \left(\mathbf{\Lambda} + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top\right)^{-1} \mathbf{y}(t) \,.$$

1. Non-convex Type-II ML loss function: Non-trivial to solve.
2. Most contributions in the literature are limited to the estimation of only a diagonal noise covariance: Limiting assumption in practice.

Our contribution: Joint estimation of Gaussian regression parameter distributions and Gaussian noise distributions with full covariance structure by solving the non-convex Type-II loss.

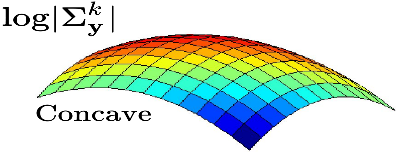
Convex Majorizing Functions

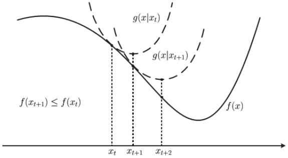
Theorem

Optimizing $\mathcal{L}^{\mathrm{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ with respect to $\boldsymbol{\Lambda}$ is equivalent to optimizing the following majorzing function:

$$\mathcal{L}^{\mathrm{conv}}_{\mathrm{noise}}(\boldsymbol{\Gamma}^k, \boldsymbol{\Lambda}) = \mathrm{tr}\left(\left(\boldsymbol{\Sigma}^k_{\mathbf{y}}\right)^{-1} \boldsymbol{\Lambda} \right) + \mathrm{tr}(\mathbf{M}^k_{\mathrm{N}} \boldsymbol{\Lambda}^{-1}) \,,$$

where $\mathbf{M}^k_{\mathrm{N}} := \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))^\top$ *and* $\boldsymbol{\Sigma}^k_{\mathbf{y}} = \boldsymbol{\Lambda}^k + \mathbf{L}\boldsymbol{\Gamma}^k\mathbf{L}^\top$.





[Sun et al., '17][Razzaviyan et al., '13]

$$\boxed{\mathcal{L}^{\mathrm{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = \log|\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top| + \frac{1}{T} \sum_{t=1}^{T} \mathbf{y}(t)^\top \left(\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top\right)^{-1} \mathbf{y}(t)}$$

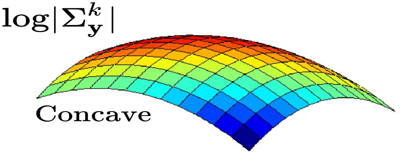
Convex Majorizing Functions

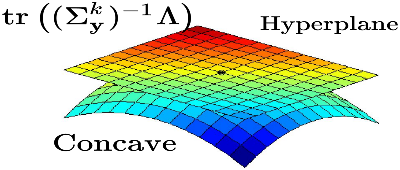
Theorem

Optimizing $\mathcal{L}^{\mathrm{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ with respect to $\boldsymbol{\Lambda}$ is equivalent to optimizing the following majorzing function:

$$\mathcal{L}^{\mathrm{conv}}_{\mathrm{noise}}(\boldsymbol{\Gamma}^k, \boldsymbol{\Lambda}) = \mathrm{tr}\left(\left(\boldsymbol{\Sigma}^k_{\mathbf{y}}\right)^{-1} \boldsymbol{\Lambda}\right) + \mathrm{tr}(\mathbf{M}^k_{\mathrm{N}} \boldsymbol{\Lambda}^{-1}) \ ,$$

where $\mathbf{M}^k_{\mathrm{N}} := \frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))^\top$ *and* $\boldsymbol{\Sigma}^k_{\mathbf{y}} = \boldsymbol{\Lambda}^k + \mathbf{L}\boldsymbol{\Gamma}^k\mathbf{L}^\top$.





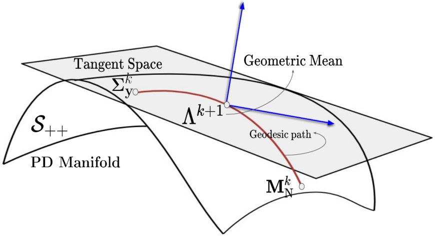
$$\mathcal{L}^{\mathrm{II}}(\boldsymbol{\Gamma}, \boldsymbol{\Lambda}) = \log|\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top| + \frac{1}{T}\sum_{t=1}^{T} \mathbf{y}(t)^\top \left(\boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^\top\right)^{-1} \mathbf{y}(t)$$

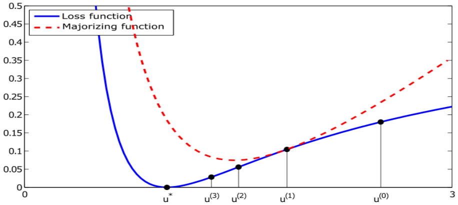
Theorem (Geometric mean)

$\mathcal{L}_{\text{noise}}^{\text{conv}}(\mathbf{\Gamma}^k, \mathbf{\Lambda})$ is geodesically convex with respect to the PD manifold, and its optimal solution with respect to $\mathbf{\Lambda}$ can be attained according to the following update rule:

$$\mathbf{\Lambda}^{k+1} \leftarrow (\mathbf{\Sigma}_{\mathbf{y}}^k)^{\frac{1}{2}} \left((\mathbf{\Sigma}_{\mathbf{y}}^k)^{-1/2} \mathbf{M}_{\mathrm{N}}^k (\mathbf{\Sigma}_{\mathbf{y}}^k)^{-1/2} \right)^{\frac{1}{2}} (\mathbf{\Sigma}_{\mathbf{y}}^k)^{\frac{1}{2}} \, .$$

which leads to a majorization-minimization (MM) algorithm with convergence guarantees $\rightsquigarrow$ Full-structural noise (FUN) learning algorithm.





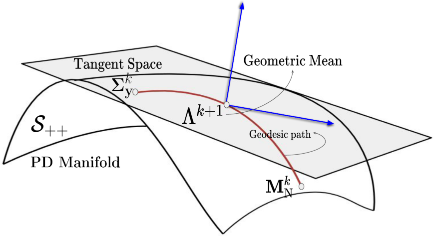
Riemannian Update on the Manifold of PD Matrices

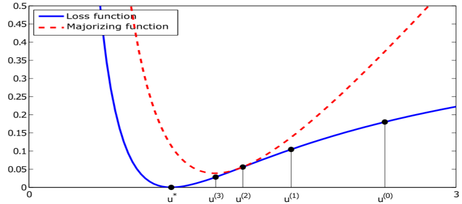
Theorem (Geometric mean)

$\mathcal{L}_{\text{noise}}^{\text{conv}}(\mathbf{\Gamma}^{k}, \mathbf{\Lambda})$ *is geodesically convex with respect to the PD manifold, and its optimal solution with respect to* $\mathbf{\Lambda}$ *can be attained according to the following update rule:*

$$\mathbf{\Lambda}^{k+1} \leftarrow (\mathbf{\Sigma}_{\mathbf{y}}^{k})^{\frac{1}{2}} \left((\mathbf{\Sigma}_{\mathbf{y}}^{k})^{-1/2} \mathbf{M}_{\mathrm{N}}^{k} (\mathbf{\Sigma}_{\mathbf{y}}^{k})^{-1/2} \right)^{\frac{1}{2}} (\mathbf{\Sigma}_{\mathbf{y}}^{k})^{\frac{1}{2}} \ .$$

which leads to a majorization-minimization (MM) algorithm with convergence guarantees $\rightsquigarrow$ *Full-structural noise (FUN) learning algorithm.*





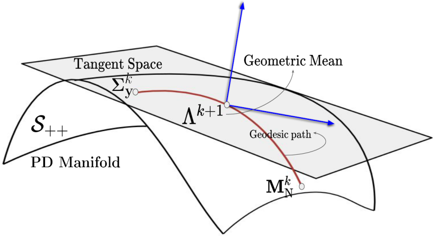
Riemannian Update on the Manifold of PD Matrices

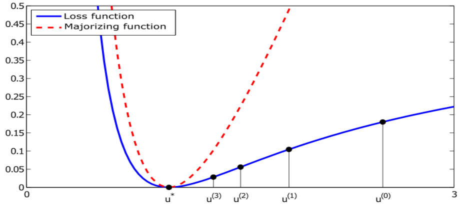
Theorem (Geometric mean)

$\mathcal{L}^{\text{conv}}_{\text{noise}}(\mathbf{\Gamma}^k, \mathbf{\Lambda})$ *is geodesically convex with respect to the PD manifold, and its optimal solution with respect to* $\mathbf{\Lambda}$ *can be attained according to the following update rule:*

$$\mathbf{\Lambda}^{k+1} \leftarrow (\mathbf{\Sigma}_{\mathbf{y}}^k)^{\frac{1}{2}} \left((\mathbf{\Sigma}_{\mathbf{y}}^k)^{-1/2} \mathbf{M}_{\mathrm{N}}^k (\mathbf{\Sigma}_{\mathbf{y}}^k)^{-1/2} \right)^{\frac{1}{2}} (\mathbf{\Sigma}_{\mathbf{y}}^k)^{\frac{1}{2}} \ .$$

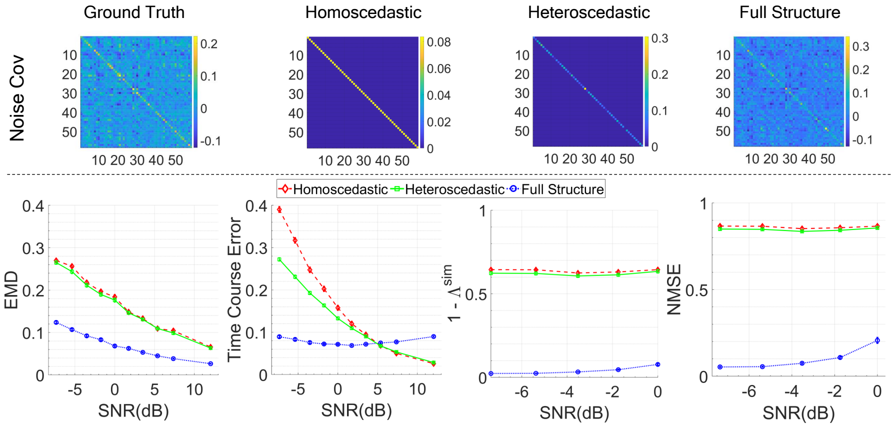
which leads to a majorization-minimization (MM) algorithm with convergence guarantees $\rightsquigarrow$ *Full-structural noise (FUN) learning algorithm.*





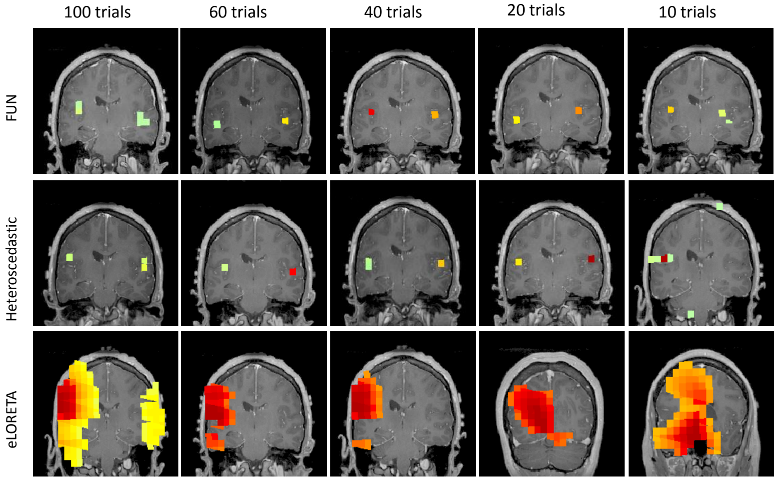
Numerical Results

Conclusion I: FUN learning consistently outperforms its homoscedastic and heteroscedastic counterparts according to all evaluation metrics.



Real Data Analysis of Auditory Evoked Fields (AEF)

Conclusion II: FUN learning can provide accurate reconstruction at the expected locations of the auditory cortex even under extreme SNR conditions - superior to benchmarks.



Thank you For yoUr attentioN!

References

A. Hashemi, Y. Gao, C. Cai, S. Ghosh, K.-R. Müller, S. S. Nagarajan and S. Haufe
Joint Learning of Full-structure Noise in Hierarchical Bayesian Regression Models.
Preprint submitted to TMI, 2022. Draft is available on bioRxiv.

A. Hashemi, C. Cai, K.-R. Müller, S. S. Nagarajan and S. Haufe
Joint Hierarchical Bayesian Learning of Full-structure Noise for Brain Source Imaging.
Medical Imaging meets NeurIPS (Med-NeurIPS) Workshop, 2020.

A. Hashemi, C. Cai, G. Kutyniok, K.-R. Müller, S. S. Nagarajan and S. Haufe
Unification of sparse Bayesian learning algorithms for electromagnetic brain imaging with the majorization minimization framework.
NeuroImage 239, 2021.

C. Cai, A. Hashemi, M. Diwakar, S. Haufe, K. Sekihara, S. S. Nagarajan
Robust estimation of noise for electromagnetic brain imaging with the Champagne algorithm.
NeuroImage 225, 2021.

C. Cai, L. Hinkley, Y. Gao, A. Hashemi, S. Haufe, K. Sekihara, S. S. Nagarajan
Empirical Bayesian Localization of Event-related Time-frequency Neural Activity Dynamics.
NeuroImage 258, 2021.

A. Hashemi and S. Haufe
Improving EEG Source Localization Through Spatio-Temporal Sparse Bayesian Learning.
26th IEEE European Signal Processing Conference (EUSIPCO), 2018.