Joint Hierarchical Bayesian Learning of Full-structure Noise for Brain Source Imaging

Ali Hashemi

Technische Universität Berlin

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

Medical Imaging meets NeurIPS (Med-NeurIPS) Workshop
December 12th, 2020

Electromagnetic Brain Source Imaging (BSI)

$\mathbf{Y} = \mathbf{L}\mathbf{X} + \mathbf{E}$	**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)

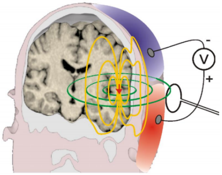


Figure by Lauri Parkkonen

Ill-posed inverse problem:

1. Type-I MAP methods: ℓ_1, ℓ_2, $\ell_{1,2}$-norms, sparsity in transformed domains (Gabor).
2. Type-II ML approaches: different sparse Bayesian learning (SBL) variants.

Hierarchical Bayesian regression:

- Source distribution: $p(\mathbf{X}|\boldsymbol{\Gamma}) = \prod_{t=1}^{T} \mathcal{N}(0, \boldsymbol{\Gamma})$, $\boldsymbol{\Gamma}$: Hyper-parameters
- Noise distribution: $\mathbf{E} = [\mathbf{e}(1), \ldots, \mathbf{e}(T)]$, $\mathbf{e}(t) \sim \mathcal{N}(0, \boldsymbol{\Lambda})$, $t = 1, \cdots, T$
  $\boldsymbol{\Lambda}$: Noise covariance with full structure
- Measurement distribution: $p(\mathbf{Y}|\mathbf{X}) = \prod_{t=1}^{T} \mathcal{N}(\mathbf{L}\mathbf{x}(t), \boldsymbol{\Lambda})$

Type-II Bayesian Learning

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

$$\bar{\mathbf{x}}(t) = \boldsymbol{\Gamma}\mathbf{L}^{\top}(\boldsymbol{\Sigma}_{\mathbf{y}})^{-1}\mathbf{y}(t) \quad \boldsymbol{\Sigma}_{\mathbf{x}} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{L}^{\top}(\boldsymbol{\Sigma}_{\mathbf{y}})^{-1}\mathbf{L}\boldsymbol{\Gamma} \quad \boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Lambda} + \mathbf{L}\boldsymbol{\Gamma}\mathbf{L}^{\top},$$

as a result of learning $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ through minimizing $-\log p(\mathbf{Y}|\boldsymbol{\Gamma},\boldsymbol{\Lambda})$:

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

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.

Convex Majorizing Functions

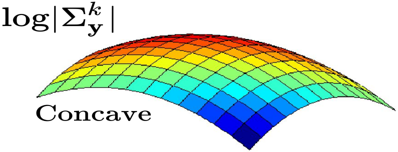
Theorem

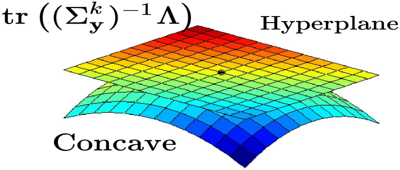
Optimizing $\mathcal{L}^{\mathrm{II}}(\mathbf{\Gamma}, \mathbf{\Lambda})$ *with respect to* $\mathbf{\Lambda}$ *is equivalent to optimizing the following majorzing function:*

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

where $\mathbf{C}_{\mathrm{N}}^k$ *and* $\mathbf{M}_{\mathrm{N}}^k$ *are defined as:*

$$\mathbf{C}_{\mathrm{N}}^k := \left(\mathbf{\Sigma}_{\mathbf{y}}^k\right) \ , \quad \mathbf{M}_{\mathrm{N}}^k := \frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))^\top \ .$$





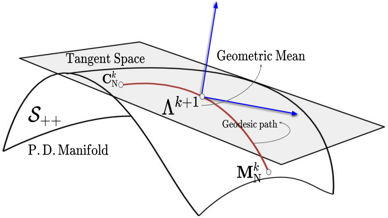
MM Framework with Convergence Guarantees

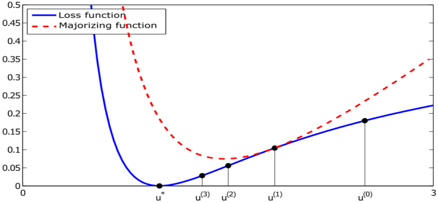
Theorem

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

$$\boldsymbol{\Lambda}^{k+1} \leftarrow (\mathbf{C}^k_{\mathrm{N}})^{\frac{1}{2}} \left((\mathbf{C}^k_{\mathrm{N}})^{-1/2} \mathbf{M}^k_{\mathrm{N}} (\mathbf{C}^k_{\mathrm{N}})^{-1/2} \right)^{\frac{1}{2}} (\mathbf{C}^k_{\mathrm{N}})^{\frac{1}{2}} \ ,$$

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





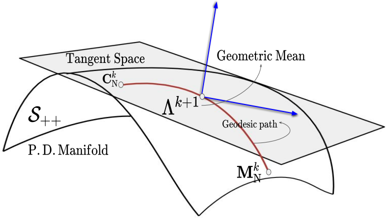
MM Framework with Convergence Guarantees

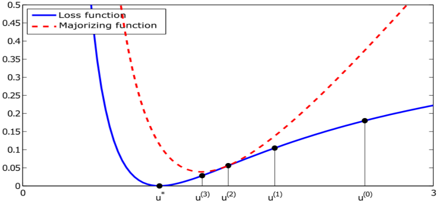
Theorem

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

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

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



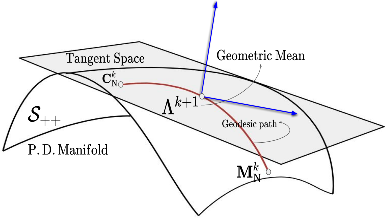


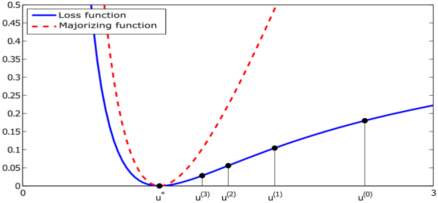
Theorem

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

$$\boldsymbol{\Lambda}^{k+1} \leftarrow (\mathbf{C}^k_{\mathrm{N}})^{\frac{1}{2}} \left((\mathbf{C}^k_{\mathrm{N}})^{-1/2} \mathbf{M}^k_{\mathrm{N}} (\mathbf{C}^k_{\mathrm{N}})^{-1/2} \right)^{\frac{1}{2}} (\mathbf{C}^k_{\mathrm{N}})^{\frac{1}{2}} ,$$

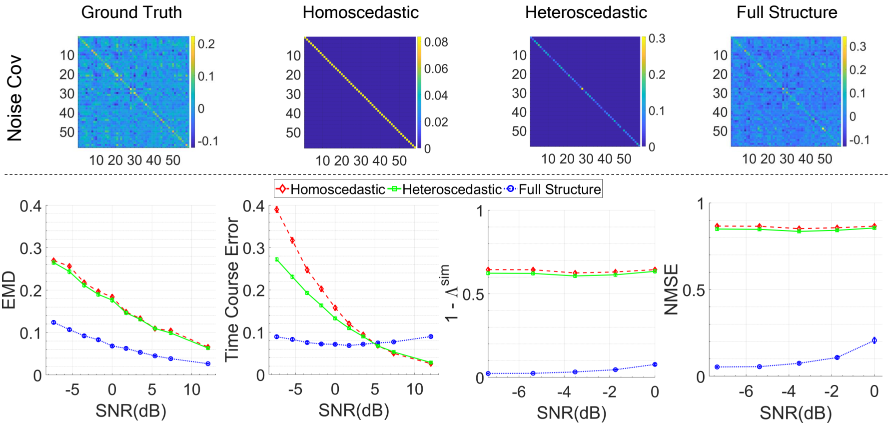
which leads to a majorization-minimization (MM) algorithm with convergence guarantees $\rightsquigarrow$ Full-structural noise (FUN) learning algorithm.



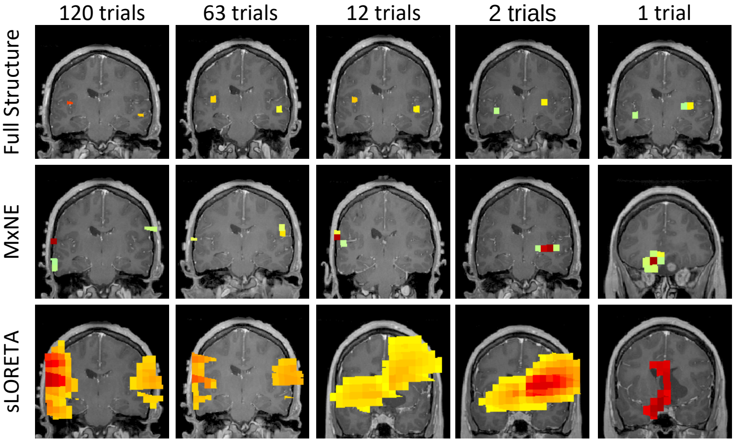


Numerical Results

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



Conclusion II: FUN learning can provide accurate reconstruction even under extreme SNR conditions - superior to benchmarks.



Thank you For yoUr attentioN!

References

A. Hashemi, C. Cai, K.-R. Müller, S. S. Nagarajan and S. Haufe
Joint Learning of Full-structure Noise in Hierarchical Bayesian Regression Models.
Preprint, 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.
Preprint, 2020. Draft is available on bioRxiv.

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, 2020.

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

K. Sekihara and S. S. Nagarajan
Electromagnetic Brain Imaging: A Bayesian Perspective.
Springer, 2015.