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

Ali Hashemi¹,²,³, Chang Cai⁴,⁵, Klaus-Robert Müller¹,⁶,⁷,⁸,*
Srikantan S Nagarajan⁴, and Stefan Haufe³,⁹,*

¹Machine Learning Group, Technische Universität Berlin, Germany.
²Institut für Mathematik, Technische Universität Berlin, Germany.
³Charité – Universitätsmedizin Berlin, Germany.
⁴National Engineering Research Center for E-Learning, Central China Normal University, China.
⁵Department of Radiology and Biomedical Imaging, University of California, San Francisco, USA.
⁶Berliner Zentrum für Maschinelles Lernen, Berlin, Germany.
⁷Department of Artificial Intelligence, Korea University, South Korea.
⁸Max Planck Institute for Informatics, Saarbrücken, Germany.
⁹Bernstein Center for Computational Neuroscience, Berlin, Germany.
*Corresponding authors

Abstract

Many problems in human brain imaging involve hierarchical Bayesian (type-II maximum likelihood) regression models for observations with latent variables for source and noise, where parameters of priors for source and noise terms need to be estimated jointly from data. One example are biomagnetic inverse problems, where crucial factors influencing accuracy of brain source estimation are not only the noise level but also its correlation structure. Importantly, existing approaches have not addressed estimation of a full-structure noise covariance matrix. Using ideas from Riemannian geometry, we derive an efficient algorithm for updating both source and a full-structure noise covariance along the manifold of positive definite matrices. Our results demonstrate that the novel framework significantly improves upon state-of-the-art techniques in the real-world scenario with fully-structured noise covariance.

1 Introduction

Having precise knowledge of the noise distribution is a fundamental requirement for obtaining accurate solutions in many hierarchical regression problems ubiquitous in various brain imaging applications [1]. In many applications however, it is impossible to separately estimate this noise distribution since distinct "noise-only" (baseline) measurements are not feasible. An alternative, therefore, is to design estimators that jointly optimize over the regression coefficients as well as over parameters of the noise distribution. This has been pursued in the literature before [2,7]. However, most contributions are limited to the estimation of only a diagonal noise covariance (i.e., independent between different measurements). Considering a diagonal noise covariance is a limiting assumption in practice as the noise interference in many realistic scenarios are highly correlated across measurements; and thus, have non-trivial off-diagonal elements. Here we argue that explicitly modeling the noise dependency structure is a key factor to improve the accuracy of a large class of regression problems including those introduced above.

Here, we focus on the problem of electromagnetic brain source imaging (BSI) as our main application. The goal of BSI is to reconstruct brain activity from magneto- or electroencephalography (M/EEG), which can be formulated as a sparse Bayesian learning (SBL) problem. Specifically, it can be cast as a linear Bayesian regression model with independent Gaussian scale mixture priors on the parameters and noise. As a departure from the classical SBL approaches, here we specifically consider Gaussian
noise with full covariance structure. Prominent source of correlated noise in this context are, for example, eye blinks, heart beats, muscular artifacts and line noise. Algorithms that can accurately estimate noise with full covariance structure are expected to achieve more accurate regression models and predictions in this setting.

2 Type-II Bayesian Regression for Brain Source Imaging

We focus on imaging of brain sources from M/EEG data. This inverse problem can be described by a linear forward model $Y = LX + E$, in which a known design matrix $L \in \mathbb{R}^{M \times N}$, called lead field matrix, maps the electrical activity of the brain, $X$, to the sensor measurements, $Y$. The measurement matrix $Y \in \mathbb{R}^{M \times T}$ captures the activity of $M$ sensors at $T$ time instants, $y(t) \in \mathbb{R}^{M \times 1}$, $t = 1, \ldots, T$, while the source matrix, $X \in \mathbb{R}^{N \times T}$, consists of the unknown activity of $N$ sources at the same time instants, $x(t) \in \mathbb{R}^{N \times 1}$, $t = 1, \ldots, T$. Here we assume a zero-mean Gaussian prior with full covariance $\Gamma$ for the underlying source distribution, $x(t) \sim \mathcal{N}(0, \Gamma)$, $t = 1, \ldots, T$. Besides, by treating different time points as independent samples, we have $p(X|\Gamma) = \prod_{t=1}^{T} \mathcal{N}(0, \Gamma)$. Similarly, the matrix $E = [e(1), \ldots, e(T)] \in \mathbb{R}^{M \times T}$ represents $T$ time instances of zero-mean Gaussian noise with full covariance $\Lambda$, $e(t) \sim \mathcal{N}(0, \Lambda)$, $t = 1, \ldots, T$, which is assumed to be independent of the source activations. This leads to the following expression for the distribution of the measurements: $p(Y|X) = \prod_{t=1}^{T} \mathcal{N}(Lx(t), \Lambda)$.

The goal is to infer the underlying brain activity $X$ from the M/EEG measurement $Y$ given the lead field matrix $L$. Two main categories of algorithms for this problem include Maximum-a-Posteriori (MAP) estimation or Type-I learning [8–11] and maximum-posterior distribution learning) [12–16]. Here, we focus on Type-II learning, which assumes a family of prior distributions $p(X|\Theta)$ parameterized by a set of hyper-parameters $\Theta$. The parameters of the Type-II model are the unknown source and noise covariances, i.e., $\Theta = \{\Gamma, \Lambda\}$. These hyper-parameters can be learned from the data along with the model parameters using a hierarchical Bayesian approach through minimizing the negative log of the marginal likelihood $p(Y|\Gamma, \Lambda)$, which is given by [17]:

$$
\mathcal{L}^{\text{II}}(\Gamma, \Lambda) = -\log p(Y|\Gamma, \Lambda) = \log|\Lambda + L\Gamma L^\top| + \frac{1}{T} \sum_{t=1}^{T} y(t)^\top (\Lambda + L\Gamma L^\top)^{-1} y(t)
$$

where $|\cdot|$ denotes the determinant of a matrix. Given the final solution of the hyperparameters $\Theta^{\text{II}} = \{\Gamma^{\text{II}}, \Lambda^{\text{II}}\}$, the posterior source distribution is obtained by plugging these estimates into the following equation: $p(X|\Gamma, \Lambda) = \prod_{t=1}^{T} \mathcal{N}(\bar{x}(t), \Sigma_x)$, where

$$
\bar{x}(t) = \Gamma L^\top (\Sigma_y)^{-1} y(t), \quad \Sigma_x = \Gamma - \Gamma L^\top (\Sigma_y)^{-1} L \Gamma, \quad \Sigma_y = \Lambda + L\Gamma L^\top.
$$

3 Proposed Method: Full-structure Noise (FUN) Learning

We present two theorems that result in a novel full-structure noise learning algorithm for optimization of $\mathcal{L}^{\text{II}}$. The Type-II cost function in $\mathcal{L}^{\text{II}}$ is non-convex. Building on the majorization-minimization (MM) class of algorithms [18], we construct convex surrogate functions that majorizes $\mathcal{L}^{\text{II}}(\Gamma, \Lambda)$, which benefits from guaranteed convergence properties.

**Theorem 1.** Let $\Lambda^k$ and $\Sigma_y^k$ be fixed values obtained in the $(k)$-th iteration of the optimization algorithm minimizing $\mathcal{L}^{\text{II}}(\Gamma, \Lambda)$. Then, optimizing the non-convex Type-II loss in $\mathcal{L}^{\text{II}}(\Gamma, \Lambda)$, with respect to $\Gamma$ and $\Lambda$ is equivalent to optimizing the following convex functions, which majorize $\mathcal{L}^{\text{II}}$.

$$
\mathcal{L}^{\text{conv}}_{\text{source}}(\Gamma, \Lambda^k) = \text{tr}(\left(\Sigma_y^k\right)^{-1} \Gamma) + \text{tr}(M_N^k \Gamma^{-1}) \quad \text{and} \quad \mathcal{L}^{\text{conv}}_{\text{noise}}(\Gamma, \Lambda^k) = \text{tr}(\left(\Sigma_y^k\right)^{-1} \Lambda) + \text{tr}(M_N^k \Lambda^{-1})
$$

$$
C_N^k := \frac{1}{2} \sum_{t=1}^{T} (y(t) - L\bar{x}(t))(y(t) - L\bar{x}(t))^\top, \quad M_N^k := \frac{1}{2} \sum_{t=1}^{T} (y(t) - L\bar{x}(t))(y(t) - L\bar{x}(t))^\top.
$$

Note that $C_N^k$ is well-defined when the rank of the lead field matrix, $L$, is less than the number of sources, which accrues if the span of the lead field is projected to the lower dimension than the number of sensors. This condition is surprisingly relaxed when $\Gamma$ has a diagonal structure.
We applied the FUN learning approach on the synthetic datasets to recover the locations and time with latent variables for source and noise, where parameters of priors for source and noise terms with non-uniform noise levels across channels, while with respect to source reconstruction performance and noise covariance estimation accuracy. Here, we as a result of Theorem 2, the source and noise covariance matrices that optimize cost function Theorem 2. The cost functions $\mathcal{L}^{\text{conv}}_{\text{source}}(\Gamma, \Delta^k)$ and $\mathcal{L}^{\text{conv}}_{\text{noise}}(\Gamma^k, \Delta)$ are both strictly geodesically convex with respect to the P.D. manifold, and their optimal solution with respect to $\Gamma$ and $\Delta$, respectively, can be attained according to the following two update rules:

\[
\Gamma^{k+1} \leftarrow (C^k_S)^{1/2} \left( (C^k_S)^{-1/2} M^k_S (C^k_S)^{-1/2} \right)^{1/2} (C^k_S)^{1/2}, \tag{2}
\]

\[
\Lambda^{k+1} \leftarrow (C^k_N)^{1/2} \left( (C^k_N)^{-1/2} M^k_N (C^k_N)^{-1/2} \right)^{1/2} (C^k_N)^{1/2}. \tag{3}
\]

As a result of Theorem 2, the source and noise covariance matrices are expected to be correlated, e.g., wireless communication [25, 28], robust portfolio optimization in finance [29], graph learning [30], thermal field reconstruction [31–33], and brain functional imaging [34–38].

4 Performance Evaluation

We applied the FUN learning approach on the synthetic datasets to recover the locations and time courses of the active brain sources. In addition to our proposed approach, two other Type-II learning schemes, namely homoscedastic and heteroscedastic SBL [22, 23], were also included as benchmarks with respect to source reconstruction performance and noise covariance estimation accuracy. Here, we note that heteroscedasticity refers to the common phenomenon that measurements are contaminated with non-uniform noise levels across channels, while homoscedasticity only accounts for uniform noise levels. Source reconstruction performance was evaluated according to the earth mover’s distance (EMD) [24], and the error in the reconstruction of the source time courses between each simulated source and the best (in terms of absolute correlations) matching reconstructed source. To evaluate the accuracy of the noise covariance matrix estimation, the following two metrics were calculated: the Pearson correlation, $\Lambda^{\text{sim}}$, and the normalized mean squared error (NMSE) between the original and estimated noise covariance matrices. Each simulation was carried out 100 times using different instances of $X$ and $E$, and the mean and standard error of the mean (SEM) of each performance measure across repetitions was calculated. Figure 1 summarizes the performance result, showing that FUN learning consistently outperforms its homoscedastic and heteroscedastic counterparts according to all evaluation metrics in particular in low-SNR settings. Consequently, as the SNR decreases, the gap between FUN learning and the two other variants increases. The convergence behaviour of all three noise estimation approaches is also shown, which indicates that the FUN learning approach reaches lower negative log-likelihood values.

Broader Impact

Many problems in medical imaging involve hierarchical Bayesian regression models for observations with latent variables for source and noise, where parameters of priors for source and noise terms need to be estimated jointly from data. This paper describes a novel inference algorithm for such hierarchical Bayesian regression models with concurrent estimation of regression parameter distributions and Gaussian noise distributions with full covariance structure. Capitalizing upon Riemannian geometry of positive definite matrices, we derived an efficient inference algorithm with guaranteed convergence properties. The benefits of our proposed framework were evaluated within an extensive set of simulations in the context of electromagnetic brain source imaging inverse problem and showed significant improvement upon state-of-the-art techniques in the realistic scenario where the noise has full covariance structure. Although this paper only discusses applications in EEG/MEG brain source imaging, FUN learning may also prove useful in other domains in which model residuals are expected to be correlated, e.g., wireless communication [25, 28], robust portfolio optimization in finance [29], graph learning [30], thermal field reconstruction [31–33], and brain functional imaging [34–38].
Figure 1: Source reconstruction performance of the simulated data with five active sources in presence of full-structure noise at 0 (dB) SNR (upper panel), and for a wide range of SNRs (lower panel). Topographic maps depict the locations of the ground-truth active brain sources along with the source reconstruction results of three noise learning schemes. For each algorithm, the estimated noise covariance matrix is also plotted above the topographic maps. The source and noise reconstruction performance in terms of EMD, time course correlation (Corr), $\Lambda_{\text{sim}}$ and NMSE is summarized in the associated table next to the upper panel example as well as in the plots in the second row for statistical analysis with 100 times repetition. MATLAB code for producing the results in the simulation study is also released as an open source package in a publicly accessible Github repository.

References


[38] H. Wei, A. Jafarian, P. Zeidman, V. Litvak, A. Razi, D. Hu, K. J. Friston, Bayesian fusion and multimodal DCM for EEG and fMRI, NeuroImage 211 (2020) 116595.