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

Ali Hashemi[1,2,3], **Chang Cai**[4,5], **Klaus-Robert Müller**[1,6,7,8,*], **Srikantan S Nagarajan**[5], **and Stefan Haufe**[3,9,*]

[1]Machine Learning Group, Technische Universität Berlin, Germany.
[2]Institut für Mathematik, Technische Universität Berlin, Germany.
[3]Charité – Universitätsmedizin Berlin, Germany.
[4]National Engineering Research Center for E-Learning, Central China Normal University, China.
[5]Department of Radiology and Biomedical Imaging, University of California, San Francisco, USA.
[6]Berliner Zentrum für Maschinelles Lernen, Berlin, Germany.
[7]Department of Artificial Intelligence, Korea University, South Korea.
[8]Max Planck Institute for Informatics, Saarbrücken, Germany.
[9]Bernstein 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 the accuracy of brain source estimates is not only influenced by the noise level but also its correlation structure. Importantly, existing approaches have not addressed the estimation of full-structure noise covariance matrices. Using Riemannian geometry, we derive an efficient algorithm for updating source and 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 where the noise has full-structure 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

Medical Imaging Meets NeurIPS (Med-NeurIPS 2020), Vancouver, Canada.

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

The goal is to infer the underlying brain activity $\mathbf{X}$ from the M/EEG measurement $\mathbf{Y}$ given the lead field matrix $\mathbf{L}$. Two main categories of algorithms for this problem include Maximum-a-Posteriori (MAP) estimation or *Type-I learning* [8–11] and maximum-marginal-likelihood estimation (Type-II learning) [12–16]. Here, we focus on Type-II learning, which assumes a family of prior distributions $p(\mathbf{X}|\mathbf{\Theta})$ parameterized by a set of hyper-parameters $\mathbf{\Theta}$. The parameters of the Type-II model are the unknown source and noise covariances, i.e., $\mathbf{\Theta} = \{\mathbf{\Gamma}, \mathbf{\Lambda}\}$. These hyper-parameters can be learned from the data along with the model parameters using a hierarchical Bayesian approach [13] through minimizing the negative log of the marginal likelihood $p(\mathbf{Y}|\mathbf{\Gamma}, \mathbf{\Lambda})$, which is given by [17]:

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

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

$$\bar{\mathbf{x}}(t) = \mathbf{\Gamma}\mathbf{L}^\top(\mathbf{\Sigma}_{\mathbf{y}})^{-1}\mathbf{y}(t) \;, \qquad \mathbf{\Sigma}_{\mathbf{x}} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{L}^\top(\mathbf{\Sigma}_{\mathbf{y}})^{-1}\mathbf{L}\mathbf{\Gamma} \text{ , and} \qquad \mathbf{\Sigma}_{\mathbf{y}} = \mathbf{\Lambda} + \mathbf{L}\mathbf{\Gamma}\mathbf{L}^\top \;.$$

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 (1). The Type-II cost function in (1) is non-convex. Building on the *majorization-minimization* (MM) class of algorithms [18], we construct convex surrogate functions that *majorizes* $\mathcal{L}^{\mathrm{II}}(\mathbf{\Gamma}, \mathbf{\Lambda})$. We then show the minimization equivalence between the constructed majoring functions and (1), which benefits from guaranteed convergence properties.

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

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

$$\mathbf{C}_{\mathrm{S}}^k := \left(\mathbf{L}^\top\left(\mathbf{\Sigma}_{\mathbf{y}}^k\right)^{-1}\mathbf{L}\right)^{-1} \;, \qquad \mathbf{C}_{\mathrm{N}}^k := \left(\mathbf{\Sigma}_{\mathbf{y}}^k\right) \;,$$

$$\mathbf{M}_{\mathrm{S}}^k := \tfrac{1}{T}\textstyle\sum_{t=1}^{T} \bar{\mathbf{x}}^k(t)\bar{\mathbf{x}}^k(t)^\top \;, \qquad \mathbf{M}_{\mathrm{N}}^k := \tfrac{1}{T}\textstyle\sum_{t=1}^{T}(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))(\mathbf{y}(t) - \mathbf{L}\bar{\mathbf{x}}^k(t))^\top \;.$$

[1] Note that $\mathbf{C}_{\mathrm{S}}^k$ is well-defined when the rank of the lead filed matrix, $\mathbf{L}$, is less than the number of sources, which accrues if the span of the lead filed is projected to the lower dimension than the number of sensors. This condition is surprisingly relaxed when $\mathbf{\Gamma}$ has a diagonal structure.

For optimization of the cost functions $\mathcal{L}_{\text{source}}^{\text{conv}}(\mathbf{\Gamma}, \mathbf{\Lambda}^k)$ and $\mathcal{L}_{\text{noise}}^{\text{conv}}(\mathbf{\Gamma}^k, \mathbf{\Lambda})$ with respect to $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$, respectively, we consider solutions that lie in the Riemannian manifold of positive definite (P.D.) matrices. This consideration enables us to invoke efficient methods from Riemannian geometry [19–21], leading to the following theorem.

Theorem 2. *The cost functions $\mathcal{L}_{\text{source}}^{\text{conv}}(\mathbf{\Gamma}, \mathbf{\Lambda}^k)$ and $\mathcal{L}_{\text{noise}}^{\text{conv}}(\mathbf{\Gamma}^k, \mathbf{\Lambda})$ are both strictly geodesically convex with respect to the P.D. manifold, and their optimal solution with respect to $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$, respectively, can be attained according to the two following update rules:*

$$\mathbf{\Gamma}^{k+1} \leftarrow (\mathbf{C}_{\mathrm{S}}^k)^{\frac{1}{2}} \left((\mathbf{C}_{\mathrm{S}}^k)^{-1/2} \mathbf{M}_{\mathrm{S}}^k (\mathbf{C}_{\mathrm{S}}^k)^{-1/2} \right)^{\frac{1}{2}} (\mathbf{C}_{\mathrm{S}}^k)^{\frac{1}{2}} \;, \tag{2}$$

$$\mathbf{\Lambda}^{k+1} \leftarrow (\mathbf{C}_{\mathrm{N}}^k)^{\frac{1}{2}} \left((\mathbf{C}_{\mathrm{N}}^k)^{-1/2} \mathbf{M}_{\mathrm{N}}^k (\mathbf{C}_{\mathrm{N}}^k)^{-1/2} \right)^{\frac{1}{2}} (\mathbf{C}_{\mathrm{N}}^k)^{\frac{1}{2}} \;. \tag{3}$$

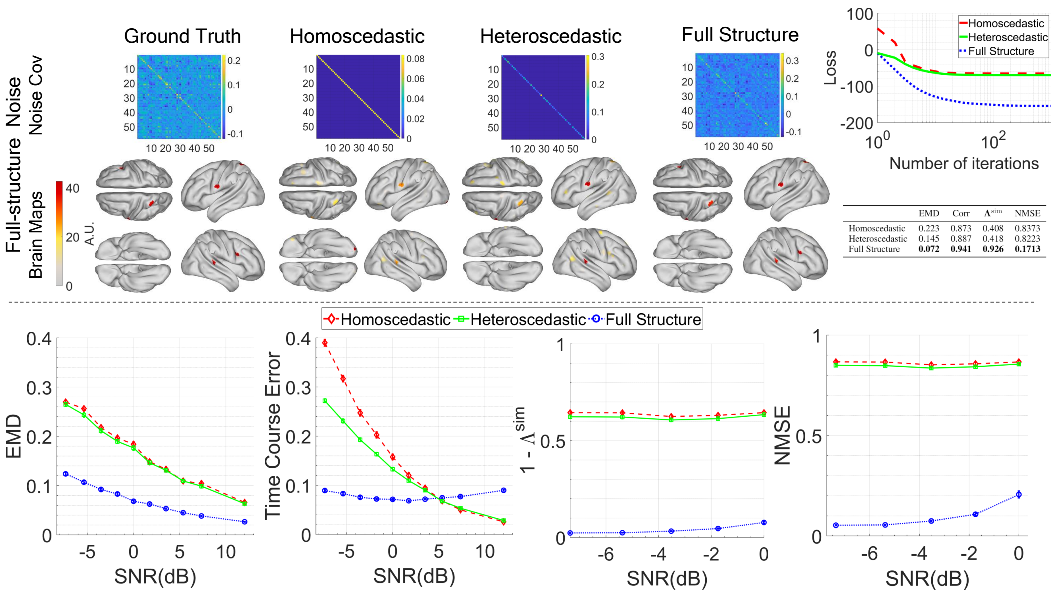
As a result of Theorem 2, the source and noise covariance matrices that optimize cost function $\mathcal{L}^{\text{II}}(\mathbf{\Gamma}, \mathbf{\Lambda})$ can be obtained by alternating between (2) and (3) until convergence. Note that, in brain source imaging, the assumption of full source covariance is often relaxed to independent sources through a diagonal covariance matrix $\mathbf{\Gamma} = \text{diag}(\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_N]^\top$. This simplification interestingly leads to sparsity of the resulting source distributions. With this assumption, Theorem 2 describes a novel algorithm for sparse Bayesian learning with full-structure noise covariance matrices.

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, $\mathbf{\Lambda}^{\text{sim}}$, 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 $\mathbf{X}$ and $\mathbf{E}$, 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].



	EMD	Corr	$\mathbf{\Lambda}^{\mathrm{sim}}$	NMSE
Homoscedastic	0.223	0.873	0.408	0.8373
Heteroscedastic	0.145	0.887	0.418	0.8223
Full Structure	**0.072**	**0.941**	**0.926**	**0.1713**

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), $\mathbf{\Lambda}^{\mathrm{sim}}$ 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

[1] L. Bungert, M. Burger, Y. Korolev, C.-B. Schoenlieb, Variational regularisation for inverse problems with imperfect forward operators and general noise models, arXiv preprint arXiv:2005.14131.

[2] Z. J. Daye, J. Chen, H. Li, High-dimensional heteroscedastic regression with an application to eQTL data analysis, Biometrics 68 (1) (2012) 316–326.

[3] A. Dalalyan, M. Hebiri, K. Meziani, J. Salmon, Learning heteroscedastic models by convex programming under group sparsity, in: International Conference on Machine Learning, 2013, pp. 379–387.

[4] H. B. Petersen, P. Jung, Robust instance-optimal recovery of sparse signals at unknown noise levels, arXiv preprint arXiv:2008.08385.

[5] M. Massias, O. Fercoq, A. Gramfort, J. Salmon, Generalized concomitant multi-task lasso for sparse multimodal regression, in: International Conference on Artificial Intelligence and Statistics, 2018, pp. 998–1007.

[6] Z. Zhang, B. D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE Journal of Selected Topics in Signal Processing 5 (5) (2011) 912–926.

[7] A. Hashemi, S. Haufe, Improving EEG source localization through spatio-temporal sparse Bayesian learning, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 1935–1939.

[8] R. D. Pascual-Marqui, C. M. Michel, D. Lehmann, Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain, International Journal of psychophysiology 18 (1) (1994) 49–65.

[9] S. Haufe, V. V. Nikulin, A. Ziehe, K.-R. Müller, G. Nolte, Combining sparsity and rotational invariance in EEG/MEG source reconstruction, NeuroImage 42 (2) (2008) 726–738.

[10] A. Gramfort, M. Kowalski, M. Hämäläinen, Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods, Physics in Medicine and Biology 57 (7) (2012) 1937.

[11] H. Janati, T. Bazeille, B. Thirion, M. Cuturi, A. Gramfort, Group level MEG/EEG source imaging via optimal transport: minimum Wasserstein estimates, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 743–754.

[12] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, Journal of Machine Learning Research 1 (Jun) (2001) 211–244.

[13] D. P. Wipf, B. D. Rao, Sparse Bayesian learning for basis selection, IEEE Transactions on Signal Processing 52 (8) (2004) 2153–2164.

[14] D. P. Wipf, B. D. Rao, An empirical Bayesian strategy for solving the simultaneous sparse approximation problem, IEEE Transactions on Signal Processing 55 (7) (2007) 3704–3716.

[15] D. Wipf, S. Nagarajan, Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions, IEEE Journal of Selected Topics in Signal Processing 4 (2) (2010) 317–329.

[16] W. Wu, S. Nagarajan, Z. Chen, Bayesian machine learning: EEG\MEG signal processing measurements, IEEE Signal Processing Magazine 33 (1) (2016) 14–36.

[17] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, S. S. Nagarajan, Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG, NeuroImage 49 (1) (2010) 641–655.

[18] Y. Sun, P. Babu, D. P. Palomar, Majorization-minimization algorithms in signal processing, communications, and machine learning, IEEE Transactions on Signal Processing 65 (3) (2017) 794–816.

[19] P. Petersen, S. Axler, K. Ribet, Riemannian geometry, Vol. 171, Springer, 2006.

[20] M. Berger, A panoramic view of Riemannian geometry, Springer Science & Business Media, 2012.

[21] J. Jost, J. Jost, Riemannian geometry and geometric analysis, Vol. 42005, Springer, 2008.

[22] A. Hashemi, C. Cai, G. Kutyniok, K.-R. Müller, S. Nagarajan, S. Haufe, Unification of sparse Bayesian learning algorithms for electromagnetic brain imaging with the majorization minimization framework, bioRxiv.

[23] 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) 117411.

[24] Y. Rubner, C. Tomasi, L. J. Guibas, The earth mover's distance as a metric for image retrieval, International Journal of Computer Vision 40 (2) (2000) 99–121.

[25] R. Prasad, C. R. Murthy, B. D. Rao, Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach, IEEE Transactions on Signal Processing 63 (20) (2015) 5369–5382.

[26] P. Gerstoft, C. F. Mecklenbräuker, A. Xenaki, S. Nannuru, Multisnapshot sparse Bayesian learning for DOA, IEEE Signal Processing Letters 23 (10) (2016) 1469–1473.

[27] S. Haghighatshoar, G. Caire, Massive MIMO channel subspace estimation from low-dimensional projections, IEEE Transactions on Signal Processing 65 (2) (2017) 303–318.

[28] M. B. Khalilsarai, T. Yang, S. Haghighatshoar, G. Caire, Structured channel covariance estimation from limited samples in Massive MIMO, in: ICC 2020-2020 IEEE International Conference on Communications (ICC), IEEE, 2020, pp. 1–7.

[29] Y. Feng, D. P. Palomar, et al., A signal processing perspective on financial engineering, Foundations and Trends® in Signal Processing 9 (1–2) (2016) 1–231.

[30] S. Kumar, J. Ying, J. V. de Miranda Cardoso, D. P. Palomar, A unified framework for structured graph learning via spectral constraints., Journal of Machine Learning Research 21 (22) (2020) 1–60.

[31] A. Hashemi, M. Rostami, N.-M. Cheung, Efficient environmental temperature monitoring using compressed sensing, in: 2016 Data Compression Conference (DCC), IEEE, 2016, pp. 602–602.

[32] A. Flinth, A. Hashemi, Thermal source localization through infinite-dimensional compressed sensing, arXiv preprint arXiv:1710.02016.

[33] A. Flinth, A. Hashemi, Approximate recovery of initial point-like and instantaneous sources from coarsely sampled thermal fields via infinite-dimensional compressed sensing, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 1720–1724.

[34] M. Cai, N. W. Schuck, J. W. Pillow, Y. Niv, A Bayesian method for reducing bias in neural representational similarity analysis, in: Advances in Neural Information Processing Systems, 2016, pp. 4951–4959.

[35] M. Shvartsman, N. Sundaram, M. Aoi, A. Charles, T. Willke, J. Cohen, Matrix-normal models for fMRI analysis, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 1914–1923.

[36] M. B. Cai, N. W. Schuck, J. W. Pillow, Y. Niv, Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias, PLoS computational biology 15 (5) (2019) e1006299.

[37] M. B. Cai, M. Shvartsman, A. Wu, H. Zhang, X. Zhu, Incorporating structured assumptions with probabilistic graphical models in fMRI data analysis, Neuropsychologia (2020) 107500.

[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.