Sparse wavelet-based solutions for the M/EEG inverse problem

Samy Mokhtari*, Jean-Michel Badier†, Christian Bénar†, Bruno Torrésani*
* *Aix Marseille Univ, CNRS, I2M, Marseille, France.*
† *Aix Marseille Univ, CNRS, INS, Inst Neurosci Syst, Marseille, France.*
{samy.mokhtari, jean-michel.badier, christian.benar, bruno.torresani}@univ-amu.fr

***Abstract*—This paper is concerned with variational and Bayesian approaches to neuro-electromagnetic inverse problems (EEG and MEG). The strong indeterminacy of these problems is tackled by introducing sparsity inducing regularization/priors in a transformed domain, namely a spatial wavelet domain. Sparsity in the wavelet domain allows to reach "data compression" in the cortical sources domain. Spatial wavelets defined on the mesh graph of the triangulated cortical surface are used, in combination with sparse regression techniques, namely LASSO regression or sparse Bayesian learning, to provide localized and compressed estimates for brain activity from sensor data. Numerical results on simulated and real MEG data are provided, which outline the performances of the proposed approach in terms of localization.**

***Index Terms*—MEG inverse problem, compression, spatial wavelets, non-smooth optimization, Sparse Bayesian Learning**

I. Introduction

MEG and EEG inverse problems [1] are extremely under-determined, the number of measurements being much smaller than the number of unknowns (by a factor ~ 50). In such a case, the obtained solution dramatically depends on the choice of regularization and/or prior information. Reference solutions, routinely used by practitioners, often suffer from known biases. For example, the Minimum-Norm Estimate (MNE) framework tends to overestimate and oversmooth the support of the cortical activity, while other sparsity-enforcing approaches (e.g. Minimum-Current Estimate) can yield very localized solutions, which may be too localized in many situations. Last but not least, the absence of ground truth makes objective comparison of different methods very difficult.

The goal of this paper is to investigate solutions exploiting sparsity in a spatial wavelet domain, with the objective of describing spatially extended (i.e. patch-like) solutions without prior assumption on the patch size. Limiting to a particular instance of spatial wavelet transform, i.e. spectral graph wavelets defined on the triangulated cortical surface, we explore a wavelet-based MNE solver paired with a Sparse Bayesian Learning approach, and analyze its performances in comparison with some state of the art methods : MNE, MCE, and sparse total variation penalization.

II. Problem statement and state of the art

Given a MEG (magnetoencephalography) dataset consisting in J_0 sensors measurements during L time steps, and summarized in the form of a matrix $\mathbf{Z} \in \mathbb{R}^{J_0 \times L}$, MEG inverse problem consists in estimating the underlying cortical current densities $\mathbf{S} \in \mathbb{R}^{N \times L}$ explaining (in an "optimal" way) the data. Typical orders of magnitude for the dimensions at stake are $J_0 \approx 250$ and $N \approx 15,000$. With a suitable whitening matrix $\boldsymbol{\Upsilon} \in \mathbb{R}^{J_0 \times J}$ and dimension reduction onto $J < J_0$ spatial channels, the measurements and cortical sources are connected by the following observation equation

$$\mathbf{Z} = \mathbf{G}\mathbf{S} + \mathbf{B} \tag{1}$$

with $\mathbf{Z} = \boldsymbol{\Upsilon}\mathbf{Z_0}$ the whitened data, $\mathbf{G} = \boldsymbol{\Upsilon}\mathbf{G_0}$ the whitened leadfield matrix, and $\mathbf{B}$ the matrix of observation noise and background activity, assumed i.i.d. with zero mean and unit variance (as a result of whitening). The leadfield matrix $\mathbf{G}$ summarizes the propagation of the electromagnetic field from cortex to sensors.

We here focus on variational and Bayesian approaches for distributed source models with constrained orientation (see [2]). Reference approaches also include dipole fitting [3], [4] and various forms of beamformers [5], [6]. Variational methods generally lead to optimization problems involving a quadratic data fidelity term and a penalty term $f(\mathbf{S})$:

$$\mathbf{S}_* = \underset{\mathbf{S}\in\mathbb{R}^{N\times L}}{\arg\min} \left[\frac{1}{2} \|\mathbf{Z} - \mathbf{G}\mathbf{S}\|_F^2 + f(\mathbf{S}) \right] , \tag{2}$$

where $\|\cdot\|_F$ stands for the Frobenius norm.

Quadratic regularizations $f(\mathbf{S}) = \|\mathbf{D}\mathbf{S}\|_F^2$ yield closed-form solutions ($\mathbf{D} = \sqrt{\lambda}\,\mathsf{I}$ for MNE [7] and $\mathbf{D}$ diagonal for weighted MNE [8]). These methods turn out to be strongly affected by *source leakage* (oversmoothing and overestimation of spatial extent) and *depth bias* (under-estimation of deep sources).

The strong under-determination of the optimization problem (2) has pushed for the introduction of sparsity inducing regularization/priors, which are implemented in variational formulations through a non-smooth, generally convex regularization $f(\mathbf{S})$. Successful examples include the minimum current estimate (MCE [9], $f(\mathbf{S}) = \lambda\|\mathbf{S}\|_1$), mixed norm estimate that promote sparsity in space and persistence in time (MxNE [10], [11], $f(\mathbf{S}) = \lambda\|\mathbf{S}\|_{21}$), total variation (VB-SCCD [12], $f(\mathbf{S}) = \lambda\|\nabla\mathbf{S}\|_1$), sparse total variation (sVB-SCCD [13], $f(\mathbf{S}) = \lambda\|\nabla\mathbf{S}\|_1 + \mu\|\mathbf{S}\|_1$) and similar approaches involving time persistence. The problem is then solved through dedicated numerical algorithms (see [13] for a review).

These methods involve regularization parameters, whose tuning and interpretation may rely on a Maximum A Posteriori (MAP) estimation in a Bayesian point of view. Quite recently, empirical Bayes methods such as Sparse Bayesian Learning (SBL, see [14] and [15] for a review) have received increasing interest in the M/EEG inverse problem literature for their ability to enforce sparse solutions. SBL involves joint optimization of the objective function in (2) in the case $f(\mathbf{S}) = \|\mathbf{\Gamma}_S^{-1/2}\mathbf{S}\|_F^2$, with $\mathbf{\Gamma_S} = \mathsf{diag}(\underline{\gamma_S})$ and $\underline{\gamma_S} \in \mathbb{R}^N$, and treats $\mathbf{S}$ as a latent variable. Marginalizing with respect to $\mathbf{S}$ results in the non-convex optimization problem

$$\underline{\gamma_S}^* = \underset{\underline{\gamma_S}\in(\mathbb{R}_+)^N}{\arg\min} \left[\mathsf{Tr}\left(\mathbf{C_Z}\mathbf{\Sigma_Z}(\underline{\gamma_S})^{-1}\right) + \ln\det(\mathbf{\Sigma_Z}(\underline{\gamma_S}))\right] \ , \quad (3)$$

where $\mathbf{\Sigma_Z}(\underline{\gamma_S}) = \mathsf{I_J} + \mathbf{G}\mathbf{\Gamma}_S\mathbf{G}^T$ is the posterior data covariance matrix, and $\mathbf{C_Z} = \frac{1}{L}\mathbf{Z}\mathbf{Z}^T$ the sample covariance matrix. Given $\underline{\gamma_S}^*$, the corresponding weighted MNE solution reads :

$$\mathbf{S}_* = \mathbf{\Gamma}_S^*\mathbf{G}^T\left[\mathbf{\Sigma}_Z(\underline{\gamma_S}^*)\right]^{-1}\mathbf{Z} \quad (4)$$

The so-obtained solutions can be shown to be sparse, with support size not exceeding the rank of $\mathbf{G}$.

III. Spatial wavelet based approaches

Sparse solutions such as MCE or SBL have been proven efficient for estimating very focal brain activity, but tend to underestimate the spatial support in the case of more extended activity. Enforcing sparsity in a transformed domain has been shown to improve the situation (see e.g. [13], where spatial gradient is used as transformation). We focus here on spatial wavelet transforms, which provide multiscale representations for functions on the cortical surface or graph. Several constructions of such wavelets have been proposed in literature. In the current work, we focus on the Spectral Graph Wavelets (SGW) construction of [16] which, unlike several other constructions, provides a redundant multiscale representation.

A. Laplacian and gradient on the cortical surface, discrete total variation and spectral graph wavelets

Available data for the M/EEG inverse problem generally involve a discretization of the cortical surface, in the form of a triangular mesh, which may be described as a weighted graph $(\mathcal{E}, \mathcal{V}, w)$, with edge set $\mathcal{E} = \{e_1, \dots e_E\}$, vertex set $\mathcal{V} = \{v_1, \dots v_N\}$, and a weight function $w : \mathcal{E} \longmapsto \mathbb{R}_+$. This defines a symmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, whose nonzero elements are associated with neighboring vertices, and are defined by $\mathbf{A}_{kk'} = w(e_{k,k'})$, $e_{k,k'}$ being the edge connecting vertices k and k'. A gradient matrix $\nabla \in \mathbb{R}^{E\times N}$ may also be defined on the graph as follows: for any edge $e = e_{k,k'}$, set $\nabla_{ek} = \mathbf{A}_{kk'}$ and $\nabla_{ek'} = -\mathbf{A}_{kk'}$. The corresponding TV norm of $\mathbf{S} \in \mathbb{R}^N$ is then defined as $\|\nabla\mathbf{S}\|_1$. This notion is used in the VB-SCCD and sVB-SCCD solvers.

The un-normalized graph Laplacian is then defined as the $N \times N$ matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D} = \mathsf{diag}\{d_1, \dots d_N\}$ is the degree matrix: $d_k = \sum_{k'} \mathbf{A}_{kk'}$, the sum running over neighboring vertices. $\mathbf{L}$ is symmetric positive semi-definite. Its eigenvectors $\{\chi_n\}_{n=1}^N$ may thus be interpreted as Fourier modes for functions defined on the graph, and the corresponding eigenvalues $0 = \lambda_1 \leq \cdots \leq \lambda_N$ as (spatial) frequencies. Denoting by $f \mapsto \widehat{f}$ the corresponding Graph Fourier Transform, a scaling operator T_h and a wavelet operator T_g^s on $\mathbb{R}^N$ can be defined [16] as follows: for all $f \in \mathbb{R}^N$ and all Fourier modes $l \in \{1, \dots, N\}$

$$\widehat{T_h[f]}(l) = h(\lambda_l)\langle\chi_l, f\rangle \ , \quad \widehat{T_g^s[f]}(l) = g(s\lambda_l)\langle\chi_l, f\rangle \ , \quad (5)$$

where h and g denote respectively a low-pass and band-pass kernel functions defined on the continuous extension $\mathbb{R}_+$ of the spectrum of $\mathbf{L}$, and $s \in \mathbb{R}_+^*$ refers to a dilation parameter. Scaling functions ϕ_n and scaled wavelets $\psi_{s,n}$ can be defined as $\phi_n = T_h[\delta_n]$ and $\psi_{s,n} = T_g^s[\delta_n]$ respectively, δ_n being the Dirac mass at vertex n. For suitably chosen g, h and scales $\{s_1, \dots s_{N_s}\}$, these functions form a frame of $\mathbb{R}^N$, of cardinality $N_W = N(N_s+1)$. The corresponding matrix reads

$$\mathbf{W} = \begin{pmatrix}\phi_1 & \dots & \phi_N & \psi_{s_1,1} & \dots & \psi_{s_{N_s},N}\end{pmatrix}^T \ . \quad (6)$$

This matrix $\mathbf{W}$ allows defining the wavelet analysis operator $\mathcal{W}$, whose adjoint $\mathcal{W}^*$ is the synthesis operator with matrix $\mathbf{W}^T$. In synthesis based approaches, a function f on the graph is expressed as $f = \mathbf{W}^T\mathbf{X}$.

It is here worth highlighting that the cutoff frequency of the low-pass kernel h, the number of detail coefficients N_s, and the wavelets quality factor are hyper-parameters of the wavelet-based method hereafter described.

B. Synthesis-based variational formulation

Given this spectral graph wavelet construction, we introduce sgw-MNE, the analogue of the MNE problem (2). From a synthesis point of view, writing $\mathbf{S} = \mathbf{W}^T\mathbf{X}$ and $\mathbf{G_W} = \mathbf{G}\mathbf{W}^T$:

$$\mathbf{X}_* = \underset{\mathbf{X}\in\mathbb{R}^{N_W\times L}}{\arg\min}\left[\frac{1}{2}\|\mathbf{Z} - \mathbf{G_W}\mathbf{X}\|_F^2 + \lambda\|\mathbf{X}\|_F^2\right] \ , \quad (7)$$

The problem has a closed-form solution $\mathbf{X}_*$, and the corresponding source estimate then reads $\mathbf{S}_* = \mathbf{W}^T\mathbf{X}_*$.

As for the above hyper-parameter λ, it can be estimated, thanks to the Bayesian point of view, from the signal to noise ratio (SNR). Indeed, defining the signal and noise levels in the sensor domains by the traces of the corresponding covariance matrices $\mathbf{\Sigma_Z} = \mathbf{\Sigma_B} + \mathbf{G}\mathbf{\Gamma_S}\mathbf{G}^T$ and $\mathbf{\Sigma_B}$, the corresponding SNR ρ (ratio of signal and noise standard deviations) satisfies $\rho^2 = 1 + \|\mathbf{G}\|_F^2/\lambda\mathsf{Tr}(\mathbf{\Sigma_B})$. For the sake of completeness, a sgw-MCE solution can also be obtained by replacing the quadratic penalization in (7) with a term $\lambda\|\mathbf{X}\|_1$, and a similar heuristics can be derived to connect λ to ρ in the MCE case.

C. Sparse Bayesian Learning formulation

To make up for the excessive sparsity enforced by a straightforward SBL approach, we designed a SBL algorithm that enforces sparsity in the graph wavelet domain instead of the spatial domain. The implementation of the SBL approach within the wavelet-based variational formulation (7) is straightforward, as it follows an optimization problem similar to (3), with now $\mathbf{\Sigma_Z}(\underline{\gamma_X}) = \mathsf{I_J} + \mathbf{G_W}\mathbf{\Gamma}_X\mathbf{G_W}^T$.

Several algorithms are described in [15] for minimizing such objective functions, such as Expectation-Maximization and convex-bounding Champagne and variants. They all require an initial value for $\underline{\gamma_X}$, which we obtain from a preliminary sgw-MNE computation followed by the evaluation of the estimated source variances. We term this approach sgw-SBL.

IV. Results

The proposed approach is first illustrated on simulated data, numerically generated on a brain geometry based on a MEG dataset acquired by the DYNAMAP group at INS Marseille in an auditory evoked potentials protocol. Available data include sensor data (245 MEG sensors, 180 trials, sampled at 2034.51 Hz). The dataset also involves a triangulated cortical surface (interface of white and gray matters), with 14995 vertices. 3D coordinates are available, as well as the adjacency matrix of the mesh graph, which permits the computation of graph Gradient and Laplacian. A leadfield matrix $\mathbf{G}$ was computed using the OpenMEEG software. A baseline covariance matrix $\mathbf{\Sigma}_B$ was estimated from pre-stimulus data, and used for data pre-whitening, and projection onto the subspace spanned by eigenvectors with significant eigenvalues.

The simulation protocol is as follows : $N_p = 100$ connected patches on the cortical surface were generated with random locations across the brain, with two different sizes (10 and 100 vertices per patch). For each patch, simulated sources $\mathbf{S}_{\mathsf{sim}}$ were set to a constant value β inside the patch and zero outside. Corresponding sensor data were computed as $\mathbf{Z}_{\mathsf{sim}} = \mathbf{G}\mathbf{S}_{\mathsf{sim}} + B$, with B a realization of sensor baseline data, simulated as iid samples from a multivariate normal distribution $\mathcal{N}(0, \mathbf{\Sigma_B})$. The constant β was tuned to match a given PSNR (ratio of peak signal value to noise standard deviation), in the sensor domain. After whitening and dimension reduction, source estimates were computed with several approaches : MNE, sgw-SBL, MCE and sVB-SCCD.

A. Wavelet design

In this paper, we use a mesh graph with binary weights, and limit ourselves to the SGW design described in section 8.1 of [16], with a bump shaped, piecewise polynomial, band-pass kernel g, and a low-pass kernel h with very fast decay. The cutoff frequency of the low-pass kernel is set to $\lambda_{min} = \lambda_{max}/K$, with λ_{max} the largest eigenvalue of the Laplacian $\mathbf{L}$, and $K > 1$ a user-defined constant. The quality factor (center frequency divided by bandwidth) of the bandpass kernel approximately equals 1.38. The wavelets frame bounds are approximately equal to 0.71 and 1.41.

B. Evaluation metrics

Objective evaluation of MEG source reconstruction methods is a difficult problem. For real data, the absence of ground truth imposes relying on expert knowledge. No real consensus exists either for result evaluation on simulated data. Indeed, many metrics have been introduced in literature, most of which having a tendency to favor specific approaches or cases.

If one is mostly interested in finding the (approximate) spatial support of the brain activity, the obtained solution has to be thresholded, and the result depends on the thresholding strategy. Methods can be assessed using standard statistical tools (ROC or PR curves, AUC,...), which are not necessarily appropriate in this context, or custom metrics. Since we are primarily interested here in quantitative source estimates, and not only support, we will not follow this avenue here.

In the M/EEG inverse problem literature (see for instance [17], [18]), dedicated metrics such as the Dipole Localization Error (DLE), the Spatial Dispersion (SD) and the Resolution Index (RI) are often put forward. These are generally defined for linear source estimation methods where a closed-form solution exists. In [19], slightly different definitions were introduced for the DLE and SD metrics in order to deal with non-linear source estimation methods, but there does not seem to exist a real consensus in literature. In addition, it has also been stressed out that such metrics, while well-suited for evaluating the localization of a single or a few separable sources, become less relevant for extended sources and complex patterns, especially with non-linear methods.

In this paper, due to limited space, we decided to restrict ourselves to the following metrics:

- a Spatial Dispersion at "peak-time" defined as follows :

$$SD_{i_{max}}(t_{max}) = \sqrt{\frac{\sum_{k=1}^{N} d_{i_{max},k}^2 |\mathbf{S}_{k,t_{max}}|^2}{\sum_{k=1}^{N} |\mathbf{S}_{k,t_{max}}|^2}} \ , \quad (8)$$

  where $t_{\max}$ denotes the time achieving maximal value in sensor domain, $i_{\max}$ the vertex with maximal value at $t_{\max}$, and $d_{i,k}$ the Euclidean distance between vertices.
- the first Wasserstein distance [20] between suitably normalized "energy maps", which are defined by considering an average over a time index range of interest $[l_0, l_1]$:

$$S_{amp,k} \propto \sqrt{\frac{1}{l_1 - l_0 + 1} \sum_{l=l_0}^{l_1} \mathbf{S}_{k,l}^2} \ . \quad (9)$$

In addition, to assess the ability to recover a correct order of magnitude, we also record the ℓ^2 norms of amplitudes S_{amp} for each solver, and compare these values with the reference.

C. Simulation results

We collect in this section a sample of results comparing the various methods. We provide 3 tables including summary statistics (mean and median, and standard deviation and interquartile distance) of evaluation metrics described above. For these tables, sgw-SBL results were obtained with a convex-bounding Champagne algorithm. Table I displays the obtained values for the SD metric (normalized by the reference spatial dispersion) for all four methods under consideration, and for the two simulated patch sizes. Similarly, Table II gives values for the Wasserstein distance between the simulated and estimated normalized energy maps. Looking at these two first evaluation metrics, it can be seen that for the large support simulation, sgw-SBL outperforms consistently

and significantly the other approaches, while sVB-SCCD also yields acceptable results. For smaller support size simulations, MCE yields better results with respect to the Wasserstein distance, but sgw-SBL is not far behind and even produces good results with respect to the SD metric. This is not really surprizing, since MCE promotes sparsity in the spatial domain, and can therefore be expected to perform well for focal sources. To better challenge MCE in this case, the wavelets hyper-parameters (low-pass cutoff frequency, number of detail coefficients) may be tuned to favor highly localized sources.

The results displayed in TABLE III also advocate, in the large support case, for sVB-SCCD and sgw-SBL when it comes to the ability to recover a correct order of magnitude for the l^2 norm of the sources amplitudes S_{amp}. Interestingly, sgw-SBL is able to avoid any over-estimation of the l^2 norm.

TABLE I
SUMMARY STATISTICS FOR THE NORMALIZED SPATIAL DISPERSION : SMALL V./ LARGE SUPPORT

$\frac{SD_{est}}{SD_{ref}}$	MNE	sgw-SBL	MCE	sVB-SCCD
mean	6.27 / 2.40	**2.67** / 0.72	2.79 / **1.12**	2.82 / 1.20
median	5.88 / 2.34	2.36 / 0.64	**2.09** / 0.97	2.30 / 0.94
std	2.22 / 0.43	**1.23** / **0.29**	2.22 / 0.73	1.98 / 0.82
IQD	2.44 / 0.5	1.31 / 0.5	2.87 / 0.55	1.4 / 0.93

TABLE II
SUMMARY STATISTICS FOR THE WASSERSTEIN METRIC : SMALL V./ LARGE SUPPORT

W_1	MNE	sgw-SBL	MCE	sVB-SCCD
mean	6.96 / 1.34	6.47 / **0.88**	**5.54** / 1.87	6.65 / 1.11
median	6.04 / 1.05	5.55 / **0.63**	**4.08** / 1.56	5.62 / 0.79
std	5.19 / 0.96	5.08 / 0.74	4.92 / 1.65	5.25 / 0.95
IQD	4.58 / 0.82	4.67 / 0.64	4.19 / 1.25	4.82 / 0.78

TABLE III
SUMMARY STATISTICS FOR THE RATIO OF L^2-NORMS : SMALL V./ LARGE SUPPORT

$\frac{\lVert S_{amp}^{est} \rVert_2}{\lVert S_{amp}^{ref} \rVert_2}$	MNE	sgw-SBL	MCE	sVB-SCCD
mean	0.04 / 0.13	0.14 / 0.54	**1.12** / 3.25	0.40 / **1.09**
median	0.04 / 0.13	0.14 / 0.55	**1.14** / 3.28	0.37 / **1.08**
std	0.01 / 0.02	0.02 / 0.08	0.45 / 0.97	0.19 / 0.43
IQD	0.01 / 0.02	0.02 / 0.1	0.6 / 1.27	0.2 / 0.45

To add some experimental basis to this preliminary numerical validation, we also display the solution obtained with a sgw-SBL solver based on an Expectation-Maximization algorithm, with the real data acquired during the auditory evoked potentials protocol. Despite the lack of ground truth solution in this case, Fig. 1 tends to confirm the performance of sgw-SBL, as the localization of the estimated sources is close to the auditory cortex in the right hemisphere (see [21]). Furthermore, the order of magnitude of the sources amplitudes is much higher with sgw-SBL compared with what can be obtained with MNE.

V. CONCLUSIONS

We described in this paper sgw-SBL, a spatial wavelet-based method to the MEG inverse problem, where sparsity is enforced in the spectral graph wavelet domain via a sparse Bayesian learning approach. The goal is to estimate spatially extended brain activity, with quantitatively relevant amplitude.

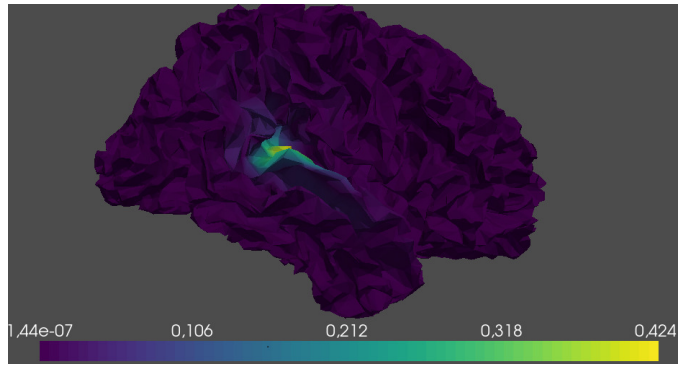

Fig. 1. Real data (auditory evoked potentials): estimated sources with sgw-SBL (EM solver)

Numerical results on simulated data show that sgw-SBL indeed achieves the best results when the spatial extent of the brain activity is large enough, and even comes as a close second behind MCE for smaller support activity, which are generally better accounted for by approaches favoring more classical spatial sparsity. Such results, well known in image processing, seem to remain true in the non-standard setting considered here (spectral graph wavelets).

Due to space limitation, we restricted the discussion to a limited set of methods. More thorough results investigating a larger family of objective functions (including in particular l^1 penalized least square in wavelet domain and analysis-based approaches), a more sophisticated modeling of time dependence, and the influence of hyper-parameters (SGW parameters, regularization parameters...) and SBL algorithms will be presented in a forthcoming publication.

Further work will also include different wavelet constructions, including SGW with different graph choice (for example taking into account physical edge lengths), or wavelet bases, such as intertwining wavelets on graphs [22], lifting based constructions [23] or diffusion wavelets [24] (we recall that SGW is redundant, which induces a significant increase of the problem dimensionality). The integration of spatial wavelets into different Bayesian inference schemes such as Bernoulli-Gauss models [25], Maximum Entropy on the Mean approach (MEM, see [26], [27]) or Variational Bayes [28] would also be worth investigating.

ACKNOWLEDGEMENTS

This work has benefited from support from the French government through the BMWs project (ANR-20-CE45-0018), and the *Institut Convergence* ILCB (ANR-16-CONV-0002).

We also wish to thank M. Kowalski, G. Mebarki and C. Mélot for fruitful discussions.

REFERENCES

[1] S. Baillet, "Magnetoencephalography for brain electrophysiology and imaging," Nature Neuroscience, vol. 20, no. 3, pp. 327–339, 03 2017.

[2] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," IEEE Signal Processing Magazine, vol. 18, no. 6, pp. 14–30, 2001.

[3] J. Mosher, P. Lewis, and R. Leahy, "Multiple dipole modeling and localization from spatio-temporal MEG data," IEEE Transactions on Biomedical Engineering, vol. 39, no. 6, pp. 541–557, June 1992.

[4] K. Jerbi, S. Baillet, J. Mosher, G. Nolte, L. Garnero, and R. Leahy, "Localization of realistic cortical activity in MEG using current multipoles," NeuroImage, vol. 22, no. 2, pp. 779 – 793, 2004.

[5] B. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, "Localization of brain electrical activity via linearly constrained minimum variance spatial filtering," IEEE Transactions on Biomedical Engineering, vol. 44, no. 9, pp. 867–880, 1997.

[6] K. Sekihara and S. S. Nagarajan, Adaptive Spatial Filters for Electromagnetic Brain Imaging, ser. Series in Biomedical Engineering, J. H. Nagel, Ed. Springer-Verlag Berlin Heidelberg, 2008.

[7] M. S. Hämäläinen and R. J. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum norm estimates," Medical and Biological Engineering & Computing, vol. 32, no. 1, pp. 35–42, 1994.

[8] M. Fuchs, M. Wagner, T. Köhler, and H.-A. Wischmann, "Linear and nonlinear current density reconstructions," Journal of clinical Neurophysiology, vol. 16, no. 3, pp. 267–295, 1999.

[9] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," IEEE Transactions on Biomedical Engineering, vol. 42, no. 6, pp. 608–615, June 1995.

[10] W. Ou, M. S. Hämäläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," NeuroImage, vol. 44, no. 3, pp. 932 – 946, 2009.

[11] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods," Physics in Medicine and Biology, vol. 57, no. 7, pp. 1937–1961, mar 2012.

[12] L. Ding, "Reconstructing cortical current density by exploring sparseness in the transform domain," Physics in Medicine and Biology, vol. 54, no. 9, pp. 2683–2697, apr 2009.

[13] H. Becker, L. Albera, P. Comon, R. Gribonval, and I. Merlet, "Fast, variation-based methods for the analysis of extended brain sources," in 2014 22nd European Signal Processing Conference (EUSIPCO), 2014, pp. 41–45.

[14] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, "Robust bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," NeuroImage, vol. 49, no. 1, pp. 641–655, 2010.

[15] 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, vol. 239, p. 118309, 2021.

[16] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," Applied and Computational Harmonic Analysis, vol. 30, no. 2, pp. 129 – 150, 2011.

[17] A. Molins, S. Stufflebeam, E. Brown, and M. Hämäläinen, "Quantification of the benefit from integrating MEG and EEG data in minimum ℓ^2-norm estimation," NeuroImage, vol. 42, no. 3, pp. 1069 – 1077, 2008.

[18] O. Hauk, M. Stenroos, and M. S. Treder, "Towards an objective evaluation of eeg/meg source estimation methods – the linear approach," NeuroImage, vol. 255, p. 119177, 2022.

[19] J. G. Samuelsson, N. Peled, F. Mamashli, J. Ahveninen, and M. S. Hämäläinen, "Spatial fidelity of MEG/EEG source estimates: A general evaluation approach," NeuroImage, vol. 224, p. 117430, 2021.

[20] G. Peyré and M. Cuturi, "Computational optimal transport: With applications to data science," Foundations and Trends® in Machine Learning, vol. 11, no. 5-6, pp. 355–607, 2019.

[21] Wikipedia contributors, "Auditory cortex — Wikipedia, the free encyclopedia," 2023, [Online; accessed 12-June-2023].

[22] L. Avena, F. Castell, A. Gaudillière, and C. Mélot, "Intertwining wavelets or multiresolution analysis on graphs through random forests," Applied and Computational Harmonic Analysis, 2018.

[23] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," SIAM Journal on Mathematical Analysis, vol. 29, no. 2, p. 511–546, 1998.

[24] R. R. Coifman and M. Maggioni, "Diffusion wavelets," Applied and Computational Harmonic Analysis, vol. 21, no. 1, pp. 53–94, 2006, special Issue: Diffusion Maps and Wavelets.

[25] P. Barbault, M. Kowalski, and C. Soussen, "Parameter estimation in sparse inverse problems using bernoulli-gaussian prior," in ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2022, pp. 5413–5417.

[26] J.-M. Lina, R. A. Chowdhury, E. Lemay, E. Kobayashi, and C. Grova, "Wavelet-based localization of oscillatory sources from magnetoencephalography data," IEEE Transactions on Biomedical Engineering, vol. 61, no. 8, pp. 2350–2364, 2014.

[27] M.-C. Roubaud, J.-M. Lina, J. Carrier, and B. Torrésani, "Space-Time Extension of the MEM Approach for Electromagnetic Neuroimaging," in IEEE Int. Workshop on Machine Learning for Signal Processing, Aalborg, Denmark, Sep. 2018.

[28] V. P. Oikonomou and I. Kompatsiaris, "A novel bayesian approach for EEG source localization," Computational Intelligence and Neuroscience, vol. 2020, p. 8837954, Oct. 2020.