HAL
open science

M/EEG source localization with multi-scale time-frequency dictionaries

Yousra Bekhti, Daniel Strohmeier, Mainak Jas, Roland Badeau, Alexandre Gramfort

To cite this version:

Yousra Bekhti, Daniel Strohmeier, Mainak Jas, Roland Badeau, Alexandre Gramfort. M/EEG source localization with multi-scale time-frequency dictionaries. 6th International Workshop on Pattern Recognition in Neuroimaging (PRNI), Jun 2016, Trento, Italy. hal-01306605

HAL Id: hal-01306605

https://hal.science/hal-01306605v1

Submitted on 21 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

M/EEG source localization with multi-scale time-frequency dictionaries

Yousra Bekhti*, Daniel Strohmeier†, Mainak Jas*, Roland Badeau*, Alexandre Gramfort*
*LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France
†Institute of Biomedical Engineering and Informatics, Technische Universität Ilmenau, Ilmenau, Germany

***Abstract*—Magnetoencephalography (MEG) and electroencephalography (EEG) source localization is a challenging ill-posed problem. To identify an appropriate solution out of an infinite set of possible candidates, the problem requires setting certain constraints depending on the assumptions or a priori knowledge about the source distribution. Different constraints have been proposed so far, including those that impose sparsity on the source reconstruction in both standard and time-frequency domains. Source localization in the time-frequency domain has already been investigated using Gabor dictionary in both a convex (TF-MxNE) and non-convex way (Iterative Reweighted TF-MxNE). The iterative reweighted (ir)TF-MxNE solver has been shown to outperform TF-MxNE in both source recovery and amplitude bias. However, the choice of an optimal dictionary remains unsolved. Due to a mixture of signals, *i.e.* short transient signals (right after the stimulus onset) and slower brain waves, the choice of a single dictionary explaining simultaneously both signals types in a sparse way is difficult. In this work, we introduce a method to improve the source estimation relying on a multi-scale dictionary, *i.e.* multiple dictionaries with different scales concatenated to fit short transients and slow waves at the same time. We compare our results with irTF-MxNE on realistic simulation, then we use somatosensory data to demonstrate the benefits of the approach on in terms of reduced leakage (time courses mixture), temporal smoothness and detection of both signals types.**

keywords— Inverse problem; MEEG; iterative reweighted optimization algorithm; multi-scale dictionary; Gabor transform.

I. Introduction

Magneto-/electroencephalography (M/EEG) allow for non-invasive analysis of functional brain imaging with high temporal and good spatial resolution. Various approaches to tackle the source localization problem from M/EEG data have been proposed in the literature. The distributed-source approach models the brain activity with a fixed number of candidate dipoles distributed over the brain, and estimates their amplitudes and orientations. As the number of candidate dipoles that can explain the measured data is much larger than the number of sensors, source localization is an ill-posed problem. This implies that there is not a unique solution. Literature shows that adding supplementary constraints such as sparse regularization of priors to the neural activation helps to tackle the problem. Those approaches are based on Bayesian modeling [1]–[4], or regularized regression [5]–[7]. These methods implicitly assume stationarity of the source activation. In contrast, the Time-Frequency Mixed Norm Estimate (TF-MxNE) [8], Spatio-Temporal Unifying Tomography (STOUT) [9] and the iterative reweighted TF-MxNE (irTF-MxNE) [10] improve reconstruction of transient and non-stationary sources by promoting structured sparsity in the time-frequency (TF) domain. Those methods compute a sparse group LASSO on the TF coefficients. TF-MxNE and STOUT apply a composite convex penalty, the sum of an $\ell_{2,1}$-mixed-norm and an ℓ_1-norm penalty, on the Gabor transform of the source time courses. On the other hand, irTF-MxNE applies a composite *non-convex* penalty, the sum of an $\ell_{2,0.5}$-quasinorm and an $\ell_{0.5}$-quasinorm penalty on the TF. The non-convex penalties have been shown to outperform convex approaches both in terms of source recovery and amplitude bias [11], [12]. However, the choice of an optimal dictionary for decomposing the mixed data remains difficult.

To address this issue, we show in this paper how to incorporate in the iterative reweighted optimization algorithm a multi-scale dictionary while keeping computational efficiency. The optimization problem is solved in the same way as irTF-MxNE [10] *i.e.* each iteration is a weighted TF-MxNE, which we solve using block coordinate descent (BCD) and an active set strategy [13]. We compare irTF-MxNE with and without a multi-scale dictionary on simulated and real MEG data. We demonstrate the benefit of the multi-scale dictionary in terms of reconstructed source time courses and temporal unmixing of activations.

Notation The transpose of a matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ is indicated by $\mathbf{A}^{\mathbf{T}}$, and the Hermitian conjugate by $\mathbf{A}^{\mathcal{H}}$. $\mathbf{A}[i,:]$ and $\mathbf{A}[:,j]$ correspond to the ith row the jth column respectively, and $\mathbf{A}[i,j]$ to the element in the ith row and j^{th} column of $\mathbf{A}$. $\|\mathbf{A}\|_{Fro}$ indicates the Frobenius norm, and $\|\mathbf{A}\|_{\mathbf{W};p,q}$ the weighted mixed norm with $\|\mathbf{A}\|_{\mathbf{W};p,q} = \left[\sum_i \left(\sum_j \mathbf{W}[i,j]|\mathbf{A}[i,j]|^q\right)^{p/q}\right]^{1/p}$. The mixed norm $\|\mathbf{A}\|_{p,q} = \|\mathbf{A}\|_{\mathbf{I};p,q}$, and $\|\mathbf{A}\|_p = \left(\sum_i \sum_j |\mathbf{A}[i,j]|\right)^{1/p}$

II. Materials and Methods

A. Inverse problem in the TF domain

Using a dictionary of TF atoms, such as a tight Gabor frame, $\mathbf{\Phi} \in \mathbb{C}^{T\times C}$ (T samples, C atoms), the neuronal activation $\mathbf{X} \in \mathbb{R}^{S\times T}$ (S sources) can be modeled as a linear combination of atoms, $\mathbf{X} = \mathbf{Z}\mathbf{\Phi}^{\mathcal{H}}$, where $\mathbf{Z} \in \mathbb{C}^{S\times C}$ is the TF coefficient matrix. A Gabor frame $\mathbf{\Phi}$ is tight when the Euclidean norm of the input signal and the vector of TF coefficients are proportional ($\|\mathbf{Z}\|_2^2 = A_{\mathbf{\Phi}}\|\mathbf{X}\|_2^2$ where

$A_{\mathbf{\Phi}} > 0$). The MEG/EEG measurements $\mathbf{M} \in \mathbb{R}^{N \times T}$ (N sensors) follows the forward model:

$$\mathbf{M} = \mathbf{G}\mathbf{X} + \mathbf{E} = \mathbf{G}\mathbf{Z}\mathbf{\Phi}^{\mathcal{H}} + \mathbf{E} \tag{1}$$

where $\mathbf{G} \in \mathbb{R}^{N \times S}$ stands for the forward operator, each source s can have one or three orientations. $\mathbf{E} \in \mathbb{R}^{N \times T}$ is the measurement noise, which can be assumed to be additive white noise: $\mathbf{E}[:, j] \sim \mathcal{N}(0, \mathbf{I})$ for all j after spatial whitening [14]. Estimating the coefficients $\mathbf{Z}$ given the measurement $\mathbf{M}$ is an ill-posed problem and constraints have to be imposed on $\mathbf{Z}$ to obtain a unique source estimate. For analyzing evoked responses, we assume that the neuronal activation is spatially sparse and temporally smooth. This corresponds to a row sparsity [8], which we promote by applying a composite non-convex regularization $\mathcal{R}(\mathbf{Z})$. The associated regularized regression problem is:

$$\mathbf{Z}^{\star} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{M} - \mathbf{G}\mathbf{Z}\mathbf{\Phi}^{\mathcal{H}}\|_{Fro}^2 + \mathcal{R}(\mathbf{Z}) \tag{2}$$

with

$$\mathcal{R}(\mathbf{Z}) = \lambda_{space}\|\mathbf{Z}\|_{2,0.5} + \lambda_{time}\|\mathbf{Z}\|_{0.5}$$

where $\lambda_{space} > 0$, $\lambda_{time} > 0$. A large regularization parameter λ_{space} will lead to a spatially very sparse solution, while a large λ_{time} will promote sources with smooth time series.

B. Fast iterative reweighted TF-MxNE with tight frames

Given a dictionary $\mathbf{\Phi}$, the optimization problem in Eq. (2) can be solved by iteratively minimizing convex surrogate problems [10]. The regularization term at each iteration k is a weighted convex mixed norm that can be written as:

$$\mathcal{R}(\mathbf{Z}) = \lambda_{space}\|\mathbf{Z}\|_{\mathbf{W}_1^{(k)};2,1} + \lambda_{time}\|\mathbf{Z}\|_{\mathbf{W}_2^{(k)};1} \tag{3}$$

with $\forall s, c$,

$$\mathbf{W}_1^{(k)}[s,c] = \left(2\sqrt{\|\hat{\mathbf{Z}}^{(k-1)}[s,:]\|_2 + \epsilon^{(k-1)}}\right)^{-2}$$
$$\mathbf{W}_2^{(k)}[s,c] = \left(2\sqrt{|\hat{\mathbf{Z}}^{(k-1)}[s,c]| + \epsilon^{(k-1)}}\right)^{-1},$$

where $\mathbf{W_1}$ and $\mathbf{W_2}$ are the weights applied to the TF coefficients, and $\hat{\mathbf{Z}}^{(k-1)}$ are the estimated coefficients at iteration $k-1$. $\epsilon^{(k-1)} \in \mathbb{R}^+$ is used to prevent infinite weights. Here, ϵ is set to 0 and infinite weights are handled as in [10].

For solving Eq. (2), we use BCD [15]. The algorithm boils down to sequentially computing a gradient step and the proximity operator (PO) of the $\ell_{2,1} + \ell_1$ norm for each block s of coefficients. Here a block maps to a location in the brain. One update of a block of coefficients is given by first a gradient step:

$$\mathbf{R} = \mathbf{M} - \mathbf{G}\hat{\mathbf{X}} \tag{4}$$

$$\bar{\mathbf{X}}[s,:] = \hat{\mathbf{X}}[s,:] + \mu[s]\mathbf{G}[:,s]^{\mathbf{T}}\mathbf{R} \tag{5}$$

$$\bar{\mathbf{Z}}[s,:] = \bar{\mathbf{X}}[s,:]\mathbf{\Phi} \tag{6}$$

followed by the computation of the PO of the weighted $\ell_{2,1} + \ell_1$ norm [8], [10]:

$$\tilde{\mathbf{Z}}[s,c] = \bar{\mathbf{Z}}[s,c]\left(1 - \frac{\mu[s]\lambda_{time}\mathbf{W}_2^{(k)}[s,c]}{|\bar{\mathbf{Z}}[s,c]|}\right)^{+} \tag{7}$$

$$\hat{\mathbf{Z}}[s,c] = \tilde{\mathbf{Z}}[s,c]\left(1 - \frac{\mu[s]\lambda_{space}\sqrt{\mathbf{W}_1^{(k)}[s,c]}}{\|\tilde{\mathbf{Z}}[s,:]\|_2}\right)^{+} \tag{8}$$

with $(a)^+ = \max(a, 0)$. When $\mathbf{\Phi}$ is a tight frame, $\mu[s]$ is given by $\mu[s] = \sqrt{\mathbf{A}_{\mathbf{\Phi}}}(\|\mathbf{G}[:,s]^{\mathbf{T}}\mathbf{G}[:,s]\|)^{-1}$. Finally:

$$\hat{\mathbf{X}}[s,:] = \hat{\mathbf{Z}}[s,:]\mathbf{\Phi}^{\mathcal{H}} \tag{9}$$

Eq. (7) and (8) are respectively solutions of the PO for the weighted ℓ_1 norm and for the weighted $\ell_{2,1}$ norm. As the ℓ_1 PO shrinks coefficients towards zero, if a block of coefficients were set to zero by the $\ell_{2,1}$ PO it would also be set to zero after the application of the ℓ_1 PO. As a consequence, it is possible to know just by applying the $\ell_{2,1}$ PO to $\bar{\mathbf{X}}[s,:]$ if the set of coefficients $\tilde{\mathbf{Z}}[s,:]$ will be set to zero. Note that this is just a sufficient condition and we may have to compute all steps to know if the block is set to zero. This is summarized in the following lemma.

Lemma 1: Let $\mathbf{\Phi}$ be a frame with constant $A_{\mathbf{\Phi}}$, if $\|\bar{\mathbf{X}}[s,:]\|_2 \leq \mu[s]\lambda_{space}\sqrt{\mathbf{W}_1^{(k)}[s,c]}/\sqrt{A_{\mathbf{\Phi}}}$ then $\hat{\mathbf{Z}}[s,c] = 0$, $\forall c$.

Computing the TF decomposition at each iteration can be costly. The consequence of the lemma is that for a lot of source locations one can avoid computing their TF decomposition during the optimization just by computing the ℓ_2 norm of the time courses after the gradient step. To speed up the computation even more, we combine the BCD scheme with an active set strategy [13], which primarily updates sources that are likely to be active, while keeping the remaining sources inactive.

C. Inverse problem with multi-scale tight Gabor frames

A Gabor tight frame is computed by setting two parameters: the length of the window that defines the time/frequency resolution, and an overlap parameter that defines the time step from one window to another. Each source waveform is a sparse linear combination of atoms from this dictionary. Fixing those parameters is then critical for having an optimal dictionary. Learning the dictionary might be a solution to avoid fixing the parameters, or the need to have an overcomplete dictionary covering a broad range of scales. However, learning both $\mathbf{Z}$ and $\mathbf{\Phi}$ simultaneously is a non-convex optimization problem, for which one needs to alternate between a convex optimization for the two variables [16].

Let us define a multi-scale TF dictionary, where we concatenate Q tight Gabor frames $\mathbf{\Phi}_q$, $1 \leq q \leq Q$, with different resolutions. One can realize that this union of tight frames $\mathbf{\Phi} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_Q]$ is also a tight frame with $A_{\mathbf{\Phi}} = \sum_q A_{\mathbf{\Phi}_q}$. The strategy presented previously is therefore still relevant for a multi-scale dictionary, where the activation $\mathbf{Z}$ is a concatenation of $\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_Q$.

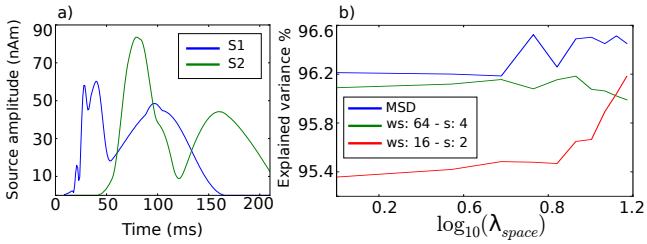

Fig. 1. (a) Simulated source time courses in S1 (blue) and S2 (green). (b) The explained variance for irTF-MxNE using two different dictionaries: long window size (ws) 64 with time shift (s) 4 (green), and small ws 16 with s 2 (red). The combination of the two dictionaries is shown in blue. This shows how the multi-scale dictionary (MSD) improves the explained variance.

III. Results

We first evaluate the accuracy of irTF-MxNE with and without multi-scale on realistic simulations. We then apply our new solver on MEG somatosensory data.

A. Simulation

We generated a realistic simulation dataset based on a fixed-orientation source model with 7549 cortical locations and 102 magnetometers. Two of these locations were selected to be active in the primary and secondary somatosensory cortex (S1 and S2). The corresponding time courses are shown in Fig. 1-a in blue (S1) and green (S2). We have both a transient source around 40 ms and slow waves afterwards around 70, 100 and 150 ms. irTF-MxNE solver improves the source recovery [10]. Therefore, we do not compare the solvers presented here over active set size or an F_1 measure, as both solvers are already able to recover all the sources. We evaluate our approach by computing the explained variance between simulated source courses and the source estimation from each solver as follows:

$$\theta = 1 - \frac{\|GX_{sim} - GX_{est}\|_{Fro}^2}{\|GX_{sim}\|_{Fro}^2} \tag{10}$$

Fig. 1-b shows the explained variance for the irTF-MxNE with different dictionaries over a logarithmic grid of λ_{space}. The first Gabor dictionary is constructed with 64 samples (64 ms) window and 4 samples time shift (green), the second Gabor dictionary is constructed with 16 samples window and 2 samples time shift (red) and the third one is the combination of the two dictionaries (blue). We observe that the irTF-MxNE solver using the combination of two dictionaries outperforms the solver with each dictionary separately in terms of explained variance measure over all parameters range. Higher values of $\log(\lambda_{space}) > 1.2$ impose high penalization on the active set size, resulting in a too sparse source estimate, where the solution is not explaining the measurement anymore. The results show on simulation a source reconstruction improvement, where it leads to a larger explained variance.

B. Experimental results with MEG somatosensory data

To demonstrate the advantage of irTF-MxNE with a multi-scale dictionary over the basic irTF-MxNE, we tested different parameters for different solvers on a MEG dataset: MIND (for details [17]). Source estimation was first performed using

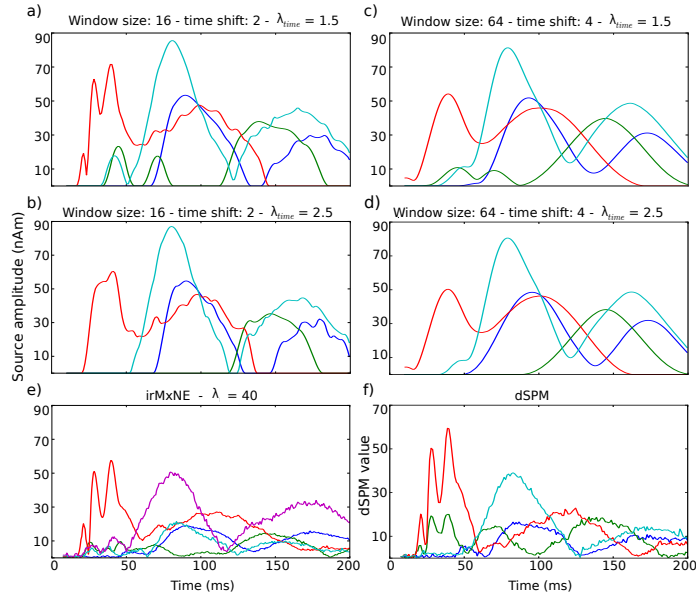

Fig. 2. Source reconstruction using somatosensory data with different solvers. (a) - (b) irTF-MxNE on a small window dictionary with $\lambda_{time} = 1.5$ and $\lambda_{time} = 2.5$ respectively. (c) - (d) irTF-MxNE on a long window dictionary with $\lambda_{time} = 1.5$ and $\lambda_{time} = 2.5$ respectively. From (a) to (d) $\lambda_{space} = 28.5$ (e) irMxNE for $\lambda = 40$ and (f) dSPM activation for the four activated sources.

several solvers: irTF-MxNE, irMxNE [18] and dSPM [19]. Regarding irTF-MxNE, two dictionaries were tested. A dictionary with a 64 samples window and a 4 samples time shift, which leads to smooth source courses; and a dictionary with a 16 samples window and a 2 samples time shift, which helps capture short transient sources. After inspection of the residual, results showed that at least four sources are necessary to capture all evoked components. We have therefore fixed the parameters of the irTF-MxNE solvers so we obtained only four sources while explaining as much variance as possible. After that, we experimented with two different parameters $\lambda_{time} = 1.5$ and $\lambda_{time} = 2.5$ to show their impact on the smoothness of the different time sources obtained. Fig.2 (a-b) demonstrates the four time courses obtained with irTF-MxNE using the short window dictionary for the selected values of λ_{time}. We show that for high values of λ_{time} (b), the solver is not able to capture the short transient component around 40 ms. While for a small value (a), the unmixing is not reliable since the light blue and the green source estimates are catching the activity from the red source. Additionally, the time courses are not smooth. On the other hand, Fig. 2 (c-d) demonstrate the four time courses obtained with irTF-MxNE using the long window dictionary for the selected λ_{time}. They confirm that both parameters are not able to capture the transient effect after the stimulus, although the time courses are smooth. These four subfigures reveal that a combination of the two dictionaries should be critical to acquire source estimates with high precision. Moreover, Fig 2-e displays the amplitudes obtained with irMxNE for five sources, as for irMxNE, one is not able to obtain the four relevant sources unmixed [6]. We notice that the light blue

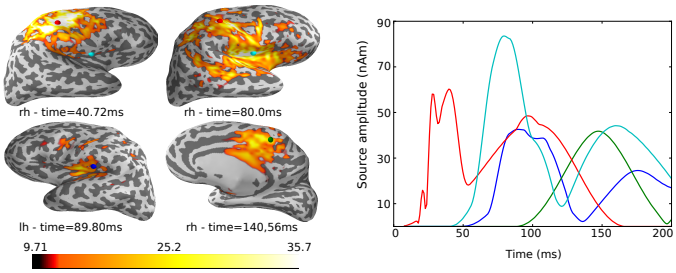

Fig. 3. Source reconstruction using somatosensory data with a multi-scale irTF-MxNE. The solver estimates four sources for $\lambda_{space} = 28.5$ and $\lambda_{time} = 1.3$. The source locations marked with spheres in right (rh) and left (lh) hemisphere, and their corresponding activation are color-coded.

source in Fig. 2-(a) to (d) appears as two separate sources in (e): light blue and purple. If we increase the λ parameter, we increase the amplitude bias due to the l_1 norm of the solver. If we set it too high ($\lambda = 50$) we obtain four sources, but the blue source which is relevant to the study would be removed and the duplicated purple source is kept. The last panel Fig. 2-f displays the source estimates for dSPM values corresponding to the four locations of the sources obtained with the irTF-MxNE. These subfigures show that none of irMxNE or dSPM solvers are able to obtain smooth sources without any leakage between the time courses.

Source estimation was then achieved using irTF-MxNE with the combination of the two dictionaries. Fig. 3 shows source reconstruction using the multi-scale irTF-MxNE for the regularization parameters $\lambda_{space} = 28.5$ and $\lambda_{time} = 1.5$. Each source's location is marked by a sphere in Fig. 3 left, and its amplitude over time is color-coded in on the right panel. The results show a suitable succession of the sources. The transient source (red) is the only source explaining the event related field until 48 ms. This red source corresponds to the contralateral primary somatosensory cortex (cS1) located in the postcentral gyrus of the parietal lobe (right hemisphere (rh)). The red sphere on the lateral view coincides with the smeared dSPM activation around 40 ms. The second source (light blue) corresponds to the secondary somatosensory cortex (cS2), and also occurs with dSPM activation around 80 ms. About 100 ms after stimulus, additional cortical sources are activated, such as ipsilateral secondary somatosensory cortex (iS2) (blue-lh), and contralateral medial wall (green-rh).

IV. Discussion and Conclusion

We have presented an improvement in the irTF-MxNE solver using a multi-scale dictionary to capture the mixture of the MEG/EEG data. The non-convex optimization problem is solved by iteratively solving the convex weighted TF-MxNE problem using block coordinate descent combined with active set strategy as in [10]. The benefits of the multi-scale irTF-MxNE have been shown on simulated and MEG somatosensory data. Both experiments confirm that multi-scale irTF-MxNE improves the source estimates, in terms of reduced mixing of the time courses, smoothness and detection of both short transients and slower waves. In contrast, both solvers are efficient regarding active set size and amplitude bias, which is due to the non-convexity of the methods. Hence, the multi-scale irTF-MxNE should be applied to data where a mixture of signals coexist, and when the aim is to acquire focal sources with non-stationary and smooth time courses. This solver will be made available in the MNE-Python package [20].

V. Acknowledgement

This work was supported by the ANR THALAMEEG ANR-14-NEUC-0002-01, the NIH R01 MH106174, and the DFG HA 2899/21-1.

References

[1] K. Friston et al., "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, 2008.
[2] K. Sekihara and S. S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," in *Electromagnetic Brain Imaging*. Springer, 2015, pp. 119–137.
[3] A. Bolstad, B. Van Veen, and R. Nowak, "Space–time event sparse penalization for magneto-/electroencephalography," *NeuroImage*, vol. 46, no. 4, pp. 1066–1081, 2009.
[4] F. Lucka, S. Pursiainen, M. Burger, and C. H. Wolters, "Hierarchical Bayesian inference for the EEG inverse problem using realistic FEM head models: depth localization and source separation for focal primary currents," *NeuroImage*, vol. 61, no. 4, pp. 1364–1382, 2012.
[5] W. Ou et al., "A distributed spatio-temporal EEG/MEG inverse solver," *NeuroImage*, vol. 44, no. 3, pp. 932–946, 2009.
[6] A. Gramfort, M. Kowalski, and M. Hämäläinen, "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods," *Phys. Med. Biol.*, vol. 57, no. 7, p. 1937, 2012.
[7] S. Haufe, V. V. Nikulin, A. Ziehe, K.-R. Müller, and G. Nolte, "Combining sparsity and rotational invariance in EEG/MEG source reconstruction," *NeuroImage*, vol. 42, no. 2, pp. 726–738, 2008.
[8] A. Gramfort et al., "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations," *NeuroImage*, vol. 70, pp. 410–422, 2013.
[9] S. Castaño-Candamil, J. Höhne, J.-D. Martínez-Vargas, X.-W. An, G. Castellanos-Domínguez, and S. Haufe, "Solving the EEG inverse problem based on space–time–frequency structured sparsity constraints," *NeuroImage*, vol. 118, pp. 598–612, 2015.
[10] D. Strohmeier, A. Gramfort, and J. Haueisen, "MEG/EEG source imaging with a non-convex penalty in the time-frequency domain," in *PRNI, 2015 International Workshop on*. IEEE, 2015, pp. 21–24.
[11] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.
[12] I. Daubechies et al., "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, no. 1, pp. 1–38, 2010.
[13] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of statistical software*, vol. 33, no. 1, p. 1, 2010.
[14] D. Engemann, D. Strohmeier, E. Larson, and A. Gramfort, "Mind the noise covariance when localizing brain sources with M/EEG," in *PRNI, 2015 International Workshop on*. IEEE, 2015, pp. 9–12.
[15] P. Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Mathematical Programming*, vol. 125, no. 2, pp. 263–295, 2010.
[16] J. Montoya-Martínez, A. Artés-Rodríguez, M. Pontil, and L. K. Hansen, "A regularized matrix factorization approach to induce structured sparse-low-rank solutions in the EEG inverse problem," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, pp. 1–13, 2014.
[17] M. Weisend et al., "Paving the way for cross-site pooling of magnetoencephalography (MEG) data," in *International Congress Series*, vol. 1300. Elsevier, 2007, pp. 615–618.
[18] D. Strohmeier, J. Haueisen, and A. Gramfort, "Improved MEG/EEG source localization with reweighted mixed-norms," in *PRNI, 2014 International Workshop on*. IEEE, 2014, pp. 1–4.
[19] A. M. Dale et al., "Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity," *Neuron*, vol. 26, no. 1, pp. 55–67, 2000.
[20] A. Gramfort et al., "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, 2014.