MEG/EEG source imaging with a non-convex penalty in the time-frequency domain

Daniel Strohmeier, Alexandre Gramfort, Jens Haueisen

HAL Id: hal-01183552

https://hal-imt.archives-ouvertes.fr/hal-01183552

Submitted on 11 Aug 2015

MEG/EEG source imaging with a non-convex penalty in the time-frequency domain

Daniel Strohmeier*, Alexandre Gramfort†‡, Jens Haueisen*§

*Institute of Biomedical Engineering and Informatics, Technische Universität Ilmenau, Ilmenau, Germany
Email: daniel.strohmeier@tu-ilmenau.de
†Institut Mines-Telecom, Telecom ParisTech, CNRS LTCI, Paris, France
‡NeuroSpin, CEA Saclay, Bat. 145, Gif-sur-Yvette Cedex, France
§Biomagnetic Center, Department of Neurology, Jena University Hospital, Jena, Germany

Abstract—Due to the excellent temporal resolution, MEG/EEG source imaging is an important measurement modality to study dynamic processes in the brain. As the bioelectromagnetic inverse problem is ill-posed, constraints have to be imposed on the source estimates to find a unique solution. These constraints can be applied either in the standard or a transformed domain. The Time-Frequency Mixed Norm Estimate applies a composite convex regularization functional promoting structured sparsity in the time-frequency domain by combining an $\ell_{2,1}$-mixed-norm and an ℓ_1-norm penalty on the coefficients of the Gabor TF decomposition of the source signals, to improve the reconstruction of spatially sparse neural activations with non-stationary and transient signals. Due to the ℓ_1-norm based constraints, the resulting source estimates are however biased in amplitude and often suboptimal in terms of source selection. In this work, we present the iterative reweighted Time-Frequency Mixed Norm Estimate, which employs a composite non-convex penalty formed by the sum of an $\ell_{2,0.5}$-quasinorm and an $\ell_{0.5}$-quasinorm penalty. The resulting non-convex problem is solved with a reweighted convex optimization scheme, in which each iteration is equivalent to a weighted Time-Frequency Mixed-Norm Estimate solved efficiently using a block coordinate descent scheme and an active set strategy. We compare our approach to alternative solvers using simulations and analysis of MEG data and demonstrate the benefit of the iterative reweighted Time-Frequency Mixed Norm Estimate with regard to active source identification, amplitude bias correction, and temporal unmixing of activations.

Keywords—MEG; EEG; inverse problem; structured sparsity; Gabor transform; iterative reweighted optimization algorithm

I. INTRODUCTION

Functional neuroimaging allows for the noninvasive examination of brain function by identifying neuronal regions involved in a specific cognitive task. To examine dynamic processes in the brain, MEG/EEG source imaging can be applied due to its excellent temporal resolution. The MEG/EEG inverse problem is however ill-posed in the sense of Hadamard and constraints have to be imposed to obtain unique source estimates. These constraints reflect assumptions on the neuronal activation such as the number of active regions, and the spatial or temporal characteristics. Several source imaging methods promoting spatial sparsity of the neuronal activation

have been proposed based on Bayesian modeling [1], [2] or regularized regression [3], [4]. These methods implicitly assume stationarity of the source activation. In contrast, the Time-Frequency Mixed Norm Estimate (TF-MxNE) [5] improves the reconstruction of transient and non-stationary sources by promoting structured sparsity in the time-frequency (TF) domain. TF-MxNE computes a sparse group lasso in the TF domain by applying 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. This involves an automatic denoising of the source estimates improving smoothness of the time courses. Due to the ℓ_1-norm based constraints, the resulting source estimates are biased in amplitude and often suboptimal in terms of source selection [6]. Regularized regression with non-convex penalties, such as the ℓ_p-quasinorm with $0 < p < 1$, has been shown to outperform convex approaches both in terms of amplitude bias and source recovery [6], [7]. The resulting non-convex problems can be solved e.g. using iterative reweighted convex optimization schemes such as iterative reweighted ℓ_2 [7]–[9] and iterative reweighted ℓ_1 [6], [10], [11]. In this work, we present the novel iterative reweighted Time-Frequency Mixed Norm Estimate (irTF-MxNE) for solving the MEG/EEG inverse problem. This regularized regression technique applies a composite non-convex penalty in the TF domain, which combines an $\ell_{2,0.5}$-quasinorm and an $\ell_{0.5}$-quasinorm penalty on the TF transformed source activations. We explain how to solve the corresponding non-convex optimization problem using iterative reweighted convex optimization. Each iteration of this approach is a weighted TF-MxNE, which we solve efficiently using block coordinate descent and an active set strategy. We compare irTF-MxNE and TF-MxNE using simulations and analysis of MEG data and demonstrate the benefit of irTF-MxNE in terms of active source identification, amplitude bias correction and temporal unmixing of activations.

Notation: The transpose of a vector $\mathbf{a} \in \mathbb{C}^N$ or matrix $\mathbf{A} \in \mathbb{C}^{N \times M}$ is indicated by $\mathbf{a}^T$ and $\mathbf{A}^T$, and the Hermitian conjugate by $\mathbf{A}^{\mathcal{H}}$. The scalar $\mathbf{a}[i]$ indicates the ith element of $\mathbf{a}$. $\mathbf{A}[i, :]$ corresponds to the ith row, $\mathbf{A}[:, j]$ to the jth column, and $\mathbf{A}[i, j]$ to the element in the ith row and j^{th} column of $\mathbf{A}$. $\|\mathbf{A}\|_{\text{Fro}}$ indicates the Frobenius norm, $\|\mathbf{A}\|$ the spectral norm, and $\|\mathbf{A}\|_{\mathbf{W};p,q}$ the weighted mixed norm with

$$\|\mathbf{A}\|_{\mathbf{W};p,q} = \left(\sum_i \left(\sum_j \mathbf{W}[i,j] \, |\mathbf{A}[i,j]|^q \right)^{\frac{p}{q}} \right)^{\frac{1}{p}} .$$

II. MATERIALS AND METHODS

A. The M/EEG inverse problem with TF dictionaries

Using a tight Gabor frame $\mathbf{\Phi} \in \mathbb{C}^{T \times C}$ (T samples, C Gabor atoms) the neuronal activation $\mathbf{X} \in \mathbb{R}^{S \times T}$ (S sources) can be modeled as a linear combination of time-frequency atoms, $\mathbf{X} = \mathbf{Z}\mathbf{\Phi}^{\mathcal{H}}$. The MEG/EEG measurements $\mathbf{M} \in \mathbb{R}^{N \times T}$ (N sensors) can thus be computed based on the MEG/EEG forward model given in Eq. (1):

$$\mathbf{M} = \mathbf{GX} + \mathbf{E} = \mathbf{GZ}\mathbf{\Phi}^{\mathcal{H}} + \mathbf{E} \ , \qquad (1)$$

where $\mathbf{G} \in \mathbb{R}^{N \times S}$ is the gain matrix, $\mathbf{Z} \in \mathbb{C}^{S \times C}$ is the TF co-efficient matrix, and $\mathbf{E}$ is the measurement noise, which can be assumed to be additive white Gaussian noise, $\mathbf{E}[:, j] \sim \mathcal{N}(0, \mathbf{I})$ for all j after spatial whitening [12]. Estimating $\mathbf{Z}$ given $\mathbf{M}$ is an ill-posed inverse 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 TF coefficient matrix with a block row structure with intra-row sparsity [5], which we promote by applying a composite non-convex regularization functional $\mathcal{R}(\mathbf{Z})$. The associated regularized regression problem is given in Eq. (2).

$$\widehat{\mathbf{Z}} = \arg\min_{\mathbf{Z}} \frac{1}{2} \|\mathbf{M} - \mathbf{GZ}\mathbf{\Phi}^{\mathcal{H}}\|_{\text{Fro}}^2 + \lambda \mathcal{R}(\mathbf{Z}) \qquad (2)$$

with

$$\mathcal{R}(\mathbf{Z}) = (1-\rho)\mathcal{R}_1(\mathbf{Z}) + \rho\mathcal{R}_2(\mathbf{Z}) = $$
$$= (1-\rho)\sum_s \sqrt{\|\mathbf{Z}[s,:]\|_2} + \rho \sum_{s,c} \sqrt{|\mathbf{Z}[s,c]|}$$

where $\lambda > 0$ balances the data fit and penalty, and $0 \le \rho \le 1$ controls the trade-off between the two non-convex penalties.

B. The iterative reweighted TF-MxNE

Following the framework of Majorization-Minimization [6], we solve the optimization problem in Eq. (2) by solving iteratively reweighted convex surrogate optimization problems. Using local linear approximations of the non-convex penalties at $\widehat{\mathbf{Z}}^{(k-1)}$, the estimate obtained in the (k-1)$^{\text{th}}$ iteration, $\widehat{\mathbf{Z}}^{(k)}$ is computed using Eq. (3):

$$\widehat{\mathbf{Z}}^{(k)} = \arg\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{M} - \mathbf{GZ}\mathbf{\Phi}^{\mathcal{H}}\|_{\text{Fro}}^2 + \lambda \mathcal{R}^*(\mathbf{Z}) \qquad (3)$$

with

$$\mathcal{R}^*(\mathbf{Z}) = (1-\rho)\|\mathbf{Z}\|_{\mathbf{W}_1^{(k)};2,1} + \rho\|\mathbf{Z}\|_{\mathbf{W}_2^{(k)};1,1}$$

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

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

Eq. (3) corresponds to a weighted TF-MxNE with weight matrices $\mathbf{W}_1$ and $\mathbf{W}_2$, where $\epsilon^{(k-1)} \in \mathbb{R}^+$ is used to prevent infinite weights [6].

For solving Eq. (3), we use a block coordinate descent (BCD) scheme [13]. The BCD subproblem per source location can be solved in closed form using the proximity operator associated to $\mathcal{R}^*(\mathbf{Z})$. This is a hierarchical group penalty [14] and the corresponding proximity operator is equivalent to

applying the proximity operators for $\|\mathbf{Z}\|_{\mathbf{W};1,1}$ and $\|\mathbf{Z}\|_{\mathbf{W};2,1}$ successively [4], [5]. The resulting closed form solution for the sth source location is given in Eq. (4), where we use intermediate solutions $\overline{\mathbf{Z}}[s,:]^{(k)}$ and $\widetilde{\mathbf{Z}}[s,:]^{(k)}$ to improve readability.

$$\mathbf{R}^{(k-1)} = \mathbf{M} - \mathbf{G}\widehat{\mathbf{Z}}^{(k-1)}\mathbf{\Phi}^{\mathcal{H}}$$
$$\overline{\mathbf{Z}}[s,:]^{(k)} = \widehat{\mathbf{Z}}[s,:]^{(k-1)} + \boldsymbol{\mu}[s]\mathbf{G}[:,s]^T \mathbf{R}^{(k-1)}\mathbf{\Phi}$$
$$\widetilde{\mathbf{Z}}[s,c]^{(k)} = \overline{\mathbf{Z}}[s,c]^{(k)}\left(1 - \frac{\boldsymbol{\mu}[s]\lambda\rho\mathbf{W}_2^{(k)}[s,c]}{|\overline{\mathbf{Z}}[s,c]^{(k)}|}\right)^+ \qquad (4)$$
$$\widehat{\mathbf{Z}}[s,c]^{(k)} = \widetilde{\mathbf{Z}}[s,c]^{(k)}\left(1 - \frac{\boldsymbol{\mu}[s]\lambda(1-\rho)\sqrt{\mathbf{W}_1^{(k)}[s,c]}}{\|\widetilde{\mathbf{z}}[s,:]^{(k)}\|_2}\right)^+$$

with $(a)^+ = \max(a, 0)$ and $\frac{0}{0} = 0$ by definition. Since $\mathbf{\Phi}$ is a tight frame, $\boldsymbol{\mu}[s]$ is given by $\boldsymbol{\mu}[s] = (\|\mathbf{G}[:,s]^T\mathbf{G}[:,s]\|)^{-1}$.

Since $\mathcal{R}(\mathbf{Z})$ promotes spatial sparsity, most blocks in $\widehat{\mathbf{Z}}$ should be zero. To speed up computation, we combine the BCD scheme with an active set strategy proposed in [15], which primarily updates sources, that are likely to be active, while keeping the remaining sources inactive. Pseudo code for the proposed BCD scheme with active set approach for solving TF-MxNE can be found in Algorithm 1.

Algorithm 1 TF-MxNE with BCD and active set approach

1: **for** $s = 1$ to S **do**
2: $\quad \widehat{\mathbf{Z}}[s,:] \longleftarrow$ Solve Eq. (4)
3: **end for**
4: $\widehat{\mathcal{A}} = \{s \mid \|\widehat{\mathbf{Z}}[s,:]\|_2 > 0\}$
5: **repeat**
6: $\quad$ **repeat**
7: $\qquad$ **for** $s \in \widehat{\mathcal{A}}$ **do**
8: $\qquad\quad \widetilde{\mathbf{Z}}[s,:] \longleftarrow$ Solve Eq. (4)
9: $\qquad$ **end for**
10: $\quad$ **until** convergence
11: $\quad \widetilde{\mathcal{A}} = \{s \mid \|\widetilde{\mathbf{Z}}[s,:]\|_2 > 0\}$
12: $\quad$ **for** $s = 1$ to S **do**
13: $\qquad \widehat{\mathbf{Z}}[s,:] \longleftarrow$ Solve Eq. (4)
14: $\quad$ **end for**
15: $\quad \widehat{\mathcal{A}} = \{s \mid \|\widehat{\mathbf{Z}}[s,:]\|_2 > 0\}$
16: **until** $\widehat{\mathcal{A}} = \widetilde{\mathcal{A}}$

Using Alg. 1, we can derive the pseudo code for irTF-MxNE given in Alg. 2. We terminate irTF-MxNE when the minimum change in the cost function is below a user specified threshold.

Algorithm 2 irTF-MxNE with BCD and active set approach

1: Require: $\widehat{\mathbf{Z}}^{(0)}$, $\epsilon^{(0)}$
2: **repeat**
3: $\quad \mathbf{W}_1^{(k)}, \mathbf{W}_2^{(k)} \longleftarrow$ Using Eq. 3 with $\widehat{\mathbf{Z}}^{(k-1)}$ and $\epsilon^{(k-1)}$
4: $\quad \widehat{\mathbf{Z}}^{(k)} \longleftarrow$ Solve Alg. 1 with $\mathbf{W}_1^{(k)}$ and $\mathbf{W}_2^{(k)}$
5: **until** convergence

In Alg. 2, $\epsilon^{(k)}$ can be selected constant [6] or reduced iteratively using an annealing strategy [8]. Here, we set $\epsilon = 0$ and handle infinite weights by a twofold approach. While computing $\widehat{\mathbf{Z}}^{(k)}$ with TF-MxNE, we restrict the source space

to sources with $\|\widehat{\mathbf{Z}}^{(k-1)}[s,:]\|_2 \neq 0$ and set $\widehat{\mathbf{Z}}^{(k)}[s,c] = 0$ if $\mathbf{W}_2^{(k)}[s,c] = \infty$. We found empirically that this procedure does not affect the solution for the problem at hand, but significantly decreases the computation time. The irTF-MxNE solution depends on the initialization of $\mathbf{W}_1$ and $\mathbf{W}_2$, which we set to $\mathbf{W}_1 = \mathbf{W}_2 = \mathbf{1}$. Due to the convexity of the surrogate problems, $\mathbf{Z}$ can be freely selected and we use $\mathbf{Z} = \mathbf{0}$ in the first iteration and warm starts using the previous estimates for subsequent iterations to speed up computation. The first iteration of irTF-MxNE is thus equivalent to solving TF-MxNE. To make the regularization parameter comparable between different data sets, we compute $\lambda_{\max}$, i.e. the lowest value of λ for a given ρ such that $\widehat{\mathbf{Z}} = \mathbf{0}$ if $\lambda \geq \lambda_{\max}$ [16], and normalize λ. We compensate the bias towards superficial sources inherent to source reconstruction approaches penalizing the source amplitude by applying the depth weighting proposed in [17]. To correct the amplitude bias resulting from the inherent shrinkage of ℓ_p-quasinorms with $0 < p \leq 1$, we rescale the estimated source activations using the debiasing procedure proposed in [5].

III. RESULTS

A. Simulation

The simulation setup is based a fixed-orientation source model computed with a three-shell boundary element model using a set of 4699 cortical locations and 102 magnetometers. We simulated trials of auditory evoked fields (AEF) by activating two dipolar sources, one in each transverse temporal gyrus, with Gabor functions peaking at 100 ms and 110 ms (peak amplitude 50 nAm), $\mathbf{X}_{\mathrm{sim}}$, and adding background activity generated by ten randomly chosen sources activated with filtered white noise (peak amplitude 50 nAm). We averaged 30 trials resulting in an SNR ≈ 2.0 after whitening with SNR $= \|\mathbf{M}_{\mathrm{signal}}\|_{\mathrm{Fro}}^2 / \|\mathbf{M}_{\mathrm{noise}}\|_{\mathrm{Fro}}^2$. We applied TF-MxNE and irTF-MxNE with different values of λ, normalized to $\lambda_{\max}$, and ρ, and computed the root mean square error in the source space, RMSE $= \|\mathbf{X}_{\mathrm{sim}} - \mathbf{X}_{\mathrm{est}}\|_{\mathrm{Fro}}$, the F_1-score, $F_1 = (2 \cdot TP)/(2 \cdot TP + FN + FP)$ with TP true positives, FN false negatives, and FP false positives, and the active set size to evaluate the source reconstruction accuracy. The results are presented in Fig. 1. By looking on the F_1-score and active set size, we see that irTF-MxNE outperforms TF-MxNE in terms of active source recovery. While irTF-MxNE is able to identify the true support as indicated by $F_1 = 1.0$ for a range of regularization parameters, the F_1-score for TF-MxNE is $F_1 < 1.0$ for the whole λ-path. The RMSE for both TF-MxNE and irTF-MxNE with $\rho = 0$, which is equivalent to computing MxNE and irMxNE (imposing structured sparsity in the time domain), is significantly higher compared to $\rho > 0$ demonstrating the advantage of imposing structured sparsity in the TF domain for reconstructing smooth nonstationary source signals [5]. Both approaches benefit from the debiasing procedure, which however has a higher effect on TF-MxNE indicating a higher amplitude bias.

B. Experimental MEG data

We evaluate the source reconstruction performance on MEG data recorded during left auditory stimulation with pure tones of 500 Hz [18] using a 306-channel Elekta Neuromag Vectorview system (Elekta Neuromag Oy, Helsinki, Finland).
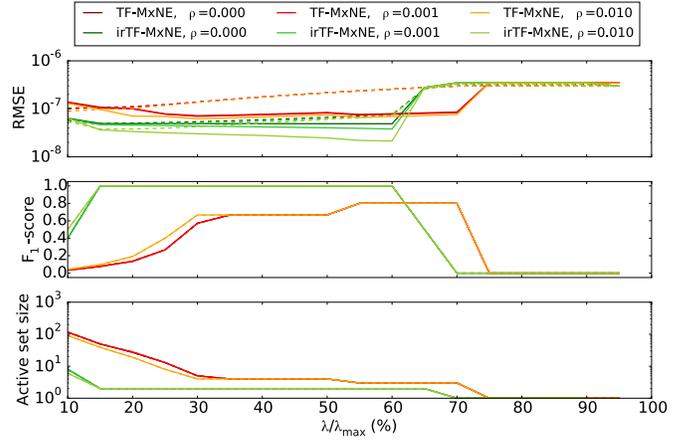

Fig. 1: RMSE, F_1-score, and active set size for TF-MxNE and irTF-MxNE without (dashed) and with (solid) debiasing for simulated AEF data.

Signal preprocessing consisted of signal-space projection for suppressing environmental noise, baseline correction and spatial whitening based on prestimulus data, and threshold-based epoch rejection. The evoked fields were then obtained by averaging the remaining 96 artifact-free epochs. The gain matrix was computed using a set of 7498 cortical locations and a three-shell boundary element model. Source estimation was performed with TF-MxNE and irTF-MxNE (with and without debiasing) on the time interval from 0 ms to 250 ms with fixed orientation constraint. The mean Goodness of Fit (GOF) around the N100m component (from 90 ms to 150 ms) and the active set size as a function of λ ($\rho = 0.01$) are presented in Fig. 2. The results confirm that the debiasing procedure has a significant effect on TF-MxNE due to the amplitude bias. TF-MxNE with debiasing and irTF-MxNE (with and without debiasing) allow for similar mean GOF values. The reconstructed source estimates are however clearly sparser. Fig. 3 shows source reconstruction results for TF-MxNE and irTF-MxNE for selected regularization parameters.
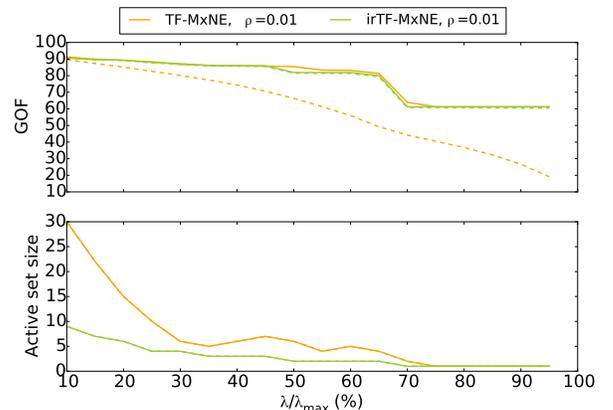

Fig. 2: Mean GOF and active set size for TF-MxNE and irTF-MxNE without (dashed) and with (solid) debiasing for experimental AEF data.
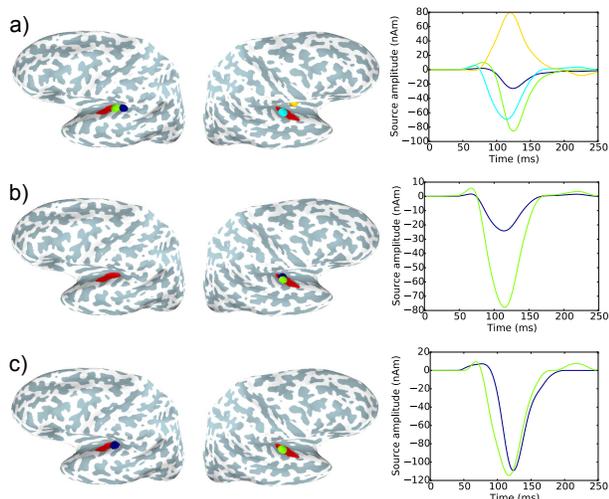
Fig. 3: Source reconstruction using AEF data evoked by left auditory stimulation: (a) TF-MxNE with $\lambda/\lambda_{max} = 55\%$, (b) TF-MxNE with $\lambda/\lambda_{max} = 70\%$, and (c) irTF-MxNE with $\lambda/\lambda_{max} = 55\%$. The dipole locations, marked with spheres, and the corresponding dipole activations are color-coded. The transverse temporal gyri are highlighted in red.

The transverse temporal gyrus (Heschl's gyrus) on each hemisphere is highlighted. We can see that irTF-MxNE with $\lambda/\lambda_{max} = 55\%$ shows activation in both primary auditory cortices represented by single dipoles in both Heschl gyri peaking around 110 ms (Fig. 3c). A similar reconstruction is obtained using TF-MxNE with $\lambda/\lambda_{max} = 55\%$. The activation on both hemispheres is however split into two correlated dipoles, which are partly located outside of the primary auditory cortex (Fig. 3a). Increasing λ to $\lambda/\lambda_{max} = 70\%$ to reduce the number of active dipoles, eliminates both dipoles on the left hemisphere, while the activation on the right hemisphere is still represented by two correlated dipoles (Fig. 3b).

IV. DISCUSSION AND CONCLUSION

In this work, we presented irTF-MxNE, an MEG/EEG inverse solver based on regularized regression with a composite non-convex penalty. We solve the non-convex optimization problem by solving iteratively reweighted convex surrogate optimization problems and presented an algorithm based on block coordinate descent and an active set strategy, which to our experience converges faster than the Fast Iterative Shrinkage-Thresholding algorithm (FISTA) proposed in [5]. Moreover, the BCD scheme is memory-efficient. Extending irTF-MxNE to source reconstruction with a loose orientation constraint or free orientation is not presented here but is straightforward using an additional weighted ℓ_2-norm over orientations [5]. The first iteration of the proposed irTF-MxNE approach is equivalent to computing a standard TF-MxNE solution. Thus, irTF-MxNE can be seen as a postprocessing for TF-MxNE improving source recovery. According to the results obtained on the real MEG data, this postprocessing is to be preferred to increasing λ for TF-MxNE, since the latter approach may eliminate plausible sources prior to erasing spurious activity. We obtained empirical evidence based on simulations and analysis of MEG data, that the source reconstruction is improved by imposing structured sparsity not in the time domain but in the TF domain. Moreover, we showed that the iterative reweighted application of TF-MxNE improves the active source identification and amplitude bias. In conclusion, the proposed inverse solver is a promising tool for M/EEG source analysis.

REFERENCES

[1] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, Feb 2008.

[2] D. Wipf and S. Nagarajan, "A unified Bayesian framework for MEG/EEG source imaging," *NeuroImage*, vol. 44, no. 3, pp. 947–966, Feb 2009.

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

[4] 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, pp. 1937–1961, Apr 2012.

[5] A. Gramfort, D. Strohmeier, J. Haueisen, M. S. Hämäläinen, and M. Kowalski, "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations," *NeuroImage*, vol. 70, pp. 410 – 422, Apr 2013.

[6] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, Dec 2008.

[7] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Commun. Pur. Appl. Math.*, vol. 63, no. 1, pp. 1–38, Oct 2010.

[8] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *IEEE International Conf. on Acoustics, Speech and Signal Processing (ICASSP 2008)*, 2008, pp. 3869–3872.

[9] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar 1997.

[10] A. Rakotomamonjy, "Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms," *Signal Process.*, vol. 91, no. 7, pp. 1505 – 1526, July 2011.

[11] D. Strohmeier, J. Haueisen, and A. Gramfort, "Improved MEG/EEG source localization with reweighted mixed-norms," in *International Workshop on Pattern Recognition in Neuroimaging*, June 2014, pp. 1–4.

[12] D. A. Engemann and A. Gramfort, "Automated model selection in covariance estimation and spatial whitening of MEG and EEG signals," *NeuroImage*, vol. 108, pp. 328–342, Mar 2015.

[13] P. Tseng, "Approximation accuracy, gradient methods, and error bound for structured convex optimization," *Math. Program.*, vol. 125, pp. 263–295, Oct 2010.

[14] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for hierarchical sparse coding," *J. Mach. Learn. Res.*, vol. 12, pp. 2297–2334, 2011.

[15] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Stat. Softw.*, vol. 33, no. 1, pp. 1–22, 2010.

[16] M. Vincent and N. R. Hansen, "Sparse group lasso and high dimensional multinomial classification," *Comput. Stat. Data An.*, vol. 71, pp. 771–786, Mar 2014.

[17] 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, Aug 2008.

[18] M. Weisend, F. Hanlon, R. Montaño, S. Ahlfors, A. Leuthold, D. Pantazis, J. Mosher, A. Georgopoulos, M. S. Hämäläinen, and C. Aine, "Paving the way for cross-site pooling of magnetoencephalography (MEG) data," *Int. Congress Series*, vol. 1300, pp. 615–618, Aug 2007.