Improved MEG/EEG source localization with reweighted mixed-norms

Daniel Strohmeier, Jens Haueisen, Alexandre Gramfort

HAL Id: hal-01044748
https://hal.science/hal-01044748

Submitted on 4 Aug 2014

Improved MEG/EEG source localization with reweighted mixed-norms

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

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

Abstract—MEG/EEG source imaging allows for the non-invasive analysis of brain activity with high temporal and good spatial resolution. As the bioelectromagnetic inverse problem is ill-posed, a priori information is required to find a unique source estimate. For the analysis of evoked brain activity, spatial sparsity of the neuronal activation can be assumed. Due to the convexity, ℓ_1-norm based constraints are often used for this, which however lead to source estimates biased in amplitude and often suboptimal in terms of source selection. As an alternative, non-convex regularization functionals such as ℓ_p-quasinorms with $0 < p < 1$ can be used. In this work, we present a MEG/EEG inverse solver based on a $\ell_{2,0.5}$-quasinorm penalty promoting spatial sparsity as well as temporal stationarity of the brain activity. For solving the resulting non-convex optimization problem, we propose the iterative reweighted Mixed Norm Estimate, which is based on reweighted convex optimization and combines a block coordinate descent scheme and an active set strategy to solve each surrogate problem efficiently. We provide empirical evidence based on simulations and analysis of MEG data that the proposed method outperforms the standard Mixed Norm Estimate in terms of active source identification and amplitude bias.

Keywords—*MEG; EEG; bioelectromagnetic inverse problem; structured sparsity; iterative reweighted optimization algorithm*

I. Introduction

MEG/EEG source imaging delivers insights into the active brain with high temporal and good spatial resolution in a non-invasive way by solving the ill-posed bioelectromagnetic inverse problem. In order to render its solution unique, constraints have to be imposed reflecting assumptions on the neuronal sources. Several non-linear source reconstruction methods were introduced, which have in common to favor source configurations consisting of a small set of focal sources to explain the MEG/EEG data. Among these is regression with ℓ_1-norm regularization, known as LASSO or Minimum Current Estimate in MEG/EEG signal processing. It is a convex surrogate for the optimal, but NP hard regularized regression with a ℓ_0-norm. Due to its convexity, this approach allows for fast algorithms with guaranteed global convergence. However, the resulting source estimates are biased in amplitude and often suboptimal in terms of source selection [1]. Regularized

regression based on logarithmic or ℓ_p-quasinorm penalties with $0 < p < 1$ yields sparser and less biased estimates than the standard LASSO at the expense of convexity [1]. The resulting optimization problem can be solved with e.g. iterative reweighted LASSO, *i.e.*, solving iterative reweighted convex surrogate problems [1], [2].

The aforementioned approaches are applicable to single measurement vector problems. Hence, they cannot account for the temporal characteristics of the brain activity. To overcome this drawback, the Mixed-Norm Estimate (MxNE) was introduced [3]. MxNE is a variant of Group LASSO (GLASSO), *i.e.*, regularized regression with a $\ell_{2,1}$-mixed-norm, adapted for MEG/EEG source reconstruction, in which the neuronal activation at each source location is treated as a separate group. The ℓ_2-norm per group imposes stationarity on the source estimates, while the ℓ_1-norm over groups promotes spatial sparsity. However, GLASSO estimates are suboptimal in terms of amplitude bias and feature selection just as the standard LASSO. Hence, regularized regression using $\ell_{2,p}$-quasinorm penalties with $0 < p < 1$ is proposed, which can be solved by iterative reweighted GLASSO [4].

In this paper, we propose a sparse MEG/EEG source imaging approach based on regularized regression with a $\ell_{2,0.5}$-quasinorm penalty. We solve the non-convex optimization problem by iterative reweighted MxNE. Each MxNE iteration is solved efficiently by combining a block coordinate descent scheme and an active set strategy. The resulting algorithm is applicable to MEG/EEG inverse problems with and without orientation constraint, running in a few seconds on real MEG/EEG problems. We provide empirical evidence using simulations and analysis of MEG data that the proposed method outperforms MxNE in terms of active source identification and amplitude bias.

Notation

We mark vectors with bold letters, $\mathbf{a} \in \mathbb{R}^N$ and matrices with capital bold letters, $\mathbf{A} \in \mathbb{R}^{N \times M}$. The transpose of a vector or matrix is denoted by $\mathbf{a}^T$ and $\mathbf{A}^T$. $\mathbf{a}_i$ identifies the ith element of a vector, $\mathbf{A}_{i,}$ the ith row, and $\mathbf{A}_{.,j}$ the jth column, and $\mathbf{A}_{i,j}$ the element in the ith row and j[th] column of a matrix. $\|\mathbf{A}\|_{\text{Fro}}$ indicates the Frobenius norm, $\|\mathbf{A}\|$ the spectral norm, and $\|\mathbf{A}\|_{p,q}$ the $\ell_{p,q}$-mixed-norm or quasinorm of $\mathbf{A}$ with

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

II. Materials and Methods

A. The inverse problem

The MEG/EEG forward problem describes the linear relationship between the MEG/EEG measurements $\mathbf{M} \in \mathbb{R}^{N \times T}$ (N number of sensors, T number of time instants) and the neuronal activation $\mathbf{X} \in \mathbb{R}^{(P \cdot O) \times T}$ (P number of source locations, O number of dipoles per source location). The model reads:

$$\mathbf{M} = \mathbf{GX} + \mathbf{E} \ , \tag{1}$$

where $\mathbf{G} \in \mathbb{R}^{N \times (P \cdot O)}$ is the gain matrix and $\mathbf{E} \sim \mathcal{N}(0, \lambda \mathbf{I})$ is the measurement noise, which is assumed to be additive, white, and Gaussian. This assumption is acceptable on the basis of a proper spatial whitening of the data.

To find a unique solution to the ill-posed MEG/EEG inverse problem, priors have to be imposed on $\mathbf{X}$, $\mathcal{P}(\mathbf{X}) \sim \exp(-\Omega(\mathbf{X}))$. Here, we set $\Omega(\mathbf{X}) = \|\mathbf{X}\|_{2,0.5}^{0.5}$, i.e., we assume, that only a few focal sources with stationary activation are active during a cognitive task by promoting spatial sparsity and temporal stationarity. The maximum a posteriori estimate is obtained by solving:

$$
\begin{aligned}
\widehat{\mathbf{X}} &= \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{GX}\|_{\text{Fro}}^2 + \lambda\|\mathbf{X}\|_{2,0.5}^{0.5} \\
&= \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{GX}\|_{\text{Fro}}^2 + \lambda\sum_i \sqrt{\|\mathbf{X}_{i,.}\|_2}
\end{aligned}
\tag{2}
$$

Regularization with a $\ell_{2,0.5}$-quasinorm imposes a block row sparsity structure on $\mathbf{X}$, where each block corresponds to the neuronal activation at a specific source location. The ℓ_2-norm per group promotes temporal stationarity of the source estimates, while the $\ell_{0.5}$-quasinorm over groups promotes spatial sparsity. When a free or loose orientation constraint is used, we need to combine the three dipoles belonging to a specific source location by applying an additional ℓ_2-norm penalty over orientations [5]. The resulting $\ell_{2,2,0.5}$-quasinorm is equivalent to a $\ell_{2,0.5}$-quasinorm by reshaping $\mathbf{X}$.

B. Iterative reweighted Mixed Norm Estimate

Based on the framework of Difference of Convex functions programming or Majorization-Minimization algorithms, it was shown, that regularized regression problems with a $\ell_{2,p}$-quasinorm penalty can be solved by iterative reweighted GLASSO [1], [2], [4]. For the MEG/EEG source reconstruction problem in Eq. (2), we hence propose the iterative reweighted MxNE (irMxNE) optimization schemes in Eq. (3).

$$
\begin{aligned}
\widehat{\mathbf{X}}^{k+1} &= \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{GX}\|_{\text{Fro}}^2 + \lambda\sum_i \frac{\|\mathbf{X}_{i,.}\|_2}{2\sqrt{\|\mathbf{X}_{i,.}^k\|_2}} \\
&= \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{GX}\|_{\text{Fro}}^2 + \lambda\sum_i w_i^k\|\mathbf{X}_{i,.}\|_2
\end{aligned}
\tag{3}
$$

For sources with $\|\mathbf{X}_{i,.}^k\|_2 = 0$, the optimization problem in Eq. (3) has an infinite regularization term, which may impair the stability. A smoothing parameter ϵ is thus typically added to avoid weights to become zero [1], [2]. In contrast, we reformulate the weighted MxNE subproblems to apply the weights by scaling the gain matrix as given in Eq. (4). Besides

avoiding an additional smoothing parameter, this reformulation allows to restrict the source space in each MxNE iteration to sources with $\|\mathbf{X}_{i,.}^k\|_2 > 0$ decreasing the computation time.

$$
\begin{aligned}
\widetilde{\mathbf{X}}^{k+1} &= \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{M} - \mathbf{GW}^k\mathbf{X}\|_{\text{Fro}}^2 + \lambda\sum_i \|\mathbf{X}_{i,.}\|_2, \\
\widehat{\mathbf{X}}^{k+1} &= \mathbf{W}^k\widetilde{\mathbf{X}}^{k+1}
\end{aligned}
\tag{4}
$$

with $\mathbf{W}^k$ being a diagonal matrix with $\mathbf{W}_{i,i}^k = 1/w_i^k$.

For solving the weighted MxNE subproblems, we use a block coordinate descent (BCD) scheme [6]. The convergence of each weighted MxNE iteration is controlled by means of the dual gap [3]. In practice, we terminate the BCD procedure, when the current solution $\mathbf{X}^k$ is ϵ-optimal with $\epsilon = 10^{-6}$, i.e., the dual gap $\eta^k < 10^{-6}$. Moreover, we apply an active set strategy [7], which significantly reduces the computation time of the weighted MxNE subproblems. We terminate the irMxNE when $\|\mathbf{X}^k - \mathbf{X}^{k-1}\|_\infty < \tau$ with a user specified tolerance level τ, which we set to 10^{-6} throughout this paper. Due to the non-convexity of the optimization problem in Eq. (2), the results depend on the initialization. For iterative reweighted LASSO, Gasso et al. showed that initialization with $\mathbf{w}^1 = \mathbf{1}$ provides reasonable results [2]. As analyzing the effect of initialization is beyond the scope of this paper, we use the proposed initialization. Consequently, the first iteration of irMxNE is equivalent to solving a standard MxNE problem. Pseudo code for irMxNE with fixed orientation constraint is provided in Algorithm 1.

Algorithm 1 Iterative reweighted MxNE

Require: Measurements $\mathbf{M}$, gain matrix $\mathbf{G}$, regularization parameter $\lambda > 0$, dual gap threshold $\epsilon > 0$, irMxNE stop criterion threshold τ, number of sources S, number of iterations per MxNE iteration I, and the number of MxNE iterations K.

1: Initialization: $\mathbf{X}^0 = \mathbf{0}$, $\mathbf{w}^1 = \mathbf{1}$
2: **for** $k = 1$ to K **do**
3:   $\mathbf{G}^k = \mathbf{GW}^k$ with $\mathbf{W}^k = \text{diag}(\mathbf{w}^k)^{-1}$
4:   **for** $s = 1$ to S **do**
5:     $\mathbf{l}_s = \|(\mathbf{G}_{.,s}^k)^T\mathbf{G}_{.,s}^k\| = \|\mathbf{G}_{.,s}^k\|^2$
6:   **end for**
7:   Initialization: $\mathbf{Y}^0 = \mathbf{0}$, $\mathbf{R} = \mathbf{M}$, $\eta^0 > \epsilon$, $0 \prec \boldsymbol{\mu} \prec \mathbf{l}^{-1}$
8:   **for** $i = 1$ to I **do**
9:     **for** $s = 1$ to S **do**
10:      $\mathbf{Y}_{s,.}^i = \mathbf{Y}_{s,.}^{i-1} + \mu_s(\mathbf{G}_{.,s}^k)^T\mathbf{R}$
11:      $\mathbf{Y}_{s,.}^i = \mathbf{Y}_{s,.}^i \left(1 - \frac{\mu_s\lambda}{\|\mathbf{Y}_{s,.}^i\|_2}\right)^+$
12:      $\mathbf{R} = \mathbf{R} - \mathbf{G}_{.,s}^k\left(\mathbf{Y}_{s,.}^i - \mathbf{Y}_{s,.}^{i-1}\right)$
13:    **end for**
14:    Compute dual gap η^i
15:    **if** $\eta^i < \epsilon$ **then**
16:      break
17:    **end if**
18:   **end for**
19:   $\mathbf{X}^k = \mathbf{W}^k\mathbf{Y}^i$
20:   $\mathbf{w}^{k+1} = 1 \left/ \left(2(\sum_i (\mathbf{X}_{s,i}^k)^2)^{1/4}\right)\right.$
21:   **if** $\|\mathbf{X}^k - \mathbf{X}^{k-1}\|_\infty < \tau$ **then**
22:     break
23:   **end if**
24: **end for**

C. Simulation setup

To provide a reproducible and reasonably fast comparison of MxNE and irMxNE, we generated a simulation data set with 20 sensors and 200 sources with fixed orientation. Five sources were randomly selected to be active each with a random activation (50 samples) drawn from a standard normal distribution. To make the comparison independent of the forward model and the sources' spatial configuration, the columns of the linear forward operator were drawn from a multivariate standard normal distribution and normalized to 1. In a second simulation, we built a correlated gain matrix from a multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma}$ being a Toeplitz matrix generated from a vector $\mathbf{v}$ with $\mathbf{v}_k = 0.95^{k-1}$, $k = 1 : 200$. White Gaussian noise was added to set the signal-to-noise ratio (SNR), which we define here as $\|M_{\text{signal}}\|^2_{\text{Fro}}/\|M_{\text{noise}}\|^2_{\text{Fro}}$. For comparing the success rate of recovering the true active set, we compute the F_1-score according to Eq. (5), where $\mathcal{A}(\mathbf{X})$ is the active set of X with $\mathcal{A}(\mathbf{X}) = \{j : \|\mathbf{X}_{j,.}\|_2 > 0\}$, $|\mathcal{A}(\mathbf{X})|$ the number of elements in $\mathcal{A}(\mathbf{X})$, $\mathbf{X}^*$ the simulated source activation, and $\widehat{\mathbf{X}}$ the source estimate.

$$F_1 = 2 \frac{|\mathcal{A}(\widehat{\mathbf{X}}) \cap \mathcal{A}(\mathbf{X}^*)|}{|\mathcal{A}(\widehat{\mathbf{X}})| + |\mathcal{A}(\mathbf{X}^*)|} \qquad (5)$$

D. Experimental MEG data

For evaluating the performance of MxNE and irMxNE on real MEG data, we use data recorded during left auditory stimulation, which is part of a publicly available MEG/EEG dataset [8]. The MEG data were recorded using a 306-channel Elekta Neuromag Vectorview system (Neuromag Elekta LTD, Helsinki) with a bandpass of 0.1 - 172 Hz and digitized at 600 samples/s. For simplicity, we report results for the magnetometer data. Signal preprocessing consisted of lowpass filtering at 40 Hz, signal-space projection for suppressing environmental noise, baseline correction and spatial whitening based on 200 ms prestimulus data, and averaging. The gain matrix was computed using a set of 7498 cortical locations and a boundary element model. Source estimation was performed using an analysis window from 60 - 120 ms containing the N100m component of the auditory evoked field with fixed orientation constraint and depth compensation ($\rho = 0.8$).

III. RESULTS

A. Simulation study

The F_1-scores for MxNE and irMxNE are presented in Fig. 1 by showing the mean over 100 repetitions. With the uncorrelated gain matrix and SNR = 10, both methods reach high F_1-scores, but only irMxNE allows for F_1-score = 1, i.e., an exact reconstruction of the active set in all repetitions. This F_1-score is achieved for a range of regularization parameters, which facilitates the choice of λ. For SNR = 2, the maximum F_1-score for MxNE is reduced, whereas irMxNE is still able to allow for an exact reconstruction. However, the range of regularization parameters, for which irMxNE reaches F_1-score ≈ 1, is smaller compared to SNR = 10. With the correlated gain matrix and SNR = 10, irMxNE reaches a F_1-score > 0.8, which is significantly higher than the F_1-score for MxNE. For SNR = 2, the F_1-scores for both approaches are clearly reduced (F_1-scores < 0.4). However, the F_1-score for irMxNE is approximately 3 times higher.
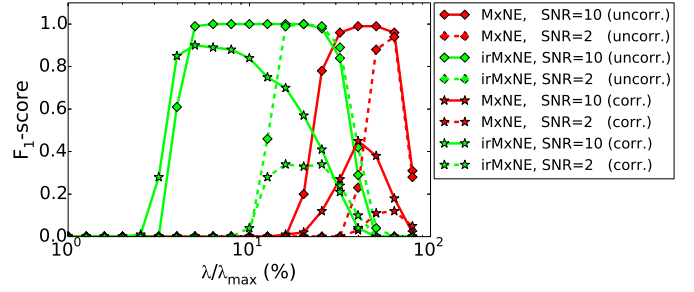

Fig. 1: F_1-score as a function of the regularization parameter λ for MxNE (red) and irMxNE (green) using simulated data with SNR = 10 (solid) and SNR = 2 (dashed). We present the mean over 100 repetitions for an uncorrelated (diamond) and a correlated gain matrix (star).

In summary, irMxNE outperforms MxNE in terms of maximum F_1-score in all simulations. MxNE requires high regularization parameters to reach its maximum F_1-score, which increases the amplitude bias.

B. Experimental MEG data

Based on the real MEG data, we analyzed the convergence speed of a single weighted MxNE iteration. We compare the applied BCD scheme with the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [9], an iterative proximal gradient method. We can see from Fig. 2, that the BCD scheme outperforms FISTA in terms of computation time. Applying the active set strategy reduced computation time for MxNE by a factor of 100. This allows to compute irMxNE on real MEG/EEG source localization problems in a few seconds.
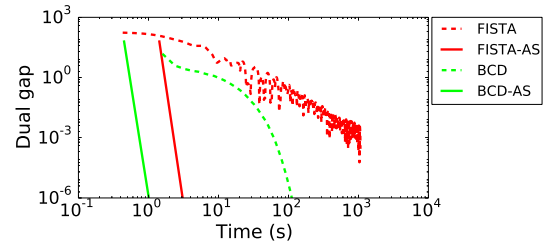

Fig. 2: Dual gap for MxNE with loose orientation constraint on real MEG data using BCD and FISTA without (dashed) and with active set strategy (solid).

Subsequently, we applied both MxNE and irMxNE on the real MEG data. Fig. 3a shows the active set as a function of the regularization parameter λ, which is given as a percentage of λ_{max} (see [3] for details). The corresponding explained variances with and without debiasing are illustrated in Fig. 3b. Debiasing is performed here by estimating a scaling factor for each active source as proposed in [5]. Source reconstruction with irMxNE yields sparser source estimates for all regularization parameters. Moreover, the number of active sources reconstructed with irMxNE is stable for a wide range of regularization parameters, whereas the size of the active set obtained with MxNE is more dependent on λ. The explained variance for MxNE without debiasing is highly dependent on
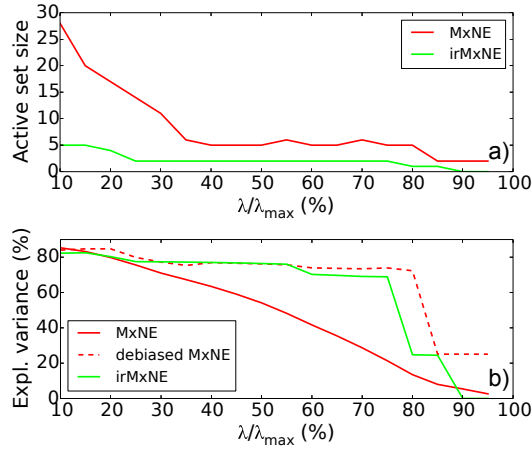
Fig. 3: Active set size and explained variance before (dashed) and after debiasing (solid) obtained with MxNE and irMxNE for the N100m component after left auditory stimulation.

the regularization parameter, whereas the explained variance for irMxNE is more stable and significantly higher for higher λ. Compared to MxNE with debiasing, irMxNE yields a comparable explained variance both in terms of amount and stability with the advantage of providing a sparser source model. The results indicate, that the amplitude bias is reduced using irMxNE. Fig. 4 illustrates reconstructed source locations and activations for selected regularization parameters. MxNE with $\lambda/\lambda_{max} = 40\%$ shows activation in both auditory cortices each represented by several highly correlated dipoles. Increasing λ eliminates the sources on the right hemisphere, whereas two highly correlated dipoles are reconstructed on the left hemisphere explaining only 30% of the total variance. In contrast, irMxNE with $\lambda/\lambda_{max} = 25\%$ reconstructs single dipoles in both auditory cortices, which explain approximately 80% of the total variance.

IV. DISCUSSION AND CONCLUSION

In this work, we presented irMxNE, a MEG/EEG inverse solver based on regularized regression with a $\ell_{2,0.5}$-quasinorm penalty. The resulting non-convex optimization problem is solved by iteratively solving convex surrogate optimization problems, which allows for fast algorithms and convergence control at each iteration. The combination of a block coordinate descent scheme and an active set approach significantly decreases the computation time making the proposed method run in a few seconds on real size MEG/EEG source reconstruction problems. Due to the non-convexity, the irMxNE solution depends on the initialization. Analyzing the convergence of iterative reweighted GLASSO algorithms is an open research question [4], which is beyond the scope of this contribution. By choosing the initialization such that the first iteration is equivalent to computing a standard MxNE, the irMxNE source estimate is at least as sparse as the MxNE solution. Moreover, we obtained empirical evidence based on simulations and analysis of MEG data, that irMxNE outperforms MxNE in terms of active source identification and amplitude bias. In conclusion, the presented inverse solver is a promising approach for M/EEG source analysis.

(a) MxNE: $\lambda/\lambda_{max} = 40\%$

(b) MxNE: $\lambda/\lambda_{max} = 85\%$

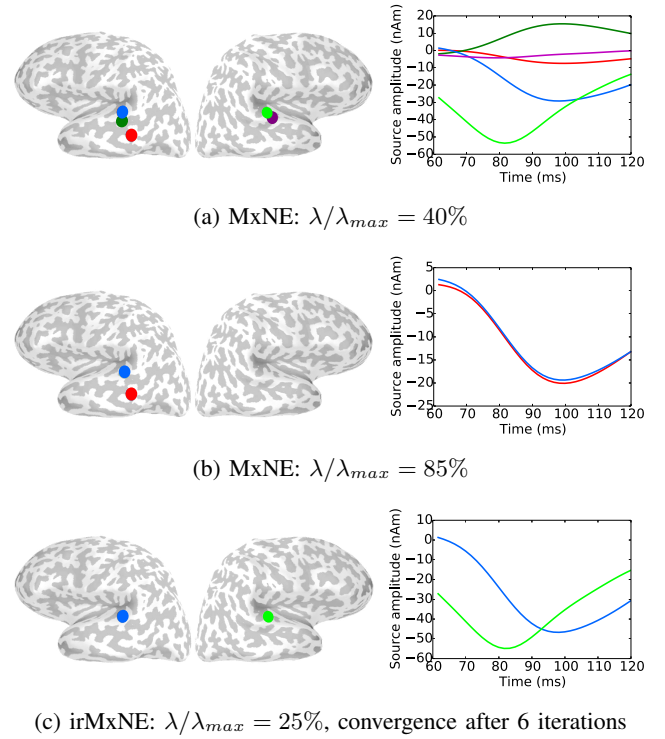(c) irMxNE: $\lambda/\lambda_{max} = 25\%$, convergence after 6 iterations

Fig. 4: Comparison of MxNE and irMxNE using real MEG data. The images show the estimated source location and activation for MxNE with debiasing (a, b) and irMxNE (c) for the N100m component after left auditory stimulation.

REFERENCES

[1] 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.

[2] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and dc programming," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.

[3] 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.

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

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

[7] V. Roth and B. Fischer, "The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms," in *Proc. of the 25th Int. Conf. on Machine learning (ICML '08)*, Helsinki, Finland, July 2008, pp. 848–855.

[8] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446 – 460, Feb. 2014.

[9] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, Jan. 2009.