Rank estimation and tensor decomposition using physics-driven constraints for brain source localization

Nasrin Taheri[1,2], Amar Kachenoura[1], Ahmad Karfoul[1], Xu Han[1], Karim Ansari-Asl[2], Isabelle Merlet[1]
Lotfi Senhadji[1], and Laurent Albera[1]
[1]Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France
[2]Department of Electrical Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran

***Abstract*—This paper deals with the tensor-based Brain Source Imaging (BSI) problem, say finding the precise location of distributed sources of interest by means of tensor decomposition. This requires to estimate accurately the rank of the considered tensor to be decomposed. Therefore, a two-step approach, named R-CPD-SISSY, is proposed including a rank estimation process and a source localization procedure. The first step consists in using a modified version of a recent method, which estimates both the rank and the loading matrices of a tensor following the canonical polyadic decomposition model. The second step uses a recent physics-driven tensor-based BSI method, named STS-SISSY, in order to localize the brain regions of interest. This second step uses the estimated rank during the first step. The performance of the R-CPD-SISSY algorithm is studied using realistic synthetic interictal epileptic recordings.**

I. Introduction

Brain Source Imaging (BSI) aims at both reconstructing the brain electrical activity from scalp ElectroEncephaloGraphy (EEG) recordings and identifying the position of the active source regions. In pre-surgical evaluations of patients suffering from drug-resistant partial epilepsy, brain source imaging techniques help to delineate the epileptogenic zones to be resected in order to control the epileptic seizures. To this end, several tensor-based approaches have been presented over the last decades [1]–[4]. In classical approaches, first the tensor is decomposed in order to identify the spatial signature of the data and the location of distributed sources is derived. These approaches suffer from several drawbacks including inaccuracy, difficulties in identifying close and correlated sources, and high computational complexity. The accuracy of these two-step source localization methods highly depends on the effieciency of the first step and more particularly on that of the rank estimation procedure. Indeed, in the case of epileptic EEG recordings, as the number of epileptic patches is unknown, this value should be estimated.

Unfortunately, estimating the rank of a given tensor is usually hard and in some cases NP hard. This problem has been adressed in the literature [5]–[7]. The R-CPD method [8] [9] appears to be more robust with respect to the presence of noise than the other approaches. Another way of improving the source localization process consists in combining the two steps of classical tensor-based BSI methods leading to the STS-SISSY algorithm [10].

In this paper, we propose to improve the performance of the R-CPD method by using the Minimum Description Length (MDL) approach [11]–[13]. Then, we combine this modified R-CPD technique with STS-SISSY in order to design a new two-step tensor-based BSI appraoch named R-CPD-SISSY. Computer results show the good behavior of R-CPD-SISSY especially for low SNR values.

II. Problem Formulation

We consider $N \times T$ scalp EEG recordings generated by D brain current dipoles forming the source space. The electrical activity is denoted by matrix $\tilde{\mathbf{S}} \in \mathbb{R}^{D \times T}$. The EEG recordings are the superposition of these dipole signals: $\mathbf{X} = \mathbf{G}\tilde{\mathbf{S}}$ where the lead-field matrix $\mathbf{G} \in \mathbb{R}^{N \times D}$ characterizes the propagation of the activity of each dipole within the head volume. Considering that the EEG recordings are mainly generated by R distributed sources and refering the average signal $\boldsymbol{s}_r$ to the r-th source, data can be modeled as [14]:

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{h}_r \mathbf{s}_r^T + \mathbf{X}_b \tag{1}$$

where the lead-field vector $\mathbf{h}_r = \mathbf{G}\boldsymbol{\psi}_r$ is the spatial mixing vector of the r-th distributed source for which $\mathbf{s}_r$ denotes the average activity of its dipoles, and where $\boldsymbol{\psi}_r$ indicates the sparse coefficient vector whose nonzero elements describe the contribution of the associated grid dipoles to the distributed source. The signals $\mathbf{X}_b$ characterize the background activity emitted by dipoles of non-distributed sources. In this paper, different intensities of noisy background are added to the simulated epileptic data to obtain noisy simulated EEGs with different SNR values. The BSI problem consists in identifying the $D \times R$ sparse matrix $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_R]$.

III. Methods

In order to build a tensor from matrix $\mathbf{X}$, an additional dimension from the measurements is collected. As EEG data contain repeated interictal epileptic spikes during the time, it is possible to stack the spike-like signals observed at P different

978-9-0827-9703-9/19/$31.00

time samples along the third dimension of the tensor $\boldsymbol{\mathcal{X}}$ and to construct a Space-Time-Spike (STS) tensor:

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} \mathbf{h}_r \circ \mathbf{s}_r \circ \mathbf{c}_r \tag{2}$$

where $\mathbf{c}_r \in \mathbb{R}^K$ represents the spike amplitudes of the different realizations. This model is the Canonical Polyadic Decomposition (CPD) and the matrices $\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_R]$, $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_R]$ and $\mathbf{C} = [\mathbf{c}_1, ..., \mathbf{c}_R]$ denote the loading matrices of the space-time-spike tensor $\boldsymbol{\mathcal{X}}$. The two steps of the R-CPD-SISSY are presented below.

A. Rank estimation using MDL

The R-CPD method [8] extracts the loading matrices of $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{C}$ by solving the following minimization problem:

$$\min_{\mathbf{H},\mathbf{S},\mathbf{C}} \ \|\mathbf{H}\|_{2,1} + \|\mathbf{S}\|_{2,1} + \|\mathbf{C}\|_{2,1} + \|\mathbf{H}\|_{1,2} + \|\mathbf{S}\|_{1,2} + \|\mathbf{C}\|_{1,2}$$
$$s.t. \qquad \boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \mathbf{h}_r \circ \mathbf{s}_r \circ \mathbf{c}_r \tag{3}$$

The notations $\|.\|_{2,1}$ and $\|.\|_{1,2}$ are the mixed-norms $L_{2,1}$ and $L_{1,2}$. In [5], it is proved that minimizing the mixed-norms led to minimizing the nuclear norm. The nuclear norm is considered as a good convex envelope of the matrix rank. For a given matrix $\mathbf{X}$ $(I \times J)$ the mixed norms $\|.\|_{2,1}$ and $\|.\|_{1,2}$ are computed as $Tr[\mathbf{X}^T\mathbf{A}\mathbf{X}]$ and $Tr[\mathbf{X}\mathbf{B}\mathbf{X}^T]$, respectively, where $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices given by $A_{ii} = 1/\sqrt{\sum_{j=1}^{J} X_{ij}^2}$ and $B_{jj} = 1/\sqrt{\sum_{i=1}^{I} {X_{ij}}^2}$. The (i,j)-th entry of $\mathbf{X}$ is denoted as X_{ij}. The minimization problem (3) can be solved by minimizing the following Augmented Lagrangian function:

$$\begin{aligned}\boldsymbol{\mathcal{L}} = \lambda_{1,2}\{\ &\mathbf{Tr}(\mathbf{H}\mathbf{B}^{(1)}\mathbf{H}^T) + \mathbf{Tr}(\mathbf{S}\mathbf{B}^{(2)}\mathbf{S}^T) + \mathbf{Tr}(\mathbf{C}\mathbf{B}^{(3)}\mathbf{C}^T)\} \\ +\lambda_{2,1}\{\ &\mathbf{Tr}(\mathbf{H}^T\mathbf{A}^{(1)}\mathbf{H}) + \mathbf{Tr}(\mathbf{S}^T\mathbf{A}^{(2)}\mathbf{S}) + \mathbf{Tr}(\mathbf{C}^T\mathbf{A}^{(3)}\mathbf{C})\} \\ + < \boldsymbol{\mathcal{Y}}, \boldsymbol{\mathcal{X}} - &\sum_{r=1}^{\hat{R}} \mathbf{h}_r \circ \mathbf{s}_r \circ \mathbf{c}_r > + \frac{\rho}{2}\|\boldsymbol{\mathcal{X}} - \sum_{r=1}^{\hat{R}} \mathbf{h}_r \circ \mathbf{s}_r \circ \mathbf{c}_r\|_F^2\end{aligned} \tag{4}$$

where $\lambda_{1,2}$, $\lambda_{2,1}$ and ρ are penalty parameters and $\boldsymbol{\mathcal{Y}}$ denotes the tensor multiplier. The penalty parameters are used to manage a trade-off between data-fit and prior knowledge. They also depend on the noise level, since the difference between recovered and estimated data is expected to be larger as the SNR value decreases. The update functions of loading matrices are calculated by vanishing the derivatives of the $\boldsymbol{\mathcal{L}}$ with respect to the $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{C}$. The loading matrices and the tenor multiplier are updated as following:

$$\begin{aligned} vec(\mathbf{H}) \ = \ &[\ \mathbf{I} \otimes (\ 2\lambda_{2,1}\, \mathbf{A}^{(1)}\ ) \ + \ \ 2\lambda_{1,2}\, \mathbf{B}^{(1)} \\ &+ \ \rho\, (\mathbf{C} \odot \mathbf{S})^T (\mathbf{C} \odot \mathbf{S}) \ \otimes \ \mathbf{I}\ ]^{-1} \\ &vec[\ (\mathbf{Y}^{(1)} \ + \ \rho \mathbf{X}^{(1)})(\ \mathbf{C} \ \odot \ \mathbf{S}\ )\ ] \end{aligned} \tag{5}$$

$$\begin{aligned} vec(\mathbf{S}) \ = \ &[\ \mathbf{I} \otimes (\ 2\lambda_{2,1}\, \mathbf{A}^{(2)}\ ) \ + \ \ 2\lambda_{1,2}\, \mathbf{B}^{(2)} \\ &+ \ \rho\, (\mathbf{C} \odot \mathbf{H})^T (\mathbf{C} \odot \mathbf{H}) \ \otimes \ \mathbf{I}\ ]^{-1} \\ &vec[\ (\mathbf{Y}^{(2)} \ + \ \rho \mathbf{X}^{(2)})(\ \mathbf{C} \ \odot \ \mathbf{H}\ )\ ] \end{aligned} \tag{6}$$

$$\begin{aligned} vec(\mathbf{C}) \ = \ &[\ \mathbf{I} \otimes (\ 2\lambda_{2,1}\, \mathbf{A}^{(3)}\ ) \ + \ \ 2\lambda_{1,2}\, \mathbf{B}^{(3)} \\ &+ \ \rho\, (\mathbf{S} \odot \mathbf{H})^T (\mathbf{S} \odot \mathbf{H}) \ \otimes \ \mathbf{I}\ ]^{-1} \\ &vec[\ (\mathbf{Y}^{(3)} \ + \ \rho \mathbf{X}^{(3)})(\ \mathbf{C} \ \odot \ \mathbf{H}\ )\ ] \end{aligned} \tag{7}$$

$$\boldsymbol{\mathcal{Y}}^{k+1} \ = \ \boldsymbol{\mathcal{Y}}^{k} \ + \ \rho_k\ [\boldsymbol{\mathcal{X}} - (\sum_{r=1}^{R} \mathbf{h}_r \circ \mathbf{s}_r \circ \mathbf{c}_r)^{k+1}] \tag{8}$$

where $\mathbf{X}^{(i)}$ and $\mathbf{Y}^{(i)}$ denote the $i-th$ unfolding matrices of the tensors $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$, respectively. In (8), k represents the iteration number. The parameter ρ in $(k+1)-th$ iteration is updated as $\rho_{k+1} \ = \ \mu\rho_k$ for $\mu > 1$.

Using the modified MDL (iMDL) [15], the rank of the third-order tensor $\boldsymbol{\mathcal{X}}$ can be estimated. First, the singular values σ_i of matrix $\mathbf{H}$ are sorted in descending order, i.e $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_R$. Then, the R_{est} can be calculated as follows:

$$\begin{aligned} R_{est} = arg\min_{n} -2log\left(\frac{\prod_{i=1+n}^{N} {\sigma_i}^{1/(N-n)}}{\frac{1}{N-n}\sum_{i=1+n}^{N} \sigma_i}\right)^{\hat{R}(N-n)} \\ +r(2N-n)log(\hat{R}) \end{aligned} \tag{9}$$

where σ_i represents the i-th highest singular value and $(N \times \hat{R})$ is the size of the estimated matrix $\hat{\mathbf{H}}$ and where $\hat{R}$ denotes the Over-Estimated(O-E) rank. Using this method, R_{est} is computed by finding the break point of singular value curve of matrix $\hat{\mathbf{H}}$.

B. Source localization using STS-SISSY

The single-step STS-SISSY algorithm [10] utilizes the fused Lasso regularization to solve the BSI problem [14]:

$$\min_{\mathbf{H},\boldsymbol{\Psi}} \quad ||\mathbf{X}^{(1)} - \mathbf{H}(\mathbf{C} \odot \mathbf{S})^\top|| + \lambda\ (\|\mathbf{T}\boldsymbol{\Psi}\|_1 + \alpha\|\boldsymbol{\Psi}\|_1)$$
$$s.t. \quad \mathbf{H} = \mathbf{G}\boldsymbol{\Psi} \tag{10}$$

where $\mathbf{X}^{(1)} \in \mathbb{R}^{N \times TR}$ is the mode-1 unfolding matrix of the tensor $\boldsymbol{\mathcal{X}}$, $\mathbf{T}$ denotes the total variational operator, λ and α stand respectively for the regularization and penalty parameters. This problem was solved using Alternating Direction Method of Multipliers (ADMM) [16] wherein all matrices were updated at each iteration in an alternative way until convergence or a maximal number of iterations are reached. At the end of the ALS algorithm [17], in addition to the matrices $\mathbf{H}$, $\mathbf{S}$, and $\mathbf{C}$, estimates of all distributed sources are directly available and correspond to the columns of the matrix (for more detailes see in [10]). Despite the efficiency of this method, it needs the true rank to be efficient. Therefore, we use STS-SISSY with the estimated rank of the previous step.

IV. Data and Performance Criteria

The performance of R-CPD-SISSY algorithm is assessed on realistic simulated interictal epileptic EEG data and compared to STS-SISSY method in term of source localization accuracy. To this end, scalp epileptic EEG for $N = 91$ electrodes, $T = 180$ time instants at the sampling rate of 256 Hz, and $K = 50$ epileptiform spikes are generated. The source space is composed of 19626 dipoles located on the cortical surface and the lead-field matrix is generated using the ASA software. The epileptic source regions are modeled by two epileptic patches: one located on the Inferior Frontal gyrus (InfFr) and another patche on the Occipital-Temporal gyrus (OccTe). Each patch includes 100 grid dipoles corresponding to a cortical area of approximately 5 cm^2. To study the performance of the proposed method, 50 realizations of the data are generated. The performance of R-CPD-SYSSY is evaluated in term of rank estimation and source localization accuracy: i) two performance criteria, namely the Accuracy Rate (AR) and the Average Rank Estimation Error (AREE) are used to evaluate the rank estimation results, and ii) the Dipole Localization Error (DLE) [1] is employed to calculate the accuracy of the source localization.

A. Performance criteria for rank estimation

These criteria are utilised to assess the rank estimation performance. The AR computes the number of times the estimation is done exactly. This criterion is calculated as:

$$AR \ : \ Times \ of \ R \ = \ R_{est} \tag{11}$$

Regarding the AREE criterion, it measures the error between the exact rank and the estimated one:

$$AREE \ : \ \frac{1}{nb \, of \, realizations} \sum_{times=1}^{nb \, of \, realizations} |R - R_{est}| \tag{12}$$

B. Performance criterion for source localization

To quantify the efficiency of different source localization methods, the accuracy of the approaches are computed in term of Dipole Localization Error (DLE) [1], which presents a measure of similarity between the original and the estimated source configuration. The DLE is defined as:

$$DLE = \frac{1}{2L} \sum_{m \in \mathcal{M}} \min_{l \in \hat{\mathcal{M}}} \|\mathbf{p}_m - \mathbf{p}_l\|_2 + \frac{1}{2\hat{L}} \sum_{l \in \hat{\mathcal{M}}} \min_{m \in \mathcal{M}} \|\mathbf{p}_m - \mathbf{p}_l\|_2$$

where $\mathcal{M}$ and $\hat{\mathcal{M}}$ denote the original and estimated sets of indices of all dipoles of an active patch, respectively. L and $\hat{L}$ are the numbers of original and estimated active dipoles. $\mathbf{p}_m$ is the position of the m-th source dipole. The DLE is averaged over the 50 realizations.

V. Simulations

To assess the efficiency of the proposed method, different intensities of background EEG are added to the simulated epileptic data with SNR = [-15, -10, -5, 0, 5, 10] dB. We first evaluate the performance of R-CPD-SISSY, in term of rank estimation by considering different Over-Estimated (O-E) ranks including 3, 8 and 20. To do so, the regularization parameters in (4), $\lambda_{1,2}$, $\lambda_{2,1}$ and ρ are tuned. To balance the weights of L_{12} and L_{21}, we take $\lambda_{1,2} = \frac{\|H\|_{2,1}}{\|H\|_{1,2}} \lambda_{2,1}$. Then to get more accurate results, a large range of values of between 1 to 2000 are given to $\lambda_{2,1}$ to choose the one which provides higher AR and less AREE for each values of SNR. Fig.1 and Fig.2 provide the values of AR and AREE as a function of SNR for the different tested (O-E) ranks, respectively. In our experiment, the true rank R (in (11), and (12)) is 2 related to the two epileptic sources.

As can be seen in Fig.1, when the (O-E) rank is 3, the estimated rank, by the step one of the R-CPD-SISSY, is equal to the true rank for all 50 realizations and all values of SNR. Consequently, the corresponding AREE in Fig.2 is zero. In the cases of (O-E) ranks equal to 8 and 20, the proposed algorithm seems to be very efficient, except for very low SNR=-15 dB. We can also see in Fig.2 that the AREE values are small, for all SNRs. This shows that, even if the rank was not estimated exactly, the obtained value is close to the true rank. Secondly, to evaluate the accuracy of source localization, the DLE of R-CPD-SISSY is computed and compared to the that of STS-SISSY, for different values of SNR and (O-E) ranks. Note that, for R-CPD-SISSY the estimated rank is given to the source localization step while the (O-E) rank is given directly to STS-SISSY method. The DLE results of both approaches are presented in Fig. 3 as function of SNR for different (O-E) ranks. As can be seen, the R-CPD-SISSY clearly outperform STS-SISSY. Indeed, the DLE values for R-CPD-SISSY approach are significantly less than those of STS-SISSY, especially for low SNR and larger (O-E) ranks.

This can be also confirmed in Fig. 4, where we illustrate an example of the BSI results of the both methods for SNR=- 5 dB and (O-E) rank=20. Compared to the ground truth (top of the figure), the two patches are better localized using R-CPD-SISSY, even if in STS-SISSY two best factors are manually chosen among the 20 ones, to reach the best localization results.

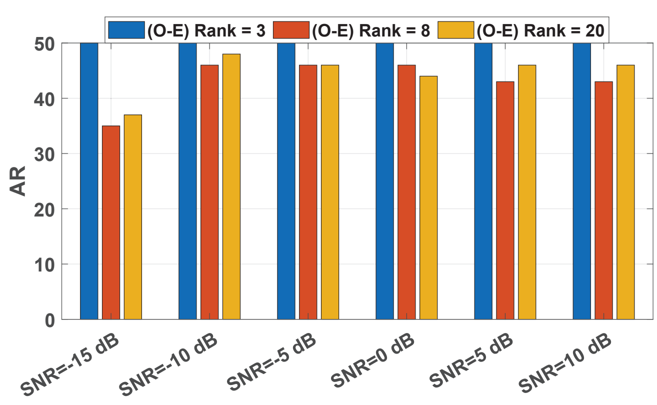

Fig. 1. AR values for different SNR values and different (O-E) Ranks.

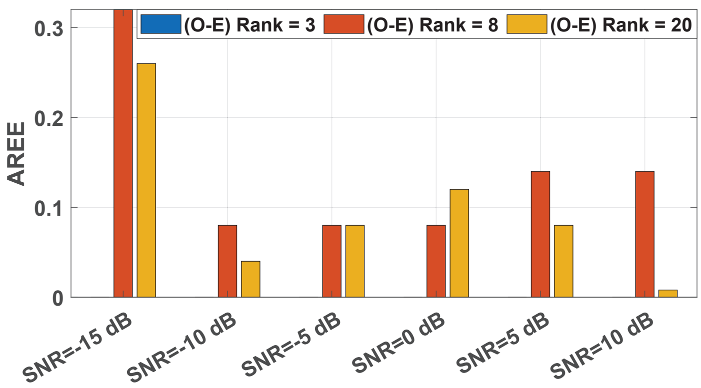

Fig. 2. AREE for different SNR values and different (O-E) Ranks.

VI. CONCLUSION

A two-step tensor-based approach, was designed to localize the extended epileptic sources precisely, with estimating the number of sources. To do so, we estimate the rank in the first step through a tensor decomposition approach with considering the group sparsity constraints (R-CPD) followed by MDL. Then, the estimated rank is used in the second step to localize the sources efficiently. Simulations on realistic epileptic EEG data demonstrated the superior performance of R-CPD-SISSY for recovery of distributed sources by giving a robust estimated rank to the source localization procedure. In other words, contrary to the state-of-the-art BSI methods, R-CPD-SISSY can solve the BSI problem without any knowledge of the expected number of sources. Forthcoming work will include the automatic estimation of the penalty parameters by exploiting, for example, the balance between the data-fit terms and the constraint ones in the cost function.

REFERENCES

[1] H. Becker, L. Albera, P. Comon, J.-C. Nunes, R. Gribonval, J. Fleureau, P. Guillotel, and I. Merlet. Sissy: An efficient and automatic algorithm for the analysis of eeg sources based on structured sparsity. *Neuroimage*, 157:157–172, 2017.

[2] B. Hunyadi, D. Camps, L. Sorber, W. Van Paesschen, M. De Vos, S. Van Huffel, and L. De Lathauwer. Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing*, 2014(1):139, 2014.

[3] M. Morup, L. K. Hansen, C. S. Herrmann, J. Parnas, and S. M. Arnfred. Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG. *NeuroImage*, 29:938–947, 2006.

[4] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener. Multiway analysis of epilepsy tensors. *Bioinforcs*, 23(13):i10–i18, 2007.

[5] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.

[6] C. J. Hillar and L. Lim. Most tensor problems are np-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

[7] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

[8] X. Han, L. Albera, A. Kachenoura, L. Senhadji, and H. Shu. Low rank canonical polyadic decomposition of tensors based on group sparsity. In *25th European Signal Processing Conference (EUSIPCO)*, pages 668–672, Kos, Greece, 2017.

[9] X. Han. *Robust low-rank tensor approximations using group sparsity*. PhD thesis, l'université de Rennes 1, 2019.

[10] H. Becker, A. Karfoul, L. Albera, R. Gribonval, Julien Fleureau, P. Guillotel, A. Kachenoura, L. Senhadji, and I. Merlet. Tensor decomposition exploiting structural constraints for brain source imaging. pages 181–184, Cancun, Mexico, 2015.

[11] G. Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[12] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[13] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2):387–392, 1985.

[14] L. Albera, H. Becker, A. Karfoul, R. Gribonval, A. Kachenoura, L. Senhadji, A. Hernandez, and I. Merlet. Localization of spatially distributed brain sources after a tensor-based preprocessing of interictal epileptic eeg data. *to appear in Proc. 37th International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

[15] D. Guimarães and R. de Souza. Simple and efficient algorithm for improving the mdl estimator of the number of sources. *Sensors*, 14(10):19477–19492, 2014.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1 – 122, 2010.

[17] R. Bro. *Multi-way analysis in the food industry: Models, algorithms and applications*. PhD thesis, University of Amsterdam (NL), 1998.

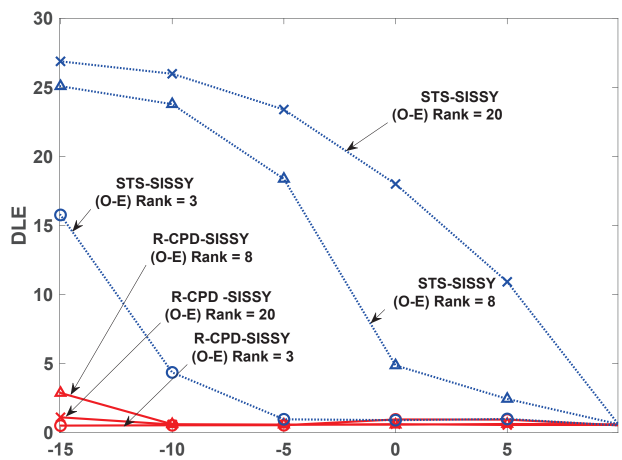

Fig. 3. DLE values for two approaches as a function of SNR for (O-E) ranks.

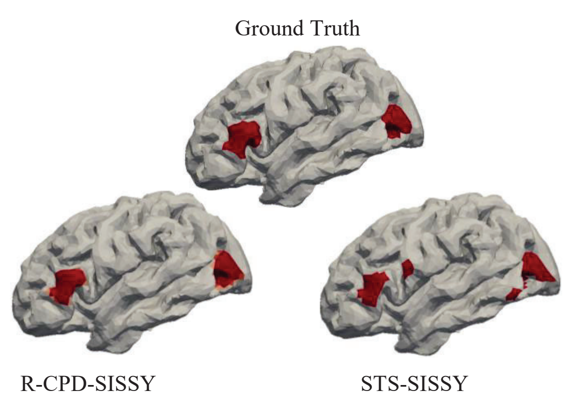

Fig. 4. Example of source localization result of R-CPD-SISSY and STS-SISSY: (O-E) rank = 20 and SNR = -5dB.