Scalable and Robust Tensor Decomposition of Spontaneous Stereotactic EEG Data

Jian Li, *Student Member, IEEE*, Justin P. Haldar, *Senior Member, IEEE*, John C. Mosher, *Senior Member*, *IEEE*, Dileep R. Nair, Jorge A. Gonzalez-Martinez, Richard M. Leahy, *Fellow, IEEE*

Abstract—*Objective*: Identification of networks from resting brain signals is an important step in understanding the dynamics of spontaneous brain activity. We approach this problem using a tensor-based model. *Methods:* We develop a rank-recursive Scalable and Robust Sequential Canonical Polyadic Decomposition (SRSCPD) framework to decompose a tensor into several rank-1 components. Robustness and scalability are achieved using a warm start for each rank based on the results from the previous rank. *Results:* In simulations we show that SRSCPD consistently outperforms the multi-start alternating least square (ALS) algorithm over a range of ranks and signal-to-noise ratios (SNRs), with lower computation cost. When applying SRSCPD to resting in-vivo stereotactic EEG (SEEG) data from two subjects with epilepsy, we found components corresponding to default mode and motor networks in both subjects. These components were also highly consistent within subject between two sessions recorded several hours apart. Similar components were not obtained using the conventional ALS algorithm. *Conclusion:* Consistent brain networks and their dynamic behaviors were identified from resting SEEG data using SRSCPD. *Significance:* SRSCPD is scalable to large datasets and therefore a promising tool for identification of brain networks in long recordings from single subjects.

Index Terms—Tensor decomposition, dynamic functional connectivity, stereotactic EEG, optimization

I. INTRODUCTION

EXPLORING functional connectivity (FC) in resting brain signals is a rich approach to studying brain networks [1]. Of particular recent interest is the dynamic nature of FC [2]. The most commonly used strategy for decoding dynamic FC (DFC) is to compute correlation or coherence using a sliding window [3], [4]. However, using a long temporal window to obtain robust FC estimates inevitably leads to over-smoothing of dynamic changes [5], [6]. To overcome this difficulty, principal component analysis (PCA)-based and independent component analysis (ICA)-based approaches have been proposed. Although they do not introduce temporal smoothing, a limitation of those methods is that either the time series of each network is required to be independent (temporal ICA) [7] or the spatial modes of the networks are disjoint (spatial ICA) [8], [9] or orthogonal (PCA) [10], whereas real

networks can overlap and be correlated in both space and time [11].

Tensors are a generalization of matrices in which we can represent data with more than two indices. For example, here we use a third-order tensor to represent SEEG data in terms of space, time and frequency (in the following "tensors" refers to tensors of order 3 or higher). As with matrices, tensors can be represented as a sum of rank-1 components. Structured data often then admit to low rank models with respect to this tensor representation. One of the two popular tensor models is the canonical polyadic (CP) form [12], which has been shown [13]–[15] to also be equivalent to parallel factors analysis (PARAFAC) [16] and the canonical decomposition (CANDECOMP) [17].

CP is a model that can capture structure inherent in multidimensional data when represented in a low rank tensor of order 3 or higher. Conversely, standard 2D matrix decomposition methods, such as PCA or ICA, applied to matricized or unfolded tensors [15] are typically not able to capture this structure using a similar low-rank model. Moreover, CP decomposition has a unique solution under less restrictive conditions than the orthogonality or independence assumptions implicit in PCA or ICA [18], [19]. This latter property is particularly appealing when analyzing SEEG or other brain data that can be represented as a third or higher-order tensor, since we can avoid the restrictive assumptions of orthogonality or independence between components.

Among the many algorithms for computing the CP decomposition, alternating least squares (ALS), [16][17] is the most widely used because of its simplicity relative to alternatives [14], [15], [20], [21]. Comparisons in these papers show that, in general, ALS provides solutions of similar quality to other algorithms. While some are more robust to over-factoring than ALS, especially in ill-conditioned cases, the expense is higher computational complexity in both memory and time [15].

ALS-based CP decomposition has been widely used in EEG analysis by transforming the raw EEG recordings to a time-frequency representation using short-time Fourier or Morlet wavelet transforms and applying a 3-way CP decomposition (channel by time by frequency) [22]. Möck [23] applied CP to event-related potentials (ERP). Miwakeichi *et al.* [24] analyzed both spontaneous and evoked EEG recordings and showed that theta activity was predominant during a task condition, while alpha activity was observed continuously during both rest and task conditions. CP decomposition has also been applied to ictal EEG recordings from patients with

epilepsy. The extracted components have been used to localize the seizure onset zone [25]–[27] as well as to remove artifacts [25], [26]. When applied to group data analysis, CP decomposition and the extracted components/features can be used for a variety of purposes: classification [28], cross-modality comparison [29], and hypothesis testing [30].

The application of CP decomposition to functional MRI (fMRI) data has also been explored, usually at the group-level with the goal of finding common factors among subjects [14]. Individual analysis using CP is rarely performed because fMRI data lacks the rich spectral information present in EEG data. Andersen *et al.* [31] applied CP decomposition to finger-tapping fMRI data. Beckmann *et al.* [32] proposed a variation of CP which extends probabilistic independent component analysis (PICA) to higher orders by adding an independence constraint in the spatial dimension. While originally applied to task fMRI data, Damoiseaux *et al.* [33] applied tensor-PICA to resting fMRI data and identified 10 networks consistent with previous findings.

In previous studies where CP decomposition was applied to either EEG or fMRI data [24]–[29], [31], [32], [34], [35], the size of the datasets is relatively small (the largest dimension has an order of 10^4 elements or less). To explore dynamic functional connectivity, especially large-scale dynamics, we need to explore much larger datasets. There are two issues that have limited studies of this kind: scalability and robustness.

Scalability: The majority of the studies cited above either truncate the data into short temporal segments (e.g. in EEG ictal and event related potential recordings) or heavily down-sample the data in the spatial domain (e.g. in fMRI studies) or sometimes both. The degrees of freedom (DOF), approximately reflected by the largest dimension in the CP model for a typical 10-minute SEEG recording used in our experiments is larger ($\geq 10^5$) than in these previous studies. Moreover, as we will show in Section II below, the computational complexity increases approximately quadratically as the DOF increase when using ALS to estimate rank. Additionally, to achieve similar quality of solutions relative to our algorithm, ALS has to be applied with multi-start as shown in Section III, resulting in an even higher complexity. Therefore, in order to compute CP decompositions on data of this size, a fast and efficient algorithm is required.

Robustness (against local minima): It is well known that the ALS algorithm on a CP model is not guaranteed to converge to a global minimum or a stationary point, even when multi-start is applied during the optimization [14], [15]. The local minimum problem becomes more severe as the number of components increases. Performance is further compromised when a larger number of components than necessary are fit to the data (over-factoring), resulting in splitting rank-1 components into two or more factors.

Several techniques have been explored to improve the robustness and efficiency of the ALS algorithm. For example, Rajih *et al.* [36] added a line search after each major ALS iteration. Navasca *et al.* [37] applied Tikhonov regularization on each sub-problem in ALS iteration. However, similar to the

ALS alternatives reviewed above, these modifications result in significantly higher computation cost limiting their practical utility, particularly for large scale problems.

In 2D matrix scenarios, Haldar *et al.* [38] proposed an incremented-rank PowerFactorization (IRPF) approach to solve minimum-rank matrix recovery problems, where higher-rank solutions were obtained recursively using lower-rank results as warm initializations, resulting in a substantially improved performance compared to the standard convex optimization approach. Performance was also theoretically characterized later in [39]. In this work, we extend IRPF to higher-order tensors, with the goal of resolving the scalability and robustness issues discussed above. We refer to our approach as "scalable and robust sequential CP decomposition" (SRSCPD). As we show below, this algorithm is more robust than ALS and can be extended to large-scale problems. An outline of the remainder of the paper is as follows. Section II will briefly summarize the notation that will be used in later sections. Section III will describe the SRSCPD framework and the experimental design for both simulation and in-vivo data. Section IV will present the experiment results and conclusions follow in Section V.

II. NOTATION AND PRELIMINARIES

We first define some necessary notation and review the ALS algorithm which we use as part of our SRSCPD framework in section III. We largely follow the notational conventions and definitions of Kolda and Bader [14].

A. Scalar, Vector, Matrix and Tensor

A scalar is denoted by a lowercase letter, e.g. x; a vector by a bold lowercase letter, e.g. $\boldsymbol{x}$; a matrix by a bold uppercase letter, e.g. $\boldsymbol{X}$; and a tensor by a bold script letter, e.g. $\boldsymbol{\mathcal{X}}$. The number of dimensions is called the *order* and each dimension is referred as a *mode*. We use a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ in the following with individual elements denoted by $x_{i,j,k}$. The notation and algorithms extend naturally to higher-order tensors.

A rank-1 tensor can be expressed as the outer product of vectors. i.e. $\boldsymbol{\mathcal{X}} = \boldsymbol{a} \circ \boldsymbol{b} \circ \boldsymbol{c}$ where "$\circ$" represents the vector outer product. We use the norm:

$$\|\boldsymbol{\mathcal{X}}\| = \sqrt{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{i,j,k}^2} \qquad (1)$$

B. Matricization

Tensors can be unfolded or "matricized" into matrix form [14]. Matricization along dimension n is denoted by $\boldsymbol{X}_{(n)}$ so that a third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$ can be matricized into $\boldsymbol{X}_{(1)} \in \mathbb{R}^{I \times JK}$ or $\boldsymbol{X}_{(2)} \in \mathbb{R}^{J \times IK}$ or $\boldsymbol{X}_{(3)} \in \mathbb{R}^{K \times IJ}$.

C. Kronecker, Khatri-Rao and Hadamard Product

We use the definitions of the following matrix products defined in [14] and repeated here for convenience. The Kronecker product $\boldsymbol{X} \otimes \boldsymbol{Y}$ of matrix $\boldsymbol{X} \in \mathbb{R}^{I \times J}$ and $\boldsymbol{Y} \in \mathbb{R}^{K \times L}$ is defined as

$$X \otimes Y = \begin{bmatrix} x_{11}Y & x_{12}Y & \cdots & x_{1J}Y \\ x_{21}Y & x_{22}Y & \cdots & x_{2J}Y \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1}Y & x_{I2}Y & \cdots & x_{IJ}Y \end{bmatrix} \quad (2)$$

where x_{ij} is the (i,j)-th element of X.

The Khatri-Rao product $X \odot Y$ of matrix $X \in \mathbb{R}^{I \times K}$ and $Y \in \mathbb{R}^{J \times K}$ is the column-wise Kronecker product of X and Y

$$X \odot Y = [x_1 \otimes y_1 \quad x_2 \otimes y_2 \quad \cdots \quad x_K \otimes y_K] \quad (3)$$

where x_i is the i-th column of X.

The Hadamard product $X * Y$ of matrix $X \in \mathbb{R}^{I \times J}$ and $Y \in \mathbb{R}^{I \times J}$ is the element-wise matrix product

$$X * Y = \begin{bmatrix} x_{11}y_{11} & x_{12}y_{12} & \cdots & x_{1J}y_{1J} \\ x_{21}y_{21} & x_{22}y_{22} & \cdots & x_{2J}y_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1}y_{I1} & x_{I2}y_{I2} & \cdots & x_{IJ}y_{IJ} \end{bmatrix}. \quad (4)$$

In the following we use the property of the Khatri-Rao product [14]:

$$(X \odot Y)^\dagger = (X^T X * Y^T Y)^\dagger (X \odot Y)^T \quad (5)$$

where $X^\dagger$ represents the Moore-Penrose pseudo-inverse of X

D. CP Decomposition

CP decomposes a tensor into a sum of rank-1 tensors or components. For a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

$$\mathcal{X} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r + \mathcal{E} \quad (6)$$

where $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$, $c_r \in \mathbb{R}^K$, R is the *rank* or the number of components and $\mathcal{E}$ is the error tensor. If we group the components in each mode into a matrix, i.e. let $A = [a_1 a_2 \cdots a_R] \in \mathbb{R}^{I \times R}$ and similarly for $B \in \mathbb{R}^{J \times R}$ and $C \in \mathbb{R}^{K \times R}$, then the CP decomposition can be expressed as [14]

$$X_{(1)} = A(C \odot B)^T + E_{(1)} \quad (7)$$

or

$$X_{(2)} = B(C \odot A)^T + E_{(2)} \quad (8)$$

or

$$X_{(3)} = C(B \odot A)^T + E_{(3)} \quad (9)$$

where A, B, C are called the *loading matrices* for the three modes respectively.

E. Computation of CP decomposition and the ALS algorithm

Suppose we want to find the best rank R approximation of $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ via

$$\min_{\widehat{\mathcal{X}}} \| \mathcal{X} - \widehat{\mathcal{X}} \| + g(\widehat{\mathcal{X}}) \quad (10)$$

where $\widehat{\mathcal{X}} = \sum_{r=1}^{R} \lambda_r a_r \circ b_r \circ c_r$, λ_r represents the scale of component r and a_r, b_r and c_r have unit norm. $g(\widehat{\mathcal{X}}) = \mu_1 g_1(A) + \mu_2 g_2(B) + \mu_3 g_3(C)$ is a regularizing function

with (μ_1, μ_2, μ_3) the corresponding regularization parameters. The ALS algorithm solves this problem in an alternating fashion. We first solve for A with B and C fixed, then solve for B with A and C fixed, and so on. This procedure is repeated until some convergence criterion is satisfied. Note that, for quadratic regularizers, each sub-problem reduces to ordinary least square. Specifically, assume B and C are fixed and we are solving for A. Using the equivalent matrix expression discussed above, we can write the optimization problem as

$$\widehat{A} = \operatorname*{argmin}_{A} \| X_{(1)} - A(C \odot B)^T \|_F + \mu_1 g_1(A) \quad (11)$$

The solution with $\mu_1 = 0$ (without regularization) reduces to a regular least square solution:

$$\widehat{A} = X_{(1)}[(C \odot B)^T]^\dagger \quad (12)$$

Using the property in Eq. (5), we can rewrite as

$$\widehat{A} = X_{(1)}(C \odot B)(C^T C * B^T B)^\dagger \quad (13)$$

This expression is almost always preferable to (12) because it achieves a much lower computational complexity by only calculating the pseudo-inverse of an $R \times R$ matrix. Finally, we normalize each component and set λ_r equal to the normalization factor for the rth component, $r = 1, \dots, R$. For the case $\mu_1 \neq 0$, the solution in (12) is replaced by the solution to (11), which will be closed form if $g_1(A)$ is quadratic but may require an iterative solution in other cases. The full ALS algorithm is shown in Algorithm I.

ALGORITHM I: CP-ALS

Algorithm *CP-ALS* $(\mathcal{X}, R, \{A^*, B^*, C^*, \lambda^*\})$
*Initialize $A \in \mathbb{R}^{I \times R}, B \in \mathbb{R}^{J \times R}, C \in \mathbb{R}^{K \times R}, \lambda \in \mathbb{R}^R$
  While not converged†
    $A \leftarrow \operatorname*{argmin}_{A} \| X_{(1)} - A(C \odot B)^T \|_F + \mu_1 g_1(A)$
    $B \leftarrow \operatorname*{argmin}_{B} \| X_{(2)} - B(C \odot A)^T \|_F + \mu_2 g_2(B)$
    $C \leftarrow \operatorname*{argmin}_{C} \| X_{(3)} - C(B \odot A)^T \|_F + \mu_3 g_3(C)$
  End While
  Normalize A, B, C such that each column has unit norm and set λ_r equal to the normalization factor for the rth component, $r = 1, \dots, R$.
  Return A, B, C and λ
End Algorithm

---

* The initialization is typically performed using either random matrices or the R leading singular vectors of the matricized $\mathcal{X}$. We define specific initializations below for our SRSCPD algorithm.

† Algorithm convergence is determined when the mean of the absolute difference of the loading matrices between two adjacent iterations over all modes is less than some small constant, e.g. 10^{-5}.

III. MATERIALS AND METHODS
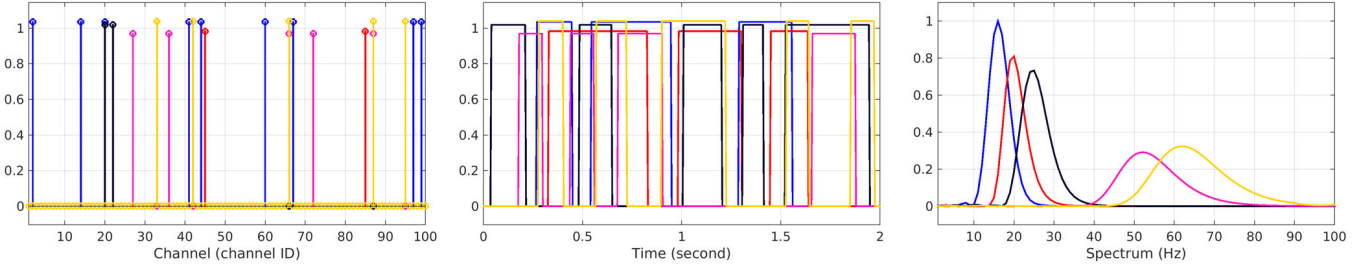
A. SRSCPD Framework

Figure 1: An example of the simulated data with 5 components. Each component is represented by a distinct color in all three modes. From left to right: The channel (spatial) mode shows the activated channels that participate in each network; The time (temporal) mode shows the block activation pattern for each network; The spectral mode shows the Morlet wavelet frequency spectrum for each network.

The best rank-r approximation of a matrix with respect to the Frobenius norm is given by the leading r factors of the SVD. This is not the case for CP decomposition of a higher-order tensor. Kolda [40] showed an example where the best rank-1 approximation is not part of the best rank-2 approximation of a tensor. As a result, components in the CP decomposition for a given desired rank should be found simultaneously. Smilde *et al.* [41] (example 4.3) showed that the naïve sequential CP, in which a rank-1 tensor is fit to the residue at each iteration, failed to extract the correct components even when the data are known to be perfectly trilinear. Interestingly, this greedy sequential approach is still frequently used, simply because it is the most tractable approach to fitting tensor models to large datasets [42], [43].

The determination of tensor rank is NP-hard [44]. Many metrics have been proposed to help find the correct rank, e.g. the core consistency diagnostic (CORCONDIA) [45], difference in fit (DIFFIT) [46] and automatic relevance determination (ARD) [47]. All these metrics require a set of decomposition results for all ranks up to the maximum rank R. Obtaining such a set of solutions using CP decomposition is quadratically more complex than finding a rank-1 approximation, as we need to compute the decompositions for each rank $r = 1, 2, ..., R$ separately $(1 + 2 + \cdots + R = O(R^2))$. This represents a significant challenge to use of higher rank tensor models and was the primary motivation for our development of the SRSCPD framework.

The SRSCPD framework is built on the original ALS algorithm. Our goal is to compute a rank-recursive set of decompositions from rank 1 to rank R. Our approach uses the result for rank r to initialize the decomposition for rank $r + 1$. Initialization for the additional component is found by fitting a rank-1 tensor to the residual from the rank r fit. In contrast, the original ALS algorithm does not use any information from rank r when fitting a model of rank $r + 1$. This "warm start" greatly improves convergence speed relative to the standard ALS algorithm except in very low rank cases (see simulation result in Fig. 4). The warm start may also help to avoid poor local minima in this non-convex optimization problem. As a result, we are able to address the problems with robustness and scalability for large-scale datasets.

The full SRSCPD framework is shown in Algorithm II, for a third-order tensor example. The inputs of the algorithm are a tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ and the desired maximum rank R. For each iteration r, a rank-r approximation is calculated using the

original *CP-ALS* algorithm with initializations $\{A^*, B^*, C^*, \lambda^*\}$. The initializations are formed by concatenating the solutions $\{A^{r-1}, B^{r-1}, C^{r-1}, \lambda^{r-1}\}$ from the previous iteration $r - 1$ with the rank-1 approximation $\{a', b', c', \lambda'\}$ of the residue tensor $\mathcal{X}_{res}$, where $\mathcal{X}_{res}$ is obtained by subtracting the reconstructed tensor using $\{A^{r-1}, B^{r-1}, C^{r-1}, \lambda^{r-1}\}$ from the original data tensor $\mathcal{X}$.

SRSCPD is flexible in the sense that techniques that have been proposed to improve the ALS algorithm can be directly incorporated. For example, one can add a line search along the estimated gradient descend direction for each mode at the end of each major iteration of ALS [36]. Moreover, constraints and regularization terms can be applied to each of the ALS sub-problems, e.g. non-negativity, sparsity, and smoothness.

ALGORITHM II: SRSCPD-ALS

Algorithm *SRSCPD-ALS* $(\mathcal{X}, R)$

$a^1, b^1, c^1, \lambda^1 \leftarrow CP\text{-}ALS(\mathcal{X}, 1)$

$\mathcal{X}_{res} \leftarrow \mathcal{X} - Tensor\_Recon(a^1, b^1, c^1, \lambda^1)$

$a', b', c', \lambda' \leftarrow CP\text{-}ALS(\mathcal{X}_{res}, 1)$

$A^* \leftarrow [a^1 \ a']; B^* \leftarrow [b^1 \ b']; C^* \leftarrow [c^1 \ c']; \lambda^* \leftarrow \begin{bmatrix} \lambda^1 \\ \lambda' \end{bmatrix}$

For $r = 2, 3, ..., R$

$\quad A^r, B^r, C^r, \lambda^r \leftarrow CP\text{-}ALS(\mathcal{X}, r, \{A^*, B^*, C^*, \lambda^*\})$

$\quad \mathcal{X}_{res} \leftarrow \mathcal{X} - Tensor\_Recon(A^r, B^r, C^r, \lambda^r)$

$\quad a', b', c', \lambda' \leftarrow CP\text{-}ALS(\mathcal{X}_{res}, 1)$

$\quad A^* \leftarrow [A^r \ a']; B^* \leftarrow [B^r \ b']; C^* \leftarrow [C^r \ c']; \lambda^* \leftarrow \begin{bmatrix} \lambda^r \\ \lambda' \end{bmatrix}$

End For

Return a set of solutions

$\quad \{a^1, b^1, c^1, \lambda^1\}, \{A^2, B^2, C^2, \lambda^2\}, ..., \{A^R, B^R, C^R, \lambda^R\}$

End Algorithm

B. Simulation

We simulated SEEG data [48] with 100 channels, 200 Hz sampling rate, 2 second duration for ranks from $R = 1$ to 10. In each component a total of N channels were co-activated where N was chosen randomly between 2 and 10. For each of the active channels for each component we generated a time series to represent a block activation pattern with the signal switching on and off, respectively, in active and inactive blocks. The number of active blocks over the 2-second period was selected randomly between 2 and 5 and both the minimum block length and the minimum interval between any adjacent activated blocks was set to 0.1 second. Within each active block, the signals in each component were unit amplitude sinusoids with frequencies chosen randomly

between 10 and 80 Hz. Finally, we added white Gaussian noise to the simulated data with a range of SNRs.

The third-order tensor $\mathcal{X}$ was generated by calculating the magnitude squared of the complex Morlet wavelet transform (MWT) coefficients of the simulated data matrix with center frequency 1 Hz, time resolution full-width-half-maximum (FWHM) of 2 seconds [49] in a linearly spaced frequency range from 1 to 100 Hz with interval 1 Hz. Thus, the final tensor $\mathcal{X}$ has the dimensions of $\mathbb{R}^{I \times J \times K}$, where $I = 100, J = 400, K = 100$. An example of the model used to simulate the data is shown in Fig. 1. Note that overlaps between components may occur in any of the three modes.

We first compared the robustness of the decomposition using the SRSCPD framework against ALS using 1, 2 and 5 random initializations (generated from a standard uniform distribution in the interval (0,1) using MATLAB (The MathWorks, Inc., Natick, MA, USA) function "rand"). The same convergence criterion was used for both algorithms and in all cases, we computed solutions from rank 1 to R. In both algorithms we used a non-negativity constraint on all loading matrices, because the squared magnitude of the wavelet coefficients are naturally non-negative and the constraint helps avoid degeneracy [50]. Let $\boldsymbol{A} \in \mathbb{R}^{I \times R}$, $\boldsymbol{B} \in \mathbb{R}^{J \times R}$, $\boldsymbol{C} \in \mathbb{R}^{K \times R}$ be the loading matrix in each of the three modes as described in Section II. Then in each sub-problem of the ALS, we used the following cost function for $\boldsymbol{A}$ (likewise for $\boldsymbol{B}$ and $\boldsymbol{C}$)

$$\widehat{\boldsymbol{A}} = \underset{\boldsymbol{A}}{\operatorname{argmin}} \left\| \boldsymbol{X}_{(1)} - \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T \right\|_F \ s.t. \boldsymbol{A} \geqslant 0 \qquad (14)$$

where "$\geqslant$" denotes the element-wise inequality.

Since we know the ground truth under the simulated settings, we assessed the quality of the solutions using the averaged congruence product (ACP) [51]. ACP is a measure of correlations between components. Specifically, let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be the column-wise normalized ground truth loading matrices and $\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{B}}, \widehat{\boldsymbol{C}}$ their estimated counterparts. Then the ACP is defined per [51] as

$$\text{ACP} = \max_{\boldsymbol{P}} \operatorname{tr}\left((\boldsymbol{A}^T \widehat{\boldsymbol{A}}) * (\boldsymbol{B}^T \widehat{\boldsymbol{B}}) * (\boldsymbol{C}^T \widehat{\boldsymbol{C}}) \, \boldsymbol{P} \right) \qquad (15)$$
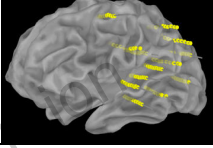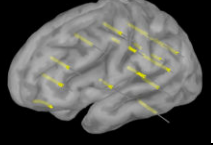
where $\boldsymbol{P}$ is a permutation matrix accounting for the ambiguity of the ordering of the solutions [16] and $\operatorname{tr}(\boldsymbol{X})$ indicates the trace of $\boldsymbol{X}$.

We evaluated the ACP of the solutions obtained from both ALS and SRSCPD as a function of R for SNR = 10. For each R, we ran 100 Monte Carlo trials and boxplots of ACP were generated. For each simulated tensor, we repeated ALS M times, where $M = 1, 2,$ and 5, each time using a different random initialization. The final solution was selected as that which has the lowest cost. We also box-plotted the Frobenius norm error as shown in Eq. 10 for each trial. Additionally, we recorded the run time for each of the methods. We then repeated the above study, but instead of varying R we conducted the experiment as a function of SNR with $R = 5$.

C. Application to In-Vivo SEEG Dataset

We performed retrospective analysis of patient data collected under an Institutional Review Board approved protocol for SEEG evaluation and monitoring in the Epilepsy

TABLE I: SUMMARY OF THE PATIENT DATA

Subject ID	1		2	
# of Channels	69		113	
Epilepsy Type	Posterior cingulate epilepsy		Left fronto-parietal	
Data Segment Time	09:23:00 – 09:33:00	Same Day: 13:22:02 – 13:32:02	21:50:30 – 22:00:30	Next Day: 08:35:20 – 08:45:20
Implantation Scheme				

Center, Cleveland Clinic, OH, USA. The SEEG evaluation performs invasive pre-surgical electrophysiological mapping for patients who have pharmaco-resistant focal epilepsy. For each patient, the implantation was performed using multi-lead depth electrodes, with each electrode comprising typically 10 contacts spaced a few millimeters apart (AdTech, Racine, Wisconsin; Integra, Plainsboro, New Jersey; or PMT, Chanhassen, Minnesota). The electrode locations were determined after a multidisciplinary patient management conference where the hypotheses about the epileptogenic zone were drawn based on available noninvasive data: clinical history, video EEG, MRI, PET, ictal SPECT and MEG. The electrodes were implanted according to the Talairach stereotactic method using orthogonal or oblique trajectories [52]. The implantation schemes are shown in Table I. The SEEG signals were recorded using a common reference on a Nihon Kohden EEG system with a sampling rate of 1000 Hz.

We chose two 10-minute data segments a minimum of 4 hours apart (see Table I) for each patient. The segments were selected using annotated video of the patients for periods of physical inactivity (e.g. reading, watching TV). Using co-registration of the post-implant X-ray CT to the patients' MR image, we selected the subset of the SEEG contacts that were in gray matter.

For each data segment we applied the MWT with center frequency 1 Hz and time resolution FWHM of 2 seconds in a linearly spaced frequency range of 1 to 100 Hz with interval 1 Hz. We computed the squared magnitude of the wavelet coefficients and temporally down-sampled the resulting envelope data in each frequency band by a factor of 5, resulting in a new envelope sampling rate of 200 Hz. We performed a flattening of the power spectrum to compensate for its "1/f" characteristics to emphasize the higher frequency components. The resulting data were then represented as a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, where I is the number of channels (see Table I), $J = 120,000$ is the number of time points and $K = 100$ is the number of frequency bins.

We applied SRSCPD to each tensor with three additional constraints: non-negativity on all three modes due to the non-negative squared magnitude of the wavelet coefficients; a sparsity constraint on the spatial (channel) mode as we assumed that in each network only a small set of channels would be involved; and a smoothness constraint on the
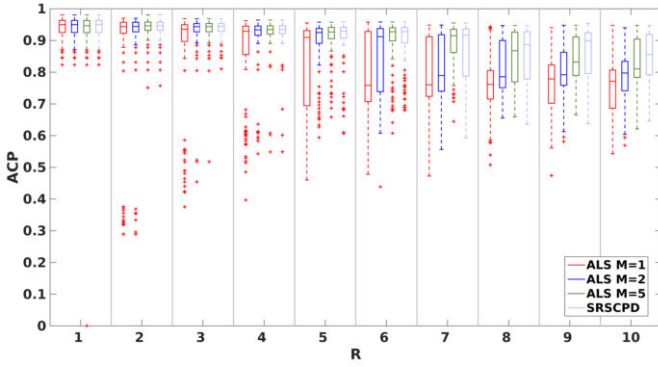
Figure 2: Simulation results. Boxplots of ACP over 100 Monte Carlo trials are shown as a function of R. M denotes the number of random initializations when using the original ALS algorithm.

spectral mode reflecting the limited frequency resolution of the MWT. Let $\boldsymbol{A} \in \mathbb{R}^{I \times R}$, $\boldsymbol{B} \in \mathbb{R}^{J \times R}$, $\boldsymbol{C} \in \mathbb{R}^{K \times R}$ be the loading matrices for each of the three modes. Then in the sub-problems of the ALS, we use the following cost functions:

$$\widehat{\boldsymbol{A}} = \underset{\boldsymbol{A}}{\mathrm{argmin}} \left\| \boldsymbol{X}_{(1)} - \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T \right\|_F^2 + \mu_1 \sum_i \|\boldsymbol{A}_{\cdot i}\|_{l_1}, s.t. \boldsymbol{A} \geqslant 0 \quad (16)$$

$$\widehat{\boldsymbol{B}} = \underset{\boldsymbol{B}}{\mathrm{argmin}} \left\| \boldsymbol{X}_{(2)} - \boldsymbol{B}(\boldsymbol{C} \odot \boldsymbol{A})^T \right\|_F^2, s.t. \boldsymbol{B} \geqslant 0 \quad (17)$$

$$\widehat{\boldsymbol{C}} = \underset{\boldsymbol{C}}{\mathrm{argmin}} \left\| \boldsymbol{X}_{(3)} - \boldsymbol{C}(\boldsymbol{B} \odot \boldsymbol{A})^T \right\|_F^2 + \mu_3 \sum_k \|\nabla \boldsymbol{C}_{\cdot k}\|_{l_2}^2, s.t. \boldsymbol{C} \geqslant 0 \quad (18)$$

where $\boldsymbol{X}_{\cdot i}$ denotes the ith column of $\boldsymbol{X}$, μ_1 and μ_3 are the regularization parameters which were set to 0.2 empirically. $\|\cdot\|_{l_1}$ and $\|\cdot\|_{l_2}$ denotes the l_1 norm and l_2 norm respectively. ∇ is the finite difference operator on the columns of $\boldsymbol{C}$. We solved each of the convex sub-problems (16) – (18) using Auslender and Teboulle's single-projection algorithm [53] in the TFOCS toolbox [54].

Finally, the rank of the tensor was estimated based on the decomposition results using CORCONDIA [45]. This rank metric used the fact that the trilinearity of components starts decreasing in the case of over-factoring (the number of fitted
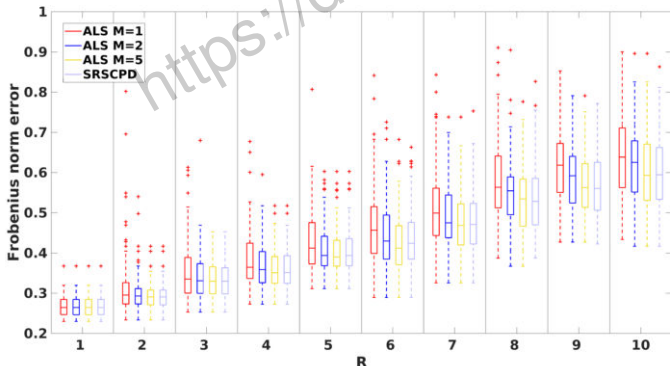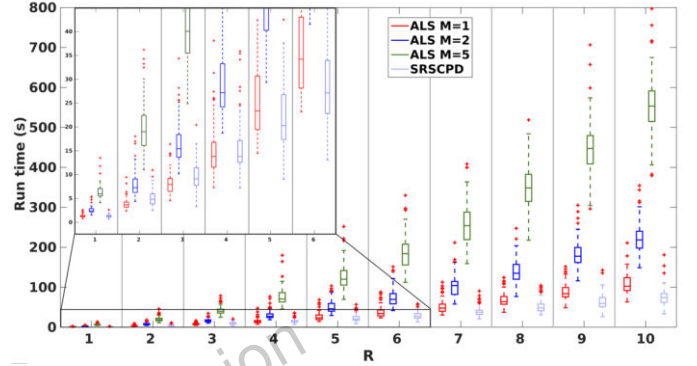

Figure 4: Simulation results. Boxplots of the run time in seconds over 100 Monte Carlo trials are shown as a function of R. M denotes the number of random initializations when using original ALS algorithm. Top-left panel shows the zoomed-in results for a better comparison for lower rank data.

components is greater than the actual rank).

IV. RESULTS

A. Simulation

Fig. 2 shows performance of ALS vs SRSCPD as a function of rank R. For small R all results are similar. However, for larger ranks we see that the ALS results are strongly dependent on initialization and that performance for $M = 5$ is significantly better than for $M = 2$ and $M = 1$. SRSCPD benefits from using the results of the lower rank as an initialization, resulting in overall improved performance (higher median ACP) relative to all three versions of ALS.

Fig. 3 shows the corresponding Frobenius norm error. As with Fig. 2, ALS with $M = 1$ and 2 shows larger error than $M = 5$, with the difference increasing with rank. In contrast to the ACP metric in Fig. 2, the error for ALS with $M = 5$ in Fig. 3 is very similar to that for SRSCPD and sometimes smaller for higher ranks. Closer examination of these results revealed that in cases where this occurs, ALS fails to find one or more of the weaker components that SRSCPD does find. Instead, part of the noise in the data is fit to one of the tensor components. This in turn leads to a lower squared error in the fit even though the extracted components are a poorer fit to the ground truth as measured with ACP.

Fig. 4 shows the run time as a function of R (the run time was measured using MATLAB with Dell Precision T3610 computer, Intel Xeon E5-1650 v2 CPU). As expected, the

Figure 3: Simulation results. Boxplots of the Frobenius norm error over 100 Monte Carlo trials are shown as a function of R. M denotes the number of random initializations when using the original ALS algorithm.
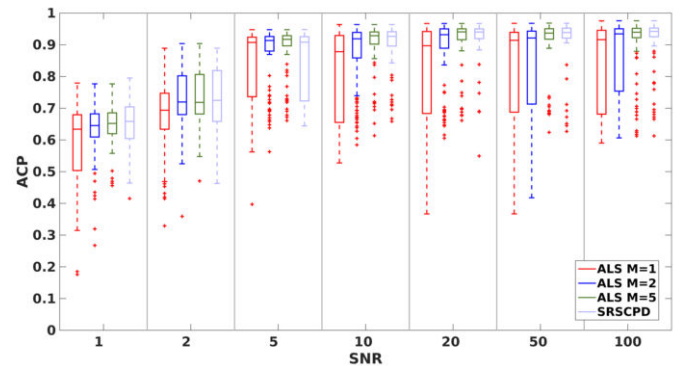

Figure 5: Simulation results. Boxplots of the ACP over 100 Monte Carlo trials are shown as a function of SNR. M denotes the number of random initializations when using the original ALS algorithm.
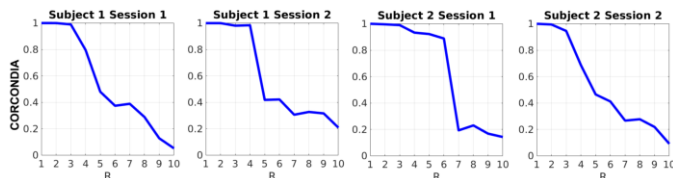
Figure 6: CORCONDIA rank metric are shown as a function of rank R for two sessions of both subjects.

ratios of the run time among the ALS methods are approximately proportional to M, the number of different initializations. The cost of SRSCPD is significantly lower than that for ALS with $M = 2$ and 5 restarts. As the rank increases ($R > 4$), the cost for SRSCPD is even lower than that for ALS without restart, $M = 1$. The reason for this is that the warm start in SRSCPD produces a better initialization that not only results in improved performance (Fig. 2) but also faster

convergence of the ALS sub-problems as shown in Fig. S1 in the supplemental material.

Fig. 5 shows that as the SNR increases, ACP also improves for all methods. SRSCPD shows generally similar performance to ALS with $M = 5$ restarts and is substantially better than results for $M = 1$. However, for lower SNRs, the performance of ALS with $M = 5$ restarts is superior to SRSCPD.

B. In-Vivo SEEG Dataset

1) Estimation of Rank

Fig. 6 shows plots of the CORCONDIA rank metric as a function of R for the two sessions for both subjects. Per recommendations in [45], rank should be chosen so that the CORCONDIA value is higher than 0.9. Based on the plots in
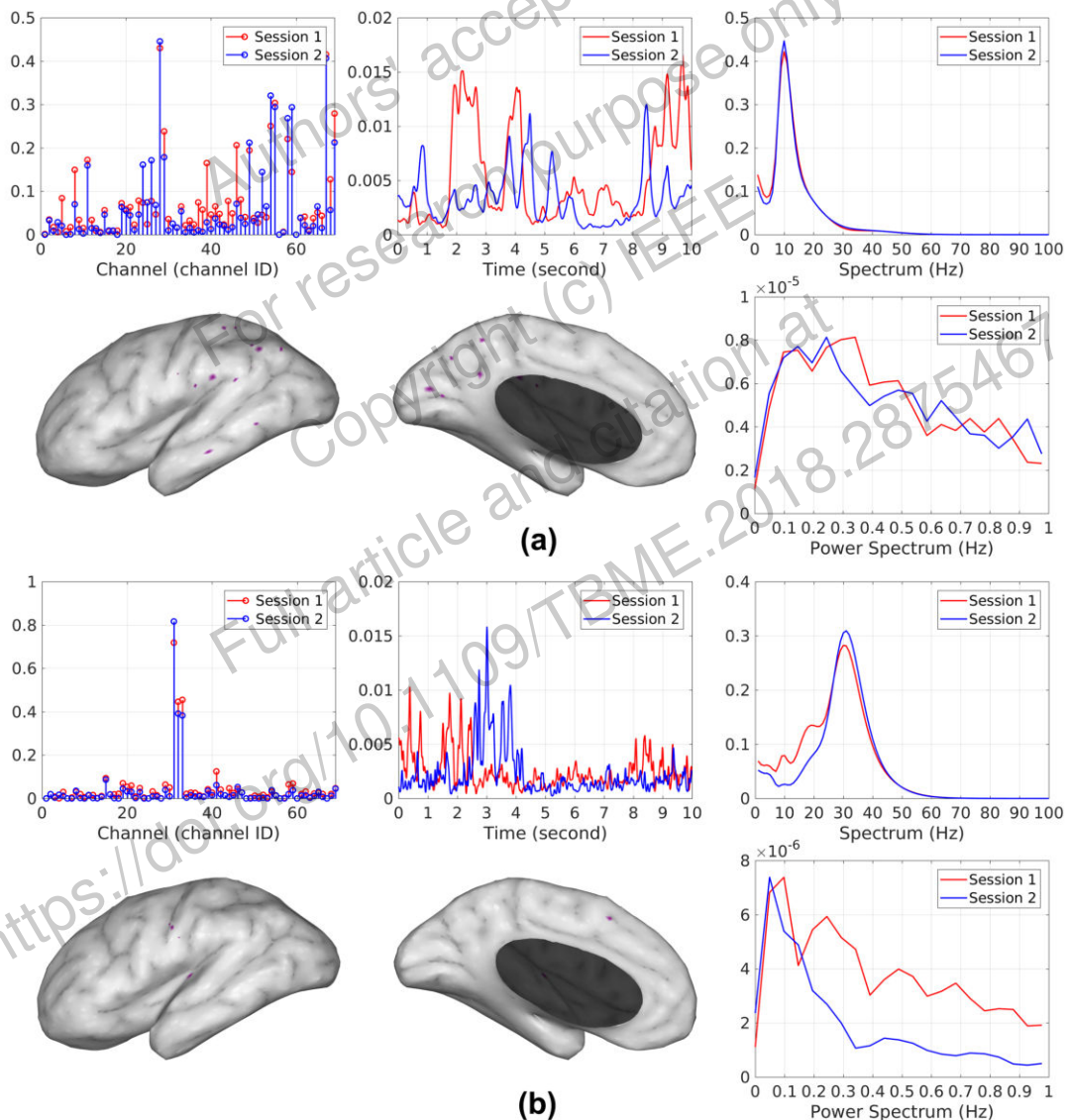

Figure 7: Two consistent components for Subject 1 show: (a) locations consistent with activity in the default mode network within the alpha band and (b) Motor activity in the beta frequency range. For each pair of consistent components, we show modes for session 1 in red and session 2 in blue in the top row. From left to right: The channel (spatial) mode shows the activated channels that participate in each network; The time (temporal) mode shows the dynamic variations of each network (only the first 10 seconds is shown for better visualization); The spectral mode shows the frequency-dependent component of the tensor. In the bottom row the left and middle sub-figures show the spatial distribution of the activated channels mapped onto the subject's smoothed cortical surface. For visualization purposes, a contact or channel is defined as activated if the value of the (normalized) channel mode at that contact exceeds a threshold of 0.05 in both sessions. The right sub-figure shows the Welch power spectrum of the temporal mode.

Fig. 6, we selected $R = 3$ and 4 for the two sessions of Subject 1 and $R = 5$ and 3 for the two sessions of Subject 2.

2) Intra-subject Network Comparison

We found corresponding consistent components (CCs) across the two sessions as the pair of components that had the largest product of spatial congruence and spectral congruence, i.e. for each component $(\boldsymbol{a}_i^1, \boldsymbol{b}_i^1, \boldsymbol{c}_i^1)$ in the first session, we found $(\boldsymbol{a}_j^2, \boldsymbol{b}_j^2, \boldsymbol{c}_j^2)$ in the second session such that

$$j = \underset{j}{\text{argmax}}\, \text{tr}((\boldsymbol{a}_i^{1T}\boldsymbol{a}_j^2) * ((\boldsymbol{c}_i^{1T}\boldsymbol{c}_j^2)), i = 1, \ldots, R \quad (19)$$

where $\boldsymbol{a}_i^k, \boldsymbol{b}_i^k, \boldsymbol{c}_i^k$ represent the ith spatial, temporal and spectral component of the kth ($k = 1, 2$) session, respectively. Note that we do not expect temporal congruence across sessions. For Subject 1, we found three CCs with large congruence product (>0.6). Fig. 7 shows two of these components, the third is included in the supplemental material, Fig. S2. The congruence was 0.892 (spatial mode) and 0.997 (spectral mode) for component (a) and 0.984 (spatial mode)

and 0.967 (spectral mode) for (b).

The first network for Subject 1 includes contacts in angular gyrus, mid-temporal gyrus, and precuneus all consistent with activation in the Default Mode Network (DMN) [55]. Electrodes were not present in this subject for other regions that are typically included in the DMN, such as the superior frontal gyrus (SFG). The spectrum for this component is dominated by alpha rhythms, which is consistent with studies of the DMN in the EEG/MEG literature [56]–[58]. The second network we found for this subject contains contacts mainly in the somatosensory and motor cortex with a peak in the beta band, which is consistent with sensorimotor rhythms as reported in the EEG/ECoG literature [59], [60]: beta activity appears most strongly in both motor and somatosensory cortex in movement preparation periods and steady contraction periods following a movement. These subjects were monitored during periods of natural inactivity when we might expect some amount of motor activity, for example page-turning while reading.
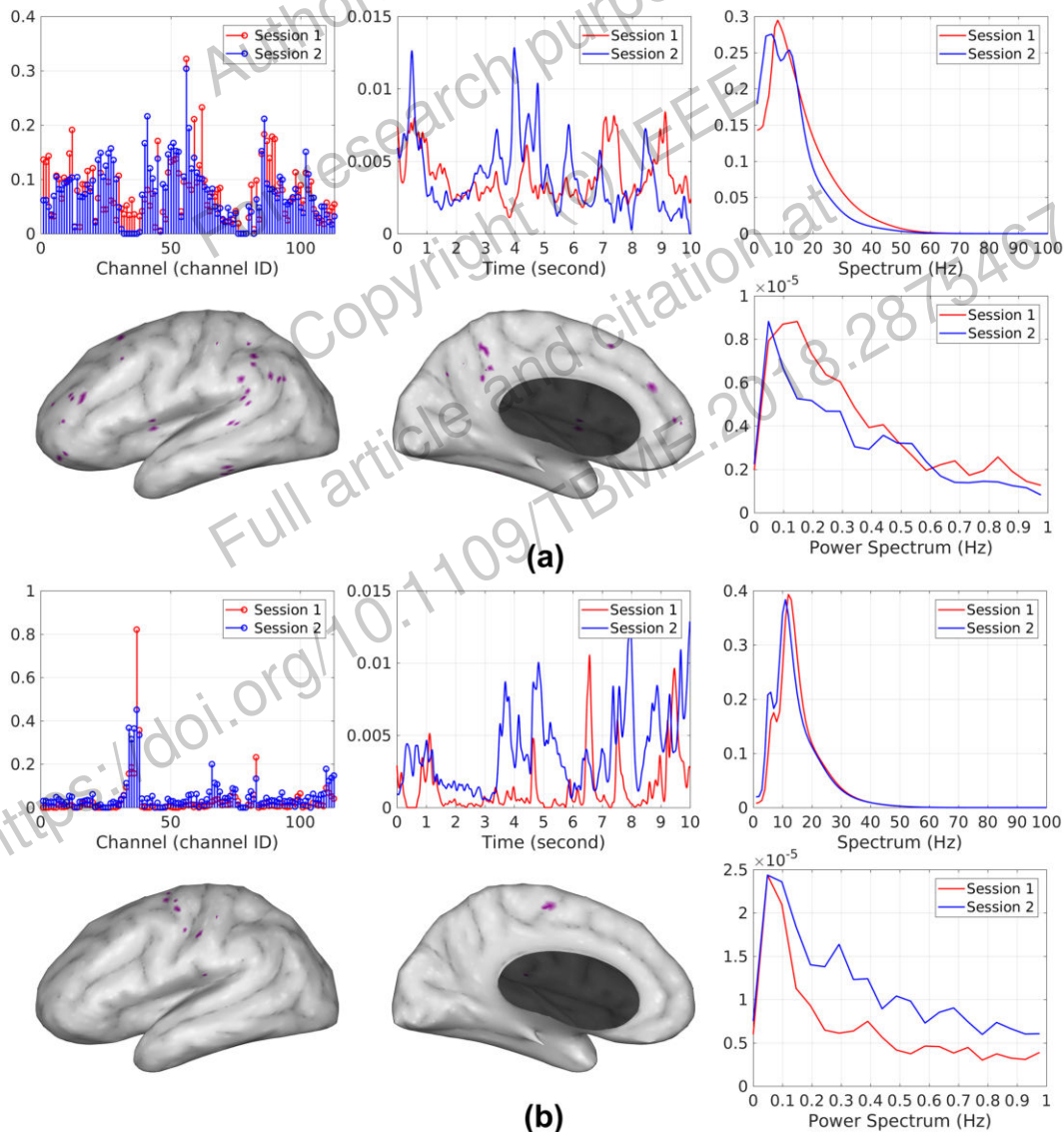

Figure 8: Two consistent components for Subject 2: (a) default mode network in alpha frequency and (b) motor network centered on the mu range. Details as for Fig. 6.

The temporal mode shows the power envelope of the SEEG signals, i.e. the dynamic variation in power across time of the spectral model. The bottom right sub-figure shows the power spectrum of this envelope estimated using the Welch method [61] after high-pass filtering with a cutoff frequency 0.02 Hz to remove DC drift. For both components, the power spectral density peaks at a frequency of approximately 0.1 Hz (0.05 Hz – 0.15 Hz), which is similar to the dominant frequency found in resting fMRI BOLD (Blood Oxygen Level Dependent) oscillations [62].

For Subject 2, we also found three CCs. Fig. 8 shows two of the three, the remaining one is shown in Fig. S3. Again, the similarity between sessions was high with congruence 0.652 (spatial mode) and 0.959 (spectral mode) for the DMN and 0.799 (spatial mode) and 0.957 (spectral mode) for the somatomotor network. In this subject there are electrodes in the SFG, unlike the first subject, and we now observe that the DMN does indeed include SFG. As with the first subject we see a second strong component in the somatosensory and motor cortex. However, unlike the first subject, the signal in this case is predominantly mu rather than beta. Again, this is consistent with periods of natural inactivity, where mu rhythms are typically observed in somatomotor cortex in parallel with alpha activity in the visual cortex during resting [63].

3) Inter-subject network comparison

When comparing the results between Subject 1 and Subject 2, we found that in both cases the DMN activity is dominated by alpha activity while the motor network is predominantly beta or mu, despite the fact the two subjects had different electrode implantation schemes and different locations of their epileptogenic zones (see Table II).

4) Artifact Detection

SRSCPD not only can be used to identify brain networks, it can also detect artifacts. For example, in the supplemental material, Fig. S4 shows a component that was mismatched (i.e. low spatial and spectral congruence) between the two sessions for Subject 1. Similarly, Fig. S5 shows one of the components that was mismatched between sessions for Subject 2. These mismatched components are likely artifacts as they mostly contain a burst or bursts of activity on a very limited number of channels and the time courses do not look physiological in nature.

5) Comparison to Results Using ALS Algorithm

We also applied the traditional ALS algorithm to the same datasets and compared the components obtained from the two methods. ALS did not find as many functionally distinct networks as SRSCPD did from each individual session based on the same rank selection criterion. The rank was estimated to be $R = 2$ and 3 for the two sessions of Subject 1 and $R = 3$ and 3 for the two sessions of Subject 2 using the ALS algorithm as shown in Fig. S6. Moreover, we only found one CC (a DMN shown in Fig. S7) between the two sessions in one of the subjects. No other CCs were found as the maximum congruence product (Eq. 19) was less than 0.15 between all other pairs of components. These results show that ALS is not as robust as the SRSCPD algorithm especially when SNR is low. See supplemental materials (Fig. S6 – S16) for details.

V. CONCLUSION

We have described a novel framework for decomposition of electrophysiological data using a third-order tensor. Our SRSCPD approach is based on the original ALS algorithm, using a warm-start in a rank-recursive search to both improve the quality of results and reduce computation cost relative to conventional ALS with multi-start. The SRSCPD framework is scalable to large datasets due to its use of the warm start. We have shown its application to SEEG data in two subjects with epilepsy and found two consistent brain networks across different sessions of recordings several hours apart and in two different subjects. In contrast, ALS found only one consistent network (the default mode) in one subject. This consistency shows promise for the use of SRSCPD for robust identification of spontaneous brain network activity from invasively recorded EEG in individual subjects.

REFERENCES

[1] K. J. Friston, "Functional and Effective Connectivity: A Review," *Brain Connect.*, vol. 1, no. 1, pp. 13–36, 2011.

[2] M. I. Rabinovich, K. J. Friston, and P. Varona, *Principles of Brain Dynamics: Global State Interactions*. MIT Press, 2012.

[3] C. Chang and G. H. Glover, "Time-frequency dynamics of resting-state brain connectivity measured with fMRI," *Neuroimage*, vol. 50, no. 1, pp. 81–98, 2010.

[4] D. A. Handwerker, V. Roopchansingh, J. Gonzalez-Castillo, and P. A. Bandettini, "Periodic changes in fMRI connectivity," *Neuroimage*, vol. 63, no. 3, pp. 1712–1719, 2012.

[5] R. M. Hutchison *et al.*, "Dynamic functional connectivity: Promise, issues, and interpretations," *Neuroimage*, vol. 80, pp. 360–378, 2013.

[6] M. G. Preti, T. A. Bolton, and D. Van De Ville, "The dynamic functional connectome: State-of-the-art and perspectives," *Neuroimage*, vol. 160, pp. 41–54, 2017.

[7] S. M. Smith *et al.*, "Temporally-independent functional modes of spontaneous brain activity," *Proc. Natl. Acad. Sci.*, vol. 109, no. 8, pp. 3131–3136, 2012.

[8] C. F. Beckmann and S. M. Smith, "Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging," *IEEE Trans. Med. Imaging*, vol. 23, no. 2, pp. 137–152, 2004.

[9] V. D. Calhoun and T. Adali, "Multisubject independent component analysis of fMRI: A decade of intrinsic networks, default mode, and neurodiagnostic discovery," *IEEE Rev. Biomed. Eng.*, vol. 5, pp. 60–73, 2012.

[10] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak, "Functional Connectivity: The Principal-Component Analysis of Large (PET) Data Sets," *J. Cereb. Blood Flow Metab.*, vol. 13, no. 1, pp. 5–14, 1993.

[11] F. I. Karahanoğlu and D. Van De Ville, "Transient brain activity disentangles fMRI resting-state dynamics in terms of spatially and temporally overlapping networks," *Nat. Commun.*, vol. 6, p. 7751, 2015.

[12] F. L. Hitchcock, "The Expression of a Tensor or a Polyadic as a Sum of Products," *J. Math. Phys.*, vol. 6, no. 1–4, pp. 164–189, 1927.

[13] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemom.*, vol. 14, no. 3, pp. 105–122, 2000.

[14] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[15] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.

[16] R. a Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis," *UCLA Work. Pap. Phonetics*, vol. 16, no. 10, pp. 1–84, 1970.

[17] J. D. Carroll and J. J. Chang, "Analysis of individual differences in

multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[18] J. B. Kruskal, "Rank, decomposition, and uniqueness for 3-way and N-way arrays," *Multiway Data Analysis*. pp. 7–18, 1989.

[19] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *J. Chemom.*, vol. 14, no. 3, pp. 229–239, 2000.

[20] N. M. Faber, R. Bro, and P. K. Hopke, "Recent developments in CANDECOMP/PARAFAC algorithms: A critical review," *Chemom. Intell. Lab. Syst.*, vol. 65, no. 1, pp. 119–137, 2003.

[21] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Comput. Stat. Data Anal.*, vol. 50, no. 7, pp. 1700–1734, 2006.

[22] F. Cong, Q. H. Lin, L. D. Kuang, X. F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of EEG signals: A brief review," *J. Neurosci. Methods*, vol. 248, pp. 59–69, 2015.

[23] J. Mocks, "Topographic components model for event-related potentials and some biophysical considerations," *IEEE Trans. Biomed. Eng.*, vol. 35, no. 6, pp. 482–484, 1988.

[24] F. Miwakeichi, E. Martinez-Montes, P. A. Valdes-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis," *Neuroimage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[25] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. 10–18, 2007.

[26] M. De Vos *et al.*, "Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone," *Neuroimage*, vol. 37, no. 3, pp. 844–854, 2007.

[27] W. Deburchgraeve *et al.*, "Neonatal seizure localization using PARAFAC decomposition," *Clin. Neurophysiol.*, vol. 120, no. 10, pp. 1787–1796, 2009.

[28] H. Lee, Y.-D. Kim, A. Cichocki, and S. Choi, "Nonnegative Tensor Factorization for Continuous EEG Classification," *Int. J. Neural Syst.*, vol. 17, no. 04, pp. 305–317, 2007.

[29] K. Vanderperren *et al.*, "Single trial ERP reading based on parallel factor analysis," *Psychophysiology*, vol. 50, no. 1, pp. 97–110, 2013.

[30] 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*, vol. 29, no. 3, pp. 938–947, 2006.

[31] A. H. Andersen and W. S. Rayens, "Structure-seeking multilinear methods for the analysis of fMRI data," *Neuroimage*, vol. 22, no. 2, pp. 728–739, 2004.

[32] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject FMRI analysis," *Neuroimage*, vol. 25, no. 1, pp. 294–311, 2005.

[33] J. S. Damoiseaux *et al.*, "Consistent resting-state networks across healthy subjects," *Proc. Natl. Acad. Sci.*, vol. 103, no. 37, pp. 13848–13853, 2006.

[34] M. Morup, L. K. Hansen, J. Parnas, and S. M. Arnfred, "Decomposing the time-frequency representation of EEG using nonnegative matrix and multi-way factorization," *Tech. Univ. Denmark Tech. Rep.*, 2006.

[35] M. Barnathan, V. Megalooikonomou, C. Faloutsos, S. Faro, and F. B. Mohamed, "TWave: High-order analysis of functional MRI," *Neuroimage*, vol. 58, no. 2, pp. 537–548, 2011.

[36] M. Rajih, P. Comon, and R. A. Harshman, "Enhanced Line Search: A Novel Method to Accelerate PARAFAC," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1128–1147, 2008.

[37] C. Navasca, L. De Lathauwer, and S. Kindermann, "Swamp reducing technique for tensor decomposition," in *European Signal Processing Conference*, 2008, pp. 1–5.

[38] J. P. Haldar and D. Hernando, "Rank-constrained solutions to linear matrix equations using powerfactorization," *IEEE Signal Process. Lett.*, vol. 16, no. 7, pp. 584–587, 2009.

[39] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank Matrix Completion using Alternating Minimization," *Proc. forty-fifth Annu. ACM Symp. Theory Comput.*, pp. 665–674, 2012.

[40] T. G. Kolda, "Orthogonal Tensor Decompositions," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 243–255, 2001.

[41] A. Smilde, R. Bro, and P. Geladi, *Multi-Way Analysis with Applications in the Chemical Sciences*. 2004.

[42] T. G. Kolda, B. W. Bader, and J. P. Kenny, "Higher-order web link analysis using multilinear algebra," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2005, pp. 242–249.

[43] T. Zhang and G. H. Golub, "Rank-One Approximation to High Order Tensors," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 2, pp. 534–550, 2001.

[44] J. Hastad, "Tensor rank is NP-complete," *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.

[45] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *J. Chemom.*, vol. 17, no. 5, pp. 274–286, 2003.

[46] M. E. Timmerman and H. A. L. Kiers, "Three-mode principal components analysis: Choosing the numbers of components and sensitivity to local optima," *Br. J. Math. Stat. Psychol.*, vol. 53, no. 1, pp. 1–16, 2000.

[47] M. Morup and L. K. Hansena, "Automatic relevance determination for multi-way models," *J. Chemom.*, vol. 23, no. 7–8, pp. 352–363, 2009.

[48] J. Bancaud *et al.*, *La stéréoencéphalographie dans l'épilepsie. Informations neuro-physio-pathologiques apportées par l'investigation fonctionnelle stéréotaxique*. Paris: Masson, 1965.

[49] F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, and R. M. Leahy, "Brainstorm: A user-friendly application for MEG/EEG analysis," *Comput. Intell. Neurosci.*, vol. 2011, p. 8, 2011.

[50] L. Lim, "Optimal solutions to non-negative PARAFAC / multilinear NMF always exist," in *Workshop on Tensor Decompositions and Applications, Centre International de rencontres Mathmatiques, Luminy, France*, 2005.

[51] G. Tomasi and R. Bro, "PARAFAC and missing values," *Chemom. Intell. Lab. Syst.*, vol. 75, no. 2, pp. 163–180, 2005.

[52] J. Gonzalez-Martinez *et al.*, "Stereotactic placement of depth electrodes in medically intractable epilepsy," *J. Neurosurg.*, vol. 120, no. 3, pp. 639–644, Mar. 2014.

[53] A. Auslender and M. Teboulle, "Interior Gradient and Proximal Methods for Convex and Conic Optimization," *SIAM J. OPTIM. c Soc. Ind. Appl. Math.*, vol. 16, no. 3, pp. 697–725, 2006.

[54] S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comput.*, vol. 3, no. 3, pp. 165–218, 2011.

[55] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network: Anatomy, function, and relevance to disease," *Ann. N. Y. Acad. Sci.*, vol. 1124, pp. 1–38, 2008.

[56] D. L. Schomer and F. Lopes da Silva, *Niedermeyer's Electroencephalography: basic principles, clinical applications, and related fields*. 2011.

[57] R. I. Goldman, J. M. Stern, J. Engel, and M. S. Cohen, "Simultaneous EEG and fMRI of the alpha rhythm," *Neuroreport*, vol. 13, no. 18, pp. 2487–2492, 2002.

[58] M. J. Brookes *et al.*, "Investigating the electrophysiological basis of resting state networks using magnetoencephalography," *Proc. Natl. Acad. Sci.*, vol. 108, no. 40, pp. 16783–16788, 2011.

[59] S. N. Baker, "Oscillatory interactions between sensorimotor cortex and the periphery," *Curr. Opin. Neurobiol.*, vol. 17, no. 6, pp. 649–655, 2007.

[60] Y. Zhang, Y. Chen, S. L. Bressler, and M. Ding, "Response preparation and inhibition: The role of the cortical sensorimotor beta rhythm," *Neuroscience*, vol. 156, no. 1, pp. 238–246, 2008.

[61] P. D. Welch, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Aver. aging Over Short, Modified Periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.

[62] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI," *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, 1995.

[63] M. Steriade, "Cellular Substrates of Brain Rhythms," *Basic principles, clinical applications, and related fields*, pp. 27–62, 1980.