Sparse tensor dimensionality reduction with application to clustering of functional connectivity

Gaëtan Frusque, Julien Jung, Pierre Borgnat, Paulo Gonçalves

Sparse tensor dimensionality reduction with application to clustering of functional connectivity

Gaëtan Frusque [a], Julien Jung[b], Pierre Borgnat[c], and Paulo Gonçalves[a]

[a]Univ Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, LIP UMR 5668, F-69342, Lyon, France
[b]HCL, Neuro. Hosp., Functional Neurology and Epileptology Dept & Lyon Neurosc. Res. Cent., INSERM, CNRS, Lyon, France
[c]Univ Lyon, ENS de Lyon, UCB Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France

1. ABSTRACT

Functional connectivity (FC) is a graph-like data structure commonly used by neuroscientists to study the dynamic behaviour of the brain activity. However, these analyses rapidly become complex and time-consuming. In this work, we present complementary empirical results on two tensor decomposition previously proposed named modified High Order Orthogonal Iteration (mHOOI) and High Order sparse Singular Value Decomposition (HOsSVD). These decompositions associated to k-means were shown to be useful for the study of multi trial functional connectivity dataset.

Keywords: dynamic networks, graph decomposition, clustering, dimensionality reduction, sparsity, tensor decompositions, HOOI, functional connectivity, iEEG.

2. INTRODUCTION

Epilepsy is one of the most common neurological disorders in the world population. iEEG electrodes[1] are used to exhibit the stages of a seizure distinguished by similar pattern in different areas of the brain. Functional Connectivities (FC) that quantify along time these temporal similarities are calculated between all pairs of signals, usually by means of the spectral coherence or the Phase Locking Value.[2] Considering electrodes as nodes and FCs as weights on the edges, neuroscientists try to find in the data, graph-like structures evolving through time (see Fig. 1). It is generally assumed that the FC dynamics are comparable from one seizure to another for a same patient. Another asset of iEEG monitoring is the possibility to record several seizures. In Frusque et al.,2019,[3] we proposed a procedure able to extract relevant ensembles of simultaneously activated components, viewed as the edges of sub-graphs of a FC network inferred from the iEEG signals.

Precisely, for each seizure, the recorded iEEG signals yield L pairwise Functional Connectivity measures (FC) as time series of T samples. The corresponding data matrix $\mathbf{X} \in \mathbb{R}^{L \times T}$ is referred to as an Epoch. As we observe S different seizures for a same patient, the resulting epochs are stacked in a 3-modes tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times T \times S}$. The data analysis process, presented Fig. 2, corresponds to first apply a dimensionality reduction method in order to denoise and reduce the dimensionality of the dataset, then apply k-means to the reduction. Notice that directly applying an extension of k-means to the tensor $\boldsymbol{\mathcal{X}}$ corresponds to maximize the following criteria :

$$\underset{\mathbf{A}}{\operatorname{argmax}} \qquad || \mathbf{A}^t \mathbf{X}_{(L)} ||_F^2 \tag{1}$$

Where $\mathbf{A}$ corresponds to a normalized cluster assignment matrix[3] and $\mathbf{X}_{(L)} \in \mathbb{R}^{L \times TS}$ is the mode-L matricization of $\boldsymbol{\mathcal{X}}$. Such generalisation does not take advantage of the tensor structure of the dataset, moreover k-means is known to perform poorly on high dimension (here TS).[4] We compared and proposed[3] several reduction methods,

Further author information:
Gaetan Frusque.: E-mail: gaetan.frusque@ens-lyon.fr, Telephone: +33 (0)6 32 14 45 18
Julien Jung.: E-mail: julien.jung@chu-lyon.fr,
Pierre Borgnat.: E-mail: Pierre.Borgnat@ens-lyon.fr,
Paulo Gonçalves.: E-mail: paulo.goncalves@ens-lyon.fr

transforming the tensor $\boldsymbol{\mathcal{X}}$ to a factor matrix $\mathbf{F} \in \mathbb{R}^{L \times K}$ with $K << TS$. The final data analysis process can then be summarized by maximizing the following criteria where $\mathbf{A}$ as the same structure as in (1).

$$\underset{\mathbf{A}}{\operatorname{argmax}} \qquad || \mathbf{A}^t \mathbf{F} ||_F^2 \tag{2}$$

The objective of this presentation is to present complementary empirical results on two tensor decomposition proposed in[3] named modified High Order Orthogonal Iteration (mHOOI) and High Order sparse Singular Value Decomposition (HOsSVD). Indeed both methods have an original algorithmic implementation without theoretical proof of convergence and convergence speed yet. Also HOsSVD which is a sparse tensor decomposition has not been compared to tensor decomposition with similar constraints. After recalling the notation, we present mHOOI, HOsSVD and similar decompositions for comparison. Finally the convergence property of HOsSVD and mHOOI are exhibited via a model of functional connectivity as well as the performance of HOsSVD to decompose a dataset in sparse components via the same model and real epileptic datasets.

3. NOTATIONS

Tensors are denoted with bold case calligraphic letters $\boldsymbol{\mathcal{X}}$, matrices and vector are denoted respectively in bold upper-case and lower-case $\mathbf{X}$, $\mathbf{x}$, and scalars by lower-case letters x. Notice that : l, t, s, k, n will be used as indices, and L, T, S, K, N will be reserved to denote their index upper bounds. Here they correspond respectively to the FC mode, time mode, seizure (or trial) mode, number of factors and number of clusters. Then, $\mathbf{x}_{:t}$, resp. $\mathbf{x}_{l:}$, corresponds to the column t, resp. to the row l, of the matrix $\mathbf{X} \in \mathbb{R}^{L \times T}$. The matrices $\mathbf{X}_{l::}$, $\mathbf{X}_{:t:}$ or $\mathbf{X}_{::s}$ correspond to the slices of the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times T \times S}$. The slice for each trial $\mathbf{X}_{::\mathbf{s}}$ is referred to as an epoch.

The matricization corresponds to the matrix representation of a tensor. The mode-L matricization of $\boldsymbol{\mathcal{X}}$ is noted $\mathbf{X}_{(L)} \in \mathbb{R}^{L \times TS}$. Writing $[\mathbf{A}, \mathbf{B}]$ the concatenation of two matrices $\mathbf{A}$ and $\mathbf{B}$ with the same number of rows, the mode-1 matricization reads:

$$\mathbf{X}_{(L)} = [\mathbf{X}_{::1}, \mathbf{X}_{::2}, ..., \mathbf{X}_{::S}]. \tag{3}$$

Considering $\mathbf{x}, \mathbf{y} \in \mathbb{R}^L$, $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{l=1}^{L} x_l y_l$ is the scalar product between two vectors and $|| \mathbf{x} ||_F = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ the Frobenius norm. These notations can be extended to D-mode arrays,[5,6] e.g. $\langle \mathbf{X}, \mathbf{Y} \rangle = \langle vec(\mathbf{X}), vec(\mathbf{Y}) \rangle$ or $|| \boldsymbol{\mathcal{X}} ||_F = || vec(\mathbf{X}_{(L)}) ||_F$. The L_1-norm for a matrix corresponds to:

$$|| \mathbf{X} ||_1 = \sum_{l=1}^{L} \sum_{t=1}^{T} | x_{lt} |. \tag{4}$$

Finally we define the Kronecker product between two matrix, noted $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ with $\mathbf{A} \in \mathbb{R}^{L \times T}$, $\mathbf{B} \in \mathbb{R}^{K \times N}$ and $\mathbf{C} \in \mathbb{R}^{LK \times TN}$, as:

$$\mathbf{C} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1T}\mathbf{B} \\ \vdots & & \vdots \\ a_{L1}\mathbf{B} & \dots & a_{LT}\mathbf{B} \end{pmatrix}. \tag{5}$$

4. REDUCTION METHODS

In order to reduce the tensor data $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times T \times S}$ to a relevant factor matrix $\mathbf{F} \in \mathbb{R}^{L \times K}$, we propose the following decomposition :

$$\mathbf{X}_{(L)} \approx \mathbf{F}(\mathbf{w} \otimes \mathbf{V})^t \tag{6}$$

where $\mathbf{w} \in \mathbb{R}^S$ is a vector performing a weighted average of the S epochs in order to extract common patterns, $\mathbf{F}$ is the factor matrix and $\mathbf{V} \in \mathbb{R}^{T \times K}$ corresponds to the temporal activation profile of the different factors. $\mathbf{F}$ is scaled such that $|| \mathbf{w} ||_F^2 = || \mathbf{v}_{:k} ||_F^2 = 1$ (to remove scaling indeterminacy and transfer all the energy in our factor

matrix). Without constraints on $\mathbf{F}$, $\mathbf{V}$ and $\mathbf{w}$, the High Order Orthogonal Iteration (HOOI) or the proposed approach mHOOI can be used to perform this decomposition. However it as be shown in[3] that constraining $\mathbf{F}$ to be sparse appears to be important because, in comparison to the case without this constraint, $\mathbf{F}$ becomes closer to a cluster assignment matrix. Moreover, it limits the complexity of clusters by reducing the number of FC they contain. This is important in the context of epileptic data where a large number of FC measurements can be passively implied in a neurological process (during the discharge of the seizure for example). As well it seems relevant to also impose $\mathbf{V}$ to be sparse so as to select specific temporal steps of FC activation, eliminating thus period where there is no common activation of FC clusters. To perform this decomposition we proposed a new algorithm HOsSVD, this can also be done thanks to a state of the art method like the sparse Tucker decomposition proposed in.[7]

4.1 mHOOI

The proposed modified HOOI (noted mHOOI) is as follows, iterating 2 steps in alternance:

(A) At iteration $(i+1)$ we assume $\mathbf{W}_i^{(1)}$ to be known. From this vector we compute the common Epoch corresponding to $\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I}) \in \mathbb{R}^{L \times T}$, with $\mathbf{I} \in \mathbb{R}^{T \times T}$ the identity matrix (it corresponds to the mode-L matricization of the contraction product between the tensor $\boldsymbol{\mathcal{X}}$ and the vector $\mathbf{W}^{(1)t}$[5]). We then compute the low rank approximations $\mathbf{U}_{i+1}^{(K)}$ and $\mathbf{V}_{i+1}^{(K)}$ as the K first components of the singular value decomposition of $\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I}) = \mathbf{U}_{i+1}\mathbf{\Lambda}_{i+1}\mathbf{V}_{i+1}^t$.

(B) In order to update $\mathbf{W}_{i+1}^{(1)}$ we first filter the tensor $\boldsymbol{\mathcal{X}}$ by projecting it on the subspaces spanned by the matrices $\mathbf{U}_{i+1}^{(K)}$ and $\mathbf{V}_{i+1}^{(K)}$. $\mathbf{W}_{i+1}^{(1)}$ is then obtained by a special HOOI decomposition on the filtered tensor with $K_L = L$, $K_T = T$ and $K_S = 1$. The optimal decomposition can be found analytically. Both the filtering and the decomposition, in order to get $\mathbf{W}_{i+1}^{(1)}$, reduce to the problem of computing the dominant left singular vector of $\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K)} \otimes \mathbf{U}_{i+1}^{(K)})$.

We summarize in Algorithm 1, this modified HOOI method (mHOOI) and how to obtain the corresponding lower dimension factor matrix $\mathbf{F}$. We use an angular metric between two successive estimates of $\mathbf{W}^{(1)}$ as the stopping criterion of the iterative procedure.

4.2 HOsSVD

The previous algorithm requires several SVD and the final matrix F is not sparse in general. As we already noticed, sparsity constraints is wanted to get FC graphs with few edges, and because the SVD is impacted by the curse of dimensionality it is desirable to limit its use. Following the same rationale, and in order to perform a decomposition as proposed in Eq. (6), we develop a sparse version of mHOOI, where we replace the computation of the SVD in step **(A)**, by the sparse SVD :

$$\underset{\mathbf{F},\mathbf{Z}}{\operatorname{argmin}} \; || \mathbf{X}_{(L)} - \mathbf{F}\mathbf{Z}^t ||_F^2 + \gamma_1 || \mathbf{F} ||_1, \tag{7}$$

$$s.t. \qquad \gamma_2 || \mathbf{Z}_{:k} ||_1 + || \mathbf{Z}_{:k} ||_F^2 \leq 1 \tag{8}$$

where the meta-parameters γ_1 and γ_2 allow for tuning the trade-off between accuracy and sparsity of the approximation. To solve this matrix factorization (7), we use the SPAMS library on matlab.[8] We call this decomposition a High Order sparse SVD (HOsSVD). Since the matrix $\mathbf{U}_{i+1}$ is not necessarily orthogonal anymore, we compute the subspace spanned by this matrix by performing its QR decomposition, and we retain only the first K column vectors of the $\mathbf{Q}$ part. The same goes for $\mathbf{V}_{i+1}$. Writing $qr(\mathbf{V}_{i+1}, K)$ this operation, Algorithm 2 presents the computation of $\mathbf{F}$, the lower dimension factor matrix stemming from this high order sparse SVD reduction (HOsSVD).

4.3 State of the art methods

Both proposed approach mHOOI and HOsSVD will be compared by state of the art approach. Notice the following decomposition :

$$\underset{\mathbf{U},\mathbf{V},\mathbf{W}}{\operatorname{argmin}} \; || \mathbf{X}_{(L)} - \mathbf{U}\mathbf{G}_{(L)}(\mathbf{W} \otimes \mathbf{V})^t ||_F^2, \tag{9}$$

Where $\mathbf{U} \in \mathbb{R}^{L \times K_L}$, $\mathbf{V} \in \mathbb{R}^{T \times K_T}$, $\mathbf{W} \in \mathbb{R}^{L \times K_S}$ and $\mathbf{G}_{(L)} \in \mathbb{R}^{K_L \times K_T K_S}$ is a dense matrix. The High order orthogonal Iteration (HOOI) algorithm provide a solution to this optimisation problem (eq. 9).[9,10] Now by noting $\mathbf{F} = \mathbf{U}\mathbf{G}_{(L)}$ in the case $K_L = K$, $K_T = K$ and $K_S = 1$ we get a solution of the decomposition (eq. 6) similar to the mHOOI algorithm 1.

In,[7] they propose a decomposition which constraints matrices $\mathbf{U}$, $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{G}_{(L)}$ to be non negative and optionally sparse. By fixing $K_L = K$, $K_T = K$ and $K_S = 1$ and imposing sparsity on $\mathbf{U}$ and $\mathbf{V}$ we get a decomposition similar to HOsSVD. The factor matrix will be $\mathbf{F} = \mathbf{U}$ after scaling $\mathbf{U}$ such as $|| \mathbf{w} ||_F^2 = || \mathbf{v}_{:k} ||_F^2 = \mathbf{G}_{(L)} : k = 1$ for all $k \in 1, ..., K$. This decomposition will is called SN-Tucker.

Algorithm 1 Estimation of $\mathbf{F}$ via mHOOI

Require: $\boldsymbol{\mathcal{X}}$, K, angular tolerance $\epsilon > 0$, et i_{max}.
  $i = 0$
  $\mathbf{W}_0^{(1)} = left\_1\_SVD(\mathbf{X}_{(3)})$
  while $i < i_{max}$ or $acos(< \mathbf{W}_i^{(1)}, \mathbf{W}_{i-1}^{(1)} >) > \epsilon$ **do**
    (A). $[\mathbf{U}_{i+1}^{(K)}, \mathbf{\Lambda}_{i+1}, \mathbf{V}_{i+1}^{(K)}] = K\_SVD(\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I}))$
    (B). $\mathbf{W}_{i+1}^{(1)} = left\_1\_SVD(\mathbf{X}_{(S)}(\mathbf{V}_{i+1}^{(K)} \otimes \mathbf{U}_{i+1}^{(K)}))$
    $i = i + 1$
  end while
  $\mathbf{F} = \mathbf{U}_i^{(K)} \mathbf{\Lambda}_i$

Algorithm 2 Estimation of $\mathbf{F}$ via HOsSVD

Require: $\boldsymbol{\mathcal{X}}$, K, γ_1, γ_2, angular tolerance $\epsilon > 0$, et i_{max}.
  $i = 0$
  $\mathbf{W}_0^{(1)} = left\_1\_SVD(\mathbf{X}_{(3)})$
  while $i < i_{max}$ or $acos(< \mathbf{W}_i^{(1)}, \mathbf{W}_{i-1}^{(1)} >) > \epsilon$ **do**
    (A). $\mathbf{U}_{i+1}$, $\mathbf{V}_{i+1}$ minimising (7) with $\mathbf{X}_{(L)}(\mathbf{W}_i^{(1)} \otimes \mathbf{I})$.
    $\mathbf{A} = qr(\mathbf{U}_{i+1}, K)$, $\mathbf{B} = qr(\mathbf{V}_{i+1}, K)$
    (B). $\mathbf{W}_{i+1}^{(1)} = left\_1\_SVD(\mathbf{X}_{(S)}(\mathbf{B} \otimes \mathbf{A}))$
    $i = i + 1$
  end while
  $\mathbf{F} = \mathbf{U}_i^{(K)}$

5. RESULTS AND APPLICATIONS

In this section we present different empirical results on the convergence speed of mHOOI, the convergence of HOsSVD and his performances on a model of FC. First, we present the model, then the different empirical results. Finally the HOsSVD method is applied on real epileptic data.

Matlab code for HOOI, mHOOI, and HOsSVD can be found here *. These codes uses the MATLAB Tensor toolbox Version 2.6 † and the sparse SVD is done using the SPAMS toolbox version 2.6 ‡? Finally the SN-Tucker algorithm come from the toolbox §

5.1 Model

Epileptic seizure implies a pathological FC that starts in a focal brain onset, then spreads to the other connected regions, and sometimes split to give rise to new FC components. Fig. 3a displays a characteristic example of actual FC time series measured by the phase lock value (PLV).[2] The model we propose is aimed at reproducing the global structured pattern of FCs' activation, and the uncertainties of the measures. More precisely, we consider a matrix $\mathbf{X} \in \mathbb{R}^{L \times T}$, where x_{lt} is set to a high value if the FC of index $l \in \{1, \ldots, L\}$ is active at time $t \in \{1, \ldots, T\}$, and to a low value otherwise (to account for the non ON-OFF discrepancy of the PLV measurement, we choose values equal to 0.7 and 0.2, respectively). A cluster C_n, $n \in \{1, \ldots, N\}$, is composed of all FCs that are activated over the same period of time T_n, figure 3b present clusters of FC activation. This is for the deterministic part of

*Algorithms used in this article: FCTensDec[3]
†MATLAB Tensor Toolbox Version 2.6[5]
‡SPAMS toolbox version 2.6[8]
§www2.imm.dtu.dk/pubdb/views/edoc_download.php/4718/zip/imm4718.zip[7]

the model, defining the structural and temporal pattern of FCs activation, common to all seizures. Superimposed to it, we add a seizure dependant random component, composed of four uncertainty sources:

i Random duration: each activation period T_n is uniformly distributed between a minimum duration (here, 5 time steps) and $\frac{T}{N}$ time steps (the binary variable α allows to able (1) or disable (0) this random mode).

ii Activation error: with probability $\beta \in [0, 1]$, each FC of a given group C_n incurs the risk to be replaced by any other randomly chosen FC.

iii Connectivity noise: we add to x_{lt} a Uniform noise in the interval $[0, \sigma]$ if the FC is inactivated, $[1 - \sigma, 1]$ else.

iv Jitter: All FCs of the same group start activating with independent jitters, uniformly distributed in $\left[-\frac{\delta}{2}, \frac{\delta}{2}\right]$.

Figure 3c displays one realization of these synthetic FC time series. Compared to real data, the model succeeds in reproducing a realistic global pattern. More importantly, as it allows a control of the nature and the intensity of variability between epochs, it will serve to evaluate the sensitivity and the robustness of the different dimension reduction methods for clustering, with respect to each source of uncertainty.

To this end, we simulate different seizures of a same patient as i.i.d. realizations of our model with the same set of parameters $b = [\alpha, \beta, \sigma, \delta]$. As for real data, the S modeled seizures are then stacked in a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{L \times T \times S}$. Figure 4 displays 4 realisations of the model $\mathbf{X}_{::i}$ corresponding to the particular choice $[\alpha = 1, \beta = 0.2, \sigma = 0.6, \delta = 0.3]$, $L = 66$, $T = 100$ and $S = 4$.

5.2 On the convergence speed of mHOOI vs HOOI

To empirically observe the convergence towards the optimal solution, with faster performance than HOOI algorithm we propose the following experiment. We consider tensors from the model presented above, with different configuration of noise vector. We fix $K_L = K_T = 3$ and $K_S = 1$. We compare the performance of both HOOI and mHOOI algorithm by computing the total variance at each iteration $tvar(t) = || \mathbf{G}_i ||_F^2$ for $HOOI$, $tvar(t) = || \mathbf{F}_i ||_F^2$ (since they are matrices containing all the variance) and the log differential of the total variance between two consecutive iterates $ldvar(t) = log(tvar(t) - tvar(t-1)))$. Fig. 5 shows the mean of $tvar(t)$ for the first 100 iterations using 100 realisations of the tensor model $\boldsymbol{\mathcal{X}}$ associated tot he noise vector b_i as input of the algorithm. A zoom of the last iteration is provided as well as the first value of $ldvar(t)$ is shown. Notice the lack of value of $ldvar$ is due to the fact that in average $tvar(t) - tvar(t-1)$ is negative, however it is not relevant since it only happen when this difference attains the precision of the machine. The new proposed algorithm has always better performances than HOOI, with exponential convergence, on every scenario.

5.3 On the convergence of HOsSVD

We consider again the model presented above, so as to study the convergence of the HOsSVD alone. Since the objective of SN-Tucker and HOsSVD are not exactly the same it is useless to compare them like mHOOI and HOOI.

For a set of hyper parameters γ_1 and γ_2 we compute the first 100 iterations using 100 realisations of the tensor model $\boldsymbol{\mathcal{X}}$ associated to the noise vector b_i as input of the HOsSVD algorithm (2). We say that the algorithm converge on 100 iteration if the angular tolerance criterion is attain at the 100^{th} iteration or before. The table 1 shows for different noise vector b_i and hyper parameters the percentage over 120 experiments that converge on 100 iterations. Few situations lead to potential no convergence of HOsSVD algorithm: mostly cases b_2 with $\gamma_1 = 1$, and b_9 with $\gamma_1 = 1$ converge $\approx 95\%$ to $\approx 98.5\%$ of the time. There is two kind of non convergence, the first kind happening twice for b_2 with $\gamma_1 = 1$, $\gamma_2 = 1$ is due to high sparsity constraints that lead to decomposition in null matrices. The second kind of non convergence happen more frequently, where the algorithm oscillates between several solutions. The figure 6 shows the convergence critera for 12 realisations with the noise vector b_{10} and $\gamma_1 = 1$, $\gamma_2 = 1$, where we can observe the second kind of non convergence. Notice convergence always happen where there is connectivity noise (which is always the case, with different probability distribution, in real dataset).

Finally in order to show that the convergence of this algorithm is fast, the table 1 shows for different noise vector b_i and hyper parameters the percentage over 130 experiments that converge on 5 iterations. We observe fast

convergence when there is connectivity noise in the model. The figure 7 shows the logarithm of the convergence critera for the same 12 realisations as the figure 6 highlighting the exponential convergence of the algortihm as the mHOOI algorithm (notice the algorithm is stopped when the angular tolerance criteria is attained, explaining the stationary phase).

5.4 On the Performances of HOsSVD vs SN-Tucker

5.4.1 Experimental setup

Using the FC model, we compare HOsSVD and SN-Tucker methods derived in the previous section. Each method yields a lower dimension factor matrix $\mathbf{F}$ that serves as an input to k-means. Performance refer here to the ability at retrieving the FC clusters of the global pattern. We use a set of hyper-parameters values to compare both method (HOsSVD and SN-Tucker), notice the effect of these parameters does not impact both methods equally, however they are chosen to cover the principal behaviours of these algorithm. We set K to 3 since there are 3 activated FC clusters in the model. Regarding k-means algorithm, we used k-means++ version[11] that we stopped after 1000 iterations and repeated 120 times with different seeds. The number of sought groups is set to $N = 4$. Finally, to assess the clustering performance of each method, we use the Adjusted Rand Index (ARI) score,[12] computed between the resulting grouping and the ground truth figure 3b: the score equals 1 for a perfect match and 0 if the correspondence does not outperform a random grouping.

5.4.2 Result

Under these experimental conditions, we evaluate all methods for various configurations of uncertainty and hyper-parameters. Fig. 8 displays the ARI scores (mean and variance estimated out of 120 independent realizations of $\boldsymbol{\mathcal{X}}$) for 10 different combinations of (i) random duration, (ii) activation error and (iii) connectivity and (iv) jitter noise. For each noise configuration, we highlight the hyper-parameters which give the best performance for both method by a horizontal line.

HOsSVD outperform SN-Tucker for all configurations considering the best performance of each method along the set of hyper-parameters. Also we can observe more variability of the ARI score using HOsSVD methods with different hyper parameters, showing a stronger impact of the hyper parameters on the final result. in fact, for high constraints of sparsity, our proposed approach seems to give more relevant decompositions than SN-Tucker, indeed it is less likely to get identical factors using HOsSVD than SN-Tucker. This can be shown figure 9 where for a realisation on the noise configuration $b = [0, 0, 0.7, 0.3]$ we show the factor matrix F for every hyper parameters using HOsSVD method, we can see for high constraints of sparsity non equivalent factors and a result close to an assignment cluster matrix. Contrary to the previous case, the factor matrix F obtained using SN-Tucker method does not seem relevant, with null factors for high constraints of sparsity.

5.5 Application on real data

5.5.1 Data

We consider real iEEG data from 3 patients with focal epilepsy.[1,13] The electrodes used are distributed on stems implanted in the brain. The activity of the brain is recorded via 5 to 10 electrodes per stem. The space between two consecutive electrodes is 3.5 mm. Over a recording time of 15 days, 4 seizures were selected for the first patient, 4 for the second patient and 2 for the third. We only have clinical knowledge of the seizure for the first patient. Each seizure is delimited in time by a window of 100 seconds centered on the beginning of the seizures. The signal is sampled at 256 Hz. Only a thirds of the initial contacts are selected to avoid too strong spatial correlations. The functional connectivity metric used is PLV.[2] A strong PLV between two signals means that their phases are similar. The FC (corresponding to each pair of electrodes) were calculated over a sliding rectangular window of 4 seconds duration, with a time step of one second. After eliminating the points that suffer from border effects, the data is formatted as a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{528\times96\times4}$ for the first patient, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{1035\times96\times4}$ for the second and $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{946\times96\times2}$ for the third. The FCs of each seizure are plotted in 11, 12 and 13 respectively for the patient 1, 2 and 3.

5.5.2 Methods

From the tensor $\boldsymbol{\mathcal{X}}$, we obtain the matrix $\mathbf{F}$ (using Algorithm 2), where the parameters, $K = 4$, $\epsilon = 10^{-3}$ and $\gamma_1 = 1$, $\gamma_2 = 1$ for the first patient, $\gamma_1 = 0.8$, $\gamma_2 = 1.6$ for the second patient, $\gamma_1 = 1.2$, $\gamma_2 = 0.8$ for the third patient are empirically fixed (It is current to consider between 3 and 5 different steps in an epileptic seizure justifying $K = 4$; for γ_1 and γ_2 we search for the highest value of parameters where results seem coherent). Fig. 14, 15, 16 (top) show the temporal activation profiles of the components of $\mathbf{F}$ (corresponding to the matrix $\mathbf{V}^{(K)}$ from Algorithm 2). We apply k-means to $\mathbf{F}$ to identify the $N = 5$ corresponding FC groups. Since $\mathbf{F}_{HOsSVD}$ is close to an assignment matrix (a solution $\mathbf{A}^*$ of Eq. (2)), we retain the 4 groups of smaller sizes, which can be associated with the 4 activation periods while the the 5th group corresponds to the unsynchronized FCs. Figure 14, 15, 16 (bottom) materialize the positions of the 33 electrodes projected on the transverse plane (according to the repair of Tailarach) for each patient. FC is represented by a link between the pair of electrodes that are in phase.

5.5.3 Results

Patient 1 with 4 epileptic seizures and clinical knowledge Fig. 14: There are 4 activation periods that can be easily associated with 4 steps of the seizures: before seizure, start, propagation and end of the seizure. The time interval around time 50s is particularly interesting as it shows no activated FC. This is likely to correspond to the functional decoupling at the early start of the seizure, a short period when iEEG activities in different areas of the brain are suddenly decorrelated. The four FC groups can be associated to four snapshots of a time-varying graph: before the seizure, only two electrodes interact in what could correspond to the epileptogenic zone. At the beginning of the seizure: spreading of FC activation with the appearance of a cluster of FC localized around the epileptogenic zone. During the crisis, other FC appear spontaneously in the other hemisphere, at the same time the FC are diffused in the left hemisphere. At the end of the crisis, the two hemispheres interact with the appearance of common FC. The graphs obtained are in agreement with the clinical results, the focus of the beginning of the seizure being close to the graphs "before-seizure" and "seizure start". The propagation of seizures in the right hemisphere of the brain is well represented by the graphs "Propagation" and "Seizure end".

Patient 2 with 4 epileptic seizures without clinical knowledge Fig. 15: As well as the previous patient 4 activation periods that can be easily associated with 4 steps of the seizures: before seizure, start, propagation and end of the seizure. There is more period without activation, certainly because FC patterns of each seizure seem more erratic than the previous patient. The four FC groups can as well be associated to four snapshots of a time-varying graph: at step 1 only two electrodes interact, at step 2 and 3 spreading of FC activation in all the brain. The last step present we return to a situation close to the initial one.

Patient 3 with 2 epileptic seizures without clinical knowledge Fig 16: Contrary to the two previous seizure, it is more difficult to distinct activation period, this can be explained by the fact that we only have access to two epileptic seizure for this patient, which mean less robust results. However the graph corresponding to each activation are localized spatially, which indicate a coherency in the obtained results.

6. CONCLUSION

We have presented new tensor decompositions with original algorithmic implementation. We shown through a general model of FC the fast empirical convergence of these algorithms compared to the already existing HOOI. The advantage of using HOsSVD instead of a state of the art tensor non negative sparse tensor decomposition to find sparse FC common to each seizure is exhibited. Finally, applied to real iEEG data recorded during an epileptic seizure, our method allowed us to identify FC activation groups corresponding to significant time periods of the evolution of the seizure.

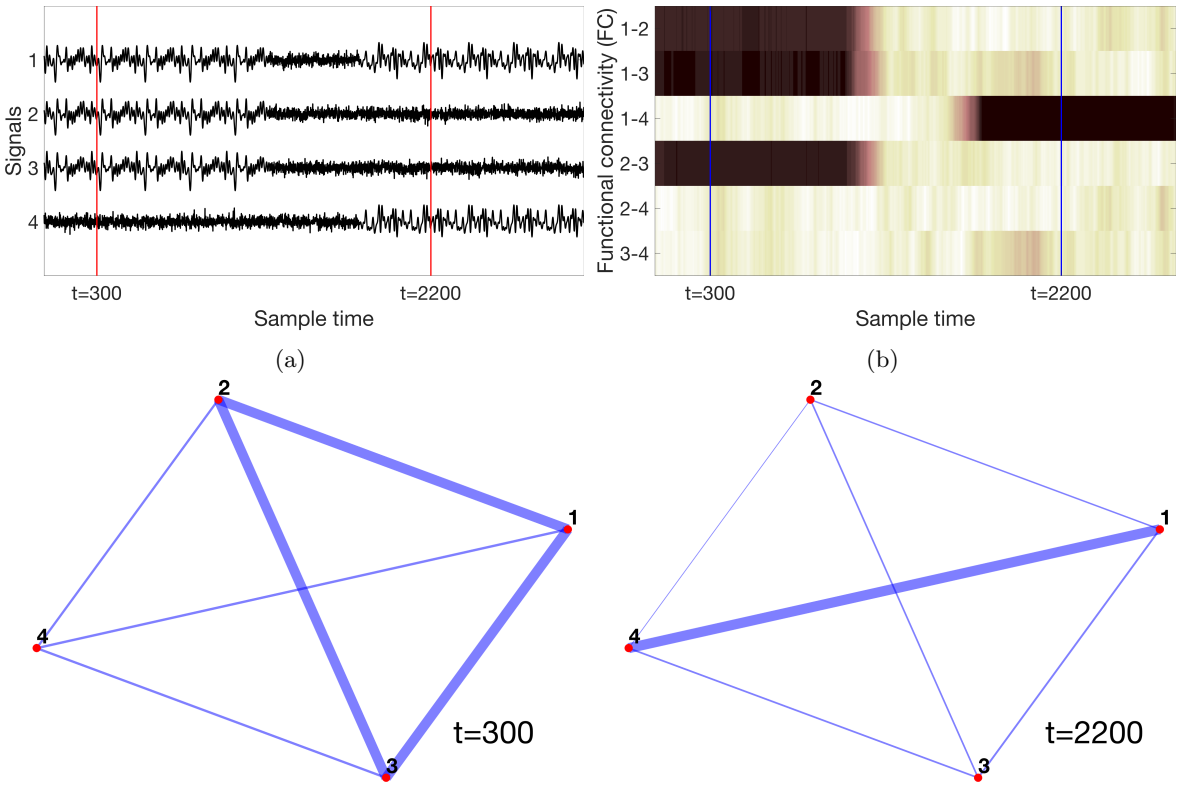
As a perspective of this work, we would like to generalise the proposed tensor reduction to the general case where the seizure mode can be decomposed in several factor ($K_S > 1$).

7. ACKNOWLEDGEMENT

This work was supported by the ACADEMICS grant of the IDEXLYON, project of the Université de Lyon, PIA operated by ANR-16-IDEX-0005.

Matlab code from[3] including HOsSVD and mHOOI algorithm can be found here [¶]:

[¶] FCTensDec



(a) (b)

(c) (d)

Figure 1: (a) 4 simulated iEEG signals, (b) FC computed between the 6 pairs of signals by PLV, (c) & (d) Graphs with electrodes 1,2,3 and 4 as nodes and PLV measures as the weight of edges, at t=300 and t=2200 (red/blue bars on Fig. (a) and (b)).

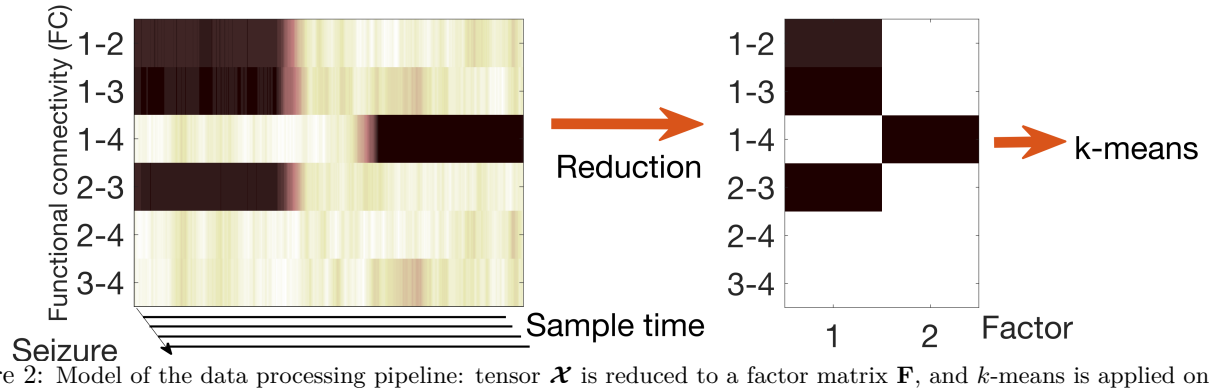
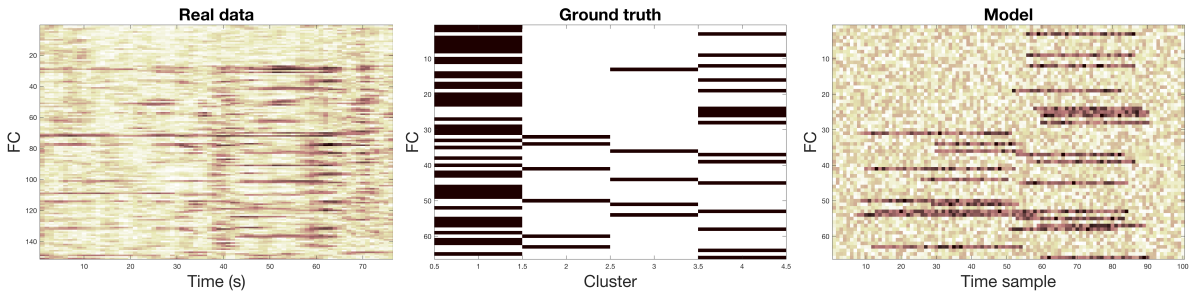
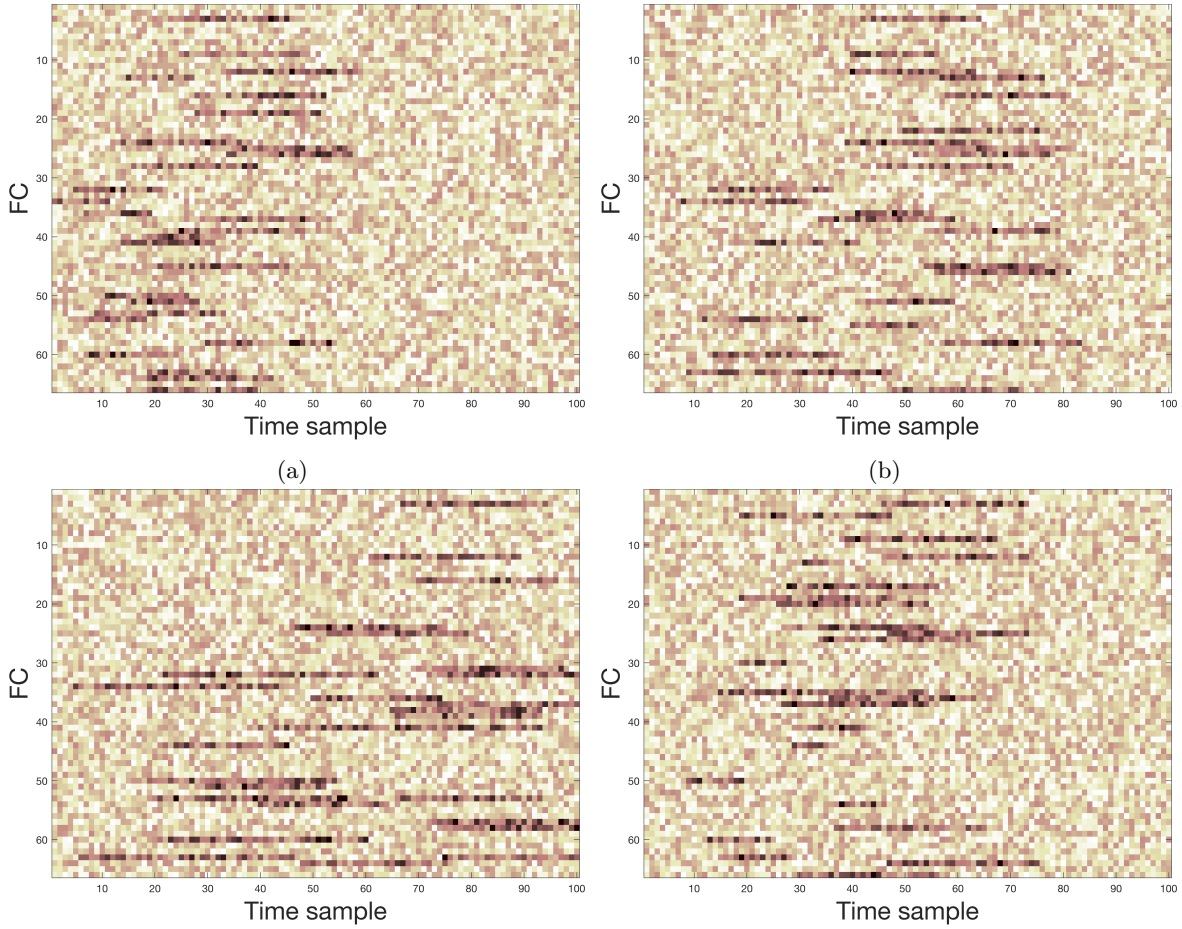


Figure 2: Model of the data processing pipeline: tensor $\boldsymbol{\mathcal{X}}$ is reduced to a factor matrix $\mathbf{F}$, and k-means is applied on $\mathbf{F}$



(a) (b) (c)

Figure 3



(a) (b)

(c) (d)

Figure 4: Example of 4 realisation our model with $b = [1, 0.2, 0.6, 0.3]$

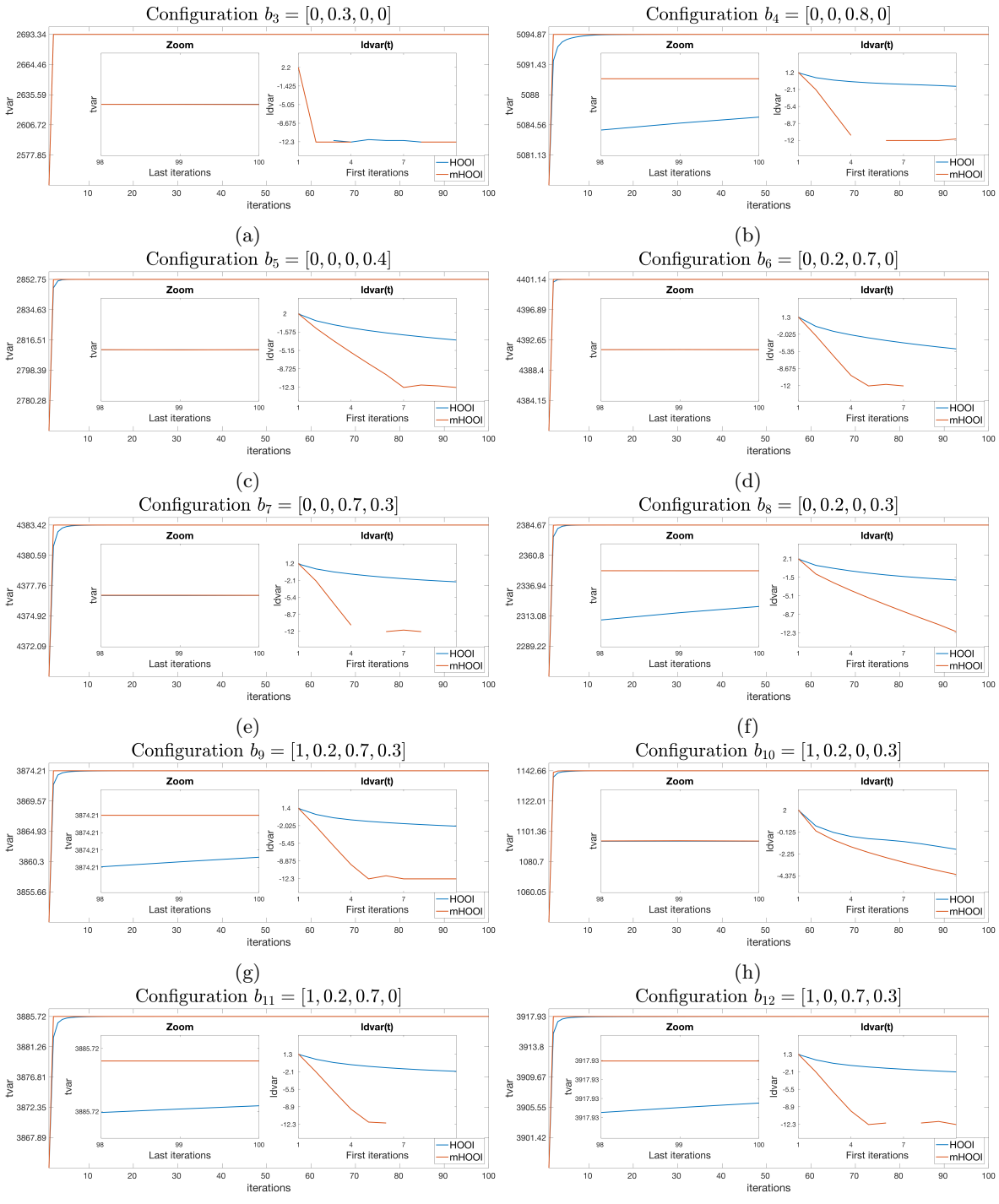


Figure 5: $tvar(t)$ for the first 100 iterations using the tensor $\boldsymbol{\mathcal{X}}_{model}$ as input of the algorithm with a zoom of the last iteration and the first value of $ldvar(t)$ for different configurations

	$[0.1, 0.1]$	$[0.1, 2]$	$[2, 0.1]$	$[0.4, 0.4]$	$[0.4, 1]$	$[1, 0.4]$	$[0.7, 0.7]$	$[1, 0.7]$	$[0.7, 1]$	$[1, 1]$
$b_1 = [0, 0, 0, 0]$	100	100	100	100	100	100	100	100	100	100
$b_2 = [1, 0, 0, 0]$	100	100	100	100	100	100	100	100	100	98.48
$b_3 = [0, 0.3, 0, 0]$	100	97.73	100	100	100	98.48	100	100	100	100
$b_4 = [0, 0, 0.8, 0]$	100	100	100	100	100	100	100	100	100	100
$b_5 = [0, 0, 0, 0.4]$	100	100	100	100	100	100	100	99.24	100	100
$b_6 = [0, 0.2, 0.7, 0]$	100	100	100	100	100	100	100	100	100	100
$b_7 = [0, 0, 0.7, 0.3]$	100	100	100	100	100	100	100	100	100	100
$b_8 = [0, 0.2, 0, 0.3]$	100	100	100	100	100	100	100	99.24	100	100
$b_9 = [1, 0.2, 0.7, 0.3]$	100	100	100	100	100	100	100	100	100	97.73
$b_{10} = [1, 0.2, 0, 0.3]$	100	100	100	100	99.24	99.24	100	96.21	100	94.7
$b_{11} = [1, 0.2, 0.7, 0]$	100	100	100	100	100	100	100	100	100	99.24
$b_{12} = [1, 0, 0.7, 0.3]$	100	100	100	100	100	100	100	100	100	100

Table 1: % of 120 experiment which converge at least before **100** iteration using HOsSVD algorithm with parameters γ_1, γ_2 and with the noise configuration b_i.

	$[0.1, 0.1]$	$[0.1, 2]$	$[2, 0.1]$	$[0.4, 0.4]$	$[0.4, 1]$	$[1, 0.4]$	$[0.7, 0.7]$	$[1, 0.7]$	$[0.7, 1]$	$[1, 1]$
$b_1 = [0, 0, 0, 0]$	100	100	100	100	100	100	100	100	100	100
$b_2 = [1, 0, 0, 0]$	65.91	24.24	34.09	31.06	12.88	15.91	10.61	11.36	10.61	10.61
$b_3 = [0, 0.3, 0, 0]$	91.67	14.39	2.27	16.67	0.76	2.27	8.33	9.85	9.85	8.33
$b_4 = [0, 0, 0.8, 0]$	100	100	100	100	100	100	100	100	100	100
$b_5 = [0, 0, 0, 0.4]$	100	99.24	97.73	99.24	97.73	98.48	96.97	96.97	96.97	92.42
$b_6 = [0, 0.2, 0.7, 0]$	100	100	100	100	100	100	100	100	100	91.67
$b_7 = [0, 0, 0.7, 0.3]$	100	100	100	100	100	100	100	100	100	100
$b_8 = [0, 0.2, 0, 0.3]$	48.48	22.73	24.24	27.27	15.15	26.52	27.27	14.39	15.91	10.61
$b_9 = [1, 0.2, 0.7, 0.3]$	100	100	100	100	100	100	100	97.73	94.7	7.58
$b_{10} = [1, 0.2, 0, 0.3]$	2.27	1.52	0.76	1.52	0.76	0.76	0.76	0	0	0
$b_{11} = [1, 0.2, 0.7, 0]$	100	100	100	99.24	100	100	100	76.52	72.73	9.09
$b_{12} = [1, 0, 0.7, 0.3]$	100	100	100	100	100	100	100	96.21	94.7	30.3

Table 2: % of 120 experiment which converge at least before **5** iteration using HOsSVD algorithm with parameters γ_1, γ_2 and with the noise configuration b_I.

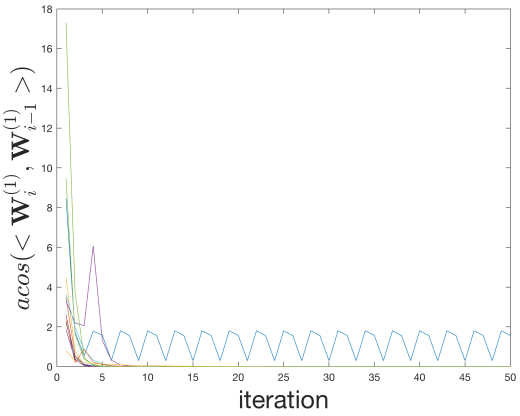


Figure 6: Convergence criteria for 12 realisation the experiment configuration b_{10} and $[\gamma_1, \gamma_2] = [1, 1]$.

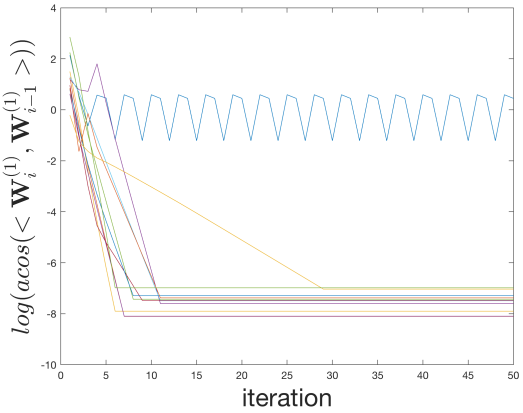


Figure 7: Log of the convergence criteria for 12 realisation of the experiment configuration b_{10} and $[\gamma_1, \gamma_2] = [1, 1]$.

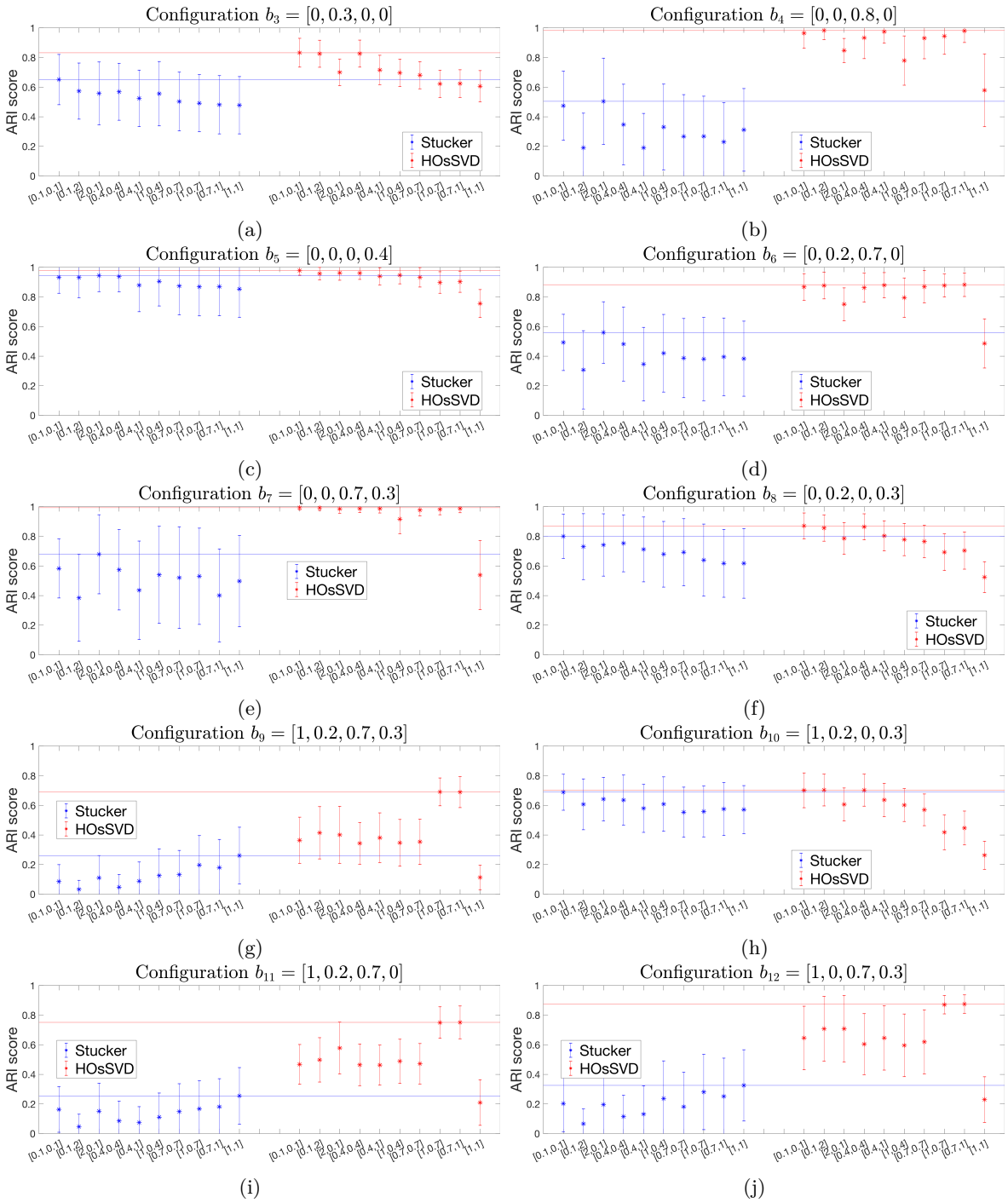


Figure 8: ARI score using SN-Tucker and HOsSVD algorithme in function of the sparsity parameters and the configurations

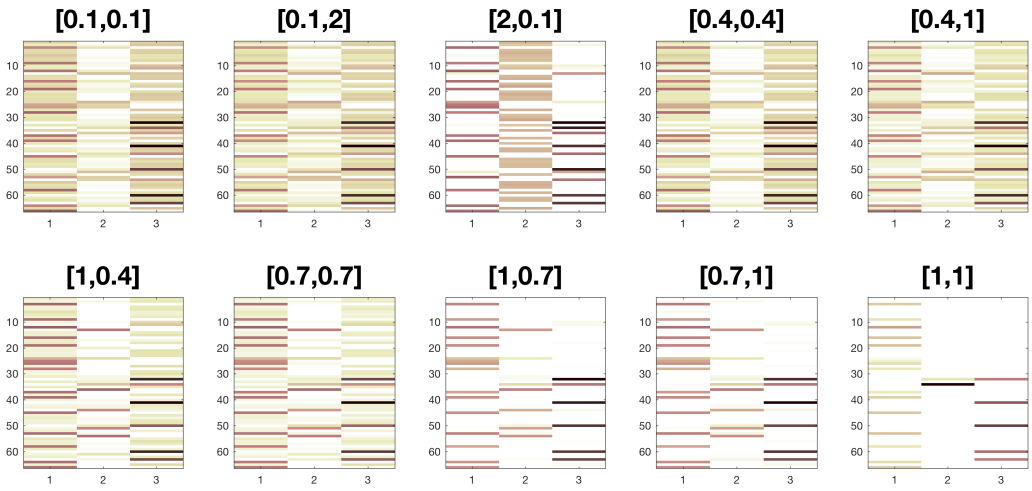


Figure 9: F reduction for all sparsity configurations using HOsSVD

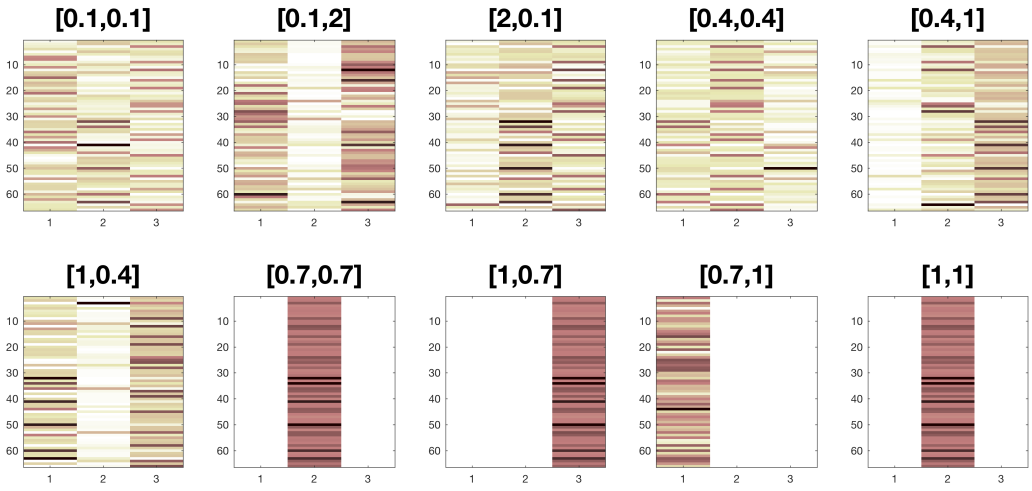
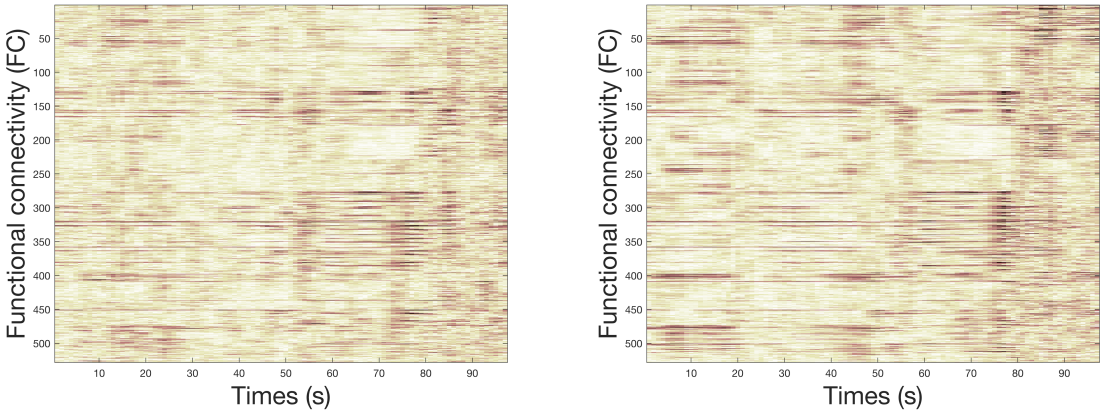
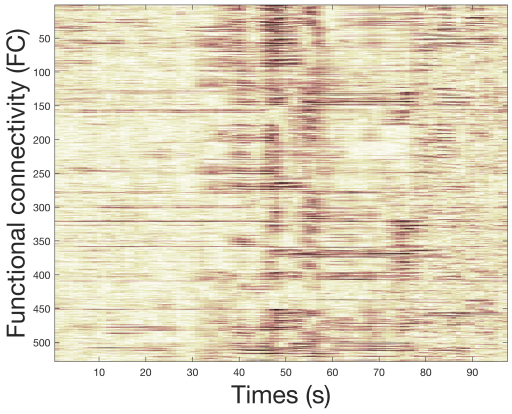


Figure 10: F reduction for all sparsity configurations using SN-Tucker

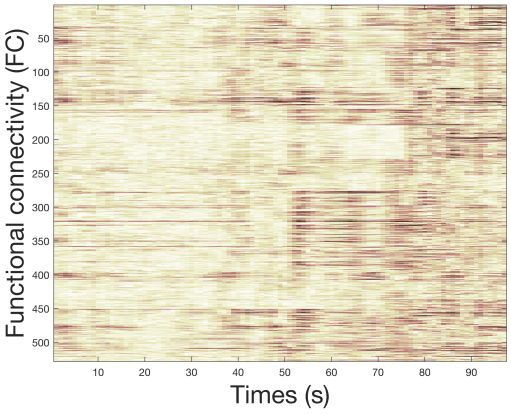


(a)

(b)

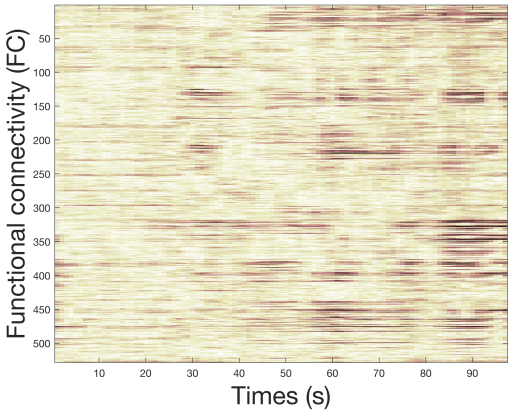


(c)

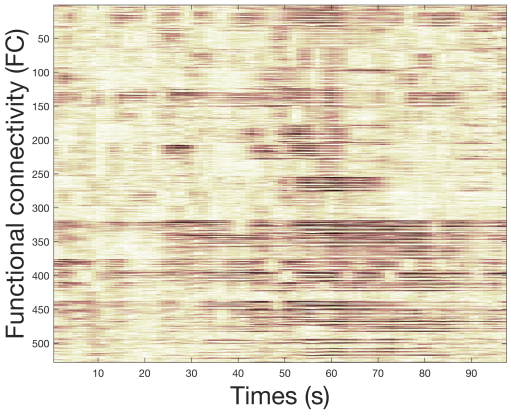


(d)

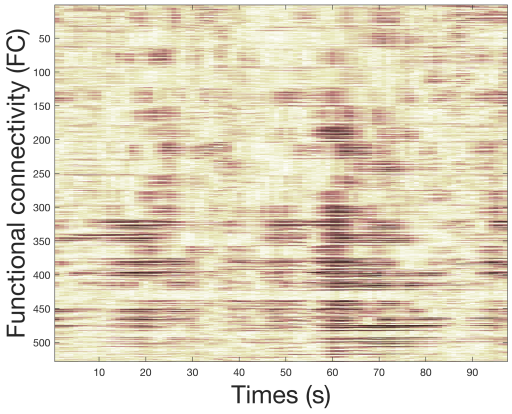
Figure 11: Functional connectivity of all 4 seizures for the patient 1



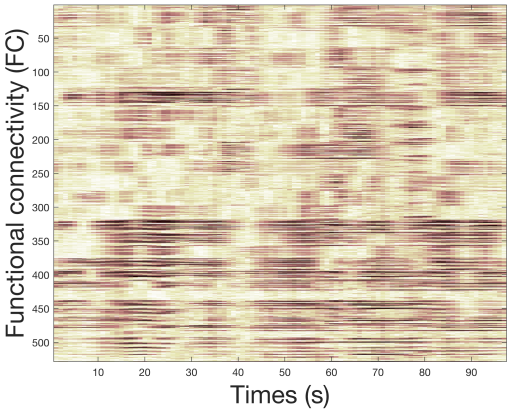
(a)



(b)

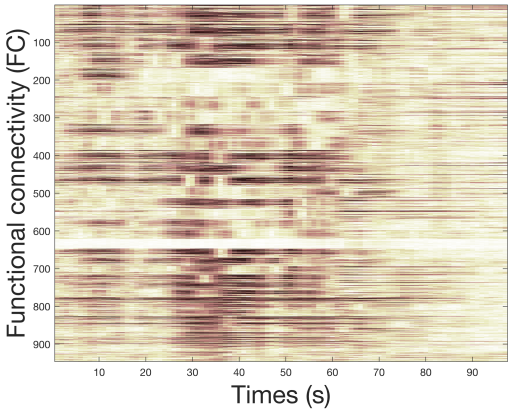


(c)

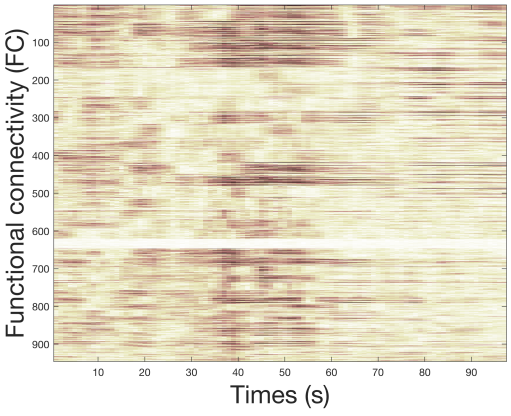


(d)

Figure 12: Functional connectivity of all 4 seizures for the patient 2



(a)



(b)

Figure 13: Functional connectivity of the 2 seizures for the patient 3

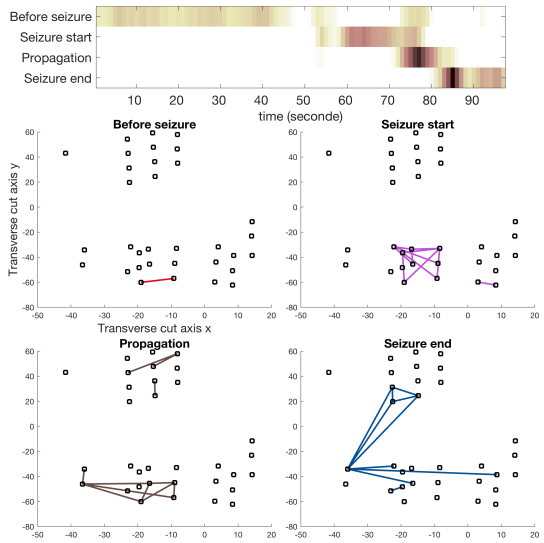


Figure 14: patient **1**) : (top) activation profile of HOsSVD (variable **V** of algorithm 2 (bottom) Cluster of FC corresponding to the 4 activation steps of the seizures

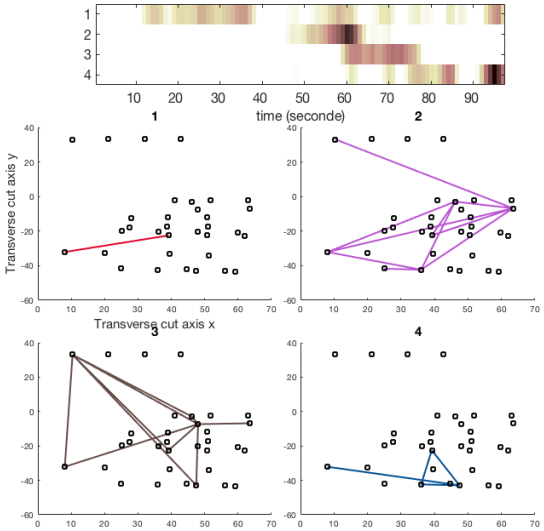


Figure 15: patient **2**) :(top) activation profile of HOsSVD (variable **V** of algorithm 2 (bottom) Cluster of FC corresponding to the 4 activation steps of the seizures.

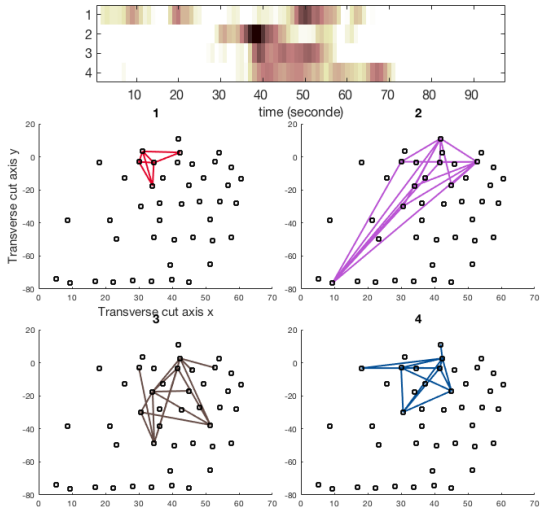


Figure 16: patient **3**) : (top) activation profile of HOsSVD (variable **V** of algorithm 2 (bottom) Cluster of FC corresponding to the 4 activation steps of the seizures.

REFERENCES

[1] Guenot, M., Isnard, J., Ryvlin, P., Fischer, C., Ostrowsky, K., Mauguiere, F., and Sindou, M., "Neurophysiological monitoring for epilepsy surgery: the talairach seeg method," *Stereotactic and functional neurosurgery* **77**(1-4), 29–32 (2001).

[2] van Mierlo, P., Papadopoulou, M., Carrette, E., Boon, P., Vandenberghe, S., Vonck, K., and Marinazzo, D., "Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization," *Progress in neurobiology* **121**, 19–35 (2014).

[3] Frusque, G., Jung, J., Borgnat, P., and Gonçalves, P., "Sparse tensor dimensionality reduction with application to clustering of functional connectivity," (2019).

[4] Jain, A. K., "Data clustering: 50 years beyond K-means," *Pattern recognition letters* **31**(8), 651–666 (2010).

[5] Kolda, T. G. and Bader, B. W., "Tensor decompositions and applications," *SIAM review* **51**(3), 455–500 (2009).

[6] De Lathauwer, L., De Moor, B., and Vandewalle, J., "A multilinear singular value decomposition," *SIAM journal on Matrix Analysis and Applications* **21**(4), 1253–1278 (2000).

[7] Mørup, M., Hansen, L. K., and Arnfred, S. M., "Algorithms for sparse nonnegative tucker decompositions," *Neural computation* **20**(8), 2112–2131 (2008).

[8] Mairal, J., Bach, F., Ponce, J., Sapiro, G., Jenatton, R., and Obozinski, G., "SPAMS: A SPArse Modeling Software, v2. 3," *URL http://spams-devel. gforge. inria. fr/downloads. html* (2014).

[9] Zhang, A. and Xia, D., "Tensor SVD: Statistical and Computational Limits," *IEEE Transactions on Information Theory* (2018).

[10] De Lathauwer, L., De Moor, B., and Vandewalle, J., "On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors," *SIAM journal on Matrix Analysis and Applications* **21**(4), 1324–1342 (2000).

[11] Arthur, D. and Vassilvitskii, S., "k-means++: The advantages of careful seeding," in [*Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*], 1027–1035, Society for Industrial and Applied Mathematics (2007).

[12] Hubert, L. and Arabie, P., "Comparing partitions," *Journal of classification* **2**(1), 193–218 (1985).

[13] Chauvel, P., Rheims, S., McGonigal, A., and Kahane, P., "French guidelines on stereoelectroencephalography (seeg): Editorial comment.," *Neurophysiologie clinique= Clinical neurophysiology* **48**(1), 1 (2018).