

# Robust Multi-dimensional Model Order Estimation Using LineAr Regression of Global Eigenvalues (LaRGE)

Alexey A. Korobkov, Marina K. Diugurova, Jens Haueisen, *Member, IEEE*, and Martin Haardt, *Fellow, IEEE*

**Abstract**—The efficient estimation of an approximate model order is very important for real applications with multi-dimensional data if the observed low-rank data is corrupted by additive noise. In this paper, we present a novel robust method for model order estimation of noise-corrupted multi-dimensional low-rank data based on the LineAr Regression of Global Eigenvalues (LaRGE). The LaRGE method uses the multi-linear singular values obtained from the HOSVD of the measurement tensor to construct global eigenvalues. In contrast to the Modified Exponential Test (EFT) that also exploits the approximate exponential profile of the noise eigenvalues, LaRGE does not require the calculation of the probability of false alarm. Moreover, LaRGE achieves a significantly improved performance in comparison with popular state-of-the-art methods. It is well suited for the analysis of biomedical data. The excellent performance of the LaRGE method is illustrated via simulations and results obtained from EEG recordings.

**Index Terms**—eigenvalue, global eigenvalue, tensor, the rank of the tensor, the model order of multi-dimensional data

## I. INTRODUCTION

**M**ULTI-DIMENSIONAL MODELS are widespread in a variety of applications, for example, radar, sonar, channel modeling in wireless communications, image processing, the estimation of MIMO channels parameters, blind source separation and many more [1]. According to these models the measured signals or the data can be stacked into multi-dimensional arrays or tensors. Moreover, in biomedical data processing multi-dimensional models have been widely used recently. For example, biomedical signals like Electroencephalograms (EEG), Magnetoencephalograms (MEG) or Electrocardiograms (ECG) are recorded from many sensors simultaneously. Therefore, it is natural to use multi-dimensional models or tensors for representing these signals.

Different types of tensor decompositions are used for the extraction of features from the data or to denoise recorded signals. However, in biomedical signal processing the most frequently used decompositions of multi-dimensional data

are an approximate low-rank Canonical Polyadic (CP) decomposition [2] also known as Parallel Factor (PARAFAC) analysis [3] or Canonical Decomposition (CANDECOMP) [4] and the truncated Multi-Linear Singular Value Decomposition (MLSVD) [5] also known as Higher Order Singular Value Decomposition (HOSVD).

According to the CP model, a tensor is decomposed into the minimum number  $R$  of rank-one components. Hence, the proper choice of the model order affects the accuracy of the processing and subsequently also the interpretable results. This problem is very important when measured biomedical data are processed. However, in most cases, the observed data are corrupted by noise. Therefore, the problem of estimating the order of an approximate low-rank model is a non-trivial task.

The first attempts to develop methods for overcoming this problem were made at the beginning of the 1970s. In [6] Akaike's information criterion (AIC) was proposed. This criterion takes the observed data structure into account. In [7] another criterion was proposed that penalizes the over parameterization more strongly than AIC. Schwarz proposed a Bayesian information criterion (BIC) for the estimation of the order of linear models that describe independent and identically distributed observations [8]. Wax and Kailath implemented the AIC and the Minimum Description Length (MDL) scheme that has been derived from the BIC for the detection of the number of signals in a multi-channel time series [9]. However, these methods often fail when the number of the snapshots of the observed data is small.

To overcome this problem, the Exponential Fitting Test (EFT) based on the geometrical profile of noise-only eigenvalues has been developed and presented in [10]. The EFT allows the probability of false alarm to be controlled and pre-defined, which is a crucial point for systems such as RADARs. Moreover, in [11] the EFT was improved for multidimensional signals via the Modified Exponential Fitting Test (M-EFT) that outperforms all other schemes for the cases when the data are corrupted by Gaussian noise. Furthermore, the authors showed how the classical AIC and MDL methods can be improved. In [12] the authors presented the  $N$ -dimensional ( $N$ -D) EFT that outperforms the other techniques for cases with Gaussian noise. For colored noise a new Closed-Form PARAFAC-based Model Order Selection (CFP-MOS) was proposed. Moreover, the AIC and MDL techniques were extended to the multi-dimensional cases. Several methods for estimating the order of CP models were presented in [13], [14], [15]. In [16], the authors propose a technique to find the "knee" of an error

Alexey A. Korobkov and Marina K. Diugurova are with the Department for Radio-Electronic and Telecommunication Systems, Institute for Radio-Electronics, Photonics, and Digital Technologies, Kazan National Research Technical University n.a. A.N. Tupolev-KAI, Kazan, Russia, e-mail: korobkov@inbox.ru.

Jens Haueisen is with the Institute of Biomedical Engineering and Informatics, Ilmenau University of Technology, Ilmenau, Germany, email: jens.haueisen@tu-ilmenau.de.

Martin Haardt is with the Communications Research Laboratory, Ilmenau University of Technology, Ilmenau, Germany, email: martin.haardt@tu-ilmenau.de.

curve to determine the number of clusters. This technique can also be used to estimate the model order. In [17], the authors implemented the  $k$ -means clustering algorithm to separate the 1-mode singular values of the covariance tensor into two clusters: the signal cluster with large singular values and the noise cluster with the remaining smaller singular values. The number of elements in the signal cluster is assumed as the model order.

The estimation of the model order in biomedical signal processing problems is often based on the results of a visual inspection of the singular values or some assumptions about the structure of the data. To avoid this visual inspection and to provide a systematic approach for model order estimation in practical applications, we propose a novel robust method based on the LineAR Regression of Global Eigenvalues (LaRGE). In addition to the initial results presented in [18], in this paper, we provide a more detailed derivation of the equations of the LaRGE method. Moreover, we introduce a heuristic alternative by incorporating a penalty function (LaRGE-PF) that shows an improved performance in the case of relatively small tensors. We compare the performance of LaRGE and LaRGE-PF with the classical AIC and MDL schemes and their  $N$ -dimensional extensions. Furthermore, we provide detailed results of the model order estimation and the subsequent low-rank tensor decomposition of measured EEG data during Intermittent Photic Stimulation.

In Section II the multi-dimensional data model is presented. Section III describes the proposed LaRGE and LaRGE-PF schemes. As a benchmark, extensions of two classical methods AIC and MDL (or BIC) to the multi-dimensional cases are reviewed in Section IV. Section V describes the simulation scenarios and compares the performance of the LaRGE, LaRGE-PF, AIC, MDL,  $N$ -D AIC, and  $N$ -D MDL. Inspired by the good performance of LaRGE and LaRGE-PF, we also implemented them for model order estimation of EEG recordings. The results are presented in Section VI. In Section VII, we conclude the paper.

## II. DATA MODEL AND NOTATION

In this paper, we use the following notation,  $a$ ,  $\mathbf{a}$ ,  $\mathbf{A}$ , and  $\mathcal{A}$  are used to denote scalars, column vectors, matrices, and tensors, respectively. Moreover,  $\mathbf{a}(i)$  defines the element ( $i$ ) of a vector  $\mathbf{a}$ . The same applies to a matrix  $\mathbf{A}(i, j)$  and a tensor  $\mathcal{A}(i, j, k)$ . The tensor  $\mathcal{I}_{D,R}$  is  $D$ -dimensional super-diagonal tensor of size  $R \times R \times \dots \times R$ , which is equal to one if all  $D$  indices are equal and zero otherwise. The  $d$ -mode product between a  $D$ -way tensor of size  $M_d$  along mode  $d = 1, 2, \dots, D$  represented as  $\mathcal{A} \in \mathbb{C}^{M_1 \times M_2 \times \dots \times M_D}$  and a matrix  $\mathbf{U} \in \mathbb{C}^{J \times M_d}$  is written as  $\mathcal{A} \times_d \mathbf{U}$ . It is computed by multiplying all  $d$ -mode vectors of  $\mathcal{A}$  with  $\mathbf{U}$ , whereas the  $d$ -mode vectors of  $\mathcal{A}$  are obtained by varying the  $d$ -th index from 1 to  $M_d$  and keeping all other indices fixed. Aligning all  $d$ -mode vectors as the columns of a matrix yields the  $d$ -mode unfolding of  $\mathcal{A}$  which is denoted by  $[\mathcal{A}]_{(d)} \in \mathbb{C}^{M_d \times M_{d+1} \dots M_D \cdot M_1 \dots M_{d-1}}$ .

The CP decomposition of a  $D$ -way noiseless tensor  $\mathcal{X}_0 \in \mathbb{C}^{M_1 \times M_2 \times M_3 \times \dots \times M_D}$  is represented as

$$\mathcal{X}_0 = \mathcal{I}_{D,R} \times_1 \mathbf{F}_1 \times_2 \mathbf{F}_2 \times_3 \mathbf{F}_3 \times \dots \times_D \mathbf{F}_D, \quad (1)$$

where  $\mathbf{F}_d \in \mathbb{C}^{M_d \times R}$  ( $d = 1, 2, 3, \dots, D$ ) are the factor matrices and  $R$  is the order of the CP model or the rank of the tensor  $\mathcal{X}_0$ .

In practice, the recorded data are corrupted by noise. The tensor that is constructed from observations can be defined as

$$\mathcal{X} = \mathcal{X}_0 + \mathcal{N}, \quad (2)$$

where  $\mathcal{N} \in \mathbb{C}^{M_1 \times M_2 \times M_3 \times \dots \times M_D}$  is the additive noise tensor. Therefore, (1) can be rewritten as

$$\mathcal{X} = \mathcal{I}_{D,R} \times_1 \mathbf{F}_1 \times_2 \mathbf{F}_2 \times_3 \mathbf{F}_3 \times \dots \times_D \mathbf{F}_D + \mathcal{N}. \quad (3)$$

Obviously, the rank of the tensor  $\mathcal{X}$  is not equal to  $R$ . In general, it is bigger. Therefore, after a CP decomposition of the tensor with observations  $\mathcal{X}$ , we obtain the estimates  $\hat{\mathbf{F}}_d \in \mathbb{C}^{M_d \times R}$  ( $d = 1, 2, 3, \dots, D$ ) of factor matrices  $\mathbf{F}_d \in \mathbb{C}^{M_d \times R}$  ( $d = 1, 2, 3, \dots, D$ )

$$\mathcal{X} \approx \mathcal{I}_{D,R} \times_1 \hat{\mathbf{F}}_1 \times_2 \hat{\mathbf{F}}_2 \times_3 \hat{\mathbf{F}}_3 \times \dots \times_D \hat{\mathbf{F}}_D. \quad (4)$$

On the other side, the HOSVD model of a tensor  $\mathcal{X}$  is defined by

$$\mathcal{X} = \mathcal{S} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \times \dots \times_D \mathbf{U}_D, \quad (5)$$

where  $\mathcal{S} \in \mathbb{C}^{M_1 \times M_2 \times M_3 \times \dots \times M_D}$  is the core tensor and  $\mathbf{U}_r \in \mathbb{C}^{M_r \times M_r}$ , ( $d = 1, 2, 3, \dots, D$ ) are the unitary factor matrices.

As shown in [10], [11], [12] the  $d$ -mode singular values that are contained in the core tensor  $\mathcal{S}$  play a major role in the problem of multi-dimensional model order estimation. The  $d$ -mode singular values can be computed via the Singular Value Decomposition (SVD) of the  $d$ -mode unfolding of the tensor  $\mathcal{X}$  according to

$$[\mathcal{X}]_{(d)} = \mathbf{U}_d \cdot \Sigma_d \cdot \mathbf{V}_d^H, \quad (6)$$

where  $\mathbf{U}_d \in \mathbb{C}^{M_d \times M_d}$ ,  $\mathbf{V}_d \in \mathbb{C}^{\tilde{M}_d \times \tilde{M}_d}$  are unitary matrices and  $\Sigma_d \in \mathbb{C}^{M_d \times \tilde{M}_d}$  is a diagonal matrix that has the  $d$ -mode singular values  $\sigma_i^{(d)}$  on the main diagonal and  $\tilde{M}_d = \frac{M}{M_d}$ .

## III. LARGE METHOD FOR MULTI-DIMENSIONAL MODEL ORDER ESTIMATION

### A. Global Eigenvalues

As is shown in (5), the  $d$ -mode singular values represent the internal structure of the data and should be exploited for model order estimation. However, tensors with real data may have different  $d$ -ranks in different modes. Moreover, when an approximate low-rank CP decomposition is computed, an estimate of the rank of the noiseless model that is equal for all  $d$ -mode unfoldings has to be found. Therefore, we can define *global eigenvalues* as proposed in [11] and [12] as the product of all  $d$ -mode singular values with the same indices as follows

$$\tilde{\lambda}_i^{[G]} = \prod_{d=1}^D \left( \sigma_i^{(d)} \right)^2, i = 1, 2, 3, \dots, M^{[G]}, \quad (7)$$

where  $M^{[G]} = \min(M_d)$ ,  $d = 1, 2, 3, \dots, D$ .

Assuming that the global eigenvalues include the information about the signals and the noise, the set of global eigenvalues can be divided into two subsets. One subset contains the global eigenvalues that represent the noise in the observations.

These global eigenvalues are called *noise global eigenvalues*. The other subset contains the *signal global eigenvalues*. The number of the signal global eigenvalues corresponds to the model order or the rank of the tensor  $\mathcal{X}_0$  as well as the number of sources or components in the noisy observations.

### B. LaRGE method

In [11] and [12], it was shown that the noise global eigenvalues have approximately an exponential profile and there is at least one noise global eigenvalue in the set of all global eigenvalues. Starting from the smallest global eigenvalue, the noise global eigenvalues follow a straight line on a logarithmic scale (due to their approximate exponential distribution)

$$\lambda_i^{[G]} = \ln(\tilde{\lambda}_i^{[G]}). \quad (8)$$

Therefore, we want to determine a linear approximation  $\hat{\lambda}_i^{[G]}$  of the actual noise global eigenvalues profile  $\lambda_i^{[G]}$  on a logarithmic scale based on the least squares criterion

$$\min \left( \sum_{i=M^{[G]}-k}^{M^{[G]}} (\hat{\lambda}_i^{[G]} - \lambda_i^{[G]})^2 \right), \quad (9)$$

where  $\hat{\lambda}_i^{[G]} = a_1 \cdot i + a_2$ ,  $i = M^{[G]}, M^{[G]} - 1, M^{[G]} - 2, \dots, M^{[G]} - k$ . We sequentially approximate the profile of the noise global eigenvalues via a straight line starting from the smallest global eigenvalues with indices  $i = M^{[G]}$  and  $i = M^{[G]} - 1$  for  $k = 1$ , where the  $k$  is the step index. Using this linear approximation the value of the next largest noise eigenvalue can be predicted.

In each step  $k = 1, 2, \dots, M^{[G]} - 1$  the absolute prediction error  $\Delta_i^{[G]}$  is calculated according to

$$\Delta_{M^{[G]}-k}^{[G]} = \Delta_i^{[G]} = \lambda_i^{[G]} - \hat{\lambda}_i^{[G]}. \quad (10)$$

A small value of the absolute prediction error indicates that a noise global eigenvalue is found. On the contrary, a big prediction error can be indicative of a signal global eigenvalue.

To normalize the scale of values of the prediction errors, we define the relative prediction error

$$\delta_{M^{[G]}-k} = \frac{\Delta_{M^{[G]}-k}^{[G]}}{|\hat{\lambda}_{M^{[G]}-k}^{[G]}|} = \frac{\lambda_{M^{[G]}-k}^{[G]} - \hat{\lambda}_{M^{[G]}-k}^{[G]}}{|\hat{\lambda}_{M^{[G]}-k}^{[G]}|}. \quad (11)$$

We consider the standard deviation of the approximation error  $\sigma_{M^{[G]}-k}^2$  to automatically find the first signal global eigenvalue

$$\sigma_{M^{[G]}-k} = \sqrt{\frac{1}{k} \sum_{i=M^{[G]}-k}^{M^{[G]}} (\Delta_i^{[G]} - \mu_{\Delta_{M^{[G]}-k}^{[G]}})^2}, \quad (12)$$

where the  $\mu_{\Delta_{M^{[G]}-k}^{[G]}}$  is the mean value of  $\Delta_i^{[G]}$  for  $i = M^{[G]} - k, \dots, M^{[G]}$  that is defined as

$$\mu_{\Delta_{M^{[G]}-k}^{[G]}} = \frac{1}{k} \sum_{i=M^{[G]}-k}^{M^{[G]}} \Delta_i^{[G]}. \quad (13)$$

The standard deviation  $\sigma_{M^{[G]}-k-1}$  takes into account all errors in previous steps. Intuitively, starting from the smallest

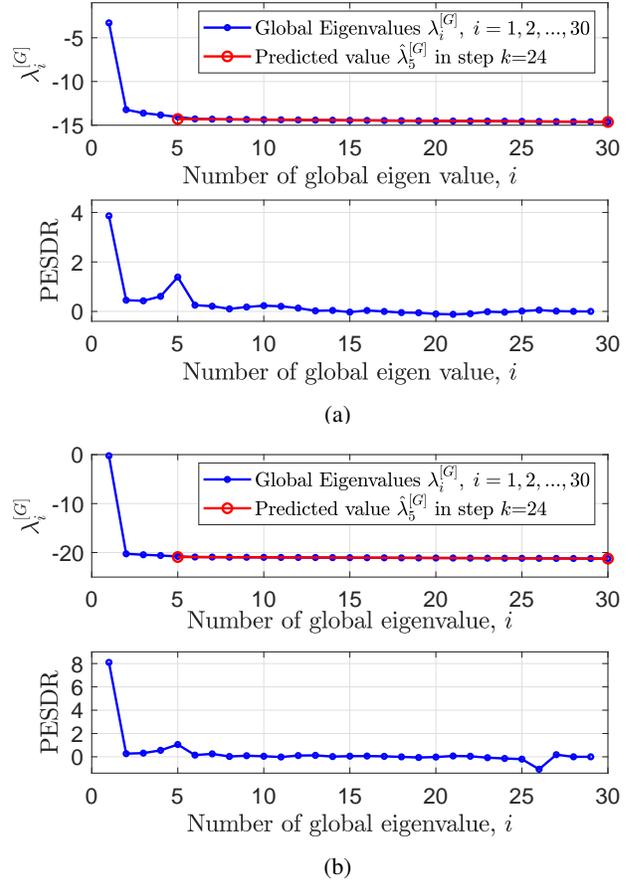


Fig. 1: The profile of the global eigenvalues (top) and the Prediction Error to Standard Deviation Ratio (bottom) computed for a random tensor  $\mathcal{X} \in \mathbb{R}^{100 \times 350 \times 30}$  with a known rank 5. (a) Uncorrelated case, SNR=-3 dB, (b) correlated case, SNR=11 dB

noise global eigenvalue, the first signal global eigenvalue can be automatically found by comparing the relative prediction error and the standard deviation of the approximation errors in the previous steps. To this end, we consider the Prediction Error to Standard Deviation Ratio (PESDR)

$$\text{PESDR}_k = \frac{\delta_{M^{[G]}-k}}{\sigma_{M^{[G]}-k-1}}, \quad k = 1, 2, \dots, M^{[G]} - 1. \quad (14)$$

To illustrate the behavior of the PESDR curve with respect to the profile of the global eigenvalues, we have constructed two random tensors with a known rank equal to five and dimensions  $100 \times 350 \times 30$  for the CP model (1). The first tensor has uncorrelated columns in all factor matrices. The correlation of all factor matrices for the second tensor is defined by the vector  $r_2 = [0.6 \ 0.7 \ 0.3]^T$  according to [19]. Gaussian noise is added to each tensor according to the CP model (3) with SNR=-3 dB and SNR=11 dB for the uncorrelated and the correlated case, respectively. The SNR is defined as

$$\text{SNR} = 10 \cdot \log_{10} \frac{\mathbb{E}\{\|\mathcal{X}_0\|_F^2\}}{\mathbb{E}\{\|\mathcal{N}\|_F^2\}}, \quad \text{dB}, \quad (15)$$

where the  $\mathbb{E}\{\cdot\}$  denotes the expected value.

Figure 1 depicts the profile of the global eigenvalues and the computed PESDR coefficients for the constructed tensors. The

straight line fits the profile of the noise global eigenvalues. It is seen that the first number  $i$ , where the PESDR curve exceeds a certain threshold  $\rho$  corresponds to the rank of the noisy tensor.

To formalize this criterion for the model order or the rank estimation, we can consider the following hypotheses:

$$H_k^{(N)} : \lambda_{M^{[G]}-k}^{[G]} \text{ is a noise global eigenvalue} \quad (16)$$

$$H_k^{(S)} : \lambda_{M^{[G]}-k}^{[G]} \text{ is a signal global eigenvalue.} \quad (17)$$

In each step  $k = 1, 2, \dots, M^{[G]} - 1$  or for each index  $i = M^{[G]} - k$  the PESDR is compared to the threshold  $\rho$  according to

$$H_k^{(N)} : \text{PESDR}_k < \rho \quad (18)$$

$$H_k^{(S)} : \text{PESDR}_k \geq \rho \quad (19)$$

The first index  $k$  for which the hypothesis  $H_k^{(N)}$  fails determines the estimated model order or the rank of the tensor and is denoted by  $\hat{R}_{\text{LaRGE}} = M^{[G]} - k$ .

### C. LaRGE with penalty function (LaRGE-PF)

According to the global eigenvalues approach [11], [12] the number of global eigenvalues that are used for model order estimation is equal to the smallest dimension of the tensor. In real-world applications, the smallest dimension can be relatively small compared to the rank. In the first steps of estimating the standard deviation of the approximation error  $\sigma_{M^{[G]}-k-1}$ , the resulting estimate tends to zero. Hence, the PESDR can have outliers that exceed the threshold  $\rho$ . To avoid such a misclassification, we additionally ensure that  $\sigma_{M^{[G]}-k-1}$  exceeds a threshold  $\epsilon$ . In this paper, we set  $\epsilon = 1.2 \cdot 10^{-3}$ .

Moreover, we also consider the slightly modified "PESDR with penalty function (PESDR-PF)" criterion that penalizes the standard deviation in the first steps

$$\begin{aligned} \text{PESDR-PF}_k &= \frac{\delta_{M^{[G]}-k}}{\text{PF}(M^{[G]}, k-1) \cdot \sigma_{M^{[G]}-k-1}} = \\ &= \frac{1}{\log_{10}(M^{[G]} - k - 1)} \cdot \frac{\delta_{M^{[G]}-k}}{\sigma_{M^{[G]}-k-1}}. \end{aligned} \quad (20)$$

In the first estimation steps, the indices of the corresponding global eigenvalues  $k$  are small and the argument of the logarithm ( $M^{[G]} - k$ ) in the denominator of the penalty function is relatively big. Therefore, the values of the penalty function (PF) have big weights. Hence, the denominator of the ratio becomes bigger and outliers are damped. As the number of steps increases, i.e., as the index  $k$  increases, the value of the penalty function decreases and the damping effect is reduced.

### D. Threshold determination

The threshold  $\rho$  is related to the statistical distribution of the absolute prediction error  $\Delta_i^{[G]}$  or the relative prediction error  $\delta_i^{[G]}$ . This statistical distribution is not known for applications with measured real data. Therefore, numerical methods can instead be used for the threshold determination.

In this paper, the threshold  $\rho$  is obtained by the Monte Carlo simulations. For this purpose, we define the probability of false

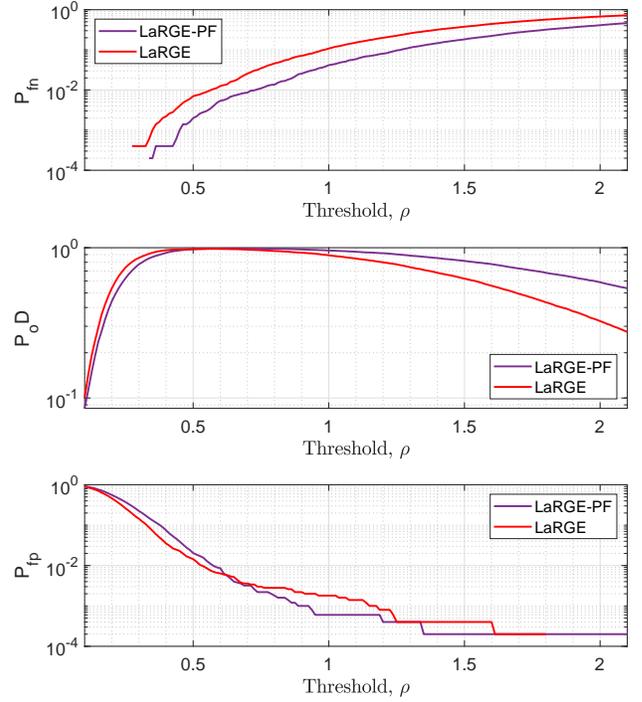


Fig. 2: The functional dependencies of  $P_{\text{fn}}$  (top),  $P_{\text{od}}$  (middle),  $P_{\text{fp}}$  (bottom) and the threshold  $\rho$  for the noisy CP model, SNR=8 dB,  $\mathcal{X} \in \mathbb{R}^{25 \times 30 \times 35}$ , the true rank is  $R = 5$ , 5000 trials.

alarm or false positive  $P_{\text{fp}}$ , the probability of missing or false negative  $P_{\text{fn}}$ , and the probability of detection  $P_{\text{od}}$ .

The probability of false positive  $P_{\text{fp}}$  is the probability that the LaRGE method falsely classifies a noise global eigenvalue as a signal global eigenvalue and is defined as

$$P_{\text{fp}} = \Pr(H_k^{(S)} | H_k^{(N)}) = \Pr(\hat{R}_{\text{LaRGE}} > R | R). \quad (21)$$

A big value of this probability means that the rank or the model order can be overestimated.

The probability of false negative  $P_{\text{fn}}$  expresses the probability that the proposed method mistakenly determines the model order smaller than the actual order. This probability is defined as

$$P_{\text{fn}} = \Pr(H_k^{(N)} | H_k^{(S)}) = \Pr(\hat{R}_{\text{LaRGE}} < R | R). \quad (22)$$

The model order or the rank can be underestimated if the value of  $P_{\text{fn}}$  is big.

The probability of detection is the probability of correct model order estimation and is considered as

$$\text{PoD} = \Pr(\hat{R} = R). \quad (23)$$

By considering a large number of trials we have obtained the functional dependency of the probabilities of false positive  $P_{\text{fp}}(\rho)$ , false negative  $P_{\text{fn}}(\rho)$ , and the probability of detection  $P_{\text{od}}(\rho)$  as a function of the threshold  $\rho$ .

To this end, we generate a data tensor  $\mathcal{X}_0$  for the CP model (1). All elements of the factor matrices are drawn from a zero-mean real-valued Gaussian distribution. Moreover, all columns of the factor matrices  $\mathbf{F}_d$ ,  $d = 1, 2, 3, \dots, D$  are not

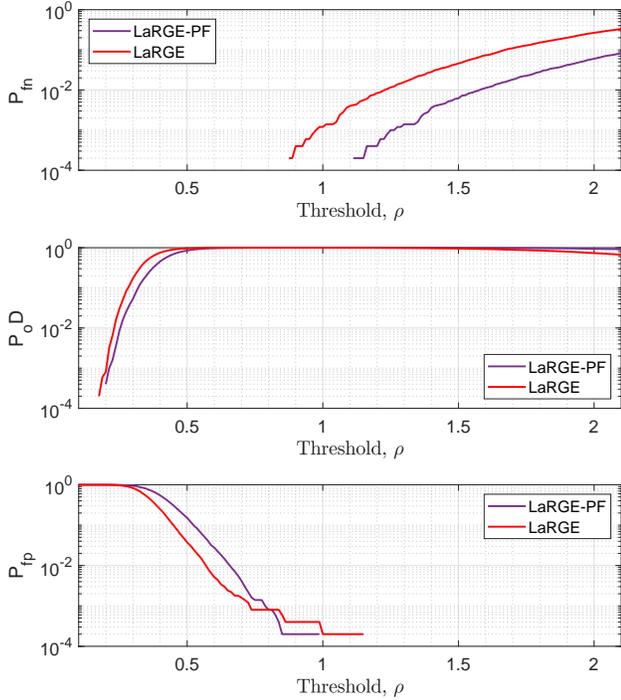


Fig. 3: The functional dependencies of  $P_{\text{fn}}$  (top), PoD (middle),  $P_{\text{fp}}$  (bottom) and the threshold  $\rho$  for the noisy CP model, SNR=0 dB,  $\mathcal{X} \in \mathbb{R}^{60 \times 100 \times 70}$ , the true rank is  $R = 5$ , 5000 trials.

correlated. For all simulations, noiseless tensors of rank  $R = 5$  are constructed. The elements of the noise tensor  $\mathcal{N}$  are drawn independently from a zero-mean Gaussian distribution with the variance  $\sigma_n^2$  and the SNR that is defined as in (15).

For the first simulation, the size of the noisy tensor is set to  $M_1 = 25, M_2 = 30, M_3 = 35$ . The SNR is fixed at SNR=8 dB, and the value of the threshold  $\rho$  varies in the range from 0.1 to 2.1. The curves of the probabilities  $P_{\text{fp}}, P_{\text{fn}}$ , and PoD of the LaARGE method obtained after 5000 Monte Carlo trials are shown in Figure 2.

For the second simulation, the size of the noisy tensor is set to  $M_1 = 60, M_2 = 100, M_3 = 70$  and SNR=0 dB. Figure 3 shows the curves of the probabilities  $P_{\text{fp}}, P_{\text{fn}}$ , and PoD of the LaARGE method obtained after 5000 Monte Carlo trials for this scenario.

For the third simulation, the size of the noisy tensor (SNR=-9 dB) is significantly increased and is set to  $M_1 = 78, M_2 = 1000, M_3 = 102$ . The resulting curves of the probabilities  $P_{\text{fp}}, P_{\text{fn}}$ , and PoD of the LaARGE method are depicted in Figure 4.

For the fourth scenario, a four-dimensional noisy tensor (SNR=-9 dB) of size  $M_1 = 60, M_2 = 60, M_3 = 60, M_4 = 60$  is generated. The obtained curves of the probabilities  $P_{\text{fp}}, P_{\text{fn}}$ , and PoD of the LaARGE method are shown in Figure 5.

Using the results of the Monte Carlo simulations, a numerical value for the threshold  $\rho$  can be obtained by setting the probabilities  $P_{\text{fp}}$  and  $P_{\text{fn}}$  to an appropriate value. In this paper, we want to keep these probabilities below 1% as used in [10] and [11]. Simulations for small tensors show that the probability  $P_{\text{fn}} = 0.01$  corresponds to the threshold  $\rho = 0.57$

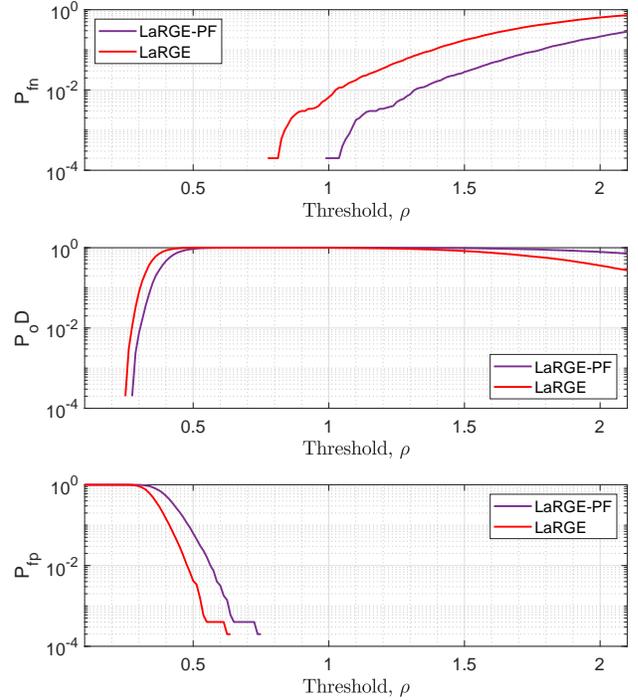


Fig. 4: The functional dependencies of  $P_{\text{fn}}$  (top), PoD (middle),  $P_{\text{fp}}$  (bottom) and the threshold  $\rho$  for the noisy CP model, SNR=-9 dB,  $\mathcal{X} \in \mathbb{R}^{78 \times 1000 \times 102}$ , the true rank is  $R = 5$ , 5000 trials.

(Figure 2). Moreover, a smaller value of the false negative probability  $P_{\text{fn}}$  of the LaARGE-PF method eventually provides a gain in terms of the correct detection probability PoD as shown in Figure 2.

On the other hand, for bigger tensors (Figure 3) the threshold  $\rho = 0.57$  corresponds to  $P_{\text{fp}} = 0.01$ . Furthermore, Figure 3 shows that for an implementation of LaARGE-PF for tensors of size  $60 \times 5100 \times 70$  the probabilities  $P_{\text{fp}}$  and  $P_{\text{fn}}$  are equal to zero, and hence, a correct detection probability PoD = 100% is obtained for the values of  $\rho$  from 1 to 1.2. Both LaARGE and LaARGE-PF can provide a probability of correct detection PoD = 100% for tensors of larger sizes or dimensions as shown in Figures 4 and 5.

For instance, LaARGE provides PoD = 100% for the threshold  $\rho$  in the range from 0.63 to 0.77 and LaARGE-PF in the range from 0.75 to 0.98 for tensors of size  $78 \times 1000 \times 102$ . For 4-dimensional tensors of size  $60 \times 60 \times 60 \times 60$  a probability of correct detection PoD = 100% can be obtained for the threshold  $\rho$  in the range from 0.5 to 0.75 using LaARGE and in the range from 0.6 to 0.92 using LaARGE-PF.

#### IV. CLASSICAL METHODS FOR MULTI-DIMENSIONAL MODEL ORDER ESTIMATION

To compare the performance of the proposed LaARGE method to the state-of-the-art, we use AIC and MDL as well as their multi-dimensional extensions as a benchmark in this paper.

The classical AIC and MDL methods are based on information theory and these methods have originally been applied

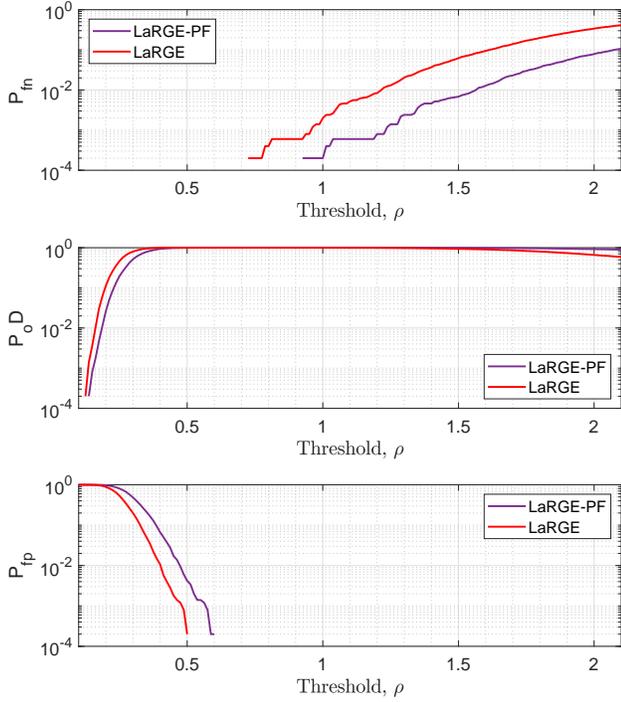


Fig. 5: The functional dependencies of  $P_{\text{in}}$  (top),  $P_{\text{PoD}}$  (middle),  $P_{\text{ip}}$  (bottom) and the threshold  $\rho$  for the noisy CP model, SNR= -9 dB,  $\mathcal{X} \in \mathbb{R}^{60 \times 60 \times 60 \times 60}$ , the true rank is  $R = 5$ , 5000 trials.

to one dimensional data. According to [9] the AIC criterion is given by

$$\text{AIC}_i = -2N^{[s]}(M^{[\sigma]} - i)L(i) + 2i(2M^{[\sigma]} - i), \quad (24)$$

where  $N^{[s]}$  and  $M^{[\sigma]}$  are the number of snapshots and the number of the eigenvalues  $\sigma_i$  of the covariance matrix  $\mathbf{R}_{xx}$  of the observation  $\mathbf{X}$ , respectively. Note that the  $L(i)$  in (24) is the log-likelihood that represents the residual error and is defined as

$$L(i) = \log_{10} \left[ \frac{g(i)}{a(i)} \right]. \quad (25)$$

The functions  $g(i)$  and  $a(i)$  are the geometric and arithmetic means of the eigenvalues  $\sigma_i$  of the covariance matrix  $\mathbf{R}_{xx}$ , respectively. The second term in (24) is a penalty for overfitting.

Contrary to the AIC criterion, the MDL criterion penalizes the overfitting more strongly

$$\text{MDL}_i = -N^{[s]}(M^{[\sigma]} - i)L(i) + \frac{1}{2}i(2M^{[\sigma]} - i) \log_{10}(N^{[s]}). \quad (26)$$

Minimizing the following cost functions, the model order can be estimated as

$$\hat{R}_{\text{AIC}} = \arg \min_i \{ \text{AIC}_i \}, \quad (27)$$

$$\hat{R}_{\text{MDL}} = \arg \min_i \{ \text{MDL}_i \}. \quad (28)$$

In [11] and [12], the authors have proposed an extension of the AIC (24) and MDL (26) methods to multidimensional data. This can be achieved by replacing the eigenvalues  $\sigma_i$  of the covariance matrix  $\mathbf{R}_{xx}$  by the global eigenvalues  $\lambda_i^{[G]}$  defined in equation (8). Replacing the eigenvalues of the covariance matrix by the global eigenvalues in (24) - (26), with some straightforward manipulations, we obtain

$$\text{AIC}_i^{[G]} = -2N^{(s)}(M^{[G]} - i)L^{[G]}(i) + 2i(2M^{[G]} - i), \quad (29)$$

$$\text{MDL}_i^{[G]} = -N^{(s)}(M^{[G]} - i)L^{[G]}(i) + \frac{1}{2}i(2M^{[G]} - i) \log_{10}(N^{(s)}), \quad (30)$$

where

$$L^{[G]}(i) = \log \left[ \frac{\left( \prod_{k=i+1}^{M^{[G]}} \lambda_k^{[G]} \right) \frac{1}{M^{[G]} - i}}{\frac{1}{M^{[G]} - i} \cdot \sum_{k=i+1}^{M^{[G]}} \lambda_k^{[G]}} \right]. \quad (31)$$

Finally, the model order is estimated by the minimization of the costs functions

$$\hat{R}_{N-D \text{ AIC}} = \arg \min_i \{ \text{AIC}_i^{[G]} \}, \quad (32)$$

$$\hat{R}_{N-D \text{ MDL}} = \arg \min_i \{ \text{MDL}_i^{[G]} \}. \quad (33)$$

## V. EXPERIMENTAL RESULTS. SYNTHETIC SIMULATIONS

In this section, we present the results of simulations to demonstrate the performance of the proposed LaARGE and LaARGE-PF methods in comparison with the classical AIC, MDL,  $N$ -D AIC, and  $N$ -D MDL methods. We set the threshold  $\rho = 0.57$  for the LaARGE and LaARGE-PF methods and compare the probability of correct detection  $\text{PoD} = \Pr(\hat{R} = R)$  versus the SNR.

A data tensor for the CP model (1) with factor matrices  $\mathbf{F}_d, d = 1, 2, 3, \dots, D$  is generated. The factor matrices contain elements drawn from a zero-mean Gaussian distribution with unit variance  $\sigma_s^2$ . Two different scenarios are used for the simulations with respect to the correlation of the columns in the factor matrices. In the first scenario, the correlation between the columns in the factor matrices  $\mathbf{F}_d, d = 1, 2, 3, \dots, D$  is parameterized by the vectors  $\mathbf{r}_1 = [0 \ 0 \ 0]^T$  and  $\mathbf{r}_1 = [0 \ 0 \ 0]^T$  for 3-D and 4-D tensors, respectively. The second scenario corresponds to the correlated case, and the correlation of the columns of factor matrices is parameterized by the vectors  $\mathbf{r}_2 = [0.6 \ 0.3 \ 0.9]^T$  and  $\mathbf{r}_2 = [0.6 \ 0.3 \ 0.9 \ 0.1]^T$  for 3-D and 4-D tensors, respectively [20]. The rank of the generated noiseless tensors is fixed at  $R = 5$ . According to the model, (3) the elements of the noise tensor  $\mathcal{N}$  are drawn independently from a zero-mean Gaussian distribution with variance  $\sigma_n^2$ .

In Figure 6 we observe the performance of the classical methods AIC, MDL, 3-D AIC, 3-D MDL and the proposed LaARGE and LaARGE-PF methods for a tensor with the following dimensions  $M_1 = 25, M_2 = 30, M_3 = 35$ . The obtained

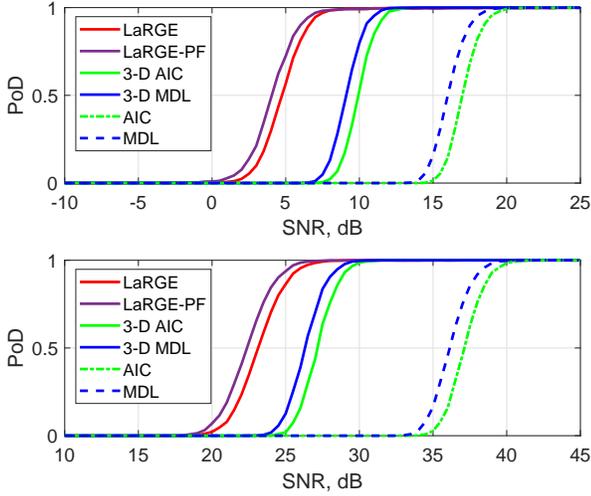


Fig. 6: Probability of correct detection PoD versus the SNR for the noisy CP model, threshold  $\rho = 0.57$ ,  $\mathcal{X} \in \mathbb{R}^{25 \times 30 \times 35}$ , the true rank is  $R = 5$ , 5000 Monte Carlo trials, uncorrelated case (top), correlated case (bottom).

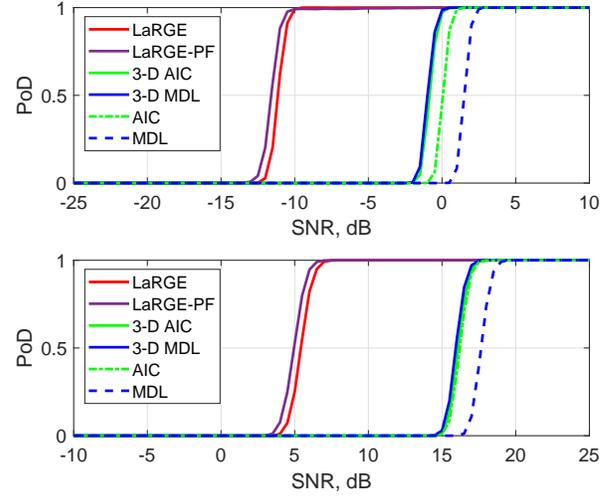


Fig. 8: Probability of correct detection PoD versus the SNR for the noisy CP model, threshold  $\rho = 0.57$ ,  $\mathcal{X} \in \mathbb{R}^{78 \times 1000 \times 102}$ , the true rank is  $R = 5$ , 5000 Monte Carlo trials, uncorrelated case (top), correlated case (bottom).

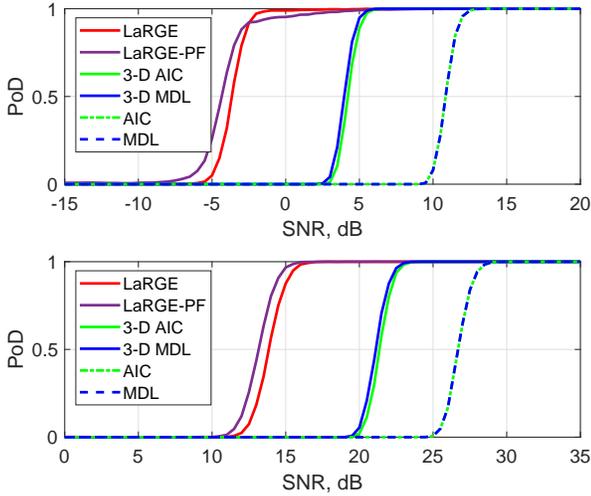


Fig. 7: Probability of correct detection PoD versus the SNR for the noisy CP model, threshold  $\rho = 0.57$ ,  $\mathcal{X} \in \mathbb{R}^{60 \times 100 \times 70}$ , the true rank is  $R = 5$ , 5000 Monte Carlo trials, uncorrelated case (top), correlated case (bottom).

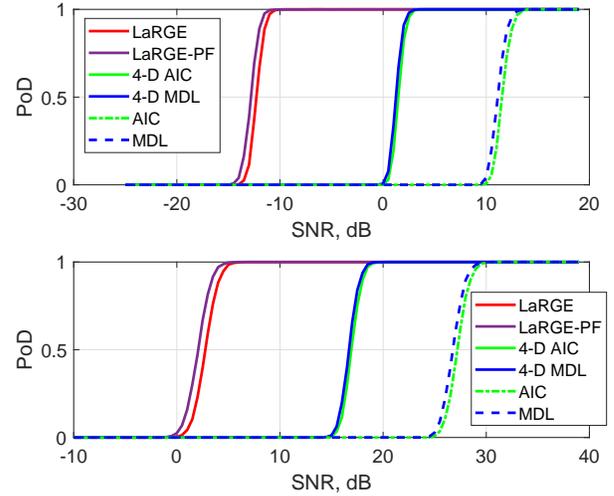


Fig. 9: Probability of correct detection PoD versus the SNR for the noisy CP model, threshold  $\rho = 0.57$ ,  $\mathcal{X} \in \mathbb{R}^{60 \times 60 \times 60 \times 60}$ , the true rank is  $R = 5$ , 5000 Monte Carlo trials, uncorrelated case (top), correlated case (bottom).

results confirm the better performance of the 3-D extension of AIC and MDL. Also, the proposed LaRGE and LaRGE-PF methods outperform the classical methods significantly.

Similarly, in Figure 7 we evaluate the probability of correct detection of the AIC, MDL, 3-D AIC, 3-D MDL, LaRGE, and LaRGE-PF methods for a tensor with the following dimensions  $M_1 = 60$ ,  $M_2 = 100$ ,  $M_3 = 70$ . The LaRGE method shows a better performance in comparison with LaRGE-PF for SNRs up to 5 dB in the uncorrelated case.

Figure 8 depicts a scenario where the size of the tensor of rank  $R = 5$  has the dimensions  $M_1 = 78$ ,  $M_2 = 1000$ ,  $M_3 = 102$ . Again, the LaRGE and LaRGE-PF methods outperform AIC, MDL, 3-D AIC, and 3-D MDL. We can see that the 3-D AIC, 3-D MDL, AIC, and MDL methods show similar results. This effect can be explained as follows.

For AIC and MDL, the  $d$ -mode singular values are used from the biggest dimension of the tensor. In our case, the largest number of  $d$ -mode singular values is equal to 1000 (for  $d = 2$ ). However, for 3-D AIC and 3-D MDL the number of global eigenvalues is limited by the smallest tensor dimension. In our case, this number is 78. Nevertheless, LaRGE and LaRGE-PF enhance the performance significantly.

We have also performed Monte Carlo simulations for a 4-D tensor scenario with dimensions  $M_1 = 60$ ,  $M_2 = 60$ ,  $M_3 = 60$ ,  $M_4 = 60$  and rank  $R = 5$ . Figure 9 shows the results for this scenario. Again, we can clearly see that LaRGE and LaRGE-PF outperform the methods based on AIC and MDL by more than 12 dB. For small size tensors, the advantage of LaRGE-PF is more pronounced than for big 3-D tensors or 4-D tensors in comparison with the LaRGE method.

## VI. REAL BIOMEDICAL SIGNAL PROCESSING

### A. Preprocessing and Decomposition

The proposed LaRGE and LaRGE-PF methods presented in this paper are used to analyze measured EEG data. These data were recorded from twelve healthy volunteers, numbered 1 to 12 in this paper, at the Biomagnetic Center of the University Hospital in Jena, Germany, for the investigation of the Photic Driving (PD) effect [21],[22]. This effect occurs when the brain is stimulated by Intermittent Photic Stimulation (IPS). The PD effect is used to assess the effects of medications and for the diagnosis of several neurophysiological diseases like Alzheimer, schizophrenia, and some forms of epilepsy.

The recorded EEG signals were filtered in the frequency range from 3 Hz to 40 Hz in a preprocessing step. The individual frequency of alpha rhythm  $f_\alpha$  were determined for each volunteer prior to the main experiment. The resulting alpha frequencies for volunteers 1 to 12 are, in this order, 9.6, 10.7, 10.4, 10.8, 10.7, 10.8, 7.5, 10.8, 11.0, 10.7, 12.2, and 10.3 Hz. To investigate the PD effect, twenty stimulation frequencies  $f_{stim} = [0.40 \ 0.45 \ 0.50 \ 0.55 \ 0.60 \ 0.70 \ 0.80 \ 0.90 \ 0.95 \ 1.00 \ 1.05 \ 1.10 \ 1.30 \ 1.60 \ 1.90 \ 1.95 \ 2.00 \ 2.05 \ 2.10 \ 2.30] \cdot f_\alpha$  were generated. Each stimulation of the particular frequency was presented in 30 stimulation trains. Each train consisted of 40 periods with pulse/cycle duration of 0.5. Between each train and frequency block there were resting periods of 4 seconds and 30 seconds, respectively.

To investigate the proposed LaRGE method for the model order estimation, two different cases are considered in this paper. 3-D tensors are constructed for the first case. To this end, the Fast Fourier Transform is used for obtaining the frequency distribution of the EEG signals. As a result, we have different complex tensors with dimensions channels  $\times$  frequency  $\times$  trains for each stimulation frequency and volunteer. The frequency and train dimensions have fixed sizes: 281 and 30, respectively. The size of the channel dimension varies, because a small number of non-functional EEG channels are excluded in the preprocessing step. The resulting 3-D tensors have dimensions channels $\times$ frequency $\times$ train.

For the second case that is considered in this paper, the wavelet decomposition is used to construct 4-D tensors. The second and third dimensions of these tensors represent the frequency-time distribution obtained after the complex Morlet decomposition. The wavelet coefficients between 3.003 Hz (1000/333 Hz) and 15.15 Hz (1000/66 Hz) are selected for the analysis. This frequency range spans the frequencies of the theta and alpha rhythms [23]. The size of the time dimension varies from 5 seconds up to 20 seconds depending on the stimulation frequency. The resulting 4-D tensors have dimensions channels $\times$ frequency $\times$ time $\times$ train.

In the first step of the model order estimation, the HOSVD of the constructed tensors is computed. Next, the global eigenvalues are obtained and the model order is estimated using the LaRGE method, the LaRGE method with penalty function (LaRGE-PF), and classical AIC and MDL methods ( $N$ -D AIC,  $N$ -D MDL, AIC, and MDL). To obtain the factor matrices of the dominant components, the CP decomposition of each tensor is computed via SECSI [20] in the 3-D case and

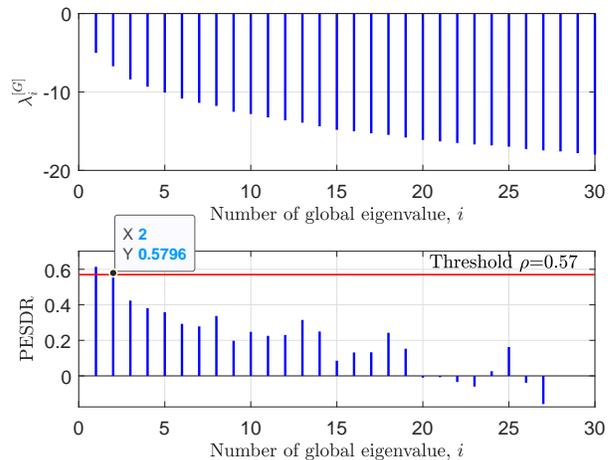


Fig. 10: The profile of the global eigenvalues (top) and Prediction Error to Standard Deviation Ratio (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{103 \times 281 \times 30}$  of volunteer 4 at stimulation frequency  $0.5 \cdot f_\alpha$ .

via SECSI-GU [24] in the 4-D case using the estimated rank of LaRGE and LaRGE-PF. The threshold  $\rho = 0.57$  is fixed for both LaRGE and LaRGE-PF to ensure that the probability  $P_{in}$  is not greater than 0.01. To eliminate the scaling ambiguity, the loading factors of the components are computed if the rank is more than one according to

$$\lambda_r = \prod_{d=1}^D \|\mathbf{F}(d, r)\|_F, r = 1 \dots \hat{R}, \quad (34)$$

where  $\hat{R}$  is the estimated rank and,  $D$  is the number of tensor dimensions.

In the rest of this section, we present some results that have been obtained during the analysis of 3-D and 4-D tensors. In the following figures, the profiles of global eigenvalues and the PESDR curves are shown. Moreover, the channel, frequency, and train signatures are presented for the 3-D case, and the channel, frequency, time, and train signatures are presented for the 4-D case. All components of the factor matrices are sorted in the order of decreasing loading factors computed according to (34).

### B. Processing of 3-D tensors

Figure 10 depicts the profile of the global eigenvalues and the PESDR coefficients for the tensor of volunteer 4 at stimulation frequency  $0.5 \cdot f_\alpha$ . The LaRGE method estimates the rank 2 for this case. Figure 11 depicts the dominant components of the factor matrices obtained via the CP decomposition. During the whole stimulation, the theta rhythm dominates which is evidenced by the distributions of both components in the spatial and frequency domains that are presented in Figure 11. The other state-of-the-art model order estimation schemes considered in this paper have estimated the rank as follow  $\hat{R}_{LaRGE-PF} = 3$ ,  $\hat{R}_{3-D AIC} = 10$ ,  $\hat{R}_{3-D MDL} = 13$ ,  $\hat{R}_{AIC} = 49$ ,  $\hat{R}_{MDL} = 43$ .

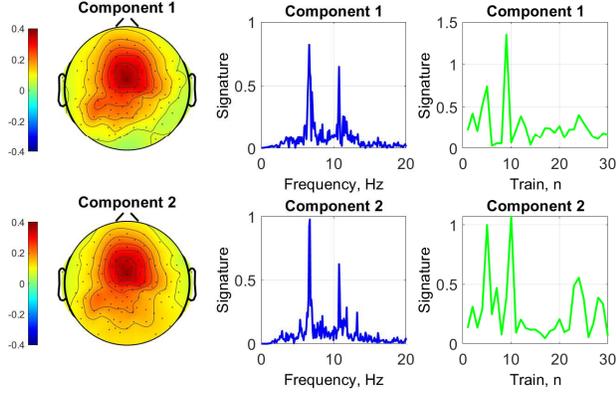


Fig. 11: Channel, frequency, and train signature for volunteer 4 and stimulation frequency  $0.5 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{103 \times 281 \times 30}$  and estimated rank 2. The loading factors of the components are  $\lambda_1 = 0.2941$ ,  $\lambda_2 = 0.2382$ .

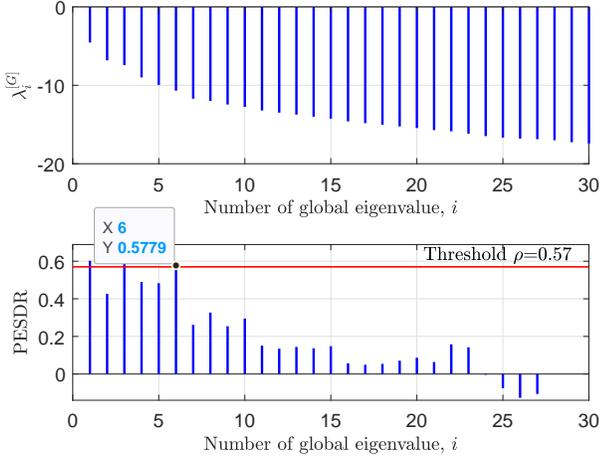


Fig. 12: The profile of the global eigenvalues (top) and Prediction Error to Standard Deviation Ratio (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{102 \times 281 \times 30}$  of volunteer 11 at stimulation frequency  $0.9 \cdot f_\alpha$ .

The results of processing the EEG signals that were recorded from volunteer 11 at stimulation frequency  $0.9 \cdot f_\alpha$  are presented in Figure 12 and Figure 13.

It is seen that the slope of the profile of the global eigenvalues changes between the sixth and the seventh global eigenvalue (see difference between 6th and 7th PESDR coefficient in Figure 12). Therefore, a change of the slope of the global eigenvalues can be detected using the PESDR reliably to separate the signal and noise global eigenvalues.

The stimulation frequency  $f_{\text{stim}} = 0.90 \cdot f_\alpha$  is close to the individual alpha frequency. Note the variability in the train dimension across the six components which indicates variability in the individual processing of each stimulus sequence (Fig. 13). The slightly variable topographies and frequency signatures for the 6 components also hint at a variable stimulation response possibly involving slightly different generator structures. The relatively high model order of six supports this

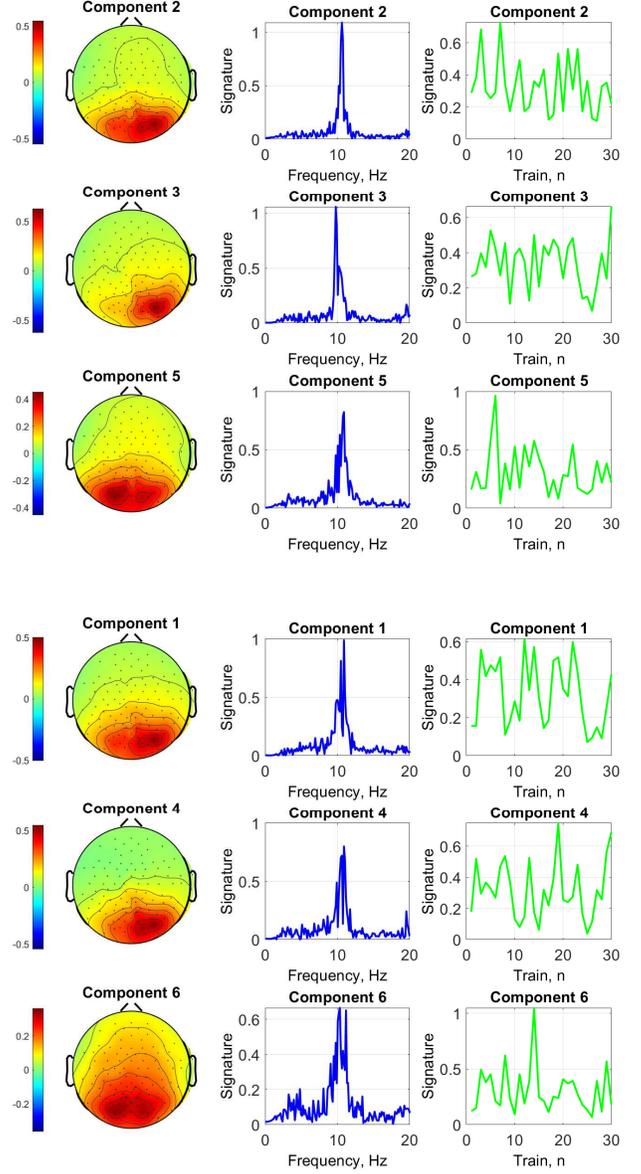


Fig. 13: Channel, frequency, and train signature for volunteer 11 and stimulation frequency  $0.9 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{102 \times 281 \times 30}$  and estimated rank 6. The loading factors of the components are  $\lambda_2 = 0.2996$ ,  $\lambda_3 = 0.2746$ ,  $\lambda_5 = 0.2729$ ,  $\lambda_1 = 0.2354$ ,  $\lambda_4 = 0.2303$ ,  $\lambda_6 = 0.2260$ .

interpretation.

### C. Processing of 4-D tensors

The results of processing a 4-D tensor with dimensions channel×frequency×time×train for volunteer 1 are depicted in Figures 14 and 15. The LaRGE method reliably estimates the model order as 2. Moreover, the bottom plot of Figure 14 shows the PESDR curve that is obtained according to LaRGE-PF. The suppression effect of the penalty function is clearly seen in this figure.

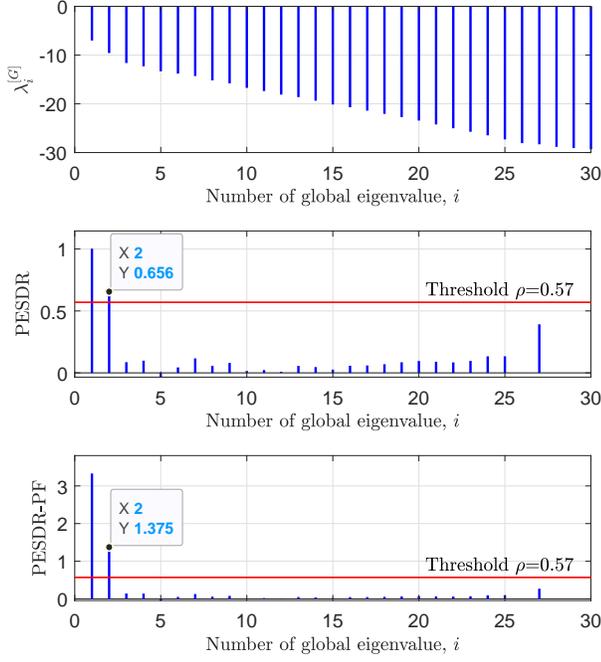


Fig. 14: The profile of the global eigenvalues (top), PESDR curve (middle) and PESDR-PF curve (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{268 \times 838 \times 105 \times 30}$  of volunteer 1 at stimulation frequency  $0.40 \cdot f_\alpha$ .

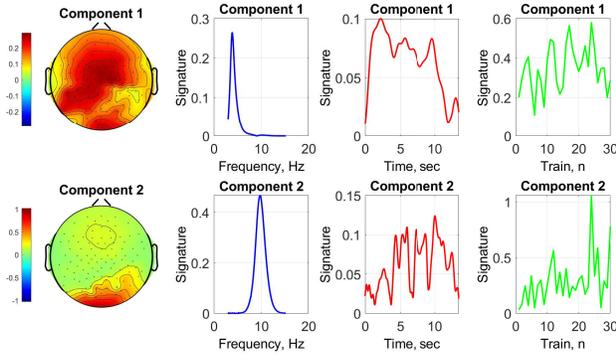


Fig. 15: Channel, frequency, time and train signature for volunteer 1 and stimulation frequency  $0.40 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{268 \times 838 \times 105 \times 30}$  and estimated rank 2. The loading factors of the components are  $\lambda_1 = 0.1809$   $\lambda_2 = 0.1037$ .

The dominant components of the factor matrices after the tensor decomposition show response in the alpha and theta frequencies. The alpha rhythm, as expected, is observed in the occipital region of the brain while the theta rhythm is dominant in the central region. Moreover, the theta rhythm has a resonance effect that is visible in the frequency and time dimensions in Figure 15 for component one. The other state-of-the-art model order estimation schemes considered in this paper estimate the model order of this tensor as  $\hat{R}_{4-D \text{ AIC}} = 24$ ,  $\hat{R}_{4-D \text{ MDL}} = 25$ ,  $\hat{R}_{\text{AIC}} = 205$ ,  $\hat{R}_{\text{MDL}} = 203$ .

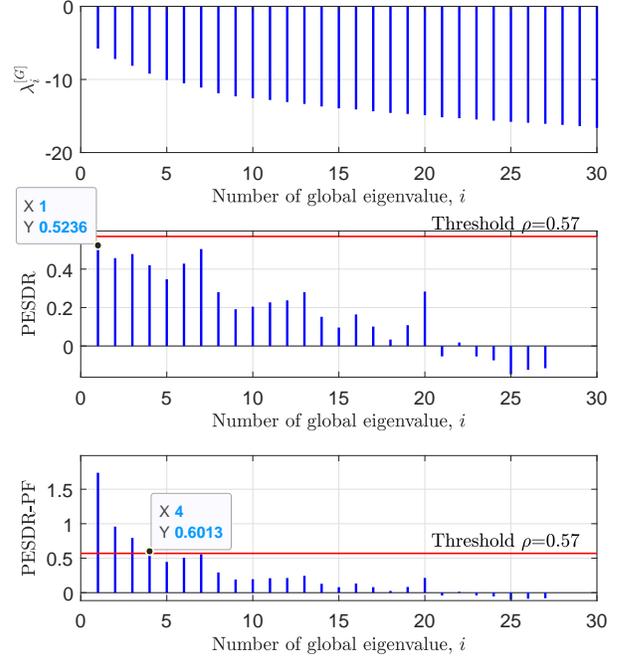


Fig. 16: The profile of the global eigenvalues (top), PESDR curve (middle) and PESDR-PF curve (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{106 \times 281 \times 30}$  of volunteer 8 at stimulation frequency  $0.50 \cdot f_\alpha$ .

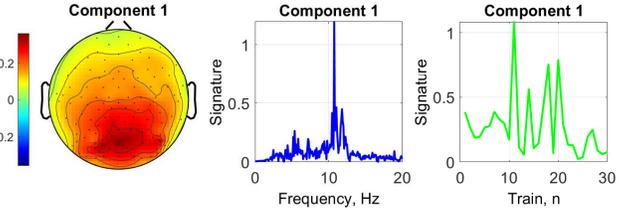


Fig. 17: Channel, frequency, time and train signature for volunteer 8 and stimulation frequency  $0.50 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{106 \times 281 \times 30}$  and estimated rank 1.

#### D. Special cases

A number of interesting cases could be observed during our investigations. Fig. 16 depicts the profile of the global eigenvalues and the PESDR curves for LaRGE and LaRGE-PF. None of the PESDR coefficients computed according to the LaRGE method exceeds the threshold  $\rho = 0.57$ . In such a case, the model order can be estimated as one (see Figure 17). Or the threshold can be decreased to the value equal to the largest PESDR coefficient. Therefore, the threshold can be updated to the new value  $\rho = 0.5236$  in the considered case. Alternatively, the estimated model order based on LaRGE-PF can be used (see Figure 18).

The profile of the global eigenvalues and the PESDR curves for a tensor with EEG data recorded from volunteer 11 at stimulation frequency  $f_{\text{stim}} = 0.6 \cdot f_\alpha$  are depicted in Figure 19. As in the previous case, none of the PESDR coefficients

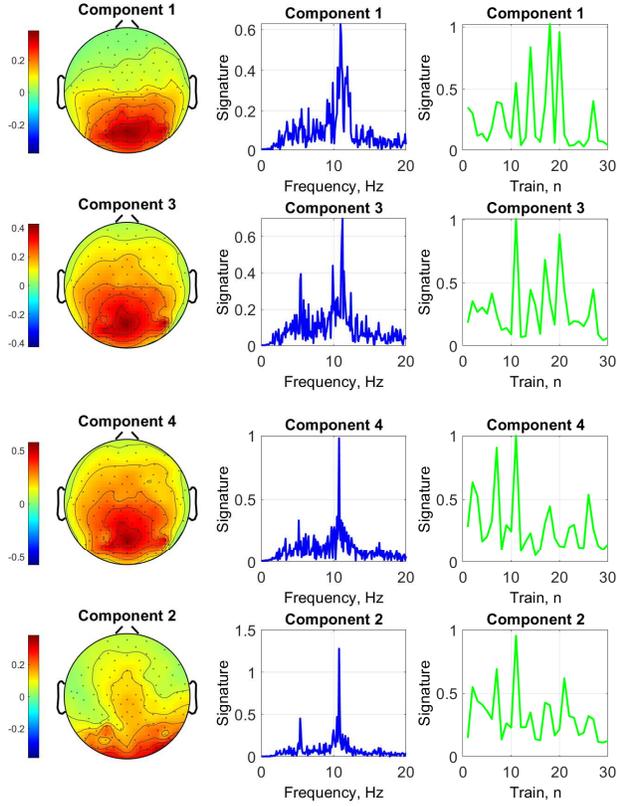


Fig. 18: Channel, frequency, and train signature for volunteer 8 and stimulation frequency  $0.5 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{106 \times 281 \times 30}$  and estimated rank 4. The loading factors of the components are  $\lambda_1 = 0.2476, \lambda_3 = 0.2239, \lambda_4 = 0.2090, \lambda_2 = 0.1728$ .

exceeds the threshold. However, the PESDR coefficients with numbers 4 and 18 are very close to the threshold. As proposed before, the model order estimated with LaRGE-PF can be used in this case.

The factor matrices obtained after the CP decomposition with rank 4 are depicted in Figure 20. In all topographic plots, one "bad" channel are clearly revealed. "Bad" channel means that the corresponding sensor had an imperfect connection with the surface of the skin during the measurements. Such channels contain artifacts with low or high amplitudes and should be removed from the observations during preprocessing (which failed for this channel in this case). After removing this channel, the tensor was constructed and the model order was estimated again. In Figure 21 the profile of the global eigenvalues and the PESDR curve are depicted after removing the "bad" channel. The eighteenth PESDR coefficient becomes significantly smaller and the first coefficient exceeds the threshold now. The obtained factor matrices after channel removal and subsequent tensor decomposition are depicted in Figure 22.

Another extraordinary case was found for the 4-D tensors. In Figure 23, the profile of the global eigenvalues and the PESDR curves are presented for the 4-D tensor with data that were recorded from volunteer 3 at the stimulation frequency  $f_{\text{stim}} = 1.00 \cdot f_\alpha$ . The LaRGE method overestimates the model

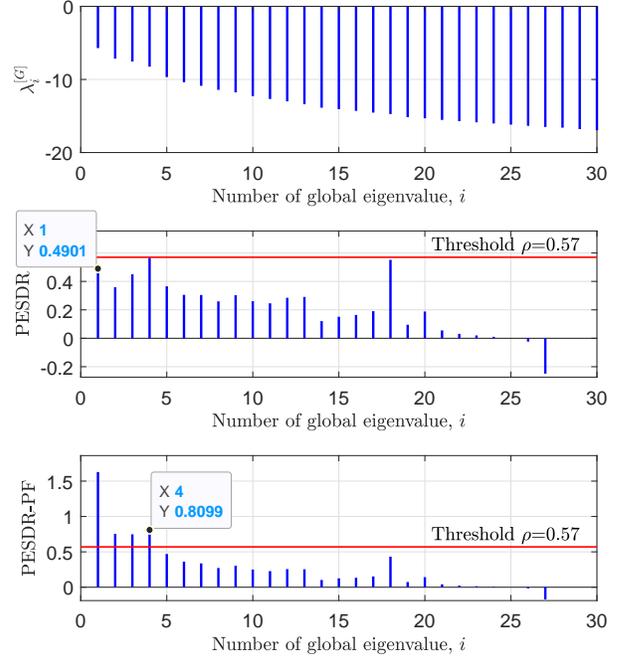


Fig. 19: The profile of the global eigenvalues (top), PESDR curve (middle) and PESDR-PF curve (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{103 \times 281 \times 30}$  of volunteer 11 at stimulation frequency  $0.60 \cdot f_\alpha$ .

order as 26. At the same time, LaRGE-PF estimates the model order as two. Using the topographic plots, three "bad" channels were found and removed. Then the analysis was repeated.

In Figures 24 and 25 the profile of the global eigenvalues, the PESDR curve, and the resulting factor matrices after the 4-D tensor decomposition are shown. Component one represents the expected resonance effect on the individual volunteer's alpha frequency.

## VII. CONCLUSION

In this paper, we have presented the robust LaRGE and LaRGE-PF methods of model order estimation based on the global eigenvalues of noise-corrupted low-rank tensors. Using the HOSVD of a measurement tensor, these global eigenvalues can be computed. Starting from the smallest global eigenvalues, the PESDR is calculated using linear regression. If it is larger than a predefined threshold  $\rho$  we have found the smallest signal global eigenvalue, i.e., the approximate model order  $\hat{R}$ . To prevent the misclassification in the first steps of this procedure, the modified LaRGE method with a heuristic penalty function (LaRGE-PF) has been proposed. LaRGE-PF suppresses the outliers of the PESDR curve for the tensors with small dimensions.

Monte Carlo simulations have been conducted to determine the best value of the threshold. To this end, the probabilities of false positive, false negative, and correct detection have been defined. Moreover, we have compared LaRGE and LaRGE-PF with the specified threshold  $\rho = 0.57$  with classical AIC,

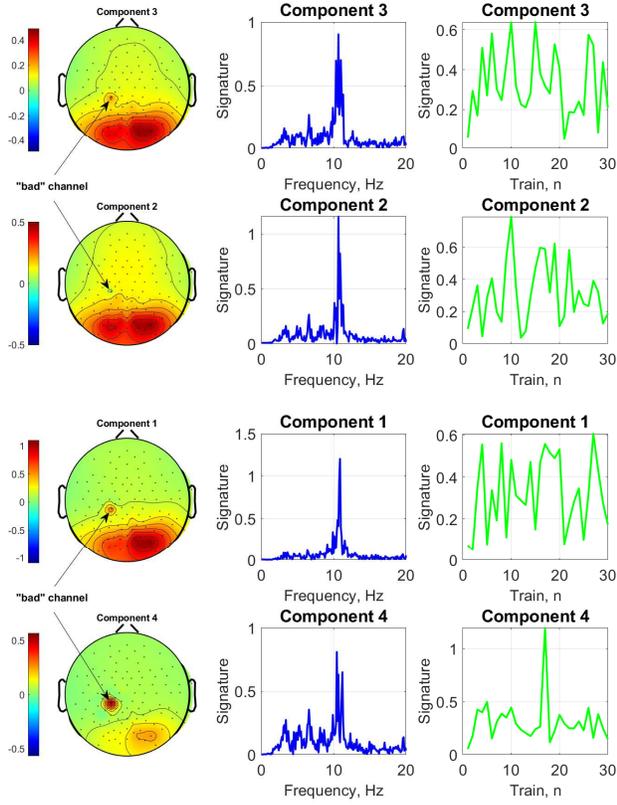


Fig. 20: Channel, frequency, and train signature for volunteer 11 and stimulation frequency  $0.6 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{103 \times 281 \times 30}$  and estimated rank 4. The loading factors of the components are  $\lambda_3 = 0.2883$ ,  $\lambda_2 = 0.2830$ ,  $\lambda_1 = 0.2751$ ,  $\lambda_4 = 0.2092$ . The arrows indicate a "bad" channel

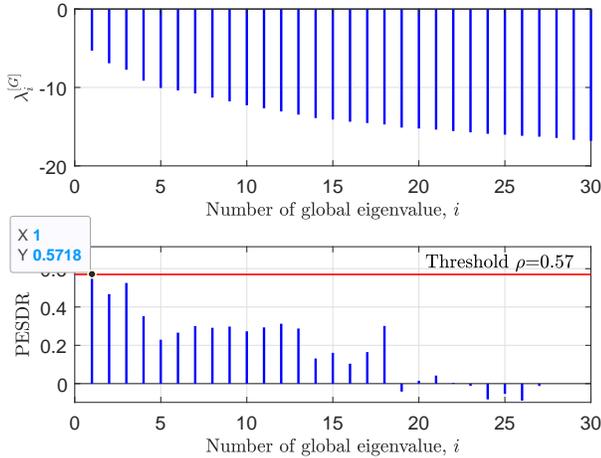


Fig. 21: The profile of the global eigenvalues (top) and PESDR curve (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{102 \times 281 \times 30}$  of volunteer 11 at stimulation frequency  $0.60 \cdot f_\alpha$  after "bad" channel removal.

MDL, 3-D AIC, 3-D MDL, 4-D AIC, and 4-D MDL methods. LaRGE and LaRGE-PF show an increased robustness and outperform the classical methods significantly.

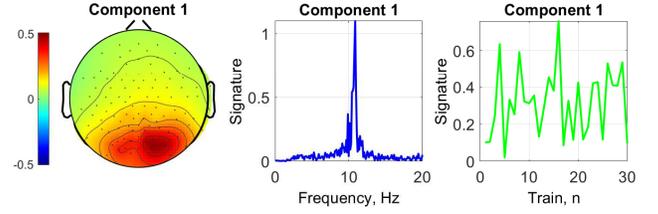


Fig. 22: Channel, frequency, time and train signature for volunteer 11 and stimulation frequency  $0.60 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{102 \times 281 \times 30}$  and estimated rank 1.

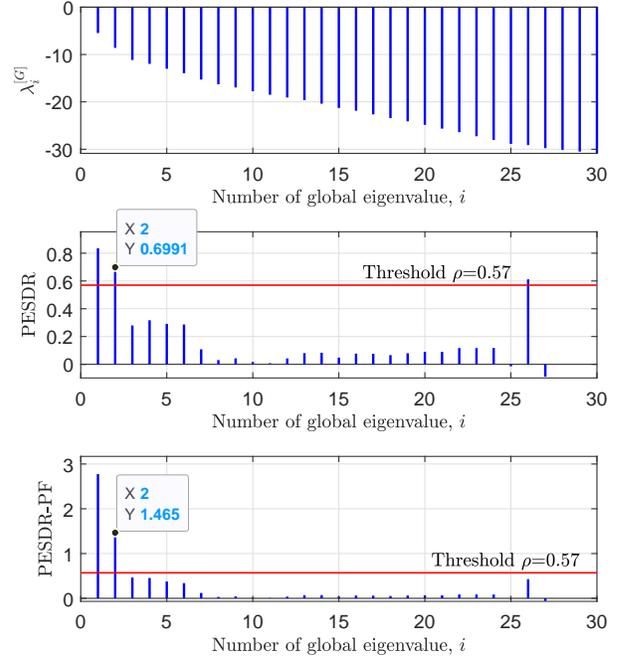


Fig. 23: The profile of the global eigenvalues (top), PESDR curve (middle) and PESDR-PF curve (bottom) computed for the tensor with real EEG data  $\mathcal{X} \in \mathbb{C}^{268 \times 429 \times 91 \times 30}$  of volunteer 3 at stimulation frequency  $1.00 \cdot f_\alpha$  before "bad" channels removal.

In this paper, we have also used the LaRGE and LaRGE-PF methods to estimate the model order of measured EEG data during an experimental investigation of the photic driving effect. The dominant components can be reliably detected by exploiting the PESDR curve.

If "bad" channels have been identified from the extracted dominant components, PESDR coefficients with a high amplitude could be observed. This effect might be very important for practical applications such as the automatic detection and removal of artifactual EEG channels and could be investigated in future research. Moreover, LaRGE could be extended to coupled tensors decompositions.

## REFERENCES

- [1] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiifa, and H. A. Phan, "Tensor decompositions for signal processing applica-

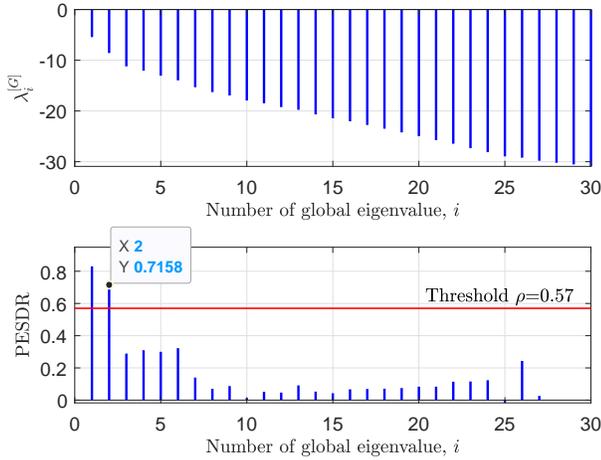


Fig. 24: The profile of the global eigenvalues (top) and PESDR curve (bottom) computed for the tensor with real EEG data Dimensions:  $\mathcal{X} \in \mathbb{C}^{268 \times 429 \times 88 \times 30}$  of volunteer 3 at stimulation frequency  $1.00 \cdot f_\alpha$  after "bad" channels removal.

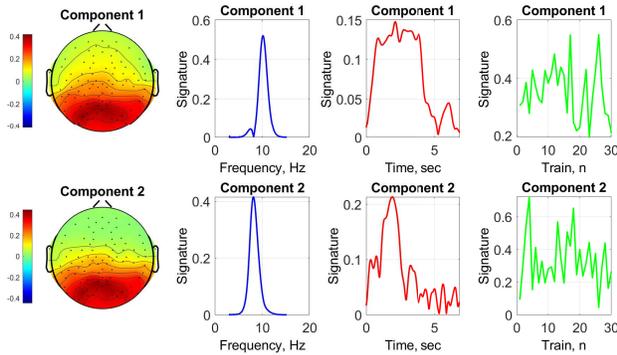


Fig. 25: Channel, frequency, time and train signature for volunteer 3 and stimulation frequency  $1.00 \cdot f_\alpha$  after the CP decomposition of the tensor with real EEG data Dimensions:  $\mathcal{X} \in \mathbb{C}^{268 \times 429 \times 88 \times 30}$  and estimated rank 1. The loading factors of the components are  $\lambda_1 = 0.3298$ , and  $\lambda_2 = 0.1777$

tions: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, March 2015.

[2] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, September 2009.

[3] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970.

[4] 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, pp. 283–319, 1970.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Analysis Applications*, vol. 21, pp. 1253–1278, 2000.

[6] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, p. 716–723, December 1974.

[7] J. Rissanen, "Modeling by shortest data description," *Automatic*, vol. 14, no. 5, p. 465–471, 1978.

[8] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, p. 461–464, 1978.

[9] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, April 1985.

[10] A. Quinlan, J.-P. Barbot, P. Larzabal, and M. Haardt, "Model order

selection for short data: An exponential fitting test (EFT)," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–11, 2006.

[11] J. P. C. L. da Costa, M. Haardt, F. Roemer, and G. Del Galdo, "Enhanced model order estimation using higher-order arrays," in *Proc. 2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, Nov 2007, pp. 412–416.

[12] J. P. C. L. da Costa, F. Roemer, M. Haardt, and R. T. de Sousa Jr, "Multi-dimensional model order selection," *EURASIP Journal on Advances in Signal Processing*, 2011.

[13] K. Liu, J. P. C. L. da Costa, H.-C. So, L. Huang, and J. Ye, "Detection of number of components in CANDECOMP/PARAFAC models via minimum description length," *Digit. Signal Process.*, vol. 51, pp. 110–123, 2016.

[14] T. Yokota, N. Lee, and A. Cichocki, "Robust multilinear tensor rank estimation using higher order singular value decomposition and information criteria," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1196–1206, March 2017.

[15] S. Pouryazdian, S. Beheshti, and S. Krishnan, "CANDECOMP/PARAFAC model order selection based on reconstruction error in the presence of Kronecker structured colored noise," *Digit. Signal Process.*, vol. 48, pp. 12–26, 2016.

[16] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *16th IEEE International Conference on Tools with Artificial Intelligence*, Nov 2004, pp. 576–584.

[17] R. Yu and E. A. Ince, "A reduced complexity tensor approach for order selection and frequency estimation," *IEEE Access*, pp. 1–1, 2021.

[18] A. A. Korobkov, M. K. Diugurova, J. Haueisen, and M. Haardt, "Multi-dimensional model order estimation using linear regression of global eigenvalues (LaRGE) with applications to EEG and MEG recordings," in *Proc. 28th European Signal Processing Conference (EUSIPCO 2020)*, Jan. 2021, pp. 1005–1009.

[19] R. Edlin, C. McCabe, C. Hulme, P. Hall, and J. Wright, *Correlated Parameters and the Cholesky Decomposition*. Cham: Springer International Publishing, 2015, pp. 119–132.

[20] F. Roemer and M. Haardt, "A semi-algebraic framework for approximate cp decompositions via simultaneous matrix diagonalizations (SECSI)," *Signal Processing*, vol. 93, no. 9, pp. 2722–2738, 2013.

[21] K. Schwab, C. Ligges, T. Jungmann, B. Hilgenfeld, J. Haueisen, and H. Witte, "Alpha entrainment in human electroencephalogram and magnetoencephalogram recordings," *Neuroreport*, vol. 17, pp. 1829–1833, 2006.

[22] C. Salchow, D. Strohmeier, S. Klee, D. Jannek, K. Schiecke, H. Witte, A. Nehorai, and J. Haueisen, "Rod driven frequency entrainment and resonance phenomena," *Frontiers in Human Neuroscience*, vol. 10, p. 413, 2016.

[23] K. Naskovska, S. Lau, A. A. Korobkov, J. Haueisen, and M. Haardt, "Coupled cp decomposition of simultaneous meg-eeeg signals for differentiating oscillators during photic driving," *Frontiers in Neuroscience*, vol. 14, p. 261, 2020.

[24] F. Roemer, C. Schroeter, and M. Haardt, "A semi-algebraic framework for approximate CP decompositions via joint matrix diagonalization and generalized unfoldings," in *Proc. 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers*, 2012, pp. 2023–2027.