GRAPH REGULARIZED TENSOR FACTORIZATION FOR SINGLE-TRIAL EEG ANALYSIS

Hayato Maki, Hiroki Tanaka, Sakriani Sakti, and Satoshi Nakamura

Nara Institute of Science and Technology, Japan

maki.hayato.lt3@is.naist.jp

ABSTRACT

This study proposes a tensor factorization algorithm for electroencephalographies (EEGs) that incorporates the geometric structure of the electrode location. The purpose is removing noise caused by EEG activities which are irrelevant to stimuli presented to a subject from single-trial event-related potential (ERP) data. Canonical polyadic decomposition (CPD) is extended by adding a regularization term that controls the spatial smoothness of the decomposed components on a scalp. An initialization method using geometrical information is also proposed. The geometric structure of an EEG signal is expressed as an undirected graph where the similarities between electrodes are defined by their relative distances on a scalp. The effectiveness is demonstrated in a noise-removing experiment using pseudo-ERP, where the proposed method achieved better performance than the conventional CPD.

1. INTRODUCTION

Electroencephalography (EEG) is an electrical recording of brain activities and has been widely used in such diverse research areas as psychology, cognitive science, medicine, and engineering. An event-related potential (ERP) is a time-locked brain response to stimuli and usually analyzed after trial-wise averaging to weaken noise that is irrelevant to stimuli presented to a subject [1]. Noise-removal (signal source separation) problems to realize single-trial ERP analysis have gathered much attention from researchers because such an averaging procedure conceals variabilities among trials.

Matrix factorization algorithms including independent component analysis (ICA) [2, 3] have been studied to tackle this problem. However, such algorithms can only be applied to a two-way array, i.e., a matrix, whereas ERP data usually have more than two modes, e.g., time, electrodes, frequencies, trials, subjects, and experimental conditions, which naturally leads us to expresses such data as a multi-dimensional array, or a tensor. Therefore, the application of tensor factorization algorithms to ERP data has been studied recently [4, 5]. Canonical polyadic decomposition (CPD, also known as CANDECOMP or PARAFAC) [6, 7] has been commonly used for its high interpretability and simplicity [8, 9, 10, 11].

However, it is well known that EEG data are noisy and sensitive to the outliers caused by body movements such as eye blinks. The amount of data is usually small, which causes the above statical noise-removal methods to slip into over-fitting and fail to find meaningful components.

Regularization can be used to avoid over-fitting by adding the norm of parameters or incorporating prior knowledge of the data [12]. For example, the geometrical information of face images was utilized to regularize the nonnegative matrix factorization [13]. The same regularization technique using the geometric structure of the EEG electrode location [14, 15] was used to create a common spatial patterns filter, which is widely used in two-class classification problems with EEG data. An EEG is a summation of an electrical recording of numerous neuron activities from which we can reasonably assume that multiple electrodes will observe signals that resemble each other if they are closely located on a scalp. However, existing CPD applications to EEG data have failed to utilize the spatial smoothness of EEGs and often identify spatially bumpy components [8, 11].

In this study, a new tensor factorization algorithm that is based on CPD and incorporates geometrical information of the EEG electrode location is proposed. In the rest of paper, the standard CPD is reviewed at first. Next, the graph regularized canonical polyadic decomposition (GCPD) is introduced. An initialization method using the graphical structure is also proposed. Finally, the effectiveness of the conventional and proposed methods are demonstrated in a noise-removing experiment.

2. BRIEF REVIEW OF CPD

A scalar, a vector, a matrix, and a tensor are respectively denoted by a standard Italic letter, a boldface Italic lowercase letter, a standard Roman capital letter, and an Euler script letter as a, $\boldsymbol{a}$, A, and $\mathcal{A}$. The i-th entry of a vector $\boldsymbol{a}$ is denoted by $\boldsymbol{a}(i)$. An element of a matrix that is specified by i-th row and j-th column is denoted by $\mathrm{A}(i,j)$. Similarly, an element of a tensor that is specified by the index of each mode $n_1, n_2, \ldots, n_N$ is denoted as $\mathcal{A}(n_1, n_2, \ldots, n_N)$.

A N-th order rank-1 tensor is the outer product of N vectors:

$$\begin{gathered}\mathcal{A} = \boldsymbol{a}^{(1)} \circ \boldsymbol{a}^{(2)} \circ \cdots \circ \boldsymbol{a}^{(N)} \\ \mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}, \quad \boldsymbol{a}^{(n)} \in \mathbb{R}^{I_n \times 1} \\ 1 \leq n \leq I_N,\end{gathered} \tag{1}$$

where $\circ$ is outer product of vectors.

CPD is formulated as a tensor decomposition into a sum of rank-1 tensors as depicted in Fig. 1:

$$\begin{aligned}\mathcal{X} \approx \hat{\mathcal{X}} &= \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} \\ &= \mathcal{I} \times_1 \mathrm{A}_1 \times_2 \mathrm{A}_2 \times_3 \cdots \times_N \mathrm{A}_N \\ \mathrm{A}_n &= \left(\boldsymbol{a}_1^{(n)}, \boldsymbol{a}_2^{(n)}, \ldots, \boldsymbol{a}_R^{(n)}\right) \in \mathbb{R}^{I_n \times R},\end{aligned} \tag{2}$$

where vector $\boldsymbol{a}_r^{(n)}$ is called the r-th basis of the mode-n, each rank-1 tensor $\boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(I_N)}$ is called the r-th component, $\mathcal{I}$ is the

This work was supported by JSPS KAKENHI Grant Numbers JP17H06101, JP17K00237, and JP16K16172.

identity tensor, and $\times_n$ is the n-mode product [16]. The constant R and the matrix A_n is called the CPD rank and the factor matrix of the n-th mode, respectively. The above decomposition is called rank-R CPD. This decomposition is unique under mild conditions up to the ambiguities for scaling among modes and for permutation among components [17]. We assume that input tensor $\mathcal{X}$ can be approximated by reconstructed tensor $\hat{\mathcal{X}}$.

Bases of each mode can be found by minimizing the following objective function $f_1(\mathcal{X}, \hat{\mathcal{X}})$:

$$f_1(\mathcal{X}, \hat{\mathcal{X}}) = \|\mathcal{X} - \hat{\mathcal{X}}\|_{\mathrm{F}} + \sum_{n=1}^{N} \lambda_n \|\mathrm{A}_n\|_{\mathrm{F}}, \tag{3}$$

where the constant λ_n is the norm regularization parameter, $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm. The first term of (3) is a divergence between $\mathcal{X}$ and $\hat{\mathcal{X}}$, and the second term is a norm regularization term. After the initialization of each factor matrix, the alternating gradient descent algorithm can be used for solving the minimization problem, which continues to update one factor matrix while fixing the others until convergence as follows:

$$\mathrm{A}_n \leftarrow \mathrm{A}_n - \eta \nabla_{\mathrm{A}_n} f_1, \tag{4}$$

where $\nabla_{\mathrm{A}_n} f_1$ is a gradient with respect to A_n and η is a learning rate.

The initialization of factor matrices is usually done using random numbers or higher order singular value decomposition (HOSVD) [18]. When HOSVD is used for initialization, an input tensor is unfolded into the matrices along with each mode at first. The mode-n unfolding (matricization) of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as the rearrangement of the tensor elements into the matrix $\mathcal{X}_n \in \mathbb{R}^{I_n \times I_1 I_2 \ldots I_{(n-1)} I_{(n+1)} \cdots I_N}$ along with the n-th mode [16]. Then, a normal SVD is applied for each unfolded matrix, and each factor matrix is initialized with the left singular vectors that correspond to the R largest singular values.

After the updating converges, noise free tensor $\mathcal{X}^*$ is obtained:

$$\mathcal{X}^* = \sum_{r \in P} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)}, \tag{5}$$

where P is a set of component indexes that correspond to the signal of interest.

3. GRAPH REGULARIZED TENSOR FACTORIZATION

In this section, CPD is extended to GCPD, which uses the geometrical information of the electrode location for both regularization and initialization. First, we explain a method that models the geometrical information of the electrode location using a graph. Second, we introduce a regularization method for CPD that constrains the spatial smoothness of the components. Third, we explain how to use the graph structure for initialization.

3.1. Obtaining Adjacency Matrix

The local structure between multiple nodes (e.g., electrodes) of the n-th mode can be expressed by an undirected graph and its corresponding adjacency matrix W_n, whose element $\mathrm{W}_n(i,j)$ expresses the similarity between the i-th and j-th nodes. One of the methods to measure the similarity between the two nodes is heat kernel weighting:

$$\mathrm{W}_n(i,j) = \exp\left(\frac{-\|\boldsymbol{z}_i - \boldsymbol{z}_j\|^2}{\sigma}\right), \tag{6}$$

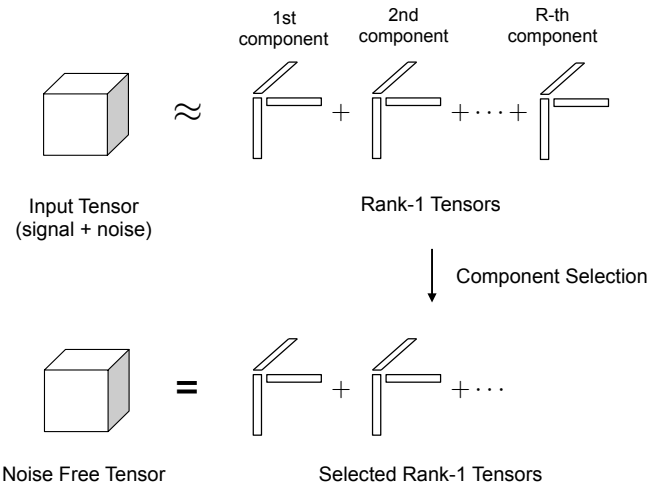

Fig. 1. Schematic image of CPD and the noise removal procedure. CPD decomposes an input tensor into the sum of the R rank-1 tensors. Then, a subset of the components are selected and summed to make a noise free tensor.

where $\boldsymbol{z}_i$ is the vector to express the i-th node and the constant σ is a variance parameter [19]. $\mathrm{W}_n(i,j)$ increases when the i-th and j-th nodes are similar and vice versa.

EEG electrodes are usually placed on a scalp according to the International 10-20, 10-10, or 10-5 systems [20, 21, 22]. The systems assume that the electrodes are placed on a sphere and define the i-th electrode position using a three-dimensional vector as $\boldsymbol{z}_i = (x_i, y_i, z_i)$ setting the sphere's center as the origin.

The spatial structure of EEG has been mentioned so far but the temporal smoothness can be also considered in the same way measuring distances between time indexes to construct a graph.

3.2. Graph Regularization

Given an adjacency matrix, the smoothness of the bases on a given graph can be measured using the following term:

$$\begin{aligned} S_n(\mathrm{A}_n, \mathrm{W}_n) &= \sum_{r=1}^{R} \sum_{i,j=1}^{I_N} \|\mathrm{A}_n(i,r) - \mathrm{A}_n(j,r)\|^2 \mathrm{W}_n(i,j) \\ &= \mathrm{Tr}(\mathrm{A}_n^\top \mathrm{L}_n \mathrm{A}_n) \qquad (7) \\ \mathrm{L}_n &= \mathrm{D}_n - \mathrm{W}_n \qquad (8) \end{aligned}$$

where $\mathrm{Tr}(\cdot)$ is the trace of a matrix, the matrix D_n is a diagonal matrix, whose diagonal elements are a column (or equivalently a row) sum of W_n. The matrix L_n is called a graph Laplacian [23]. In spectral graph theory, this term is called the graph Laplacian quadratic form, and increases when the difference between $\mathrm{A}_n(i,r)$ and $\mathrm{A}_n(j,r)$ is big although $\mathrm{W}_n(i,j)$ is big (e.g., the i-th and j-th nodes are close to each other on the graph) [19, 24].

Adding this term to Eq. (3), the objective function of GCPD is formulated as follows:

$$f_2(\mathcal{X}, \hat{\mathcal{X}}) = \|\mathcal{X} - \hat{\mathcal{X}}\|_{\mathrm{F}} + \sum_{n=1}^{N} \left(\lambda_n \|\mathrm{A}_n\|_{\mathrm{F}} + \psi_n S_n(\mathrm{A}_n, \mathrm{W}_n)\right), \tag{9}$$

where the constant ψ_n is a graph regularization parameter. By minimizing this function we expect that bases can be found that respect

the intrinsic graphical structure. A gradient of f_2 with respect to A_n is given as follows:

$$\nabla_{\mathrm{A}_n} f_2 = (\hat{\mathcal{X}}_n - \mathcal{X}_n)\mathrm{B}_n + (\lambda_n \mathrm{I} + \psi_n \mathrm{L}_n)\mathrm{A}_n \quad (10)$$

$$\mathrm{B}_n = \mathrm{A}_N \odot \mathrm{A}_{N-1} \odot \cdots \odot \mathrm{A}_{n+1} \odot \mathrm{A}_{n-1} \odot \cdots \odot \mathrm{A}_1, \quad (11)$$

where I is the identity matrix, $\odot$ is the product (column-wise Kronecker product) [25], and the matrix $\hat{\mathcal{X}}_n$ is the mode-n unfolding of $\hat{\mathcal{X}}$, which is calculated as:

$$\hat{\mathcal{X}}_n = \mathrm{A}_n \mathrm{B}_n^\top. \quad (12)$$

Not all the modes have to have the graphical structure because W_n can be set to the zero matrix if the graphical structure of the n-th mode is not defined. Iterative updating is performed as in Eq. (4).

3.3. Initialization Using Graph Fourier Bases

We propose to incorporate graphical smoothness not only for regularization but also initialization. The bases of each mode found by GCPD should respect its geometric structure: in other words, such bases make smoothness defined by Eq. (7) small. Therefore, the proposed initialization method is given by solving the following minimization problem:

$$\mathrm{A}_n^* = \arg\min S(\mathrm{A}_n, \mathrm{W}_n). \quad (13)$$

The solution of this minimization problem is given by the following eigenvalue problem [19]:

$$\mathrm{L}_n \boldsymbol{b}^{(n)} = p \boldsymbol{b}^{(n)}. \quad (14)$$

Let the dimension of the n-th mode be I_n. Since the graph Laplacian is a I_n–by–I_n real symmetric matrix, it has a complete set of the I_n eigenvectors $\{\boldsymbol{b}_r^{(n)}\}_{r=0,\ldots,I_n-1}$ associated with eigenvalues p_r that satisfy $0 = p_0 < p_1 \leq \cdots \leq p_{I_n-1}$. The eigenvectors of the graph Laplacian is called the graph Fourier bases (GFB) [24]. The GFBs corresponding to a smaller eigenvalue reduce the following term:

$$S'(\boldsymbol{b}_r^{(n)}, \mathrm{W}_n) = \sum_{i,j=1}^{I_N} \|\boldsymbol{b}_r^{(n)}(i) - \boldsymbol{b}_r^{(n)}(j)\|^2 \mathrm{W}_n(i,j). \quad (15)$$

In other words, the following relation holds:

$$0 = S'(\boldsymbol{b}_0^{(n)}, \mathrm{W}_n) < \cdots \leq S'(\boldsymbol{b}_{I_{N-1}}^{(n)}, \mathrm{W}_n). \quad (16)$$

In our proposed method, each factor matrix is initialized with the R GFBs corresponding to the R smallest eigenvalues.

While HOSVD initializes the factor matrices based on the orthogonality and the variance explained by the direction of the singular vectors and fails to consider the graph structures, the GFBs are also orthogonal as well as they consider the smoothness on a given graph. Moreover, they depend only on a graph Laplacian matrix. Therefore, they always produce the same initialization result with the same Laplacian matrix (the same EEG electrode location) even if a subject or an experimental condition is different. HOSVD is dependent on input data, which means the quality of initialization will be largely affected by the data amount and the artifacts in the data. This can be an advantage for the proposed method because EEGs are easily affected by such artifacts as body movements and the amount of data is often severely limited.

4. EXPERIMENTAL EVALUATION

A noise removal experiment is described in this section. Since it is difficult to discriminate which decomposed components really correspond to a *signal* (an EEG signal caused by brain activities of interest) or a *noise* (a signal caused by other brain activities or artifacts), a pseudo ERP data were created using real EEG signals to make it possible to evaluate the proposed method and the standard CPD objectively.

4.1. Data Acquisition

All EEG signals were recorded in a soundproof room with 25 scalp electrodes placed according to the International 10-20 system at a sampling rate of 1000 Hz. Three healthy subjects aged from 23 to 26 years old without any neurological disorders participated in the experiment. All experimental procedures were approved by the Ethical Review Board of Nara Institute of Science and Technology. The recorded EEG signals were down-sampled to 200 Hz, and a band-pass filter was applied between 0.01 Hz and 30 Hz.

An auditory oddball paradigm was used to elicit ERP of P300 from the subjects. A random sequence including 2 kHz and 1 kHz sound stimuli was presented to each subject by earphones. All the sound stimuli had a duration of 200 milliseconds and the intervals between them were 1.4 seconds. 2- and 1 kHz sounds were respectively presented 50 and 200 times. Subjects counted the number of 2 kHz sound stimuli (target trials) while ignoring 1 kHz sound (non-target trials). It is well documented that the ERP of P300 appears more conspicuously when a subject is presented a target stimuli than a non-target stimuli [1]. After repeating this procedure twice (100 target and 400 non-target trials in total), the subjects were told to relax without a task or a stimulus to record their resting state EEG for two minutes.

4.2. Validation Using Pseudo-ERP Data

From the resting state EEG signal of each subject, 100 epochs of 850 milliseconds duration (170 time samples) were extracted randomly. All the target-trials obtained from the oddball paradigm were averaged across trials. Then, the resulting averaged ERP that stands for *signal* was added to the each of 100 resting state epochs that stands for *noise* to make the 100 trials of pseudo-ERP data. The phase of the added ERP was randomly shifted by adding a random number sampled from the normal distribution with mean 0 ms and standard deviation 15 ms to its onset, and the amplitude was also randomly changed by multiplying a random number sampled from the normal distribution with mean 1 and standard deviation 0.4. The averaged non-target trial ERP was also added to another 100 resting state epochs as same as the target-trial ERP. Concatenating the two data, a third order tensor $\mathcal{X} \in \mathbb{R}^{170 \times 25 \times 200}$ with modes of $time \times space(electrodes) \times trials$ was made for each subject.

The noise-removing performances of conventional CPD, GCPD with HOSVD initialization, and GCPD with GFB initialization were compared. The rank of CPD and GCPD were set to 20 without tuning. The noise removal performances were measured by root mean squareRMSE).

4.3. Component Selection

As with other source separation techniques, CPD decomposes an input signal into multiple components without identifying the components of interest. Therefore, they must be selected based on prior knowledge [26].

Table 1. RMSE of all subjects

	RMSE		
	Subject 1	Subject 2	Subject 3
CPD	3.96 (s.d. 3.15)	7.12 (s.d. 6.29)	6.13 (s.d. 4.57)
GCPD with HOSVD	3.41 (s.d. 3.03)	7.12 (s.d. 6.29)	6.39 (s.d. 6.50)
GCPD with GFB	2.87 (s.d. 2.05)	6.88 (s.d. 4.60)	4.05 (s.d. 4.36)

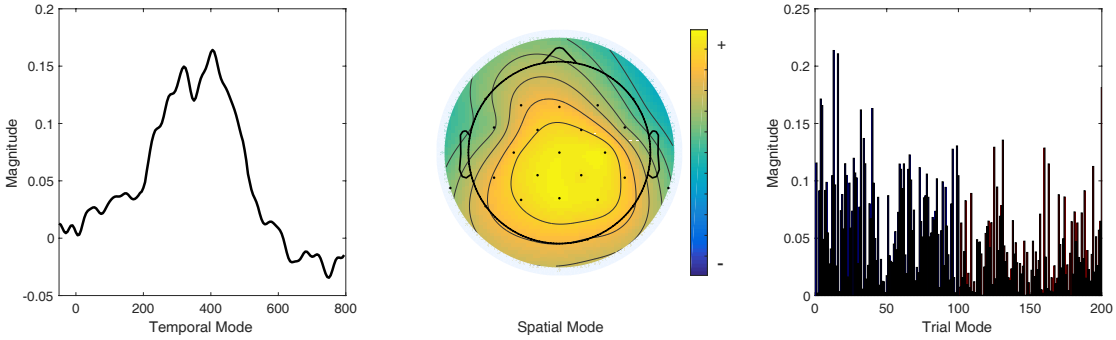

Fig. 2. Each mode of the component that had the highest ICV. (Left) the temporal mode, (middle) the spatial mode, and (right) the trial mode with the highest ICV. From the first to one-hundredth elements correspond to target trials (blue bars) and the rest to non-target trials (red bars)

Target-trials elicit larger ERP than non-target trials. If a component corresponds to an ERP, the magnitude of bases of its trial modes bases differ among the set of elements corresponding to target trials and the set of the non-target trials. On the other hand, if a component doesn't correspond to an ERP, the magnitude of the trial mode bases will take similar values among elements. Based on this idea, the inter-condition variance (ICV) of the r-th component is defined as follows:

$$\mathrm{ICV}(r) = \sum_{i \in P_T} \left| \mathrm{A}^{trial}(i,r) \right| - \sum_{i \notin P_T} \left| \mathrm{A}^{trial}(i,r) \right|, \qquad (17)$$

where P_T is a set of component indexes that corresponds to the target trials. Components with a high ICV were chosen as ERP components.

A previous work adopted a similar approach [8] and selected components based on the p-value of a statical test to the magnitude of the subject mode to investigate the difference between subjects with reading and with attention disabilities. ICV is simpler and more instinctive than the p-value of a statical significant test, which needs a sufficient amount of data to find significant differences and assumptions the about data, e.g., distribution or variance.

We did not use only ICV to select components but also manually inspected each component, and found that the result of the component selection by ICV agreed well with that of the manual selection. We selected the five components that had the highest ICV in this study.

4.4. Result

The noise-removal performance of GCPD was superior to CPD for all subjects when it was used with GFB initialization as summarized in Table 1. The averaged RMSE values with the standard deviation (s.d.) were shown. The average RMSE for each subject was calculated across all the 200 trials and the 25 electrodes (5000 RMSEs in total).

Fig. 2 illustrates the bases of the component with the highest ICV obtained by GFB-GCPD from the subject 3. The temporal mode basis shows the clear appearance of P300. The spatial mode basis shows the amplitude distribution on the scalp. It is spatially smooth and has the highest activity near the center of the scalp, which is physiologically plausible. The trial mode basis is shown for both of trial and non-target trials. The mode is important for single-trial analysis because it is possible to know that which trials affect to a cognitive state of a subject more strongly than others. From the first to one-hundredth elements correspond to target trials (shown in blue bars) and the rest to non-target trials (in red). It can be seen that magnitude are larger at the target-trials than the non-target ones, which is why this component is selected by ICV.

5. CONCLUSION

This paper proposed a tensor factorization method that incorporates the geometrical structure of the EEG electrode location.

Although our proposed algorithm's noise-removal performance was significantly superior to the conventional method, several questions remain uninvestigated. First, the decomposed components corresponding to ERP were automatically selected in a heuristic manner using the difference of the trial mode bases between experimental conditions. Even though this approach worked well in our study, a better component selection method might exist. Second, we solely measured the electrode similarities based on electrode location, but they can also be measured in other ways, for example, based on the brain's functional connectivity. Lastly, an initialization method using GFB was proposed. GFBs are orthogonal to each other as same as bases given by HOSVD initialization. However, it has been pointed out recently that orthogonality might not be a good property [27]. It is worth investigating an non-orthogonal initialization bases considering a graph structure.

6. REFERENCES

[1] Steven J Luck, *An introduction to the event-related potential technique*, MIT press, 2014.

[2] Tzyy-Ping Jung, Scott Makeig, Marissa Westerfield, Jeanne Townsend, Eric Courchesne, and Terrence J Sejnowski, "Analysis and visualization of single-trial event-related potentials," *Human Brain Mapping*, vol. 14, no. 3, pp. 166–185, 2001.

[3] Steven Lemm, Gabriel Curio, Yevhen Hlushchuk, and Klaus-Robert Muller, "Enhancing the signal-to-noise ratio of ICA-based extracted ERPs," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 4, pp. 601–607, 2006.

[4] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, John Wiley & Sons, 2009.

[5] Fengyu Cong, Qiu-Hua Lin, Li-Dan Kuang, Xiao-Feng Gong, Piia Astikainen, and Tapani Ristaniemi, "Tensor decomposition of EEG signals: a brief review," *Journal of Neuroscience Methods*, vol. 248, pp. 59–69, 2015.

[6] J Douglas Carroll and Jih-Jie Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of eckart-young decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[7] Richard A Harshman, "Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, 1970.

[8] Fengyu Cong, Anh Huy Phan, Qibin Zhao, Tiina Huttunen-Scott, Jukka Kaartinen, Tapani Ristaniemi, Heikki Lyytinen, and Andrzej Cichocki, "Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from EEG collected by low-density array," *International Journal of Neural Systems*, vol. 22, no. 06, 2012.

[9] Morten Mørup, Lars Kai Hansen, Christoph S Herrmann, Josef Parnas, and Sidse M Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, vol. 29, no. 3, pp. 938–947, 2006.

[10] Fumikazu Miwakeichi, Eduardo Martınez-Montes, Pedro A Valdés-Sosa, Nobuaki Nishiyama, Hiroaki Mizuhara, and Yoko Yamaguchi, "Decomposing EEG data into space–time–frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[11] Martin Weis, Florian Romer, Martin Haardt, Dunja Jannek, and Peter Husar, "Multi-dimensional space-time-frequency component analysis of event related EEG data using closed-form parafac," in *Proceedings of the 34th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'09)*. IEEE, 2009, pp. 349–352.

[12] Andreĭ Nikolaevich Tikhonov and Vasiliĭ Arsenin, *Solutions of ill-posed problems*, vol. 14, Winston Washington, DC, 1977.

[13] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.

[14] Hiroshi Higashi, Andrzej Cichocki, and Toshihisa Tanaka, "Regularization using geometric information between sensors capturing features from brain signals," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'12)*, 2012, pp. 721–724.

[15] Fabien Lotte and Cuntai Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 2, pp. 355–362, 2011.

[16] Tamara G Kolda and Brett W Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[17] Joseph B Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and Its Applications*, vol. 18, no. 2, pp. 95–138, 1977.

[18] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, "A multilinear singular value decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1253–1278, 2000.

[19] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proceedings of Advances in Neural Information Processing Systems 15 (NIPS'02)*, 2002, pp. 585–591.

[20] Robert Oostenveld and Peter Praamstra, "The five percent electrode system for high-resolution EEG and ERP measurements," *Clinical Neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.

[21] Marc R Nuwer, "Recording electrode site nomenclature," *Journal of Clinical Neurophysiology*, vol. 4, no. 2, pp. 121–134, 1987.

[22] Jaakko Malmivuo and Robert Plonsey, *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*, Oxford University Press, USA, 1995.

[23] Fan RK Chung, *Spectral Graph Theory*, Number 92. American Mathematical Soc., 1997.

[24] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[25] Age Smilde, Rasmus Bro, and Paul Geladi, *Multi-Way Analysis: Applications in the Chemical Sciences*, John Wiley & Sons, 2005.

[26] Piia Astikainen, Fengyu Cong, Tapani Ristaniemi, and Jari K Hietanen, "Event-related potentials to unattended changes in facial expressions: detection of regularity violations or encoding of emotions?," *Frontiers in Human Neuroscience*, vol. 7, 2013.

[27] Daichi Kitamura and Nobutaka Ono, "Efficient initialization for nonnegative matrix factorization based on nonnegative independent component analysis," in *Proceedings of the 16th IEEE International Workshop on Acoustic Signal Enhancement (IWAENC'16)*, 2016, pp. 1–5.