FROM BASIS COMPONENTS TO COMPLEX STRUCTURAL PATTERNS

Anh Huy Phan[‡], Andrzej Cichocki[‡], Petr Tichavský[•†], Rafal Zdunek[§] and Sidney Lehky[‡⋆]*

[‡]Brain Science Institute, RIKEN, Wakoshi, Japan
[•]Institute of Information Theory and Automation, Prague, Czech Republic
[§]Wroclaw University of Technology, Poland
[⋆]Computational Neurobiology Lab, The Salk Institute, USA

ABSTRACT

A novel approach is proposed to extract high-rank patterns from multiway data. The method is useful when signals comprise collinear components or complex structural patterns. Alternating least squares and multiplication algorithms are developed for the new model with/without non negativity constraints. Experimental results on synthetic data and real-world dataset confirm the validity of the proposed model and algorithms.

Index Terms— CANDECOMP/PARAFAC (CP), Kronecker tensor decomposition (KTD), PARALIND, block term decomposition, rank-overlap

1. PROBLEM FORMULATION

Tensor decompositions have become a valuable tool for analysis of data with multiple modes. The approach has found numerous applications such as in chemometrics, telecommunication, analysis of fMRI data, time-varying EEG spectrum, data mining, classification, clustering and compression [1–3]. One major application of matrix/tensor decompositions is to extract latent components from signals. Such analyses allow a deeper understanding of signals, or discrimination of one group of signals from others. For high order data (tensor), the decompositions retrieve multiple factor matrices, each of which comprises hidden components for each mode. For example, we can have temporal, spectral and spatial components for an order-3 tensor including time × frequency × channel. Most existing decompositions assume one-to-one interaction among basis components in factor matrices. For example, the well-known CANDECOMP/PARAFAC (CP) decomposition [4,5] approximates an order-N tensor $\mathcal{Y}$ of size $I_1 \times I_2 \times \cdots \times I_N$ by R rank-one tensors in which a component in a factor matrix $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ has only one interaction with components in other factor matrices, that is,

$$\mathcal{Y} \approx \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} = [\![\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!] \quad (1)$$

where $\circ$ denotes the outer vector product. This simple assumption unfortunately often cannot entirely characterize the data: hidden activities in real-world data can be generated by several sources, and they themselves can activate several other activities. For example, motor imagery activities in a BCI system yield several rhythms which usually occur over the motor and sensory cortices of the brain. This linear dependence problem also commonly appears in fluorescence excitation-emission data, or in flow injection analysis (FIA) [6].

A well-known tensor decomposition which establishes multi-interactions among components in factor matrices is the Tucker decomposition [7]. However, this decomposition allows too many interactions among components. One component can be associated with all components in all factor matrices except its own factor. To deal with the linear dependence problem and especially the rank-overlap problem, R. Bro *et. al.* proposed the structured CP decomposition (CPD) with additional dependence matrices $\mathbf{Q}_n$ which often consist of zeros and ones [1, 6], that is,

$$\mathcal{Y} \approx [\![\mathbf{A}^{(1)}\mathbf{Q}_1, \mathbf{A}^{(2)}\mathbf{Q}_2, \ldots, \mathbf{A}^{(N)}\mathbf{Q}_N]\!]. \quad (2)$$

A partial uniqueness of FIA-type PARALIND with doubly linear dependence was investigated in [6]. In another direction, L. de Lathauwer [8] restricted interactions among components in sub tensors instead of completely spreading entries of the core tensor in the Tucker decomposition, which leads to the rank-$(L_r, L_r, 1)$ block term decomposition (BTD) for order-3 tensors, and the generalized rank-$L_r \circ$ rank-1 BTD [9]

$$\mathcal{Y} \approx \sum_{r=1}^{R} \left(\mathbf{A}_r^{(1)} \mathbf{A}_r^{(2)T}\right) \circ \boldsymbol{a}_r^{(3)}. \quad (3)$$

Indeed the rank-$(L_r, L_r, 1)$ BTD is a particular PARALIND with a specific dependence matrix. BTD leads to rank-overlap in the third factor matrix in which each component is replicated L_r times. Note that because of rank-overlap, estimation of the first two factor matrices from order-3 tensors is not unique. Partial uniqueness conditions of PARALIND for order-3 tensors can be found in [10, 11].

In this paper, we formulate new structured CP decompositions for the decomposition of an order-N tensor into multiple Kronecker terms [12]. The first new CP-like model has factor matrices in the Kronecker form (product) which allows factorization of large-scale tensors with (very) high rank, say $R > 1000$. The second CP model is a particular case of the

*Also affiliated with the EE Dept., Warsaw University of Technology and with Systems Research Institute, Polish Academy of Science, Poland.

†The work of P. Tichavsky was supported by Grant Agency of the Czech Republic through the project 102/09/1278.

former one, but it can be considered as generalization of the PARALIND model for rank-overlap problems [6], or the BTD model [10]. The models have been verified on both synthetic and real data, using a real application for clustering.

2. NOVEL DECOMPOSITIONS

For a given multiway data $\mathcal{Y}$ of size $I_1 \times I_2 \times \cdots \times I_N$, we first consider the decomposition of this tensor into multiple Kronecker product terms (KTD) of smaller scale tensors $\mathcal{A}_p$ and $\mathcal{X}_p$, for $p = 1, 2, \ldots, P$ [12]

$$\mathcal{Y} \approx \sum_{p=1}^{P} \mathcal{A}_p \otimes \mathcal{X}_p, \tag{4}$$

where $\otimes$ denotes the generalized Kronecker product between two tensors $\mathcal{A}_p$ of size $J_{p1} \times J_{p2} \times \cdots \times J_{pN}$ and $\mathcal{X}_p$ of size $K_{p1} \times K_{p2} \times \cdots \times K_{pN}$ such that $I_n = J_{pn} K_{pn}$ [12]. Properties of the Kronecker tensor products are discussed in detailed in [12, 13]. Fig. 1 illustrates KTD of an order-3 tensors. Tensors $\mathcal{X}_p$ tile the entire data $\mathcal{Y}$. Small patches represent details of the tensor well, while large patterns capture smooth structures. Following this, we introduce a variation of this decomposition, and establish a connection from it to the well-known CP decomposition (1).

2.1. CPD with factor matrices in the Kronecker form

We assume that $\mathcal{A}_p$ and $\mathcal{X}_p$ are rank-L_p and rank-M_p tensors in the Kruskal form [3], respectively, that is

$$\mathcal{A}_p = [\![\mathbf{U}_p^{(1)}, \mathbf{U}_p^{(2)}, \ldots, \mathbf{U}_p^{(N)}]\!], \quad \mathcal{X}_p = [\![\mathbf{V}_p^{(1)}, \mathbf{V}_p^{(2)}, \ldots, \mathbf{V}_p^{(N)}]\!],$$

where $\mathbf{U}_p^{(n)} \in \mathbb{R}^{J_{p1} \times L_p}$ and $\mathbf{V}_p^{(n)} \in \mathbb{R}^{K_{pn} \times M_p}$, for all n. The new model is established in the following lemma.

Lemma 2.1 (CPD with factors in the Kronecker form)**.** *If $\mathcal{A}_p$ and $\mathcal{X}_p$ are rank-L_p and rank-M_p tensors in the Kruskal form, respectively, then the KTD model of $\mathcal{Y}$ is equivalent to a CP decomposition of rank $R = \sum_{p=1}^{P} L_p M_p$ given by*

$$\mathcal{Y} \approx [\![\mathbf{W}^{(1)}, \mathbf{W}^{(2)}, \ldots, \mathbf{W}^{(N)}]\!], \tag{5}$$

$$\mathbf{W}^{(n)} = \left[\mathbf{U}_1^{(n)} \otimes \mathbf{V}_1^{(n)}, \mathbf{U}_2^{(n)} \otimes \mathbf{V}_2^{(n)}, \ldots, \mathbf{U}_P^{(n)} \otimes \mathbf{V}_P^{(n)}\right] \in \mathbb{R}^{I_n \times R}. \tag{6}$$

The result implies an alternative interpretation and new application of the KTD model in which sub tensors have low-rank. Because the factors in the Kronecker form have a much smaller number of parameters than in the full CP model, the new model is suitable for high rank and large-scale tensors (say $R \geq 1000$).

2.2. Structured CPD with overlapped components

We consider a further simplification of the KTD model for collinear components or rank-overlap. A particular case is one in which $\mathcal{A}_p$ and $\mathcal{X}_p$ have singleton modes, i.e., $J_n = 1$ or $K_n = 1$ for all $n = 1, 2, \ldots, N$. For simplicity, we assume $K_1 = \ldots = K_S = 1$. This also means $J_n = I_n$, for $n = 1, 2, \ldots, S$. $\mathbf{U}^{(n)} = \mathbf{1}_{L_p}^T$, for $n = S+1, \ldots, N$ and $\mathbf{V}^{(n)} = \mathbf{1}_{R_p}^T$, for $n = 1, \ldots, S$.

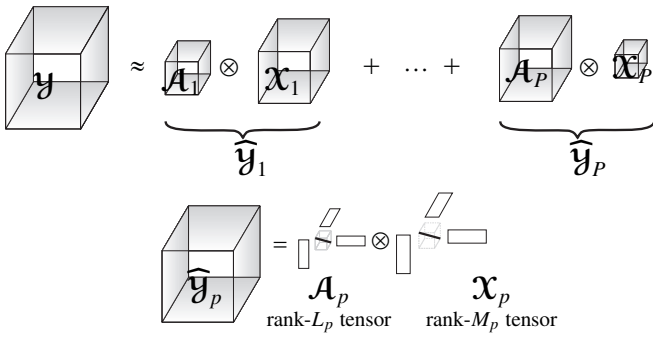


Fig. 1. Illustration of the tensor decomposition of an order-3 tensor $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ into P terms of Kronecker tensor products of $\mathcal{A}_p$, $\mathcal{X}_p$, and its particular case when $\mathcal{A}_p$ and $\mathcal{X}_p$ are rank-L_p and rank-M_p tensors.

Lemma 2.2. *When all $\mathcal{X}_p$ have S singleton modes ($1 \leq S < N$) with $K_1 = \ldots = K_S = 1$ and $K_n = I_n$ for $n = S+1, \ldots, N$, the KTD model in (4) and the CPD in (5) are simplified into a structured CPD of factor matrices $\mathbf{W}^{(n)}$ given by*

$$\mathbf{W}^{(n)} = \begin{cases} \widetilde{\mathbf{U}}^{(n)} \mathbf{Q}_L, & n = 1, 2, \ldots, S, \\ \widetilde{\mathbf{V}}^{(n)} \mathbf{Q}_M, & n = S+1, \ldots, N, \end{cases} \tag{7}$$

$$\widetilde{\mathbf{U}}^{(n)} = \left[\mathbf{U}_1^{(n)}, \mathbf{U}_2^{(n)}, \ldots, \mathbf{U}_P^{(n)}\right], \tag{8}$$

$$\widetilde{\mathbf{V}}^{(n)} = \left[\mathbf{V}_1^{(n)}, \mathbf{V}_2^{(n)}, \ldots, \mathbf{V}_P^{(n)}\right], \tag{9}$$

$$\mathbf{Q}_L = \texttt{blkdiag}(\mathbf{I}_{L_1} \otimes \mathbf{1}_{M_1}^T, \mathbf{I}_{L_2} \otimes \mathbf{1}_{M_2}^T, \ldots, \mathbf{I}_{L_P} \otimes \mathbf{1}_{M_P}^T), \tag{10}$$

$$\mathbf{Q}_M = \texttt{blkdiag}(\mathbf{1}_{L_1}^T \otimes \mathbf{I}_{M_1}, \mathbf{1}_{L_2}^T \otimes \mathbf{I}_{M_2}, \ldots, \mathbf{1}_{L_P}^T \otimes \mathbf{I}_{M_P}). \tag{11}$$

For this new decomposition, each component (column) of $\mathbf{U}_p^{(n)}$ is replicated M_p times in $\mathbf{W}^{(n)}$ for $n \leq S$, and each component of $\mathbf{V}_p^{(n)}$ is replicated L_p times in $\mathbf{W}^{(n)}$ for $n > S$. Such behavior is related to the rank-overlap problem which often exists in real-world signals such as chemical data, flow injection analysis (FIA) data [1, 6], or spectral tensors of EEG signals. The structured CPD in Lemma 2.2 is a particular case of the CP model with linearly dependent loadings (PARALIND) in (2) [6] in which the dependency matrices are fixed and given in Lemma 2.2. Discussions on uniqueness of the CPD with linearly dependent loadings can be found in [11, 14].

If all $\mathcal{X}_p$ are rank-1, i.e., $M_p = 1$, for all p, the KTD model and the structured CPD model in Lemma 2.2 is simplified into the rank-$(L_r, L_r, 1)$ block term decomposition (BTD) for order-3 tensors [8], and the rank-$L_r \circ$ rank-1 BTD [9] (see (3)).

3. ALGORITHMS

Algorithms for the generalized model (4) without specific rank have been developed in [12]. For two simplified tensor decompositions in Lemmas 2.1 and 2.2, CPD-like algorithms can be efficiently used such as the ALS algorithm [6] (often with line search extrapolation methods [15–17]) or the fast damped Gauss-Newton algorithm [18]. For space reasons, we present the ALS algorithm for the structured CPD in Lemma 2.2 which sequentially updates $\widetilde{\mathbf{U}}^{(n)}$ and $\widetilde{\mathbf{V}}^{(n)}$

$$\widetilde{\mathbf{U}}^{(n)} \leftarrow \mathbf{G}_n \mathbf{Q}_L^T \left(\mathbf{Q}_L \mathbf{\Gamma}_n \mathbf{Q}_L^T\right)^{-1}, \quad (n = 1, 2, \ldots, S), \tag{12}$$

Table 1. MSAE (in dB) in decomposition of random tensors in Example 4.1.

SNR (dB)	Bipolar data: Approx. Error	Bipolar data: MSAE (dB) $\mathbf{U}_p^{(n)}$	Bipolar data: MSAE (dB) $\mathbf{V}_p^{(n)}$	Nonnegative data: Approx. Error	Nonnegative data: MSAE (dB) $\mathbf{U}_p^{(n)}$	Nonnegative data: MSAE (dB) $\mathbf{V}_p^{(n)}$
0	$7.07\ 10^{-1}$	40.41	41.62	$7.07\ 10^{-1}$	31.00	34.95
10	$3.02\ 10^{-1}$	50.01	52.54	$3.01\ 10^{-1}$	41.37	44.74
Inf	$6.28\ 10^{-8}$	136.31	155.96	$8.57\ 10^{-6}$	71.92	89.13

$$\widetilde{\mathbf{V}}^{(n)} \leftarrow \mathbf{G}_n \mathbf{Q}_M^T \left(\mathbf{Q}_M \mathbf{\Gamma}_n \mathbf{Q}_M^T\right)^{-1}, \quad (n = S+1, \ldots, N), \quad (13)$$

where $\mathbf{G}_n = \mathbf{Y}_{(n)} \left(\mathbf{W}^{(N)} \odot \cdots \odot \mathbf{W}^{(n+1)} \odot \mathbf{W}^{(n-1)} \odot \cdots \odot \mathbf{W}^{(1)}\right)$, $\mathbf{\Gamma}_n = \left(\mathbf{W}^{(1)T}\mathbf{W}^{(1)}\right) \circledast \cdots \circledast \left(\mathbf{W}^{(n-1)T}\mathbf{W}^{(n-1)}\right) \circledast \left(\mathbf{W}^{(n+1)T}\mathbf{W}^{(n+1)}\right) \circledast \cdots \circledast \left(\mathbf{W}^{(N)T}\mathbf{W}^{(N)}\right)$, $\odot$ and $\circledast$ denote the Khatri-Rao and Hadamard product, respectively. With non negativity constraints, we can derive the multiplicative algorithm as follows [19–22]

$$\widetilde{\mathbf{U}}^{(n)} \leftarrow \widetilde{\mathbf{U}}^{(n)} \circledast \left(\mathbf{G}_n \mathbf{Q}_L^T\right) \oslash \left(\widetilde{\mathbf{U}}^{(n)} \mathbf{Q}_L \mathbf{\Gamma}_n \mathbf{Q}_L^T\right), \ (n = 1, 2, \ldots, S)$$
$$\widetilde{\mathbf{V}}^{(n)} \leftarrow \widetilde{\mathbf{V}}^{(n)} \circledast \left(\mathbf{G}_n \mathbf{Q}_M^T\right) \oslash \left(\widetilde{\mathbf{V}}^{(n)} \mathbf{Q}_M \mathbf{\Gamma}_n \mathbf{Q}_M^T\right), \ (n = S+1, \ldots, N)$$

where $\oslash$ denote (element-wise) Hadamard division.

4. SIMULATIONS

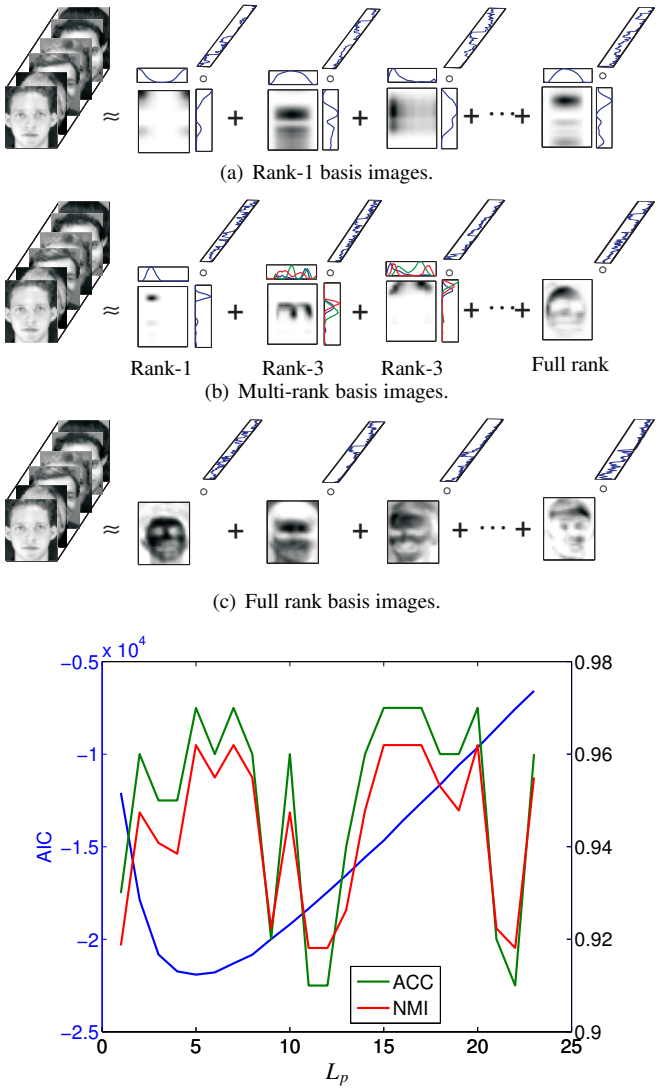
4.1. Synthetic data

This example aimed to verify the correctness of the proposed model in Lemma 2.2 for order-6 data synthetic tensors. The tensors $\mathcal{Y}$ of size $I_n = 10$ for all n were composed from order-3 tensors $\mathcal{A}_p$ and $\mathcal{X}_p$ with 3 singleton modes, i.e., $S = 3$, and corrupted by Gaussian noise of SNR = 0 and 10 dB. Ranks of sub tensors were set to $[L_p] = [2, 2, 3]$ and $[M_p] = [1, 2, 2]$. We generated 60 noiseless tensors (30 tensors consisting of nonnegative values), and 30 noisy tensors from each noiseless one for each noise level. Hence, there were in total 3660 tensors. Factor matrices were initialized using random values.

Quality of the decompositions was evaluated through mean square angular error (MSAE) between an estimated component and its original one over multiple runs MSAE $= -10\log_{10}(E[\text{acos}^2 \frac{\boldsymbol{a}^T\hat{\boldsymbol{a}}}{\|\boldsymbol{a}\|\,\|\hat{\boldsymbol{a}}\|}])$ dB [23, 24]. Table 1 summarized the average approximation errors, and MSAEs for all components of $\tilde{\mathbf{U}}^{(n)}$ and $\tilde{\mathbf{V}}^{(n)}$. Note that an MSAE higher than 40 dB, 50 dB means that two components are different by a mutual angle on average less than 0.57^o, 0.18^o, respectively. The high performance of MSAE (in dB) confirms the efficiency of the algorithm and the correctness of the proposed model.

4.2. Factorization of the ORL face database

This example demonstrated the superiority and the advantages of the structured CPD in Lemma 2.2 or BTD to the standard CPD in feature extraction. We used 100 faces for the first 10 subjects in the ORL face database [25]. Images were down-sampled to the size of 28×23 to give an order-3 tensor of size $28 \times 23 \times 100$. By applying nonnegative CPD, we can extract rank-one basis images $\boldsymbol{u}_r^{(1)} \boldsymbol{u}_r^{(2)T}$, each of which was associated with a vector feature $\boldsymbol{u}_r^{(3)}$, and expressed local structures as illustrated in Fig. 2(a). It is obvious that such local rank-one structures did not fully reveal common patterns among all faces such as eyes, mouth, eyebrow, forehead, nose which can be extracted by nonnegative matrix factorization [2, 21, 22]. Fig. 2(c) illustrates full rank basis images which expressed global and more complex structures, obtained through the structured CPD with $L_p = 23$ for all p. An important observation is that by varying rank of the basis images from rank-one to full-rank, we can accordingly extract local simple or global complex structures. For example, in Fig. 2(b), we extracted one rank-one basis image ($L_1 = 1$),



(a) Rank-1 basis images.

(b) Multi-rank basis images.

(c) Full rank basis images.

(d) AIC, Accuracy and NMI.

Fig. 2. Example for clustering the ORL face dataset by different feature extraction models. Complexity of basis images was varied according to their ranks (rank-1 to full-rank).

two rank-three basis images ($L_2 = L_3 = 3$) for eyebrow-nose and forehead, and one last full-rank basis image ($L_P = 23$) for common background of facial images.

Based on the above observation, we considered simple decompositions with identical L_p and $M_p = 1$ for all p to extract $P = 10$ nonnegative vector features from the facial tensor. The model was simplified to rank-$(Lp, Lp, 1)$ BTD [8] with nonnegativity constraints

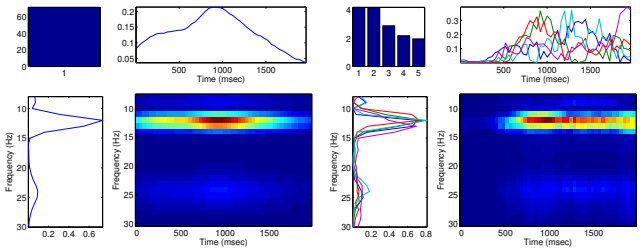
$$\mathcal{Y} \approx \sum_{p=1}^{P} [\![\mathbf{U}_p^{(1)}, \mathbf{U}_p^{(2)}, \boldsymbol{v}_p \mathbf{1}_{L_p}^T]\!], \tag{14}$$

where $\mathbf{U}_p^{(1)} \in \mathbb{R}^{28 \times L_p}$, $\mathbf{U}_p^{(2)} \in \mathbb{R}^{23 \times L_p}$ and $\boldsymbol{v}_p \in \mathbb{R}^{100}$, for all $p = 1, \ldots, P$. The ranks L_p were varied from within the range [1, 23]. Vector features $\boldsymbol{v}_p$ were associated with basis images $\mathbf{F}_p = \mathbf{U}_p^{(1)} \mathbf{U}_p^{(2)T}$ of size 28×23 for $p = 1, \ldots, P$. The multiplicative algorithm in section 3 was applied. Since the decomposition could get stuck in local minima, for each rank L_p, we factorized the tensor 40 times to find 40 sets of 10 vector features. On the basis of extracted features, categorical labels of images were first predicted using the k-means algorithm. The final labels were then defined as the most frequent predicted values. In Fig. 2(d), in addition to the accuracy and normalized mutual information (NMI), we provide the Akaike Information Criterion (AIC) for balancing the approximation error and the number of estimated parameters for different models [26]. The results indicate that with $L_p = 5, 6, 7$, we can achieve the clustering accuracy of 97% compared with the accuracy of 92% using the simple CPD with $L_p = M_p = 1$ for all p. Moreover the selected model also coincided with that having the minimum AIC.

4.3. Factorization of EEG Motor Imagery Data

In this simulation, we emphasized the efficiency of high rank structural pattern extraction as compared with the traditional rank-1 basis component for factorization of high order tensor which involves left/right motor imagery (MI) movements. We analyzed the EEG MI dataset [27] for subjects 1 and 2. EEG signals were transformed into the time-frequency domain using the complex Morlet wavelets CMOR6-1 giving an order-4 tensor of dimension 23 *frequency bins* (8-30 Hz) $\times$ 50 *time frames* $\times$ 62 *channels* $\times$ 200 *trials* [28].

It is well-known that in preparation and imagination of movement the mu and beta rhythms are desynchronized over the contralateral sensory and motor areas [29]. However, in practice, the EEG signals cannot comprise only one rhythm at a specific frequency band during a fixed time frame, but there exists groups of similar rhythms in the signals. Fig. 3 compares rank-1 ($L_p = 1$) and rank-5 ($L_p = 5$) spectrograms extracted from the MI tensor. Rank-one spectral maps can extract only the major activity of the MI signals, but cannot capture a group of similar components. It is obvious that although five spectral components in Fig. 3(b) explain rhythms in the same frequency range, they are not identical. Moreover,



(a) Rank-1 spectrogram. (b) Rank-5 spectrogram.

Fig. 3. Comparison between basis spectral-temporal maps which have rank-1 (a) and rank-5 (b) extracted from the MI tensor.

their associated temporal components convey more information than the rank-one temporal component in Fig. 3(a). Finally, according to the appearances of the spectrogram, the higher rank spectrogram is more natural and realistic than the one of rank-one. On the basis of extracted features from $P = 10$ rank-one patterns *channel* $\times$ *trial*, i.e., $S = 2$, MI activities were clustered with respect to left and right movements. However, instead of using all the features, we selected only 2-3 components related to the MI activities, that is, their corresponding spatial components distributed over the contralateral sensory and motor areas. The clustering accuracies using CPD were 71% and 79.5% for subjects 1 and 2, respectively. These accuracies can be improved up to 83.5% for subject 1 with rank-5 basis spectrograms, and 85% for subject 2 using rank-7 basis spectrograms, respectively.

5. CONCLUSIONS

Here we introduce a method to decompose tensors into the sum of Kronecker product terms of smaller scale tensors. This allows us to represent multiway data in terms of components of variable complexity depending on the rank of the smaller tensors, which is set during the decomposition process. The method can be used when there are collinear components (that is, rank-overlap exists). An issue that needs to be better understood for this method is the large, non-systematic fluctuations in classification accuracy of real data when the rank of extracted components was varied (Fig. 2(d)). One controversy related to biological signal processing in the brain is the extent to which sensory signals are decomposed into either local or global components. For example, for face processing in the brain, there is a debate whether populations of neurons decompose the visual input into a set of local features (mouth, nose, eyes, etc.) or into a more holistic representation [30]. The algorithm developed here allows the creation of basis sets with graded degrees of complexity ranging from highly local to highly global, expanding the conceptual possibilities when analyzing or modeling neurophysiological and brain imaging data. Selection of a suitable model for real-world data is still an open issue, but can be efficiently solved using the greedy method in [31].

6. REFERENCES

[1] R. Bro, *Multi-way Analysis in the Food Industry - Models, Algorithms, and Applications*, Ph.D. thesis, University of Amsterdam, Holland, 1998.

[2] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, Chichester, 2009.

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

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

[5] J.D. Carroll and J.J. Chang, "Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckart–Young decomposition," *Psychometrika*, vol. 35, no. 3, pp. 283–319, 1970.

[6] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, "Modeling multi-way data with linearly dependent loadings," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 324–340, 2009.

[7] L.R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279–311, 1966.

[8] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms – Part I: Lemmas for partitioned matrices," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1022–1032, 2008, Special Issue on Tensor Decompositions and Applications.

[9] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(lr,lr,1) terms and a new generalization," Tech. Rep., University of Leuven, 2012.

[10] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms – Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008, Special Issue on Tensor Decompositions and Applications.

[11] X. Guo, S. Miron, D. Brie, and A. Stegeman, "Uni-mode and partial uniqueness conditions for candecomp/parafac of three-way arrays with linearly dependent loadings," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 1, pp. 111–129, 2012.

[12] A.-H. Phan, A. Cichocki, P. Tichavský, D. P. Mandic, and K. Matsuoka, "On revealing replicating structures in multi-way data: A novel tensor decomposition approach," in *Latent Variable Analysis and Signal Separation*, vol. 7191 of *Lecture Notes in Computer Science*, pp. 297–305. Springer Berlin Heidelberg, 2012.

[13] S. Ragnarsson, *Structured Tensor Computations: Blocking, Symmetries And Kronecker Factorizations*, Ph.D. thesis, Cornell University, 2012.

[14] A. Stegeman and T. Lam, "Improved uniqueness conditions for canonical tensor decompositions with linearly dependent loadings," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 4, pp. 1250–1271, 2012.

[15] G. Tomasi, *Practical and Computational Aspects in Chemometric Data Analysis*, Ph.D. thesis, Frederiksberg, Denmark, 2006.

[16] M. Rajih, P. Comon, and R. A. Harshman, "Enhanced line search: A novel method to accelerate PARAFAC," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1128–1147, 2008.

[17] Y. Chen, D. Han, and L. Qi, "New ALS methods with extrapolating search directions and optimal step size for complex-valued tensor decompositions," *IEEE Trans. Signal Process*, vol. 59, no. 12, pp. 5888–5898, 2011.

[18] A.-H. Phan, P. Tichavský, and A. Cichocki, "Low complexity damped Gauss-Newton algorithms for CANDECOMP/PARAFAC," *SIAM J. Matrix Anal. Appl.* vol 34(1), pp. 126–147, 2013.

[19] M.E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorthm suitable for volume ECT," *IEEE Trans. Med. Imag.*, vol. 5, pp. 61–66, 1986.

[20] H Lantéri, R. Soummer, and C. Aime, "Comparison between ISRA and RLA algorithms: Use of a Wiener filter based stopping criterion," *Astronomy and Astrophysics Supplemantary Series*, vol. 140, pp. 235–246., 1999.

[21] D.D. Lee and H.S. Seung, "Learning of the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[22] C. J. Lin, "Projected gradient methods for non-negative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756–2779, October 2007.

[23] P. Tichavský and Z. Koldovský, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Trans. Signal Process*, vol. 59, no. 3, pp. 1037–1047, 2011.

[24] P. Tichavský, A. H. Phan, and Z. Koldovský, "Cramér-Rao-Induced Bounds for CANDECOMP/PARAFAC tensor decomposition," *IEEE Trans. Signal Process*, accepted for publication, 2013.

[25] F. Samaria and A.C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, 1994.

[26] P. Stoica and Y. Selén, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[27] Center for Brain-Like Computing and Machine Intelligence, "Data set for single trial EEG classification in BCI," http://bcmi.sjtu.edu.cn/data1/.

[28] A.-H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *Nonlinear Theory and Its Applications, IEICE*, vol. 1, pp. 37–68 (invited paper), 2010.

[29] G. Pfurtscheller and Lopes F. H. da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin Neurophysiol*, vol. 110, pp. 1842–1857, 1997.

[30] D.Y. Tsao and M.S. Livingstone , "Mechanisms of face perception," *Annu. Rev. Neurosci.*, vol. 31, pp. 411–437, 2008.

[31] A. J. Brockmeier, J. C. Principe, A.-H. Phan, and A. Cichocki, "A greedy algorithm for model selection for tensor decompositions," ICASSP 2013.