Journal of Neuroscience Methods 248 (2015) 59–69

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational neuroscience

Tensor decomposition of EEG signals: A brief review

Fengyu Cong[a,b,*], Qiu-Hua Lin[c], Li-Dan Kuang[c], Xiao-Feng Gong[c], Piia Astikainen[d], Tapani Ristaniemi[b]

[a] Department of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China
[b] Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland
[c] School of Information and Communication Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China
[d] Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

HIGHLIGHTS

- EEG signals are naturally born with multi modes.
- EEG signals can be represented by the high-order multi-way array, tensor.
- Tensor of EEG can be exploited by tensor decomposition for multi-way analysis.

ARTICLE INFO

Article history:
Received 29 December 2014
Received in revised form 17 February 2015
Accepted 12 March 2015
Available online 1 April 2015

Keywords:
Event-related potentials
EEG
Tensor decomposition
Canonical polyadic
Tucker
Brain
Signal

ABSTRACT

Electroencephalography (EEG) is one fundamental tool for functional brain imaging. EEG signals tend to be represented by a vector or a matrix to facilitate data processing and analysis with generally understood methodologies like time-series analysis, spectral analysis and matrix decomposition. Indeed, EEG signals are often naturally born with more than two modes of time and space, and they can be denoted by a multi-way array called as tensor. This review summarizes the current progress of tensor decomposition of EEG signals with three aspects. The first is about the existing modes and tensors of EEG signals. Second, two fundamental tensor decomposition models, canonical polyadic decomposition (CPD, it is also called parallel factor analysis-PARAFAC) and Tucker decomposition, are introduced and compared. Moreover, the applications of the two models for EEG signals are addressed. Particularly, the determination of the number of components for each mode is discussed. Finally, the N-way partial least square and higher-order partial least square are described for a potential trend to process and analyze brain signals of two modalities simultaneously.

Contents

* Corresponding author at: Department of Biomedical Engineering, Dalian University of Technology, Dalian 116024, China. Tel.: +86 158 411 92277.
E-mail addresses: cong@dlut.edu.cn, Fengyu.Cong@aliyun.com (F. Cong).

http://dx.doi.org/10.1016/j.jneumeth.2015.03.018

1. Introduction

Data processing and analysis plays a fundamental role in brain research using brain imaging tools. The recorded brain imaging data can be represented by a one-way series (called as a vector), a two-way array (called as a matrix), and a multi-way array (called as a tensor). Different signal processing and analysis methods are applied in terms of different ways of data representation.

1.1. One-way and two-way brain signals

For example, in the early stage of electroencephalography (EEG) studies, EEG data were represented by a time series and all data samples were carried by a vector (Berger, 1929). Since then, power spectrum analysis of the time series has been often applied for investigating EEG oscillations (Cohen, 2014; Niedermeyer and Lopes da Silva, 2005). Recently, time-frequency analysis (TFA) of the time series has been very attractive for analyzing single-trial EEG (Herrmann et al., 2013). Nowadays, multiple electrodes are often used to collect EEG data in the experiment. Therefore, EEG recordings naturally include two modes of time and space, at least. Subsequently, a matrix with the two modes has been extensively used to represent the EEG data. In some professional software for EEG data processing and analysis, EEG data are displayed on the screen (Delorme and Makeig, 2004). The horizontal axis of the plane (the row of a matrix) is for the time mode, and the vertical axis of the plane is for the space mode (the column of the matrix). As a result, the two-way signal processing methods including principal component analysis (PCA) and independent component analysis (ICA) have been performed on the matrix to remove artifacts and to extract brain activities of interest (Dien, 2012; Onton and Makeig, 2006; Vigario and Oja, 2008).

1.2. Multi-way nature of EEG signals

Indeed, in EEG experiments, usually there are more modes than the two modes of time and space. For instance, analysis of EEG signals may compare responses recorded in different subject groups (e.g. comparison of responses in a healthy control group and a clinical group). Thus, at least one more mode appears and it is the subject. Furthermore, in an experiment to elicit event-related potentials (ERPs), there are modes of EEG trial (since several stimulus presentations are required) and stimulus presentation condition. This means the brain data collected by EEG techniques can be naturally fit into a multi-way array including multiple modes.

However, the mostly applied computing tools for brain research are oriented for one-way or two-way data. Consequently, in order to facilitate the two-way signal processing methods, the extra modes besides the two modes of time and space are often concatenated (data are horizontally connected in a plane) or stacked (data are vertically connected in a plane) with the time or the space mode for generating a matrix (Calhoun and Adali, 2012; Cong et al., 2013b, 2014a; Delorme and Makeig, 2004; Dien, 2012; Eichele et al., 2011). This is often called unfolding a multi-way array into a matrix. For EEG data, such unfolding inevitably loses some potentially existing interactions between/among the folded modes, such as time, frequency and space modes. The interactions can be of research interest. Consequently, in order to appropriately reveal the interactions among multiple modes, the signal processing methods particularly for a multi-way array are naturally promising tools.

1.3. Multi-way array is a tensor, a new way to represent and analyze data

A multi-way array is named as a tensor (Cichocki et al., 2009; Kolda and Bader, 2009). For a matrix, matrix decomposition can be applied for data processing and analysis. Analogously, for a tensor, tensor decomposition can be applied as well. Tensor decomposition inherently exploits the interactions among multiple modes of the tensor. It was first defined in the field of mathematics (Hitchcock, 1927), and has been glorious in the fields of psychometrics and chemometrices for multi-way data analysis (Kroonenberg, 2008; Smilde et al., 2004). The existing key reviews for tensor decomposition often include its history, models, algorithms and various applications (Acar and Yener, 2009; Cichocki et al., 2015; Comon, 2014; Comon et al., 2009; Khoromskij, 2011; Kolda and Bader, 2009; Lu et al., 2011; Morup, 2011).

Recently, tensor decomposition has become surprisingly attractive for signal processing (Cichocki et al., 2015). Indeed, it has already been applied for analysis of ERPs in 1980s (Mocks, 1988a,b). In the past ten years, there have been many reports about tensor decomposition for processing and analyzing EEG signals. However, there is no review particularly for tensor decomposition of EEG signals yet. Therefore, this study is devoted to summarizing previous reports concerned with tensor decomposition of EEG signals and discussing the key issues regarding the application.

In the rest of the paper, next, tensors of EEG signals and tensor decomposition models are introduced. Subsequently, how tensor decomposition can be applied for processing and analyzing tensors of EEG signals is addressed. Finally, the potential trend for analyzing data of two modalities in brain research is stated. Since superiority

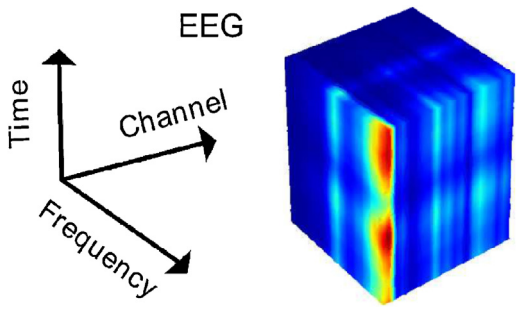

Fig. 1. A demo of the third-order tensor of EEG (Zhao et al., 2011).

of tensor decomposition in contrast to matrix decomposition has been discussed in many previous reports, it is not shown here. Furthermore, this review is targeted to fundamental concepts of tensor decomposition and its application. Thus, the algorithms of tensor decomposition (Cichocki et al., 2015; Kolda and Bader, 2009; Zhou et al., 2014) are not included.

2. Multiple modes and high-order tensors of EEG data

EEG data can be divided into three types including spontaneous EEG (Niedermeyer and Lopes da Silva, 2005), ERPs (Luck, 2005), and ongoing EEG (Cong et al., 2013a). No matter which type of EEG is collected, the raw recordings are continuous and the length of the recording in time can be dozens of minutes or even a few hours. Off-line, the continuous EEG data are usually segmented in terms of stimulus types. Then, in addition to the two modes of time and space, the additional mode segment (also called as epoch or trial) is naturally born for the ERP data. Furthermore, if the EEG data of one segment is transformed into the time-frequency domain, another mode called as frequency is yielded. Moreover, particularly in ERPs, there are often two or more experimental conditions, for example, different types of changes in a sound in the experiment for mismatch negativity (MMN) (Naatanen et al., 2011). This means that one more mode exists and it is named as the mode of condition. For the within-subject analysis, all participants belong to one group, and then, there is the additional mode of subject. Regarding the between-subject analysis in an ERP experiment, there are two or more groups of subjects. Therefore, another mode is the group.

As a result, in an EEG experiment, potentially, there could be even 7 modes including time, frequency, space, trial, condition, subject and group. For the conventional ERP study, the mode of trial disappears after averaging EEG data over single trials is applied to obtain the conventionally defined ERP waveforms. Due to these modes, the high-order tensors including some of the seven modes do naturally exist in the experiment. Obviously, such a tensor may be very big in sizes.

It is not straightforward to visualize the tensors with more than three modes. For demonstration, Fig. 1 shows a third-order tensor including the three modes of time, frequency and space, and it is adapted from the previous study (Zhao et al., 2011). The third-order tensor consists of the time-frequency representation (TFR) of the multiple channels' EEG data. It allows the observation of the temporal, spectral and spatial evolutions of the brain activity simultaneously. This is hard to realize using a two-mode matrix of the TFR data when any of the three modes is concatenated or stacked with any of the left two modes.

EEG data represented by a time-series can be transformed into many other domains besides the time-frequency domain, and after the transformation the EEG data can then be represented by other means instead of the waveform. For example, if the power spectrum is used to analyze the time-series of EEG data, the data are represented by strength of many frequency components in the frequency domain. Strength of a frequency component can be called as a feature or signature. Various types of EEG features have already been used for the computer-aided diagnosis of neurological disorders (Adeli and Ghosh-Dastidar, 2010) and brain–computer interface (BCI) (Tan and Nijholt, 2010). Therefore, there is another mode called as the mode of feature for composing tensors of EEG data after some transforms are applied (Acar et al., 2007a,b).

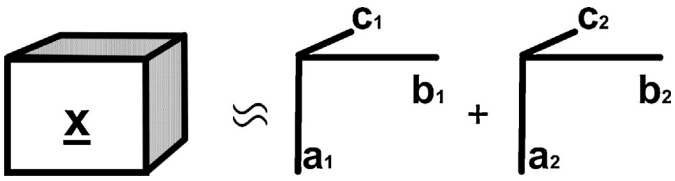

Fig. 2. Two-component CPD of a third-order tensor [Adapted from Bro (1998)].

3. Models of tensor decomposition

Two most fundamental models for tensor decomposition are canonical polyadic (CP) (Hitchcock, 1927) and Tucker (Tucker, 1966). The CP model is also known as parallel factor analysis (PARAFAC) (Harshman, 1970) and canonical decomposition (CANDECOMP) (Carroll and Chang, 1970). Please refer to the Appendix for basic mathematical operations used in tensor decomposition.

3.1. Canonical polyadic decomposition (CPD)

3.1.1. Definition

Given a third-order tensor, a two-component canonical polyadic decomposition (CPD) is shown in Fig. 2 (Bro, 1998).

$$\underline{\mathbf{X}} = \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 + \underline{\mathbf{E}} \approx \mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1 + \mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2 = \underline{\mathbf{X}}_1 + \underline{\mathbf{X}}_2. \quad (1)$$

where the outer product is defined in Section 6.2.

As an example, a third-order tensor is shown in Fig. 1. After the two-component CPD is applied on the tensor, two temporal, two spectral, and two spatial components are extracted, as shown in Fig. 2. In this application, the first temporal component $\mathbf{a}_1$, the first spectral component $\mathbf{b}_1$, and the first spatial component $\mathbf{c}_1$ are associated with one another, and their outer product produces rank-one tensor $\underline{\mathbf{X}}_1$. The second components in the time, frequency, and space modes are associated with one another, and their outer product generates rank-one tensor $\underline{\mathbf{X}}_2$. The sum of rank-one tensors $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ approximates original tensor $\underline{\mathbf{X}}$. Therefore, CPD is the sum of a number of rank-one tensors plus the error tensor.

Generally, for a given Nth-order tensor $\underline{\mathbf{X}} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$, the CPD is defined as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)} + \underline{\mathbf{E}} = \sum_{r=1}^{R} \underline{\mathbf{X}}_r + \underline{\mathbf{E}} = \hat{\underline{\mathbf{X}}} + \underline{\mathbf{E}} \approx \hat{\underline{\mathbf{X}}}, \quad (2)$$

where $\underline{\mathbf{X}}_r = \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)}$, $r = 1, 2, \cdots, R$; $\hat{\underline{\mathbf{X}}}$ approximates tensor $\underline{\mathbf{X}}$, $\underline{\mathbf{E}} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$; and $||\mathbf{u}_r^{(n)}||_2 = 1$, for $n = 1, 2, \cdots, N-1$. $\mathbf{U}^{(n)} = \left[\mathbf{u}_1^{(n)}, \mathbf{u}_2^{(n)}, \cdots, \mathbf{u}_R^{(n)}\right] \in \Re^{I_n \times R}$ denotes a component matrix for mode #n, and $n = 1, 2, \cdots, N$.

In the tensor-matrix product form, Eq. (2) transforms into

$$\underline{\mathbf{X}} = \underline{\mathbf{I}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_N \mathbf{U}^{(N)} + \underline{\mathbf{E}} = \hat{\underline{\mathbf{X}}} + \underline{\mathbf{E}}, \quad (3)$$

where $\underline{\mathbf{I}}$ is an identity tensor, which is a diagonal tensor with a diagonal entry of one.

For tensor decomposition, additional constraints can be applied to some or even all modes according to the properties of each mode.

For example, when the data are nonnegative, the nonnegative constraint can be used (Cichocki et al., 2009). This constraint has been very extensively applied (Cichocki et al., 2015).

3.1.2. Relationship between matrix decomposition and CPD

If $N = 2$, Eq. (2) degenerates into a matrix decomposition as

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E}. \tag{4}$$

Indeed, Eq. (4) is the model for blind source separation (BSS) (Cichocki and Amari, 2003; Comon, 1994; Comon and Jutten, 2010; Hyvarinen et al., 2001). Therefore, we can state that Eq. (4) is a blind separation of a mixture with two modes, and Eq. (2) is a blind separation of a mixture with N modes (Cichocki et al., 2009, 2015). Moreover, CPD and matrix decomposition can be regarded as the sum of rank-one tensor and rank-one matrix, respectively. Furthermore, without additional assumptions constrained, no unique matrix decomposition is available for Eq. (4). This is the main difference between the matrix decomposition and CPD because CPD is unique with only very mild condition required (Kolda and Bader, 2009).

3.2. Tucker decomposition

3.2.1. Definition

For a given Nth-order tensor $\underline{\mathbf{X}} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$, the Tucker decomposition is expressed as follows:

$$\begin{aligned} \underline{\mathbf{X}} &= \sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_N=1}^{R_N} g_{r_1 r_2 \cdots r_N} \mathbf{a}_{r_1}^{(1)} \circ \mathbf{a}_{r_2}^{(2)} \circ \cdots \circ \mathbf{a}_{r_N}^{(N)} + \underline{\mathbf{E}} \\ &= \sum_{r_1=1}^{R_1}\sum_{r_2=1}^{R_2}\cdots\sum_{r_N=1}^{R_N} g_{r_1 r_2 \cdots r_N} \underline{\mathbf{X}}_{r_1 r_2 \cdots r_N} + \underline{\mathbf{E}}, \end{aligned} \tag{5}$$

where rank-one tensor $\underline{\mathbf{X}}_{r_1 r_2 \cdots r_N} = \mathbf{a}_{r_1}^{(1)} \circ \mathbf{a}_{r_2}^{(2)} \circ \cdots \circ \mathbf{a}_{r_N}^{(N)}$, $I_n \geq R_n$, and $g_{r_1 r_2 \cdots r_N}$ composes the core tensor $\underline{\mathbf{G}} \in \Re^{R_1 \times R_2 \times \cdots \times R_N}$. To avoid unnecessary confusion, we denote the component in the Tucker decomposition by **a** instead of **u** in the CPD.

Therefore, the Tucker decomposition is the sum of the $R_1 \times R_2 \times \cdots \times R_N$ scaled rank-one tensors plus the error tensor. Each rank-one tensor is the outer product of N components from N component matrices (each component matrix among the N matrices just contributes one component for the outer product).

In the tensor-matrix form, Eq. (5) is transformed into

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \times_3 \cdots \times_N \mathbf{A}^{(N)} + \underline{\mathbf{E}} = \hat{\underline{\mathbf{X}}} + \underline{\mathbf{E}}, \tag{6}$$

where $\mathbf{A}^{(n)} = \left[\mathbf{a}_1^{(n)}, \mathbf{a}_2^{(n)}, \cdots, \mathbf{a}_{R_n}^{(n)}\right] \in \Re^{I_n \times R_n}$ $(n = 1, 2, \cdots, N)$ denotes the component matrix. To avoid confusion, we denote the component matrix in the Tucker decomposition by **A** instead of **U** in the CPD.

In theory, the Tucker decomposition does not possess unique solutions even though it is subjected to the permutation and variance indeterminacies (Cichocki et al., 2015; Kolda and Bader, 2009). In practice, when additional assumptions are introduced on the different modes, the Tucker decomposition can be unique (Zhou and Cichocki, 2012).

3.2.2. Difference between CPD and Tucker decomposition

In terms of the two fundamental tensor decomposition models, four key differences exist between the CPD and Tucker decomposition.

In the CPD, the number of components in each mode remains invariant. However, in the Tucker decomposition, the numbers of components in the different modes can be different.

Both the CPD and Tucker decomposition are derived in terms of the sum of rank-one tensors, and each rank-one tensor is the outer product of N components from N component matrices with N modes. However, in the CPD, each of the N components must be the component #r of each component matrix. In the Tucker decomposition, each component comes from each component matrix, and no limitation is imposed on the order of the chosen component in the component matrix. For example, in Fig. 2, the first temporal component $\mathbf{a_1}$, first spectral component $\mathbf{b_1}$, and first spatial component $\mathbf{c_1}$ are associated with one another, but any of them is not associated with $\mathbf{a_2}$, $\mathbf{b_2}$, or $\mathbf{c_2}$. In the Tucker decomposition, any component from the different component matrices can be associated with one another.

The core tensor in the CPD is the identity tensor, but that in the Tucker decomposition it can be any tensor with compatible sizes.

In theory, the CPD can be unique under very mild conditions (Kolda and Bader, 2009); however, the Tucker decomposition is generally not unique without imposing additional constraints (Kolda and Bader, 2009; Zhou and Cichocki, 2012).

4. Applications of tensor decomposition of EEG signals

When tensor decomposition is applied to process and analyze EEG signals, the purpose is often source localization of brain activity (Becker et al., 2014), research questions of cognitive neuroscience (Cong et al., 2012b) or clinical neuroscience (Acar et al., 2007a,b), or related to brain–computer interface (Cichocki et al., 2008). In all cases, two issues are essential to concern. The first is whether the CP model or the Tucker model will be applied. The second is whether the tensor is for individual-level analysis or for group-level analysis. Regarding the former analysis, the tensor usually includes the data of one segment (or called as epoch, trial) of one subject in an EEG experiment. As for the latter analysis, the tensor tends to contain the data of multiple trials of one subject or multiple subjects. Therefore, in this review, we categorize the applications of tensor decomposition on EEG based on the above-mentioned two issues. Furthermore, it is necessary to consider one additional issue: how to select the number of components for tensor decomposition. This will be discussed in the end of this section.

Hereinafter, we use 'sample' to denote one subject or one trial/epoch in an EEG experiment.

4.1. Canonical polyadic decomposition of tensor of one sample

As shown in Fig. 1, the TFR of EEG data of multiple channels consists of a third-order tensor. Such a tensor has been decomposed using the CP model (Acar et al., 2007a,b; De Vos et al., 2007b; Deburchgraeve et al., 2009; Miwakeichi et al., 2004; Wang et al., 2012).

Consequently, the spatial component can be used for the source localization (De Vos et al., 2007a,b); the temporal and the spectral components reveal the temporal and spectral characteristics of EEG data simultaneously, and the results can be used for diagnosis of disease (Acar et al., 2007a,b; Wang et al., 2012).

4.2. Canonical polyadic decomposition of tensor of multiple samples

CPD on tensor of one sample can be used for visual inspection. For different samples, if tensor decomposition is applied individually, it is very difficult to straightforwardly make group-level statistical analysis for multiple samples. This is because of the inherent variance indeterminacy of an extracted component by tensor decomposition (Kolda and Bader, 2009). Therefore, it is very significant and important to investigate tensor decomposition on the tensor of multiple samples. The application includes

feature extraction, feature selection and feature analysis. Indeed, these procedures are very extensively applied in the field of pattern recognition and machine learning (Bishop, 2006).

4.2.1. Feature extraction

When a tensor includes the mode of sample (subject or trial), an extracted component in the mode contains a signature of each sample. Usually, the sample mode is arranged as the last mode of the tensor in Eqs. (2) and (5). Given a Nth-order tensor $\underline{\mathbf{X}} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$, I_N is the size of all samples in the sample mode. CPD of the tensor $\underline{\mathbf{X}}$ reads

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N-1)} \circ \mathbf{f}_r + \underline{\mathbf{E}} = \underline{\mathbf{I}} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \cdots \times_{N-1} \mathbf{U}^{(N-1)} \times_N \mathbf{F} + \underline{\mathbf{E}}, \quad (7)$$

where for $n = 1, 2, \cdots, N-1$, $||\mathbf{u}_r^{(n)}||_2 = 1$, and $\mathbf{u}_r^{(n)} (\in \Re^{I_n \times 1})$ denotes the component #r in the mode #n, and it is common to all samples carried by the feature component $\mathbf{f}_r \in \Re^{I_N \times 1}$. Consequently, if the sample mode includes multiple conditions or/and multiple groups of subjects, data analysis can be applied to those signatures carried by the feature component $\mathbf{f}_r$ ($r \in [1, R]$).

The idea shown in Eq. (7) has been applied in many EEG studies (Cong et al., 2012b; de Munck and Bijma, 2009; Lee et al., 2007; Mocks, 1988a,b; Morup et al., 2007; Morup et al., 2006, 2008; Paulick et al., 2014; Schaefer et al., 2013; Vanderperren et al., 2013; Wang et al., 2008, 2012; Weis et al., 2009).

For instance, when a fourth-order ERP tensor of the TFRs includes the time, frequency, space, and sample modes, it can be decomposed into four component matrices of the four modes by CPD with nonnegativity constraints as follows:

$$\underline{\mathbf{X}} = \underline{\mathbf{I}} \times_1 \mathbf{U}^{(t)} \times_2 \mathbf{U}^{(s)} \times_3 \mathbf{U}^{(c)} \times_4 \mathbf{F} + \underline{\mathbf{E}} = \sum_{r=1}^{R} \mathbf{u}_r^{(t)} \circ \mathbf{u}_r^{(s)} \circ \mathbf{u}_r^{(c)} \circ \mathbf{f}_r + \underline{\mathbf{E}}, \quad (8)$$

where $\underline{\mathbf{X}} \in \Re_+^{I_t \times I_s \times I_c \times I}$, $\mathbf{U}^{(t)} = \left[\mathbf{u}_1^{(t)}, \mathbf{u}_2^{(t)}, \cdots, \mathbf{u}_R^{(t)}\right] \in \Re_+^{I_t \times R}$, $\mathbf{U}^{(s)} = \left[\mathbf{u}_1^{(s)}, \mathbf{u}_2^{(s)}, \cdots, \mathbf{u}_R^{(s)}\right] \in \Re_+^{I_s \times R}$, and $\mathbf{U}^{(c)} = \left[\mathbf{u}_1^{(c)}, \mathbf{u}_2^{(c)}, \cdots, \mathbf{u}_R^{(c)}\right] \in \Re_+^{I_c \times R}$, respectively, denote the temporal component matrix, the spectral component matrix, and the channel/spatial component matrix; and $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_R] \in \Re_+^{I \times R}$ represents the multi-domain feature component matrix. It should be noted that all the elements in Eq. (8) are nonnegative.

Fig. 3 shows the demonstration of feature extraction using CPD on a fourth-order tensor including TFR of ERP data in an MMN experiment (Cong et al., 2012b). The first three components of each mode are shown in the figure. The sizes of the tensor were 71(frequency bins) by 60 (temporal points) by 9 (channels) by 42 (2 groups with 21 children in each). The number of extracted components for each mode was 36 (Cong et al., 2012b). Each row of Fig. 3 represents the temporal, spectral and spatial components of a feature, and each feature takes the variance in signatures of 42 children. In other words, the temporal, spectral and spatial components are common to all 42 children, and the variance in signatures of 42 children is revealed given those common components. Furthermore, it is worth emphasizing again that the components and the feature in each row in Fig. 3 are not associated with any other components and other features in other rows. This is the special property of CPD in contrast to Tucker decomposition as discussed in Section 3.2.2 (Kolda and Bader, 2009).

4.2.2. Feature selection and analysis

After a tensor is factorized by CPD, the next is how to select and analyze the extracted features. Usually, the purpose of feature analysis determines the method for feature selection. It is well known that the collected EEG data include the brain activities of interest, brain activities of no interest, and interference, as well as noise. Therefore, the features extracted from EEG data (even when major artefacts are removed in a preprocessing phase) using tensor decomposition can still include features of interest and features of no interest. It is certainly expected that the features of brain activities of interest are used for the further feature analysis, motivating the feature selection step.

Generally speaking, there are two categories for data analysis in terms of two different research questions. One is that the statistical test methods, like analysis of variance (ANOVA), are applied on the selected features to examine the pre-defined hypothesis. The other is that classifiers for pattern recognition (Bishop, 2006) are performed and utilize the previously validated hypothesis. For the former, results can be reported using p values to denote the significance of difference in signatures between/among factors (e.g., conditions, groups) for testing hypothesis (Cong et al., 2012b; Cong et al., 2014a,b; Morup et al., 2007). As for the latter, results are shown using accuracy values in percentage for task or disease prediction, detection, diagnosis and classification (Cong et al., 2010; Lee et al., 2007; Vanderperren et al., 2013). Consequently, the feature selection for the former category is usually based on the properties of EEG signals and the group labels of subjects are not used (Cong et al., 2012b, 2014a,b), and for the latter it is often in light of adaptive feature selection algorithms and the group labels of subjects have to be used (Cong et al., 2010; Lee et al., 2007).

For the former category, all samples' data are decomposed and the feature selection should be relevant in terms of the properties of EEG signals. For instance, Fig. 3 shows three features of 36 ones as mentioned above. Since MMN is a very well-defined component, its properties in the time, frequency and spatial domains are known. The temporal, spectral and spatial components are used to select the feature which is relevant to MMN (Cong et al., 2012b). Therefore, features themselves are not used in the feature selection process. In other words, the feature can be selected when the associated components of the brain signal meet the temporal, spectral and spatial properties of MMN given the example in Fig. 3. Here, the second feature shown in the second row of Fig. 3 was selected as the feature of MMN for the further data analysis (Cong et al., 2012b).

For the latter category, all samples are divided into two parts for the training and testing. The data of the training part are first decomposed and some features are selected in terms of adaptive feature selection algorithms to train the classifier to recognize the label of each trained sample. Therefore, all features are used to select the feature of interest. This is very different from the former category. After the successfully training, the data of new samples (i.e., the testing part) are projected on to the common components of the trained samples to obtain the features of new samples. Using the trained classifier, the labels of new samples are obtained (Cong et al., 2010; Lee et al., 2007; Zhang et al., 2013, 2014). In terms of the successfully training and testing, the classifier would be useful for the future research or application, for example, for diagnosis of brain disorders (Adeli and Ghosh-Dastidar, 2010).

Indeed, the two categories correspond to two different types of research questions. The former is for the why-questions in the field of neuroscience. The latter is for how-questions in the field of engineering and the underlying why-questions in the field of neuroscience have been successfully answered and validated. As a result, the former appears more often in the field of neuroscience and the latter in the field of computer science. To some extent, the

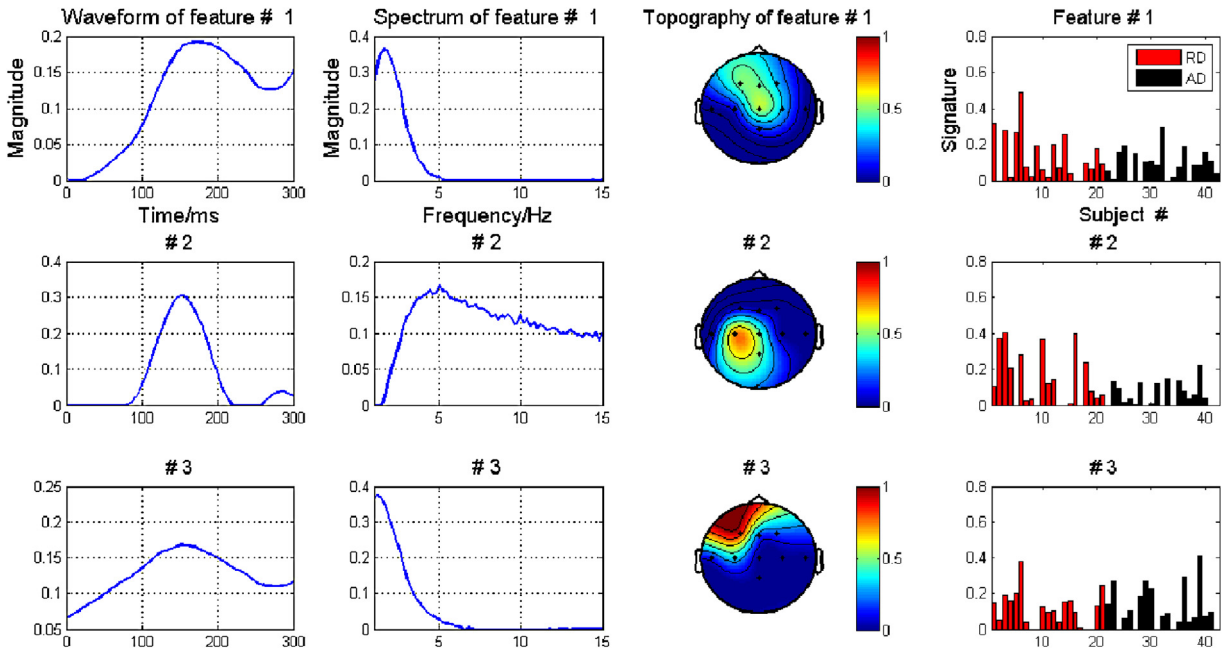

Fig. 3. Example of CPD of a fourth-order ERP tensor including TFR of MMN data with the nonnegativity constraints. The first three components in each mode are shown.

two categories could be combined together, particularly for feature selection (Vanderperren et al., 2013).

4.3. Tucker decomposition of tensor of multiple samples

Among various types of brain imaging data, Tucker decomposition is mostly applied on EEG of multiple samples (Latchoumane et al., 2012; Morup et al., 2007; Phan and Cichocki, 2010, 2011). The application also includes feature extraction, feature selection and feature analysis.

When the sample mode is arranged as the last mode of a Nth-order tensor $\underline{\mathbf{X}} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$ and Tucker decomposition is performed on the tensor in light of Eq. (6), features can be carried by the component matrix of the sample mode (Morup et al., 2007), or by the core tensor (Cong et al., 2013c). The following feature selection and analysis can also be grouped into two categories like those for CPD. As they have been illustrated in the previous subsection, they are not explained in details here.

Fig. 4 shows the example of Tucker decomposition on the fourth-order tensor of MMN (used in Fig. 3) with the nonnegativity constraints. The numbers of temporal components, spectral components, and spatial components are 8, 4, and 6. The features are carried by the core tensor $\underline{\mathbf{G}} \in \Re^{8\times4\times6\times42}$ (Cong et al., 2013c). Therefore, the number of all features is 192. As mentioned above, the second feature extracted by CPD shown in Fig. 3 is the feature of MMN which is of interest for the further analysis. Interestingly, the temporal component #8 in Fig. 4a, the spectral component #3 in Fig. 4b, and the spatial component #4 in Fig. 4c are very similar to the corresponding components shown in the second row of Fig. 3. Then, the feature of MMN by Tucker decomposition is represented by $\underline{\mathbf{G}}(8, 3, 4, :)$ (here, ':' means that all samples are taken). It should be noted that each extracted component in each mode in Fig. 4 is interacted with any other component in any other mode. This is one of the key differences between CPD and Tucker decomposition as discussed in Section 3.2.2 (Kolda and Bader, 2009).

Because of the four differences listed in the Section 3.2.2, the Tucker decomposition is in general more flexible than the CPD (Cichocki et al., 2015; Kolda and Bader, 2009), which is validated by results shown in Figs. 3 and 4. Using CPD 36 components in each mode were extracted, and using Tucker decomposition much fewer components were needed to factorize the tensor for extracting the feature of interest. Moreover, in order to select the feature of interest in terms of prior knowledge of EEG signals, it is necessary to have such information in each mode except the sample mode when Tucker decomposition is used. This is not essential regarding CPD of EEG signals.

4.4. Determination of the number of extracted components for each mode

When tensor decomposition is applied to process and analyze EEG signals, one critical issue is to determine an important parameter, i.e., the number of extracted components for each mode in Eqs. (2) and (5). For the CP model only one parameter is required, and for the Tucker model the number of selected parameters is equal to the number of modes. Methods to determine the parameter include the difference of fit (DIFFIT, the definition will be introduced next) (Timmerman and Kiers, 2000), the cross-validation of component models (Bro and Kiers, 2003; Bro et al., 2008), the Bayesian learning based Automatic relevance determination (ARD) (Morup and Hansen, 2009), and model order selection (He et al., 2010).

It has been generally acknowledged that both tensor decomposition and matrix decomposition can be regarded as solutions of blind source separation. Therefore, in order to determine the number of extracted components by CPD, it is beneficial to consider the selected number of components when the matrix decomposition is applied on EEG signals. Independent component analysis (ICA) (Comon, 1994) is a matrix decomposition method and has been very extensively applied to process and analyze EEG data in the open source toolbox, EEGLAB (Delorme and Makeig, 2004). When ICA is applied on EEG signals, the number of extracted components is regarded as the number of sources in the brain (Makeig et al., 1997, 1999; Vigario and Oja, 2008). No matter in EEG data or in other functional brain imaging data in the time-spatial domain, a source includes a spatial component and a temporal component and the out-product of the two components yield a pattern of the source. Therefore, the number of sources is indeed the number of sources' patterns. The number of estimated sources usually in

functional brain imaging data is often several dozens (Cong et al., 2013b; Delorme and Makeig, 2004; Delorme et al., 2012; Li et al., 2007). Such estimation for ICA can be the appropriate reference to determine the number of extracted components when CPD is applied on EEG signals.

However, the methods in terms of cross-validation of component models and Bayesian learning based ARD often suggested just several components to extract (Acar et al., 2007a,b, 2011; De Vos et al., 2007b; Deburchgraeve et al., 2009; Miwakeichi et al., 2004; Morup et al., 2006, 2008; Nazarpour et al., 2008), which does not very much conform to the physiological properties of EEG signals.

Model order selection methods theoretically match the models of tensor decomposition for determining the number of sources (He et al., 2010). Nevertheless, they do not produce precise results when the signal-to-noise (SNR) ratio is low (e.g., less than 0 dB) (Cong et al., 2012a). In practice, EEG signals, especially EEG data, often possess very low SNR. Therefore, an empirical method in terms of the explained variance calculated by the sum of a certain number of eigenvalues over the sum of all eigenvalues is sometimes applied for matrix decomposition (Arrubla et al., 2014; Breuer et al., 2014). This can also be applied for tensor decomposition.

Surprisingly, DIFFIT for CPD yields similar estimation in matrix decomposition regarding the number of extracted components from EEG data (Cong et al., 2012b, 2013c, 2014b). It is briefly introduced hereinafter. DIFFIT is in light of the change of the fit of a CP model or Tucker model with the increase of the number of extracted components (Timmerman and Kiers, 2000). Here, the example of the CP model is taken. The fit of a CP model reads

$$\mathrm{fit}(m) = 1 - \frac{\left\|\underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}}_m\right\|_F}{\left\|\underline{\mathbf{X}}\right\|_F}, \tag{9}$$

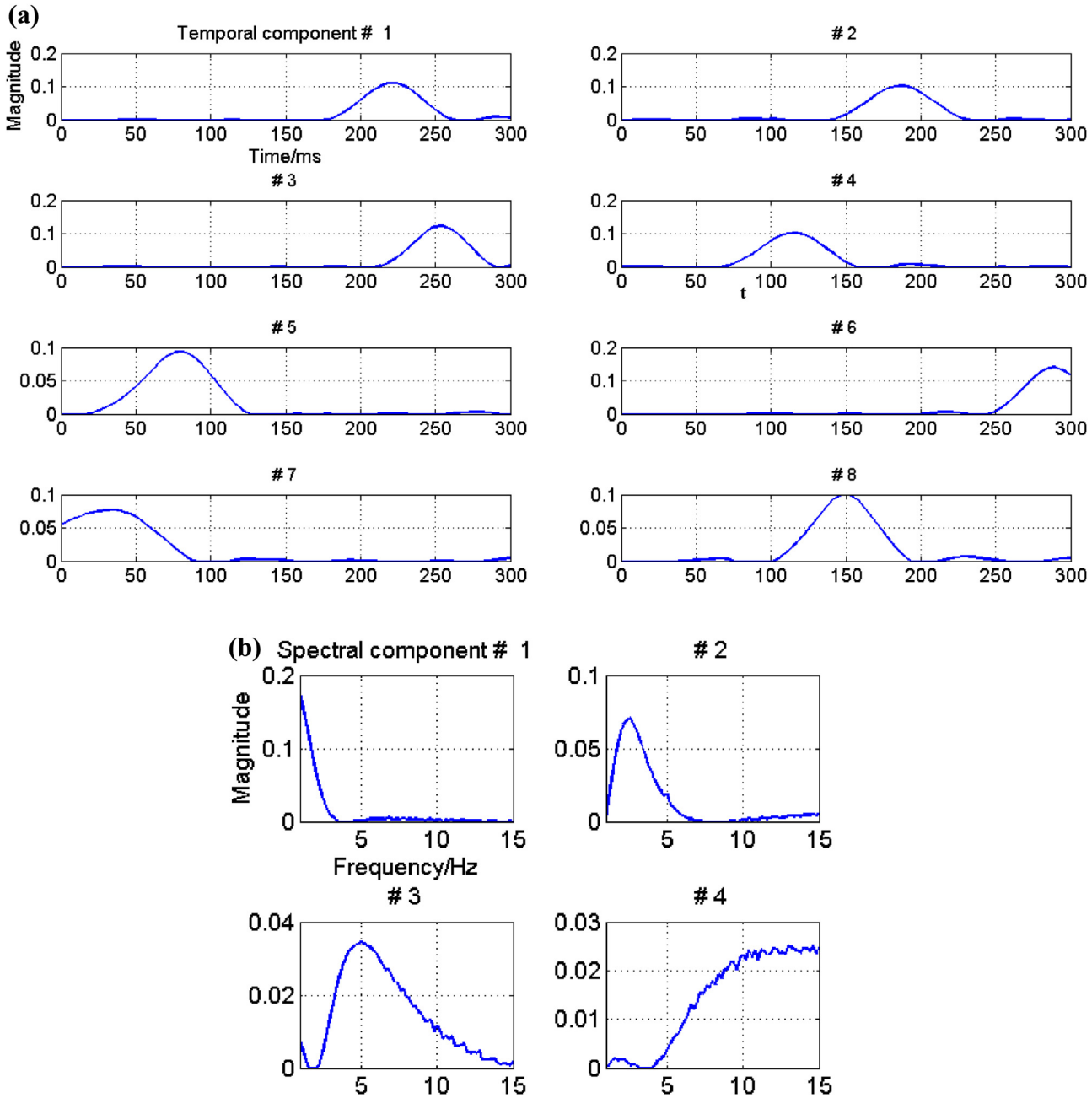

Fig. 4. Example of Tucker decomposition of the fourth-order ERP tensor (used in Fig. 3) with the nonnegativity constraints: (a) eight temporal components, (b) four spectral components, (c) six spatial components, and (d) all extracted features.

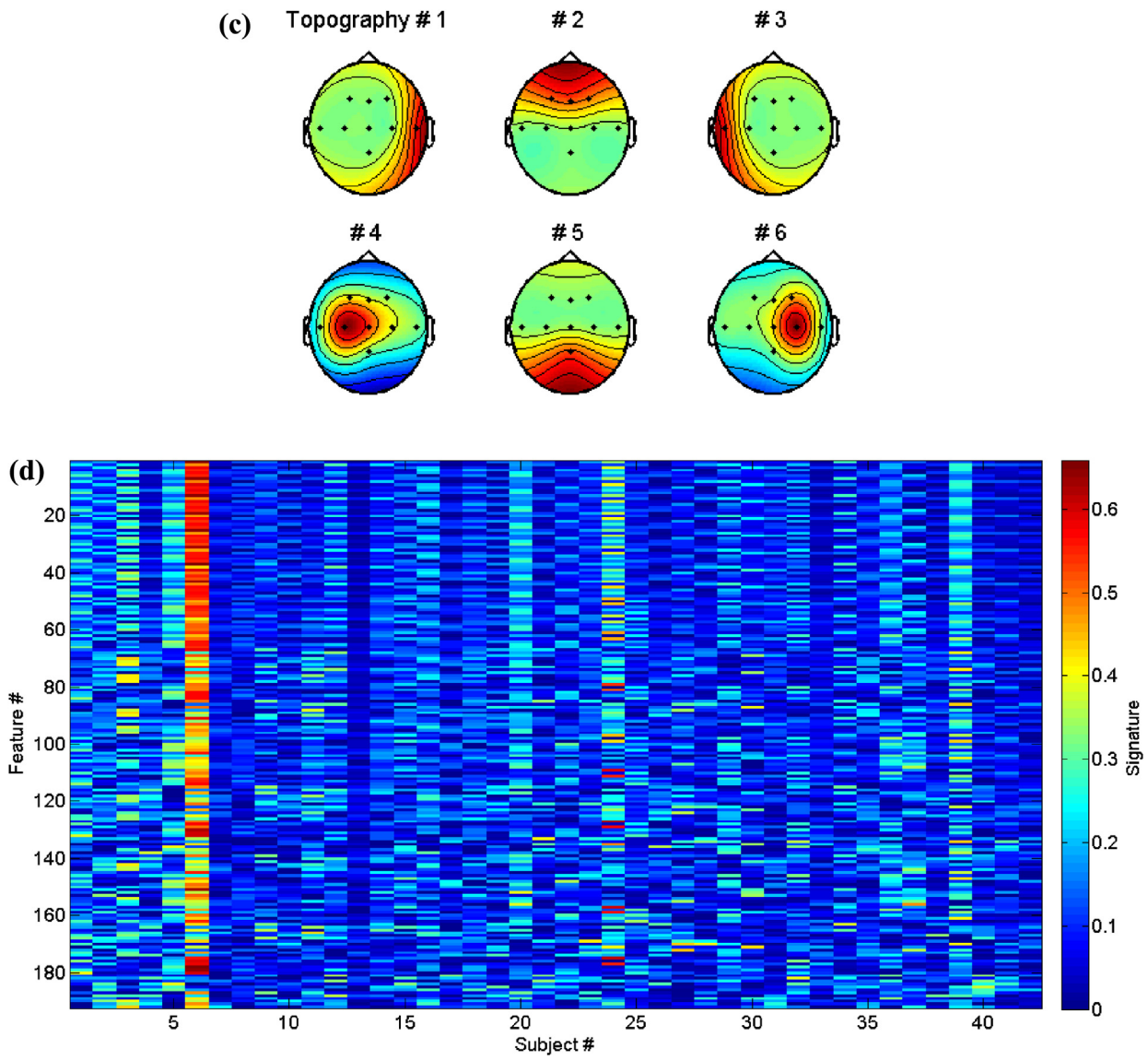

Fig. 4. (*Continued*).

where $\widehat{\underline{\mathbf{X}}}_{\mathbf{m}}$ is the approximation for the raw data tensor $\underline{\mathbf{X}}$ by a rank-m CP model (m plays the same role as R in Eq. (2)), $\|\bullet\|_F$ is the Frobenius norm, $m = 1, \ldots, M$, and fit(m) is monotonically rising. Then, the difference fit of two adjacent fits is given by

$$\text{dif}(m) = \text{fit}(m) - \text{fit}(m-1) \tag{10}$$

where $m = 2, \ldots, M$. Next, the ratio of the adjacent difference fits is defined as

$$\text{DIFFIT}(m) = \frac{\text{dif}(m)}{\text{dif}(m+1)}, \tag{11}$$

where $m = 2, \ldots, M-1$. The model with the largest DIFFIT value is regarded as the appropriate CP model for the raw data tensor.

Fig. 5 shows one example of DIFFIT for CPD and it is adapted from the previous study (Cong et al., 2012b). The m in Eq. (11) ranges from 2 to 50. DIFFIT suggests the appropriate number of extracted components is 36.

The drawback of DIFFIT is that it is very time consuming. In case the study using tensor decomposition is offline, DIFFIT is an appropriate option.

5. Potential trend: Tensor decomposition for data of two modalities simultaneously

Recently, research of the simultaneous EEG and functional magnetic resonance imaging (fMRI) has becomes very attractive in order to overcome the inherent shortcoming of EEG (low spatial resolution) and fMRI (low temporal resolution). Therefore, how to process the data of the two modalities becomes a very significant research question (Ullsperger and Debener, 2010). Currently, most of the data processing methods for the data of the two modalities are in terms of matrix decomposition (Sui et al., 2012). Nevertheless, when EEG data and fMRI data are represented by two tensors, the tensor decomposition methods for the data of two modalities can be applied (Martinez-Montes et al., 2004).

Indeed, as long as data of one modality is denoted by a tensor (data of the other mode can be represented by a vector, a matrix or a tensor), the data of two modalities can be factorized by the tensor decomposition methods (Acar et al., 2007a,b; Eliseyev and Aksenova, 2013; Eliseyev et al., 2012; Zhao et al., 2013a,b). The tensor decomposition methods for the data of two modalities are called N-way partial least squares (PLS) (Andersson and Bro, 2000; Bro, 1996) and higher-order partial least squares (HOPLS) (Zhao et al.,

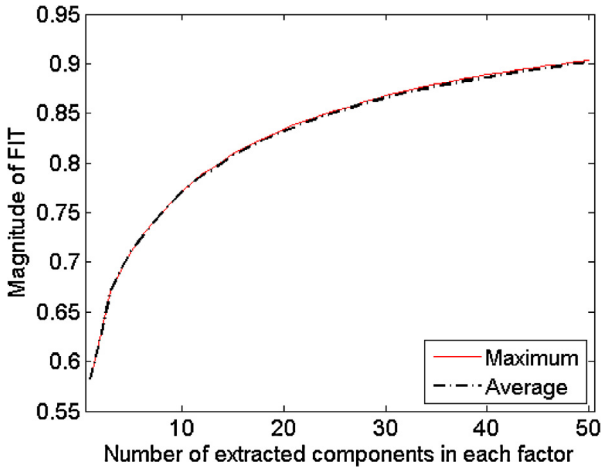

Fig. 5. One example of DIFFIT for CPD and it is adapted from the previous study (Cong et al., 2012b). The m in Eq. (11) of this study ranges from 2 to 50 in this case. For each m, the CPD with the nonnegativity constraints were run 100 times for each on the fourth-order ERP tensor mentioned in Section 4.2.1. The solid curve represents the maximal fit among 100 runs for each model, and the dash dot curve denotes the averaged fit over 100 runs for each model. DIFFIT is performed on the averaged curve.

2013a). The solutions of N-way PLS and HOPLS have been thoroughly discussed previously (Andersson and Bro, 2000; Zhao et al., 2013a). For simplicity of this study, they are not introduced here. The theoretical models of PLS, N-way PLS and HOPLS are described hereinafter.

5.1. Standard partial least squares

For brain research, the standard PLS seeks the common latent vectors from brain data $\mathbf{X} \in \Re^{I\times L}$ and behavior data $\mathbf{Y} \in \Re^{I\times N}$ and the constraint is that those latent vectors mostly explain the covariance between the brain and the behavior data (Krishnan et al., 2011). The standard PLS reads

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}_\mathbf{X}, \tag{12}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{C}^T + \mathbf{E}_\mathbf{Y}, \tag{13}$$

where $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_R] \in \Re^{I\times R}$ $\left(\mathbf{T}^T\mathbf{T} = \mathbf{I}, \mathbf{I}\right.$ is the identity matrix$\left.\right)$ includes extracted orthonormal latent vectors from brain data $\mathbf{X}$; $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_R] \in \Re^{I\times R}$ consists of latent vectors from behavior data $\mathbf{Y}$; $\mathbf{P}$ and $\mathbf{C}$ correspond to loadings; and $\mathbf{E}_\mathbf{X}$ and $\mathbf{E}_\mathbf{Y}$ represent the residuals from brain data $\mathbf{X}$ and behavior data $\mathbf{Y}$, respectively. Under the framework of PLS, and $\mathbf{U}$ column-wisely has the maximum covariance with $\mathbf{T}$, and the simplest model to mostly represent $\mathbf{Y}$ is given by

$$\mathbf{U} \approx \mathbf{T}\mathbf{D}, \tag{14}$$

where $\mathbf{D}$ is a diagonal matrix and its elements are $d_{\mathrm{rr}} = \mathbf{u}_r^T\mathbf{t}_r / \mathbf{t}_r^T\mathbf{t}_r$ $(r = 1, \cdots, R)$, the regression coefficients to predict $\mathbf{Y}$.

5.2. N-way partial least squares

When the brain data are represented by a tensor $\underline{\mathbf{X}}$ instead of a matrix $\mathbf{X}$, the N-way PLS may substitute the standard PLS to decompose the brain data $\underline{\mathbf{X}}$ and the behavior data $\mathbf{Y}$ together (Andersson and Bro, 2000; Bro, 1996). For example, when the brain data is represented by a four-way tensor $\underline{\mathbf{X}}$ conforming to the four modes of subject by time by space by type of stimulus, i.e., $\underline{\mathbf{X}} \in \Re^{I\times L\times K\times J}$. Using N-way PLS, the CP model is exploited to decompose the tensor $\underline{\mathbf{X}}$ into R components in each mode (Andersson and Bro, 2000; Bro, 1996). Hence the N-way PLS reads as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{t}_r \circ \mathbf{p}_r \circ \mathbf{q}_r \circ \mathbf{s}_r + \underline{\mathbf{E}}_{\underline{\mathbf{X}}}, \tag{15}$$

$$\mathbf{Y} = \sum_{r=1}^{R} d_{\mathrm{rr}}\mathbf{t}_r\mathbf{c}_r^T + \mathbf{E}_\mathbf{Y}, \tag{16}$$

where $\mathbf{t}_r \in \Re^{I\times 1}$ is a latent vector whose elements are comparable for all I subjects, and $\mathbf{p}_r \in \Re^{L\times 1}$ $\mathbf{q}_r \in \Re^{K\times 1}$ and $\mathbf{s}_r \in \Re^{J\times 1}$ are loading vectors which are common for all subjects. They sometimes also are named as factors. Under N-way PLS, $\forall r$, $r = 1, \cdots, R$, the vectors $\mathbf{p}$, $\mathbf{q}$, $\mathbf{s}$, and $\mathbf{c}$ satisfy

$$\{\mathbf{p}, \mathbf{q}, \mathbf{s}, \mathbf{c}\} = arg \underbrace{\max}_{\mathbf{p},\mathbf{q},\mathbf{s},\mathbf{c}} [cov(\mathbf{t}, \mathbf{u})], \tag{17}$$

$$\text{s.t.} \quad t_i = \sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{j=1}^{J} x_{ilkj}p_l q_k s_j, = \mathbf{Yc}$$
$$\text{and} \quad \|\mathbf{p}\|_2^2 = \|\mathbf{q}\|_2^2 = \|\mathbf{s}\|_2^2 = 1.$$

5.3. Higher-order partial least squares

Regarding the tensor $\underline{\mathbf{X}}$, the Tucker model can be applied under HOPLS (Zhao et al., 2013a) which is expressed as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \underline{\mathbf{G}}_r \times_1 \mathbf{t}_r \times_2 \mathbf{P}_r \times_3 \mathbf{Q}_r \times_4 \mathbf{S}_r + \underline{\mathbf{E}}_{\underline{\mathbf{X}},R}, \tag{18}$$

$$\mathbf{Y} = \sum_{r=1}^{R} d_{\mathrm{rr}}\mathbf{t}_r\mathbf{c}_r^T + \mathbf{E}_{\mathbf{Y},R}, \tag{19}$$

where, R is the number of latent vectors; $\mathbf{t}_r \in \Re^{I\times 1}$ is the rth latent vector; $\mathbf{P}_r \in \Re^{L\times R_2}$ $(L \geq R_2)$, $\mathbf{Q}_r \in \Re^{K\times R_3}$ $(K \geq R_3)$, and $\mathrm{S}_r \in \Re^{J\times R_2}$ $(J \geq R_4)$ are the corresponding loading matrices which are common for all I subjects; $\underline{\mathbf{G}}_r \in \Re^{1\times R_2\times R_3\times R_4}$ is the rth core tensor and reveals the interaction between the rth latent vector and the corresponding loading matrices.

Although only few studies using tensor decomposition to process and analyze the simultaneous EEG and fMRI data (Martinez-Montes et al., 2004), theoretically, N-way PLS and HOPLS are very promising tools for exploiting the data of two modalities in brain research. In practice, when the two methods are applied on the data of a large group of subjects, a super-computer is required for computing. In case of the data processing for one subject, the normal workstation is suitable.

Acknowledgements

Cong thanks the financial support from TEKES in Finland (MUSCLES projects, 40334/10), the Fundamental Research Funds for the Central Universities [DUT14RC(3)037] and for XingHai Scholar in Dalian University of Technology in China, and National Natural Science Foundation of China (grant nos. 81471742 & 61379012). All authors appreciate the anonymous reviewers for the invaluable comments to improve the study. Cong F. thanks Prof. Andrzej Cichocki (RIKEN, Japan) very much for assistance of studying tensor decomposition.

Appendix A.

To avoid complicated mathematics in the tensor decomposition algorithms, we only introduce the basic products of tensor decomposition (Cong et al., 2012b).

A.1. Inner and outer products

The inner and outer products are very fundamental in tensor decomposition. Given two column vectors $\mathbf{a} = [a_1, a_2, a_3]^T$ and $\mathbf{b} = [b_1, b_2, b_3]^T$, their inner product is

$$x = \mathbf{a}^T \times \mathbf{b} = a_1 \times b_1 + a_2 \times b_2 + a_3 \times b_3,$$

and their outer product is

$$\mathbf{X} = \mathbf{a} \circ \mathbf{b} = \begin{bmatrix} a_1 \times b_1 & a_1 \times b_2 & a_1 \times b_3 \\ a_2 \times b_1 & a_2 \times b_2 & a_2 \times b_3 \\ a_3 \times b_1 & a_3 \times b_2 & a_3 \times b_3 \end{bmatrix} = \mathbf{a} \times \mathbf{b}^T, \quad x_{ij} = a_i \times b_j,$$

where the symbol "$\circ$" denotes the outer product of the vectors.

A.2. Outer product of multiple vectors

Given three vectors $\mathbf{a} \in \Re^{I \times 1}$, $\mathbf{b} \in \Re^{J \times 1}$, and $\mathbf{c} \in \Re^{K \times 1}$, their outer product yields a third-order rank-one tensor

$$\underline{\mathbf{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \in \Re^{I \times J \times K},$$

where $x_{ijk} = a_i b_j c_k$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $k = 1, \ldots, K$. This tensor is called a rank-one tensor.

A.3. Mode-n tensor matrix product

The mode-n product $\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_n \mathbf{A}$ of tensor $\underline{\mathbf{G}} \in \underline{\Re}^{J_1 \times J_2 \times \cdots \times J_N}$ and matrix $\underline{\mathbf{A}} \in \mathbb{R}^{In \times J_n}$ is tensor $\underline{\mathbf{X}} \in \underline{\Re}^{J_1 \times J_2 \times \cdots \times J_{n-1} \times I_n \times J_{n+1} \times \cdots \times J_N}$ with the following elements:

$$x_{j_1 j_2 \ldots j_{n-1} i_n j_{n+1} \ldots j_N} = \sum_{j_n=1}^{J_n} g_{j_1 j_2 \ldots j_N} a_{i_n, j_n}.$$

References

Acar E, Aykut-Bingol C, Bingol H, Bro R, Yener B. Multiway analysis of epilepsy tensors. Bioinformatics (Oxford, England) 2007a;23(13):i10-i18, http://dx.doi.org/10.1093/bioinformatics/btm210.

Acar E, Bingol CA, Bingol H, Bro R, Yener B. Seizure recognition on epilepsy feature tensor. In: Conference proceedings: . . .Annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. Conference, 2007; 2007b. p. 4273–6, http://dx.doi.org/10.1109/IEMBS.2007.4353280.

Acar E, Dunlavy DM, Kolda T, Morup M. Scalable tensor factorizations for incomplete data. Chemom Intell Lab Syst 2011;106(1):41–56.

Acar E, Yener B. Unsupervised multiway data analysis: a literature survey. IEEE Trans Knowl Data Eng 2009;21(1):6–20.

Adeli H, Ghosh-Dastidar S. Automated EEG-based diagnosis of neurological disorders—inventing the future of neurology. Florida, USA: CRC Press; 2010.

Andersson CA, Bro R. The N-way toolbox for MATLAB. Chemom Intell Lab Syst 2000;52:1–4.

Arrubla J, Neuner I, Dammers J, Breuer L, Warbrick T, Hahn D, et al. Methods for pulse artefact reduction: experiences with EEG data recorded at 9.4 T static magnetic field. J Neurosci Methods 2014;232:110–7.

Becker H, Albera L, Comon P, Haardt M, Birot G, Wendling F, et al. EEG extended source localization: tensor-based vs. conventional methods. NeuroImage 2014;96:143–57.

Berger H. Ueber das Elektrenkephalogramm des Menschen. Archives fur Psychiatrie Nervenkrankheiten 1929;87:527–70.

Bishop CM. Pattern recognition and machine learning, vol. 1st. Singapore: Springer; 2006.

Breuer L, Dammers J, Roberts TPL, Jon Shah N. Ocular and cardiac artifact rejection for real-time analysis in MEG. J Neurosci Methods 2014;233:105–14.

Bro R. Multiway calibration. Multilinear PLS. J Chemometr 1996;10(1):47–61.

Bro R. Multi-way analysis in the food industry—models, algorithms, and applications. Holland: University of Amsterdam; 1998 (PhD thesis).

Bro R, Kiers HAL. A new efficient method for determining the number of components in PARAFAC models. J Chemom 2003;17:274–86.

Bro R, Kjeldahl K, Smilde AK, Kiers HAL. Cross-validation of component models: a critical look at current methods. Anal Bioanal Chem 2008;390:1241–51.

Calhoun VD, Adali T. Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Rev Biomed Eng 2012;5:60–73, http://dx.doi.org/10.1109/RBME.2012.2211076; 10.1109/RBME.2012.2211076.

Carroll JD, Chang J. Analysis of individual differences in ultidimensional scaling via an n-way generalization of 'Eckartâ Young' decomposition. Psychometrika 1970;35:283–319.

Cichocki A, Amari S. Adaptive blind signal and image processing: learning algorithms and applications, vol. revised. Chichester: John Wile & Sons Inc; 2003.

Cichocki A, Mandic D, Caiafa C, Phan A-H, Zhou G, Zhao Q, et al. Tensor decompositions for signal processing applications from two-way to multiway component analysis. IEEE Signal Process Mag 2015;32(2):145–63.

Cichocki A, Washizawa Y, Rutkowski TM, Bakardjian H, Phan AH, Choi S, et al. Noninvasive BCIs: multiway signal-processing array decompositions. Computer 2008;41(10):34–42.

Cichocki A, Zdunek R, Phan AH, Amari S. Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis. John Wiley; 2009.

Cohen MX. Analyzing neural time series data: theory and practice. MIT Press; 2014.

Comon P. Independent component analysis, a new concept? Signal Process 1994;36(3):287–314.

Comon P. Tensors: a brief introduction. IEEE Signal Process Mag 2014;31(3):44–53.

Comon P, Jutten C. Handbook of blind source separation: independent component analysis and applications, vol. 1st. Academic Press; 2010.

Comon P, Luciani X, De Almeida ALF. Tensor decompositions, alternating least squares and other tales. J Chemom 2009;23:393–405.

Cong F, Alluri V, Nandi AK, Toiviainen P, Fa R, Abu-Jamous B, et al. Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. IEEE Trans Multimedia 2013a;15(5):1060–9.

Cong F, He Z, Hamalainen J, Leppanen PHT, Lyytinen H, Cichocki A, et al. Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection. J Neurosci Methods 2013b;212(1):165–72.

Cong F, Nandi AK, He Z, Cichocki A, Ristaniemi T. Fast and effective model order selection method to determine the number of sources in a linear transformation model. In: Proc. The 2012 European Signal Processing Conference (EUSIPCO-2012); 2012a. p. 1870–4.

Cong F, Phan AH, Astikainen P, Zhao Q, Wu Q, Hietanen JK, et al. Multi-domain feature extraction for small event-related potentials through nonnegative multi-way array decomposition from low dense array EEG. Int J Neural Syst 2013c;23(2)., http://dx.doi.org/10.1142/S0129065713500068, 1350006, 18 pages.

Cong F, Phan AH, Lyytinen H, Ristaniemi T, Cichocki A. Classifying healthy children and children with attention deficit through features derived from sparse and nonnegative tensor factorization using event-related potential. Lect Notes Comput Sci 2010;6365:620–8 (In V. Vigneron et al. (Eds.): LVA/ICA 2010).

Cong F, Phan AH, Zhao Q, Huttunen-Scott T, Kaartinen J, Ristaniemi T, et al. Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from EEG collected by low-density array. Int J Neural Syst 2012b;22(6)., http://dx.doi.org/10.1142/S0129065712500256, 1250025, 19 pages.

Cong F, Puolivali T, Alluri V, Sipola T, Burunat I, Toiviainen P, et al. Key issues in decomposing fMRI during naturalistic and continuous music experience with independent component analysis. J Neurosci Methods 2014a;223:74–84.

Cong F, Zhou G, Astikainen P, Zhao Q, Wu Q, Nandi AK, et al. Low-rank approximation based nonnegative multi-way array decomposition on event-related potentials. Int J Neural Syst 2014b;24(8)., http://dx.doi.org/10.1142/S012906571440005X, 1440005, 19 pages.

de Munck JC, Bijma F. Three-way matrix analysis, the MUSIC algorithm and the coupled dipole model. J Neurosci Methods 2009;183:63–71.

De Vos M, De Lathauwer L, Vanrumste B, Van Huffel S, Van Paesschen W. Canonical decomposition of ictal scalp EEG and accurate source localisation: principles and simulation study. Comput Intell Neurosci 2007a:58253, http://dx.doi.org/10.1155/2007/58253.

De Vos M, Vergult A, De Lathauwer L, De Clercq W, Van Huffel S, Dupont P, et al. Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone. NeuroImage 2007b;37(3):844–54, http://dx.doi.org/10.1016/j.neuroimage.2007.04.041.

Deburchgraeve W, Cherian P, De Vos M, Swarte R, Blok J, Visser G, et al. Neonatal seizure localization using PARAFAC decomposition. Clin Neurophysiol 2009;120:1787–96 (official journal of the International Federation of Clinical Neurophysiology).

Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9–21, http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Delorme A, Palmer J, Onton J, Oostenveld R, Makeig S. Independent EEG sources are dipolar. PLoS ONE 2012;7(2):e30135, http://dx.doi.org/10.1371/journal.pone.0030135.

Dien J. Applying principal components analysis to event-related potentials: a tutorial. Dev Neuropsychol 2012;37(6):497–517.

Eichele T, Rachakonda S, Brakedal B, Eikeland R, Calhoun VD. EEGIFT: group independent component analysis for event-related EEG data. Comput Intell Neurosci 2011;2011:129365, http://dx.doi.org/10.1155/2011/129365; 10.1155/2011/129365.

Eliseyev A, Aksenova T. Recursive N-way partial least squares for brain–computer interface. PLoS ONE 2013;8(7):e69962, http://dx.doi.org/10.1371/journal.pone.0069962.

Eliseyev A, Moro C, Faber J, Wyss A, Torres N, Mestais C, et al. L1-penalized N-way PLS for subset of electrodes selection in BCI experiments. J Neural Eng 2012;9(4.):045010.

Harshman RA. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. In: UCLA working papers in phonetics; 1970. p. 1–84.

He Z, Cichocki A, Xie S, Choi K. Detecting the number of clusters in n-way probabilistic clustering. IEEE Trans Pattern Anal Mach Intell 2010;32(11):2006–21, http://dx.doi.org/10.1109/TPAMI.2010.15.

Herrmann CS, Rach S, Vosskuhl J, Struber D. Time-frequency analysis of event-related potentials: a brief tutorial. Brain Topogr 2013:1–13, http://dx.doi.org/10.1007/s10548-013-0327-5.

Hitchcock FL. The expression of a tensor or a polyadic as a sum of products. J Math Phys 1927;6(1):164–89.

Hyvarinen A, Karhunen J, Oja E. Independent component analysis. New York, NY: John Wile & Sons Inc; 2001.

Khoromskij B. Tensors-structured numerical methods in scientific computing: survey on recent advances. Chemom Intell Lab Syst 2011;110(1):1–19.

Kolda T, Bader B. Tensor decompositions and applications. SIAM Rev 2009;51(3):455–500.

Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. NeuroImage 2011;56(2):455–75, http://dx.doi.org/10.1016/j.neuroimage.2010.07.034.

Kroonenberg PM. Applied multiway data analysis. Wiley; 2008.

Latchoumane CF, Vialatte FB, Sole-Casals J, Maurice M, Wimalaratna SR, Hudson N, et al. Multiway array decomposition analysis of EEGs in Alzheimer's disease. J Neurosci Methods 2012;207(1):41–50, http://dx.doi.org/10.1016/j.jneumeth.2012.03.005.

Lee H, Kim YD, Cichocki A, Choi S. Nonnegative tensor factorization for continuous EEG classification. Int J Neural Syst 2007;17(4):305–17.

Li YO, Adali T, Calhoun VD. Estimating the number of independent components for functional magnetic resonance imaging data. Human Brain Mapp 2007;28(11):1251–66, http://dx.doi.org/10.1002/hbm.20359.

Lu H, Plataniotis K, Venetsanopoulos A. A survey of multilinear subspace learning for tensor data. Pattern Recognit 2011;44(7):1540–51.

Luck SJ. An introduction to the event-related potential technique. The MIT Press; 2005.

Makeig S, Jung TP, Bell AJ, Ghahremani D, Sejnowski TJ. Blind separation of auditory event-related brain responses into independent components. Proc Natl Acad Sci USA 1997;94(20):10979–84.

Makeig S, Westerfield M, Jung TP, Covington J, Townsend J, Sejnowski TJ, et al. Functionally independent components of the late positive event-related potential during visual spatial attention. J Neurosci 1999;19(7):2665–80 (the official journal of the Society for Neuroscience).

Martinez-Montes E, Valdes-Sosa PA, Miwakeichi F, Goldman RI, Cohen MS. Concurrent EEG/fMRI analysis by multiway Partial Least Squares. NeuroImage 2004;22(3):1023–34, http://dx.doi.org/10.1016/j.neuroimage.2004.03.038.

Miwakeichi F, Martinez-Montes E, Valdes-Sosa PA, Nishiyama N, Mizuhara H, Yamaguchi Y. Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis. NeuroImage 2004;22(3):1035–45, http://dx.doi.org/10.1016/j.neuroimage.2004.03.039.

Mocks J. Topographic components model for event-related potentials and some biophysical considerations. IEEE Trans Biomed Eng 1988b;35:482–4.

Mocks J. Decomposing event-related potentials: a new topographic components model. Biol Psychol 1988a;26:199–215.

Morup M. Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdiscip Rev: Data Min Knowl Discovery 2011;1(1):24–40.

Morup M, Hansen LK. Automatic relevance determination for multiway models. J Chemom 2009;23:352–63.

Morup M, Hansen LK, Arnfred SM. ERPWAVELAB a toolbox for multi-channel analysis of time-frequency transformed event related potentials. J Neurosci Methods 2007;161(2):361–8, http://dx.doi.org/10.1016/j.jneumeth.2006.11.008.

Morup M, Hansen LK, Arnfred SM, Lim LH, Madsen KH. Shift-invariant multilinear decomposition of neuroimaging data. NeuroImage 2008;42:1439–50.

Morup M, Hansen LK, Herrmann CS, Parnas J, Arnfred SM. Parallel Factor Analysis as an exploratory tool for wavelet transformed event-related EEG. NeuroImage 2006;29(3):938–47, http://dx.doi.org/10.1016/j.neuroimage.2005.08.005.

Naatanen R, Kujala T, Kreegipuu K, Carlson S, Escera C, Baldeweg T, et al., Ponton C. The mismatch negativity: an index of cognitive decline in neuropsychiatric and neurological diseases and in ageing. Brain 2011;134(Pt 12):3432–50, http://dx.doi.org/10.1093/brain/awr064 (a journal of neurology).

Nazarpour K, Wongsawat Y, Sanei S, Chambers JA, Oraintara S. Removal of the eye-blink artifacts from EEGs via STF-TS modeling and robust minimum variance beamforming. IEEE Trans Biomed Eng 2008;55(9):2221–31.

Niedermeyer E, Lopes da Silva F. Electroencephalography: basic principles, clinical applications, and related fields. Baltimore, MD: Williams & Wilkins; 2005.

Onton J, Makeig S. Information-based modeling of event-related brain dynamics. Prog Brain Res 2006;159:99–120, http://dx.doi.org/10.1016/S0079-6123(06)59007-7.

Paulick C, Wright MN, Verleger R, Keller K. Decomposition of 3-way arrays: a comparison of different PARAFAC algorithms. Chemom Intell Lab Syst 2014;137:97–109.

Phan AH, Cichocki A. Tensor decomposition for feature extraction and classification problem. IEICE Trans Fundam Electron Commun Comput Sci 2010;1(1):37–68.

Phan AH, Cichocki A. Extended HALS algorithm for nonnegative Tucker decomposition and its applications for multiway analysis and classification. Neurocomputing 2011;74(11):1956–69.

Schaefer RS, Desain P, Farquhar J. Shared processing of perception and imagery of music in decomposed EEG. NeuroImage 2013;70:317–26.

Smilde A, Bro R, Geladi P. Multi-way analysis with applications in the chemical sciences. Wiley; 2004.

Sui J, Adali T, Yu Q, Chen J, Calhoun VD. A review of multivariate methods for multimodal fusion of brain imaging data. J Neurosci Methods 2012;204(1):68–81, http://dx.doi.org/10.1016/j.jneumeth.2011.10.031; 10.1016/j.jneumeth.2011.10.031.

Tan DS, Nijholt A. Brain–computer interfaces: applying our minds to human–computer interaction. London: Springer; 2010.

Timmerman ME, Kiers HA. Three-mode principal components analysis: choosing the numbers of components and sensitivity to local optima. Br J Math Stat Psychol 2000;53(Pt 1):1–16.

Tucker LR. Some mathematical notes on three-mode factor analysis. Psychometrika 1966;31(3):279–311.

Ullsperger M, Debener S. Simultaneous EEG and fMRI: recording, analysis, and application, vol. 1st. Oxford University Press; 2010.

Vanderperren K, Mijovic B, Novitskiy N, Vanrumste B, Stiers P, Van den Bergh BR, et al. Single trial ERP reading based on parallel factor analysis. Psychophysiology 2013;50(1):97–110, http://dx.doi.org/10.1111/j.1469-8986.2012.01405.x; 10.1111/j.1469-8986.2012.01405.x.

Wang J, Li X, Lu C, Voss LJ, Barnard JPM, Sleigh JW. Characteristics of evoked potentialmultiple EEG recordings in patients with chronic pain bymeans of parallel factor analysis. Comput Intell Neurosci 2012;2012(279560):1–10.

Wang Z, Maier A, Logothetis NK, Liang H. Single-trial decoding of bistable perception based on sparse nonnegative tensor decomposition. Comput Intell Neurosci 2008;2008(642387):1–10.

Weis M, Romer F, Haardt M, Jannek D, Husar P. Multi-dimensional space-time-frequency component analysis of event related EEG data using closed-form PARAFAC. Proc IEEE Int Conf Acoust, Speech Signal Process 2009;2009:349–52.

Vigario R, Oja E. BSS and ICA in neuroinformatics: from current practices to open challenges. IEEE Rev Biomed Eng 2008;1:50–61.

Zhang Y, Zhou G, Jin J, Wang M, Wang X, Cichocki A. L1-regularized multiway canonical correlation analysis for SSVEP-based BCI. IEEE Trans Neural Syst Rehabil Eng 2013;21(6):887–96.

Zhang Y, Zhou G, Zhao Q, Onishi A, Jin J, Wang X, et al. Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIs. Lect Notes Comput Sci 2014;7062:287–95.

Zhao Q, Caiafa CF, Mandic DP, Chao ZC, Nagasaka Y, Fujii N, et al. Higher-order partial least squares (HOPLS): a generalized multi-linear regression method. IEEE Trans Pattern Anal Mach Intell 2013a;35(7):1660–73.

Zhao Q, Caiafa CF, Mandic DP, Zhang L, Ball T, Schulze-Bonhage A, et al. Multilinear subspace regression: an orthogonal tensor decomposition approach. Adv Neural Inf Process Syst 2011;24:1269–77.

Zhao Q, Zhou G, Adali T, Zhang L, Cichocki A. Kernelization of tensor-based models for multiway data analysis. IEEE Signal Process Mag 2013b;30(4):137–48.

Zhou G, Cichocki A. Fast and unique Tucker decompositions via multiway blind source separation. Bull Pol Acad Sci—Tech Sci 2012;60(3):389–405.

Zhou G, Cichocki A, Xie S. Nonnegative matrix and tensor factorizations: an algorithmic perspective. IEEE Signal Process Mag 2014;31(3):54–65.