JYU DISSERTATIONS 321

Xiulin Wang

Coupled Nonnegative Matrix/Tensor Factorization in Brain Imaging Data

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION TECHNOLOGY

JYU DISSERTATIONS 321

Xiulin Wang

Coupled Nonnegative Matrix/Tensor Factorization in Brain Imaging Data

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi joulukuun 2. päivänä 2020 kello 11.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
on December 2, 2020 at 11 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2020

Editors
Timo Männikkö
Faculty of Information Technology, University of Jyväskylä
Ville Korkiakangas
Open Science Centre, University of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-8407-6

ISBN 978-951-39-8407-6 (PDF)
URN:ISBN:978-951-39-8407-6
ISSN 2489-9003

Hold fast to dreams
For if dreams die
Life is a broken-winged bird
That cannot fly.
Hold fast to dreams
For when dreams go
Life is a barren field
Frozen with snow.

— Langston Hughes (1902-1967)

ABSTRACT

Wang, Xiulin
Coupled Nonnegative Matrix/Tensor Factorization in Brain Imaging Data
Jyväskylä: University of Jyväskylä, 2020, 58 p. (+included articles)
(JYU Dissertations
ISSN 2489-9003; 321)
ISBN 978-951-39-8407-6 (PDF)

Continuous advancement of brain imaging techniques has witnessed data analysis methods evolving from matrix component analysis to tensor component analysis, from individual analysis to group analysis regarding the analysis of brain data with multi-set/multi-modal, multi-coupling and multi-way characteristics. Coupled matrix/tensor factorization is robust in merging the advantages of analysis methods, including multi-way retainability, flexible coupling settings, mild uniqueness conditions, and applicability of various constraints, which is relatively difficult for most existing methods. Therefore, this dissertation aims to develop efficient coupled nonnegative matrix/tensor factorization algorithms, which can be used for the analysis of brain imaging data at the group level.

First, aiming at constrained group analysis of data from multiple sources, we design a flexible model of coupled nonnegative matrix factorization with sparse regularization and adopt alternating direction method of multipliers (ADMM) for optimization. Then, to reduce the high computational cost of large-scale problems, we propose three efficient coupled nonnegative tensor factorization algorithms, which are respectively based on fast hierarchical alternating least squares (fHALS), alternating proximal gradient (APG) and a combination of APG and low-rank approximation.

Experiments using synthetic and real-world data are conducted to demonstrate the performances of the proposed algorithms. Specifically, for multi-subject simulated functional magnetic resonance imaging data, the proposed ADMM-based algorithm can achieve better performance than its competitors and extract both common and individual patterns while correcting the disorders of common patterns. For multi-subject ongoing electroencephalography data, the proposed fHALS-based algorithm can effectively extract brain activities of interest associated with the musical stimulus. For multi-subject event-related potential data, the proposed APG-based algorithms can obtain higher decomposition accuracy and more robust multi-domain feature extraction stability, and low-rank approximation can greatly improve computation efficiency without losing the accuracy.

Overall, according to data characteristics , we have developed efficient coupled nonnegative matrix/tensor decomposition algorithms, which have been successfully applied to the group analysis of brain imaging data.

Keywords: Brain imaging data, coupled constraint, group analysis, nonnegative matrix/tensor factorization, sparse regularization

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Wang, Xiulin
Liitoksellisten matriisien / tensorien epänegatiivinen yhteishajotelma aivokuvantamisaineistossa
Jyväskylä: University of Jyväskylä, 2020, 58 s. (+artikkelit)
(JYU Dissertations
ISSN 2489-9003; 321)
ISBN 978-951-39-8407-6 (PDF)

Aivokuvantamismenetelmien jatkuvaa kehitystä on seurannut aineistojen analyysiin tarkoitettujen menetelmien kehittyminen matriisikomponenttianalyysistä tensorikomponenttianalyysiin ja yksilötason analyysistä ryhmätason analyysiin . Liitoksellisten matriisien / tensorien yhteishajotelma luotettavasti säilyttää rakenteiden ulottuvaisuuden, sallii joustavat rakenteiden liitokset, on yksikäsitteinen lievin ehdon, ja mahdollistaa monien eri reunaehtojen huomioimisen . Tämän väitöskirjan tavoite on kehittää liitoksellisten matriisien / tensorien epänegatiivisen yhteishajotelman laskemiseksi tehokkaita algoritmeja, jotka soveltuvat aivokuvantamisaineistojen ryhmätason analyysiin.

Tavoitteenamme usean datalähteen ryhmätason analyysi reunaehdot huomioon ottaen, esittelemme joustavan liitoksellisten matriisien epänegatiivisen yhteishajotelman, jossa käytetään harvan ratkaisun tuottavaa regularisaatiota, ja joka ratkaistaan käyttämällä kertoimien vuorottelevan suunnan menetelmää (ADMM). Vähentääksemme korkeaa laskennallista vaativuutta , ehdotamme kolmea tehokasta liitoksellisten tensorien epänegatiivista yhteishajotelmaa varten suunniteltua algoritmia, jotka vastaavasti perustuvat kolmeen eri menetelmään: nopeaan, hierarkkiseen ja vuorottelevaan pienimmän neliösumman menetelmään (fHALS), vuorottelevan proksimaalisen gradientin menetelmään (APG), ja yhdistelmään joka koostuu APG:sta ja matala-asteisesta approksimaatiosta (LRA).

Esitettyjen algoritmien suorituskykyä arvioidaan sekä simuloidulla että oikealla aineistolla. Usean koehenkilön simuloidulle fMRI-aineistolle ehdotettu ADMM-pohjainen algoritmi pystyy saavuttamaan parempia tuloksia kuin vaihtoehdot. Usean koehenkilön jatkuvassa elektroen-kefalografia-aineistossa ehdotettu fHALS-perustainen algoritmi onnistuu eristämään kiinnostavia aivotoimintoja. Usean koehenkilön herätevasteaineistossa ehdotetut APG-perustaiset algoritmit saavuttavat tarkempia hajotelmia ja löytävät olennaisia piirteitä aineistosta vakaammin. Lisäksi matala-asteinen approksimaatio voi huomattavasti lisätä laskennallista tehokkuutta kuitenkaan menettämättä tarkkuutta.

Tiivistäen, olemme siis kehittäneet tehokkaita algoritmeja liitoksellisten matriisien / tensorien epänegatiiviselle yhteishajotelmalle, ja soveltaneet niitä onnistuneesti ryhmätason analyysiin aivokuvantamisaineistossa.

Avainsanat: Aivokuvantamisaineisto, liitosrajoite, ryhmäanalyysi, matriisien / tensorien epänegatiivinen hajotelma, harva regularisaatio

Author

Xiulin Wang
Faculty of Information Technology
University of Jyväskylä
Finland
Email: xiulin.wang@foxmail.com
ORCID: https://orcid.org/0000-0002-8884-0973

Supervisor

Professor Fengyu Cong
School of Biomedical Engineering
Dalian University of Technology
China
Faculty of Information Technology
University of Jyväskylä
Finland

Docent Zheng Chang
Faculty of Information Technology
University of Jyväskylä
Finland

Professor Timo Hämäläinen
Faculty of Information Technology
University of Jyväskylä
Finland

Professor Tapani Ristaniemi
Faculty of Information Technology
University of Jyväskylä
Finland

Reviewers

Professor Guoxu Zhou
School of Automation
Guangdong University of Technology
China

Assistant Professor Zheng Zhang
Department of Electrical & Computer Engineering
University of California
United States

Opponent

Professor Qibin Zhao
Tensor Learning Team
RIKEN Center for Advanced Intelligence Project (AIP)
Japan

ACKNOWLEDGEMENTS

I would like to acknowledge the people who journeyed with me in the last several years. Without their help and support, it is impossible to finish this dissertation.

First and foremost, I would like to express my most profound appreciation to my supervisors: Professor Tapani Ristaniemi, Professor Fengyu Cong, Docent Zheng Chang and Professor Timo Hämäläinen, for their unwavering supports and guidance during my doctoral research. I am incredibly lucky to have four excellent supervisors. Their rich experience, unparalleled knowledge, and ingenious suggestions benefited me not only in academic but also in life. Their patient with people, their wisdom in life and their research expertise are the goals I have always pursued. Tapani and Fengyu mentored me throughout almost my entire doctoral study. To Tapani, he always gave quick feedback on my various needs and provided insightful suggestions on my reports; his smile was very contagious and he could always liberate us from tense scientific research atmosphere with his particular humor. To Feng yu, he taught me how to become a qualified doctoral student; he encouraged and supported me to attend advanced seminars and conferences worldwide; he patiently unwrapped my academic confusions with creative ideas; he gave me warm consolations when I was down and guided me on how to fight it. Zheng and Timo became my supervisors in the last part of my doctoral stage. To Zheng, he is more like a big brother who can share his personal research experience with us and give the straightforward and very efficient advice. I also thank him for serving as the custos of my thesis defense. To Timo, he actively and conscientiously reminded me to carry out all the steps of preparing for the thesis pre-examination and defense in time, which really provided a lot of help for my graduation. I also thank them for providing me with such an opportunity for doctoral research in the University of Jyväskyä.

I would like to express my sincere thanks to the reviewers of my dissertation, Professor Guoxu Zhou and Assistant Professor Zheng Zhang, for their constructive comments and valuable suggestions. I am very grateful to Professor Qibin Zhao for accepting the invitation and being my opponent in my public defense. I would like to thank Associate Professor Xiaofeng Gong, who supervised me when I was a master student and encouraged me to start my doctoral research.

I am particularly grateful to Zhonghua Chen, Jia Liu, Wenya Liu, Yanru Liu, Ye Ren, Xiaobang Sun, Lili Tian, Biying Wang, Deqing Wang, Xiaoshuang Wang, Yingying Wang, Huashuai Xu, Rui Yan, Tiantian Yang, Dongdong Zhou, Guanghui Zhang, Yongjie Zhu, and many other friends, who had made my doctoral study in Jyväskylä so wonderful. The life with you here granted me the happiest time and the most precious memories since my carefree childhood. Similar thanks to Guoqiang Hu, Yuxing Hao, Xiaoyu Wang, Wei Zhao, and all other colleagues in the ASAP laboratory of Dalian University of Technology for their help and support during my doctoral research. Also, I want to thank Heinilä Erkka, who helped me translate the Finnish abstract and summary in my dissertation.

As a co-cultivated doctoral student, I would like to thank the University of Jyväskylä and Dalian University of Technology in China for establishing such an important training platform. I would like to thank the China Scholarship Council, and the Faculty of Information Technology at University of Jyväskylä for providing the funding that has enabled my doctoral research, as well as the funding for my travels for conferences.

Finally and most importantly, I would like to express my deepest gratitude to my parents, my sister, my grandpa, and all the other family members for their continuous support and endless love. They gave me unlimited courage and freedom to see the outside world. I hope I am the one who makes them proud.

This work is dedicated to the memory of Professor Tapani Ristaniemi, whose integrity, wisdom and grace will always be missed and remembered.

Jyväskylä, July 2020
Xiulin Wang

LIST OF ACRONYMS

ADMM	alternating direction method of multipliers
APG	alternating proximal gradient
BCD	block coordinate decent
CNMF	couple nonnegative matrix factorization
CNTF	couple nonnegative tensor factorization
CP	canonical polyadic
EEG	electroencephalography
ERP	event-related potential
fHALS	fast hierarchical alternating least squares
fMRI	functional magnetic resonance imaging
HALS	hierarchical alternating least squares
ICA	independent component analysis
MCC	maximum correlation coefficient
MEG	magnetoencephalography
NMF	nonnegative matrix factorization
NTF	nonnegative tensor factorization
PARAFAC	parallel factor analysis
PCA	principle component analysis
sMRI	structural magnetic resonance imaging

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF INCLUDED ARTICLES

PI Xiulin Wang, Wenya Liu, Fengyu Cong, and Tapani Ristaniemi. Group Nonnegative Matrix Factorization with Sparse Regularization in Multi-set Data. *28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands*, 2020.

PII Xiulin Wang, Chi Zhang, Tapani Ristaniemi and Fengyu Cong. Generalization of Linked Canonical Polyadic Tensor Decomposition for Group Analysis. *16th International Symposium on Neural Networks (ISNN 2019), Moscow, Russia*, http://doi.org/10.1007/978-3-030-22808-8_19, 2019.

PIII Xiulin Wang, Tapani Ristaniemi and Fengyu Cong. Fast Implementation of Double-coupled Nonnegative Canonical Polyadic Decomposition. *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK*, http://doi.org/10.1109/ICASSP.2019.8682737, 2019.

PIV Xiulin Wang, Wenya Liu, Petri Toiviainen, Tapani Ristaniemi and Fengyu Cong. Group analysis of ongoing EEG data based on fast double-coupled nonnegative tensor decomposition. *Journal of neuroscience methods, 330, p.108502*, http://doi.org/10.1016/j.jneumeth.2019.108502, 2020.

PV Xiulin Wang, Tapani Ristaniemi and Fengyu Cong. Fast Learnings of Coupled Nonnegative Tensor Decomposition Using Optimal Gradient and Low-rank Approximation. *Submitted to signal processing*, 2020.

1 INTRODUCTION

1.1 Brain imaging data

In the field of cognitive neuroscience, the advent of advanced brain imaging techniques has enabled researchers to explore the brain from various perspectives. For example, from the view of brain function, electroencephalography (EEG) is a collection of potentials along the scalp that reflect electrical activities of the brain; magnetoencephalography (MEG) maps brain activities by recording magnetic fields generated by electrical currents in the brain; functional magnetic resonance imaging (fMRI) measures hemodynamic responses related to neural activities of the brain. From the view of brain structure, structural MRI (sMRI) and diffusion tensor imaging (DTI) present structural information (mainly gray matter, white matter and cerebrospinal fluid) about the brain, and DTI can additionally provide brain structural connectivity. Incidentally, a large amount of brain imaging data are generated with the following characteristics:

(1) Multi-set/Multi-modal: Regarding multiple datasets in brain imaging data, such as those from multiple subjects, multiple modalities, multiple trials, multiple groups or multiple tasks, following the article (Chen et al., 2016), two major categories are defined in this dissertation. (i) Multiple datasets collected from the same type of brain data are termed as *multi-set*, e.g., fMRI data collected from multiple subjects under a simple visual paradigm (Calhoun et al., 2001). (ii) Multiple datasets collected from different types of brain data are termed as *multi-modal*, e.g., fMRI, sMRI and EEG data collected from patients with schizophrenia and healthy controls during an auditory oddball task (Adali et al., 2015). Each type of data can record brain information from specific aspect. However, it sometimes fails to provide complete information of interest. Thus the brain studies in cognitive neuroscience are more inclined to the joint analysis of the data, which can promise to provide us with more insights of the brain and enhance our understandings of the brain (Sui et al., 2012; Karahan et al., 2015; Lahat et al., 2015; Adali et al., 2015).

(2) Multi-coupling: Benefiting from the development of brain imaging technologies, we now have access to multi-set/multi-modal brain data that are coupled in various types. For data collected from the same phenomenon, it is reasonable to expect them (collected from different subjects or trials) to have the same or highly correlated information, which makes the data linked together. For example, in the group analysis of ongoing EEG data, the authors found that there were significant correlations among subjects only in spatial and spectral modes, while the correlation in temporal mode was almost non-existent (Wang et al., 2020). The complementary strengths among multi-modal data can also be seen as a type of coupling. For example, EEG has excellent temporal but poor spatial resolution. On the contrary, fMRI provides high spatial but low temporal resolution. Combining fMRI and EEG data can naturally achieve extremely high spatial-temporal accuracy, which has been applied to clinical medicine such as epilepsy and sleep disorders (Ritter and Villringer, 2006; Gotman et al., 2004). Besides, further incorporation of sMRI data that captures brain structural information can obtain more comprehensive information of the brain across the structure and function (Adali et al., 2015; Levin-Schwartz et al., 2014; Acar et al., 2017a). There are also some other combining types such as fMRI and MEG (Plis et al., 2010), sMRI and EEG (Calhoun et al., 2006), and positron emission tomography (PET) and sMRI (Specht et al., 2009).

(3) Multi-way: In addition to two intrinsic ways (also known as modes or dimensions) of time and space, most brain imaging data generally have many other modes, such as frequency, subject, trial, group and condition. Thus the data can be naturally represented as multi-way arrays, i.e., tensors (Cichocki, 2013; Cong et al., 2015a; Zhou et al., 2016). For example, EEG data of multiple channels per subject can be expressed as a two-way matrix with space and time. EEG data from multiple subjects can be arranged into a three-way array with space, time and subject. Further considering time-frequency representation, a fourth-order tensor of space, time, frequency and subject can be generated(Cong et al., 2012b). For event-related potential (ERP) data, the authors generated a fifth-order tensor of space, time, frequency, subject and condition (Mørup et al., 2006a; Wang et al., 2018).

1.2 Conventional methods

The development of data imaging techniques not only produces a large amount of multiple brain datasets, but also promotes the continuous upgrade of data analysis methods from two-way (matrix-based) component analysis to multi-way (tensor-based) component analysis. Next, a review of two-way and multi-way component analysis methods and their applications in brain imaging data is provided.

Two-way component analysis: Matrix factorization methods, such as in-

dependent component analysis (ICA), principle component analysis (PCA) and nonnegative matrix factorization (NMF), can recover factors matrices of interest from the observed matrix via imposing specific constraints and have been widely used to analyze brain imaging data. For example, Cong et al. employed ICA to decompose ongoing EEG and finally built the connections between brain activities and naturalistic music (Cong et al., 2013). Salimi-Khorshidi et al. combined ICA and hierarchical fusion of classifiers and achieved automatic denoising of fMRI data (Salimi-Khorshidi et al., 2014). Using PCA and regression analysis, Zhong et al. proposed an fMRI connectivity analysis approach to detect functional connectivity between the brain regions (Zhong et al., 2009). Chen et al. applied constrained NMF for early detection of Alzheimer disease using clinical EEG recordings (Chen et al., 2006).

Although the methods above have been proved to be effective in many studies, their main contributions are concentrated on the analysis of a single dataset. When faced with the group analysis of multi-set/multi-modal data, they cannot sufficiently utilize the coupling information between datasets, which limits their broader usage to some extent (Sui et al., 2012; Chen et al., 2016; Gong et al., 2015). As such, group analysis methods for multiple datasets have drawn great interest from researchers. Calhoun et al. first proposed the group ICA method and applied it to multi-subject fMRI data for making group inferences, in which the data were concatenated together along the temporal dimension by assuming sharing a common source subspace (Calhoun et al., 2001). Using different assumptions and data grouping strategies, joint ICA was applied to simultaneously analyze fMRI and sMRI data collected from schizophrenia patients and healthy controls (Calhoun et al., 2006). Linked ICA was applied to analyze the datasets of Alzheimer's patients and age-matched controls, which combines two structural MRI data: morphological data and diffusion data (Groves et al., 2011). Canonical correlation analysis (CCA) and its extension to multiple datasets, namely multi-set CCA (MCCA), achieve JBSS of multiple datasets through maximizing the correlation among extracted sources (Li et al., 2009). Correa et al. applied them to the data fusion of two or three modalities, such as fMRI, EEG, and sMRI data obtained from patients with schizophrenia and healthy controls (Correa et al., 2008, 2010a), and concurrent EEG and fMRI data acquired in an auditory task (Correa et al., 2010b). Examples of similar joint analysis methods include but are not limited to: independent vector analysis (IVA, (Lee et al., 2008; Adali et al., 2014)), partial least squares (PLS, (Krishnan et al., 2011)), parallel ICA (pICA, (Liu and Calhoun, 2007; Liu et al., 2009b,a; Jagannathan et al., 2010)) and group NMF (GNMF, (Lee and Choi, 2009; Shin and Oh, 2012)). For various variants and combinations of the methods mentioned above, as well as more data fusion or joint analysis methods, please refer to (Sui et al., 2012; Lahat et al., 2015; Chen et al., 2016).

Multi-way component analysis: In addition to two-way component analysis methods, multi-way (tensor) methods are also widely applied in the analysis of brain imaging data by virtue of retaining the multi-way nature of the data(Cichocki, 2013; Cong et al., 2015a; Zhou et al., 2016; Mørup, 2011; Hunyadi et al., 2017). Importantly, more modes such as spatial, temporal and spectral in-

formation of brain data can be simultaneously considered through tensor component analysis. It can reveal the structural connections between various modes and present more interpretable physical and physiological significance (Cong et al., 2015a; Mørup et al., 2006b; Cong et al., 2015b). For example, Beckmann and Smith applied tensor probabilistic ICA to multisubject/multisession fMRI studies, in which the components of interest in spatial, temporal, and subject/session domains were directly estimated (Beckmann and Smith, 2005). Acar et al. applied parallel factor analysis (PARAFAC, (Harshman et al., 1970)) to epilepsy study by generating an EEG tensor of temporal, spectral and spatial modes, and proved to be effective in epileptic focus localization (Acar et al., 2007). Multilinear PLS and nonnegative tensor factorization (NTF) were performed to analyze ongoing EEG elicited by natural music stimuli (Cong et al., 2012b; Wang et al., 2016). Kuang et al. combined ICA and shift-invariant tensor decomposition for the multi-subject fMRI analysis. Spyrou et al. applied PARAFAC2-based complex tensor factorization in EEG connectivity estimation using the data from patients with mild cognitive impairment or Alzheimer's disease (Spyrou et al., 2018). Zhu et al. proposed a tensor component analysis-based framework to track dynamic task-modulated functional networks using MEG data (Zhu et al., 2020).

Furthermore, increasing recognition of joint analysis has gradually drawn researchers' attention to tensor-based data fusion. For example, a coupled matrix-tensor factorization model was applied to EEG and fMRI data fusion to capture the difference of brain activity between patients with schizophrenia and healthy controls (Acar et al., 2017b). Coupled tensor decomposition was conducted to extract common task features from EEG-fMRI data in an N-Back memory task (Jonmohamadi et al., 2019) and signal sources from MEG-EEG data for differentiating oscillators during intermittent photic stimulation (Naskovska et al., 2020). Moreover, Mørup et al. developed the software ERPWAVELAB, a toolbox for multi-channel time-frequency analysis of EEG and MEG data through nonnegative multi-way (tensor) factorization based on PARAFAC and Tucker models (Mørup et al., 2007). For more tensor-based analysis about brain imaging data, please refer to (Karahan et al., 2015; Cong et al., 2015a).

1.3 Research questions and motivations

For the joint analysis of multiple datasets, traditional matrix factorization methods such as ICA and PCA only analyze each dataset separately, and then performs post-mortem analysis (such as clustering or correlation analysis) to find the group inference among datasets. Obviously, the mutual information between datasets cannot be effectively utilized. Group analysis methods can take advantage of the dependencies of extracted sources between datasets, potentially obtaining higher performance than single analysis methods originally designed for a single dataset (Chen et al., 2016; Gong et al., 2015). However, when processing the multiple datasets with inherent multi-way structures, the above-mentioned

two-way analysis methods often concatenate or stack the extra modes besides two modes to generate a matrix so that the two-way matrix analysis methods can be applied, such unfolding operation inevitably loses some potentially existing interactions between/among the folded modes (Cong et al., 2015a). Moreover, two-way analysis methods generally cannot guarantee a unique solution. In contrast, multi-way analysis methods obtain a unique solution under fairly mild conditions (Kolda and Bader, 2009).

Multi-way analysis methods often stack the matrices or tensors to generate higher-order tensor with the assumption of complete consistency among corresponding modes. For example, Cong et al. constructed a fourth-order ongoing EEG tensor with the dimension of space, time, frequency and subject, in which the spatial, temporal and frequency information among subjects was assumed to be consistent (Cong et al., 2012b). However, with the same datasets, the study in (Wang et al., 2020) only found significant correlations in spatial and spectral modes among subjects. Furthermore, if performing the tensor method directly and separately on subjects, the prior coupling information in spatial and spectral modes between subjects cannot be utilized.

From the perspective of data analysis, according to data characteristics, imposing specific constraints (such as sparseness, nonnegativity, and smoothness) on underlying extracted factors would contribute to obtaining more meaningful solutions (Cichocki, 2013). Undoubtedly, in the era of big data, the high computational cost of processing large-scale datasets is always one of the existing issues (Wang et al., 2019a; Zhou et al., 2012). Therefore, to address the problems mentioned above, in this dissertation, we mainly study coupled nonnegative matrix/tensor factorization on the group analysis of brain imaging data, in which the following advantages can be included: (i) coupled matrix/tensor factorization can be considered as a generalization of matrix/tensor factorization to multiple datasets, providing a natural framework for the joint analysis of matrices or tensors with coupling information; (ii) it considers incomplete consistency of the corresponding modes between datasets and enables the simultaneous extraction of common and individual features; (iii) it can potentially reveal underlying structures and inner-relationships between datasets with retaining the original data structure; (iv) it can take full advantage of prior information such as flexible constraints and achieve higher decomposition accuracy and more robust decomposition stability.

1.4 Structure of the dissertation

The structure of this dissertation is listed as follows.

Chapter 1 briefly introduces brain imaging data and related conventional analysis methods and interprets the main research questions and motivations of this dissertation.

Chapter 2 describes coupled nonnegative matrix/tensor factorization prob-

lems, including mathematical concepts, models and optimization strategies, and provides some verification experiments.

Chapter 3 briefly summarizes the included articles and lists the contributions of authors to the articles.

Chapter 4 presents the conclusion of this dissertation, as well as the research contributions, limits and plans.

2 COUPLED NONNEGATIVE MATRIX/TENSOR FACTORIZATION

Coupled nonnegative matrix/tensor factorization, an important extension of nonnegative matrix/tensor factorization to multi-block matrices/tensors which need to be jointly analyzed, has been successfully applied in many fields, such as biomedical signal processing (Wang et al., 2020; Jonmohamadi et al., 2019), array signal processing (Sørensen and De Lathauwer, 2013), joint blind source separation (Gong et al., 2018a), classification (Lee and Choi, 2009) and metabolic physiology (Acar et al., 2015). Following the articles PI-PV, this chapter comprehensively elaborates on the coupled nonnegative matrix/tensor decomposition, including basic notations and tensor operations, mathematical models, optimization methods, experiments and results.

2.1 Notations and operations

Tensors, also known as multi-way/multi-dimensional/multi-mode arrays, are the higher-order generalizations of scalars, vectors and matrices, which can more intuitively present the structural characteristics of high-dimensional data. Generally, tensors are denoted by calligraphic boldface uppercase letters, e.g., $\boldsymbol{\mathcal{A}} \in \mathbb{R}_{+}^{I_1 \times I_2 \times \cdots \times I_N}$. $\mathbb{R}$ and $\mathbb{R}_{+}$ represent that all elements of the tensor are real and nonnegative values, respectively. N denotes the order of the tensor, i.e., the number of ways, arrays or dimensions. In addition, matrices, vectors, scalars are denoted by boldface uppercase, boldface uppercase and lowercase letters. For example, for a matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_J] \in \mathbb{R}_{+}^{I \times J}$, the jth column vector of $\boldsymbol{A}$ is denoted as $\boldsymbol{a}_j$ and the (i, j)th element of $\boldsymbol{A}$ is denoted as a_{ij}. The operators of $(\cdot)^T$, $[\![\cdot]\!]$, $\|\cdot\|_F$ and $\|\cdot\|_1$ denote transpose, Kruskal operator, Frobenius norm and l_1-norm, respectively.

Inner product is denoted as $\langle \cdot, \cdot \rangle$. For the matrices $\boldsymbol{A} \in \mathbb{R}_{+}^{I \times J}$ and $\boldsymbol{B} \in \mathbb{R}_{+}^{I \times J}$, their inner product is defined as $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i,j} a_{ij} b_{ij}$, which can be substituted by

$\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^T)$, and $\mathrm{tr}(\cdot)$ means the trace of a matrix.
Outer product is denoted as '$\circ$'. The outer product of N vectors $\boldsymbol{a}_n \in \mathbb{R}_+^{I_n}, n = 1, 2, \cdots, N$ can generate an Nth-order rank-1 tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ as $\boldsymbol{\mathcal{A}} = \boldsymbol{a}_1 \circ \boldsymbol{a}_2 \circ \cdots \circ \boldsymbol{a}_N$, and $\boldsymbol{\mathcal{A}}_{i_1 i_2 \cdots i_N} = a_{i_1} a_{i_2} \cdots a_{i_N}$. Specially, for two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, their outer product is $\boldsymbol{a} \circ \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^T$.
Kronecker product and **Khatri-Rao product** are denoted as '$\otimes$' and '$\odot$', and defined as follows:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}_+^{I_1 I_2 \times J_1 J_2}, \boldsymbol{A} \in \mathbb{R}_+^{I_1 \times J_1}, \boldsymbol{B} \in \mathbb{R}_+^{I_2 \times J_2} \tag{1}$$

and

$$\boldsymbol{A} \odot \boldsymbol{B} = [\boldsymbol{a}_1 \otimes \boldsymbol{b}_1, \boldsymbol{a}_2 \otimes \boldsymbol{b}_2, \cdots, \boldsymbol{a}_J \otimes \boldsymbol{b}_J] \in \mathbb{R}_+^{I_1 I_2 \times J}, \boldsymbol{A} \in \mathbb{R}_+^{I_1 \times J}, \boldsymbol{B} \in \mathbb{R}_+^{I_2 \times J}. \tag{2}$$

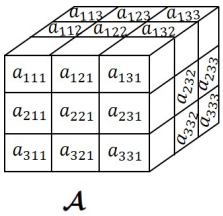
Element-wise product (also known as Hadamard product) and **element-wise division** are denoted as '$\circledast$' and '$\oslash$', and defined as follows:

$$\boldsymbol{A} \circledast \boldsymbol{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots \\ a_{21}b_{21} & a_{22}b_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}_+^{I \times J}, \boldsymbol{A} \in \mathbb{R}_+^{I \times J}, \boldsymbol{B} \in \mathbb{R}_+^{I \times J} \tag{3}$$

and

$$\boldsymbol{A} \oslash \boldsymbol{B} = \begin{bmatrix} a_{11}/b_{11} & a_{12}/b_{12} & \cdots \\ a_{21}/b_{21} & a_{22}/b_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \in \mathbb{R}_+^{I \times J}, \boldsymbol{A} \in \mathbb{R}_+^{I \times J}, \boldsymbol{B} \in \mathbb{R}_+^{I \times J}. \tag{4}$$

Moreover, for the matrices $\boldsymbol{A}^{(n)}$, $n = 1, 2, \cdots, N$, $\boldsymbol{A}^{(N)} \odot \cdots \odot \boldsymbol{A}^{(2)} \odot \boldsymbol{A}^{(1)}$ and $\boldsymbol{A}^{(N)} \odot \cdots \odot \boldsymbol{A}^{(n+1)} \odot \boldsymbol{A}^{(n-1)} \odot \cdots \odot \boldsymbol{A}^{(1)}$ are defined as $\boldsymbol{A}^{\odot}$ and $\boldsymbol{A}^{\odot -n}$, respectively. $\boldsymbol{A}^{\circledast}$ and $\boldsymbol{A}^{\circledast -n}$ are defined in the similar way.
Tensor matricization is unfolding/flattening a tensor into a matrix in a specific order. The mode-n matricization of tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ can yield a matrix defined as $\boldsymbol{A}_{(n)} \in \mathbb{R}_+^{I_n \times (I_1 I_2 \cdots I_{n-1} I_{(n+1)} \cdots I_N)}$ by arranging the mode-n fibers of $\boldsymbol{\mathcal{A}}$ as the columns of $\boldsymbol{A}_{(n)}$. Here taking a third-order tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}_+^{3 \times 3 \times 3}$ as an example, its mode-n matricizations have been illustrated in Figure 1.
Tensor vectorization aims to unfold/flatten a tensor into a vector. The vectorization of the tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ is defined by $\mathrm{vec}(\boldsymbol{\mathcal{A}}) \in \mathbb{R}_+^{I_1 I_2 \cdots I_N}$, and $[\mathrm{vec}(\boldsymbol{\mathcal{A}})]_j = \boldsymbol{\mathcal{A}}_{i_1 i_2 \cdots i_N}$ with $j = \sum_{n=1}^{N} (i_n - 1) \prod_{m=1}^{n-1} I_m + 1$.
Tensor diagonalization is denoted by $\mathrm{ddiag}(\cdot)$, which aims to extract the super-diagonal elements of a cube tensor to generate a vector. It is defined as a generalization of MATLAB command `diag` to the tensor case. For a cube tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}_+^{I \times I \times \cdots \times I}$, $\mathrm{ddiag}(\boldsymbol{\mathcal{A}}) = [\boldsymbol{\mathcal{A}}_{11\cdots 1}, \boldsymbol{\mathcal{A}}_{22\cdots 2}, \cdots, \boldsymbol{\mathcal{A}}_{II\cdots I}]^T$. Conversely, for a vector $\boldsymbol{a} = [a_1, a_2, \cdots, a_I]^T$, $\mathrm{ddiag}(\boldsymbol{a})$ can yield a super-diagonal tensor $\boldsymbol{\mathcal{A}}$ with $\boldsymbol{\mathcal{A}}_{ii\cdots i} = a_i$ and all other elements are 0.
Please refer to the review paper (Kolda and Bader, 2009) for a more detailed description of standard notations and basic tensor operations.



Mode-1: $\boldsymbol{A}_{(1)} =$

a_{111}	a_{121}	a_{131}	a_{112}	a_{122}	a_{132}	a_{113}	a_{123}	a_{133}
a_{211}	a_{221}	a_{231}	a_{212}	a_{222}	a_{232}	a_{213}	a_{223}	a_{233}
a_{311}	a_{321}	a_{331}	a_{312}	a_{322}	a_{332}	a_{313}	a_{323}	a_{333}

Mode-2: $\boldsymbol{A}_{(2)} =$

a_{111}	a_{211}	a_{311}	a_{112}	a_{212}	a_{322}	a_{113}	a_{213}	a_{313}
a_{121}	a_{221}	a_{321}	a_{122}	a_{222}	a_{322}	a_{123}	a_{223}	a_{323}
a_{131}	a_{231}	a_{331}	a_{132}	a_{232}	a_{332}	a_{133}	a_{233}	a_{333}

Mode-3: $\boldsymbol{A}_{(3)} =$

a_{111}	a_{211}	a_{311}	a_{121}	a_{221}	a_{321}	a_{131}	a_{231}	a_{331}
a_{112}	a_{212}	a_{312}	a_{122}	a_{222}	a_{322}	a_{132}	a_{232}	a_{332}
a_{113}	a_{213}	a_{313}	a_{123}	a_{223}	a_{323}	a_{133}	a_{233}	a_{333}

FIGURE 1 Example of mode-1, mode-2 and mode-3 unfolding of a third-order tensor.

2.2 Mathematical models

This section introduces the definition of the nonnegative matrix/tensor factorization model for single matrix or tensor data. Further it introduces how to formulate the joint analysis of multi-block matrices or tensors into the coupled nonnegative matrix/tensor factorization problems.

2.2.1 Nonnegative matrix factorization

Nonnegative matrix factorization (NMF), as an unsupervised and part-based learning method, has attracted increasing interests for providing a solution to extract potentially hidden factors that are meaningful and physically interpretable (Lee and Seung, 1999). In the last two decades, NMF and its variants have been successfully applied in a variety of fields including text mining (Pauca et al., 2004), document clustering (Xu et al., 2003), face recognition (Guillamet and Vitria, 2002), image/audio/biomedical signal processing (Lee and Seung, 1999; Lee et al., 2009; Kameoka et al., 2009) and blind source separation (Cichocki et al., 2006), etc. Unlike traditional matrix factorization methods(e.g., ICA or PCA), NMF can represent nonnegative input data through nonsubtractive combinations of nonnegative vectors.

Given a nonnegative matrix $\boldsymbol{X} \in \mathbb{R}_+^{I\times J}$, NMF searches for two nonnegative factor matrices $\boldsymbol{A} \in \mathbb{R}_+^{I\times R}$ and $\boldsymbol{B} \in \mathbb{R}_+^{R\times J}$ whose product approximates $\boldsymbol{X}$ as $\boldsymbol{X} \approx \boldsymbol{AB}$, as shown in Figure 2. The commonly used objective function for the NMF model can be formulated as:

$$\min_{\boldsymbol{A},\boldsymbol{B}} \; F = \frac{1}{2}\|\boldsymbol{X} - \boldsymbol{AB}\|_F^2, \quad \text{s.t., } \boldsymbol{A} \geq 0, \; \boldsymbol{B} \geq 0, \tag{5}$$

where the factor matrices $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_R]$ and $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_R]$ denote the basis matrix and coefficient matrix, respectively. Generally, $R < \min\{I, \; J\}$.

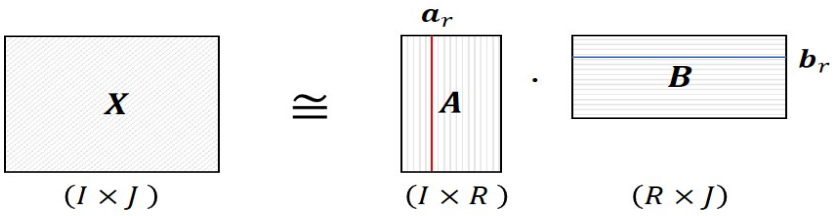


FIGURE 2 Illustration of nonnegative matrix factorization problem: finding two nonnegative factors $\boldsymbol{A} \in \mathbb{R}_+^{I\times R}$ and $\boldsymbol{B} \in \mathbb{R}_+^{R\times J}$ to approximate the input matrix $\boldsymbol{X} \in \mathbb{R}_+^{I\times J}$ such that $\boldsymbol{X} \approx \boldsymbol{AB}$.

2.2.2 Nonnegative tensor factorization

By virtue of the capability to explore the multi-way structure of data, tensor decomposition has been widely used in an ensemble of disciplines, especially signal processing and machine learning (Mørup, 2011; Cichocki et al., 2015; Sidiropoulos et al., 2017). To date, the canonical polyadic (CP) model (Hitchcock, 1927) and Tucker model (Tucker, 1966) are the two most commonly used models for tensor factorization. CP model is also known as canonical decomposition (CANDECOMP, (Carroll and Chang, 1970)) or parallel factor analysis (PARAFAC, (Harshman et al., 1970)). In theory, the CP model can be seen as a special case of the Tucker model. However, compared with the latter, it has better uniqueness even under very mild conditions (Kolda and Bader, 2009; Sidiropoulos et al., 2017). Therefore, the factorization adopted and mentioned later in this dissertation are all based on the CP model.

Generally, given an Nth higher-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1\times I_2\times\cdots\times I_N}$, CP-based tensor factorization can decompose it into a minimum number of rank-1 tensor terms as:

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} d_r \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \circ \cdots \circ \boldsymbol{a}_r^{(N)} = \left[\!\left[\boldsymbol{\mathcal{D}}; \boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \cdots, \boldsymbol{A}^{N} \right]\!\right], \tag{6}$$

where $\boldsymbol{a}_r^{(n)} \in \mathbb{R}^{I_n}$ denotes the rth column of the nth mode factor matrix $\boldsymbol{A}^{(n)} = [\boldsymbol{a}_1^{(n)}, \boldsymbol{a}_2^{(n)} \cdots, \boldsymbol{a}_R^{(n)}] \in \mathbb{R}^{I_n\times R}$. $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{R\times R\times\cdots\times R}$ represents a core tensor with nonzero entries d_r only on its super-diagonal. The minimal number R is defined as the rank of the tensor or the number of components. In analogy to NMF, nonnegative tensor factorization (NTF) also imposes nonnegativity constraint on the underlying factors, and combines the advantages of NMF and tensor factorization (Zhou et al., 2014). Figure 3 gives an illustration of the NTF model. The NTF problem of a given tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}_+^{I_1\times I_2\times\cdots\times I_N}$ in Eq. 6 can be solved by reformulating it as the following optimization problem:

$$\min_{\boldsymbol{\mathcal{D}}, \boldsymbol{A}^{(1)}, \cdots, \boldsymbol{A}^{(N)}} \mathcal{F} = \frac{1}{2} \left\| \boldsymbol{\mathcal{X}} - \left[\!\left[\boldsymbol{\mathcal{D}}; \boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \cdots, \boldsymbol{A}^{N} \right]\!\right] \right\|_F^2 \tag{7}$$

$$\text{s.t. } \boldsymbol{\mathcal{D}} \in \mathbb{R}_+^{R\times R\times\cdots\times R}, \boldsymbol{A}^{(n)} \in \mathbb{R}_+^{I_n\times R}, n = 1, 2, \cdots, N.$$

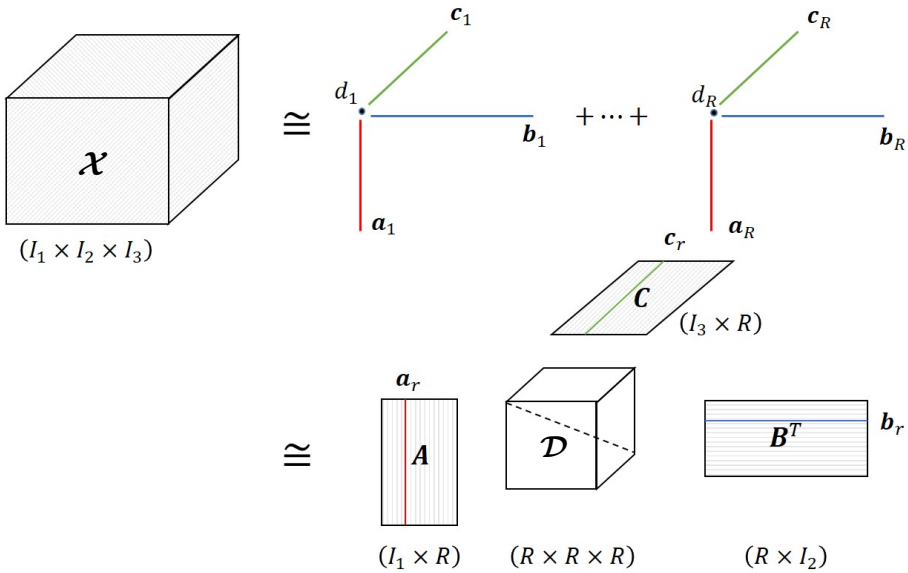


FIGURE 3 Illustration of nonnegative tensor factorization based on the canonical polyadic model: finding four factors $\boldsymbol{A} \in \mathbb{R}_+^{I_1 \times R}$, $\boldsymbol{B} \in \mathbb{R}_+^{I_2 \times R}$, $\boldsymbol{C} \in \mathbb{R}_+^{I_3 \times R}$ and $\boldsymbol{\mathcal{D}} \in \mathbb{R}_+^{R \times R \times R}$ to approximate the input third-order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}_+^{I_1 \times I_2 \times I_3}$ such that $\boldsymbol{\mathcal{X}} \approx [\![\boldsymbol{\mathcal{D}}; \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}]\!]$.

Considering the mode-n matricization of $\boldsymbol{\mathcal{X}}$, the objective function in Eq. 7 can be converted to $\frac{1}{2}\left\|\boldsymbol{X}_{(n)} - \boldsymbol{A}^{(n)}\boldsymbol{D}(\boldsymbol{A}^{\odot -n})^T\right\|_F^2$ and $\boldsymbol{D}$ is a diagonal matrix with entries from the super-diagonal entries of $\boldsymbol{\mathcal{D}}$. If one considers the vectorization of $\boldsymbol{\mathcal{X}}$, Eq. 7 can also be reformulated as $\frac{1}{2}\left\|\text{vec}(\boldsymbol{\mathcal{X}}) - \boldsymbol{A}^{\odot}\text{ddiag}(\boldsymbol{\mathcal{D}})\right\|_F^2$, where $\text{vec}(\boldsymbol{\mathcal{X}})$ denotes the vectorization of tensor $\boldsymbol{\mathcal{X}}$ and $\text{ddiag}(\boldsymbol{\mathcal{D}})$ denotes the vectorization of the super-diagonal entries from the core tensor $\boldsymbol{\mathcal{D}}$.

2.2.3 Coupled nonnegative matrix/tensor factorization

Although NMF and NTF models have shown good performance in many studies, their poor effectiveness in solving the emerging multiple matrix/tensor datasets that need to be analyzed together has currently limited their broader applications (Chen et al., 2016; Gong et al., 2015; Zhou et al., 2016). For example, for multi-subject/multi-modal biomedical data or low spatial resolution hyperspectral (LRHS) and high spatial resolution multispectral (HRMS) image data, the coupling and complementary information between them obviously cannot be fully utilized in the traditional NMF/NTF methods designed for a single dataset (Kanatsoulis et al., 2018; Li et al., 2018; Jonmohamadi et al., 2019; Wang et al., 2020). Therefore, the growing demand for joint analysis of matrix/tensor data makes coupled matrix/tensor factorization applicable in a number of fields, such as data fusion (Acar et al., 2015; Li et al., 2018), signal processing (Shin and Oh, 2012; Sørensen and De Lathauwer, 2013), joint blind source separation (Gong et al., 2018a) and neuroscience (Jonmohamadi et al., 2019; Wang et al., 2020). Note that the *coupled* (also termed as *joint*, *linked* or *group*) here means that the multi-block matrix or tensor datasets not only share the same or highly correlated com-

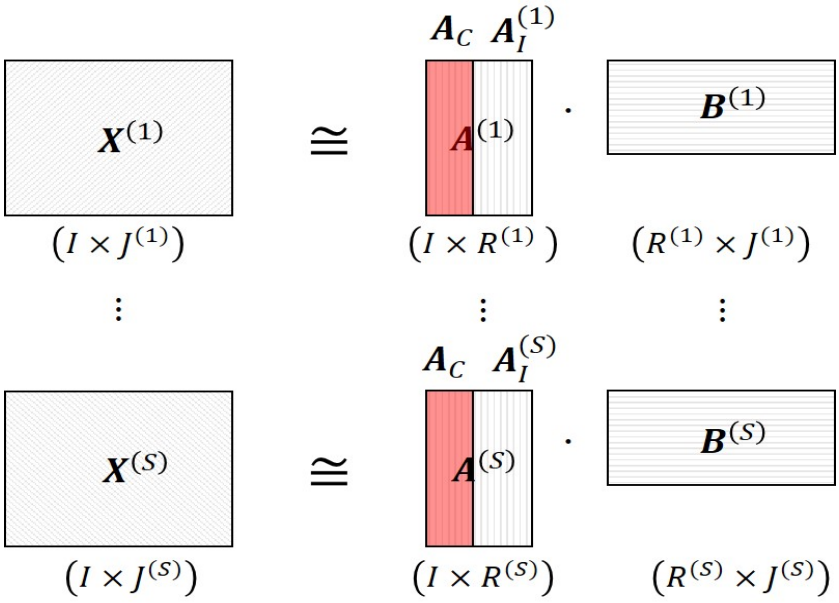


FIGURE 4 Illustration of the coupled nonnegative matrix factorization problem: finding the common components $\boldsymbol{A}_C \in \mathbb{R}_+^{I\times L}$ shared by all the matrices and the individual components $\boldsymbol{A}_I^{(s)} \in \mathbb{R}_+^{I\times(R^{(s)}-L)}$ and $\boldsymbol{B}^{(s)} \in \mathbb{R}_+^{R^{(s)}\times J^{(s)}}$ possessed by the individual matrix such that $\boldsymbol{X}^{(s)} \approx \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} = [\boldsymbol{A}_C\ \boldsymbol{A}_I^{(s)}]\boldsymbol{B}^{(s)}$, $s = 1, 2, \cdots, S$.

ponents but also have their own individual characteristics (Lee and Choi, 2009; Zhou et al., 2015; Yokota et al., 2012; Kim et al., 2015; Chen et al., 2016; Zhou et al., 2016; Gong et al., 2018a; Wang et al., 2019b).

Given a set of nonnegative matrices $\boldsymbol{X} = \{\boldsymbol{X}^{(1)}, \boldsymbol{X}^{(2)}, \cdots, \boldsymbol{X}^{(S)}\}$, each $\boldsymbol{X}^{(s)} \in \mathbb{R}_+^{I\times J^{(s)}}$ satisfies the NMF model, namely $\boldsymbol{X}^{(s)} \approx \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}$, where $\boldsymbol{A}^{(s)} \in \mathbb{R}_+^{I\times R^{(s)}}$ and $\boldsymbol{B}^{(s)} \in \mathbb{R}_+^{R^{(s)}\times J^{(s)}}$ denote the latent variable and corresponding coefficient matrix, respectively. $R^{(s)} < \min(I, J^{(s)})$ is assumed, aiming to provide a low-rank representation of $\boldsymbol{X}^{(s)}$. Furthermore, in the coupled NMF (CNMF) model, each factor matrix $\boldsymbol{A}^{(s)}$ is assumed to include two parts, i.e., $\boldsymbol{A}^{(s)} = [\boldsymbol{A}_C\ \boldsymbol{A}_I^{(s)}]$. The submatrix $\boldsymbol{A}_C \in \mathbb{R}_+^{I\times L}$ consists of a common subspace representing the coupled components shared by all the S matrices. $\boldsymbol{A}_I^{(s)} \in \mathbb{R}_+^{I\times(R^{(s)}-L)}$ consists of subspaces representing the individual characteristics, with $L \leq \min(R^{(s)})$. An illustration of the CNMF model is given in Figure 4. The objective of the CNMF problem is to extract the latent variables and meanwhile separate the common and individual components, which can be achieved by solving the following objective function:

$$\min_{\boldsymbol{A}_C,\, \boldsymbol{A}_I^{(s)},\, \boldsymbol{B}^{(s)}} \frac{1}{2}\sum_{s=1}^{S} \left\| \boldsymbol{X}^{(s)} - \left[\boldsymbol{A}_C\ \boldsymbol{A}_I^{(s)}\right]\boldsymbol{B}^{(s)} \right\|_F^2, \quad \text{s.t. } \boldsymbol{A}_C \geq 0, \boldsymbol{A}_I^{(s)} \geq 0, \boldsymbol{B}^{(s)} \geq 0. \quad (8)$$

In analogy to the CNMF model, the coupled NTF (CNTF) model considers the joint analysis of multi-block tensors with coupling information and provides better performance than the NTF model. Following the article PII, given a set of

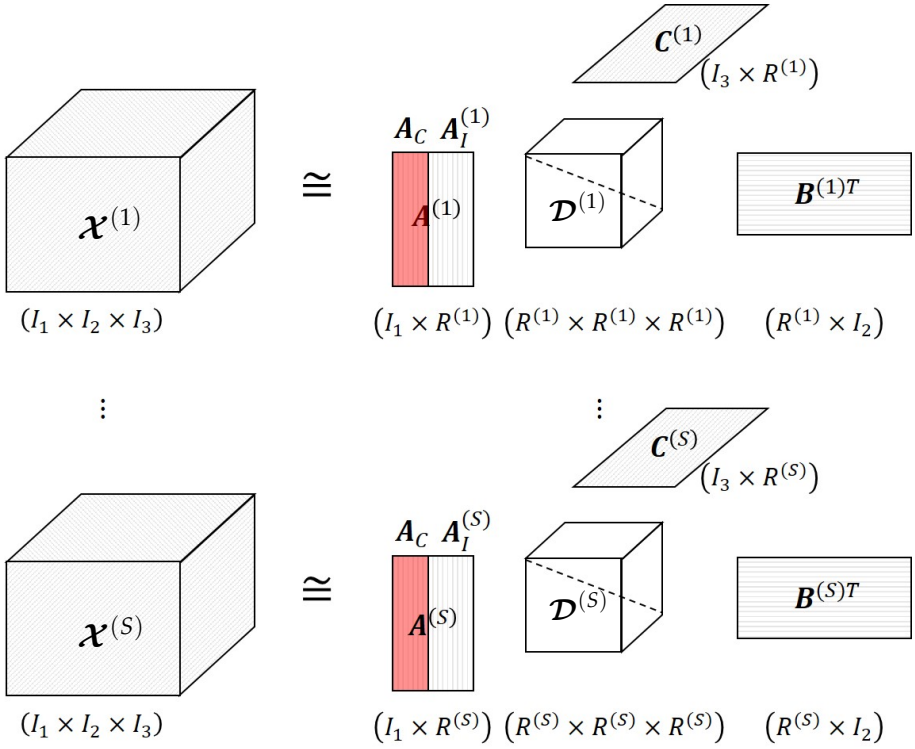


FIGURE 5 Illustration of mode-1 coupled nonnegative CP factorization model for a set of third-order tensors $\mathcal{X}^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times I_3}$: aiming to estimate the factor matrices $A^{(s)} = [A_C \ A_I^{(s)}] \in \mathbb{R}_+^{I_1 \times R^{(s)}}$, $B^{(s)} \in \mathbb{R}_+^{I_2 \times R^{(s)}}$, $C^{(s)} \in \mathbb{R}_+^{I_3 \times R^{(s)}}$ and the core tensors $\mathcal{D} \in \mathbb{R}_+^{R^{(s)} \times R^{(s)} \times R^{(s)}}$, where $A_C \in \mathbb{R}_+^{I_1 \times L}$ represents the common components shared by all the first factors in the tensors.

Nth-order nonnegative tensors $\mathcal{X}^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots I_N}$, $s = 1, 2, \cdots, S$, the generalized CNTF model can be represented as:

$$\mathcal{X}^{(s)} \approx \sum_{r=1}^{R^{(s)}} d_r^{(s)} \boldsymbol{a}_r^{(1,s)} \circ \boldsymbol{a}_r^{(2,s)} \circ \cdots \circ \boldsymbol{a}_r^{(N,s)} = \left[\!\left[\mathcal{D}^{(s)}; A^{(1,s)}, A^{(2,s)}, \cdots, A^{(N,s)} \right]\!\right], \quad (9)$$

where $\boldsymbol{a}_r^{(n,s)}$ denotes the rth column of n-mode factor matrix $A^{(n,s)} \in \mathbb{R}_+^{I_n \times R^{(s)}}$ of sth tensor ($s = 1, 2, \cdots, S$, $n = 1, 2, \cdots, N$). $\mathcal{D}^{(s)} \in \mathbb{R}_+^{R^{(s)} \times R^{(s)} \times \cdots R^{(s)}}$ represents the sth core tensor with non-zero entries $d_r^{(s)}$ only on the super-diagonal elements. Each factor matrix $A^{(n,s)}$ includes two parts: $A_C^{(n)} \in \mathbb{R}_+^{I_n \times L_n}$, $0 \leq L_n \leq R^{(s)}$ and $A_I^{(n,s)} \in \mathbb{R}_+^{I_n \times (R^{(s)} - L_n)}$. $A_C^{(n)}$ shared by all tensors represents the coupling information among them, and $A_I^{(n,s)}$ corresponds to the individual characteristics of each tensor. Figure 5 illustrates the CP-based CNTF model of third-order tensors in the case that their mode-1 factor matrices are partially linked. Therefore, in the CNTF model, aiming to extract the constrained factor matrices $A^{(n,s)}$ and core tensors $\mathcal{D}^{(s)}$ from partially linked tensors $\mathcal{X}^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots I_N}$, the objective function can be represented as follows:

$$\min_{\mathcal{D}^{(s)}, A^{(n,s)}} \mathcal{F} = \frac{1}{2} \sum_{s=1}^{S} \left\| \mathcal{X}^{(s)} - \left[\!\left[\mathcal{D}^{(s)}; A^{(1,s)}, A^{(2,s)}, \cdots, A^{(N,s)} \right]\!\right] \right\|_F^2 \quad (10)$$

$$\text{s.t. } \boldsymbol{\mathcal{D}}^{(s)} \in \mathbb{R}_{+}^{R^{(s)}\times\cdots\times R^{(s)}},\ \boldsymbol{A}^{(n,s)} = [\boldsymbol{A}_C^{(n)}\ \boldsymbol{A}_I^{(n,s)}] \in \mathbb{R}_{+}^{I_n\times R^{(s)}}.$$

Note that sometimes for simplicity, the core tensors $\boldsymbol{\mathcal{D}}^{(s)}$ will be absorbed into the factor matrices $\boldsymbol{A}^{(n,s)}$. When $N = 2$, the CNTF model will degenerate to the CNMF model. More details about the generalized CNTF model can be found in the attached article PII.

2.3 Optimization methods

Using block coordinate decent (BCD) framework (Razaviyayn et al., 2013), the NMF/NTF problem can be expressed as several sub-problems, which can be alternatively solved by various optimization methods, such as multiplicative update (MU, (Lee and Seung, 1999)), alternating proximal gradient (APG, (Guan et al., 2012; Xu, 2015)), projected gradient (PG, (Lin, 2007; Zdunek and Cichocki, 2006)), active-set (or like) (AS, (Kim and Park, 2008, 2011)), (fast) hierarchical alternating least squares (HALS/fHALS, (Cichocki et al., 2007; Cichocki and Phan, 2009)), alternating direction method of multipliers (ADMM, (Boyd et al., 2011)) and series of their variants. Following the articles PI, PIII-PV, this section introduces several effective and efficient optimization methods to solve CNMF and CNTF problems under the BCD framework.

2.3.1 Alternating direction method of multipliers

Constrained joint analysis of data from multiple sources has received widespread attention. It allows us to explore potential connections and extract meaningful hidden components. Even though imposing nonnegative constraints can naturally bring the sparse representation in matrix factorization, in practical applications, the sparseness generated in this way is still insufficient and uncontrollable (Hoyer, 2004). This part introduces an efficient CNMF with sparse regularization (CNMF-SR) model optimized by the ADMM strategy. A comprehensive review of ADMM problems can be found in (Boyd et al., 2011).

Considering the introduction of l_1-norm regularizer on factor matrix $\boldsymbol{A}^{(s)}$, the objective function in Eq. 8 can be expressed as:

$$\min_{\boldsymbol{A}^{(s)},\boldsymbol{B}^{(s)}} \frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{X}^{(s)} - \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 + \sum_{s=1}^{S}\beta^{(s)}\sum_{r=1}^{R(s)}\left\|\boldsymbol{a}_r^{(s)}\right\|_1 \tag{11}$$

$$\text{s.t. } \boldsymbol{A}^{(s)} = \left[\boldsymbol{A}_C\ \boldsymbol{A}_I^{(s)}\right] \geq 0,\ \boldsymbol{B}^{(s)} \geq 0,$$

where $\beta^{(s)} \geq 0$ is a predefined penalty parameter and here we set $\beta^{(1)} = \cdots \beta^{(S)}$.

(1) When calculating the factor matrix $\boldsymbol{A}^{(s)}$, we need to merge all the matrices $\boldsymbol{X}^{(s)}, s = 1, \cdots, S$ since the common part $\boldsymbol{A}_C$ is shared by them. Via introducing the auxiliary variables $\tilde{\boldsymbol{A}}^{(s)}$, the augmented Lagrangian function of Eq. 11 can

be expressed as:

$$\begin{aligned}\mathcal{L}_\rho\big(\boldsymbol{A}^{(s)},\tilde{\boldsymbol{A}}^{(s)},\boldsymbol{\Lambda}^{(s)}\big) = \frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{X}^{(s)}-\boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 + \sum_{s=1}^{S}\beta^{(s)}\sum_{r=1}^{R(s)}\left\|\tilde{\boldsymbol{a}}_r^{(s)}\right\|_1 \\ + \sum_{s=1}^{S}\frac{\rho^{(s)}}{2}\left\|\boldsymbol{A}^{(s)}-\tilde{\boldsymbol{A}}^{(s)}+\boldsymbol{\Lambda}^{(s)}\right\|_F^2,\end{aligned} \tag{12}$$

where $\boldsymbol{\Lambda}^{(s)} \in \mathbb{R}^{I\times R^{(s)}}$ are the dual variables and $\rho^{(s)}$ are the predefined penalty parameters. Note that $\tilde{\boldsymbol{A}}^{(s)}$ also includes two parts as $\tilde{\boldsymbol{A}}^{(s)} = \left[\tilde{\boldsymbol{A}}_C\ \tilde{\boldsymbol{A}}_I^{(s)}\right]$.

(2) When calculating the factor matrix $\boldsymbol{B}^{(s)}$, we just need to use the corresponding sth matrix $\boldsymbol{X}^{(s)}$. Via introducing the auxiliary variables $\tilde{\boldsymbol{B}}^{(s)}$, the augmented Lagrangian function of Eq. 11 can be formulated as:

$$\mathcal{L}_\mu(\boldsymbol{B}^{(s)},\tilde{\boldsymbol{B}}^{(s)},\boldsymbol{\Gamma}^{(s)}) = \frac{1}{2}\left\|\boldsymbol{X}^{(s)}-\boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 + \frac{\mu^{(s)}}{2}\left\|\boldsymbol{B}^{(s)}-\tilde{\boldsymbol{B}}^{(s)}+\boldsymbol{\Gamma}^{(s)}\right\|_F^2, \tag{13}$$

where $\boldsymbol{\Gamma}^{(s)} \in \mathbb{R}^{R^{(s)}\times J^{(s)}}$ are the dual variables and $\mu^{(s)}$ are the predefined penalty parameters.

In the BCD framework, the factor matrices $\boldsymbol{A}^{(s)}$ and $\boldsymbol{B}^{(s)}$ can be updated alternatively using the ADMM strategy, followed by the updates of $\tilde{\boldsymbol{A}}^{(s)}$, $\tilde{\boldsymbol{B}}^{(s)}$, $\boldsymbol{\Lambda}^{(s)}$ and $\boldsymbol{\Gamma}^{(s)}$, which has been summarized in **Algorithm** 1.

Algorithm 1: CNMF-SR-ADMM algorithm

Input: $\boldsymbol{X}^{(s)}$, L, and $R^{(s)}$, $s = 1,2,\cdots,S$
Initialization: $\boldsymbol{A}^{(s)}$, $\boldsymbol{B}^{(s)}$, $\tilde{\boldsymbol{A}}^{(s)}$, $\tilde{\boldsymbol{B}}^{(s)}$, $\boldsymbol{\Lambda}^{(s)}$, $\boldsymbol{\Gamma}^{(s)}$, $s = 1,2,\cdots,S$
for $k = 1$ *until convergence* **do**
  Update $\{\boldsymbol{A}_C,\tilde{\boldsymbol{A}}_C\}$ and $\{\boldsymbol{A}_I^{(s)},\tilde{\boldsymbol{A}}_I^{(s)},\boldsymbol{\Lambda}^{(s)}\}_{s=1}^{S}$ by solving argmin $\mathcal{L}_\rho\big(\boldsymbol{A}^{(s)},\tilde{\boldsymbol{A}}^{(s)},\boldsymbol{\Lambda}^{(s)}\big)$ in Eq. 12
  Update $\{\boldsymbol{B}^{(s)},\tilde{\boldsymbol{B}}^{(s)},\boldsymbol{\Gamma}^{(s)}\}_{s=1}^{S}$ by solving argmin $\mathcal{L}_\mu(\boldsymbol{B}^{(s)},\tilde{\boldsymbol{B}}^{(s)},\boldsymbol{\Gamma}^{(s)})$ in Eq. 13
end
Output: $\boldsymbol{A}^{(s)}$, $\boldsymbol{B}^{(s)}$, $s = 1,2,\cdots,S$

2.3.2 Fast hierarchical alternative least squares

In order to achieve group tensor analysis, Yokota et al. proposed a flexible CNTF model, namely linked CP tensor decomposition (LCPTD), in which multiple tensors with coupling information can be simultaneously decomposed into common factor matrices, individual factor matrices and core tensors (Yokota et al., 2012). However, this model is not suitable for linked tensors with inconsistent component numbers. Besides, hierarchical alternative least squares (HALS, (Cichocki et al., 2007)) algorithm can result in rather high computational cost for large-scale problems (Wang et al., 2019a,b). In this section, the fast hierarchical alternative

least squares (fHALS, (Cichocki and Phan, 2009)) algorithm is further introduced to solve the CNTF problem.

First, using the HALS algorithm, the minimization optimization problem in Eq. 10 can be converted into $\max(R^{(s)})$ rank-1 tensor approximation sub-problems, which can be solved sequentially and iteratively as follows:

$$\begin{aligned} \min \mathcal{F}(\boldsymbol{u}_r^{(n,s)}) &= \sum_{s=1}^{S} \left\| \boldsymbol{\mathcal{Y}}_r^{(s)} - d_r^{(s)} \boldsymbol{a}_r^{(1,s)} \circ \boldsymbol{a}_r^{(2,s)} \circ \cdots \circ \boldsymbol{a}_r^{(N,s)} \right\|_F^2 \\ &= \sum_{s=1}^{S} \left\| \boldsymbol{Y}_{r,(n)}^{(s)} - \boldsymbol{a}_r^{(n,s)} d_r^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}^T} \right\|_F^2, \end{aligned} \tag{14}$$

where $\boldsymbol{\mathcal{Y}}_r^{(s)} \doteq \boldsymbol{\mathcal{X}}^{(s)} - \sum_{k \neq r}^{R^{(s)}} d_r^{(s)} \boldsymbol{a}_k^{(1,s)} \circ \boldsymbol{a}_k^{(2,s)} \circ \cdots \circ \boldsymbol{a}_k^{(N,s)}$ and $\boldsymbol{Y}_{r,(n)}^{(s)}$ is the mode-n matricization of $\boldsymbol{\mathcal{Y}}_r^{(s)}$. Let the gradient of $\mathcal{F}(\boldsymbol{a}_r^{(n,s)})$ with respect to $\boldsymbol{a}_r^{(n,s)}$ be zero, then the solution of $\boldsymbol{a}_r^{(n,s)}$ can be calculated as:

$$\boldsymbol{a}_r^{(n,s)} = \begin{cases} \sum\limits_{s=1}^{S} \left[d_r^{(s)} \boldsymbol{Y}_{r,(n)}^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}} \right] \Big/ \sum\limits_{s=1}^{S} \left[d_r^{(s)T} d_r^{(s)} \{\boldsymbol{a}_r^{(s)T} \boldsymbol{a}_r^{(s)}\}^{\circledast_{-n}} \right], & r \leq L_n, \\ \left[\boldsymbol{Y}_{r,(n)}^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}} \right] \Big/ \left[d_r^{(s)} \{\boldsymbol{a}_r^{(s)T} \boldsymbol{a}_r^{(s)}\}^{\circledast_{-n}} \right], & r > L_n, \end{cases} \tag{15}$$

Second, note that $\boldsymbol{Y}_{r,(n)}^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}}$ needs to calculate mode-n matricization and Khatri-Rao product in each loop and thus results in rather high computation cost, then the fHALS algorithm is adopted to address such a problem. Substitute the mode-n matricization $\boldsymbol{Y}_{r,(n)}^{(s)}$ below

$$\boldsymbol{Y}_{r,(n)}^{(s)} = \boldsymbol{X}_{(n)}^{(s)} - \boldsymbol{A}^{(n,s)} \boldsymbol{D}^{(s)} \left\{ \boldsymbol{A}^{(s)} \right\}^{\odot_{-n}^T} + \boldsymbol{a}_r^{(n,s)} d_r^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}^T} \tag{16}$$

into $\boldsymbol{Y}_{r,(n)}^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}}$ in Eq. 15, and let $\boldsymbol{\zeta}_r^{(n,s)} \doteq d_r^{(s)} \boldsymbol{Y}_{r,(n)}^{(s)} \{\boldsymbol{a}_r^{(s)}\}^{\odot_{-n}}$, then we can get

$$\boldsymbol{\zeta}_r^{(n,s)} = \left[\boldsymbol{X}_{(n)}^{(s)} \boldsymbol{D}^{(s)} \left\{ \boldsymbol{A}^{(s)} \right\}^{\odot_{-n}} \right]_r - \boldsymbol{A}^{(n,s)} \left[\boldsymbol{\gamma}^{(n,s)} \right]_r + \boldsymbol{a}_r^{(n,s)} \left[\boldsymbol{\gamma}^{(n,s)} \right]_{(r,r)}, \tag{17}$$

where $\boldsymbol{\gamma}^{(n,s)} = \boldsymbol{D}^{(s)T} \boldsymbol{D}^{(s)} \left\{ \boldsymbol{A}^{(s)T} \boldsymbol{A}^{(s)} \right\}^{\circledast_{-n}}$. Therefore, the learning rule of $\boldsymbol{a}_r^{(n,s)}$ can be represented as:

$$\boldsymbol{u}_r^{(n,s)} = \begin{cases} \left[\sum\limits_{s=1}^{S} \boldsymbol{\zeta}_r^{(n,s)} \Big/ \sum\limits_{s=1}^{S} \left[\boldsymbol{\gamma}^{(n,s)} \right]_{(r,r)} \right]_+, & r \leq L_n, \\ \left[\boldsymbol{\zeta}_r^{(n,s)} / \left[\boldsymbol{\gamma}^{(n,s)} \right]_{(r,r)} \right]_+, & r > L_n, \end{cases} \tag{18}$$

where the mode-n matricization $\boldsymbol{X}_n^{(s)}$ in $\boldsymbol{\zeta}_r^{(n,s)}$ only needs to be performed once in initialization, which greatly improves the computation efficiency.

Third, for the solution of core tensor $\boldsymbol{\mathcal{D}}^{(s)}$, the minimization optimization problem in Eq. 10 can be converted into

$$\min \mathcal{F}(\boldsymbol{\mathcal{D}}^{(s)}) = \frac{1}{2}\left\|\mathrm{vec}(\boldsymbol{\mathcal{X}}^{(s)}) - A^{(s)\odot}\mathrm{ddiag}(\boldsymbol{\mathcal{D}}^{(s)})\right\|_F^2 \tag{19}$$

where $\mathrm{vec}(\boldsymbol{\mathcal{X}}^{(s)})$ denotes the vectorization of tensor $\boldsymbol{\mathcal{X}}^{(s)}$. Let the gradient of $\mathcal{F}(\boldsymbol{\mathcal{D}}^{(s)})$ with respect to $\mathrm{ddiag}(\boldsymbol{\mathcal{D}}^{(s)})$ to be zero, then the learning rule of $\boldsymbol{\mathcal{D}}^{(s)}$ can be represented as:

$$\boldsymbol{\mathcal{D}}^{(s)} = \mathrm{ddiag}\left(\left[\left(A^{(s)\odot}\right)^{-1}\mathrm{vec}\left(\boldsymbol{\mathcal{X}}^{(s)}\right)\right]_+\right) \tag{20}$$

where 'ddiag' means the tensorization from a vector to a super-diagonal tensor. To avoid the problem of rank deficiency, a small value such as eps $= 10e-16$ is usually added to the denominators of Eq. 18. The summary of fHALS-based CNTF algorithm is given in **Algorithm** 2.

Algorithm 2: CNTF-fHALS algorithm

Input: $\boldsymbol{\mathcal{X}}^{(s)}$, L_n, and $R^{(s)}$, $n = 1, 2, \cdots, N$, $s = 1, 2, \cdots, S$
Initialization: $A^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $\boldsymbol{X}_n^{(s)}$, $n = 1, 2, \cdots, N$, $s = 1, 2, \cdots, S$
for $k = 1$ *until convergence* **do**
  for $n = 1, 2, \cdots, N$ **do**
    for $r = 1, 2, \cdots, max(R^{(s)})$ **do**
      if $r \leq R^{(s)}$ **then**
        Update $\boldsymbol{a}_r^{(n,s)}$ via Eq. 18, $s = 1, 2, \cdots, S$
      end
    end
  end
  Update $\boldsymbol{\mathcal{D}}^{(s)}$ via Eq. 20, $s = 1, 2, \cdots, S$
end
Output: $A^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $n = 1, 2, \cdots, N$, $s = 1, 2, \cdots, S$

2.3.3 Alternating proximal gradient and low-rank approximation

To date, increasing recognition of joint tensor analysis has enabled coupled tensor decomposition to be used in many applications. However, due to the non-negative constraint and high-dimensional nature of tensor data, existing coupled methods often suffer from slow convergence speed and low optimization accuracy (Zhou et al., 2012; Zhang et al., 2016). APG originally proposed by Nesterov is used for smooth optimization with the convergence rate of $\mathcal{O}(\frac{1}{K^2})$ (K: the number of iterations) (Nesterov, 1983; Beck and Teboulle, 2009), and has proven to be very efficient for NMF/NTF problems (Guan et al., 2012; Xu and Yin, 2013; Xu, 2015; Zhang et al., 2016; Wang et al., 2018). In this section, the APG method and

a combination of APG method and low-rank approximation are respectively applied to solve the CNTF problem. According to the BCD framework, the CNTF problem can be converted into several sub-problems by optimizing $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{A}^{(n,s)}$ alternatively in each iteration. Each sub-problem can be regarded as a minimization problem of a continuously differentiable function, which can be solved efficiently by the APG method (Guan et al., 2012; Xu and Yin, 2013; Xu, 2015).

First, for the solution of $\boldsymbol{A}^{(n,s)}$, using the APG method, we have

$$\boldsymbol{A}^{(n,s)} = \underset{\boldsymbol{A}^{(n,s)} \geq 0}{\operatorname{argmin}} \sum_{s=1}^{S} \left[\left\langle \hat{\boldsymbol{G}}^{(n,s)}, \boldsymbol{A}^{(n,s)} - \hat{\boldsymbol{A}}^{(n,s)} \right\rangle + \frac{L_a^{(n,s)}}{2} \left\| \boldsymbol{A}^{(n,s)} - \hat{\boldsymbol{A}}^{(n,s)} \right\|_F^2 \right] \tag{21}$$

where $\hat{\boldsymbol{A}}^{(n,s)}$ denotes the extrapolated point of $\boldsymbol{A}^{(n,s)}$, and $L_a^{(n,s)}$ denotes the Lipschitz constant defined as $L_a^{(n,s)} = \left\| \boldsymbol{D}^{(s)} (\boldsymbol{A}^{(s)\odot -n})^T \boldsymbol{A}^{(s)\odot -n} \boldsymbol{D}^{(s)} \right\|$. $\hat{\boldsymbol{G}}^{(n,s)}$ denotes the block-partial gradient of Eq. 10 at $\hat{\boldsymbol{A}}^{(n,s)}$ calculated as:

$$\hat{\boldsymbol{G}}^{(n,s)} = \hat{\boldsymbol{A}}^{(n,s)} \boldsymbol{D}^{(s)} \left(\boldsymbol{A}^{(s)^T} \boldsymbol{A}^{(s)} \right)^{\circledast -n} \boldsymbol{D}^{(s)} - \boldsymbol{X}_{(n)}^{(s)} \boldsymbol{A}^{(s)^{\odot -n}} \boldsymbol{D}^{(s)} \tag{22}$$

where $\boldsymbol{D}^{(s)}$ is a diagonal matrix and its diagonal elements correspond to the super-diagonal elements of core tensor $\boldsymbol{\mathcal{D}}^{(s)}$. The updating rules of $\boldsymbol{A}_C^{(n)}$ and $\boldsymbol{A}_I^{(n,s)}$ can be written in the closed form as:

$$\boldsymbol{A}_C^{(n)} = \max \left(0,\ \hat{\boldsymbol{A}}_C^{(n)} - \frac{\sum_{s=1}^{S} \hat{\boldsymbol{G}}_C^{(n,s)}}{\sum_{s=1}^{S} L_a^{(n,s)}} \right), \boldsymbol{A}_I^{(n,s)} = \max \left(0,\ \hat{\boldsymbol{A}}_I^{(n,s)} - \frac{\hat{\boldsymbol{G}}_I^{(n,s)}}{L_a^{(n,s)}} \right) \tag{23}$$

where $\hat{\boldsymbol{A}}^{(n,s)} = \left[\hat{\boldsymbol{A}}_C^{(n,s)}\ \hat{\boldsymbol{A}}_I^{(n,s)} \right]$ and $\hat{\boldsymbol{G}}^{(n,s)} = \left[\hat{\boldsymbol{G}}_C^{(n,s)}\ \hat{\boldsymbol{G}}_I^{(n,s)} \right]$.

Second, for the solution of $\boldsymbol{\mathcal{D}}^{(s)}$, similarly, we derive

$$\boldsymbol{\mathcal{D}}^{(s)} = \underset{\boldsymbol{\mathcal{D}}^{(s)} \geq 0}{\operatorname{argmin}} \left[\left\langle \hat{\boldsymbol{\mathcal{G}}}^{(s)}, \boldsymbol{\mathcal{D}}^{(s)} - \hat{\boldsymbol{\mathcal{D}}}^{(s)} \right\rangle + \frac{L_d^{(s)}}{2} \left\| \boldsymbol{\mathcal{D}}^{(s)} - \hat{\boldsymbol{\mathcal{D}}}^{(s)} \right\|_F^2 \right] \tag{24}$$

which can also be written in the closed form as

$$\boldsymbol{\mathcal{D}}^{(s)} = \max \left(0,\ \hat{\boldsymbol{\mathcal{D}}}^{(s)} - \frac{\hat{\boldsymbol{\mathcal{G}}}^{(s)}}{L_d^{(s)}} \right) \tag{25}$$

where $\hat{\boldsymbol{\mathcal{D}}}^{(s)}$ denotes the extrapolated point and $L_d^{(s)}$ denotes the Lipschitz constant defined as $L_d^{(s)} = \left\| (\boldsymbol{A}^{(s)\odot})^T \boldsymbol{A}^{(s)\odot} \right\|$. $\hat{\boldsymbol{\mathcal{G}}}^{(s)}$ is the block-partial gradient of Eq. 10 at $\hat{\boldsymbol{\mathcal{D}}}^{(s)}$, which can be calculated as:

$$\hat{\boldsymbol{\mathcal{G}}}^{(s)} = \operatorname{ddiag} \left[\left(\boldsymbol{A}^{(s)^T} \boldsymbol{A}^{(s)} \right)^{\circledast} \operatorname{ddiag} \left(\hat{\boldsymbol{\mathcal{D}}}^{(s)} \right) - \left(\boldsymbol{A}^{(s)^{\odot}} \right)^T \operatorname{vec} \left(\boldsymbol{\mathcal{X}}^{(s)} \right) \right] \tag{26}$$

where $\operatorname{ddiag}(\boldsymbol{\mathcal{D}}^{(s)})$ denotes a vector vectorized from the super-diagonal elements of $\boldsymbol{\mathcal{D}}^{(s)}$, and the outer-loop notation $\operatorname{ddiag}(\cdot)$ means the tensorization from a vector to a super-diagonal tensor, which is the reverse operation of the inner-loop $\operatorname{ddiag}(\cdot)$.

Third, considering the update of $\boldsymbol{A}^{(n,s)}$ and $\boldsymbol{\mathcal{D}}^{(s)}$ at the kth iteration, the extrapolated points $\hat{\boldsymbol{\mathcal{D}}}_k^{(s)}$ and $\hat{\boldsymbol{A}}_k^{(n,s)}$ are defined as

$$\hat{\boldsymbol{\mathcal{D}}}_k^{(s)} = \boldsymbol{\mathcal{D}}_k^{(s)} + w_{d,k}^{(s)} \left(\boldsymbol{\mathcal{D}}_k^{(s)} - \boldsymbol{\mathcal{D}}_{k-1}^{(s)} \right) \tag{27}$$

and

$$\hat{\boldsymbol{A}}_k^{(n,s)} = \boldsymbol{A}_k^{(n,s)} + w_{u,k}^{(n,s)} \left(\boldsymbol{A}_k^{(n,s)} - \boldsymbol{A}_{k-1}^{(n,s)} \right) \tag{28}$$

where $w_{d,k}^{(s)}$ and $w_{u,k}^{(n,s)}$ denote the extrapolation weights as

$$w_{d,k}^{(s)} = \min \left(\hat{w}_k,\ \delta_w \sqrt{\frac{L_{d,k-1}^{(s)}}{L_{d,k}^{(s)}}} \right), w_{u,k}^{(n,s)} = \min \left(\hat{w}_k,\ \delta_w \sqrt{\frac{L_{u,k-1}^{(n,s)}}{L_{u,k}^{(n,s)}}} \right) \tag{29}$$

where $\delta_w < 1$ is predefined (e.g., 0.9999, (Xu, 2015)), and $\hat{w}_k = \frac{t_k - 1}{t_{k+1}}$ with $t_0 = 1$ and $t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$.

In each iteration, $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{A}^{(n,s)}$ are updated alternatively one after another until convergence, and we define it as the CNTF-APG algorithm. The main time cost is attributed to the multiplication of $\boldsymbol{X}_{(n)}^{(s)} \boldsymbol{A}^{(s)^{\odot -n}}$ in Eq. 22 and $(\boldsymbol{A}^{(s)^{\odot}})^T \mathrm{vec}(\boldsymbol{\mathcal{X}}^{(s)})$ in Eq. 26. The cost will become very serious, especially for large-scale tensor problems. In NMF/NTF problems, low-rank approximation has proven its excellent performance not only in computational efficiency but also in computational accuracy (Zhou et al., 2012; Cong et al., 2014; Zhang et al., 2016). Therefore, it will be further introduced into the CNTF problem. Suppose that $\left[\!\left[\tilde{\boldsymbol{A}}^{(1,s)}, \tilde{\boldsymbol{A}}^{(2,s)}, \cdots, \tilde{\boldsymbol{A}}^{(N,s)} \right]\!\right]$ is the rank-$\tilde{R}^{(s)}$ approximation of $\boldsymbol{\mathcal{X}}^{(s)}$ obtained by an unconstrained tensor decomposition, $\tilde{\boldsymbol{A}}^{(n,s)} \in \mathbb{R}^{I_n \times \tilde{R}^{(s)}}$, $\tilde{R}^{(s)} \leq R^{(s)}$. Thus $\mathrm{vec}(\boldsymbol{\mathcal{X}}^{(s)})$ and $\boldsymbol{X}_{(n)}^{(s)}$ can be respectively expressed as $\mathrm{vec}(\boldsymbol{\mathcal{X}}^{(s)}) = \tilde{\boldsymbol{A}}^{(s)^{\odot}} \mathrm{ddiag}(\boldsymbol{\mathcal{I}})$ and $\boldsymbol{X}_{(n)}^{(s)} = \tilde{\boldsymbol{A}}^{(n,s)} \left(\tilde{\boldsymbol{A}}^{(s)^{\odot -n}} \right)^T$. $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{\tilde{R}^{(s)} \times \cdots \times \tilde{R}^{(s)}}$ is a core tensor with all super-diagonal elements being 1. This thereby leads to

$$\boldsymbol{X}_{(n)}^{(s)} \boldsymbol{A}^{(s)^{\odot -n}} = \tilde{\boldsymbol{A}}^{(s,n)} \left(\tilde{\boldsymbol{A}}^{(s)^T} \boldsymbol{A}^{(s)} \right)^{\circledast -n}. \tag{30}$$

and

$$\left(\boldsymbol{A}^{(s)^{\odot}} \right)^T \mathrm{vec}\left(\boldsymbol{\mathcal{X}}^{(s)} \right) = \left(\boldsymbol{A}^{(s)^T} \tilde{\boldsymbol{A}}^{(s)} \right)^{\circledast} \mathrm{ddiag}(\boldsymbol{\mathcal{I}}) \tag{31}$$

By virtue of low-rank approximation, only very small matrices are involved to calculate $\boldsymbol{X}_{(n)}^{(s)} \boldsymbol{A}^{(s)^{\odot -n}}$ and $(\boldsymbol{A}^{(s)^{\odot}})^T \mathrm{vec}(\boldsymbol{\mathcal{X}}^{(s)})$. The computational complexities are reduced from $\mathcal{O}(NSR \prod_n I_n)$ and $\mathcal{O}(SR \prod_n I_n)$ to $\mathcal{O}(NSR^2 \sum_n I_n)$ and $\mathcal{O}(SR^2 \sum_n I_n)$ respectively via the substitutions in Eq. 30 and Eq. 31 (here we set $\tilde{R}^{(s)} = R^{(s)} = R$). The CNTF algorithm based on APG and low-rank approximation is termed as the lraCNTF-APG and summarized in **Algorithm** 3.

Algorithm 3: lraCNTF-APG algorithm

Input: $\boldsymbol{\mathcal{X}}^{(s)}$, L_n and $R^{(s)}$, $n = 1, 2, \cdots, N$, $s = 1, 2, \cdots, S$

Initialization: $\boldsymbol{A}^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $\boldsymbol{X}^{(s)}_{(n)}$, $n = 1, 2, \cdots, N$, $s = 1, 2, \cdots, S$

Calculate $\tilde{\boldsymbol{A}}^{(n,s)}$, $n = 1, 2, \cdots, N$, $s = 1, 2, \cdots, S$ via low-rank approximation of $\boldsymbol{\mathcal{X}}^{(s)}$, $s = 1, 2, \cdots, S$

for $k = 1$ *until convergence* **do**

  for $n = 1, 2, \cdots, N$ **do**

    Update $\boldsymbol{A}^{(n,s)}_k$ via Eq. 23 and Eq. 30, $s = 1, 2, \cdots, S$

  end

  Update $\boldsymbol{\mathcal{D}}^{(s)}_k$ via Eq. 25 and Eq. 31, $s = 1, 2, \cdots, S$

  if $F(k) \geq F(k-1)$ **then**

    $\hat{\boldsymbol{A}}^{(n,s)}_{k-1} = \boldsymbol{A}^{(n,s)}_{k-1}$, $\hat{\boldsymbol{\mathcal{D}}}^{(s)}_{k-1} = \boldsymbol{\mathcal{D}}^{(s)}_{k-1}$

    Reupdate $\boldsymbol{A}^{(n,s)}_k$ and $\boldsymbol{\mathcal{D}}^{(s)}_k$ via Eq. 23, Eq. 25, Eq. 30 and Eq. 31

  end

end

Output: $\boldsymbol{A}^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$

2.4 Experiments and results

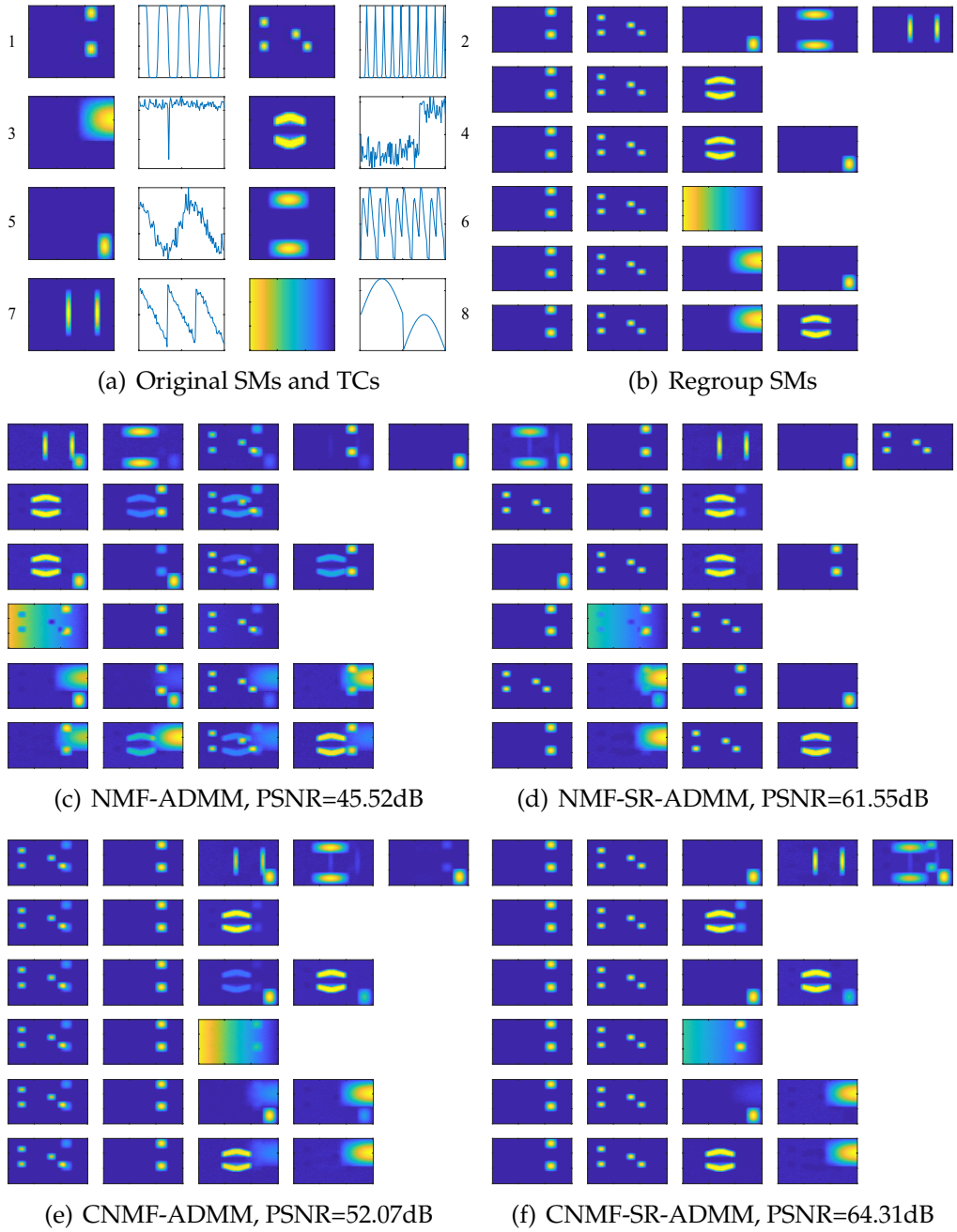
In this chapter, following the articles PI, PIV and PV, several experiments and their results are presented to demonstrate the performances of the proposed coupled nonnegative matrix/tensor factorization algorithms.

2.4.1 Synthetic multi-subject fMRI data

In this section, NMF and CNMF models with or without sparse regularization optimized by the ADMM are used in the joint analysis of multi-subject nonnegative fMRI-like data. The data are reconstructed from the benchmark simulated complex fMRI dataset [1]. Spatial maps (SM) and corresponding time courses (TC) are shown in Figure 6(a). They are adopted to generate the nonnegative fMRI-like data for 6 subjects according to the source index sets {1,2,5,6,7}, {1,2,4}, {1,2,4,5}, {1,2,8}, {1,2,3,5} and {1,2,3,4}. More information about data construction can be found in (Gong et al., 2018b). The SM images of all subjects are shown in Figure 6(b). Each row corresponds to one subject. The first two columns are shared by all the subjects, which are considered as the common components, and the remains are the individual ones.

The sparse parameters βs are selected from 0 to 5 with a total of 25 values. With varying βs, we calculate the peak signal-to-noise ratio (PSNR) of SM estimates under SNR=20dB in the NMF-SR and CNMF-SR models ($L = 0$ & $L = 2$), and the PSNR curves averaged from 30 Monte Carlo runs are shown in Figure 7.

[1] http://mlsp.umbc.edu/simulated_complex_fmri_data.html



(a) Original SMs and TCs

(b) Regroup SMs

(c) NMF-ADMM, PSNR=45.52dB

(d) NMF-SR-ADMM, PSNR=61.55dB

(e) CNMF-ADMM, PSNR=52.07dB

(f) CNMF-SR-ADMM, PSNR=64.31dB

FIGURE 6 (a) Amplitude images of 1-8 simulated fMRI-like spatial maps (1st and 3rd columns) and corresponding time courses (2nd and 4th columns). (b-f) SM images of constructed data and that of estimated ones via NMF-ADMM ($\beta = 0$, $L = 0$), NMF-SR-ADMM ($\beta = 3e-4$, $L = 0$), CNMF-ADMM ($\beta = 0$, $L = 2$) and CNMF-SR-ADMM ($\beta = 3e-4$, $L = 2$) under SNR=20dB.

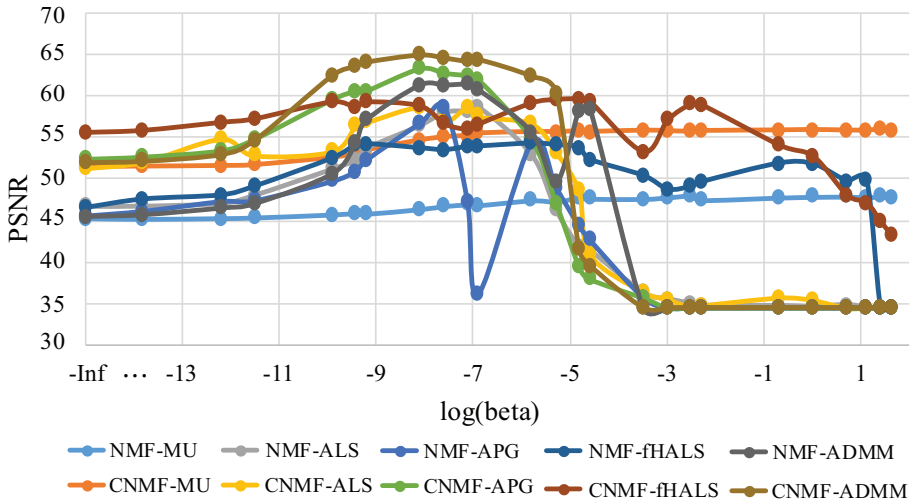


FIGURE 7 Mean PSNR of SM estimates for 6 subjects under NMF-SR ($L = 0$) and CNMF-SR ($L = 2$) models with the βs of 25 values varying from 0 to 5, SNR=20dB.

Note that in this experiment, in addition to CNMF-SR model, three other models are also considered: (1) when $\beta = 0$, CNMF-SR will degenerate into the CNMF problem; (2) when $L = 0$, CNMF-SR will degenerate into the NMF-SR problem; (3) when $L = 0$ and $\beta = 0$, CNMF-SR will degenerate into the NMF problem. From Figure 7, it can be seen that the PSNR values of all algorithms will increase and reach the highest value at some points when the sparse penalty parameter β increases, which proves that sparseness can improve the algorithm performance to a certain extent. The performances of CNMF-based methods are superior to that of NMF-based ones.

Furthermore, the SM images estimated via NMF-ADMM, NMF-SR-ADMM, CNMF-ADMM and CNMF-SR-ADMM at $\beta = 0,\ 3e-4$ and $L = 0,\ 2$ are shown in Figure 6(c)-6(f). It can be clearly seen that some of SM images obtained by NMF-ADMM and CNMF-ADMM algorithms are blurred with shadows or small outliers. With imposing adequate sparse regularization, those blurs are basically eliminated in the results of NMF-SR-ADMM and CNMF-SR-ADMM algorithms. Moreover, in Figure 6(e)-6(f), CNMF-based algorithms can extract both common and individual patterns for all the datasets, and also successfully correct the disorder problem of common patterns in the results of two NMF-based algorithms as shown in Figure 6(c)-6(d). For a more detailed description of experimental design and analysis of results, please refer to the attached article PI.

2.4.2 Real-world multi-subject ongoing EEG data

Ongoing EEG data are recorded as mixtures of stimulus-elicited EEG, spontaneous EEG and noises, which require advanced signal processing techniques for separation and analysis. Existing methods did not simultaneously consider common and individual characteristics among/within subjects when extracting stimulus-elicited brain activities from ongoing EEG. In this section, the CNTF

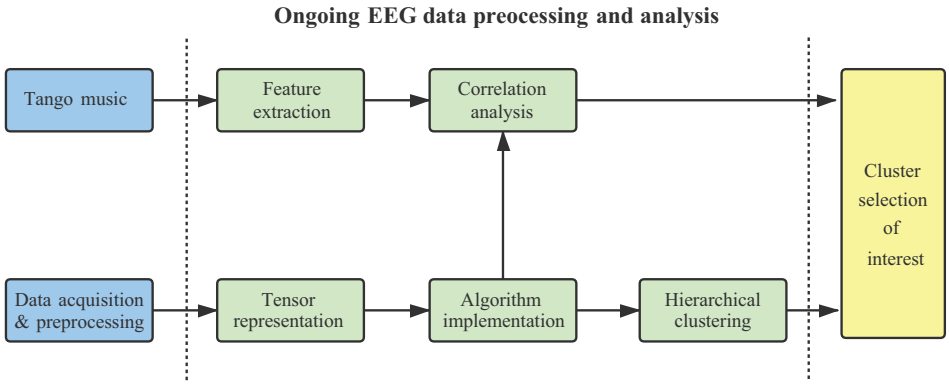


FIGURE 8 Flow chart of multi-subject ongoing EEG data processing and analysis: data acquisition & preprocessing, musical feature extraction, tensor representation, coupled factorization, correlation analysis, hierarchical clustering and cluster selection of interest.

algorithm based on the fHALS strategy (see it in Section 2.3.2) is adopted to the joint analysis of multi-subject ongoing EEG data, with imposing a partially double-coupled constraint on spatial and spectral modes. Figure 8 illustrates the comprehensive flowchart of ongoing EEG data processing and analysis, which includes the following seven steps: (1) the data were collected from 14 participants when listening to an 8.5-minute piece of modern tango (Alluri et al., 2012), and were preprocessed off-line using EEGLAB toolbox (Delorme and Makeig, 2004) and MATLAB R2016b; (2) through short-time Fourier transform (STFT), 14 third-order tensors with the size of 46 (frequency bins) $\times$ 510 (time samples) $\times$ 64 (space channels) were generated for 14 participants; (3) five long-term tonal and rhythmic musical features with the length of 510 samples were extracted from the musical stimulus, including Mode, Key Clarity, Fluctuation Centroid, Fluctuation Entropy and Pulse Clarity, which provided a bridge for analyzing the connections between musical stimulus and ongoing EEG data; (4) through CNTF-fHALS algorithm, the ongoing EEG tensors of space, time and frequency were simultaneously decomposed into common and individual components; (5) correlation analysis was performed between temporal courses of extracted temporal components and temporal courses of musical features to discover the brain activities elicited by the musical stimulus, and the significantly correlated temporal components and their corresponding spatial and spectral components will be of interest and further analyzed; (6) hierarchical clustering was adopted to merge the highly correlated spatial components (selected from step 5) within the subjects, and finally, q clusters of spatial components were obtained; (7) the cluster with more than half of the total number of subjects was selected as the cluster of interest, in which the corresponding brain activities were considered to be elicited by the musical stimulus.

Through the proposed framework, the brain activities were effectively extracted and sorted into three clusters of interest. The averaged topographies and their corresponding spectrograms of clusters #I, #II and #III are illustrated in Figure 9. Regarding cluster #I, the topography reveals that brain's centro-parietal

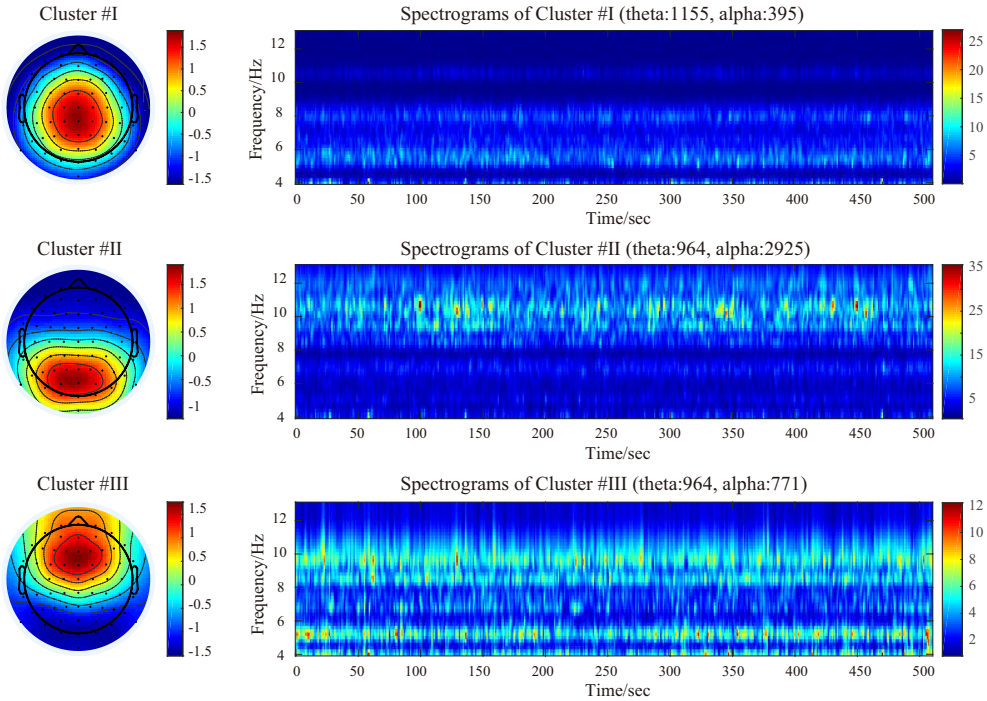


FIGURE 9 Illustration of clusters of interest #I, #II and #III obtained from 100 runs of CNTF-fHALS algorithm. Spatial information, presented by the averaged topographies (left column), indicate the activation of centro-parietal, occipito-parietal and frontal regions of the brain elicited by musical stimulus, respectively. Overall spectrogram of clusters #I, #II and #III (right column) from 100 runs illustrate the frequency oscillations over the entire period.

region is activated with quite a lot theta oscillations (4∼7 Hz, 74.52%, 1155/1550) but little alpha oscillations (around 10 Hz, 25.48%, 395/1550). Conversely, the occipito-parietal region of the brain is activated with significant alpha oscillations (8∼13 Hz, 75.21%, 2925/3889), accompanied by a small amount of theta oscillations (4∼8 Hz, 24.79%, 964/3889) in cluster #II. In addition, the topography representing the activation of the frontal region is also obtained, as shown in cluster #III, and the frequency oscillations are distributed in the range of 4 to 11 Hz (theta-55.56%, 964/1735, alpha-44.44%, 771/1735). The findings are in line with the results of previous studies, and it can be inferred that those extracted brain activities are associated with the musical stimulus. Moreover, the framework based on coupled tensor factorization provides a new perspective for the processing and analysis of multi-subject ongoing EEG data. For a more detailed description of the experimental design and analysis of results, please refer to the attached article PIV.

2.4.3 Real-world multi-subject ERP data

This section applies CNTF-APG and lraCNTF-APG algorithms to the joint analysis of multi-subject event-related potential (ERP) data[2]. The data of 21 children

[2] http://www.escience.cn/people/cong/AdvancedSP_ERP.html

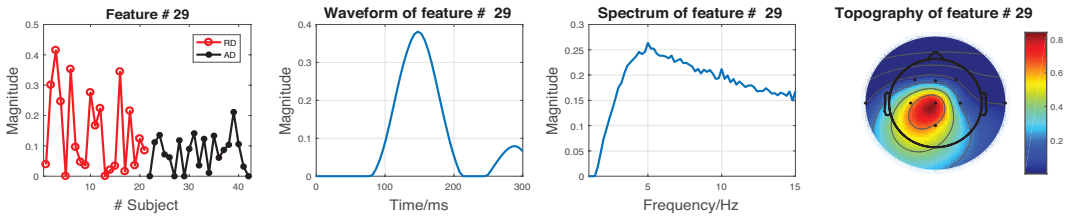


FIGURE 10 An example of multi-domain feature extracted from multi-subject ERP data and its corresponding temporal, spectral and spatial components through CNTF-APG algorithm.

with reading disability (RD) and 21 children with attention deficit (AD) are chosen, aiming to acquire multi-domain features that can discriminate between two groups. Using complex Morlet wavelet transform, third-order tensors of 42 subjects (21 RD & 21 AD) with the size of $9 \times 71 \times 60$ are generated, presenting nine channels, 71 frequency bins and 60 temporal points, to testify the effectiveness and practicality of coupled tensor factorization. Following (Cong et al., 2012a), the number of components is set as $R^{(1)} = R^{(2)} = \cdots R^{(42)} = 36$. Considering the nature of ERP data, 42 third-order ERP tensors are assumed to be coupled in spatial, spectral and temporal modes. The number of coupled components between tensors is set to 36.

ERP data are acquired through repeated presentation of stimuli, so certain properties of temporal, spectral and spatial domains can be expected in advance. According to prior knowledge given in (Cong et al., 2012a), the expected multi-domain features and its corresponding temporal, spectral and spatial components can be selected from the decomposition results of ERP data. Figure 10 shows an example illustration of multi-domain features and their corresponding components extracted by the CNTF-APG algorithm in the 1st run. For multi-domain features shown in the figure, statistical analysis using t-test reveals the significant difference between RD and AD groups with $t_{20} = 2.419, p = 0.025$. The temporal component (latency peaks around 150 ms) and spectral component (spectrum peaks around 5 Hz) closely match the property of mismatch negativity component (Cong et al., 2012a). The corresponding topography denotes that the difference between RD and AD groups may appear in the central and left hemispheres (Cong et al., 2012a).

To verify the stability of multi-domain feature extraction of CNTF-APG and lraCNTF-APG algorithms in 100 Monte Carlo experiments, we adopt the following steps: (1) the multi-domain features and their parallel three components in the 1st runs of all algorithms are selected and averaged separately as a set of template patterns, and termed as $\boldsymbol{a}_{\text{temp}}^{\text{fea}}$, $\boldsymbol{a}_{\text{temp}}^{\text{tem}}$, $\boldsymbol{a}_{\text{temp}}^{\text{spe}}$ and $\boldsymbol{a}_{\text{temp}}^{\text{spa}}$; (2) the maximum correlation coefficient (MCC) between template patterns and feature-based components of kth runs is defined as follows

$$\begin{aligned}\text{MCC}(k) = \max \Big[&\text{corr}(\boldsymbol{a}_{\text{temp}}^{\text{fea}}, \boldsymbol{A}_k^{\text{fea}}) \circledast \text{corr}(\boldsymbol{a}_{\text{temp}}^{\text{tem}}, \boldsymbol{A}_k^{\text{tem}}) \\ &\circledast \text{corr}(\boldsymbol{a}_{\text{temp}}^{\text{spe}}, \boldsymbol{A}_k^{\text{spe}}) \circledast \text{corr}(\boldsymbol{a}_{\text{temp}}^{\text{spa}}, \boldsymbol{A}_k^{\text{spa}})\Big]\end{aligned} \tag{32}$$

where k denotes the run number and `corr` is a matlab function which returns

TABLE 1 Performance illustration in multi-domain feature extraction of multi-subject ERP data based on CNTF model under 100 independent runs.

Methods	TenFit	Time	MCC-Mean	MCC-SD
CNTF-APG	0.8490	183.91	0.8898	0.0956
lraCNTF-APG	0.8490	27.45	0.8916	0.0809

a vector containing the pairwise linear correlation coefficient between $\boldsymbol{a}$ and $\boldsymbol{A}$. $\boldsymbol{A}_k^{\text{fea}}$, $\boldsymbol{A}_k^{\text{tem}}$, $\boldsymbol{A}_k^{\text{spe}}$ and $\boldsymbol{A}_k^{\text{spa}}$ represent the multi-domain features and their corresponding temporal, spectral and spatial components in the kth run, respectively. Table 1 gives the means and SDs of MCCs in 100 independent runs of two algorithms, as well as the averages of tensor fittings and running time. It can be seen that both algorithms show very good performance in terms of decomposition accuracy and multi-domain feature extraction stability. This experiment also proves that low-rank approximation can significantly improve computation efficiency without losing the decomposition accuracy. For a more detailed description of the experimental design and analysis of results, please refer to the attached article PV.

3 INCLUDED ARTICLES AND AUTHOR CONTRIBUTIONS

This chapter presents an overview of the methods and main results involved in the included articles, and provides the author contributions to the articles.

3.1 Article I: "Group Nonnegative Matrix Factorization with Sparse Regularization in Multi-set Data"

Xiulin Wang, Wenya Liu, Fengyu Cong, and Tapani Ristaniemi. Group Nonnegative Matrix Factorization with Sparse Regularization in Multi-set Data. 28th European Signal Processing Conference (EUSIPCO), Amsterdam, NL, 2020. Accepted

Method&Results: Constrained joint analysis of data from multiple sources has received widespread attention for that it allows us to explore potential connections and extract meaningful hidden components. This article formulated a flexible joint source separation model, and termed it as group nonnegative matrix factorization with sparse regularization (GNMF-SR) model. Then the authors combined alternating optimization (AO) and alternating direction method of multipliers (ADMM) to solve the proposed model, thus both common and individual patterns of particular underlying factors could be considered and simultaneously extracted with imposing the nonnegative and sparse penalties. The designed synthetic fMRI-like data were adopted to testify the performance of the proposed algorithm and its competitors. In this article, the authors found that (1) a moderate sparse penalty will improve the performance of the algorithm, and then increasing it may have a negative impact; (2) the performance of the GNMF-based methods is superior to that of NMF-based ones; (3) with sparse regularization, the performance of NMF-based and GNMF-based methods can be both significantly improved; (4) sparse penalty yields better performance improvements than group constraint for NMF-based methods; (5) ADMM-based

methods achieve better performance, but are more time consuming; (6) GNMF-based methods can extract both the common and individual patterns for all the datasets, and can also successfully correct the disorders of the common patterns that cannot be realized in the NMF-based methods.

Contributions: Xiulin Wang conceived and carried out the idea of the study, including algorithm derivation and coding, experimental implementation and analysis, and paper writing. Wenya Liu took part in the discussion and analysis of the results and revised the manuscript. Fengyu Cong and Tapani Ristaniemi supervised the whole study and revised the manuscript.

3.2 Article II: "Generalization of Linked Canonical Polyadic Tensor Decomposition for Group Analysis"

Xiulin Wang, Chi Zhang, Tapani Ristaniemi and Fengyu Cong. Generalization of Linked Canonical Polyadic Tensor Decomposition for Group Analysis. 16th International Symposium on Neural Networks (ISNN), Moscow, Russia, 2019.

Method&Results: The Linked Tensor Decomposition (LTD) model is a coupled tensor decomposition technique proposed in recent years for group analysis, specifically for joint analysis of multi-block tensors. In this article, the authors proposed a generalized linked canonical polyadic tensor decomposition (GLCPTD) model that was well suited to exploiting the linking nature in multi-block tensor analysis. An efficient algorithm based on the hierarchical alternating least squares (HALS) method was proposed and termed as the GLCPTD-HALS algorithm, which enabled the simultaneous extraction of common components, individual components and core tensors from tensor blocks. Simulation experiments of synthetic EEG data analysis and image reconstruction and denoising were conducted to demonstrate the superior performance of the proposed generalized model and its realization. In synthetic EEG data analysis, the authors found that (1) GLCPTD-HALS and NTF-HALS algorithms can successfully extract the common components as well as individual components; (2) the components learned by NTF-HALS algorithm are disordered, thus clustering and other post-ordering methods are needed; (3) although LCPTD-HALS algorithm can extract all the common components, it cannot recover all of the potential components (omitted or merged), which makes group analysis more complicated. In image reconstruction and denoising, the authors found that (1) the images reconstructed by the LPCTD model are fuzzier or distorted than those from the GLCPTD model; (2) the PSNRs obtained by the GLCPTD model are higher than those obtained by the LCPTD model, which indicates that the proposed GLCPTD model matches the real-world data more closely; (3) when designing the coupled tensor decomposition experiment, stacking face images from the same subject with different expressions is more reliable than stacking that of the same expressions with dif-

ferent subjects; (4) the excessive number of common components may affect the fitness of the estimated tensors.

Contributions: Xiulin Wang conceived and carried out the idea of the study, including algorithm derivation and coding, experimental implementation and analysis, and paper writing. Chi Zhang helped revise the manuscript. Fengyu Cong and Tapani Ristaniemi supervised the study and revised the manuscript.

3.3 Article III: "Fast Implementation of Double-coupled Nonnegative Canonical Polyadic Decomposition"

Xiulin Wang, Tapani Ristaniemi and Fengyu Cong. Fast Implementa- tion of Double-coupled Nonnegative Canonical Polyadic Decomposition. 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 2019.

Method&Results: Real-world tensor data exhibiting high order/dimensionality and various couplings are linked to each other since they share some common characteristics. To address the multi-block tensor data, the authors proposed a fast double-coupled nonnegative Canonical Polyadic decomposition (FDC-NCPD) algorithm in this study, based on the linked CP tensor decomposition (LCPTD) model and fast hierarchical alternating least squares (fHALS) algorithm. Simulation experiments of synthetic and real-world data were conducted to demonstrate the superior performance of the proposed algorithm. From the experiment results, the authors found that: (1) simultaneous extraction of common components, individual components and core tensors can be achieved with the proposed algorithm; (2) the proposed algorithm can significantly reduce calculation time while retaining the decomposition quality; (3) joint/coupled analysis can effectively utilize prior information to improve the decomposition accuracy.

Contributions: Xiulin Wang conceived and carried out the idea of the study, including algorithm derivation and coding, experimental implementation and analysis, and paper writing. Fengyu Cong and Tapani Ristaniemi supervised the whole study and revised the manuscript.

3.4 Article IV: "Group analysis of ongoing EEG data based on fast double-coupled nonnegative tensor decomposition"

Xiulin Wang, Wenya Liu, Petri Toiviainen, Tapani Ristaniemi and Fengyu Cong. Group analysis of ongoing EEG data based on fast double-coupled nonnegative tensor decomposition. Journal of neuroscience methods, 330, p.108502, 2020.

Method&Results: Aiming to discover the reliable links between brain responses and musical stimulus from ongoing EEG elicited by 512-second long modern tango music, the authors first proposed a fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm. This algorithm is based on the generalized coupled model, which is capable of simultaneously decomposing EEG tensors of space, time and frequency into common components and individual components. Then the authors proposed a comprehensive framework for the group analysis of multi-subject ongoing EEG data, which includes seven steps: data acquisition & preprocessing, musical feature extraction, tensor representation, FDC-NTD implementation, correlation analysis, hierarchical clustering and cluster selection of interest. In this article, the authors found that (1) with the proposed framework, the brain activities can be effectively extracted and sorted into the clusters of interest; (2) the proposed algorithm based on the generalized model achieved higher fittings and stronger robustness; (3) the oscillatory brain activities in the three selected clusters are mainly distributed in theta (4-8 Hz), alpha (8-13 Hz) bands, and 4-11 Hz, and located in the centro-parietal, occipito-parietal and frontal regions respectively, while their corresponding temporal courses are significantly correlated musical features of the musical stimulus. The present study, providing a solution to separate common stimulus-elicited brain activities using coupled tensor decomposition, has shed new light on the processing and analysis of ongoing EEG data at the multi-subject level.

Contributions: Xiulin Wang conceived and carried out the idea of this study, including algorithm derivation and coding, experimental implementation and analysis, and paper writing. Wenya Liu contributed to the discussion of the results and the revision of the manuscript. Petri Toiviainen provided the data and revised the manuscript. Fengyu Cong and Tapani Ristaniemi supervised the whole study and revised the manuscript.

3.5 Article V: "Fast Learnings of Coupled Nonnegative Tensor Decomposition Using Optimal Gradient and Low-rank Approximation"

Xiulin Wang, Tapani Ristaniemi and Fengyu Cong. Fast Learnings of Coupled Nonnegative Tensor Decomposition Using Optimal Gradient and Low-rank Approximation. Submitted to <signal processing>, 2020.

Method&Results: Nonnegative tensor decomposition has been widely applied in signal processing and neuroscience, etc. When it comes to group analysis of multi-block tensors, traditional tensor decomposition is insufficient to maintain feature comparability or to utilize coupled information among tensors. In this article, the authors first proposed an effective coupled nonnegative CANDE-

COMP/PARAFAC decomposition algorithm based on alternating proximal gradient (CoNCPD-APG) method. This algorithm is capable of a simultaneous decomposition of tensors from different samples with coupled information and a simultaneous extraction of common components, individual components and core tensors. Moreover, to improve the optimization efficiency, the authors further introduced the low-rank approximation to the couple factorization problem and proposed the lraCoNCPD-APG algorithm. The authors designed and carried out two experiments on face image data and real-world event-related potential (ERP) data, aiming to examine and demonstrate the superior performance of CoNCPD-APG and lraCoNCPD-APG algorithms. In image reconstruction and denoising of image data, the authors found that (1) the proposed APG-based algorithms are superior to other algorithms, as they obtain higher decomposition accuracy and stronger image reconstruction capabilities; (2) by introducing the low-rank approximation strategy, the lraCoNCPD-APG algorithm can greatly reduce execution time while maintaining decomposition performance; (3) the proposed APG-based algorithms have excellent decomposition stability. In multi-domain feature extraction of ERP data, the authors found that (1) the proposed CoNCPD-APG and lraCoNCPD-APG algorithms are superior to competitors in terms of decomposition accuracy and multi-domain feature extraction stability; (2) the MU-based couple method achieves better performance than the fHALS-based algorithms in has accuracy but has the worst multi-domain feature extraction stability; (3) the low-rank approximation in joint analysis of large-scale tensors can greatly improve computation efficiency without losing the decomposition accuracy.

Contributions: Xiulin Wang conceived and carried out the idea of this study, including algorithm derivation and coding, experimental implementation and analysis, and paper writing. Fengyu Cong and Tapani Ristaniemi supervised the whole study and revised the manuscript.

4 CONCLUSION

This dissertation mainly focuses on the group analysis of brain imaging data using coupled nonnegative matrix/tensor decomposition algorithms, which fully utilize data characteristics, including multi-set/multi-modal, multi-coupling and multi-way structures. This dissertation is a synthesis of the included articles PI-PV, the contents of which are organized as follows.

(1) For the constrained joint analysis of data from multiple sources, we designed a flexible CNMF-SR model, which can be converted into NMF, NMF-SR and CNMF models by controlling relevant parameters. Within the BCD framework, we adopted the ADMM strategy to optimize the CNMF-SR model. Experiment using Multi-subject fMRI-like data verify that the proposed algorithm has better performance than its counterparts with coupling and adequate sparse constraints. In addition to the effective elimination of the blurs of recovered images, the proposed algorithm can also extract both common patterns (sorted) and individual patterns.

(2) To discover the commonly stimulus-elicited features among subjects, we proposed a comprehensive data analysis framework based on the CNTF model. We adopted the fHALS strategy to solve the CNTF model with the double-coupled constraint. For multi-subject ongoing EEG data recorded during freely listening to tango music, more reliable associations between brain responses and musical stimulus can be found using the proposed framework. Through providing a solution of how to use coupled tensor factorization to separate commonly stimulus-elicited brain activities, this study has shed new light on the processing and analysis of ongoing EEG data at the multi-subject level.

(3) To address the slow convergence speed and low optimization accuracy of most existing methods, we proposed two advanced CNTF algorithms: CNTF-APG and its fast implementation lraCNTF-APG. We adopted multi-subject ERP data collected from two groups of 21 children with reading difficulty and 21 children with attention deficit. The results show that the proposed APG-based algorithms can accurately and steadily extract the expected multi-domain features which can better discriminate the two groups. The introduction of low-rank approximation can greatly improve computation efficiency without losing the de-

composition accuracy.

In conclusion, this dissertation proposed efficient CNMF and CNTF algorithms, which have been successfully applied to brain imaging data's joint analysis. The CNMF and CNTF algorithms provide a natural framework for the simultaneous analysis of matrices or tensors with coupling information and the simultaneous extraction of common and individual features. Moreover, The CNMF and CNTF algorithms can reveal underlying structures and inner-relationships between datasets, and take full advantage of prior information to achieve higher decomposition accuracy and decomposition stability. However, this dissertation still has some limitations to consider.

(1) The proposed algorithms achieve the joint analysis of multiple datasets by imposing 'hard' links between datasets, i.e., we assume that different datasets have (partially) shared factors. However, in some cases, there are only (partially) highly correlated or similar factors between datasets, namely the 'soft' links. Thus, selecting the 'soft' links in the CNMF and CNTF problems according to the characteristics of data will be one of our future works.

(2) The determination of the number of components and coupled components depends on the validity of potential assumptions and relevant prior knowledge. The number of components can be calculated subjectively or objectively via a series of methods (such as model order selection or PCA-based method). We have proposed how to determine the number of coupled components by calculating the correlations of corresponding components between subjects. However, its selection is still subjective to some extent, which remains an open issue and invites more discussion.

(3) So far, the algorithms proposed in this dissertation are only applied for multi-subject and single-modal brain data, which does not suffice to study and understand brain functions. Therefore, joint analysis of multi-subject and multi-modal brain imaging data using coupled matrix/tensor factorization will be one of our future research topics.

YHTEENVETO (SUMMARY IN FINNISH)

Pääosin tämä väitöskirja käsittelee aivokuvantamisaineistoille tehtyjä liitoksellisten matriisien / tensorien epänegatiivisen yhteishajotelman algoritmeja, jotka ottavat hyvin huomioon aineistojen erityisominaisuuksia mukaan lukien useat eri modaliteetit, useaulotteiset rakenteet sekä rakenteiden väliset liitokset. Mukaan liitettyjen artikkeleiden PI-PV jälkeen väitöskirja rakentuu seuraavanlaisesti.

Kehitimme reunaehdot huomioon ottavalle usean lähteen yhteisanalyysille joustavan liitoksellisten matriisien epänegatiivinen yhteishajotelman harvalla esityksellä (CNMF-SR), josta saadaan sopivilla parametreilla erikoistapauksina matriisin epänegatiivinen hajotelma (NMF), matriisin epänegatiivinen hajotelma harvalla esityksellä (NMF-SR), ja liitoksellisten matriisien epänegatiivinen yhteishajotelma (CNMF). Lohkotun laskevan koordinaattisuunnan menetelmän (BCD) puitteissa sovelsimme ADMM-strategiaa CNMF-SR-mallin ratkaisemiseksi. Testimme usean koehenkilön simuloidun fMRI-aineiston kanssa osoittavat että ehdotettu algoritmi liitos- ja harvuusrajoitteilla on tehokkaampi kuin muut vastaavat algoritmit. Sen lisäksi että algoritmit pystyy tehokkaasti vähentämään sumentumia palautetuista kuvista, se kykenee samanaikaisesti erottamaan toisistaan yksilölliset ja jaetut piirteet.

Koehenkilöiden yhteisten ärsykkeiden tuottamien aivovasteiden löytämiseksi kehitimme CNTF-malliin pohjaten kattavan analyysikehikon, jossa käytimme fHALS-strategiaa kahden (taajuuden ja paikan) tensorien välisen liitoksen sisältävän aineiston hajotelman laskemiseksi. Tämän kehikon soveltaminen usean koehenkilön tango-musiikin kuuntelua sisältävään aineistoon löysi luotettavasti yhteyksiä aivovasteiden ja musiikin piirteiden välillä. Näyttäessään kuinka liitoksellisten tensorien yhteishajotelmaa voi käyttää yksilöllisten ja jaettujen aivovasteiden erottamiseksi, tämä tutkimus on tärkeä edistysaskel usean koehenkilön sisältävän jatkuvan EEG-aineiston analyysissä.

Useimpien jo olemassaolevien menetelmien hitaan suppenemisen ja matalan ratkaisutarkkuuden vuoksi ehdotamme kahta edistynyttä liitoksellisten tensorien epänegatiivisen yhteishajotelman algoritmia: vuorottelevaan proksimaaliseen gradienttiin pohjaavan menetelmän (CNTF-APG) ja sen matala-asteiseen approksimaatioon perustuvan nopeamman version (lraCNTF-APG). EEG-herätevasteaineistossa, jossa oli mukana 21 lukemisvaikeuksista kärsivää lasta sekä 21 tarkkavaisuushäiriöstä kärsivää lasta, menetelmät löysivät tarkasti ja vakaasti aivovasteet, jotka erottivat kaksi ryhmää toisistaan. Matala-asteinen approksimaation käyttöönotto voi merkittävästi lisätä laskennallista tehokkuutta kuitenkaan vähentämättä hajotelman tarkkuutta.

Tiivistäen, tämä väitöskirja esittelee joitain tehokkaita algoritmeja liitoksellisten matriisien tai tensorien epänegatiivisen yhteishajotelman laskemiseksi, ja niitä on onnistuneesti sovellettu aivokuvantamisaineistojen analyysiin.

REFERENCES

Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., and Yener, B. (2007). Multiway analysis of epilepsy tensors. *Bioinformatics*, 23(13):i10–i18.

Acar, E., Bro, R., and Smilde, A. K. (2015). Data fusion in metabolomics using coupled matrix and tensor factorizations. *Proceedings of the IEEE*, 103(9):1602–1620.

Acar, E., Levin-Schwartz, Y., Calhoun, V. D., and Adali, T. (2017a). Acmtf for fusion of multi-modal neuroimaging data and identification of biomarkers. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 643–647. IEEE.

Acar, E., Levin-Schwartz, Y., Calhoun, V. D., and Adali, T. (2017b). Tensor-based fusion of eeg and fmri to understand neurological changes in schizophrenia. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–4. IEEE.

Adali, T., Anderson, M., and Fu, G.-S. (2014). Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging. *IEEE Signal Processing Magazine*, 31(3):18–33.

Adali, T., Levin-Schwartz, Y., and Calhoun, V. D. (2015). Multimodal data fusion using source separation: Application to medical imaging. *Proceedings of the IEEE*, 103(9):1494–1506.

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage*, 59(4):3677–3689.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202.

Beckmann, C. F. and Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject fmri analysis. *Neuroimage*, 25(1):294–311.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

Calhoun, V. D., Adali, T., Giuliani, N., Pekar, J. J., Kiehl, K., and Pearlson, G. (2006). Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data. *Human brain mapping*, 27(1):47–62.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001). A method for making group inferences from functional mri data using independent component analysis. *Human brain mapping*, 14(3):140–151.

Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319.

Chen, X., Wang, Z. J., and McKeown, M. (2016). Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods. *IEEE Signal Processing Magazine*, 33(3):86–107.

Chen, Z., Cichocki, A., and Rutkowski, T. M. (2006). Constrained non-negative matrix factorization method for eeg analysis in early detection of alzheimer disease. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 5, pages V–V. IEEE.

Cichocki, A. (2013). Tensor decompositions: a new concept in brain data analysis? *arXiv preprint arXiv:1305.0395*.

Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine*, 32(2):145–163.

Cichocki, A. and Phan, A.-H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 92(3):708–721.

Cichocki, A., Zdunek, R., and Amari, S.-i. (2006). New algorithms for nonnegative matrix factorization in applications to blind source separation. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 5, pages V–V. IEEE.

Cichocki, A., Zdunek, R., and Amari, S.-i. (2007). Hierarchical als algorithms for nonnegative matrix and 3d tensor factorization. In *International Conference on Independent Component Analysis and Signal Separation*, pages 169–176. Springer.

Cong, F., Alluri, V., Nandi, A. K., Toiviainen, P., Fa, R., Abu-Jamous, B., Gong, L., Craenen, B. G., Poikonen, H., Huotilainen, M., et al. (2013). Linking brain responses to naturalistic music through analysis of ongoing eeg and stimulus features. *IEEE Transactions on Multimedia*, 15(5):1060–1069.

Cong, F., Lin, Q.-H., Kuang, L.-D., Gong, X.-F., Astikainen, P., and Ristaniemi, T. (2015a). Tensor decomposition of EEG signals: a brief review. *Journal of neuroscience methods*, 248:59–69.

Cong, F., Phan, A. H., Zhao, Q., Huttunen-Scott, T., Kaartinen, J., Ristaniemi, T., Lyytinen, H., and Cichocki, A. (2012a). Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from eeg collected by low-density array. *International journal of neural systems*, 22(06):1250025.

Cong, F., Phan, A. H., Zhao, Q., Nandi, A. K., Alluri, V., Toiviainen, P., Poikonen, H., Huotilainen, M., Cichocki, A., and Ristaniemi, T. (2012b). Analysis of ongoing eeg elicited by natural music stimuli using nonnegative tensor factorization. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 494–498. IEEE.

Cong, F., Ristaniemi, T., and Lyytinen, H. (2015b). *Advanced Signal Processing on Brain Event-Related Potentials: Filtering ERPs in Time, Frequency and Space Domains Sequentially and Simultaneously*, volume 13. World Scientific.

Cong, F., Zhou, G., Astikainen, P., Zhao, Q., Wu, Q., Nandi, A. K., Hietanen, J. K., Ristaniemi, T., and Cichocki, A. (2014). Low-rank approximation based non-negative multi-way array decomposition on event-related potentials. *International journal of neural systems*, 24(08):1440005.

Correa, N. M., Adali, T., Li, Y.-O., and Calhoun, V. D. (2010a). Canonical correlation analysis for data fusion and group inferences. *IEEE signal processing magazine*, 27(4):39–50.

Correa, N. M., Eichele, T., Adalı, T., Li, Y.-O., and Calhoun, V. D. (2010b). Multiset canonical correlation analysis for the fusion of concurrent single trial erp and functional mri. *Neuroimage*, 50(4):1438–1445.

Correa, N. M., Li, Y.-O., Adali, T., and Calhoun, V. D. (2008). Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE journal of selected topics in signal processing*, 2(6):998–1007.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Gong, X.-F., Lin, Q.-H., Cong, F.-Y., and De Lathauwer, L. (2018a). Double coupled canonical polyadic decomposition for joint blind source separation. *IEEE Transactions on Signal Processing*, 66(13):3475–3490.

Gong, X.-F., Mao, L., Liu, Y.-L., and Lin, Q.-H. (2018b). A jacobi generalized orthogonal joint diagonalization algorithm for joint blind source separation. *IEEE Access*, 6:38464–38474.

Gong, X.-F., Wang, X.-L., and Lin, Q.-H. (2015). Generalized non-orthogonal joint diagonalization with lu decomposition and successive rotations. *IEEE Transactions on Signal Processing*, 63(5):1322–1334.

Gotman, J., Bénar, C.-G., and Dubeau, F. (2004). Combining eeg and fmri in epilepsy: methodological challenges and clinical results. *Journal of Clinical Neurophysiology*, 21(4):229–240.

Groves, A. R., Beckmann, C. F., Smith, S. M., and Woolrich, M. W. (2011). Linked independent component analysis for multimodal data fusion. *Neuroimage*, 54(3):2198–2217.

Guan, N., Tao, D., Luo, Z., and Yuan, B. (2012). Nenmf: An optimal gradient method for nonnegative matrix factorization. *IEEE Transactions on Signal Processing*, 60(6):2882–2898.

Guillamet, D. and Vitria, J. (2002). Non-negative matrix factorization for face recognition. In *Catalonian Conference on Artificial Intelligence*, pages 336–344. Springer.

Harshman, R. A. et al. (1970). Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis.

Hitchcock, F. L. (1927). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469.

Hunyadi, B., Dupont, P., Van Paesschen, W., and Van Huffel, S. (2017). Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(1):e1197.

Jagannathan, K., Calhoun, V. D., Gelernter, J., Stevens, M. C., Liu, J., Bolognani, F., Windemuth, A., Ruaño, G., Assaf, M., and Pearlson, G. D. (2010). Genetic associations of brain structural networks in schizophrenia: a preliminary study. *Biological psychiatry*, 68(7):657–666.

Jonmohamadi, Y., Muthukumaraswamy, S., Chen, J., Roberts, J., Crawford, R., and Pandey, A. (2019). Extraction of common task features in eeg-fmri data using coupled tensor-tensor decomposition. *bioRxiv*, page 685941.

Kameoka, H., Ono, N., Kashino, K., and Sagayama, S. (2009). Complex nmf: A new sparse representation for acoustic signals. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3437–3440. IEEE.

Kanatsoulis, C. I., Fu, X., Sidiropoulos, N. D., and Ma, W.-K. (2018). Hyperspectral super-resolution: A coupled tensor factorization approach. *IEEE Transactions on Signal Processing*, 66(24):6503–6517.

Karahan, E., Rojas-Lopez, P. A., Bringas-Vega, M. L., Valdes-Hernandez, P. A., and Valdes-Sosa, P. A. (2015). Tensor analysis and fusion of multimodal brain images. *Proceedings of the IEEE*, 103(9):1531–1559.

Kim, H., Choo, J., Kim, J., Reddy, C. K., and Park, H. (2015). Simultaneous discovery of common and discriminative topics via joint nonnegative matrix factorization. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 567–576.

Kim, H. and Park, H. (2008). Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM journal on matrix analysis and applications*, 30(2):713–730.

Kim, J. and Park, H. (2011). Fast nonnegative matrix factorization: An active-set-like method and comparisons. *SIAM Journal on Scientific Computing*, 33(6):3261–3281.

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

Krishnan, A., Williams, L. J., McIntosh, A. R., and Abdi, H. (2011). Partial least squares (pls) methods for neuroimaging: a tutorial and review. *Neuroimage*, 56(2):455–475.

Lahat, D., Adali, T., and Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9):1449–1477.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Lee, H. and Choi, S. (2009). Group nonnegative matrix factorization for EEG classification. In *Artificial Intelligence and Statistics*, pages 320–327.

Lee, H., Cichocki, A., and Choi, S. (2009). Kernel nonnegative matrix factorization for spectral eeg feature extraction. *Neurocomputing*, 72(13-15):3182–3190.

Lee, J.-H., Lee, T.-W., Jolesz, F. A., and Yoo, S.-S. (2008). Independent vector analysis (iva): multivariate approach for fmri group study. *Neuroimage*, 40(1):86–109.

Levin-Schwartz, Y., Calhoun, V. D., and Adalı, T. (2014). Data-driven fusion of eeg, functional and structural mri: A comparison of two models. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE.

Li, S., Dian, R., Fang, L., and Bioucas-Dias, J. M. (2018). Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Transactions on Image Processing*, 27(8):4118–4130.

Li, Y.-O., Adali, T., Wang, W., and Calhoun, V. D. (2009). Joint blind source separation by multiset canonical correlation analysis. *IEEE Transactions on Signal Processing*, 57(10):3918–3929.

Lin, C.-J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural computation*, 19(10):2756–2779.

Liu, J. and Calhoun, V. (2007). Parallel independent component analysis for multimodal analysis: application to fmri and eeg data. In *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1028–1031. IEEE.

Liu, J., Kiehl, K. A., Pearlson, G., Perrone-Bizzozero, N. I., Eichele, T., and Calhoun, V. D. (2009a). Genetic determinants of target and novelty-related event-related potentials in the auditory oddball response. *Neuroimage*, 46(3):809–816.

Liu, J., Pearlson, G., Windemuth, A., Ruano, G., Perrone-Bizzozero, N. I., and Calhoun, V. (2009b). Combining fmri and snp data to investigate connections between brain function and genetics using parallel ica. *Human brain mapping*, 30(1):241–255.

Mørup, M. (2011). Applications of tensor (multiway array) factorizations and decompositions in data mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):24–40.

Mørup, M., Hansen, L. K., and Arnfred, S. M. (2007). Erpwavelab: A toolbox for multi-channel analysis of time–frequency transformed event related potentials. *Journal of neuroscience methods*, 161(2):361–368.

Mørup, M., Hansen, L. K., Herrmann, C. S., Parnas, J., and Arnfred, S. M. (2006a). Parallel factor analysis as an exploratory tool for wavelet transformed event-related eeg. *NeuroImage*, 29(3):938–947.

Mørup, M., Hansen, L. K., Parnas, J., and Arnfred, S. M. (2006b). Decomposing the time-frequency representation of EEG using non-negative matrix and multi-way factorization. *Technical University of Denmark Technical Report*, pages 1–28.

Naskovska, K., Lau, S., Korobkov, A. A., Haueisen, J., and Haardt, M. (2020). Coupled cp decomposition of simultaneous meg-eeg signals for differentiating oscillators during photic driving. *Frontiers in Neuroscience*, 14.

Nesterov, Y. E. (1983). A method for solving the convex programming problem with convergence rate o (1/kˆ 2). In *Dokl. akad. nauk Sssr*, volume 269, pages 543–547.

Pauca, V. P., Shahnaz, F., Berry, M. W., and Plemmons, R. J. (2004). Text mining using non-negative matrix factorizations. In *Proceedings of the 2004 SIAM International Conference on Data Mining*, pages 452–456. SIAM.

Plis, S. M., Calhoun, V. D., Eichele, T., Weisend, M. P., and Lane, T. (2010). Meg and fmri fusion for nonlinear estimation of neural and bold signal changes. *Frontiers in neuroinformatics*, 4:114.

Razaviyayn, M., Hong, M., and Luo, Z.-Q. (2013). A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153.

Ritter, P. and Villringer, A. (2006). simultaneous eeg–fmri. *Neuroscience & Biobehavioral Reviews*, 30(6):823–838.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*, 90:449–468.

Shin, B. and Oh, A. (2012). Bayesian group nonnegative matrix factorization for EEG analysis. *arXiv preprint arXiv:1212.4347*.

Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., and Faloutsos, C. (2017). Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582.

Sørensen, M. and De Lathauwer, L. (2013). Coupled tensor decompositions for applications in array signal processing. In *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 228–231. IEEE.

Specht, K., Zahn, R., Willmes, K., Weis, S., Holtel, C., Krause, B. J., Herzog, H., and Huber, W. (2009). Joint independent component analysis of structural and functional images reveals complex patterns of functional reorganisation in stroke aphasia. *Neuroimage*, 47(4):2057–2063.

Spyrou, L., Parra, M., and Escudero, J. (2018). Complex tensor factorization with parafac2 for the estimation of brain connectivity from the eeg. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(1):1–12.

Sui, J., Adali, T., Yu, Q., Chen, J., and Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of neuroscience methods*, 204(1):68–81.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311.

Wang, D., Cong, F., Zhao, Q., Toiviainen, P., Nandi, A. K., Huotilainen, M., Ristaniemi, T., and Cichocki, A. (2016). Exploiting ongoing eeg with multilinear partial least squares during free-listening to music. In *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE.

Wang, D., Zhu, Y., Ristaniemi, T., and Cong, F. (2018). Extracting multi-mode erp features using fifth-order nonnegative tensor decomposition. *Journal of neuroscience methods*, 308:240–247.

Wang, X., Liu, W., Toiviainen, P., Ristaniemi, T., and Cong, F. (2020). Group analysis of ongoing eeg data based on fast double-coupled nonnegative tensor decomposition. *Journal of neuroscience methods*, 330:108502.

Wang, X., Ristaniemi, T., and Cong, F. (2019a). Fast implementation of double-coupled nonnegative canonical polyadic decomposition. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8588–8592. IEEE.

Wang, X., Zhang, C., Ristaniemi, T., and Cong, F. (2019b). Generalization of linked canonical polyadic tensor decomposition for group analysis. In *International Symposium on Neural Networks*, pages 180–189. Springer.

Xu, W., Liu, X., and Gong, Y. (2003). Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273.

Xu, Y. (2015). Alternating proximal gradient method for sparse nonnegative tucker decomposition. *Mathematical Programming Computation*, 7(1):39–70.

Xu, Y. and Yin, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences*, 6(3):1758–1789.

Yokota, T., Cichocki, A., and Yamashita, Y. (2012). Linked parafac/cp tensor decomposition and its fast implementation for multi-block tensor analysis. In *International Conference on Neural Information Processing*, pages 84–91. Springer.

Zdunek, R. and Cichocki, A. (2006). Non-negative matrix factorization with quasi-newton optimization. In *International conference on artificial intelligence and soft computing*, pages 870–879. Springer.

Zhang, Y., Zhou, G., Zhao, Q., Cichocki, A., and Wang, X. (2016). Fast nonnegative tensor factorization based on accelerated proximal gradient and low-rank approximation. *Neurocomputing*, 198:148–154.

Zhong, Y., Wang, H., Lu, G., Zhang, Z., Jiao, Q., and Liu, Y. (2009). Detecting functional connectivity in fmri using pca and regression analysis. *Brain topography*, 22(2):134–144.

Zhou, G., Cichocki, A., and Xie, S. (2012). Fast nonnegative matrix/tensor factorization based on low-rank approximation. *IEEE Transactions on Signal Processing*, 60(6):2928–2940.

Zhou, G., Cichocki, A., Zhang, Y., and Mandic, D. P. (2015). Group component analysis for multiblock data: Common and individual feature extraction. *IEEE transactions on neural networks and learning systems*, 27(11):2426–2439.

Zhou, G., Cichocki, A., Zhao, Q., and Xie, S. (2014). Nonnegative matrix and tensor factorizations: An algorithmic perspective. *IEEE Signal Processing Magazine*, 31(3):54–65.

Zhou, G., Zhao, Q., Zhang, Y., Adalı, T., Xie, S., and Cichocki, A. (2016). Linked component analysis from matrices to high-order tensors: Applications to biomedical data. *Proceedings of the IEEE*, 104(2):310–331.

Zhu, Y., Liu, J., Ye, C., Mathiak, K., Astikainen, P., Ristaniemi, T., and Cong, F. (2020). Discovering dynamic task-modulated functional networks with specific spectral modes using meg. *NeuroImage*, page 116924.

ORIGINAL PAPERS

PI

GROUP NONNEGATIVE MATRIX FACTORIZATION WITH SPARSE REGULARIZATION IN MULTI-SET DATA

by

Xiulin Wang, Wenya Liu, Fengyu Cong, and Tapani Ristaniemi 2020

28th European Signal Processing Conference (EUSIPCO), Amsterdam, Netherlands

Group Nonnegative Matrix Factorization with Sparse Regularization in Multi-set Data

Xiulin Wang[⋆,†], Wenya Liu[⋆,†], Fengyu Cong[⋆,†], Tapani Ristaniemi[†]
[⋆] *School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China*
[†] *Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland*
xiulin.wang@foxmail.com, wenyaliu0912@foxmail.com, cong@dlut.edu.cn, tapani.e.ristaniemi@jyu.fi

***Abstract*—Constrained joint analysis of data from multiple sources has received widespread attention for that it allows us to explore potential connections and extract meaningful hidden components. In this paper, we formulate a flexible joint source separation model termed as group nonnegative matrix factorization with sparse regularization (GNMF-SR) to jointly analyze the partially coupled multi-set data. With the generalized model, both common and individual patterns of particular underlying factors can be considered and simultaneously extracted with imposing the nonnegative and sparse penalties. Alternating optimization and alternating direction method of multipliers (ADMM) are combined to solve the GNMF-SR model. Using the experiment of simulated fMRI-like data, we demonstrate the ADMM-based GNMF-SR algorithm can achieve the better performance.**

***Index Terms*—Alternating direction method of multipliers, coupled, group nonnegative matrix factorization, joint analysis, sparse representation**

I. Introduction

Nonnegative matrix factorization (NMF), providing a part-based representation of nonnegative data, has been widely applied in blind source separation (BSS) problems including signal processing and machine learning [1]–[4]. With the increasing availability of sensor technologies, we are now facing a mass of data from multiple sources that need to be jointly separated [5]–[8], such as the multi-subject/multi-modal biomedical data [6], [7]. Although many studies have shown that conventional NMF methods are effective in a large of single dataset applications, their inefficiency in group analysis of multiple datasets has limited their broader usage [7]. In order to fill the gap between NMF and group analysis of multiple datasets, group nonnegative matrix factorization (GNMF) was proposed as an updated modification of the standard NMF in multi-set problems [9], [10]. In the group model, coupling information across datasets can be exploited, making it possible to achieve higher performance than BSS-based algorithms originally designed for one dataset [5], [7], [9]. Therefore, it is easy to extract the underlying patterns that are common among datasets, as well as individual patterns that exhibit internal variability [8], [9]. In addition, group analysis in multiple datasets can also automatically maintain the alignment of coupled patterns among datasets, while BSS-based algorithms need to adopt some post-aligned strategies [5], [9].

Spares representation aims to encode the data using fewer 'active' components for better interpretation of the encoding [11], [12]. Even though NMF-based methods can naturally produce a sparse representation of the data, the sparseness of extracted factors is not enough and uncontrollable [11]. Therefore, sparse regularization has been widely applied to NMF to promote the sparse representation and alleviate the factorized non-uniqueness [13]. Inspired by GNMF and sparse NMF, we formulate a group nonnegative matrix factorization with sparse regularization (GNMF-SR) model by imposing an efficient and commonly used regularizer l_1-norm for constrained joint analysis of partially coupled datasets. Obviously, the GNMF works in [9], [10] do not take the sparse characteristic of latent variables into consideration, and the sparse NMF works in [11]–[13] cannot utilize the coupled information across the datasets. In recent years, the alternating direction method of multipliers (ADMM) has become an effective and popular tool for constrained NMF problems [14]–[17], and in this study we employ ADMM to optimize the GNMF-SR model. The convergence issue of NMF-based or nonconvex optimization problems about ADMM has been widely discussed in [15]–[19], which will not be discussed in this study. A comprehensive review of ADMM can be found in [17].

The rest of this paper is organized as follows. Section 2 introduces multi-set data model, GNMF models, ADMM method and model optimization via ADMM method. In section 3, simulation experiment on synthetic fMRI-like data is conducted. The last section concludes this paper.

Notations: Scalars, vectors and matrices are respectively denoted by lowercase, boldface lowercase and boldface uppercase, e.g. x, $\boldsymbol{x}$ and $\boldsymbol{X}$. $\mathbb{R}_+$ denotes the nonnegative real number. Operators $(\cdot)^T$, $\|\cdot\|_1$ and $\|\cdot\|_F$ denote transpose, l_1-norm and Frobenius norm, respectively. $\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ denotes the inner product of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. $\langle \boldsymbol{A}, \boldsymbol{B} \rangle := \sum_{i,j} a_{ij} b_{ij}$ can be substituted by $\mathrm{tr}(\boldsymbol{A}\boldsymbol{B}^T)$ for $\boldsymbol{A}$ and $\boldsymbol{B}$ with the same size $I \times J$.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91748105 & 81471742), the Fundamental Research Funds for the Central Universities [DUT2019] in Dalian University of Technology in China and the scholarships from China Scholarship Council (Nos. 201706060262 and 201706060263). Corresponding author: Fengyu Cong, cong@dlut.edu.cn

II. Methods

In this section, we first introduce the multi-set data model, then we present the GNMF and GNMF-SR models, and last give the ADMM method and the optimization solution of GNMF-SR model.

A. Multi-set data model

Given a set of nonnegative matrices $\boldsymbol{X}^{(s)} \in \mathbb{R}_+^{I^{(s)}\times J^{(s)}}$, $s = 1, 2, \cdots, S$, the multi-set data model assumes that each data $\boldsymbol{X}^{(s)}$ can be expressed by:

$$\boldsymbol{X}^{(s)} \approx \boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)} = [\boldsymbol{A}_C^{(s)}\ \boldsymbol{A}_I^{(s)}]\boldsymbol{B}^{(s)}, \tag{1}$$

where $\boldsymbol{A}^{(s)} \in \mathbb{R}_+^{I^{(s)}\times R^{(s)}}$ and $\boldsymbol{B}^{(s)} \in \mathbb{R}_+^{R^{(s)}\times J^{(s)}}$ represent the latent variable and corresponding coefficient matrix respectively. Generally, $R^{(s)} < \min(I^{(s)}, J^{(s)})$ is assumed for providing a low-rank representation of $\boldsymbol{X}^{(s)}$. Assuming that those data are collected under the same condition, it is reasonable to expect that there will be some identical or highly correlated hidden information between the data. Therefore, in multi-set data model, we assume that each factor matrix $\boldsymbol{A}^{(s)} = [\boldsymbol{A}_C^{(s)}\ \boldsymbol{A}_I^{(s)}]$ includes two patterns: $\boldsymbol{A}_C^{(s)} \in \mathbb{R}_+^{I^{(s)}\times L}$, $0 \le L \le R^{(s)}$, a common matrix shared by all S matrices as $\boldsymbol{A}_C^{(1)} = \cdots \boldsymbol{A}_C^{(S)} = \boldsymbol{A}_C$, and $\boldsymbol{A}_I^{(s)} \in \mathbb{R}_+^{I^{(s)}\times(R^{(s)}-L)}$, which corresponds to the individual characteristic in each dataset.

B. Group nonnegative matrix factorization

Considering the coupling structure among the latent variables $\boldsymbol{A}^{(s)}$ in multi-set data model, we need to analyze S sets of $\boldsymbol{X}^{(s)}$ simultaneously, which is different from the conventional NMF problem. Using the Euclidean divergence minimization, the GNMF of $\boldsymbol{X}^{(s)}$, $s = 1, 2, \cdots, S$, can be achieved by solving the following optimization:

$$\begin{aligned} &\underset{\boldsymbol{A}^{(s)},\boldsymbol{B}^{(s)}}{\text{minimize}}\ \frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{X}^{(s)}-\boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 \\ &\text{subject to}\ \ \boldsymbol{A}^{(s)} \ge 0, \boldsymbol{B}^{(s)} \ge 0, \end{aligned} \tag{2}$$

In many applications only the underlying patterns in the variable dimension need to be sparse [20], thus in this study we only consider imposing the sparsity on factor matrix $\boldsymbol{A}^{(s)}$ as follows:

$$\begin{aligned} &\underset{\boldsymbol{A}^{(s)},\boldsymbol{B}^{(s)}}{\text{minimize}}\ \frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{X}^{(s)}-\boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 + \sum_{s=1}^{S}\beta^{(s)}\sum_{r=1}^{R(s)}\left\|\boldsymbol{a}_r^{(s)}\right\|_1 \\ &\text{subject to}\ \ \boldsymbol{A}^{(s)} \ge 0, \boldsymbol{B}^{(s)} \ge 0, \end{aligned} \tag{3}$$

where $\boldsymbol{a}_r^{(s)}$ corresponds to the rth column of $\boldsymbol{A}^{(s)}$, and note that $\boldsymbol{A}_C^{(1)} = \cdots \boldsymbol{A}_C^{(S)} = \boldsymbol{A}_C$. The penalty term $\sum_{r=1}^{R^{(s)}}\left\|\boldsymbol{a}_r^{(s)}\right\|_1$ is to impose the sparsity on factor matrix $\boldsymbol{A}^{(s)}$, and it can be reformed as $\left\langle \boldsymbol{E}, \boldsymbol{A}^{(s)}\right\rangle$, in which $\boldsymbol{E} \in \mathbb{R}_+^{I^{(s)}\times J^{(s)}}$ is a matrix whose entries are all ones. $\beta^{(s)} \ge 0$ is a predefined penalty parameter. For simplicity, we set $\beta^{(1)} = \beta^{(2)} = \cdots \beta^{(S)}$. We term the GNMF with sparse regularization in (3) as GNMF-SR model. Later we will give a detailed explanation of how to solve GNMF-SR model using ADMM algorithm.

C. Alternating direction method of multipliers

According to [17], ADMM algorithm considers the following problem:

$$\begin{aligned} &\underset{\boldsymbol{x},\boldsymbol{z}}{\text{minimize}}\ \ f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ &\text{subject to}\ \ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}. \end{aligned} \tag{4}$$

Using the scaled from, it can be updated iteratively using the following steps:

$$\begin{cases} \boldsymbol{x} := \underset{\boldsymbol{x}}{\operatorname{argmin}}\left(f(\boldsymbol{x}) + (\rho/2)\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{z}-\boldsymbol{c}+\boldsymbol{u}\right\|_2^2\right), \\ \boldsymbol{z} := \underset{\boldsymbol{z}}{\operatorname{argmin}}\left(g(\boldsymbol{z}) + (\rho/2)\left\|\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{z}-\boldsymbol{c}+\boldsymbol{u}\right\|_2^2\right), \\ \boldsymbol{u} := \boldsymbol{u} + \left(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{z}-\boldsymbol{c}\right), \end{cases} \tag{5}$$

where $\boldsymbol{u}$ denote the scaled dual variable and $\rho > 0$ denotes the preselected augmented Lagrangian parameter.

D. GNMF-SR solutions via ADMM

To solve the nonconvex optimization problem, ADMM algorithm splits it into smaller pieces so that it can be easily handled one-to-one [17]. Moreover, the problem (3) can be first converted to two sub-problems: $\boldsymbol{A}^{(s)}$ and $\boldsymbol{B}^{(s)}$ via alternating optimization strategy, and then one of sub-problems can be solved using ADMM algorithm effectively if the other is fixed [16]. Combining alternating optimization and ADMM strategies [14]–[17], [21], we introduce two auxiliary variables $\tilde{\boldsymbol{A}}^{(s)}$ and $\tilde{\boldsymbol{B}}^{(s)}$, and consider the following minimization reformation of (3) as:

$$\begin{aligned} &\frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{X}^{(s)}-\boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 + \sum_{s=1}^{S}\beta^{(s)}\sum_{r=1}^{R(s)}\left\|\tilde{\boldsymbol{a}}_r^{(s)}\right\|_1 \\ &\text{subject to}\ \ \boldsymbol{A}^{(s)} = \tilde{\boldsymbol{A}}^{(s)}, \boldsymbol{B}^{(s)} = \tilde{\boldsymbol{B}}^{(s)}, \tilde{\boldsymbol{A}}^{(s)} \ge 0, \tilde{\boldsymbol{B}}^{(s)} \ge 0. \end{aligned} \tag{6}$$

Corresponding to $\boldsymbol{A}^{(s)}$, the auxiliary variable $\tilde{\boldsymbol{A}}^{(s)}$ still consists of two parts: $\tilde{\boldsymbol{A}}_C^{(s)}$ and $\tilde{\boldsymbol{A}}_I^{(s)}$, and $\tilde{\boldsymbol{A}}_C^{(1)} = \cdots \tilde{\boldsymbol{A}}_C^{(S)} = \tilde{\boldsymbol{A}}_C$. The augmented Lagrangian function for the above problem (6) is given by:

$$\begin{aligned} &\mathcal{L}(\boldsymbol{A}^{(s)}, \boldsymbol{B}^{(s)}, \tilde{\boldsymbol{A}}^{(s)}, \tilde{\boldsymbol{B}}^{(s)}, \boldsymbol{\Lambda}^{(s)}, \boldsymbol{\Gamma}^{(s)}) \\ &= \frac{1}{2}\sum_{s=1}^{S}\left\|\boldsymbol{X}^{(s)}-\boldsymbol{A}^{(s)}\boldsymbol{B}^{(s)}\right\|_F^2 + \sum_{s=1}^{S}\beta^{(s)}\sum_{r=1}^{R(s)}\left\|\tilde{\boldsymbol{a}}_r^{(s)}\right\|_1 \\ &+ \sum_{s=1}^{S}\frac{\rho^{(s)}}{2}\left\|\boldsymbol{A}^{(s)}-\tilde{\boldsymbol{A}}^{(s)}+\boldsymbol{\Lambda}^{(s)}\right\|_F^2 \\ &+ \sum_{s=1}^{S}\frac{\mu^{(s)}}{2}\left\|\boldsymbol{B}^{(s)}-\tilde{\boldsymbol{B}}^{(s)}+\boldsymbol{\Gamma}^{(s)}\right\|_F^2, \end{aligned} \tag{7}$$

where $\boldsymbol{\Lambda}^{(s)} \in \mathbb{R}_+^{I^{(s)}\times R^{(s)}}$ and $\boldsymbol{\Gamma}^{(s)} \in \mathbb{R}_+^{R^{(s)}\times J^{(s)}}$ are termed as dual variables. $\rho^{(s)}$ and $\mu^{(s)}$ are the penalty parameters predefined by the user, and here we set $\rho^{(s)} = \left\|\boldsymbol{B}^{(s)}\right\|_F^2/R^{(s)}$ and $\mu^{(s)} = \left\|\boldsymbol{A}^{(s)}\right\|_F^2/R^{(s)}$ as suggested in [16].

$$
\begin{cases}
\boldsymbol{A}_C &= \left[\sum_{s=1}^{S} \boldsymbol{X}^{(s)}\left(\boldsymbol{B}_C^{(s)}\right)^T - \sum_{s=1}^{S} \boldsymbol{A}_I^{(s)}\boldsymbol{B}_I^{(s)}\left(\boldsymbol{B}_C^{(s)}\right)^T - \sum_{s=1}^{S} \rho^{(s)}\boldsymbol{\Lambda}_C^{(s)} + \sum_{s=1}^{S} \rho^{(s)}\tilde{\boldsymbol{A}}_C^{(s)}\right]\left[\sum_{s=1}^{S} \boldsymbol{B}_C^{(s)}\left(\boldsymbol{B}_C^{(s)}\right)^T + \sum_{s=1}^{S} \rho^{(s)}\boldsymbol{I}\right]^{-1} \\
\boldsymbol{A}_I^{(s)} &= \left[\boldsymbol{X}^{(s)}\left(\boldsymbol{B}_I^{(s)}\right)^T - \boldsymbol{A}_C^{(s)}\boldsymbol{B}_C^{(s)}\left(\boldsymbol{B}_I^{(s)}\right)^T - \rho^{(s)}\boldsymbol{\Lambda}_I^{(s)} + \rho^{(s)}\tilde{\boldsymbol{A}}_I^{(s)}\right]\left[\boldsymbol{B}_I^{(s)}\left(\boldsymbol{B}_I^{(s)}\right)^T + \rho^{(s)}\boldsymbol{I}\right]^{-1} \\
\boldsymbol{B}^{(s)} &= \left[\left(\boldsymbol{X}^{(s)}\right)^T\boldsymbol{A}^{(s)} - \mu^s\boldsymbol{\Gamma}^{(s)} + \mu^{(s)}\tilde{\boldsymbol{B}}^{(s)}\right]\left[\left(\boldsymbol{A}^{(s)}\right)^T\boldsymbol{A}^{(s)} + \mu^{(s)}\boldsymbol{I}\right]^{-1} \\
\tilde{\boldsymbol{A}}_C &= \left[\boldsymbol{A}_C + \frac{\sum_{s=1}^{S}\rho^{(s)}\boldsymbol{\Lambda}_C^{(s)}}{\sum_{s=1}^{S}\rho^{(s)}} - \frac{\sum_{s=1}^{S}\beta^{(s)}\boldsymbol{E}_C}{\sum_{s=1}^{S}\rho^{(s)}}\right]_+, \quad \tilde{\boldsymbol{A}}_I^{(s)} = \left[\boldsymbol{A}_I^{(s)} + \boldsymbol{\Lambda}_I^{(s)} - \frac{\beta^{(s)}\boldsymbol{E}_I}{\rho^{(s)}}\right]_+ \\
\tilde{\boldsymbol{B}}^{(s)} &= \left[\boldsymbol{B}^{(s)} + \boldsymbol{\Gamma}^{(s)}\right]_+, \quad \boldsymbol{\Lambda}^{(s)} = \boldsymbol{\Lambda}^{(s)} + \boldsymbol{A}^{(s)} - \tilde{\boldsymbol{A}}^{(s)}, \quad \boldsymbol{\Gamma}^{(s)} = \boldsymbol{\Gamma}^{(s)} + \boldsymbol{B}^{(s)} - \tilde{\boldsymbol{B}}^{(s)}
\end{cases}
\tag{8}
$$

For the solutions of $\left\{\boldsymbol{A}^{(s)}, \tilde{\boldsymbol{A}}^{(s)}, \boldsymbol{\Lambda}^{(s)}\right\}$, $\left\{\boldsymbol{B}^{(s)}, \tilde{\boldsymbol{B}}^{(s)}, \boldsymbol{\Gamma}^{(s)}\right\}$ in (7), we can calculate them successively via minimizing $\mathcal{L}$ with respect to one of them while fixing the others. Note that the primal variable $\boldsymbol{A}^{(s)}$ and auxiliary variable $\tilde{\boldsymbol{A}}^{(s)}$ both include the common and individual patterns, we need to calculate these two patterns separately. Furthermore, since the common pattern $\boldsymbol{A}_C$ (or $\tilde{\boldsymbol{A}}_C$) is shared by $\boldsymbol{A}^{(s)}$ (or $\tilde{\boldsymbol{A}}^{(s)}$), $s = 1, 2, \cdots, S$, we need to combine the information from all matrices from 1 to S to calculate their solutions. Different from the common pattern, the individual pattern $\boldsymbol{A}_I^{(s)}$ or $\tilde{\boldsymbol{A}}_I^{(s)}$ just needs to be calculated separately by the corresponding sth set data. Moreover, we also divide $\boldsymbol{B}^{(s)}$ into two parts $\boldsymbol{B}_C^{(s)} \in \mathbb{R}_+^{L\times J^{(s)}}$ and $\boldsymbol{B}_I^{(s)} \in \mathbb{R}_+^{(R^{(s)}-L)\times J^{(s)}}$ row-wisely. The specific solutions of primal, auxiliary and dual variables are given in (8), in which $\boldsymbol{E}_C \in \mathbb{R}_+^{I^{(l)}\times L}$ and $\boldsymbol{E}_I \in \mathbb{R}_+^{I^{(l)}\times(R^{(l)}-L)}$ are the matrices whose elements are all equal to one. We summarize the GNMF-SR algorithm based on ADMM update (termed as GNMF-SR-ADMM) in **Algorithm 1**.

Algorithm 1: GNMF-SR-ADMM algorithm

Input: $\boldsymbol{X}^{(s)}$, L, and $R^{(s)}$, $s = 1, \cdots, S$
Initialization:
$\boldsymbol{A}^{(s)}$, $\boldsymbol{B}^{(s)}$, $\tilde{\boldsymbol{A}}^{(s)}$, $\tilde{\boldsymbol{B}}^{(s)}$, $\boldsymbol{\Lambda}^{(s)}$, $\boldsymbol{\Gamma}^{(s)}$, $s = 1, \cdots, S$
for $k = 1, \cdots, MAX_k$ **do**
  According to (8);
  Update $\boldsymbol{A}_C$ and $\tilde{\boldsymbol{A}}_C$;
  for $s = 1, \cdots, S$ **do**
    Update $\boldsymbol{A}_I^{(s)}$, $\tilde{\boldsymbol{A}}_I^{(s)}$ and $\boldsymbol{\Lambda}^{(s)}$;
    Let $\boldsymbol{A}^{(s)} = \tilde{\boldsymbol{A}}^{(s)}$;
    Update $\boldsymbol{B}^{(s)}$, $\tilde{\boldsymbol{B}}^{(s)}$ and $\boldsymbol{\Gamma}^{(s)}$;
    Let $\boldsymbol{B}^{(s)} = \tilde{\boldsymbol{B}}^{(s)}$;
  end
  if *stopping criterion is satisfied* **then**
    return
  end
end
Output: $\boldsymbol{A}^{(s)}$, $\boldsymbol{B}^{(s)}$, $s = 1, 2, \cdots, S$

III. Experiment and results

In this section, we provide a comprehensive experiment of synthetic nonnegative fMRI-like data to demonstrate the performance of GNMF-SR-ADMM algorithm. Multiplicative update (MU, [1], [9]), alternating proximal gradient (APG, [4], [22]), alternative least squares (ALS, [3]) and fast hierarchical alternative least squares (fHALS, [2], [23]) are also extended to solve the GNMF-SR model for comparison. In addition, by controlling the values of β and L, three other models including NMF ($\beta = 0$, $L = 0$), NMF-SR ($L = 0$) and GNMF ($\beta = 0$) are also considered in this experiment.

All experiments are carried out with the following computer configurations: CPU: Intel Core i5-7500 @ 3.40Hz 3.41Hz; Memory: 16Gb; System:64-bit Windows 10; Matlab R2016b.

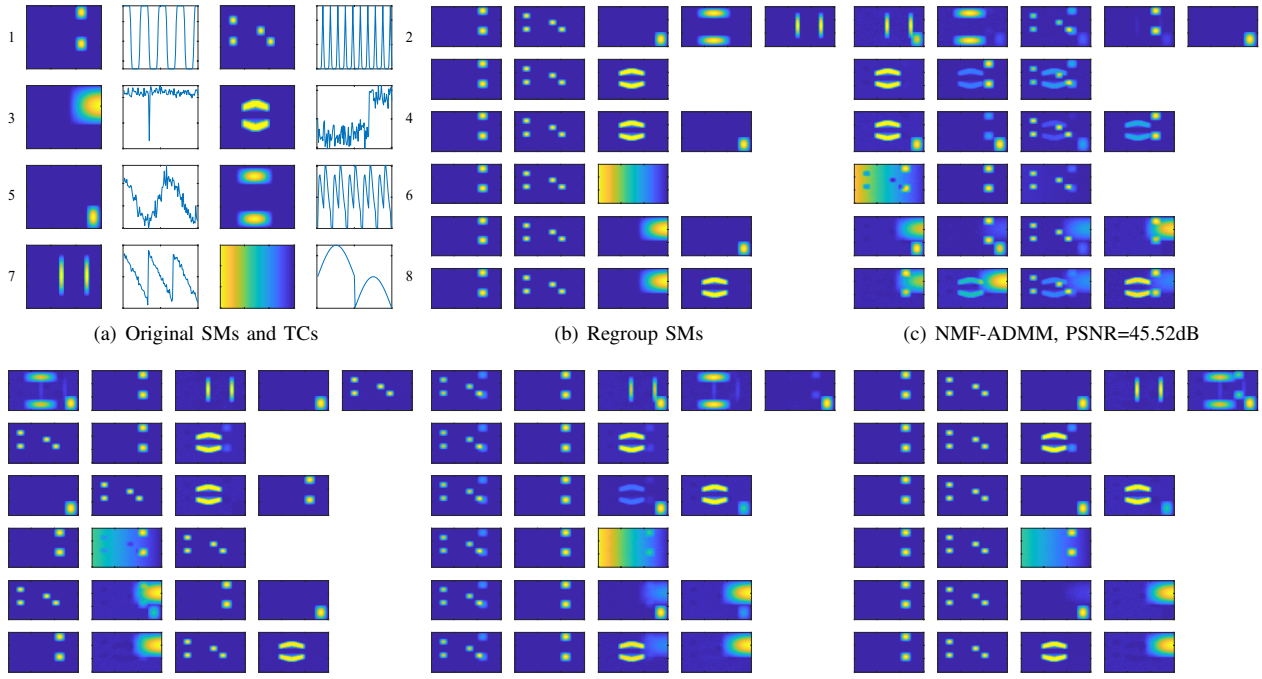
Initialization. For the initialization of factor matrices, we use the uniformly distributed pseudorandom numbers generated by Matlab function `rand`.

Termination criterion. We use the change of relative error [22] (the threshold is set by 10^{-8}), and fix the maximum number of iterations to 1000.

Evaluation index. We adopt peak signal-to-noise ratio (PSNR, [3]) and inter-symbol-interference (ISI, [24]) to evaluate the accuracy of the estimated factor matrices. Meanwhile, we use the values of objective function (Obj), relative error (RelErr) and running time to assess the data fittings.

Data construction. We apply the GNMF-SR model to the joint analysis of multi-subject nonnegative fMRI-like data, which are constructed from the benchmark simulated complex fMRI dataset[1]. The amplitude of spatial maps (SM) and corresponding time courses (TC) are shown in **Fig. 1(a)** and they are adopted to generate the nonnegative fMRI-like data for 6 subjects according to the source index set {1,2,5,6,7}, {1,2,4}, {1,2,4,5}, {1,2,8}, {1,2,3,5} and {1,2,3,4} designed in [24], and more information about data construction can be found in [24]. The SM images of all subjects are shown in **Fig. 1(b)**. Each row corresponds to one subject, and the first two columns are shared by all the subjects, which are considered as the common patterns and the remains are the individual ones.

[1]http://mlsp.umbc.edu/simulated_complex_fmri_data.html



(a) Original SMs and TCs (b) Regroup SMs (c) NMF-ADMM, PSNR=45.52dB

(d) NMF-SR-ADMM, PSNR=61.55dB (e) GNMF-ADMM, PSNR=52.07dB (f) GNMF-SR-ADMM, PSNR=64.31dB

Fig. 1. (a) Amplitude images of 1-8 simulated fMRI-like spatial maps (1st and 3rd columns) and corresponding time courses (2nd and 4th columns). (b-f) SM images of constructed data and that of estimated ones via NMF-ADMM ($\beta = 0$, $L = 0$), NMF-SR-ADMM ($\beta = 3e-4$, $L = 0$), GNMF-ADMM ($\beta = 0$, $L = 2$) and GNMF-SR-ADMM ($\beta = 3e-4$, $L = 2$) under SNR=20dB.

We fix SNR=20dB, and select 25 values for β ranging from 0 to 5. With varying βs, the PSNR curves of SM estimates averaged from 30 Monte Carlo runs in the GNMF-SR model ($L = 0$ & $L = 2$) via MU, ALS, APG, fHALS and ADMM algorithms are shown in **Fig.2**. Note that when $L = 0$ and $\beta = 0$, the GNMF-SR will degenerate into the NMF problem. From **Fig. 2**, we can see that the PSNR values of all algorithms will increase and reach the highest at some point when the sparse penalty parameter β increases, except that MU-based algorithms show the insensitivity to the settings of β between 0 and 5. The sparse penalty will have a negative effect on the algorithm performance when β increases to a certain point. We also present the specific values of PSNR, ISI, Obj, RelErr and running time for each algorithm under $\beta = 0$ and a post-selected β (which corresponds to the best performance) in **Table I**. The performance of the GNMF-based methods is superior to that of NMF-based ones. With sparse regularization, the performance of NMF-based and GNMF-based methods can be both significantly improved. Interestingly, sparse penalty yields better performance improvements than group constraint for NMF-based methods. GNMF-SR-ADMM algorithm achieves the best performance, followed by GNMF-SR-APG, NMF-SR-ADMM and GNMF-SR-fHALS algorithms. However, ADMM-based methods are more time consuming and will be improved in our future work.

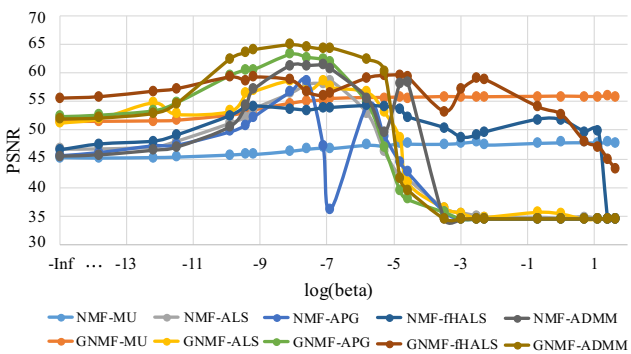


Fig. 2. Mean PSNR of SM estimates for 6 subjects under NMF-SR ($L = 0$) and GNMF-SR ($L = 2$) models with the βs of 25 values varying from 0 to 5, SNR=20dB.

Furthermore, the SM images estimated via NMF-ADMM, NMF-SR-ADMM, GNMF-ADMM and GNMF-SR-ADMM at $\beta = 0$, $3e-4$ and $L = 0$, 2 are shown in **Fig.1(c-f)**. It can be clearly seen that some of SM images obtained by NMF-ADMM and GNMF-ADMM algorithms are blurred with shadows or small outliers. By imposing adequate sparse regularization, those blurs are basically eliminated in the results of NMF-SR-ADMM and GNMF-SR-ADMM algorithms. Moreover, from **Fig. 1(e-f)**, we can denote that two group analysis methods including GNMF-ADMM and GNMF-SR-ADMM can extract both the common and individual patterns for all the datasets, and also successfully correct the disorder scenario of common patterns in the results of two NMF-based algorithms as shown in **Fig. 1(c-d)**.

TABLE I
PERFORMANCE COMPARISON ON FMRI_LIKE DATA BASED ON GNMF-SR MODEL ($L = 0, 2$, SNR=20DB)

	Method	β	PSNR	ISI	Obj	RelErr	Time/s
$L = 0$	ALS	0	46.68	0.0545	0.0103	0.3332	3.6941
		1e-3	58.70	0.0117	0.0103	0.3335	6.2381
	MU	0	45.23	0.0741	0.0102	0.3312	5.3236
		8e-2	48.02	0.0645	0.0102	0.3318	5.2624
	APG	0	45.56	0.0859	0.0101	0.3303	6.8339
		5e-4	58.61	0.0303	0.0105	0.3367	3.7993
	fHALS	0	46.50	0.0753	0.0101	0.3298	6.1396
		3e-3	54.24	0.0281	0.0101	0.3308	6.1959
	ADMM	0	45.47	0.0876	0.0101	0.3303	7.0635
		3e-4	61.27	0.0128	0.0103	0.3330	7.1136
$L = 2$	ALS	0	54.33	0.0393	0.0197	0.4245	2.9618
		3e-4	58.69	0.0152	0.0149	0.3789	4.3242
	MU	0	51.43	0.0330	0.0103	0.3340	5.2181
		4	55.91	0.0344	0.0104	0.3352	5.2229
	APG	0	52.37	0.0258	0.0103	0.3336	3.6041
		5e-4	62.70	0.0215	0.0108	0.3412	5.0651
	fHALS	0	55.58	0.0376	0.0109	0.3387	4.9561
		8e-3	59.65	0.0083	0.0103	0.3335	5.7812
	ADMM	0	51.95	0.0302	0.0103	0.3335	7.2444
		3e-4	64.94	0.0062	0.0104	0.3359	7.1910

IV. CONCLUSION

In this paper, we presented a group nonnegative matrix factorization with sparse regularization (GNMF-SR) model for the group analysis of data from multiple sources. The alternating optimization and alternating direction method of multipliers (ADMM) were developed in combination to optimize the GNMF-SR model, in which the common and individual patterns can be simultaneously extracted while aligning the common patterns. The experiment of simulated fMRI-like data demonstrated that the ADMM-based algorithms have better performance, and imposing group constraint and sparse penalty can greatly improve the performance of NMF-based algorithms.

REFERENCES

[1] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788, 1999.

[2] Andrzej Cichocki and Anh-Huy Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 3, pp. 708–721, 2009.

[3] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*, John Wiley & Sons, 2009.

[4] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan, "Nenmf: An optimal gradient method for nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, 2012.

[5] Xiao-Feng Gong, Xiu-Lin Wang, and Qiu-Hua Lin, "Generalized non-orthogonal joint diagonalization with lu decomposition and successive rotations," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1322–1334, 2015.

[6] Guoxu Zhou, Qibin Zhao, Yu Zhang, Tülay Adalı, Shengli Xie, and Andrzej Cichocki, "Linked component analysis from matrices to high-order tensors: Applications to biomedical data," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 310–331, 2016.

[7] Xun Chen, Z Jane Wang, and Martin McKeown, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 86–107, 2016.

[8] Xiulin Wang, Wenya Liu, Petri Toiviainen, Tapani Ristaniemi, and Fengyu Cong, "Group analysis of ongoing eeg data based on fast double-coupled nonnegative tensor decomposition," *Journal of neuroscience methods*, vol. 330, pp. 108502, 2020.

[9] Hyekyoung Lee and Seungjin Choi, "Group nonnegative matrix factorization for EEG classification," in *Artificial Intelligence and Statistics*, 2009, pp. 320–327.

[10] Bonggun Shin and Alice Oh, "Bayesian group nonnegative matrix factorization for EEG analysis," *arXiv preprint arXiv:1212.4347*, 2012.

[11] Patrik O Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of machine learning research*, vol. 5, no. Nov, pp. 1457–1469, 2004.

[12] Jun Xu, Lei Xiang, Guanhao Wang, Shridar Ganesan, Michael Feldman, Natalie NC Shih, Hannah Gilmore, and Anant Madabhushi, "Sparse non-negative matrix factorization (snmf) based color unmixing for breast histopathological image analysis," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 20–29, 2015.

[13] Morten Mørup and Lars Kai Hansen, "Tuning pruning in sparse non-negative matrix factorization," in *2009 17th European Signal Processing Conference*. IEEE, 2009, pp. 1923–1927.

[14] Dennis L Sun and Cedric Fevotte, "Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence," in *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2014, pp. 6201–6205.

[15] Davood Hajinezhad, Tsung-Hui Chang, Xiangfeng Wang, Qingjiang Shi, and Mingyi Hong, "Nonnegative matrix factorization using admm: Algorithm and convergence analysis," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 4742–4746.

[16] Kejun Huang, Nicholas D Sidiropoulos, and Athanasios P Liavas, "A flexible and efficient algorithmic framework for constrained matrix and tensor factorization," *IEEE Transactions on Signal Processing*, vol. 64, no. 19, pp. 5052–5065, 2016.

[17] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[18] Davood Hajinezhad and Qingjiang Shi, "Alternating direction method of multipliers for a class of nonconvex bilinear optimization: convergence analysis and applications," *Journal of Global Optimization*, vol. 70, no. 1, pp. 261–288, 2018.

[19] Yu Wang, Wotao Yin, and Jinshan Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.

[20] Evrim Acar, Gözde Gürdeniz, Morten A Rasmussen, Daniela Rago, Lars O Dragsted, and Rasmus Bro, "Coupled matrix factorization with sparse factors to identify potential biomarkers in metabolomics," in *2012 IEEE 12th International Conference on Data Mining Workshops*. IEEE, 2012, pp. 1–8.

[21] Deqing Wang, "Extracting meaningful EEG features using constrained tensor decomposition," *JYU dissertations*, 2019.

[22] Yangyang Xu and Wotao Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.

[23] Xiulin Wang, Tapani Ristaniemi, and Fengyu Cong, "Fast implementation of double-coupled nonnegative canonical polyadic decomposition," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8588–8592.

[24] Xiao-Feng Gong, Lei Mao, Ying-Liang Liu, and Qiu-Hua Lin, "A jacobi generalized orthogonal joint diagonalization algorithm for joint blind source separation," *IEEE Access*, vol. 6, pp. 38464–38474, 2018.

PII

GENERALIZATION OF LINKED CANONICAL POLYADIC TENSOR DECOMPOSITION FOR GROUP ANALYSIS

by

Xiulin Wang, Chi Zhang, Tapani Ristaniemi and Fengyu Cong 2019

16th International Symposium on Neural Networks (ISNN 2019), Moscow, Russia, http://doi.org/10.1007/978-3-030-22808-8_19

Generalization of Linked Canonical Polyadic Tensor Decomposition for Group Analysis

Xiulin Wang[1,2], Chi Zhang[1], Tapani Ristaniemi[2], and Fengyu Cong[1,2](✉)

[1]School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
xiulin.wang@foxmail.com, chizhang@dlut.edu.cn, cong@dlut.edu.cn
[2]Faculty of Information Technology,
University of Jyväskylä, Jyväskylä 40100, Finland
tapani.e.ristaniemi@jyu.fi

Abstract. Real-world data are often linked with each other since they share some common characteristics. The mutual linking can be seen as a core driving force of group analysis. This study proposes a generalized linked canonical polyadic tensor decomposition (GLCPTD) model that is well suited to exploiting the linking nature in multi-block tensor analysis. To address GLCPTD model, an efficient algorithm based on hierarchical alternating least squares (HALS) method is proposed, termed as GLCPTD-HALS algorithm. The proposed algorithm enables the simultaneous extraction of common components, individual components and core tensors from tensor blocks. Simulation experiments of synthetic EEG data analysis and image reconstruction and denoising were conducted to demonstrate the superior performance of the proposed generalized model and its realization.

Key words: Linked tensor decomposition · Hierarchical alternating least squares · Canonical polyadic · Simultaneous extraction

1 Introduction

Linked tensor decomposition (LTD) is an emerging technique for group analysis in recent years, specially designed for simultaneous analysis of multi-block tensor data. It has been successfully applied in the fields of neuroscience [1], multi-dimensional harmonic retrieval [2], array signal processing [3] and metabolic physiology [4].

Linked tensor decomposition can be seen as an extension of tensor decomposition applied to single-block tensor [5–7] in multi-block data analysis, e.g., analysis of electrophysiological (EEG) data collected from different subjects under a certain stimulus, which can be naturally linked together for sharing the similar brain activities [1]. LTD method can take full advantage of such linking/coupling information among data blocks to improve the decomposition identifiability [3]. In addition, LTD method has its advantage in imposing constraints on particular modes or components compared to its matrix counterpart [9, 10].

Any combination of constraints including independence, sparsity, smoothness and non-negativity can be added more easily and flexibly [11]. Moreover, imposing specific constraints on different modes or components would contribute to obtaining more reasonable decomposition solutions with convincing interpretations [6, 8, 11]. For example, the constraint of non-negativity is applied in the processing of ERP data with time-frequency representation [6]. Furthermore, tensor decomposition is superior to two-way matrix factorization such as solution uniqueness and component identification in some cases [12]. To unfold some of the modes in matrix factorization will inevitably loss the potential interactions under the multiway structure [13]. Therefore, it is reliable to take the high-order characteristics of tensors into consideration in data analysis.

With the LTD model, simultaneous extraction of common components, individual components and core tensors can be obtained. The notion 'linked' is based on the assumption that different data blocks share the same or highly correlated components while retaining individual information [14]. In group data analysis, e.g. face images collected from different subjects with the same expression [14], or EEG data collected from different participants under the same stimulus [8], all subjects may share the similar or even identical information, which can be regarded as linking factors among tensors. However, individual characteristics will exist in particular subjects at the same time, which may lead to inconsistent number of components for tensors. Obviously, this inconsistency does not match the linked canonical polyadic tensor decomposition (LCPTD) model in [14]. Therefore, this study aims to develop a more generalized and flexible model with inconsistent component number for linked tensor decomposition. To obtain the solution of the new model, we propose a generalized linked canonical polyadic tensor decomposition algorithm based on HALS strategy [7], which is termed as GLCPTD-HALS algorithm. The experiment results show that the generalized model is more practical in multi-block data analysis, and its realization can achieve better performance.

This paper is organized as follows. Section 2 introduces LCPTD model and its generalization. In section 3, GLCPTD-HALS algorithm is proposed. In section 4, simulation experiments are conducted to verify the performance of proposed algorithm. The last section summarizes this paper.

2 Problem Formulation

In this section, we mainly introduce the linked canonical polyadic tensor decomposition (LCPTD) model [14] and its generalization. CP model [15] is also called parallel factor analysis (PARAFAC) [16] and canonical composition (CANDECOMP) [17]. CP decomposition (CPD) can decompose a tensor into a minimal number of rank-1 tensors, and the minimum number R is termed as the rank of a tensor. It can achieve good unique identification under some mild conditions without any special constraints. Please refer to [18] for a detailed description of standard notations and basic tensor operations.

2.1 Review of LCPTD Model

To deal with multi-block tensors with coupling information, researchers in [14] proposed a model of simultaneous decomposition, namely LCPTD model, which is defined as follows:

$$\begin{aligned}\underline{\boldsymbol{X}}^{(s)} \approx \underline{\hat{\boldsymbol{X}}}^{(s)} &= \sum_{r=1}^{R} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \\ &= \left[\!\left[\underline{\boldsymbol{G}}^{(s)}; \boldsymbol{U}^{(1,s)}, \boldsymbol{U}^{(2,s)}, \cdots, \boldsymbol{U}^{(N,s)} \right]\!\right],\end{aligned} \tag{1}$$

where $\underline{\boldsymbol{X}}^{(s)} \in \Re^{I_1 \times I_2 \times \cdots I_N}$ and $\underline{\hat{\boldsymbol{X}}}^{(s)} \in \Re^{I_1 \times I_2 \times \cdots I_N}$ denote the original and estimated tensors, respectively. $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{u}_1^{(n,s)}, \boldsymbol{u}_2^{(n,s)}, \cdots, \boldsymbol{u}_R^{(n,s)}\right] \in \Re^{I_n \times R}$ denotes the n-mode factor matrix of sth tensor. S, R, N are denoted as the number, rank and order of tensors, respectively. $\underline{\boldsymbol{G}}^{(s)} \in \Re^{R \times R \times \cdots R}$ denotes the sth core tensor with non-zero entries only on the super-diagonal. $\lambda_r^{(s)}$ is the $(r, r, ..., r)$th element of $\underline{\boldsymbol{G}}^{(s)}$. The LCPTD model assumes that each factor matrix $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{U}_C^{(n)}\ \boldsymbol{U}_I^{(n,s)}\right] \in \Re^{I_n \times R}$ consists of two parts: $\boldsymbol{U}_C^{(n)} \in \Re^{I_n \times L_n}$, $0 \leq L_n \leq R$ and $\boldsymbol{U}_I^{(n,s)} \in \Re^{I_n \times (R - L_n)}$. The former shared by all tensor blocks represents the coupling (same or highly correlated) information, whereas the latter corresponds to the individual characteristics of each tensor block.

2.2 Generalization of LCPTD Model

Even though multiple data blocks are collected under the same condition, individual characteristics will exist in the particular blocks due to the individual differences. These characteristics may lead to inconsistent number of components for tensors. Obviously, this inconsistency does not match the LCPTD model. Therefore, we extend the LCPTD model to the generalized case with different component number $R^{(s)}$, termed as GLCPTD, which is defined as:

$$\begin{aligned}\underline{\boldsymbol{X}}^{(s)} \approx \underline{\hat{\boldsymbol{X}}}^{(s)} &= \sum_{r=1}^{R^{(s)}} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \\ &= \left[\!\left[\underline{\boldsymbol{G}}^{(s)}; \boldsymbol{U}^{(1,s)}, \boldsymbol{U}^{(2,s)}, \cdots, \boldsymbol{U}^{(N,s)} \right]\!\right].\end{aligned} \tag{2}$$

The generalized LCPTD model still assumes that each factor matrix $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{U}_C^{(n)}\ \boldsymbol{U}_I^{(n,s)}\right] \in \Re^{I_n \times R^{(s)}}$ consists of two parts: $\boldsymbol{U}_C^{(n)} \in \Re^{I_n \times L_n}$, $0 \leq L_n \leq min(R^{(s)})$ and $\boldsymbol{U}_I^{(n,s)} \in \Re^{I_n \times (R^{(s)} - L_n)}$, representing the same meanings with LCPTD model. $\underline{\boldsymbol{G}}^{(s)} \in \Re^{R^{(s)} \times R^{(s)} \times \cdots R^{(s)}}$ denotes the sth core tensor.

Fig. 1 illustrates the conceptual model of dual-linked tensor decomposition based on CP model (all tensors are linked together by the common parts $\boldsymbol{U}_C^{(1)}$ and $\boldsymbol{U}_C^{(2)}$).

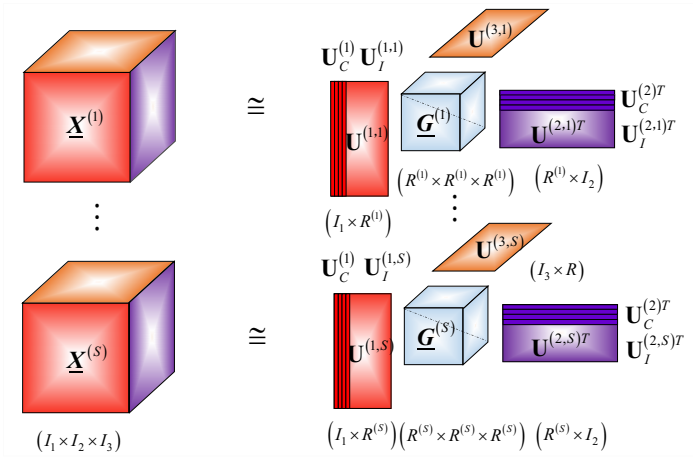


Fig. 1. Conceptual illustration of GLCPTD model with dual-linked parts [11]

3 Realization of GLCPTD Model

In this section, we aim to provide a solution of how to solve the above-mentioned GLCPTD model through HALS strategy [7]. The optimization criterion of squared Euclidean distance minimization is utilized to minimize the error between the original and estimated tensors. Therefore, the cost function can be expressed as:

$$min \sum_{s=1}^{S} \left\| \underline{\boldsymbol{X}}^{(s)} - \sum_{r=1}^{R^{(s)}} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \right\|_F^2 \tag{3}$$

$$s.t. \boldsymbol{u}_r^{(n,1)} = \cdots = \boldsymbol{u}_r^{(n,S)}, \quad r \leq L_n,$$

$$\left\| \boldsymbol{u}_r^{(n,s)} \right\| = 1, n = 1 \cdots N, r = 1 \cdots R^{(s)}, s = 1 \cdots S.$$

The above minimized optimization problem can be transformed into $max(R^{(s)})$ sub-problems via HALS strategy, which can be solved sequentially and iteratively as follows:

$$D_F^{(r)}(\lambda_r^{(s)}, \boldsymbol{u}_r^{(n,s)}) = \sum_{s=1, r \leq R^{(s)}}^{S} \left\| \underline{\boldsymbol{Y}}_r^{(s)} - \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \right\|_F^2, \tag{4}$$

where $\underline{\boldsymbol{Y}}_r^{(s)} \doteq \underline{\boldsymbol{X}}^{(s)} - \sum_{k \neq r}^{R^{(s)}} \lambda_k^{(s)} \boldsymbol{u}_k^{(1,s)} \circ \boldsymbol{u}_k^{(2,s)} \circ \cdots \circ \boldsymbol{u}_k^{(N,s)}$. For the solution of $\boldsymbol{u}_r^{(n,s)}$, we only set the derivative of $D_F^{(r)}(\lambda_r^{(s)}, \boldsymbol{u}_r^{(n,s)})$ with respect to $\boldsymbol{u}_r^{(n,s)}$ to zero. The learning rule of $\boldsymbol{u}_r^{(n,s)}$ can be formulated as:

$$\boldsymbol{u}_r^{(n,s)} = \begin{cases} \sum_s \left(\underline{\boldsymbol{Y}}_{r,(n)}^{(s)} \lambda_r^{(s)} \{\boldsymbol{u}_r^{(s)}\}^{\odot -n} \right) \Big/ \sum_s \left(\lambda_r^{(s)T} \lambda_r^{(s)} \right), & r \leq L_n, \\ \underline{\boldsymbol{Y}}_{r,(n)}^{(s)} \{\boldsymbol{u}_r^{(s)}\}^{\odot -n} \Big/ \lambda_r^{(s)T}, & r > L_n, \end{cases} \tag{5}$$

where $\underline{\boldsymbol{Y}}_{r,(n)}^{(s)}$ is the mode-n matricization of $\underline{\boldsymbol{Y}}_r^{(s)}$. $\{\boldsymbol{u}_r^{(s)}\}^{\odot -n} = \boldsymbol{u}_r^{(N,s)} \odot \cdots \odot \boldsymbol{u}_r^{(n+1,s)} \odot \boldsymbol{u}_r^{(n-1,s)} \odot \cdots \odot \boldsymbol{u}_r^{(1,s)}$ and '$\odot$' denotes the Khatri-Rao product. If $r \leq L_n$, $\boldsymbol{u}_r^{(n,s)}$ will be calculated by combining all tensor information and assigned to each s. Otherwise, it needs to be calculated separately. $\boldsymbol{u}_r^{(n,s)}$ needs to be normalized to unit variance by $\boldsymbol{u}_r^{(n,s)} \leftarrow \boldsymbol{u}_r^{(n,s)}/\|\boldsymbol{u}_r^{(n,s)}\|_2$ in each iteration. After N iterations of $\boldsymbol{u}_r^{(n,s)}$, the $(r, r, ..., r)$th element $\lambda_r^{(s)}$ of core tensors is updated as follows:

$$\lambda_r^{(s)} \leftarrow \underline{\boldsymbol{Y}}_r^{(s)} \times_1 \boldsymbol{u}_r^{(1,s)} \times_2 \boldsymbol{u}_r^{(2,s)} \cdots \times_N \boldsymbol{u}_r^{(N,s)}. \tag{6}$$

These $max(R^{(s)})$ stages are alternatively updated one after another until convergence. In order to impose non-negativity, a simple "half-rectifying" non-linear projection is applied as $\boldsymbol{u}_r^{(n,s)} \leftarrow \|\boldsymbol{u}_r^{(n,s)}\|_+$ or $\lambda_r^{(s)} \leftarrow \|\lambda_r^{(s)}\|_+$ after (5) and (6). We summarize the extended GLCPTD-HALS algorithm in Algorithm 1.

Algorithm 1: GLCPTD-HALS algorithm

Input: $\underline{\boldsymbol{X}}^{(s)}$, L_n and $R^{(s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$
Initialization: $\underline{\boldsymbol{G}}^{(s)}$, $\boldsymbol{U}^{(n,s)}$, $\boldsymbol{u}_r^{(n,s)} \leftarrow \boldsymbol{u}_r^{(n,s)}/\|\boldsymbol{u}_r^{(n,s)}\|_2$
$\underline{\boldsymbol{E}}^{(s)} = \underline{\boldsymbol{X}}^{(s)} - \sum_r^{R^{(s)}} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)}$
while *not convergence* **do**
  for $r = 1, 2, \cdots, max(R^{(s)})$ **do**
    $\underline{\boldsymbol{Y}}^{(s)} = \underline{\boldsymbol{E}}^{(s)} + \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)}$, $r \leq R^{(s)}$, $s = 1, 2, \cdots, S$
    for $n = 1, 2, \cdots, N$ **do**
      update $\boldsymbol{u}_r^{(n,s)}$, $r \leq R^{(s)}$, $s = 1, 2, \cdots, S$ via equation (5)
    end
    update $\lambda_r^{(s)}$, $s = 1, 2, \cdots, S$ via equation (6)
    $\underline{\boldsymbol{E}}^{(s)} = \underline{\boldsymbol{Y}}^{(s)} - \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)}$, $r \leq R^{(s)}$, $s = 1, 2, \cdots, S$
  end
end
Output: $\underline{\boldsymbol{G}}^{(s)}$, $\boldsymbol{U}^{(n,s)}$, $n = 1, ..., N$, $s = 1, ..., S$

4 Simulation Results

4.1 Synthetic EEG Data Analysis

In this part, we synthetically generate three types of factor matrices based on brain activities, respectively presenting topography, waveform and power spectrum, as shown in the Fig. 2 (a)-(c). Through the back projection of factor matrices, four tensor blocks representing four subjects are constructed with the SNR of 10dB, as shown in Fig. 2 (d). SNR refers to the signal-to-noise ratio, which is defined as SNR $= 10\log_{10}(\sigma_s/\sigma_n)$. σ_s and σ_n denote the levels of signal and noise, respectively. To prove the usefulness of GLCPTD model, we set the

number of components for four tensors as {4, 4, 3, 3}. Furthermore, factor matrices of topography and power spectrum consist of two common bases and one or two individual bases ($L_1 = L_2 = 2$), while the components of waveform are completely individual ($L_3 = 0$). The common bases represent that the occipital region in the mid-line and left-hemisphere of four subjects are activated with the alpha oscillations (8∼13Hz).

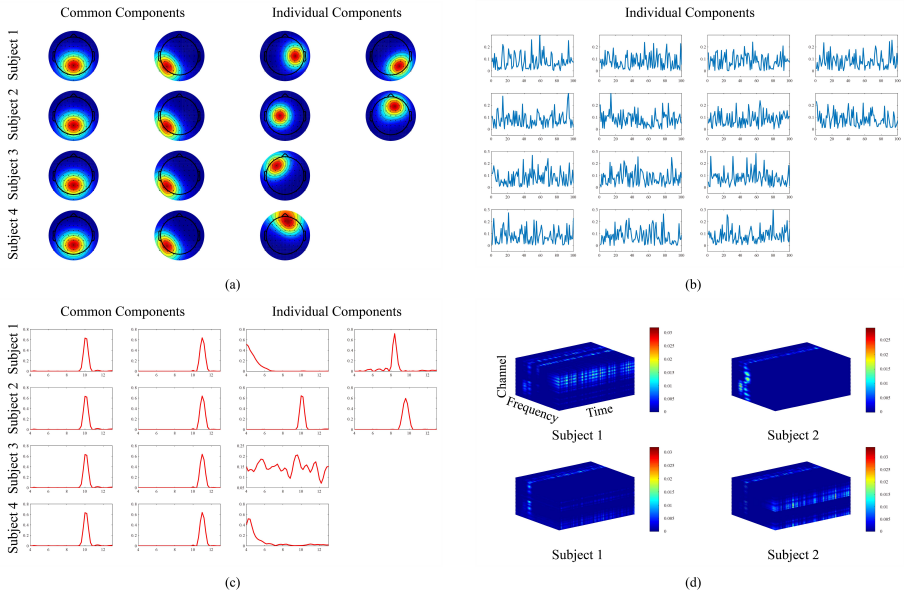


Fig. 2. Illustration of factor matrices of (a) topography, (b) waveform and (c) power spectrum and (d) tensors (frequency×time×channel) for four subjects. Factor matrices of topography and power spectrum for each subject consist of two common components and one/two individual components, while temporal components are individual for each subject.

We apply LCPTD-HALS [14], GLCPTD-HALS, and NTF-HALS [7] algorithms with nonnegative constraint to analyze the four tensor blocks. Solutions of topography learned by these algorithms are shown in Fig. 3 (a)-(d). We can see that, GLCPTD-HALS and NTF-HALS algorithms can successfully extract the common components as well as individual components. The difference is that the components learned by NTF-HALS algorithm are disordered. Clustering and other post-ordering methods need to be applied to obtain the common bases. Although LCPTD-HALS algorithm can extract all the common components, only 3 components are extracted from subject 1 or subject 2 shown in Fig. 3 (c) and 4 components are recovered from subject 3 or subject 4 shown in Fig. 3 (d). The former causes potential components to be omitted (subject 1) or merged (subject 2). The latter depends on the magnitude of the particular component

being redistributed in a certain way driven by algorithm (e.g. the 1st component of subject 3 in Fig. 2 (a) is recovered to the 1st and 4th components in Fig. 3 (d) corresponding to that its magnitude is divided into two parts from predefined 1), which makes group analysis more complicated especially when the number of components increases.

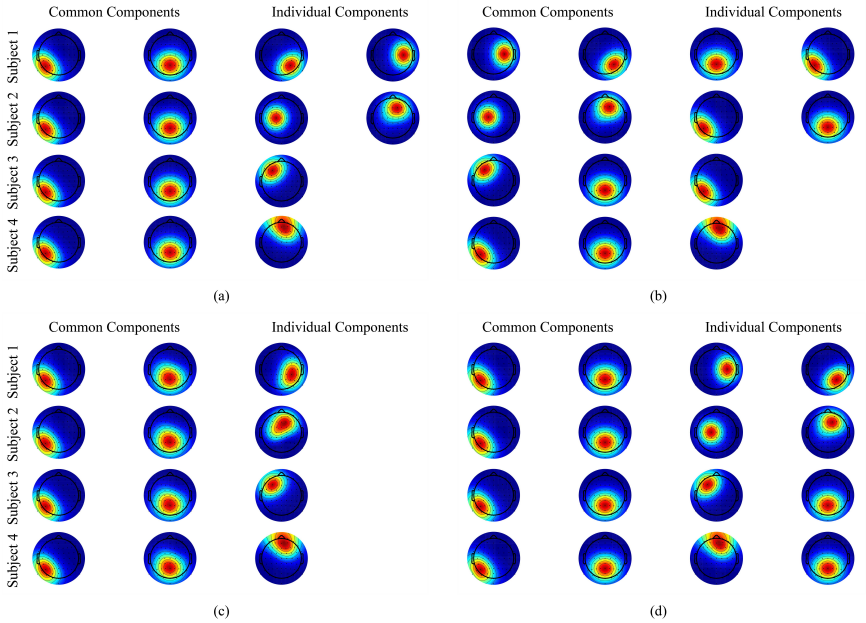


Fig. 3. Illustration of factor matrices of topography for four subjects under four conditions. (a)-(d) show the components learned by GLCPTD-HALS, NTF-HALS, LCPTD-HALS ($R = 3$) and LCPTD-HALS ($R = 4$) algorithms, respectively.

4.2 Image Reconstruction and Denoising

In this part, to examine and demonstrate the performance of the proposed algorithm, we apply the LCPTD and GLCPTD models to image reconstruction and denoising. There are 165 gray-scale images from 15 individuals in the Yale face database. Each individual has 11 images of different face expressions ('centerlight', 'glasses', 'happy', 'leftlight', 'noglasses', 'normal', 'rightlight', 'sad', 'sleepy', 'surprised', 'wink') , and the size of each image is 215×171 pixels. We construct the multi-block tensors by stacking corresponding face images under two conditions: (1) Face images from the same subject with different expressions, $I_1 = 215$, $I_2 = 171$, $I_3 = 11$, $S = 15$; (2) Face images from different subjects with the same expression, $I_1 = 215$, $I_2 = 171$, $I_3 = 15$, $S = 11$. Furthermore, 5% salt-

and-pepper noises are added to all face images. We use the peak-signal-to-noise ratio (PSNR) to measure the quality of reconstructed images.

In terms of the number of components in each tensor, we set R to 40 in the LCPTD model, which is consistent with the original parameter in [14]. Differently, in the GLCPTD model, we use the following method to calculate it: we concatenate each tensor along the first mode to generate a matrix, and perform principle component analysis (PCA) on the matrices successively; when the percentage of the total variance explained by each principle component is greater than 99.6%, the number of corresponding principle components is chosen as the number of components. In this experiment, we assume that the coupling information exists in two modes of images so that we set the number of coupled components to $L_1 = L_2$, $L_3 = 0$, and the values of $L_{1,2}$ are changed in {10, 20, 30}.

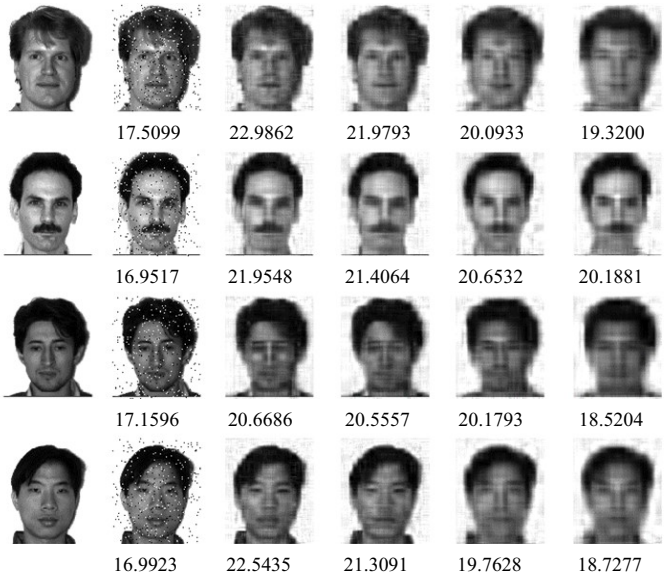


Fig. 4. Original, noisy and reconstructed face images of 'centerlight' from four subjects with PSNRs (dB). 1st column: original images, 2nd column: noisy images, 3rd column: GLCPTD model of condition I, 4th column: LCPTD model of condition I, 5th column: GLCPTD model of condition II, 6th column: LCPTD model of condition II.

By performing the LCPTD-HALS [14] and GLCPTD-HALS algorithms with nonnegative constraint on the above two models, we can compute the PSNRs of reconstructed images. Fig. 4 depict the original, noisy and reconstructed face images from subject 1-4 with the same expression of 'centerlight' ($I_{1,2} = 10$). We can see that the images reconstructed by LPCTD model/condition II are more fuzzier or distorted than those from GLCPTD model/condition I. Table 1

shows the averaged PSNRs of the reconstructed images under the two conditions. The PSNRs obtained by GLCPTD model are higher than those obtained by the LCPTD model in both conditions. It can be considered that the proposed GLCPTD model matches the real-world data more closely, which may make it more practical in real-world data analysis. The PSNRs obtained under condition I are higher than those under condition II, which means that it is more reliable to stack face images from the same subject with different expressions together. It seems that if the number of common components becomes larger, the PSNRs become smaller. The excessive number of common components may affect the fitness of the estimated tensors. However, the selection of parameter L_n is still an open issue in the current study, which will be one of our future works.

Table 1. Averaged PSNRs (dB) of reconstructed images

	Condition I			Condition II		
	$L_1, L_2 = 10$	20	30	$L_1, L_2 = 10$	20	30
LCPTD	21.3651	20.7517	19.9021	19.0809	18.8694	18.5321
GLPCTD	22.0421	21.5476	20.8134	19.9649	19.6537	19.4444

5 Conclusion

The main objective of this paper is to develop a generalized and flexible model of linked tensor decomposition which is more suitable for group analysis. We proposed the generalized LCPTD model as well as its realization, in which the common components, individual components and core tensors can be extracted simultaneously. Experiments of synthetic EEG data analysis and image reconstruction and denoising were conducted to compare the performance of proposed algorithm with LCPTD-HALS and NTF-HALS algorithms. The results illustrated the superior performance of the newly generalized model and its realization.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No. 81471742), the Fundamental Research Funds for the Central Universities [DUT16JJ(G)03] in Dalian University of Technology in China, and the scholarships from China scholarship Council (No. 201706060262).

References

1. Zhou, G.-X., Zhao, Q.-B., Zhang, Y., et al.: Linked component analysis from matrices to high-order tensors: applications to biomedical data. Proc. IEEE. **104**(2), 310–331 (2016)

2. Sorensen, M., De Lathauwer, L.: Multidimensional harmonic retrieval via coupled canonical polyadic decomposition — part II: algorithm and multirate sampling. IEEE Trans. on Signal Process. **65**(2), 528–539 (2017)
3. Gong, X.-F., Lin, Q.-H., Cong, F.-Y., De Lathauwer, L.: Double coupled canonical polyadic decomposition for joint blind source separation. IEEE Trans. on Signal Process. **66**(13), 3475–3490 (2016)
4. Acar, E., Bro, R., Smilde, A.-K.: Data fusion in metabolomics using coupled matrix and tensor factorizations. Proc. IEEE. **103**(9), 1602–1620 (2015)
5. Zhou, G.-X., Cichocki, A., Xie, S.-L.: Fast nonnegative matrix/tensor factorization based on low-rank approximation. IEEE Trans. on Signal Process. **60**(6), 2928–2940 (2012)
6. Cong, F.-Y., Zhou, G.-X., Cichocki, A., et al.: Low-rank approximation based non-negative multi-way array decomposition on event-related potentials. Int. J. Neural Syst. **24**(8), 1440005 (2014)
7. Cichocki, A., Zdunek, R., Amari, S.: Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In: 7th International Conference on Independent Component Analysis and Signal Separation, pp. 169–176. Springer, London (2007)
8. Cong, F.-Y., Phan, A.-H., Zhao, Q.-B., et al.: Analysis of ongoing EEG elicited by natural music stimuli using nonnegative tensor factorization. In: 20th European Signal Processing Conference, pp. 494–498. Elsevier, Bucharest (2012)
9. Calhoun, V.-D., Liu, J., Adali, T.: A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage **45**(1), 163–172 (2009)
10. Gong, X.-F., Wang, X.-L., Lin, Q.-H.: Generalized non-orthogonal joint diagonalization with LU decomposition and successive rotations. IEEE Trans. on Signal Process. **63**(5), 1322–1334 (2015)
11. Cichocki, A.: Tensor decompositions: a new concept in brain data analysis? arXiv Prepr. arXiv1305.0395 (2013)
12. Mørup, M.: Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. **1**(1), 24–40 (2011)
13. Cong, F.-Y., Lin, Q.-H., Kuang, L.-D., et al.: Tensor decomposition of EEG signals: A brief review. J. Neurosci. Methods **248**, 5969 (2015)
14. Yokota, T., Cichocki, A., Yamashita, Y.: Linked PARAFAC / CP tensor decomposition and its fast implementation for multi-block tensor analysis. In: 19th International Conference on Neural Information Processing, pp. 84–91. Springer, Doha (2012)
15. Hitchcock, F.-L.: The expression of a tensor or a polyadic as a sum of products. J. Math. Phys. **6**(1-4), 164–189 (1927)
16. Harshman, R.-A.: Foundations of the PARAFAC procedure: models and conditions for an explanatory multimodal factor analysis. UCLA Work. Pap. Phonetics. **16**, 1–84 (1970)
17. Carroll, J.-D., Chang, J.-J.: Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young decomposition. Psychometrika **35**(3), 283–319 (1970)
18. Kolda, T.-G., Bader, B.-W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2008)

PIII

FAST IMPLEMENTATION OF DOUBLE-COUPLED NONNEGATIVE CANONICAL POLYADIC DECOMPOSITION

by

Xiulin Wang, Tapani Ristaniemi and Fengyu Cong 2019

2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, http://doi.org/10.1109/ICASSP.2019.8682737

FAST IMPLEMENTATION OF DOUBLE-COUPLED NONNEGATIVE CANONICAL POLYADIC DECOMPOSITION

Xiulin Wang[1,2] *Tapani Ristaniemi*[2] *Fengyu Cong*[1,2,*]

[1] School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China
[2] Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland

ABSTRACT

Real-world data exhibiting high order/dimensionality and various couplings are linked to each other since they share some common characteristics. Coupled tensor decomposition has become a popular technique for group analysis in recent years, especially for simultaneous analysis of multi-block tensor data with common information. To address the multi-block tensor data, we propose a fast double-coupled nonnegative Canonical Polyadic Decomposition (FDC-NCPD) algorithm in this study, based on the linked CP tensor decomposition (LCPTD) model and fast Hierarchical Alternating Least Squares (Fast-HALS) algorithm. The proposed FDC-NCPD algorithm enables simultaneous extraction of common components, individual components and core tensors from tensor blocks. Moreover, time consumption is greatly reduced without compromising the decomposition quality when handling large-scale tensor blocks. Simulation experiments of synthetic and real-world data are conducted to demonstrate the superior performance of the proposed algorithm.

Index Terms— Tensor decomposition, coupled tensor decomposition, Hierarchical Alternating Least Squares (HALS), linked CP tensor decomposition (LCPTD)

1. INTRODUCTION

Tensor decomposition has been successfully applied to an ensemble of disciplines including blind source separation, signal processing, classification, data mining and neuroscience [1, 2, 3, 4, 5]. For instance, in EEG data analysis, spatial, temporal and spectral information can be simultaneously considered via tensor decomposition, which in turn provides solutions with convincing physiological or pathological interpretations [5]. However, when it comes to joint analysis of multi-block tensor data, such as multiset or multimodal neurophysiological data fusion [6], conventional methods meet challenges in maintaining feature/component comparability and utilizing coupled information across tensors. Joint analysis of data from different samples can potentially reveal underlying structures and inner-relationships among data [7, 8]. Furthermore, joint analysis can take full advantage of prior information to improve the accuracy and stability of solutions [9]. Therefore, increasing recognition of joint analysis makes coupled tensor decomposition more extensively utilized. Coupled tensor decomposition can jointly analyze the multi-block tensors represented by various samples, meanwhile provide a simultaneous extraction of common components, individual components and core tensors [10]. To date, coupled tensor decomposition has been applied in the fields of neuroscience, multi-dimensional harmonic retrieval, array signal processing and metabolic physiology [8, 11, 12, 13].

Compared with tensor decomposition originally designed for single tensors [1], coupled tensor decomposition can utilize shared information among tensors to improve decomposition identifiability with keeping feature/component comparability. In addition, coupled tensor decomposition has the advantage of imposing constraints on particular modes or components compared to its matrix counterparts [14]. Any combination of constraints, including independence, sparsity, orthogonality and non-negativity, can be added more easily and flexibly [14]. Moreover, it is more reliable to consider the high-order feature of tensors in data analysis [5].

In the analysis of EEG data collected from multiple subjects under the same stimulus, the time consumed by the depletion process in the coupled algorithms (assuming that shared information exists in both modes of space and frequency) would go extremely heavy, due to the high-order, high-dimensional and nonnegative nature of EEG data. To address this problem, we propose a fast double-coupled nonnegative Canonical Polyadic Decomposition (FDC-NCPD) algorithm. This algorithm is based on linked CP tensor decomposition (LCPTD) model [10] and fast Hierarchical

This work was supported by the National Natural Science Foundation of China (Grant No. 81471742), the Fundamental Research Funds for the Central Universities [DUT16JJ(G)03] in Dalian University of Technology in China, and the scholarships from China scholarship Council (No. 201706060262). The authors would like to thank Dr. P. Toiviainen and Dr. M. Huotilainen for the use of EEG data in Experiment 4.2.

* Corresponding author: cong@dlut.edu.cn

Alternating Least Squares (Fast-HALS) algorithm [15]. With the proposed algorithm, time consumption is greatly reduced without losing decomposition quality in the analysis of large-scale tensor blocks.

The rest of this paper is organized as follows. Section 2 introduces the LCPTD model. In section 3, a fast double-coupled implementation of LCPTD model is proposed. In section 4, simulation experiments are conducted to verify the performance of the proposed algorithm. The last section concludes this paper.

2. REVIEW OF LCPTD MODEL

To deal with multi-block tensors with coupling information, researchers in [10] proposed a generalized model of simultaneous decomposition, namely LCPTD model, which is defined as follows:

$$\begin{aligned}\underline{\boldsymbol{X}}^{(s)} \approx \underline{\hat{\boldsymbol{X}}}^{(s)} &= \sum_{r=1}^{R} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \\ &= [\![\underline{\boldsymbol{G}}^{(s)}; \boldsymbol{U}^{(1,s)}, \boldsymbol{U}^{(2,s)}, \cdots, \boldsymbol{U}^{(N,s)}]\!]\end{aligned} \tag{1}$$

$$\text{s.t.} \left\| \boldsymbol{u}_r^{(n,s)} \right\| = 1, n = 1 \cdots N, r = 1 \cdots R, s = 1 \cdots S,$$

where $\underline{\boldsymbol{X}}^{(s)} \in \Re^{I_1 \times I_2 \times \cdots I_N}$ and $\underline{\hat{\boldsymbol{X}}}^{(s)} \in \Re^{I_1 \times I_2 \times \cdots I_N}$ denote the original and estimated tensors, respectively. $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{u}_1^{(n,s)}, \cdots, \boldsymbol{u}_R^{(n,s)}\right] \in \Re^{I_n \times R}$ denotes the n-mode factor matrix of sth tensor. S, R, N are denoted as the number, rank and order of tensors, respectively. $\underline{\boldsymbol{G}}^{(s)} \in \Re^{R \times R \times \cdots R}$ denotes the sth core tensor with non-zero entries only on the super-diagonal. $\lambda_r^{(s)}$ is the (r, r, r)th element of $\underline{\boldsymbol{G}}^{(s)}$.

The LCPTD model assumes that each factor matrix $\boldsymbol{U}^{(n,s)} = [\boldsymbol{U}_C^{(n)} \ \boldsymbol{U}_I^{(n,s)}] \in \Re^{I_n \times R}$ consists of two parts: $\boldsymbol{U}_C^{(n)} \in \Re^{I_n \times L_n}$, $0 \le L_n \le R$ and $\boldsymbol{U}_I^{(n,s)} \in \Re^{I_n \times (R - L_n)}$. The former shared by all tensor blocks represents the coupling (same or highly correlated) information, whereas the latter corresponds to the individual characteristics of each tensor block.

3. ALGORITHM IMPLEMENTATION

In this section, an optimization criterion of squared Euclidean divergence minimization is used to evaluate the error between the original and estimated tensors. For simplicity, we assume that the element $\lambda_r^{(s)}$ in core tensors can be absorbed into the non-normalized component $\boldsymbol{u}_r^{(N,s)}$. Therefore, the cost function can be expressed in a simplified form as:

$$\min \sum_{s=1}^{S} \left\| \underline{\boldsymbol{X}}^{(s)} - \sum_{r=1}^{R} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \right\|_F^2 \tag{2}$$

$$\text{s.t.} \ \boldsymbol{u}_r^{(n,1)} = \cdots = \boldsymbol{u}_r^{(n,S)} \text{ for } r \le L_n,$$

$$\left\| \boldsymbol{u}_r^{(n,s)} \right\| = 1, n = 1 \cdots N - 1, r = 1 \cdots R, s = 1 \cdots S.$$

The above minimization problem can be transformed into R sub-problems via the HALS algorithm [16], which can be optimized sequentially and iteratively. More exactly, the learning rule of $\boldsymbol{u}_r^{(n,s)}$ can be formulated as follows:

$$\boldsymbol{u}_r^{(n,s)} = \begin{cases} \left[\sum_s \zeta_r^{(n,s)}\right] / \sum_s \gamma_r^{(n,s)}, & r \le L_n, \\ \zeta_r^{(n,s)} / \gamma_r^{(n,s)}, & r > L_n, \end{cases} \tag{3}$$

where $\zeta_r^{(n,s)} = \underline{\boldsymbol{Y}}_{r,(n)}^{(s)} \{\boldsymbol{u}_r^{(s)}\}^{\odot -n}$. $\underline{\boldsymbol{Y}}_{r,(n)}^{(s)}$ is the mode-n matricization of $\underline{\boldsymbol{Y}}_r^{(s)} \doteq \underline{\boldsymbol{X}}^{(s)} - \sum_{p \ne r}^{R} \boldsymbol{u}_p^{(1,s)} \circ \boldsymbol{u}_p^{(2,s)} \circ \cdots \circ \boldsymbol{u}_p^{(N,s)}$. '$\odot$' denotes the Khatri-Rao product. The scaling coefficients $\gamma_r^{(n,s)}$ can be calculated as:

$$\gamma_r^{(n,s)} = \begin{cases} \boldsymbol{u}_r^{(N,s)T} \boldsymbol{u}_r^{(N,s)}, & n \ne N. \\ 1, & n = N. \end{cases} \tag{4}$$

If $r \le L_n$, $\boldsymbol{u}_r^{(n,s)}$ will be calculated by combining all tensor information and assigned to each s. Otherwise, it needs to be calculated separately. The calculation of $\zeta_r^{(n,s)}$ in equation (3) which seems relatively simple may result in rather high computational cost, especially for large-scale problems [15]. To address the above issue, we further introduce the Fast-HALS algorithm [15] to LCPTD model instead of HALS algorithm (LCPTD-HALS algorithm only considered the introduction of HALS algorithm in the LCPTD model [10]). A detailed analysis of HALS and Fast-HALS algorithms can be found in [15, 16]. In the proposed algorithm, $\zeta_r^{(n,s)}$ in equation (3) can be represented as:

$$\zeta_r^{(n,s)} = \left[\underline{\boldsymbol{X}}_{(n)}^{(s)} \{\boldsymbol{U}^{(s)}\}^{\odot -n}\right]_r - \boldsymbol{U}^{(n,s)} \left[\xi_{(n)}^{(s)}\right]_r + \gamma_r^{(n,s)} \boldsymbol{u}_r^{(n,s)}, \tag{5}$$

where $\xi_{(n)}^{(s)} = (\boldsymbol{U}^{(s)T} \boldsymbol{U}^{(s)}) \circledast \oslash (\boldsymbol{U}^{(n,s)T} \boldsymbol{U}^{(n,s)})$. '$\circledast$' and '$\oslash$' are denoted as Hadamard (element-wise) product and element-wise division, respectively. $\boldsymbol{u}_r^{(n,s)}$, $n \ne N$ needs to be normalized to unit variance by $\boldsymbol{u}_r^{(n,s)} \leftarrow \boldsymbol{u}_r^{(n,s)} / \|\boldsymbol{u}_r^{(n,s)}\|_2$ in each iteration. Meanwhile, the denominators of equation (3) can be omitted due to the normalization of $\boldsymbol{u}_r^{(n,s)}$. In order to obtain the nonnegative components, a simple "half-rectifying" nonlinear projection is applied as $\boldsymbol{u}_r^{(n,s)} \leftarrow \|\boldsymbol{u}_r^{(n,s)}\|_+$ after (3). These R stages are updated alternatively one after another until convergence.

In the end, the (r, r, r)th element $\lambda_r^{(s)}$ of core tensors which has been absorbed into $\boldsymbol{u}_r^{(N,s)}$ can be obtained as $\lambda_r^{(s)} = \|\boldsymbol{u}_r^{(N,s)}\|_2$, and $\boldsymbol{u}_r^{(N,s)}$ needs to be normalized as $\boldsymbol{u}_r^{(N,s)} \leftarrow \boldsymbol{u}_r^{(N,s)} / \|\boldsymbol{u}_r^{(N,s)}\|_2$. Compared with the mode-n matricization $\underline{\boldsymbol{Y}}_{r,(n)}^{(s)}$ of $\underline{\boldsymbol{Y}}_r^{(s)}$ in (3), which is performed repeatedly in each iteration, mode-n matricization $\underline{\boldsymbol{X}}_{(n)}^{(s)}$ in (5) only needs to be executed once in the initialization before the iteration. This greatly improves the computational efficiency of the proposed algorithm. With the consideration

that tensors are only assumed to be coupled in two modes (i.d., $L_n = 0$, $n > 2$), this extended fast-HALS-based algorithm with nonnegative and CP constraints is termed as the fast double-coupled nonnegative CP Decomposition (FDC-NCPD). Please refer to [17] for the detailed description of standard notations and basic tensor operations due to the limited length of paper.

4. SIMULATION RESULTS

In this section, simulation experiments of synthetic and real-world ongoing EEG data are provided to illustrate and compare the performance of NTF-HALS [16], NTF-FastHALS [15], LCPTD-HALS [10] and FDC-NCPD algorithms.

The following experiments are done with the following computer configurations; CPU: Intel Core i5-7500 @ 3.40Hz 3.41Hz; Memory: 8.00 Gb; System: 64-bit Windows 10; Matlab R2016b.

4.1. Synthetic data

Data generation. The double-coupled nonnegative tensors with noisy disturbance are generated as follows:

$$\underline{\boldsymbol{X}}^{(s)'} = \sigma_s \frac{\underline{\boldsymbol{X}}^{(s)}}{\|\underline{\boldsymbol{X}}^{(s)}\|_F} + \sigma_n \frac{\underline{\boldsymbol{N}}^{(s)}}{\|\underline{\boldsymbol{N}}^{(s)}\|_F}, \quad s = 1 \cdots S, \tag{6}$$

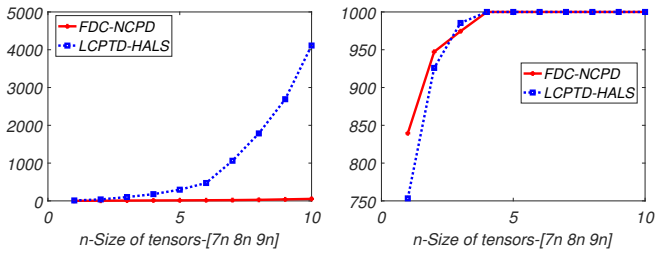
where $\underline{\boldsymbol{X}}^{(s)}$ is constructed as in equation (1). $\underline{\boldsymbol{N}}^{(s)}$ denotes noise term drawn from the standard uniform distribution on the open interval $(0, 1)$. Moreover, σ_s and σ_n represent the levels of signal and noise, respectively. Signal-to-noise ratio (SNR) is defined as $\mathrm{SNR} = 10\log_{10}(\sigma_s/\sigma_n)$.

Evaluation index. The performance index (PI) and Fit are used to evaluate the decomposition quality in the following experiments. More exactly, we use Fit value to measure the tensor reconstruction capability of algorithms, which is defined as: $Fit = \|\underline{\boldsymbol{X}}^{(s)} - \hat{\underline{\boldsymbol{X}}}^{(s)}\|_F / \|\underline{\boldsymbol{X}}^{(s)}\|_F$, where $\hat{\underline{\boldsymbol{X}}}^{(s)}$ is the reconstructed version of $\underline{\boldsymbol{X}}^{(s)}$. We use PI value to evaluate the accuracy of the estimated factor matrices [18].

Termination criteria. The iteration termination criteria for all algorithms are set as $|\mathrm{Fit}_{new} - \mathrm{Fit}_{old}| < \varepsilon, \varepsilon = 1e-6$ but no more than 1000 iterations .

4.1.1. Convergence speed

In this experiment, execution time and iteration number of LCPTD-HALS and FDC-NCPD algorithms are compared against the dimensionality of tensors $I_1 = 7n, I_2 = 8n, I_3 = 9n$ with n varying from 1 to 10. The target tensors are constructed as equation (6) under the noise scenario $\mathrm{SNR} = 20$ dB. The number of components, coupled components and tensors are fixed to $R = 4n$, $L_1 = L_2 = 2n$, $S = 10$, respectively. The performance curves averaged from 30 Monte Carlo runs are illustrated in Fig. 1.



(a) Execution time (b) Number of iterations

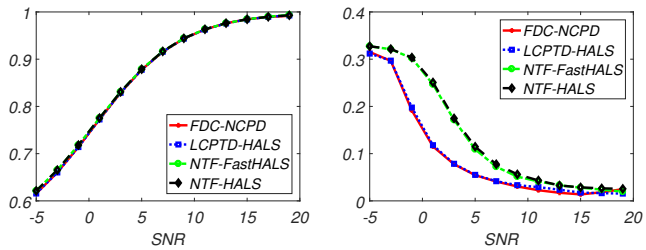
Fig. 1. Averaged execution time and iteration number versus dimensionality of tensors

From Fig.1 (a), we can see that the FDC-NCPD algorithm could greatly reduce the execution time compared with LCPTD-HALS algorithm. This advantage becomes more significant as the dimensionality of tensors increases. Furthermore, in Fig.2 (b), the difference in the number of iterations between two algorithms seems relatively small, especially when $n \geq 4$, which indicates that the FDC-NCPD algorithm outperforms significantly in terms of convergence speed in each iteration. The increase of convergence speed is in parallel with our analysis in section 3, replacing $\underline{\boldsymbol{Y}}^{(s)}_{r,(n)}$ in equation (3) by $\underline{\boldsymbol{X}}^{(s)}_{(n)}$ in equation (5) greatly reduces the calculation time.

4.1.2. Decomposition quality

In this experiment, we compare the Fit and PI performance of FDC-NCPD with LCPTD-HALS and two conventional tensor decomposition algorithms including NTF-HALS and NTF-FastHALS. The noisy double-coupled non-negative tensors are generated as equation (6) under different SNRs from -5 dB to 20 dB with a step size of 2 dB. For the dimensionality of tensors, we set $I_1 = 40, I_2 = 50, I_3 = 60$. The number of components, coupled components and tensors are fixed to $R = 30$, $L_1 = L_2 = 20$ and $S = 10$, respectively. The averaged performance curves obtained from 20 Monte Carlo runs are ploted as in Fig. 2.

As indicated in Fig.2 (a), the four algorithms provide nearly the same Fit performance under all SNRs. In Fig.2 (b), we can see that FDC-NCPD and LCPTD-HALS algorithms show better PI performance than the two conventional algorithms. Moreover, the coupled algorithms obtain equal performance in low SNRs (-5∼7 dB). When SNR is distributed in 7∼16 dB, FDC-NCPD algorithm yields slightly better performance than LCPTD-HALS algorithm. However, when SNR exceeds 17 dB, the proposed algorithm slightly underperforms its competitors. This experiment also verifies that joint/coupled analysis can effectively utilize prior information to improve the decomposition accuracy.



(a) Fit of tensors (b) PI of factor matrices

Fig. 2. Average tensor Fit and PI versus SNR

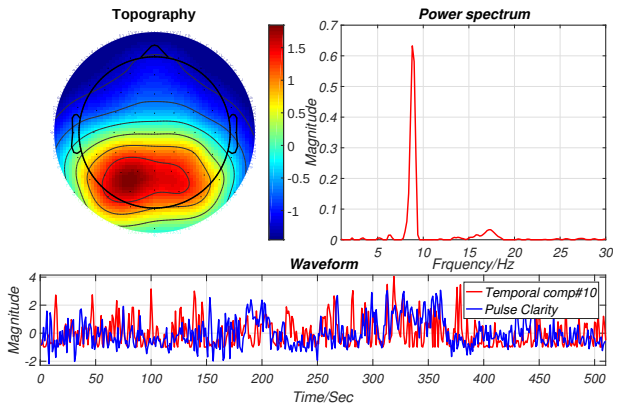


Fig. 3. The 10th temporal component and its corresponding spatial and spectral components from subject #1

4.2. Real-world ongoing EEG data

Data description. In this experiment, we apply the FDC-NCPD and LCPTD-HALS algorithms to the ongoing EEG data, collected from 14 subjects while listening to an 8.5-minute long tango music. We compare the performance of two algorithms in terms of execution time, data Fit and the number of components matched with musical features. The details of data collection, data preprocessing and music feature extraction can be found in [19]. Through short-time Fourier transform (STFT), 14 third-order tensors are formulated with size of $64 \times 146 \times 510$ (64 spatial channels, 146 frequency bins (1∼30Hz) and 510 temporal samples from EEG data of each subject).

Parameter selection. For the selection of the number of components, smoothed DIFFIT [20] is used in this study with suggested $R = 36$. We assume that the coupling information exists both in brain activation areas and frequency oscillations among participants. Thus, the number of coupled components is set as $L_1 = L_2 = 20, L_3 = 0$ (after analyzing the components obtained by conventional method, we found that there were nearly 20 spatial or frequency components among participants with correlations of more than 0.96). Random initializations are used for both factor matrices and core tensors. Termination criteria of algorithms are identical with Experiment 4.1.

Correlation analysis. The temporal, spectral and spatial components can be extracted simultaneously via the FDC-NCPD and LCPTD-HALS algorithms. Correlation analyses are conducted between the temporal courses from EEG data and the temporal courses of musical features, aiming to find the brain activities corresponding to musical stimuli. In addition, we are interested in finding brain components with the significant correlation coefficients(at level $p < 0.05$). The method for determining significant correlation thresholds can be found in [21]. Fig. 3 shows an example of 10th EEG components (topography, spectrum and waveform) extracted from subject #1, in which the temporal course of temporal component is significantly correlated with the temporal course of musical feature termed as 'Pulse Clarity' ($0.1462 > 0.1167$). In addition, the corresponding spatial and spectral components indicate that the posterior area of subject #1 is activated with an alpha oscillation (8∼13Hz) when listening to the tango music.

Table 1. Performance of two algorithms in ongoing EEG data analysis. I-total number of components matched with musical feature, II-execution time, III-data Fit.

	I	II	III
LCPTD-HALS	59.3	76442.65	0.7360
FDC-NCPD	65.6	350.97	0.7353

Results analysis. By 10 times of algorithm executions and correlation analyses, the averaged results of execution time, data Fit and the number of components matched with musical features are illustrated in Table 1. It can be noted that the FDC-NCPD algorithm extracts 6.3 interested components more than the latter on the total number of components. More importantly, the FDC-NCPD algorithm greatly reduces the execution time by nearly 200 times, while yielding equal performance on data Fit (the gap of 0.0007 can be negligible).

5. CONCLUSION

In this study, we introduced the Fast-HALS algorithm to LCPTD model and proposed the FDC-NCPD algorithm, in which the common components, individual components and core tensors can be extracted simultaneously. Simulation experiments of synthetic and real-world data were conducted, showing that the proposed algorithm can significantly reduce time consumption while retaining the decomposition quality. Besides, it can extract a larger number of interested components in the EEG data analysis. In the future studies, we can further analyze brain activation regions and frequency oscillations corresponding to the significantly correlated temporal components.

6. REFERENCES

[1] 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*, John Wiley & Sons, 2009.

[2] N.D. Sidiropoulos, L. De Lathauwer, X. Fu, et al., "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.

[3] 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, no. 1, pp. 37–68, 2010.

[4] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 24–40, 2011.

[5] F.Y. Cong, Q.H. Lin, L.D. Kuang, et al., "Tensor decomposition of eeg signals: a brief review," *Journal of neuroscience methods*, vol. 248, pp. 59–69, 2015.

[6] N.M. Correa, T. Adalı, Y.O. Li, and V.D. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE signal processing magazine*, vol. 27, no. 4, pp. 39–50, 2010.

[7] G.X. Zhou, A. Cichocki, Y. Zhang, and D.P. Mandic, "Group component analysis for multiblock data: Common and individual feature extraction," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2426–2439, 2016.

[8] G.X. Zhou, Q.B. Zhao, Y. Zhang, et al., "Linked component analysis from matrices to high-order tensors: Applications to biomedical data," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 310–331, 2016.

[9] X.F. Gong, X.L. Wang, and Q.H. Lin, "Generalized nonorthogonal joint diagonalization with lu decomposition and successive rotations.," *IEEE Trans. Signal Processing*, vol. 63, no. 5, pp. 1322–1334, 2015.

[10] T. Yokota, A. Cichocki, and Y. Yamashita, "Linked parafac/cp tensor decomposition and its fast implementation for multi-block tensor analysis," in *International Conference on Neural Information Processing*. Springer, 2012, pp. 84–91.

[11] M. Sørensen and L. De Lathauwer, "Multidimensional harmonic retrieval via coupled canonical polyadic decompositionpart ii: Algorithm and multirate sampling," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 528–539, 2017.

[12] X.F. Gong, Q.H. Lin, F.Y. Cong, and L. De Lathauwer, "Double coupled canonical polyadic decomposition for joint blind source separation," *IEEE Transactions on Signal Processing*, vol. 66, no. 13, pp. 3475–3490, 2016.

[13] E. Acar, R. Bro, and A.K. Smilde, "Data fusion in metabolomics using coupled matrix and tensor factorizations," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1602–1620, 2015.

[14] A. Cichocki, "Tensor decompositions: a new concept in brain data analysis?," *arXiv preprint arXiv:1305.0395*, 2013.

[15] A. Cichocki and A.H. Phan, "Fast local algorithms for large scale nonnegative matrix and tensor factorizations," *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 3, pp. 708–721, 2009.

[16] A. Cichocki, R. Zdunek, and S. Amari, "Hierarchical als algorithms for nonnegative matrix and 3d tensor factorization," in *International Conference on Independent Component Analysis and Signal Separation*. Springer, 2007, pp. 169–176.

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

[18] A. Souloumiac, "Joint diagonalization: Is non-orthogonal always preferable to orthogonal?," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2009 3rd IEEE International Workshop on*. IEEE, 2009, pp. 305–308.

[19] F.Y. Cong, V. Alluri, A.K. Nandi, et al., "Linking brain responses to naturalistic music through analysis of ongoing eeg and stimulus features," *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1060–1069, 2013.

[20] D.Q. Wang, Y.J. Zhu, T. Ristaniemi, and F.Y. Cong, "Extracting multi-mode erp features using fifth-order nonnegative tensor decomposition," *Journal of neuroscience methods*, vol. 308, pp. 240–247, 2018.

[21] V. Alluri, P. Toiviainen, I.P. Jääskeläinen, et al., "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *Neuroimage*, vol. 59, no. 4, pp. 3677–3689, 2012.

PIV

GROUP ANALYSIS OF ONGOING EEG DATA BASED ON FAST DOUBLE-COUPLED NONNEGATIVE TENSOR DECOMPOSITION

by

Xiulin Wang, Wenya Liu, Petri Toiviainen, Tapani Ristaniemi and Fengyu Cong
2020

Journal of neuroscience methods, 330, p.108502,
http://doi.org/10.1016/j.jneumeth.2019.108502

Journal of Neuroscience Methods 330 (2020) 108502

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

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

Group analysis of ongoing EEG data based on fast double-coupled nonnegative tensor decomposition

Xiulin Wang[a,b], Wenya Liu[b], Petri Toiviainen[c], Tapani Ristaniemi[b], Fengyu Cong[a,b,*]

[a] School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China
[b] Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland
[c] Finnish Centre of Excellence in Interdisciplinary Music Research, Department of Music, University of Jyväskylä, Jyväskylä, Finland

ARTICLE INFO

Keywords:
Coupled
Music
Nonnegative
Tensor decomposition
Ongoing EEG

ABSTRACT

Background: Ongoing EEG data are recorded as mixtures of stimulus-elicited EEG, spontaneous EEG and noises, which require advanced signal processing techniques for separation and analysis. Existing methods cannot simultaneously consider common and individual characteristics among/within subjects when extracting stimulus-elicited brain activities from ongoing EEG elicited by 512-s long modern tango music.

New method: Aiming to discover the commonly music-elicited brain activities among subjects, we provide a comprehensive framework based on fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm. The proposed algorithm with a generalized model is capable of simultaneously decomposing EEG tensors into common and individual components.

Results: With the proposed framework, the brain activities can be effectively extracted and sorted into the clusters of interest. The proposed algorithm based on the generalized model achieved higher fittings and stronger robustness. In addition to the distribution of centro-parietal and occipito-parietal regions with theta and alpha oscillations, the music-elicited brain activities were also located in the frontal region and distributed in the 4 ~ 11 Hz band.

Comparison with existing method(s): The present study, by providing a solution of how to separate common stimulus-elicited brain activities using coupled tensor decomposition, has shed new light on the processing and analysis of ongoing EEG data in multi-subject level. It can also reveal more links between brain responses and the continuous musical stimulus.

Conclusions: The proposed framework based on coupled tensor decomposition can be successfully applied to group analysis of ongoing EEG data, as it can be reliably inferred that those brain activities we obtained are associated with musical stimulus.

1. Introduction

Listening to music has proven to be an effective strategy to improve and rehabilitate the human health (Koelsch, 2012; MacDonald et al., 2013), especially for people with insomnia, depression, schizophrenia or similar illnesses (Maratos et al., 2008; Mössler et al., 2011; Jespersen et al., 2015). Therefore, revealing brain activities during listening to music has drew an increasing amount of research interest in recent decades (Cong et al., 2012b, 2013a; Wang et al., 2016; Li et al., 2016; Zhu et al., 2019). The advent of brain imaging techniques has provided researchers with the opportunity and insight to probe the brain functions elicited by listening to music. For example, Electroencephalography (EEG) is a collection of potentials along the scalp that reflect electrical activities of the brain. Since Hans Berger first introduced EEG to the world in 1929 (Berger, 1929), it has been widely used in the study of brain functions (Cong et al., 2013a; Huber et al., 2004; Herrmann, 2001) and diagnosis of neurological diseases/disorders (Jeong, 2004; Adeli and Ghosh-Dastidar, 2010; Siuly et al., 2016). Unlike spontaneous EEG recorded in resting state (Berger, 1929) or event-related potentials (ERP) acquired through repeated

* Corresponding author at: School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China.

E-mail addresses: xiulin.wang@foxmail.com (X. Wang), wenyaliu0912@foxmail.com (W. Liu), petri.toiviainen@jyu.fi (P. Toiviainen), tapani.e.ristaniemi@jyu.fi (T. Ristaniemi), cong@dlut.edu.cn (F. Cong).

URLs: http://www.xiulin.wang (X. Wang), http://www.users.jyu.fi/~ptoiviai/ (P. Toiviainen), http://www.users.jyu.fi/~riesta (T. Ristaniemi), http://www.escience.cn/people/cong (F. Cong).

https://doi.org/10.1016/j.jneumeth.2019.108502
Received 5 June 2019; Received in revised form 31 October 2019; Accepted 31 October 2019
Available online 13 November 2019

presentation of stimuli (Luck, 2014), ongoing EEG is a direct response to brain activities in naturalistic and continuous context (e.g. listening to music or watching movies) (Busch et al., 2009; Cong et al., 2012b, 2013a), which makes it possible to study brain functions during real-world experiences. In ongoing EEG experiment, the recorded data can be viewed as mixtures of stimulus-elicited EEG, spontaneous EEG and noises, but how to separate the stimulus-elicited brain activities from ongoing EEG data still remains open for research (Cong et al., 2013a, 2015b; Zhu et al., 2019). Therefore, this study is devoted to the separation and analysis of ongoing EEG data elicited by a 512-s long piece of modern tango music.

In recent years, for the data used in this paper, several studies has been tried to extract the music-elicited brain activities. Cong et al. constructed a fourth-order EEG tensor of channel × frequency × time × subject and the tensor was decomposed using nonnegative tensor factorization (NTF) (Cong et al., 2012b). It should be noted that the analysis of high-order tensors is based on the assumption that the underlying information in temporal, spatial and spectral modes are consistent among subjects (Wang et al., 2018b). However, we found that there was almost no consistent temporal information among subjects as shown in Fig. 1. In Cong et al. (2013a) and Zhu et al. (2019), the authors first adopted independent component analysis (ICA) or spatial ICA to decompose two-way ongoing EEG data represented by each subject and then applied time-frequency analysis and K-means clustering to find spatial, spectral and temporal information of interest. Apparently, such ICA-based approach did not take into account the high-correlation information in space and frequency modes among subjects as shown in Fig. 1, and did not fully utilize the inherent structural information of the tensors represented spatial, temporal and spectral modes (Cong et al., 2015a). In Wang et al. (2016), multilinear partial least squares (PLS) was performed on the tensor (represented by ongoing EEG data) and matrix (represented by musical features), however, it did not consider the nonnegative nature of EEG tensor brought by the time-frequency analysis. With the consideration of phase characteristics, Li et al. applied the ordered Parallel Factors (PARAFAC) algorithm to the ongoing EEG data elicited by the same tango music (Li et al., 2016), but no coupled information among subjects was utilized in the data processing. Tensors, also termed as multi-way arrays, are the higher-order extension of matrices (Kolda and Bader, 2009). Ongoing EEG data can be naturally represented as tensors in which the structural information of inherent interactions between different modes can be fully utilized (Cong et al., 2015a). For example, considering the time-frequency representation of EEG data in each channel, a third-order tensor of channel × time × frequency can be formed (Acar et al., 2007). Tensor decomposition allows for simultaneous consideration of spatial, temporal and spectral information, which provides convincing solutions with physiological or pathological interpretations (Cichocki, 2013). However, when it comes to the analysis of an ensemble of ongoing EEG data (e.g., the data collected from different subjects under the same tango music), it is unreasonable to represent them as a high-order tensor of channel × frequency × time × subject and apply high-order tensor decomposition owing to the incomplete consistency in channel, time and frequency patterns across subjects. Moreover, when analyzing the data through two-way matrix or individual tensor decomposition methods, potential interactions among the multi-way structure of tensors or the coupling information among tensors will inevitably be lost (Cong et al., 2015a; Mørup, 2011).

Coupled tensor decomposition, the extension of tensor decomposition to multi-block tensors, provides a natural framework for the simultaneous analysis of heterogeneous tensors with coupling information (Zhou et al., 2016; Sørensen et al., 2015; Gong et al., 2016). The crucial difference between them is that tensor decomposition processes tensors of × frequency × time individually or a higher-order tensor of channel × frequency × time × subject (generated by stacking tensors from different subjects with the consistent assumption of spatial, temporal and spectral information among the third-order tensors) (Wang et al., 2018b; Mørup et al., 2006; Cong et al., 2012a), while coupled tensor decomposition generalizes tensor decomposition to cover the sharing information across multiple tensor blocks (Sørensen et al., 2015; Gong et al., 2016; Ermiş et al., 2015; Yokota et al., 2012). Compared with its matrix counterparts (Chen et al., 2016; Gong et al., 2015; Calhoun et al., 2009), coupled tensor decomposition can achieve unique solutions and interpretable components, while circumventing the independence constraint (Hunyadi et al., 2017; Mørup, 2011). Given the data collected under the same stimulus, it is reasonable to expect identical or highly correlated stimulus-elicited information among subjects, which can be regarded as a prerequisite for applying coupled tensor decomposition. However, the inter-component similarity among subjects has rarely been considered in previous methods (Cong et al., 2013a; Wang et al., 2016; Li et al., 2016; Zhu et al., 2019). Due to individual differences, individual characteristics in each subject may lead to inconsistent number of components among tensors (Zhou et al., 2016; Ermiş et al., 2015). This inconsistency is not considered in the realization of linked canonical polyadic tensor decomposition (LCPTD) model in Yokota et al. (2012) and Wang et al. (2019). In addition, the time consumption load will be extremely heavy due to the

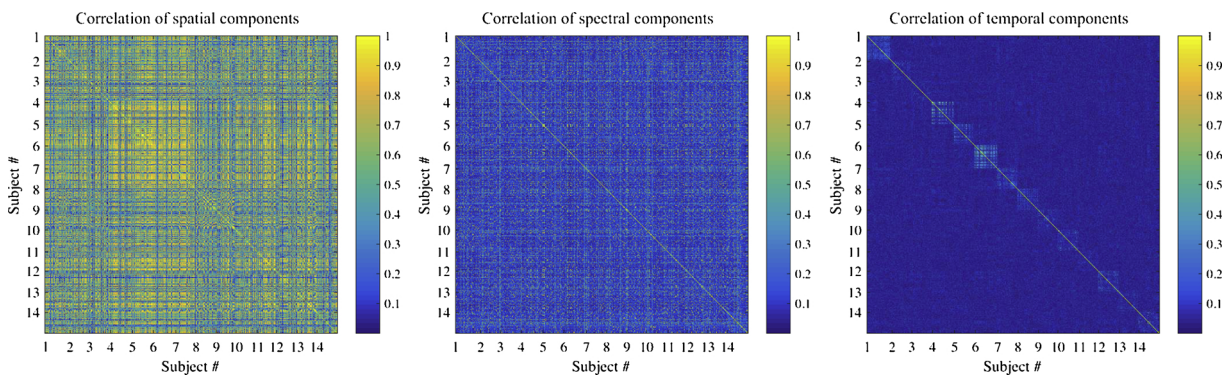


Fig. 1. Inter- and intra-subject correlations of spatial, spectral and temporal components. The spatial (spectral or temporal) components decomposed from ongoing EEG data of 14 subjects by tensor decomposition individually (here we use the fast hierarchical alternative least squares (fast-HALS) algorithm Cichocki and Phan (2009)) are concatenated together, and then the correlation coefficients are calculated.

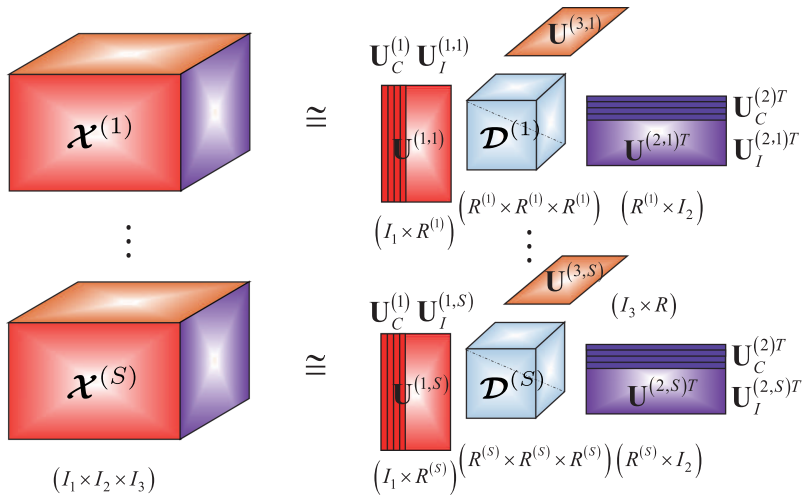


Fig. 2. Conceptual illustration of generalized LCPTD model with double-coupled constraint (adapted from Cichocki (2013)), in which the factor matrices of mode-1 and mode-2 among tensors are partially linked, as they share the same components $\boldsymbol{U}_C^{(1)}$ and $\boldsymbol{U}_C^{(2)}$, respectively.

high-dimensional and nonnegative nature of ongoing EEG data (Wang et al., 2019; Zhou et al., 2012). Our preliminary exploration on the LCPTD model in the application of ongoing EEG data processing has been reported in Wang et al. (2019). Based on the above issues, the specific contributions of this paper can be listed as follows:

First, regarding the individual differences in ongoing EEG data, this study developed a more versatile and flexible model with inconsistent component number in each tensor for coupled tensor decomposition. This model enables the simultaneous decomposition of common components and individual components among tensors.

Second, based on the model mentioned above, this study proposed an efficient data-driven coupled tensor decomposition method termed as fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm.

Third, in order to discover the reliable links between brain responses and musical stimulus, this study proposed a general framework based on coupled tensor decomposition for ongoing EEG data processing and analysis. To the best of our knowledge, this is the first attempt to apply coupled tensor decomposition to the group analysis of ongoing EEG data.

1.1. Why nonnegative and double-coupled constraints?

From the perspective of data analysis, imposing specific constraints on different modes or components during the decomposition process would contribute to obtaining more meaningful solutions (Cichocki, 2013; Wang et al., 2018b). After performing TFR, nonnegative constraint is naturally brought into the EEG data. Correspondingly, the temporal, spectral and spatial components of EEG tensor are all nonnegative, representing the specific physical meanings of time envelope, spectrum and topography, respectively (Wang et al., 2018b).

Given the ongoing EEG data collected under the same stimulus, it is reasonable to expect coupled (identical or highly correlated) components among subjects. Fig. 1 shows the inter- and intra-subject correlations of spatial, spectral and temporal components respectively, which are extracted from the ongoing EEG data of 14 subjects by conventional tensor decomposition individually (i.e., regardless of any coupled information). The detailed information of these data is described in Section 3. We can see that the correlations of some components among subjects in spatial and spectral modes are very significant, while the correlations of temporal components are almost nonexistent. Due to the sparse nature in spectral mode, of course, the correlations of spectral components are not as pronounced as correlations of the spatial components. Therefore, in this study, we consider imposing double coupled constraints in spatial and spectral modes.

2. Fast double-coupled nonnegative tensor decomposition

2.1. Basic notations and mathematical operations

Generally, scalars, vectors, matrices and tensors are respectively denoted by lowercase, boldface lowercase, boldface uppercase and calligraphic boldface uppercase letters, e.g. x, $\boldsymbol{x}$, $\boldsymbol{X}$, $\mathcal{X}$. $\mathbb{R}$ and $\mathbb{R}_+$ denote real number and nonnegative real number. Operators $(\cdot)^T$ and $\|\cdot\|_F$ denote transpose and Frobenius norm, respectively. Outer product, Khatri-Rao product, Hadamard product and element-wise division are denoted by '$\circ$', '$\odot$', '$\circledast$' and '$\oslash$', respectively. Moreover, $\boldsymbol{U}^{(1)} \odot \boldsymbol{U}^{(2)} \cdots \odot \boldsymbol{U}^{(N)}$, $\boldsymbol{U}^{(1)} \odot \cdots \odot \boldsymbol{U}^{(n-1)} \odot \boldsymbol{U}^{(n+1)} \cdots \odot \boldsymbol{U}^{(N)}$, $\boldsymbol{U}^{(1)} \circledast \boldsymbol{U}^{(2)} \cdots \circledast \boldsymbol{U}^{(N)}$ $\boldsymbol{U}^{(1)} \circledast \cdots \circledast \boldsymbol{U}^{(n-1)} \circledast \boldsymbol{U}^{(n+1)} \cdots \circledast \boldsymbol{U}^{(N)}$ are defined as $\{\boldsymbol{U}\}^{\odot}$, $\{\boldsymbol{U}\}^{\odot_{-n}}$, $\{\boldsymbol{U}\}^{\circledast}$, $\{\boldsymbol{U}\}^{\circledast_{-n}}$, respectively. The mode-n matricization of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is termed as $\mathcal{X}_{(n)}$ with the size of $I_n \times (I_1 \cdots I_{n-1} I_{n+1} \cdots I_N)$. Please refer to Kolda and Bader (2009) for a more detailed description of standard notations and basic tensor operations.

2.2. Model generalization

Aiming to process multi-block tensors with coupled information, Yokota et al. proposed the LCPTD model (Yokota et al., 2012), which can enable the simultaneous extraction of common components, individual components and core tensors. This model assumes that those tensors are linked together for sharing some common components. However, even if the tensors are generated under the same conditions, individual differences between them will present as individual characteristics, which may result in inconsistent number of components in

each tensor. This inconsistency was not considered in the LCPTD model. Therefore, in this section, we first extend a generalized LCPTD model of inconsistent component number $R^{(s)}$. Given a set of Nth-order nonnegative tensors $\mathcal{X}^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$, $s = 1, 2, \ldots, S$, the generalized nonnegative LCPTD model can be expressed as:

$$\begin{aligned} \mathcal{X}^{(s)} \approx \hat{\mathcal{X}}^{(s)} &= \sum_{r=1}^{R^{(s)}} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \\ &= \mathcal{D}^{(s)} \times_1 \boldsymbol{U}^{(1,s)} \times_2 \boldsymbol{U}^{(2,s)} \cdots \times_N \boldsymbol{U}^{(N,s)}, \end{aligned} \tag{1}$$

where the tensor $\hat{\mathcal{X}}^{(s)} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ denotes the estimated item of tensor $\mathcal{X}^{(s)}$. $\boldsymbol{u}_r^{(n,s)}$ denotes the rth column of n-mode factor matrix $\boldsymbol{U}^{(n,s)} \in \mathbb{R}_+^{I_n \times R^{(s)}}$ of sth tensor ($s = 1, 2, \ldots, S$, $n = 1, 2, \ldots, N$), and $\boldsymbol{U}^{(n,s)} = [\boldsymbol{u}_1^{(n,s)}, \boldsymbol{u}_2^{(n,s)}, \ldots, \boldsymbol{u}_R^{(n,s)}]$. $\mathcal{D}^{(s)} \in \mathbb{R}_+^{R^{(s)} \times R^{(s)} \cdots \times R^{(s)}}$ represents the sth core tensor with non-zero entries $\lambda_r^{(s)}$ only on the super-diagonal elements $(r, r, \ldots, r)$, $r = 1, 2, \cdots, R^{(s)}$. Most importantly, in generalized LCPTD model, each factor matrix $\boldsymbol{U}^{(n,s)}$ includes two parts: $\boldsymbol{U}_C^{(n)} \in \mathbb{R}_+^{I_n \times L_n}$, $0 \le L_n \le R^{(s)}$ and $\boldsymbol{U}_I^{(n,s)} \in \mathbb{R}_+^{I_n \times (R^{(s)} - L_n)}$. $\boldsymbol{U}_C^{(n)}$ shared by all tensors represents the coupling information among them, while $\boldsymbol{U}_I^{(n,s)}$ corresponds to the individual characteristics of each tensor. Fig. 2 gives the conceptual illustration of generalized double-coupled tensor decomposition model.

2.3. Model realization

In this section, aiming to extract the constrained factor matrices $\boldsymbol{U}^{(n,s)}$ and core tensors $\mathcal{D}^{(s)}$, a solution based on fast-HALS algorithm for the generalized LCPTD model is provided. Note that the scalar factor $\lambda_r^{(s)}$ of core tensor $\mathcal{D}^{(s)}$can be absorbed into one denormalized component, such as $\boldsymbol{u}_r^{(N,s)}$, so the cost function using squared Euclidean distance minimization can be represented in a simplified form as follows:

$$\min \sum_{s=1}^{S} \left\| \mathcal{X}^{(s)} - \sum_{r=1}^{R^{(s)}} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \right\|_F^2 \tag{2}$$

s. t. $\boldsymbol{u}_r^{(n,1)} = \cdots = \boldsymbol{u}_r^{(n,S)}, \quad r \le L_n,$

$\|\boldsymbol{u}_r^{(n,s)}\| = 1,\ n = 1 \cdots N-1,\ r = 1 \cdots R^{(s)},\ s = 1 \cdots S.$

The above optimization problem can be converted into max $(R^{(s)})$ suboptimization problems via HALS algorithm (Cichocki et al., 2007), in which $\boldsymbol{u}_r^{(n,s)}$ can be calculated sequentially and iteratively. To address the issue of high computation cost, we further introduce the fast-HALS algorithm (Cichocki and Phan, 2009) to the proposed model. Therefore, the updating rule of $\boldsymbol{u}_r^{(n,s)}$ can be defined as follows:

$$\boldsymbol{u}_r^{(n,s)} = \begin{cases} \left[\sum_s \zeta_r^{(n,s)}\right] / \sum_s \gamma_r^{(n,s)}, & r \le L_n, \\ \zeta_r^{(n,s)} / \gamma_r^{(n,s)}, & r > L_n, \end{cases} \tag{3}$$

where $[\cdot]_+$ means "half-rectifying" nonlinear projection to obtain non-negative components and $\zeta_r^{(n,s)}$ is defined as:

$$\zeta_r^{(n,s)} = \left[\mathcal{X}_{(n)}^{(s)} \{\boldsymbol{U}^{(s)}\}^{\odot -n}\right]_r - \boldsymbol{U}^{(n,s)} [\xi_{(n)}^{(s)}]_r + \gamma_r^{(n,s)} \boldsymbol{u}_r^{(n,s)} \tag{4}$$

with $\xi_{(n)}^{(s)} = (\boldsymbol{U}^{(s)T} \boldsymbol{U}^{(s)}) \circledast \oslash (\boldsymbol{U}^{(n,s)T} \boldsymbol{U}^{(n,s)})$ and $\mathcal{X}_{(n)}^{(s)}$ is the mode-n matricization of tensor $\mathcal{X}^{(s)}$. The scaling coefficients $\gamma_r^{(n,s)}$ can be formulated as:

$$\gamma_r^{(n,s)} = \begin{cases} \boldsymbol{u}_r^{(N,s)T} \boldsymbol{u}_r^{(N,s)}, & n \ne N. \\ 1, & n = N. \end{cases} \tag{5}$$

In each iteration, we perform the updates of $\boldsymbol{u}_r^{(n,s)}$ with the indexes $n = 1, 2, \ldots, N$, $s = 1, 2, \ldots, S$ and $r = 1, 2, \ldots, R^{(s)}$ successively, while normalizing it to unit variance by $\boldsymbol{u}_r^{(n,s)} \longleftarrow \boldsymbol{u}_r^{(n,s)} / \|\boldsymbol{u}_r^{(n,s)}\|_2$ except $n \ne N$. As illustrated in (3), the calculation of common component $\boldsymbol{u}_r^{(n,s)}$ depends on all tensors and the individual component needs to be calculated separately. These components are alternatively updated one after another until convergence. In this study, considering that those nonnegative tensors represented by ongoing EEG data are assumed to be coupled in spatial and spectral modes (i.d., $L_n = 0$, $n > 2$), the proposed algorithm is termed as fast double-coupled nonnegative tensor decomposition (FDC-NTD) algorithm. We summarize the proposed FDC-NTD algorithm in Algorithm 1, and its detailed derivation is given in the appendix.

Algorithm 1. FDC-NTD algorithm

Input: $\mathcal{X}^{(s)}$, L_1, L_2 and $R^{(s)}$, $s = 1, \cdots, S$
Initialization: $\boldsymbol{U}^{(n,s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$
$\boldsymbol{u}_r^{(n,s)} \leftarrow \boldsymbol{u}_r^{(n,s)} / \|\boldsymbol{u}_r^{(n,s)}\|_2$, $n \ne N$,
$\mathcal{X}_{(n)}^{(s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$
$\boldsymbol{E}^{(s)} = \left(\boldsymbol{U}^{(s)T} \boldsymbol{U}^{(s)}\right)^{\circledast}$, $s = 1, \cdots, S$
while *not convergence* **do**
  for $n = 1, 2, \cdots, N$ **do**
    $\boldsymbol{E}^{(s)} = \boldsymbol{E}^{(s)} \oslash \left(\boldsymbol{U}^{(n,s)T} \boldsymbol{U}^{(n,s)}\right)$, $s = 1, \cdots, S$
    for $r = 1, 2, \cdots, max(R^{(s)})$ **do**
      update $\boldsymbol{u}_r^{(n,s)}$, $r \le R^{(s)}$, $s = 1, 2, \cdots, S$
      via equation (3)
    end
    $\boldsymbol{E}^{(s)} = \boldsymbol{E}^{(s)} \circledast \left(\boldsymbol{U}^{(n,s)T} \boldsymbol{U}^{(n,s)}\right)$, $s = 1, \cdots, S$
  end
end
Output: $\boldsymbol{U}^{(n,s)}$, $n = 1, \ldots, N$, $s = 1, \ldots, S$

3. Experiments and methods

In this section, we provide a comprehensive framework for ongoing EEG data processing and analysis based on the proposed FDC-NTD algorithm, aiming to find commonly appearing brain activities elicited by naturalistic and continuous musical stimulus. Undoubtedly, such common information shared by the majority of subjects is more reliable than individual information from particular subject (Cong et al., 2013a; Lee and Choi, 2009). Through TFR and FDC-NTD algorithm, tensors with dimensions of channel × time × frequency can be constructed and decomposed into common and individual components in the spatial, spectral and temporal modes. Meanwhile, five long-term musical features can be extracted from the music. Correlation analysis and hierarchical clustering are performed together to determine the cluster of interest. Fig. 3 illustrates the overall flow diagram of ongoing EEG data processing and analysis.

3.1. Data acquisition & preprocessing

The ongoing EEG data collected from 14 participants aged from 20 to 46 years old were used in this study. All participants were right-handed and healthy, without musical expertise and any problem of hearing loss or history of neurological disorders. The musical stimulus adopted an 8.5-min piece of modern tango by Astor Piazzolla (Alluri et al., 2012). The data were recorded using BioSemi bioactive electrode caps with the sampling rate of 2048 Hz, according to the international 10–20 system. The collected EEG data were preprocessed off-line using

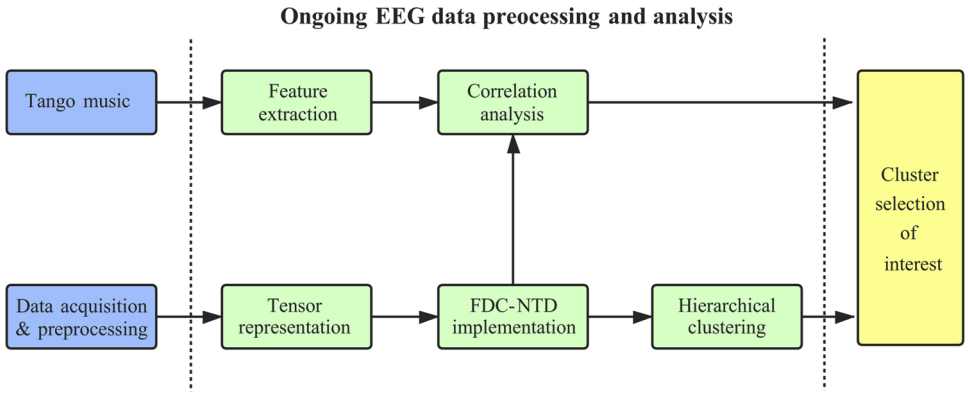


Fig. 3. FDC-NTD-based ongoing EEG data analysis includes the following steps: (1) data acquisition & preprocessing; (2) musical feature extraction; (3) tensor representation; (4) FDC-NTD implementation; (5) correlation analysis; (6) hierarchical clustering; (7) cluster selection of interest.

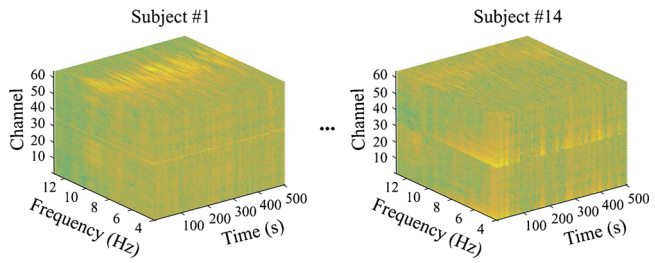


Fig. 4. Third-order tensors for 14 participants, and each tensor includes three dimensions of 64 channels, 46 frequency bins (4 ~ 13 Hz) and 510 time samples.

EEGLAB toolbox (Delorme and Makeig, 2004) and MATLAB R2016b, down-sampled to 256 Hz, and filtered by high-pass and low-pass filters with 1 Hz and 30 Hz cutoff frequencies. Detailed information about data acquisition and preprocessing can be found in our previous work (Cong et al., 2013a).

3.2. Tensor representation

TFR of the preprocessed EEG data was obtained by short-time Fourier transform (STFT). The Hamming window was adopted as the window function, with the window length of 3 s and 66.7% overlap ratio between windows. The number of Fourier points in each window was 1024, which was four times of the sampling rate. Power spectrum of EEG data are often evaluated in several frequency bands, such as delta (0.5 ~ 4 Hz), theta (4 ~ 8 Hz), alpha (8 ~ 13 Hz) and beta (13 ~ 30 Hz) (Siuly et al., 2016). According to previous work (Sammler et al., 2007; Lin et al., 2008, 2010; Shahin et al., 2009; Schaefer et al., 2011; Cong et al., 2012b, 2013a; Li et al., 2016), frequency fluctuations of brain activities elicited by musical stimulus are generally distributed in theta and alpha bands, hence in this study, frequency ranging from 4 to 13 Hz was used for further analysis. Therefore, third- order tensors including the spectrograms of EEG data with dimensions of 46 (frequency bins) × 510 (time samples) × 64 (space channels) were generated for 14 participants, as shown in Fig. 4.

3.3. Musical feature extraction

In this study, five long-term musical features (tonal and rhythmic, Fig. 5) were extracted by a frame-by-frame analysis method, providing a bridge for analyzing the connections between musical stimulus and ongoing EEG (Alluri et al., 2012; Cong et al., 2013a; Zhu et al., 2019). The duration of each frame was 3 s and overlap ratio between frames was 66.7%, which was consistent with the window settings in the STFT of EEG data. Therefore, the corresponding temporal courses with 510 samples were generated for those features. Furthermore, for the tonal features, Mode denotes the strength of major or minor mode, and Key Clarity is the measure of the tonal clarity (Alluri et al., 2012). For the rhythmic features, Fluctuation Centroid is defined as the geometric mean of the fluctuation spectrum, and it represents the global repartition of rhythm periodicities within the range of 0 ~ 10 Hz, indicating the average frequency of these periodicities (Alluri et al., 2012). Fluctuation entropy is the Shannon entropy of the fluctuation spectrum, and it represents the global repartition of rhythm periodicities. Pulse Clarity is regarded as an estimate of clarity of the pulse. The details of musical features and extraction method can be found in Latrillot and Toiviainen (2007), Alluri et al. (2012), and Cong et al. (2013a).

3.4. FDC-NTD implementation

3.4.1. Parameter initialization

The input factor matrices of spatial, spectral and temporal modes were initialized with uniformly distributed pseudorandom numbers generated by MATLAB function `rand`.

3.4.2. Termination criteria

In this study, two iteration termination criteria of FDC-NTD algorithm were adopted. (a) $|\mathrm{Fit}_{new} - \mathrm{Fit}_{old} < \varepsilon$, it means that the Fit change between the adjacent iterations should be smaller than the predefined threshold (e.g., $\varepsilon = 1\mathrm{e} - 6$. Tensor fitting is defined as $\mathrm{Fit} \triangleq \frac{1}{S}\sum_{s=1}^{S}[1 - \|\mathcal{X}^{(s)} - \hat{\mathcal{X}}^{(s)}\|_F / \|\mathcal{X}^{(s)}\|_F]$, where $\mathcal{X}^{(s)}$ and $\hat{\mathcal{X}}^{(s)}$ are original and recovered tensors respectively. Furthermore, the relative error is defined as $\mathrm{RelErr} \triangleq \sum_{s=1}^{S}[\|\mathcal{X}^{(s)} - \hat{\mathcal{X}}^{(s)}\|_F / \|\mathcal{X}^{(s)}\|_F]$. (b) The maximum number of iterations is no more than 1000.

3.4.3. Component number selection

To determine the number of components, a multi-dimensional model order selection technique termed as R-dimensional minimum description length (R-D MDL, da Costa et al., 2011) was adopted in this study. The R-D MDL method based on information theoretic criterium extended 1-D MDL (modified MDL) method to the multi-dimensional case by using the global eigenvalues, providing low computational complexity and maintaining good performance even for lower SNR scenarios (da Costa, 2010). Its optimization problem is given as follows:

$$\hat{d} = \underset{P}{\operatorname{argmin}} - I(\alpha^{(G)} - P)\log\left(\frac{g^{(G)}(P)}{\alpha^{(G)}(P)}\right) + p(P, I, \alpha^{(G)}) \tag{6}$$

where the penalty function $p(P, I, \alpha^{(G)})$ is chosen as $\frac{1}{2}P(2\alpha^{(G)} - P)\log(I)$. $\hat{d}$ denotes an estimated of the true model order

$d, g^{(G)}(P)$ and $\alpha^{(G)}(P)$ represent the geometric and arithmetic means of the P smallest global eigenvalues, respectively. I is set as $I = \max(I_1, I_2, \ldots, I_N)$, and $\alpha^{(G)}$ is the total number of adaptively defined global eigenvalues. Therefore, for the EEG tensors of 14 subjects, the number of components were respectively selected as {44, 34, 36, 38, 36, 39, 35, 35, 34, 37, 33, 36, 34, 35} via R-D MDL algorithm adapted from IPM software.[1]

Regarding the number of common components, we first ran 10 times of individual fast-HALS decompositions on each EEG tensor and then performed correlation analysis on the spatial/spectral modes between any two subjects successively. We obtained the averaged correlation coefficients $r = 0.8714$ and $r = 0.9031$ at level $p < 0.001$ on the two modes respectively. There was an average of 26 high-correlated spatial/spectral components between any two subjects (the correlation coefficients of 0.7 to 1 are considered to represent high or very high correlations; Asuero et al., 2006). Therefore, considering the hypothesis of double-coupled constraint, we set $L_1 = L_2 = 26$ and $L_3 = 0$.

3.5. Correlation analysis

After extracting the components using FDC-NTD algorithm, it is necessary to determine which ones are relevant to musical stimulus. According to our previous work Cong et al. (2012b, 2013a), correlation analysis was conducted between temporal courses of extracted temporal components and temporal courses of musical features, aiming to find the brain activities elicited by musical stimulus. Pearson correlation analysis was applied to calculate the correlation coefficient, and then Monte Carlo method and permutation test were used to determine significant thresholds of correlation and correct for multiple comparisons (Alluri et al., 2012; Groppe et al., 2011). Moreover, a threshold (at level $p < 0.05$) of correlation coefficient was calculated by a musical feature and $R^{(s)}$ temporal components from each participant. Those temporal components which are significantly correlated with temporal courses of musical features were considered to be relevant to musical stimulus, and will be of interest and further analyzed. Fig. 6 shows an example of spatial, spectral and temporal components of EEG data, represented as topography, power spectrum and waveform, respectively. The temporal component was significantly correlated with the musical feature of 'Fluctuation Centroid' (i.e., $0.1128 > 0.1064$). In addition, we can see that the occipital region of subject #11 is activated with theta oscillation.

3.6. Hierarchical clustering

It should be noted that the correlations in Fig. 1 include two parts: auto-correlation (intra-subject) and cross-correlation (inter-subject). Therefore, in addition to the high spatial cross-correlation of inter-subject, we also found the high correlations between spatial components within the subject. Different from imposing coupled constraints to address inter-subject correlations, in this study, we adopted hierarchical clustering to merge the highly correlated spatial components within the subject. Through the FDC-NTD algorithm, L_1 common spatial and L_2 spectral components from the ongoing EEG can be extracted. By virtue of the coupled constraints across subjects, we only need to cluster the L_1 common spatial components.

Moreover, clustering L_1 spatial components is simpler than clustering all of the spatial components extracted from 14 subjects by independent component analysis (ICA) individually (Cong et al., 2013a). For stable clustering, we adopted hierarchical agglomerative clustering algorithm, in which complete linkage was used to calculate the furthest distance (here we used correlation) between pairs of clusters and the pairs of clusters with the nearest distance were merged. We applied PCA to L_1 spatial components, and the number of principal components with 99% explained variance was selected as the number of clusters. Figs. 7 and 8 give some illustrations and results relevant to hierarchical clustering analysis about 26 spatial components. As shown in Fig. 7(a), when the cumulative explained variance exceeds the threshold of 99% (red dash line), 4 is selected as the number of clusters. Therefore, the hierarchical tree in the hierarchical clustering is spit into 4 clusters by cutting branches (red dash line). Fig. 7(c) shows the averaged spatial maps of each cluster. In cluster #1, there is only one component. For clusters #2, #3 and #4, the mean correlation coefficients between spatial components within each cluster are 0.9518, 0.9268 and 0.8397, and the corresponding standard deviations (SDs) are 0.0502, 0.0596 and 0.1314, respectively. This indicates that the components are highly correlated with each other in each cluster. The low correlations of averaged spatial components between clusters, as shown in Fig. 8, also demonstrate the accuracy of clustering results.

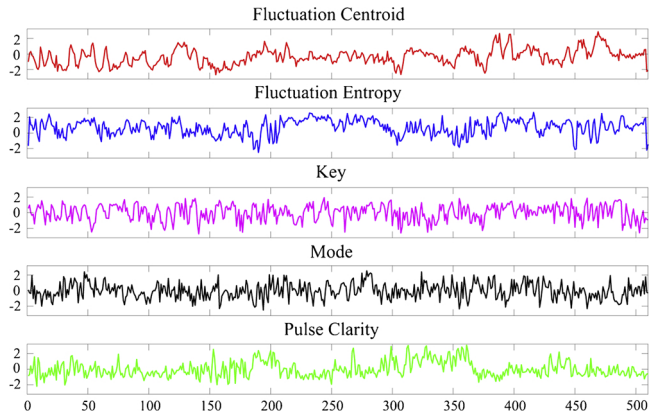


Fig. 5. Temporal courses of five musical features, including Fluctuation Centroid, Fluctuation Entropy, Key Clarity, Mode and Pulse Clarity.

3.7. Cluster selection of interest

In group analysis, phenomena commonly appearing in most subjects are more attractive than the individual ones of a particular subject.

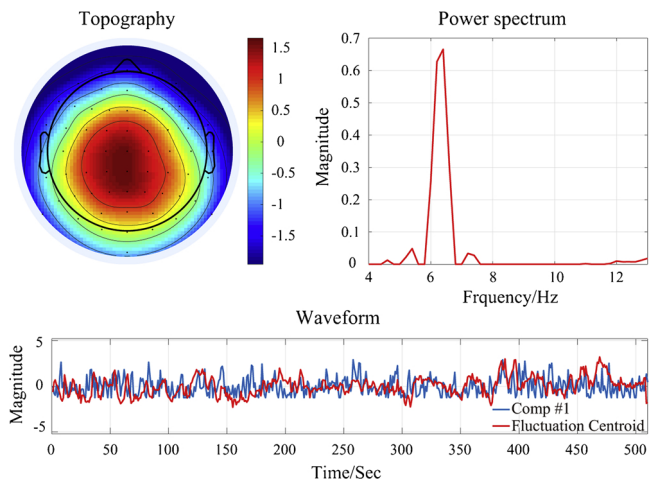


Fig. 6. The topography, power spectrum and waveform of the 1st EEG components from subject #11 of Run #1. The temporal course of Comp #1 is significantly correlated with the temporal course of 'Fluctuation Centroid' with a correlation coefficient of 0.1128 and a significant correlation threshold of 0.1064 (at level $p < 0.05$).

[1] https://lasp.unb.br/index.php/publications/softwares/

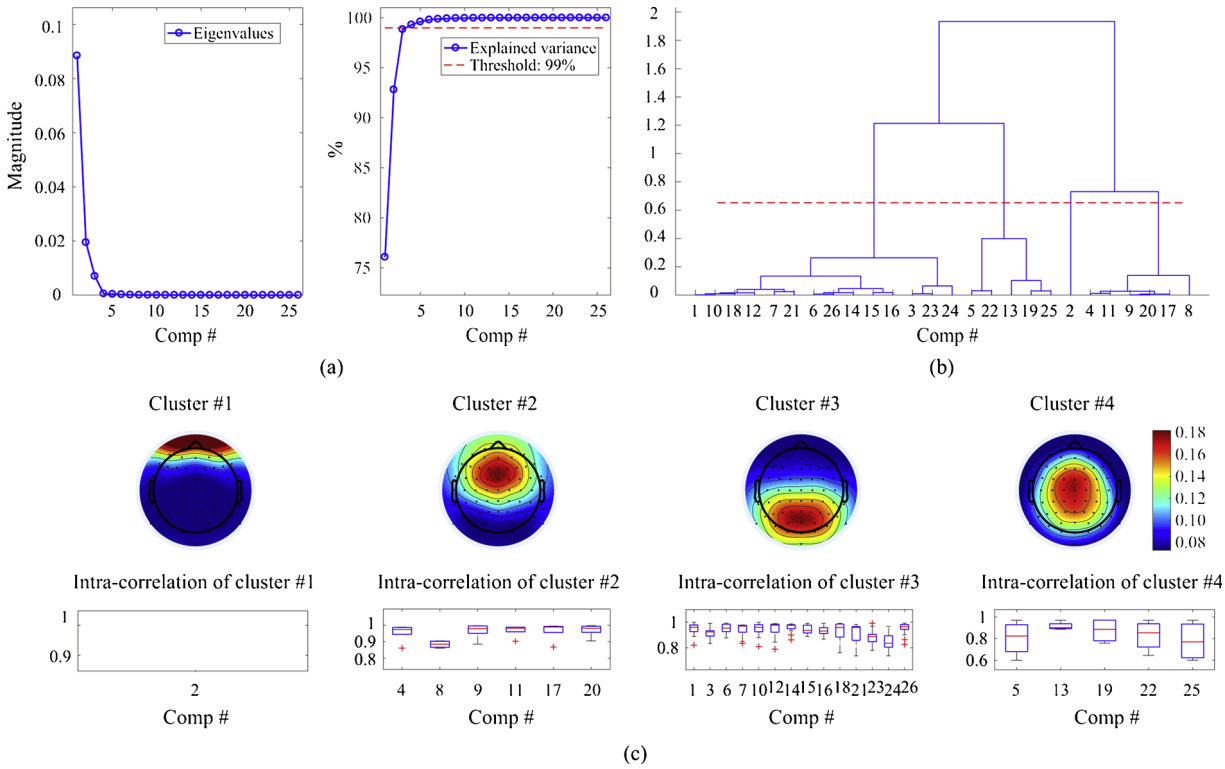


Fig. 7. Hierarchical clustering results of 26 spatial components of Run #7. (a) Selection of the number of clusters; (b) dendrogram output of hierarchical clustering; (c) averaged topographies of clusters and correlations between components within the clusters.

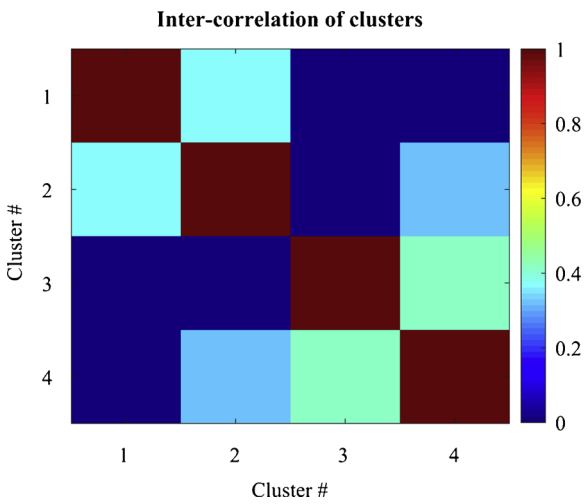


Fig. 8. Correlations of spatial maps (the averaged spatial component in each cluster) between clusters.

Therefore, in this section, our object is to determine the brain activities shared by the majority of subjects from the ongoing EEG data. Through comprehensively analyzing the results from correlation analysis (temporal components and musical features) and hierarchical clustering (common spatial components), we can obtain q clusters of spatial components whose parallel temporal components satisfy the threshold of significant correlation coefficients (with any musical feature). If the number of subjects contributing to a cluster exceeds half of the total number of subjects, the cluster will be selected as the cluster of interest and kept for further analysis in this study (Cong et al., 2013a). For each cluster of interest, the corresponding brain responses in most subjects are considered to be related to the musical stimulus. In Table 1, the subjects contributing to the 4 clusters in Fig. 7 are listed separately, and clusters #2, #3 and #4 are selected as the cluster of interest based on the predefined criterion. For the sake of simplicity, here we integrate the clustered spatial components and their corresponding temporal and spectral components into category of the cluster of interest.

4. Results

The uniqueness of the decomposition is critical to the interpretation of extracted components (Hunyadi et al., 2017). For the ongoing EEG data contaminated with noise, it is difficult to verify that the recovered

Table 1
Subject distribution of 4 clusters after comprehensively analyzing the results of correlation analysis and hierarchical clustering of Run #1. '1' and '–' indicate that the subject contributes or does not contribute to the cluster. The number of subjects contributing to cluster #2, #3 and #4 exceed half of the total number of subject.

Subject	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	Total
Cluster #1	–	–	–	–	–	–	1	–	–	1	–	–	–	1	3
Cluster #2	1	–	1	–	1	–	–	–	–	1	1	1	1	1	**8**
Cluster #3	1	1	1	–	1	1	1	1	1	1	1	1	1	–	**12**
Cluster #4	1	1	1	–	1	1	1	–	–	1	–	1	–	–	**8**

Table 2
Performance comparison between FDC-NTD and LCPTD algorithms. Evaluation indices include Obj, RelErr, Fit and Time averaged from 100 runs, and the number of occurrence of Clusters #I, #II and #III in 100 runs.

	Obj	RelErr	Fit	Time/s	Cluster #I	Cluster #II	Cluster #III
FDC-NTD	***1.1263e + 11***	***3.9202***	***0.7200***	164.06	***83***/100	100/100	***96***/100
LCPTD	1.1669e+11	3.9634	0.7169	***153.00***	82/100	100/100	93/100

Obj: objective function value; RelErr: relative error; Fit: tensor fitting; Time: running time.

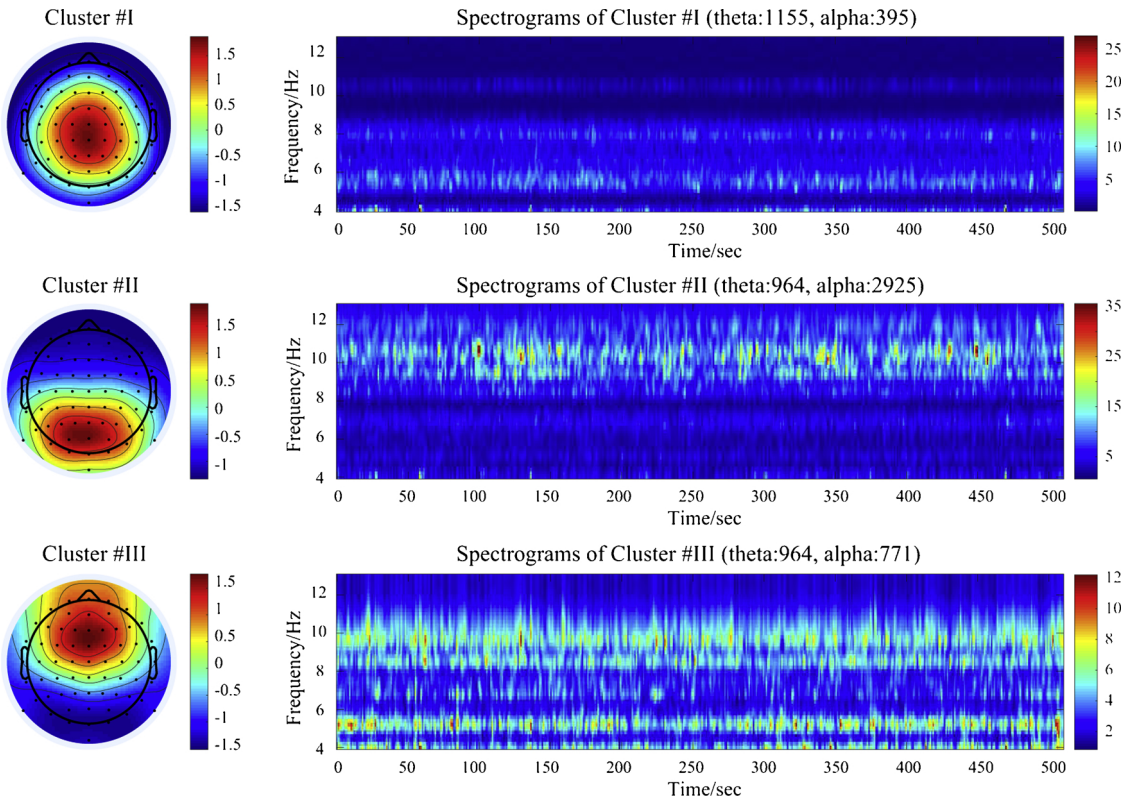


Fig. 9. Illustrations of averaged clusters of interest #I, #II and #III obtained from 100 runs. Spatial information, presented by the averaged topographies (left column), indicate the activations of centro-parietal, occipito-parietal and frontal regions of the brain elicited by musical stimulus, respectively. Overall spectrograms of clusters #I, #II and #III (right column) from 100 runs illustrate the frequency oscillations over the entire period. For cluster #I, the numbers of theta and alpha components are 1155 and 395. For cluster #II, the numbers of theta and alpha components are 964 and 2925. For cluster #III, the numbers of theta and alpha components are 964 and 771.

components are the true versions of the observed tensors. Therefore, in order to validate the reliability and stability of the solutions, we performed FDC-NTD algorithm 100 times on the generalized LCPTD model in this experiment. At the same time, to prove the validity of the generalized model, we also ran 100 times of Fast-HALS algorithm on the LCPTD model (for simplicity, we named it the LCPTD algorithm). According to Cong et al. (2012b), we chose 35 as the number of components in the LCPTD algorithm.

The experiments were carried out with the following computer configurations: CPU: Intel Core i5-7500 @ 3.40Hz 3.41Hz; Memory: 16.00 Gb; System: 64-bit Windows 10; Matlab R2016b. For data visualization, Figs. 10 and 11 were plotted using the graphics toolbox gramm[2] (Morel, 2018).

As shown in Table 2, we compare FDC-NTD and LCPTD algorithms in five aspects including objective function value (Obj), relative error

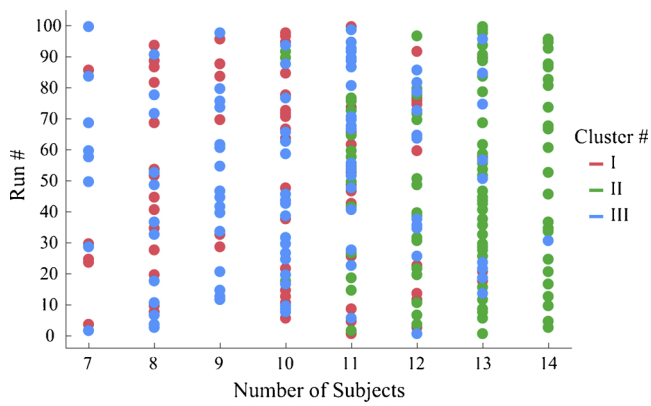


Fig. 10. Distribution of the number of the subjects contributing to the clusters #I, #II and #III in 100 runs. Cluster #1, #II and #III appeared 83, 100 and 96 times in 100 runs, respectively.

[2] https://github.com/piermorel/gramm

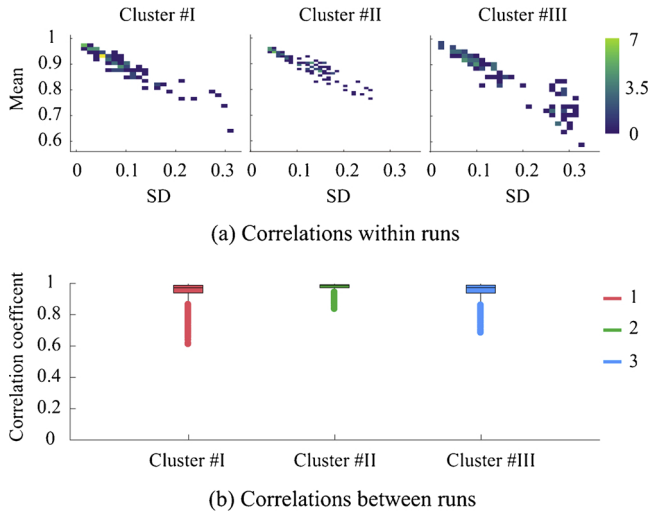


Fig. 11. Spatial correlations within and between runs for clusters #I, #II and #III in 100 runs. (a) Distribution of means and SDs of correlation coefficients calculated by the internal spatial components in each run for clusters #I, #II and #III. (b) Illustration of correlation coefficients calculated by the averaged spatial components between runs for clusters #I, #II and #III.

(RelErr), tensor fitting (Fit), running time (Time) and occurrence probabilities of clusters #I, #II and #III. For the averaged Obj, RelErr and Fit value, the FDC-NTD algorithm performs slightly better than LCPTD algorithm, but requires more consuming time. Through correlation analysis, hierarchical clustering and cluster selection of interest, three kinds of clusters of interest are obtained from 100 decomposition results. For example, the corresponding averaged topographies obtained by FDC-NTD algorithm are plotted in Fig. 9. For the FDC-NTD algorithm, the probabilities of clusters #I, #II, and #III occurring in 100 runs reach 83% (83/100), 100% (100/100) and 96% (96/100), while the occurrence probabilities of clusters #I, #II, and #III obtained by LCPTD algorithm are 82% (82/100), 100% (100/100) and 93% (93/100). It should be noted that the FDC-NTD algorithm is more stable than the LCPTD algorithm.

Fig. 9 also illustrates the overall spectrograms of clusters #I, #II and #III obtained in FDC-NTD algorithm. The spectrogram can be generated by back-projection of spectral and temporal components, presenting a qualitative and quantitative evaluation of frequency oscillations over the entire period. Regarding cluster #I, the topography reveals that the centro-parietal region of the brain was activated with quite a lot theta oscillations (4 ~ 7 Hz, 74.52%, 1155/1550) but little alpha oscillations (around 10 Hz, 25.48%, 395/1550). Conversely, the occipito-parietal region of the brain is activated with significant alpha oscillations (8 ~ 13 Hz, 75.21%, 2925/3889), accompanied by a small amount of theta oscillations (4 ~ 8 Hz, 24.79%, 964/3889) in cluster #II. In addition, we also obtained the topography representing the activation of frontal region of the brain, as shown in cluster #III. The frequency oscillations of cluster #III are distributed in the range of 4 ~ 11 Hz (theta-55.56%, 964/1735, alpha-44.44%, 771/1735).

The number of subjects contributing to the clusters #I, #II and #III in each run is visualized in Fig. 10. Regarding cluster #II, the number of subjects in each run is concentrated at 12, 13 and 14 (green circles). The number of subjects contributing to clusters #I and #III in 100 runs is mostly distributed in 7, 8, 9 and 10 (pink and blue circles). Furthermore, the gramm plot of spatial correlations within and between runs for clusters #I, #II and #III in 100 runs is illustrated in Fig. 11. From Fig. 11(a), we can see that the distribution of cluster #II is more compressed than that of clusters #I and #III. Regarding cluster #II, the averaged mean and SD of correlations in 100 runs are 0.8949 and 0.0949. For cluster #I, the averaged mean and SD of correlations in 83 runs are 0.8959 and 0.0985. For cluster #III, the averaged mean and SD of correlations in 96 runs are 0.8459 and 0.1402, which are inferior to the ones of clusters #I and #II. Fig. 11(b) illustrates the correlation coefficients between runs for clusters #I, #II and #III. We can find that the cluster #II is more stable with less outliers. Regarding cluster #II, the mean and SD of correlation coefficient are 0.9813 and 0.0182. For cluster #I, the mean and SD are 0.9570 and 0.0492. For cluster #III, the mean and SD are 0.9591 and 0.0449. From the results in Fig. 11, we can conclude that spatial components within and between runs for each cluster are highly correlated with each other. It strongly demonstrates the accuracy and stability of clustering results obtained by FDC-NTD algorithm.

5. Discussion

Tensor decomposition and group-level ICA methods have been generally used to extract stimulus-elicited components from a higher-order EEG tensor or concatenated EEG matrix of different subjects for group- level analysis in the cognitive research (Eichele et al., 2011; Cong et al., 2012b, 2015a). Only if the number of sources or the hidden information in EEG data of different subjects is consistent, the above methods will make sense to stack/concatenate the data to the tensor/matrix for analysis (Cong and He, 2013b; Wang et al., 2018b). Individual ICA method is not naturally suited to explore group inferences since the result fusion across individuals is sometimes a non-trial problem (Eichele et al., 2011). Coupled tensor decomposition, an extension of tensor decomposition to multi-block tensors, has been widely utilized to explore the potential common phenomenon across tensors (Sørensen et al., 2015; Ermiş et al., 2015; Gong et al., 2016). Therefore, in this study, a comprehensive framework based on coupled tensor decomposition applying to group-level analysis of ongoing EEG data during free listening to a 8.5-min long tango music was investigated.

The theoretical principle of coupled tensor decomposition is that the different tensors share some of the same or partially identical factor matrices (Zhou et al., 2016; Yokota et al., 2012). For the ongoing EEG data, we indeed found that there were highly correlated information in both spatial and spectral modes among the tensors represented from 14 subjects (see it in Fig. 1), which can be regarded as a prerequisite for applying coupled tensor decomposition in this study. Meanwhile, there is also individual information for each tensor, which may lead to inconsistent number of components among tensors. Considering the

common and individual information with the inconsistent number of components, we extended the LCPTD model to a general case. For the validation of the generalized model, we ran 100 times of two algorithms to compare their performance as shown in Table 2. It indicates that the our generalized LCPTD model has higher model fit and stronger robustness than the LCPTD model, and the performance development depends to a large extent on the actual number of components we use. For clusters whose occurrence rate are not 100%, this is mainly due to the inconsistency of local optimal solutions caused by random initialization in 100 algorithm implementations. Wang et al. demonstrated that sparsity regularization can improve the extraction stability of EEG components (Wang et al., 2018a), which provides a good perspective for our future work.

From the results of 100 FCD-NTD algorithm implementations, we found that the brain activities of selected clusters #I, #II and #III relevant to the musical stimulus were mainly distributed in the bands of 4∼8 Hz (theta), 8∼13 Hz (alpha) and 4∼11 Hz, and located in the centro-parietal, occipito-parietal and frontal regions respectively. For the same data in this study, such theta and alpha activities in the central and occipital regions were reported in Cong et al. (2013a), but no such activities in the frontal region. For the ICA-based method in Cong et al. (2013a), posterior K- means clustering was adopted to cluster all of the spatial components extracted from 14 subjects individually. However, the prior coupling information present in spatial and spectral modes was not employed when extract the hidden information, which may result in the failure of information extraction. In addition, clustering $L_{1,2}$ common spatial components is obviously much simpler and stable than clustering all of the spatial components in Cong et al. (2013a), where the mean correlation coefficients between spatial maps in clusters #I and #II were only 0.85 and 0.81 respectively. Regarding the clusters #I and #II, the significant theta and alpha oscillations were also reported in Li et al. (2016), where Li et al. utilized tensor decomposition imposing EEG phase characteristics to explore the brain responses to the naturalistic and continuous musical stimulus. Cong et al. extracted only alpha activity in the posterior region using fourth-order nonnegative tensor decomposition without considering the existence of individual information for each subject (Cong et al., 2012b). Compared to previous work, the proposed FDC-NTD algorithm can avoid strong constraint that imposes consistency on temporal, spatial and spectral modes between EEG tensors (Cong et al., 2012b; Li et al., 2016). In addition, it can utilize the multi-way structure of tensor-represented data and the coupled relationship across tensor blocks, and can decompose EEG tensors into common components and individual components in each mode. The extraction of common components among data makes it easier to discover the commonly appearing brain activities among majority of subjects. The high means and low SDs of correlation coefficients within/between 100 runs can demonstrate the stability and practicability of coupled tensor decomposition applied to the group-level analysis of ongoing EEG data.

Besides the studies of ongoing EEG analysis elicited by naturalistic and continuous musical stimulus, previous work on the use of EEG activities to analyze emotion and musical stimuli can provide some solid references for the results in this study (Sammler et al., 2007; Lin et al., 2008, 2010; Schmidt and Trainor, 2001). During listening to the emotional music, the spectrum power asymmetry indexes located in the brain areas corresponding to RASM12 (namely, 12 symmetric electrode pairs including Fp1-Fp2, F7-F8, F3-F4, FT7-FT8, FC3-FC4, T7-T8, P7-P8, C3-C4, TP7-TP8, CP3-CP4, P3-P4, and O1-O2) are sensitive to the brain activations associated with emotion responses (Lin et al., 2008). Lin et al. found the frontal and parietal lobes across frequency bands including theta and alpha contributed a lot in the emotion recognition during music listening (Lin et al., 2010). According to Sammler et al. (2007), the increase of theta power over the frontal midline was associated with pleasant music, while the frontal alpha asymmetry of ongoing EEG activity was used to distinguish the emotional valence of musical stimuli (Schmidt and Trainor, 2001). The previous studies regarding EEG and musical stimuli verified the plausibility of our findings to some extent.

In conclusion, we proposed a comprehensive framework based on coupled tensor decomposition for the group analysis of ongoing EEG data, elicited by naturalistic and continuous musical stimulus. Specifically, the proposed framework includes the following seven steps: data acquisition & preprocessing, musical feature extraction, tensor representation, algorithm implementation, correlation analysis, hierarchical clustering and cluster selection of interest, aiming to discover commonly appearing brain activities among subjects. The results obtained in the proposed framework illustrate that our findings are in line with the results of previous studies, and it can be inferred that those brain activities we extracted are associated with musical stimulus. Furthermore, the proposed framework based on coupled tensor decomposition in this study provides a new perspective for the processing and analysis of multi- subject ongoing EEG data. Coupled tensor decomposition methods with different optimization strategies can be applied for comparison to find more convincing solutions when processing and analyzing ongoing EEG data, which will be one of our future work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91748105 and 81471742), the Fundamental Research Funds for the Central Universities [DUT2019] in Dalian University of Technology in China and the scholarships from China Scholarship Council (Nos. 201706060262 and 201706060263).

Appendix A. Derivation of FDC-NTD algorithm

The minimized optimization problem in (2) can be converted into $\max(R^{(s)})$ rank-1 tensor approximation problems via HALS algorithm (Cichocki et al., 2007), which can be solved sequentially and iteratively as follows:

$$\min D_F^{(r)}(\boldsymbol{u}_r^{(n,s)}) = \sum_{s=1}^{S} \|\mathcal{Y}_r^{(s)} - \boldsymbol{u}_r^{(1,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)}\|_F^2 = \sum_{s=1}^{S} \|\mathcal{Y}_{r,(n)}^{(s)} - \boldsymbol{u}_r^{(n,s)} \{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}^{T}}\|_F^2, \tag{7}$$

where $\mathcal{Y}_r^{(s)} \doteq \mathcal{X}^{(s)} - \sum_{k\neq r}^{R^{(s)}} \boldsymbol{u}_k^{(1,s)} \circ \boldsymbol{u}_k^{(2,s)} \circ \cdots \circ \boldsymbol{u}_k^{(N,s)}$ and $\mathcal{Y}_{r,(n)}^{(s)}$ is the mode-n matricization of $\mathcal{Y}_r^{(s)}$. Mathematically, using the trace property of matrix, (7) can be further represented as:

$$\min D_F^{(r)}(\boldsymbol{u}_r^{(n,s)}) = \sum_{s=1}^{S} \mathrm{tr}[\left(\mathcal{Y}_{r,(n)}^{(s)} - \boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\right)\left(\mathcal{Y}_{r,(n)}^{(s)} - \boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\right)^T] \tag{8}$$

We can calculate the gradient of $D_F^{(r)}(\boldsymbol{u}_r^{(n,s)})$ in (8) with respect to $\boldsymbol{u}_r^{(n,s)}$ as:

$$\frac{\partial D_F^{(r)}(\boldsymbol{u}_r^{(n,s)})}{\partial \boldsymbol{u}_r^{(n,s)}} = \begin{cases} \sum_{s=1}^{S} \left[-2\mathcal{Y}_{r,(n)}^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} + 2\boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}\right], & r \le L_n \\ -2\mathcal{Y}_{r,(n)}^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} + 2\boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}, & r > L_n \end{cases} \tag{9}$$

For the solution of $\boldsymbol{u}_r^{(n,s)}$, we only need to set the gradient in (9) to zero. Therefore, the learning rule of $\boldsymbol{u}_r^{(n,s)}$ obtained via HALS strategy can be formulated as:

$$\boldsymbol{u}_r^{(n,s)} = \begin{cases} \sum_s [\mathcal{Y}_{r,(n)}^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}] / \sum_s \gamma_r^{(n,s)}, & r \le L_n, \\ [\mathcal{Y}_{r,(n)}^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}]/\gamma_r^{(n,s)}, & r > L_n, \end{cases} \tag{10}$$

where $\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}$ is denoted by $\gamma_r^{(n,s)}$ for simplicity, and it can be reformulated as $\{\boldsymbol{u}_r^{(s)T}\boldsymbol{u}_r^{(s)}\}^{\circledast_{-n}}$. As $\|\boldsymbol{u}_r^{(n,s)}\| = 1$, $n = 1 \cdots N-1$, the scaling coefficients $\gamma_r^{(n,s)}$ can be abbreviated as (5). It is important to note that the calculation of $\mathcal{Y}_{r,(n)}^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}$ including the mode-n matricization of $\mathcal{Y}_r^{(s)}$ and the Khatri-Rao product of $\boldsymbol{u}_r^{(m,s)}$, $m = 1 \ldots, n-1, n+1, \ldots, N$ needs to be performed in each iteration. It can result in rather high computational cost, especially for large-scale problems (Cichocki and Phan, 2009). Since the fast-HALS algorithm has been proven to be more efficient in Cichocki and Phan (2009) than HALS algorithm, in this study, we further extend it to the generalized LCPTD model. Inspired by fast-HALS algorithm, $\mathcal{Y}_r^{(s)}$ can be represented as:

$$\mathcal{Y}_r^{(s)} = \mathcal{X}^{(s)} - \sum_{r=1}^{R^{(s)}} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} + \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} = \mathcal{X}^{(s)} - \hat{\mathcal{X}}^{(s)} + \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)}. \tag{11}$$

By defining $\zeta_r^{(n,s)} \doteq \mathcal{Y}_{r,(n)}^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}}$, we substitute the mode-n matricization of $\mathcal{Y}_r^{(s)}$ in (11) into $\zeta_r^{(n,s)}$ and it can be represented as:

$$\begin{aligned} \zeta_r^{(n,s)} &= [\mathcal{X}_n^{(s)} - \boldsymbol{U}^{(n,s)}\{\boldsymbol{U}^{(s)}\}^{\underline{\odot}_{-n}^T} + \boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}]\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} \\ &= \mathcal{X}_n^{(s)}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} - \boldsymbol{U}^{(n,s)}\{\boldsymbol{U}^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} + \boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} \\ &= \left[\mathcal{X}_n^{(s)}\{\boldsymbol{U}^{(s)}\}^{\odot_{-n}}\right]_r - \boldsymbol{U}^{(n,s)}\left[\{\boldsymbol{U}^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{U}^{(s)}\}^{\odot_{-n}}\right]_r + \boldsymbol{u}_r^{(n,s)}\{\boldsymbol{u}_r^{(s)}\}^{\underline{\odot}_{-n}^T}\{\boldsymbol{u}_r^{(s)}\}^{\odot_{-n}} \\ &= [\mathcal{X}_n^{(s)}\{\boldsymbol{U}^{(s)}\}^{\odot_{-n}}]_r - \boldsymbol{U}^{(n,s)}\left[\left\{\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\right\}^{\odot_{-n}}\right]_r + \gamma_r^{(n,s)}\boldsymbol{u}_r^{(n,s)} \end{aligned} \tag{12}$$

with

$$\{\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\}^{\circledast_{-n}} = \{\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\}^{\circledast} \oslash (\boldsymbol{U}^{(n,s)^T}\boldsymbol{U}^{(n,s)}). \tag{13}$$

At last, we obtain the learning rule of $\boldsymbol{u}_r^{(n,s)}$ based on fast-HALS algorithm Cichocki and Phan (2009) in generalized LCPTD model (1) as follows:

$$\boldsymbol{u}_r^{(n,s)} = \begin{cases} \sum_s \zeta_r^{(n,s)} / \sum_s \gamma_r^{(n,s)}, & r \le L_n \\ \zeta_r^{(n,s)}/\gamma_r^{(n,s)}, & r > L_n \end{cases} \tag{14}$$

with the definitions of $\zeta_r^{(n,s)}$ in (12) and $\gamma_r^{(n,s)}$ in (5). The mode-n matricization $\mathcal{X}_n^{(s)}$ in $\zeta_r^{(n,s)}$ only needs to be performed once in initialization, which greatly improves the computation efficiency of the proposed FDC-NTD algorithm.

References

Acar, E., Aykut- Bingol, C., Bingol, H., Bro, R., Yener, B., 2007. Multiway analysis of epilepsy tensors. Bioinformatics 23, i10–i18. https://doi.org/10.1093/bioinformatics/btm210.

Adeli, H., Ghosh-Dastidar, S., 2010. Automated EEG-Based Diagnosis of Neurological Disorders: Inventing the Future of Neurology. CRC presshttps://doi.org/10.1201/9781439815328.

Alluri, V., Toiviainen, P., Jääskeläinen, I.P., Glerean, E., Sams, M., Brattico, E., 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage 59, 3677–3689. https://doi.org/10.1016/j.neuroimage.2011.11.019.

Asuero, A.G., Sayago, A., Gonzalez, A., 2006. The correlation coefficient: an overview.

Crit. Rev. Anal. Chem. 36, 41–59. https://doi.org/10.1080/10408340500526766.
Berger, H., 1929. Über das elektrenkephalogramm des menschen. Arch. Psychiatr. Nervenkrankh. 87, 527–570. https://doi.org/10.1007/BF01797193.
Busch, N.A., Dubois, J., VanRullen, R., 2009. The phase of ongoing EEG oscillations predicts visual perception. J. Neurosci. 29, 7869–7876. https://doi.org/10.1523/JNEUROSCI.0113-09.2009.
Calhoun, V.D., Liu, J., Adalı, T., 2009. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. Neuroimage 45, S163–S172. https://doi.org/10.1016/j.neuroimage.2008.10.057.
Chen, X., Wang, Z.J., McKeown, M., 2016. Joint blind source separation for neurophysiological data analysis: multiset and multimodal methods. IEEE Signal Process. Mag. 33, 86–107. https://doi.org/10.1109/MSP.2016.2521870.
Cichocki, A. 2013. Tensor decompositions: a new concept in brain data analysis? arXiv preprint arXiv:1305.0395, pp. 1–19.
Cichocki, A., Phan, A.-H., 2009. Fast local algorithms for large scale nonnegative matrix and tensor factorizations. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. 92, 708–721. https://doi.org/10.1587/transfun.E92.A.708.
Cichocki, A., Zdunek, R., Amari, S.-i., 2007. Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In: International Conference on Independent Component Analysis and Signal Separation. Springer. pp. 169–176. https://doi.org/10.1007/978-3-540-74494-8_22.
Cong, F., Alluri, V., Nandi, A.K., Toiviainen, P., Fa, R., Abu- Jamous, B., Gong, L., Craenen, B.G., Poikonen, H., Huotilainen, M., et al., 2013a. Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. IEEE Trans. Multimed. 15, 1060–1069. https://doi.org/10.1109/TMM.2013.2253452.
Cong, F., He, Z., Hämäläinen, J., Leppänen, P.H., Lyytinen, H., Cichocki, A., Ristaniemi, T., 2013b. Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection. J. Neurosci. Methods 212, 165–172. https://doi.org/10.1016/j.jneumeth.2012.09.029.
Cong, F., Lin, Q.-H., Kuang, L.-D., Gong, X.-F., Astikainen, P., Ristaniemi, T., 2015a. Tensor decomposition of EEG signals: a brief review. J. Neurosci. Methods 248, 59–69. https://doi.org/10.1016/j.jneumeth.2015.03.018.
Cong, F., Phan, A.H., Zhao, Q., Huttunen-Scott, T., Kaartinen, J., Ristaniemi, T., Lyytinen, H., Cichocki, A., 2012a. 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. 22, 1250025. https://doi.org/10.1142/S0129065712500256.
Cong, F., Phan, A.H., Zhao, Q., Nandi, A.K., Alluri, V., Toiviainen, P., Poikonen, H., Huotilainen, M., Cichocki, A., Ristaniemi, T., 2012b. Analysis of ongoing EEG elicited by natural music stimuli using nonnegative tensor factorization. In: 2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO). IEEE. pp. 494–498.
Cong, F., Ristaniemi, T., Lyytinen, H., 2015b. Advanced signal processing on brain event-related potentials: filtering ERPs in time. In: Frequency and Space Domains Sequentially and Simultaneously, vol. 13. World Scientific. https://doi.org/10.1142/9306.
da Costa, J.P.C.L., 2010. Parameter Estimation Techniques for Multi-Dimensional Array Signal Processing. Shaker Verlag.
da Costa, J.P.C.L., Roemer, F., Haardt, M., de Sousa, R.T., 2011. Multi-dimensional model order selection. EURASIP J. Adv. Signal Process. 2011, 26. https://doi.org/10.1186/1687-6180-2011- 26.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.
Eichele, T., Rachakonda, S., Brakedal, B., Eikeland, R., Calhoun, V., 2011. EEGIFT: group independent component analysis for event- related EEG data. Comput. Intell. Neurosci. 2011, 9. https://doi.org/10.1155/2011/129365.
Ermiş, B., Acar, E., Cemgil, A.T., 2015. Link prediction in heterogeneous data via generalized coupled tensor factorization. Data Min. Knowl. Discov. 29, 203–236. https://doi.org/10.1007/s10618-013-0341-y.
Gong, X., Lin, Q., Cong, F., De Lathauwer, L., 2016. Double coupled canonical polyadic decomposition for joint blind source separation. IEEE Trans. Signal Process. 66, 3475–3490. https://doi.org/10.1109/TSP.2018.2830317.
Gong, X., Wang, X., Lin, Q., et al., 2015. Generalized non-orthogonal joint diagonalization with LU decomposition and successive rotations. IEEE Trans. Signal Processing 63, 1322–1334. https://doi.org/10.1109/TSP.2015.2391074.
Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48, 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x.
Herrmann, C.S., 2001. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Exp. Brain Res. 137, 346–353. https://doi.org/10.1007/s002210100682.
Huber, R., Ghilardi, M.F., Massimini, M., Tononi, G., 2004. Local sleep and learning. Nature 430, 78. https://doi.org/10.1038/nature02663.
Hunyadi, B., Dupont, P., Van Paesschen, W., Van Huffel, S., 2017. Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 7, e1197. https://doi.org/10.1002/widm.1197.
Jeong, J., 2004. EEG dynamics in patients with Alzheimer's disease. Clin. Neurophysiol. 115, 1490–1505. https://doi.org/10.1016/j.clinph.2004.01.001.
Jespersen, K.V., Koenig, J., Jennum, P., Vuust, P., 2015. Music for insomnia in adults. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD010459.pub2.
Koelsch, S., 2012. Brain and Music. John Wiley & Sons.
Kolda, T.G., Bader, B.W., 2009. Tensor decompositions and applications. SIAM Rev. 51, 455–500. https://doi.org/10.1137/07070111X.
Latrillot, O., Toiviainen, P., 2007. MIR in Matlab: a toolbox for musical feature extraction. Proceedings of the International Conference on Music Information Retrieval.
Lee, H., Choi, S., 2009. Group nonnegative matrix factorization for EEG classification. Artificial Intelligence and Statistics. pp. 320–327.
Li, J., Ji, H., Gu, R., Hou, L., Zhang, Z., Wu, Q., Lu, R., Li, M., 2016. Explore the brain response to naturalistic and continuous music using EEG phase characteristics. In: International Conference on Intelligent Computing. Springer. pp. 294–305. https://doi.org/10.1007/978-3-319-42291-6_29.
Lin, Y.P., Wang, C.H., Jung, T.P., Wu, T.L., Jeng, S.K., Duann, J.R., Chen, J.H., 2010. EEG-based emotion recognition in music listening. IEEE Trans. Biomed. Eng. 57, 1798–1806. https://doi.org/10.1109/TBME.2010.2048568.
Lin, Y.P., Wang, C.H., Wu, T.L., Jeng, S.-K., Chen, J.-H., 2008. Support vector machine for EEG signal classification during listening to emotional music. In: 2008 IEEE 10th Workshop on Multimedia Signal Processing. IEEE. pp. 127–130. https://doi.org/10.1109/MMSP.2008.4665061.
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique. MIT Press.
MacDonald, R., Kreutz, G., Mitchell, L., 2013. Music, Health, and Wellbeing. Oxford University Press.
Maratos, A., Gold, C., Wang, X., Crawford, M., 2008. Music therapy for depression. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD004517.pub2.
Morel, P., 2018. Gramm: grammar of graphics plotting in Matlab. J. Open Source Softw. 3, 568.
Mørup, M., 2011. Applications of tensor (multiway array) factorizations and decompositions in data mining. Wiley Interdiscip. Rev.: Data Min. Knowl. Discov. 1, 24–40 10.1002/widm.1.
Mørup, M., Hansen, L.K., Parnas, J., Arnfred, S.M., 2006. Decomposing the Time-Frequency Representation of EEG using Non-Negative Matrix and Multi-Way Factorization. Technical University of Denmark Technical Report. pp. 1–28.
Mössler, K., Chen, X., Heldal, T.O., Gold, C., 2011. Music therapy for people with schizophrenia and schizophrenia-like disorders. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD004025.pub3.
Sammler, D., Grigutsch, M., Fritz, T., Koelsch, S., 2007. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. Psychophysiology 44, 293–304. https://doi.org/10.1111/j.1469-8986.2007.00497.x.
Schaefer, R.S., Vlek, R.J., Desain, P., 2011. Music perception and imagery in EEG: alpha band effects of task and stimulus. Int. J. Psychophysiol. 82, 254–259. https://doi.org/10.1016/j.ijpsycho.2011.09.007.
Schmidt, L.A., Trainor, L.J., 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. Cogn. Emotion 15, 487–500. https://doi.org/10.1080/02699930126048.
Shahin, A.J., Trainor, L.J., Roberts, L.E., Backer, K.C., Miller, L.M., 2009. Development of auditory phase-locked activity for music sounds. J. Neurophysiol. 103, 218–229. https://doi.org/10.1152/jn.00402.2009.
Siuly, S., Li, Y., Zhang, Y., 2016. EEG Signal Analysis and Classification: Techniques and Applications. Springer International Publishinghttps://doi.org/10.1007/978-3-319-47653-7.
Sørensen, M., Domanov, I., De Lathauwer, L., 2015. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-($L_{r,n}$, $L_{r,n}$, 1) terms – part II: algorithms. SIAM J. Matrix Anal. Appl. 36, 1015–1045. https://doi.org/10.1137/140956865.
Wang, D., Cong, F., Zhao, Q., Toiviainen, P., Nandi, A.K., Huotilainen, M., Ristaniemi, T., Cichocki, A., 2016. Exploiting ongoing EEG with multilinear partial least squares during free-listening to music. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE. pp. 1–6. https://doi.org/10.1109/MLSP.2016.7738849.
Wang, D., Wang, X., Zhu, Y., Toiviainen, P., Huotilainen, M., Ristaniemi, T., Cong, F., 2018a. Increasing stability of EEG components extraction using sparsity regularized tensor decomposition. In: International Symposium on Neural Networks. Springer. pp. 789–799. https://doi.org/10.1007/978-3-319 -92537-0_89.
Wang, D., Zhu, Y., Ristaniemi, T., Cong, F., 2018b. Extracting multi-mode ERP features using fifth-order nonnegative tensor decomposition. J. Neurosci. Methods 308, 240–247. https://doi.org/10.1016/j.jneumeth.2018.07.020.
Wang, X., Ristaniemi, T., Cong, F., 2019. Fast implementation of double-coupled non-negative canonical polyadic decomposition. In: International Conference on Acoustics, Speech, and Signal Processing. IEEE. pp. 8588–8592. https://doi.org/10.

1109/ICASSP.2019.8682737.
Yokota, T., Cichocki, A., Yamashita, Y., 2012. Linked PARAFAC/CP tensor decomposition and its fast implementation for multi-block tensor analysis. In: International Conference on Neural Information Processing. Springer. pp. 84–91. https://doi.org/10.1007/978-3-642-34487 -9_11.
Zhou, G., Cichocki, A., Xie, S., 2012. Fast nonnegative matrix/tensor factorization based on low-rank approximation. IEEE Trans. Signal Process. 60, 2928–2940. https://doi.org/10.1109/TSP.2012.2190410.
Zhou, G., Zhao, Q., Zhang, Y., Adalı, T., Xie, S., Cichocki, A., 2016. Linked component analysis from matrices to high-order tensors: applications to biomedical data. Proc. IEEE 104, 310–331. https://doi.org/10.1109/JPROC.2015.2474704.
Zhu, Y., Zhang, C., Toiviainen, P., Huotilainen, M., Mathiak, K., Ristaniemi, T., Cong, F., 2019. Exploring frequency-dependent brain networks from ongoing EEG using spatial ICA during music listening. bioRxiv 509802.

PV

FAST LEARNINGS OF COUPLED NONNEGATIVE TENSOR DECOMPOSITION USING OPTIMAL GRADIENT AND LOW-RANK APPROXIMATION

by

Xiulin Wang, Tapani Ristaniemi and Fengyu Cong 2020

Submitted to Signal Processing.

Fast Learnings of Coupled Nonnegative Tensor Decomposition Using Optimal Gradient and Low-rank Approximation

Xiulin Wang[a,b,*], Tapani Ristaniemi[b], Fengyu Cong[b,c,d,a]

[a] *School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China*
[b] *Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland*
[c] *School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, China*
[d] *Key Laboratory of Integrated Circuit and Biomedical Electronic System, Liaoning Province, Dalian University of Technology, Dalian, China*

Abstract

Nonnegative tensor decomposition has been widely applied in signal processing and neuroscience, etc. When it comes to group analysis of multi-block tensors, traditional tensor decomposition is insufficient to utilize the coupled information among tensors. In this study, we propose a coupled nonnegative CANDECOMP/PARAFAC decomposition algorithm based on alternating proximal gradient method (CoNCPD-APG), which is capable of a simultaneous decomposition of tensors from different samples that are partially linked and a simultaneous extraction of common components, individual components and core tensors. Due to the low optimization efficiency brought by nonnegative constraint and high-dimensional nature of the data, we further propose the lraCoNCPD-APG algorithm by combining low-rank approximation and the proposed coupled decomposition method. When processing multi-block large-scale tensors, the proposed lraCoNCPD-APG algorithm can greatly reduce the computational load without compromising the decomposition quality. Experiment results of cou-

*Corresponding author
Email addresses: `xiulin.wang@foxmail.com` (Xiulin Wang), `tapani.e.ristaniemi@jyu.fi` (Tapani Ristaniemi), `cong@dlut.edu.cn` (Fengyu Cong)
URL: `www.xiulin.wang` (Xiulin Wang), `users.jyu.fi/~riesta` (Tapani Ristaniemi), `www.escience.cn/people/cong` (Fengyu Cong)

pled nonnegative tensor decomposition problems designed for synthetic data, real-world face image and event-related potential data demonstrate the practicability and superiority of the proposed algorithms.

Keywords: Alternating proximal gradient, CANDECOMP/PARAFAC, coupled, low-rank approximation, nonnegative tensor decomposition

1. Introduction

Decomposing a tensor into a minimal number of rank-1 tensors, is known as CANDECOMP/PARAFAC (also known as Canonical Polyadic, [1, 2, 3]) decomposition (CPD). Nonnegative CPD (NCPD) provides a part-based representation of a tensor via imposing nonnegative constraints on its hidden factors, which enables us to extract more meaningful and convincing information [4]. For example, electroencephalography (EEG) data with spatial, temporal and subject features can be represented as a third-order tensor, and the underlying features can be simultaneously extracted through CPD [5]. If time-frequency representation is further considered, non-negativity will naturally be brought into the EEG data, which should be solved by NCPD [6, 7]. Although CPD/NCPD methods have received widespread attention both in theory and application, their main contribution lies in the decomposition analysis of a tensor represented by a single dataset [8, 9, 10]. Regarding emerging multi-block tensors (e.g., multi-subject/multi-modal biomedical data) that need to be analyzed together, traditional tensor decomposition methods become very difficult to identify and utilize the connections between different tensors [11, 12]. For example, the internal connections in ongoing EEG data of multiple subjects or the complementary spatiotemporal characteristics in the EEG-fMRI integrated data obviously cannot be effectively utilized in the CPD/NCPD methods [12, 13].

Coupled tensor decomposition, an extension of tensor decomposition to multi-block tensors, provides an ideal solution for how to jointly analyze two or more tensors [11, 14, 15]. Interestingly, the uniqueness condition of coupled tensor decomposition is more relaxed than that of single tensor decomposition

[15, 16]. Coupled tensor decomposition can achieve higher decomposition accuracy and stronger robustness by combining the prior coupling information among tensors[17]. It can also take full advantage of the constraints (e.g., sparsity, smoothness and nonnegativity) imposed in the optimization process, so that can obtain more unique solutions and interpretable components [17, 18]. Coupled tensor decomposition can reveal the inner-relationships among tensor data [5, 12], and maintain the potential interactions among multi-way structure of tensor data. The interactions will inevitably be lost in two-way matrix counterparts [19, 20], and coupled tensor methods can circumvent the independent constraint [16, 21]. Moreover, coupled tensor decomposition of multi-block tensors can achieve the simultaneous extraction of common components shared by all blocks and individual components corresponding to individual block [22, 23, 24].

To date, increasing recognition of joint tensor analysis has led to the consideration of coupled tensor decomposition in a number of applications. A special case, coupled matrix and tensor factorization (CMTF [25]) and its variants including advanced CMTF [26], and coupled tensor-tensor decomposition (CTTD [13, 27]) have proved their superiority over ICA-based two-way methods in EEG and fMRI data fusion. The algebraic double coupled CPD (DC-CPD) algorithm using second-order statistics in joint blind source separation (JBSS) problem exhibits more relaxed uniqueness and higher accuracy than the standard CPD [14]. Linked CPD model optimized by hierarchical alternating least squares (HALS), fast HALS and alternating direction method of multipliers (ADMM) has also achieved good performance in classification, image processing and biomedical signal processing[12, 22, 28, 29]. Common and individual feature extraction (CIFE) framework for multi-block data enables the separate extraction of common and individual components by incorporating dimensionality reduction and blind source separation (BSS) methods, and has been successfully applied to classification, clustering and linked BSS problems [11, 24]. Last but not the least, coupled tensor decomposition is also applied to data fusion of low spatial resolution hyperspectral (LRHS) and high spatial resolution multispectral (HRMS) images [30, 31], array signal processing [32], linked prediction[33] and

metabolic physiology [34].

However, due to the nonnegative constraint and high-dimensional nature of tensor data, existing coupled tensor decompsotion methods often suffer from slow convergence speed and low optimization accuracy [35, 36]. Therefore, aiming to effectively and efficiently achieve the joint analysis of tensors with coupled information, we propose two advanced coupled NCPD methods: coupled nonnegative CANDECOMP/PARAFAC decomposition algorithm based on alternating proximal gradient (CoNCPD-APG) and its fast implementation based on low-rank approximation (lraCoNCPD-APG). Specifically, our contributions in this study are listed as follows.

(1) Using the optimal gradient method, we propose an effective CoNCPD-APG algorithm for the joint analysis of multi-block tensors that are partially linked. It can realize the simultaneous decomposition of tensors with excellent decomposition accuracy.

(2) By introducing low-rank approximation, we further proposed an efficient larCoNCPD-APG algorithm, which can greatly reduce the time consumption without losing the decomposition accuracy.

(3) The designed experiments on synthetic data, real-world face image and EEG data prove the practicability and superiority of the proposed algorithms.

The rest of this paper is organized as follows. Section 2 introduces some basic preliminaries and related work. In Section 3, we present the proposed algorithms as well as some theoretical analyses. Experiments on synthetic and real-world data are designed in Section 4 to verify the performance of the proposed algorithms. The last section concludes this paper.

2. Preliminaries and related work

2.1. Notations and tensor operations

Tensor, also known as multi-way array, is the high-order generalization of vector and matrix. The order of a tensor is the number of dimensions, ways or modes of it. Generally, tensors are denoted by calligraphic boldface uppercase

letters, matrices by boldface uppercase letters, vectors by boldface uppercase letters, and scalars by lowercase letters. Table 1 gives a summary of basic notations and mathematical operations throughout this study, and please refer to [8] for a more detailed description of them.

Table 1: Basic notations and mathematical operations

Symbol	**Defination**
$\mathbb{R}$, $\mathbb{R}_+$	real number, nonnegative real number
$\circ$, $\odot$	outer product, Khatri-Rao product
$\circledast$, $\oslash$	element-wise (Hadamard) product, division
$(\cdot)^T$, $\mathrm{vec}(\cdot)$	transpose, vectorization operator
$[\![\cdot]\!]$, $\langle\cdot,\cdot\rangle$, $\|\cdot\|_F$	Kruskal operator, inner product and Frobenius norm
m, $\boldsymbol{m}$, $\boldsymbol{M}$, $\boldsymbol{\mathcal{M}}$	scalar, vector, matrix and tensor
$\boldsymbol{\mathcal{M}}_{(n)}$	mode-n matricization of tensor $\boldsymbol{\mathcal{M}}$
$\mathrm{vec}(\boldsymbol{\mathcal{M}})$	vectorization of tensor $\boldsymbol{\mathcal{M}}$
$\mathrm{ddiag}(\boldsymbol{\mathcal{M}})$	vectorization of super-diagonal elements of tensor $\boldsymbol{\mathcal{M}}$
$\mathrm{ddiag}(\boldsymbol{m})$	tensorization with $\boldsymbol{m}$ on the super-diagonal elements
$\boldsymbol{U}_{i:j,:}$	row-wise submatrix of $\boldsymbol{U}$, ith to jth row
$\boldsymbol{U}_{:,i:j}$	column-wise submatrix of $\boldsymbol{U}$, ith to jth column
$\boldsymbol{U}^{(n)}$	the n-th factor matrix
[1] $\boldsymbol{U}^{\odot}$	$\boldsymbol{U}^{(N)} \odot \boldsymbol{U}^{(N-1)} \odot \cdots \odot \boldsymbol{U}^{(2)} \cdots \odot \boldsymbol{U}^{(1)}$
[2] $\boldsymbol{U}^{\odot -n}$	$\boldsymbol{U}^{(N)} \odot \cdots \odot \boldsymbol{U}^{(n+1)} \odot \boldsymbol{U}^{(n-1)} \cdots \odot \boldsymbol{U}^{(1)}$

[1, 2] $\boldsymbol{U}^{\circledast}$ and $\boldsymbol{U}^{\circledast -n}$ are defined in the similar way.

2.2. Optimal gradient method

Accelerated/Alternating proximal gradient (APG), an accelerated version of proximal gradient (PG, [37]), was originally proposed by Nesterov for smooth

optimization with achieving the convergence rate of $\mathcal{O}(\frac{1}{K^2})$, where K is the number of iterations [38, 39]. For a minimization problem: $\min\{f(x), x \in \mathbb{R}^n\}$, assuming that $f(x) : \mathbb{R}^n \to \mathbb{R}$ is a convex function with Lipschitz continuous gradient f', there will hold that

$$\|f'(x_i) - f'(x_j)\| \leq L \|x_i - x_j\|, \forall x_i, x_j \in \mathbb{R}^n \tag{1}$$

where $L > 0$ is the Lipschitz constant. To obtain an optimal point $\ddot{x}$, two sequences are updated successively in each iteration round (assume at the kth iteration) in APG method [38, 40] as follows:

$$\begin{aligned} x_k = \underset{x}{\operatorname{argmin}} \Big\{ \phi(x, x_{k-1}) = f(x_{k-1}) \\ + \langle x - x_{k-1}, f'(x_{k-1}) \rangle + \frac{L}{2} \|x - x_{k-1}\| \Big\}, \end{aligned} \tag{2}$$

and

$$x_{k+1} = x_k + \frac{\alpha_k - 1}{\alpha_{k+1}}(x_k - x_{k-1}) \tag{3}$$

with

$$\alpha_0 = 1, \ \alpha_{k+1} = \frac{1 + \sqrt{4\alpha_k^2 + 1}}{2} \tag{4}$$

where $\phi(x, x_{k-1})$ denotes the proximal regularized function of $f(x)$ at x_{k-1} and (3) denotes an extrapolated point by combining the points of current and previous iterations. Using the Lagrange multiplier method, from (2), we have

$$x_k \leftarrow \mathcal{P}(x_{k-1} - \frac{1}{L} f'(x_{k-1})) \tag{5}$$

where $\mathcal{P}(\cdot)$ denotes a shrinkage operator predefined by the user. Since each subproblem under block coordinate descent (BCD) framework is a convex function with Lipchitz continuous gradient, APG and its variants have proven to be very efficient for nonnegative matrix/tensor factorization issues and outperform many other competitors [7, 36, 40, 41, 42]. In the sequel, we adopt APG method to solve the coupled nonnegative tensor decomposition problems.

2.3. Coupled NCPD model

Given a set of Nth-order nonnegative tensors $\boldsymbol{\mathcal{M}}^{(s)} \in \mathbb{R}_{+}^{I_1 \times I_2 \times \cdots I_N}$, $s = 1, 2, \cdots, S$, the coupled NCPD model can be expressed as:

$$\begin{aligned} \boldsymbol{\mathcal{M}}^{(s)} \approx \hat{\boldsymbol{\mathcal{M}}}^{(s)} &= \sum_{r=1}^{R^{(s)}} \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)} \\ &= \left[\!\left[\boldsymbol{\mathcal{D}}^{(s)}; \boldsymbol{U}^{(1,s)}, \boldsymbol{U}^{(2,s)}, \cdots, \boldsymbol{U}^{(N,s)} \right]\!\right] \end{aligned} \tag{6}$$

where $\hat{\boldsymbol{\mathcal{M}}}^{(s)} \in \mathbb{R}_{+}^{I_1 \times I_2 \times \cdots I_N}$ denotes the estimated item of $\boldsymbol{\mathcal{M}}^{(s)}$. $\boldsymbol{u}_r^{(n,s)} \in \mathbb{R}_{+}^{I_n}$ denotes the rth column of mode-n factor matrix of sth tensor and $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{u}_1^{(n,s)}, \boldsymbol{u}_2^{(n,s)}, \cdots, \boldsymbol{u}_R^{(n,s)}\right] \in \mathbb{R}_{+}^{I_n \times R^{(s)}}$. $\boldsymbol{\mathcal{D}}^{(s)} \in \mathbb{R}_{+}^{R^{(s)} \times \cdots \times R^{(s)}}$ represents the sth core tensor with non-zero entries $\lambda_r^{(s)}$ only on its super-diagonal elements. $\hat{\boldsymbol{\mathcal{M}}}_r^{(s)} = \lambda_r^{(s)} \boldsymbol{u}_r^{(1,s)} \circ \boldsymbol{u}_r^{(2,s)} \circ \cdots \circ \boldsymbol{u}_r^{(N,s)}$ is termed as a rank-1 tensor generated by outer product of $\boldsymbol{u}_r^{(n,s)}$, $n = 1, 2, \cdots, N$, and $\lambda_r^{(s)}$ is used to represent the scaling of rank-1 tensor. The decomposition of each tensor $\boldsymbol{\mathcal{M}}^{(s)}$ can be regarded as decomposing a high-order tensor into a minimal number of rank-1 tensors, and the minimum number $R^{(s)}$ is named as the rank of the tensor or the number of components.

In coupled NCPD model, we assume that each factor matrix includes two parts and satisfies $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{U}_C^{(n,s)}\ \boldsymbol{U}_I^{(n,s)}\right]$. $\boldsymbol{U}_C^{(n,s)} \in \mathbb{R}_{+}^{I_n \times L_n}$, $0 \leq L_n \leq \min(R^{(s)})$ represents the common information shared by all block-tensors as $\boldsymbol{U}_C^{(n,1)} = \cdots = \boldsymbol{U}_C^{(n,S)} = \boldsymbol{U}_C^{(n)}$, and $\boldsymbol{U}_I^{(n,s)} \in \mathbb{R}_{+}^{I_n \times (R^{(s)} - L_n)}$ denotes the individual part corresponding to individual tensor. L_n represents the number of coupled components between tensors in the nth mode.

3. Proposed Algorithm

This section illustrates how to use APG method or combine APG method and low-rank approximation to solve the coupled NCPD problem. In addition, we give some discussions on the properties of the proposed algorithms as well as some implementation remarks.

3.1. Coupled NCPD using APG

For the coupled NCPD model, the optimization criterion of Euclidean divergence minimization is adopted to minimize the error between the original and estimated tensors. Therefore, given a set of nonnegative tensors $\boldsymbol{\mathcal{M}}^{(s)}$, $s = 1, 2, \ldots, S$, the objective function of coupled NCPD model can be presented as follows:

$$\min_{\boldsymbol{\mathcal{D}}^{(s)}, \boldsymbol{U}^{(n,s)}} F \frac{1}{2} \sum_{s=1}^{S} \left\| \boldsymbol{\mathcal{M}}^{(s)} - \left[\!\left[\boldsymbol{\mathcal{D}}^{(s)}; \boldsymbol{U}^{(1,s)}, \cdots, \boldsymbol{U}^{(N,s)} \right]\!\right] \right\|_F^2 \tag{7}$$

$$\text{s.t., } \boldsymbol{\mathcal{D}}^{(s)} \in \mathbb{R}_+^{R^{(s)} \times \cdots \times R^{(s)}}, \ \boldsymbol{U}^{(n,s)} \in \mathbb{R}_+^{I_n \times R^{(s)}}$$

where $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{U}_C^{(n,s)} \ \boldsymbol{U}_I^{(n,s)} \right]$ and $\boldsymbol{U}_C^{(n,1)} = \cdots = \boldsymbol{U}_C^{(n,S)} = \boldsymbol{U}_C^{(n)}$. According to BCD framework, the coupled NCPD problem can be converted into several subproblems by optimizing $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{U}^{(n,s)}$ alternatively in each iteration. Each subproblem can be regarded as a minimization problem of a continuously differentiable function, which can be solved efficiently by APG methed [40, 41, 42]. Next we provide a solution for coupled NCPD problem based on APG method.

First, regarding core tensor $\boldsymbol{\mathcal{D}}^{(s)}$, we adopt the following update:

$$\boldsymbol{\mathcal{D}}^{(s)} = \underset{\boldsymbol{\mathcal{D}}^{(s)} \geq 0}{\operatorname{argmin}} \left[F(\hat{\boldsymbol{\mathcal{D}}}^{(s)}) + \left\langle \hat{\boldsymbol{\mathcal{G}}}^{(s)}, \boldsymbol{\mathcal{D}}^{(s)} - \hat{\boldsymbol{\mathcal{D}}}^{(s)} \right\rangle + \frac{L_d^{(s)}}{2} \left\| \boldsymbol{\mathcal{D}}^{(s)} - \hat{\boldsymbol{\mathcal{D}}}^{(s)} \right\|_F^2 \right] \tag{8}$$

where $\hat{\boldsymbol{\mathcal{D}}}^{(s)}$ denotes an extrapolated point and $L_d^{(s)}$ denotes the Lipschitz constant of $F'(\boldsymbol{\mathcal{D}}^{(s)})$. $\hat{\boldsymbol{\mathcal{G}}}^{(s)}$ is the block-partial gradient of (7) at $\hat{\boldsymbol{\mathcal{D}}}^{(s)}$, which can be calculated as:

$$\begin{aligned} \hat{\boldsymbol{\mathcal{G}}}^{(s)} = \operatorname{ddiag} \Big[& \left(\boldsymbol{U}^{(s)^T} \boldsymbol{U}^{(s)} \right)^{\circledast} \operatorname{ddiag} \left(\hat{\boldsymbol{\mathcal{D}}}^{(s)} \right) \\ & - \left(\boldsymbol{U}^{(s)^{\odot}} \right)^T \operatorname{vec} \left(\boldsymbol{\mathcal{M}}^{(s)} \right) \Big] \end{aligned} \tag{9}$$

where $\operatorname{vec}(\boldsymbol{\mathcal{M}}^{(s)})$ denotes the vectorization of tensor $\boldsymbol{\mathcal{M}}^{(s)}$ and $\operatorname{ddiag}(\boldsymbol{\mathcal{D}}^{(s)})$ denotes a vector vectorized from the super-diagonal elements of $\boldsymbol{\mathcal{D}}^{(s)}$. The outer-loop notation $\operatorname{ddiag}(\cdot)$ means the tensorization from a vector to a super-diagonal tensor, which is the reverse operation of the inner-loop $\operatorname{ddiag}(\cdot)$. By

keeping all the other variables and using the Lagrange multiplier method, from (8), we can obtain the solution of $\boldsymbol{\mathcal{D}}^{(s)}$ in a closed form as follows:

$$\boldsymbol{\mathcal{D}}^{(s)} = \max\left(0,\ \hat{\boldsymbol{\mathcal{D}}}^{(s)} - \frac{\hat{\boldsymbol{\mathcal{G}}}^{(s)}}{L_d^{(s)}}\right). \tag{10}$$

Second, for the solution of factor matrix $\boldsymbol{U}^{(n,s)}$ (without coupled information, i.e., $L_n = 0$), we consider the updating method as follows:

$$\begin{aligned}\boldsymbol{U}^{(n,s)} = \underset{\boldsymbol{U}^{(n,s)}\geq 0}{\operatorname{argmin}} \sum_{s=1}^{S} \Big[& F(\hat{\boldsymbol{U}}^{(n,s)}) + \left\langle \hat{\boldsymbol{G}}^{(n,s)}, \boldsymbol{U}^{(n,s)} - \hat{\boldsymbol{U}}^{(n,s)} \right\rangle \\ & + \frac{L_u^{(n,s)}}{2} \left\| \boldsymbol{U}^{(n,s)} - \hat{\boldsymbol{U}}^{(n,s)} \right\|_F^2 \Big]\end{aligned} \tag{11}$$

where $\hat{\boldsymbol{U}}^{(n,s)}$ denotes an extrapolated point of $\boldsymbol{U}^{(n,s)}$, $L_u^{(n,s)}$ denotes a Lipschitz constant. The block-partial gradient $\hat{\boldsymbol{G}}^{(n,s)}$ of (7) at $\hat{\boldsymbol{U}}^{(n,s)}$ can be expressed as:

$$\begin{aligned}\hat{\boldsymbol{G}}^{(n,s)} = & \hat{\boldsymbol{U}}^{(n,s)} \boldsymbol{D}^{(s)} \left(\boldsymbol{U}^{(s)^T} \boldsymbol{U}^{(s)} \right)^{\circledast -n} \boldsymbol{D}^{(s)} \\ & - \boldsymbol{\mathcal{M}}_{(n)}^{(s)} \boldsymbol{U}^{(s)^{\odot -n}} \boldsymbol{D}^{(s)}\end{aligned} \tag{12}$$

where $\boldsymbol{\mathcal{M}}_{(n)}^{(s)}$ denotes the mode-n matricization of $\boldsymbol{\mathcal{M}}^{(s)}$. $\boldsymbol{D}^{(s)}$ is a diagonal matrix and its diagonal elements correspond to the super-diagonal elements of core tensor $\boldsymbol{\mathcal{D}}^{(s)}$. The update rule of $\boldsymbol{U}^{(n,s)}$ can be written in the closed form as follows:

$$\boldsymbol{U}^{(n,s)} = \max\left(0,\ \hat{\boldsymbol{U}}^{(n,s)} - \frac{\hat{\boldsymbol{G}}^{(n,s)}}{L_u^{(n,s)}}\right). \tag{13}$$

However, for the factor matrix which includes $\boldsymbol{U}_C^{(n)}$ and $\boldsymbol{U}_I^{(n,s)}$, we need to calculate their solutions separately. Since $\boldsymbol{U}_C^{(n)}$ is shared by all tensors as $\boldsymbol{U}_C^{(n,1)} = \cdots = \boldsymbol{U}_C^{(n,S)} = \boldsymbol{U}_C^{(n)}$, we should combine the information of all tensors to calculate the solution of $\boldsymbol{U}_C^{(n)}$. The solution of individual part $\boldsymbol{U}_I^{(n,s)}$ only needs to consider the corresponding sth-set tensor. Therefore, we have

$$\boldsymbol{U}_C^{(n)} = \max\left(0,\ \hat{\boldsymbol{U}}_C^{(n)} - \frac{\sum_{s=1}^{S} \hat{\boldsymbol{G}}_C^{(n,s)}}{\sum_{s=1}^{S} L_u^{(n,s)}}\right), \tag{14}$$

and

$$\boldsymbol{U}_I^{(n,s)} = \max\left(0,\ \hat{\boldsymbol{U}}_I^{(n,s)} - \frac{\hat{\boldsymbol{G}}_I^{(n,s)}}{L_u^{(n,s)}}\right) \tag{15}$$

where $\hat{\boldsymbol{G}}_C^{(n,s)}$ and $\hat{\boldsymbol{G}}_I^{(n,s)}$ denote the block-partial gradients of (7) at $\hat{\boldsymbol{U}}_C^{(n,s)}$ and $\hat{\boldsymbol{U}}_I^{(n,s)}$, respectively. $\hat{\boldsymbol{U}}_C^{(n,s)}$ and $\hat{\boldsymbol{U}}_I^{(n,s)}$ denote the extrapolated points of $\boldsymbol{U}_C^{(n,s)}$ and $\boldsymbol{U}_I^{(n,s)}$. Moreover, $\hat{\boldsymbol{U}}^{(n,s)} = \left[\hat{\boldsymbol{U}}_C^{(n,s)}\ \hat{\boldsymbol{U}}_I^{(n,s)}\right]$ and $\hat{\boldsymbol{G}}^{(n,s)} = \left[\hat{\boldsymbol{G}}_C^{(n,s)}\ \hat{\boldsymbol{G}}_I^{(n,s)}\right]$.

Consider updating $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{U}^{(n,s)}$ at the kth iteration. The extrapolated points $\hat{\boldsymbol{\mathcal{D}}}_{k-1}^{(s)}$ and $\hat{\boldsymbol{U}}_{k-1}^{(n,s)}$ are defined as

$$\hat{\boldsymbol{\mathcal{D}}}_{k-1}^{(s)} = \boldsymbol{\mathcal{D}}_{k-1}^{(s)} + w_{d,k-1}^{(s)} \left(\boldsymbol{\mathcal{D}}_{k-1}^{(s)} - \boldsymbol{\mathcal{D}}_{k-2}^{(s)}\right), \tag{16}$$

and

$$\hat{\boldsymbol{U}}_{k-1}^{(n,s)} = \boldsymbol{U}_{k-1}^{(n,s)} + w_{u,k-1}^{(n,s)} \left(\boldsymbol{U}_{k-1}^{(n,s)} - \boldsymbol{U}_{k-2}^{(n,s)}\right) \tag{17}$$

where $w_{d,k-1}^{(s)}$ and $w_{u,k-1}^{(n,s)}$ denote the extrapolation weights. Since APG is not a monotone method, i.e., $F(k)$ may not be smaller than $F(k-1)$. Therefore, if $F(k) \geq F(k-1)$ after iteration k, an additional re-updating of $\boldsymbol{U}_k^{(n,s)}$ and $\boldsymbol{\mathcal{D}}_k^{(s)}$ will be taken via $\hat{\boldsymbol{\mathcal{D}}}_{k-1}^{(s)} = \boldsymbol{\mathcal{D}}_{k-1}^{(s)}$ and $\hat{\boldsymbol{U}}_{k-1}^{(n,s)} = \boldsymbol{U}_{k-1}^{(n,s)}$. In each iteration, we perform the optimization with the order $\boldsymbol{\mathcal{D}}^{(1)}, \boldsymbol{\mathcal{D}}^{(2)}, \cdots, \boldsymbol{\mathcal{D}}^{(S)}$ and $\boldsymbol{U}^{(1,1)}, \cdots \boldsymbol{U}^{(1,S)}, \cdots, \boldsymbol{U}^{(N,1)}, \cdots \boldsymbol{U}^{(N,S)}$, which are alternatively updated one after another until convergence. We term the proposed coupled NCPD algorithm based on APG update as CoNCPD-APG and summarize it in Algorithm 1. The detailed derivations and relevant parameter settings are given in the Appendix A.

3.2. Coupled NCPD using APG and low-rank approximation

In CoNCPD-APG algorithm, the time consumption of updating $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{U}^{(n,s)}$ is mainly attributed to the multiplication of $(\boldsymbol{U}^{(s)^{\odot}})^T \mathrm{vec}(\boldsymbol{\mathcal{M}}^{(s)})$ and $\boldsymbol{\mathcal{M}}_{(n)}^{(s)} \boldsymbol{U}^{(s)^{\odot -n}}$ in (9) and (12), and it will be increasingly serious especially for the tensors with large dimensionality. Specifically, in each iteration, let $R^{(s)} = R$, the computational complexity of $(\boldsymbol{U}^{(s)^{\odot}})^T \mathrm{vec}(\boldsymbol{\mathcal{M}}^{(s)})$ reaches $\mathcal{O}(SR\prod_n I_n)$ and $\boldsymbol{\mathcal{M}}_{(n)}^{(s)} \boldsymbol{U}^{(s)^{\odot -n}}$ has the complexity of $\mathcal{O}(NSR\prod_n I_n)$. The applications of low-rank approximation in nonnegative matrix/tensor factorization have demonstrated their performance improvement in terms of computational efficiency while maintaining computational accuracy [35, 36, 43]. Generally, the unconstrained CPD of a tensor converges in dozens of iterations and is consid-

Algorithm 1: CoNCPD-APG algorithm

Input: $\boldsymbol{\mathcal{M}}^{(s)}$, L_n and $R^{(s)}$, $n{=}1,\cdots,N$, $s{=}1,\cdots,S$

Initialization:

$\boldsymbol{U}^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $\boldsymbol{\mathcal{M}}^{(s)}_{(n)}$, $n=1,\cdots,N$, $s=1,\cdots,S$

for $k=1,2,\cdots,$ *MaxIt* **do**

  for $s=1,\cdots,S$ **do**

    Calculate $\hat{\boldsymbol{\mathcal{G}}}^{(s)}_{k-1}$ and $\hat{\boldsymbol{\mathcal{D}}}^{(s)}_{k-1}$ via (9) and (16)

    Update $\boldsymbol{\mathcal{D}}^{(s)}_{k}$ via (10)

  end

  for $n=1,2,\cdots,N$ **do**

    for $s=1,\cdots,S$ **do**

      Calculate $\hat{\boldsymbol{G}}^{(n,s)}_{k-1}$ and $\hat{\boldsymbol{U}}^{(n,s)}_{k-1}$ via (12) and (17)

      Update $\boldsymbol{U}^{(n,s)}_{k}$ via (13), (14) and (15)

    end

  end

  if $F(k) \geq F(k-1)$ **then**

    $\hat{\boldsymbol{\mathcal{D}}}^{(s)}_{k-1} = \boldsymbol{\mathcal{D}}^{(s)}_{k-1}$, $\hat{\boldsymbol{U}}^{(n,s)}_{k-1} = \boldsymbol{U}^{(n,s)}_{k-1}$

    Reupdate $\boldsymbol{\mathcal{D}}^{(s)}_{k}$, $\boldsymbol{U}^{(n,s)}_{k}$ via (10), (13), (14) and (15)

  end

  if *stopping criteron is satisfied* **then**

    return

    $\boldsymbol{U}^{(n,s)}_{k}$, $\boldsymbol{\mathcal{D}}^{(s)}_{k}$, $n=1,\cdots,N$, $s=1,\cdots,S$

  end

end

Output: $\boldsymbol{U}^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $n=1,\cdots,N$, $s=1,\cdots,S$

ered faster than its counterpart with nonnegative constraint. Therefore, aiming to reduce the computational complexity of CoNCPD-APG algorithm, we consider introducing the low-rank approximation of $\boldsymbol{\mathcal{M}}^{(s)}$ before performing the actual coupled decomposition. Suppose that $\left[\!\left[\tilde{\boldsymbol{U}}^{(1,s)}, \tilde{\boldsymbol{U}}^{(2,s)}, \cdots, \tilde{\boldsymbol{U}}^{(N,s)}\right]\!\right]$ is the rank-$\tilde{R}^{(s)}$ approximation of $\boldsymbol{\mathcal{M}}^{(s)}$ obtained by the unconstrained CPD, $\tilde{\boldsymbol{U}}^{(n,s)} \in \mathbb{R}^{I_n \times \tilde{R}^{(s)}}$, $\tilde{R}^{(s)} \leq R^{(s)}$, thus the cost function in (7) can be represented as the following optimization problem with fixed $\tilde{\boldsymbol{U}}^{(1,s)}, \tilde{\boldsymbol{U}}^{(2,s)}, \cdots, \tilde{\boldsymbol{U}}^{(N,s)}$ as:

$$\min_{\boldsymbol{\mathcal{D}}^{(s)}, \boldsymbol{U}^{(n,s)}} \mathcal{F} \frac{1}{2} \sum_{s=1}^{S} \left\| \left[\!\left[\tilde{\boldsymbol{U}}^{(1,s)}, \tilde{\boldsymbol{U}}^{(2,s)}, \cdots, \tilde{\boldsymbol{U}}^{(N,s)}\right]\!\right] - \left[\!\left[\boldsymbol{\mathcal{D}}^{(s)}; \boldsymbol{U}^{(1,s)}, \boldsymbol{U}^{(2,s)}, \cdots, \boldsymbol{U}^{(N,s)}\right]\!\right] \right\|_F^2 \tag{18}$$
$$\text{s.t. } \boldsymbol{\mathcal{D}}^{(s)} \in \mathbb{R}_+^{I_1 \times \cdots \times I_N},\ \boldsymbol{U}^{(n,s)} \in \mathbb{R}_+^{I_n \times R^{(s)}},\ \tilde{\boldsymbol{U}}^{(n,s)} \in \mathbb{R}^{I_n \times \tilde{R}^{(s)}},\ \tilde{R}^{(s)} \leq R^{(s)}.$$

In other words, instead of loading the original tensor $\boldsymbol{\mathcal{M}}^{(s)}$ directly into the iterations, we first split the tensor into smaller compressed matrices, such as $\tilde{\boldsymbol{U}}^{(1,s)}, \tilde{\boldsymbol{U}}^{(2,s)}, \cdots, \tilde{\boldsymbol{U}}^{(N,s)}$, and then bring them into the decomposition iterations, which can greatly reduce the time and space complexities of algorithms [35]. Via low-rank approximation, $\text{vec}(\boldsymbol{\mathcal{M}}^{(s)})$ and $\boldsymbol{\mathcal{M}}_{(n)}^{(s)}$ in (9) and (12) can be respectively expressed by $\text{vec}(\boldsymbol{\mathcal{M}}^{(s)}) = \tilde{\boldsymbol{U}}^{(s)^{\odot}} \text{ddiag}(\boldsymbol{\mathcal{I}})$ and $\boldsymbol{\mathcal{M}}_{(n)}^{(s)} = \tilde{\boldsymbol{U}}^{(s,n)} \left(\tilde{\boldsymbol{U}}^{(s)^{\odot -n}}\right)^T$, and $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{\tilde{R}^{(s)} \times \cdots \times \tilde{R}^{(s)}}$ denotes a core tensor with all superdiagonal elements being 1. This thereby leads to

$$\begin{aligned} \left(\boldsymbol{U}^{(s)^{\odot}}\right)^T \text{vec}\left(\boldsymbol{\mathcal{M}}^{(s)}\right) &= \left(\boldsymbol{U}^{(s)^{\odot}}\right)^T \tilde{\boldsymbol{U}}^{(s)^{\odot}} \text{ddiag}(\boldsymbol{\mathcal{I}}) \\ &= \left(\boldsymbol{U}^{(s)^T} \tilde{\boldsymbol{U}}^{(s)}\right)^{\circledast} \text{ddiag}(\boldsymbol{\mathcal{I}}), \end{aligned} \tag{19}$$

and

$$\begin{aligned} \boldsymbol{\mathcal{M}}_{(n)}^{(s)} \boldsymbol{U}^{(s)^{\odot -n}} &= \tilde{\boldsymbol{U}}^{(s,n)} \left(\tilde{\boldsymbol{U}}^{(s)^{\odot -n}}\right)^T \boldsymbol{U}^{(s)^{\odot -n}} \\ &= \tilde{\boldsymbol{U}}^{(s,n)} \left(\tilde{\boldsymbol{U}}^{(s)^T} \boldsymbol{U}^{(s)}\right)^{\circledast -n}. \end{aligned} \tag{20}$$

By virtue of low-rank approximation, only very small matrices are involved to perform the multiplications in (19) and (20), and the heavy cost of Khatri-Rao product can also be avoided. In addition, the computational complexities of $(\boldsymbol{U}^{(s)^{\odot}})^T \text{vec}(\boldsymbol{\mathcal{M}}^{(s)})$ and $\boldsymbol{\mathcal{M}}_{(n)}^{(s)} \boldsymbol{U}^{(s)^{\odot -n}}$ are respectively reduced to $\mathcal{O}(SR\tilde{R}\sum_n I_n)$

and $\mathcal{O}(NSR\tilde{R}\sum_n I_n)$ via the transformations of (19) and (20) (here we set $\tilde{R}^{(s)} = \tilde{R}$).

Overall, to develop an efficient coupled tensor decomposition algorithm, we further propose the lraCoNCPD-APG algorithm based on APG algorithm and low-rank approximation. The implementation of lraCoNCPD-APG algorithm includes two steps: (i) performing unconstrained CPD of tensors $\boldsymbol{\mathcal{M}}^{(s)}$ successively to achieve low-rank approximation as $\boldsymbol{\mathcal{M}}^{(s)} \approx \left[\!\left[\tilde{\boldsymbol{U}}^{(1,s)}, \tilde{\boldsymbol{U}}^{(2,s)}, \cdots, \tilde{\boldsymbol{U}}^{(N,s)}\right]\!\right]$; (ii) updating $\boldsymbol{U}_k^{(n,s)}$ and $\boldsymbol{\mathcal{D}}_k^{(s)}$ via solving the optimization problem in (18) with fixed $\tilde{\boldsymbol{U}}^{(1,s)}, \tilde{\boldsymbol{U}}^{(2,s)}, \cdots, \tilde{\boldsymbol{U}}^{(N,s)}$. The framework of lraCoNCPD-APG algorithm is presented in Algorithm 2.

Algorithm 2: larCoNCPD-APG algorithm

Input: $\boldsymbol{\mathcal{M}}^{(s)}$, L_n, and $R^{(s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$

Initialization:

$\boldsymbol{U}^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $\boldsymbol{\mathcal{M}}_{(n)}^{(s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$

Calculate $\tilde{\boldsymbol{U}}^{(n,s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$ via unconstrained CPD on $\boldsymbol{\mathcal{M}}^{(s)}$, $s = 1, \cdots, S$

for $k = 1, 2, \cdots, MaxIt$ **do**

  Repeat the 4th to 21st command lines of Algorithm 1 except the 5th and 10th lines:

  Calculate $\hat{\boldsymbol{\mathcal{G}}}_{k-1}^{(s)}$ and $\hat{\boldsymbol{G}}_{k-1}^{(n,s)}$ via (9) and (12) by introducing (19) and (20)

end

Output: $\boldsymbol{U}^{(n,s)}$, $\boldsymbol{\mathcal{D}}^{(s)}$, $n = 1, \cdots, N$, $s = 1, \cdots, S$

3.3. Remarks and discussions

3.3.1. Acceleration strategy

Even though we have introduced low-rank approximation to reduce the computation load of CoNCPD-APG algorithm, there is still some way to speed up the algorithm. In conventional NCPD problem, core tensor $\boldsymbol{\mathcal{D}}$ is generally merged into the factor matrices, so we do not need to calculate additional core

tensor, which can reduce the computation load to a certain extent. Analogously, this strategy can be extended to the coupled NCPD problem, but only for the cases where all N modes are not fully coupled between tensors [12], e.g., $\exists\, n, L_n = 0$. However, for the cases of $\forall\, n, L_n > 0$, the core tensors $\boldsymbol{\mathcal{D}}^{(s)}$ are required and defined to differentiate the magnitude of corresponding components (rank-1 tensor) between tensors. In *Experiment 1*, we design the coupled NCPD problem using synthetic data and verify the importance of core tensors in some cases. In *Experiment 3*, the scaling features of corresponding brain activities provided by core tensors extracted from ERP tensors are used to classify patient and normal people groups. Therefore, this acceleration strategy depends on the coupling constraints of tensor data.

3.3.2. Normalization

According to coupling constraints in this study, we have explained the necessity of retaining core tensors in optimization. However, in each iteration of the optimization process, we do not consider the column normalization of factor matrices (i.e., the column scalings of factor matrices are absorbed into core tensors). Because it is actually equivalent to applying additional normalization constraint to the coupled NCPD optimization problem. Then each subproblem under BCD framework will become a nonconvex problem, which is not easy to solve. Therefore, in the following experiments, we will not add such normalization constraint even in the HALS-based algorithm. Moreover, this constraint may generate a negative impact if we further consider imposing sparse constraint to the optimization problem.

3.3.3. Computational complexity

In this study, for the computational complexity, we mainly refer to the time complexity based on the multiplication operations. From Section 3, we can note that the main computational load is dominated by the updates of $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{U}^{(n,s)}$, especially the calculation of block-partial gradients $\tilde{\boldsymbol{\mathcal{G}}}^{(s)}$ and $\tilde{\boldsymbol{G}}^{(n,s)}$ in (9) and (12). The specific costs of calculating them (nth mode of sth tensor) are

Table 2: Time complexity per iteration in CoNCPD-APG and *tra*CoNCPD-APG algorithms: nth mode of sth tensor

Equation	**Operation**	**Description**	**Input size**	**Output size**	**Cost**
(9)	① $= \left(\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\right)^{\circledast}$	Hadamard product	$I_n \times R, n = 1, 2, \cdots, N$	$R \times R$	$R^2 \sum_{n=1}^{N} I_n + R^2 N$
(9)	② $= \left(\boldsymbol{U}^{(s)^{\odot}}\right)^T$	Khatri-Rao product	$I_n \times R, n = 1, 2, \cdots, N$	$R \times \prod_{n=1}^{N} I_n$	$R \prod_{n=1}^{N} I_n$
(9)	③ $=$ ② $\cdot \operatorname{vec}\left(\boldsymbol{\mathcal{M}}^{(s)}\right)$	Matrix product	$R \times \prod_{n=1}^{N} I_n$, $\prod_{n=1}^{N} I_n \times 1$	$R \times 1$	$R \prod_{n=1}^{N} I_n$
(12)	④ $= \left(\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\right)^{\circledast_{-n}}$	Hadamard product	$I_m \times R, m = 1, \cdots, N, m \neq n$	$R \times R$	$R^2 \sum_{m \neq n}^{N} I_m + R^2 (N-1)$
(12)	⑤ $= \hat{\boldsymbol{U}}^{(n,s)} \cdot$ ④	Matrix product	$I_n \times R, R \times R$	$I_n \times R$	$R^2 I_n$
(12)	⑥ $= \boldsymbol{U}^{(s)^{\odot_{-n}}}$	Khatri-Rao product	$I_m \times R, m = 1, \cdots, N, m \neq n$	$\prod_{m \neq n}^{N} I_m \times R$	$R \prod_{m \neq n}^{N} I_m$
(12)	⑦ $= \boldsymbol{\mathcal{M}}_{(n)}^{(s)} \cdot$ ⑥	Matrix product	$I_n \times \prod_{m \neq n}^{N} I_m, \prod_{m \neq n}^{N} I_m \times R$	$I_n \times R$	$R \prod_{n=1}^{N} I_n$
(19)	⑧ $= \left(\boldsymbol{U}^{(s)^T}\tilde{\boldsymbol{U}}^{(s)}\right)^{\circledast}$	Hadamard product	$I_n \times R, n = 1, 2, \cdots, N$	$R \times R$	$R^2 \sum_{n=1}^{N} I_n + R^2 N$
(20)	⑨ $= \left(\tilde{\boldsymbol{U}}^{(s)^T}\boldsymbol{U}^{(s)}\right)^{\circledast_{-n}}$	Hadamard product	$I_m \times R, m = 1, \cdots, N, m \neq n$	$R \times R$	$R^2 \sum_{m \neq n}^{N} I_m + R^2 (N-1)$
(20)	⑩ $= \tilde{\boldsymbol{U}}^{(s,n)} \cdot$ ⑨	Matrix product	$I_n \times R, R \times R$	$I_n \times R$	$R^2 I_n$

[1] Here let $R^{(s)} = \tilde{R}^{(s)} = R, s = 1, 2, \cdots, S$.

listed in Table 2. Let $R^{(s)} = R$, taking into consideration that there are N modes and S tensors, the total time complexity for each iteration of CoNCPD-APG algorithm reaches $\mathcal{O}(NSR\prod_{n=1}^{N} I_n)$. By introducing low-rank approximation, (9) and (12) can be calculated using (19) and (20), the costs of which are also given in Table 2. Let $\tilde{R}^{(s)} = R$, the overall computation load per iteration of lraCoNCPD-APG algorithm is reduced to $\mathcal{O}(NSR^2\sum_{n=1}^{N} In)$.

3.3.4. Termination criteria

In this study, we consider two iteration termination parameters: the change of relative error (RelErr) and the maximum number of iterations (MaxIt). In the CoNCPD model, we define RelErr $\triangleq \sum_{s=1}^{S}\left[\|\boldsymbol{\mathcal{M}}^{(s)} - \hat{\boldsymbol{\mathcal{M}}}^{(s)}\|_F / \|\boldsymbol{\mathcal{M}}^{(s)}\|_F\right]$. Furthermore, we stipulate $|\text{RelErr}_{\text{new}} - \text{RelErr}_{\text{old}}| < \varepsilon$, i.e., the adjacent RelErr change should be smaller than the preselected threshold. In this study, we choose $\varepsilon = 1e-8$ and MaxIt=1000 in CoNCPD algorithms and $\varepsilon = 1e-4$ and MaxIt=200 in unconstrained CPD algorithm.

4. Experiments

In this section, aiming to examine and demonstrate the superior performance of CoNCPD-APG and lraCoNCPD-APG algorithms on coupled NCPD problem, we design and perform three experiments on synthetic data, face image data and real-world electroencephalography (EEG) data. We adopt alternating least squares (ALS, [4]) algorithm to implement the low-rank approximation, which has proven to be a reasonable solver for unconstrained CPD problems [36]. The optimization strategies including fast hierarchical alternating least squares (fHALS, [28, 29, 44]), multiplicative updating (MU, [45, 46]) and ALS are used as the competitors to the proposed algorithms. Additionally, we also introduce the low-rank approximation to fHALS-based algorithm.

For algorithm performance comparison, we adopt decomposition quality indicators including RelErr, objective function value (ObjFun), tensor fitting value (TenFit) and performance index (PI), as well as execution time (Time).

TenFit $\triangleq \frac{1}{S}\sum_{s=1}^{S}\left[1-\|\boldsymbol{\mathcal{M}}^{(s)}-\hat{\boldsymbol{\mathcal{M}}}^{(s)}\|_F/\|\boldsymbol{\mathcal{M}}^{(s)}\|_F\right]$, and it is used to present the fittings between original and recovered tensors. PI is used to evaluate the recovery accuracy of factor matrices and defined as

$$\text{FacFit} \triangleq \frac{1}{2R(R-1)}\left[\sum_{i=1}^{R}\left(\sum_{j=1}^{R}\frac{|g_{ij}|}{\max_k|g_{ik}|}\right)+\sum_{i=1}^{R}\left(\sum_{j=1}^{R}\frac{|g_{ji}|}{\max_k|g_{ki}|}\right)\right] \quad (21)$$

where g_{ij} denotes the (i,j)th element of $\boldsymbol{G}=(\bar{\boldsymbol{U}})^{\dagger}\boldsymbol{U}$. $\bar{\boldsymbol{U}}$ is the recovered estimation of factor matrix $\boldsymbol{U}$ and $\dagger$ denotes the pseudo inverse operator. The small value of PI indicates an accurate estimation of the true factor matrix. The input factor matrices and core tensors are initialized with uniformly distributed pseudorandom numbers generated by matlab function `rand`. Signal-to-noise ratio (SNR) is defined as SNR $= 10\log_{10}(p_s/p_n)$, where p_s and p_n denote signal level and noise level, respectively.

The experiments are carried out with the following computer configurations: CPU-Intel Core i5-7500 @3.40Hz; Memory-16.00 Gb; System-64-bit Windows 10; Software-matlab R2016b.

Experiment 1 Synthetic data. In this part, we design an experiment to illustrate the performance of CoNCPD-APG and lraCoNCPD-APG algorithms in terms of decomposition efficiency and accuracy using synthetic data, and compare them with or without core tensors during the optimization process. Here the algorithms without optimizing core tensors are respectively termed as CoNCPD-APG-NC and lraCoNCPD-APG-NC. We construct 10 third-order tensors partially coupled in three modes according to equation (6) and set the size of tensors to $I_1 = 8n$, $I_2 = 9n$ and $I_3 = 10n$, where n ranges from 2 to 12. The number of components and coupled components of tensors are set to $R^{(s)} = \text{round}(I_2/2)$ and $L_{1,2,3} = \text{round}(I_2/4)$, and `round` is a matlab function that towards nearest integer. The performance curves including PI, Tenfit, Time and ObjFun versus the size of tensors are illustrated in Figure 1 with contributions from 50 independent runs under the SNR of 20dB.

From Figure 1, we can see that CoNCPD-APG achieves the best decomposition accuracy in terms of PI, Tenfit and ObjFun, followed by lraCoNCPD-APG,

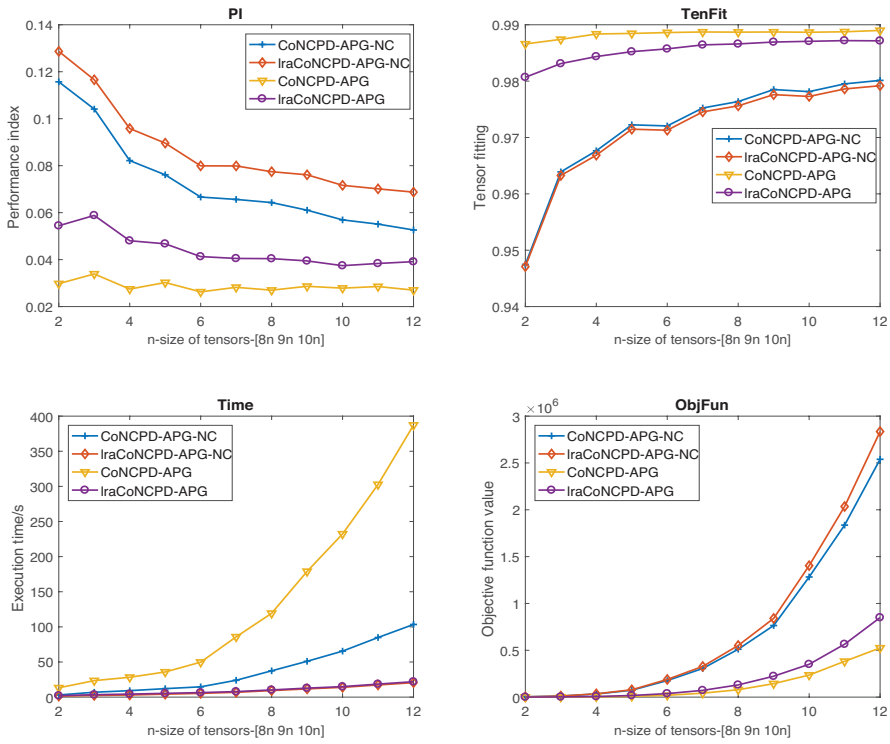


Figure 1: PI, Tenfit, Time and ObjFun curves of all compared algorithms versus the size of tensors under SNR=20dB and 50 independent runs.

CoNCPD-APG-NC and lraCoNCPD-APG-NC. However, CoNCPD-APG is the most time-consuming, followed by CoNCPD-APG-NC and then by lraCoNCPD-APG and lraCoNCPD-APG-NC. The introduction of low-rank approximation can greatly reduce the execution time, and this advantage becomes more significant as the size of tensors increases. Meanwhile, its cost is only a slight reduction in decomposition accuracy. Since the core tensors are not updated in the optimization process, the time consumption of CoNCPD-APG-NC is also alleviated, but the decomposition accuracy of CoNCPD-APG-NC and lraCoNCPD-APG-NC is reduced to some extent.

Experiment 2 Face image data. In this experiment, we use the extended Yale B face database[1] for coupled tensor decomposition analysis via image reconstruction and denoising. This database contains gray-scale face images of 38 subjects obtained under 9 poses and 64 illumination conditions [47]. For this database, we only use the cropped images under frontal pose of all illuminations[2] [48]. Each subject corresponds to 64 images, and each cropped image is resized to 32×32 pixels. Finally, we construct 31 third-order tensors by stacking corresponding face images of each subject along illumination conditions, and the size of each tensor is 32 pixels ×32 pixels ×64 conditions (Data of 7 subjects were not used because of incompleteness).

In terms of the number of components for each tensor, a simple explained variance-based method is adopted in this study. Through unfolding along the first mode, each tensor can be reconstructed into a matrix with the size of 32×2048. Then performing principle component analysis (PCA) on the matrices successively, and the number of principle components with a total explained variance of 99% is regarded as the corresponding component number. Using this method, the component number of tensors for 31 subjects are separately selected. In this experiment, we assume that the coupling information between tensors exists in their three modes, and totally 20 coupled components among

[1] http://vision.ucsd.edu/ leekc/ExtYaleDatabase/ExtYaleB.html
[2] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

Table 3: Performance comparison of image reconstruction and denoising on Yale B face database based on coupled NCPD model (with 0.1 salt&pepper noise)

Methods	ALS	fHALS	lra&fHALS	APG	lra&APG
RelErr	18.37	15.92	15.98	15.56	15.59
TenFit	0.4075	0.4866	0.4844	0.4980	0.4971
ObjFun	3.80e9	2.81e9	2.84e9	2.68e9	2.69e9
Time	101.50	94.59	19.84	95.53	12.58
PNSR	16.02	19.93	19.78	20.80	20.73

Table 4: Performance comparison of image reconstruction and denoising on Yale B face database based on coupled NCPD model (with 0.0001 salt&pepper noise)

Methods	ALS	fHALS	lra&fHALS	APG	lra&APG
RelErr	13.68	8.96	9.08	7.61	7.65
TenFit	0.5587	0.7108	0.7071	0.7544	0.7531
ObjFun	1.74e9	7.25e8	7.43e8	5.07e8	5.12e8
Time	103.84	96.34	20.59	87.70	12.51
PNSR	16.51	20.16	20.03	21.52	21.47

tensors were extracted.

We compare CoNCPD-APG and lraCoNCPD-APG algorithms with their competitors on two noisy image datasets of 0.1 and 0.0001 salt-and-pepper noise. The value of peak-signal-to-noise ratio (PSNR) is also used to measure the quality of reconstructed face images. Table 3 and 4 give the algorithm performance comparison averaged from 100 independent runs under two noise settings. From the tables, we can see that the proposed APG-based algorithms are superior to other algorithms, as they obtain lower RelErr and ObjFun values, as well as higher tensor fittings and PSNRs. It indicates that the proposed algorithms have higher decomposition accuracy and stronger image reconstruction capabilities. More importantly, by introducing low-rank approximation strategy, lraCoNCPD-APG and lraCoNCPD-fHALS algorithms can greatly reduce execution time while maintaining decomposition performance. It should be noted that the execution time of lra-based algorithms showed in tables includes the running time of the unconstrained CPD and the running time of APG/fHALS optimization in the CoNCPD problem. In addition, Figure 2 illustrates the PSNR value curves of all the compared algorithms over 100 independent runs, indicating that the proposed CoNCPD-APG and lraCoNCPD-APG algorithms have excellent decomposition stability compared with competitors.

Experiment 3 Real-world ERP data. In this experiment, we compare the proposed CoNCPD-APG and lraCoNCPD-APG algorithms with CoNCPD-MU, CoNCPD-fHALS and lraCoNCPD-fHALS algorithms in the multi-domain feature extraction of event-related potential (ERP) data [3] (ALS-based algorithm is excluded due to its poor performance). Two groups of data are chosen: 21 children with reading disability (RD) and 21 children with attention deficit (AD), aiming to acquire multi-domain features of ERP data which can better discriminate the two groups. Using complex Morlet wavelet transform, we generate the third-order tensors of 42 subjects (21 RD & 21 AD) with the size of 9 (channels) $\times$ 71 (frequency bins) $\times$ 60 (temporal points) to testify the effectiveness and

[3] http://www.escience.cn/people/cong/AdvancedSP_ERP.html

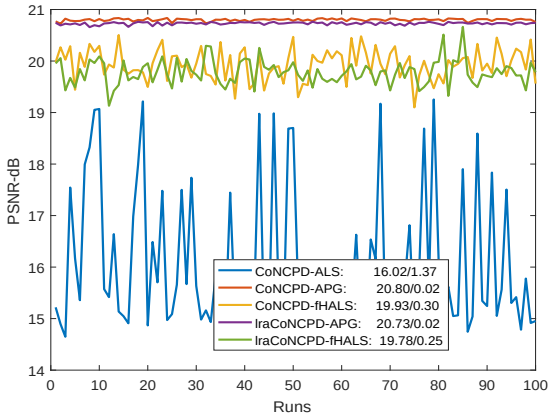


Figure 2: PSNR values of all compared algorithms for image reconstruction and denoising with 0.1 salt-and-pepper noise over 100 independent runs based on coupled NCPD model and their means and standard deviations (SDs)

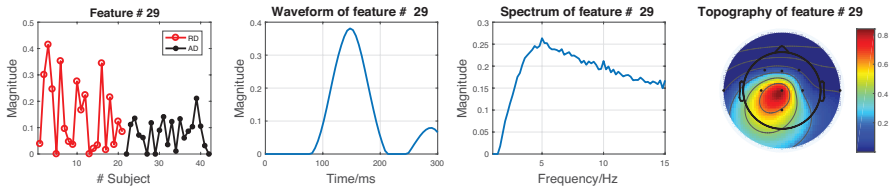


Figure 3: An example of multi-domain feature and its related temporal, spectral and spatial components of ERP data extracted by CoNCPD-APG algorithm

practicality of coupled tensor decomposition. Following [49], we set the number of components to $R^{(1)} = R^{(2)} = \cdots R^{(42)} = 36$. Considering the nature of ERP data, we assume that these third-order ERP tensors are coupled in spatial, spectral and temporal modes, and we directly set the number of coupled components to 36.

ERP data are acquired through repeated presentation of stimuli, which makes their properties in temporal, spectral and spatial domains roughly known before they are actually extracted. According to prior knowledge given in [49], we can select the expected multi-domain features and their corresponding temporal, spectral and spatial components from the decomposition results

of ERP data. Figure 3 gives an example illustration of multi-domain features and their corresponding components extracted by CoNCPD-APG algorithm in the 1st run. For multi-domain feature shown in the figure, statistical analysis using t-test reveals the significant difference between RD and AD groups with $t_{20} = 2.419$, $p = 0.025$. The relevant temporal component (latency peaks around 150 ms) and spectral component (spectrum peaks around 5 Hz) closely match the property of mismatch negativity component [49]. The corresponding topography denotes that the difference of RD and AD groups may appear in the central and left hemisphere [49].

We adopt three steps to verify the stability of multi-domain feature extraction of all the compared algorithms in 100 runs. (1) We select the multi-domain features and their parallel three components in the 1st runs of 5 algorithms. (2) We average the selected ones separately as a set of template patterns, which are termed as $\boldsymbol{u}_{\text{temp}}^{\text{fea}}$, $\boldsymbol{u}_{\text{temp}}^{\text{tem}}$, $\boldsymbol{u}_{\text{temp}}^{\text{spe}}$ and $\boldsymbol{u}_{\text{temp}}^{\text{spa}}$. (3) We define the maximum correlation coefficient (MCC) between template patterns and feature-based components of kth runs as follows

$$\begin{aligned}\text{MCC}(k) = \max\big[&\text{corr}(\boldsymbol{u}_{\text{temp}}^{\text{fea}}, \boldsymbol{U}_k^{\text{fea}}) \circledast \text{corr}(\boldsymbol{u}_{\text{temp}}^{\text{tem}}, \boldsymbol{U}_k^{\text{tem}}) \\ &\circledast \text{corr}(\boldsymbol{u}_{\text{temp}}^{\text{spe}}, \boldsymbol{U}_k^{\text{spe}}) \circledast \text{corr}(\boldsymbol{u}_{\text{temp}}^{\text{spa}}, \boldsymbol{U}_k^{\text{spa}})\big]\end{aligned} \tag{22}$$

where k denotes the run number and `corr` is a matlab function which returns a vector containing the pairwise linear correlation coefficient between $\boldsymbol{u}$ and $\boldsymbol{U}$. $\boldsymbol{U}_k^{\text{fea}}$, $\boldsymbol{U}_k^{\text{tem}}$, $\boldsymbol{U}_k^{\text{spe}}$ and $\boldsymbol{U}_k^{\text{spa}}$ represent multi-domain features and their corresponding temporal, spectral and spatial components in the kth run, respectively. Obviously, if the MCC is close to 1, it means that the extraction of multi-domain features is more stable.

Table 5 gives the average RelErr, ObjFun, TenFit and Time from 100 independent runs for ERP data, as well as the means and SDs of MCCs of multi-domain features. From the table, we see that the proposed CoNCPD-APG and lraCoNCPD-APG algorithms are superior to competitors in terms of decomposition accuracy and multi-domain feature extraction stability. Interestingly, the MU-based couple method achieves better performance than the fHALS-based

Table 5: Performance comparison of the algorithms in multi-domain feature extraction of ERP data based on coupled NCPD model

Methods	MU	fHALS	lra&fHALS	APG	lra&APG
RelErr	6.83	7.06	7.02	6.34	6.34
TenFit	0.8373	0.8319	0.8327	0.8490	0.8490
ObjFun	6.34e5	6.87e5	6.81e5	5.48	5.47e5
Time	164.30	192.71	39.35	183.91	27.45
MCC-Mean	0.7293	0.8782	0.8621	0.8896	0.8916
MCC-SD	0.1770	0.0973	0.1171	0.0956	0.0809

algorithms in accuracy but has the worst multi-domain feature extraction stability. This experiment also proves that the low-rank approximation in coupling analysis of large-scale tensors can greatly improve computation efficiency without losing the decomposition accuracy.

5. Conclusion

In this study, we considered the coupled tensor decomposition problem, aiming to solve the simultaneous decomposition of nonnegative multi-block tensors. To improve convergence speed and optimization accuracy, we first proposed a coupled nonnegative CANDECOMP/PARAFAC decomposition algorithm based as alternating proximal gradient (APG) method (CoNCPD-APG). Then by combining APG and low-rank approximation, we further proposed the lraCoNCPD-APG algorithm. We also gave some discussions on the properties of the proposed algorithms as well as some implementation remarks. Experiments of synthetic data, real-world face image data and event-related potential (ERP) data were conducted to compare the proposed algorithms with fast hierarchical alternating least squares (fHALS), multiplicative updating (MU) and alternating least squares (ALS)-based algorithms in the designed coupled NCPD

problems. The experiment results illustrated that the proposed algorithms are superior to competitors in terms of decomposition accuracy, image reconstruction capability and multi-domain feature extraction stability, and also demonstrated that the introduction of low-rank approximation can greatly improve the computation efficiency without compromising the decomposition quality. Determining the number of coupled components depends on the validity of potential assumptions and relevant prior knowledge. So far, its selection in real-world applications is still subjective to a certain extent, which remains it an open issue and will be one of our future works.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant Nos. 91748105 and 81471742), the Fundamental Research Funds for the Central Universities [DUT2019] in Dalian University of Technology in China, and the scholarships from China Scholarship Council (No. 201706060262).

Appendix A.

For completeness of this paper, in this section, some steps of algorithm derivations and relevant parameter settings will be further explained below.

Appendix A.1. CoNCPD-APG algorithm

Following [40, 41], by using Lagrange multiplier method, we obtain (10), (13), (14) and (15) from (8) and (11). When updating the core tensor $\boldsymbol{\mathcal{D}}^{(s)}$, by keeping all factor matrices $\boldsymbol{U}^{(n,s)}$ fixed, we first convert (7) to

$$F_d = \frac{1}{2}\left\|\mathrm{vec}\big(\boldsymbol{\mathcal{M}}^{(s)}\big) - \boldsymbol{U}^{(s)\odot}\mathrm{ddiag}\big(\boldsymbol{\mathcal{D}}^{(s)}\big)\right\|_F^2 \tag{A.1}$$

where $\mathrm{vec}(\boldsymbol{\mathcal{M}}^{(s)})$ denotes the vectorization of tensor $\boldsymbol{\mathcal{M}}^{(s)}$. Mathematically, the squared Frobenius norm of a matrix can be replaced by the trace of multiplication of the matrix and its transpose. Then (A.1) can be represented as:

$$F_d = \frac{1}{2}\mathrm{tr}\left[\left(\mathrm{vec}\big(\boldsymbol{\mathcal{M}}^{(s)}\big) - \boldsymbol{U}^{(s)\odot}\mathrm{ddiag}\big(\boldsymbol{\mathcal{D}}^{(s)}\big)\right)^T \left(\mathrm{vec}\big(\boldsymbol{\mathcal{M}}^{(s)}\big) - \boldsymbol{U}^{(s)\odot}\mathrm{ddiag}\big(\boldsymbol{\mathcal{D}}^{(s)}\big)\right)\right] \tag{A.2}$$

According to trace property, the block-partial gradient $\hat{\boldsymbol{\mathcal{G}}}^{(s)}$ of (A.2) with respect to $\hat{\boldsymbol{\mathcal{D}}}^{(s)}$ can be calculated by

$$\begin{aligned}\hat{\boldsymbol{\mathcal{G}}}^{(s)} &= \nabla_{\hat{\boldsymbol{\mathcal{D}}}^{(s)}} F_d \\ &= \mathrm{ddiag}\left[\left(\boldsymbol{U}^{(s)\odot}\right)^T \boldsymbol{U}^{(s)\odot}\mathrm{ddiag}\left(\hat{\boldsymbol{\mathcal{D}}}^{(s)}\right)\right] \\ &- \mathrm{ddiag}\left[\left(\boldsymbol{U}^{(s)\odot}\right)^T \mathrm{vec}\left(\boldsymbol{\mathcal{M}}^{(s)}\right)\right]\end{aligned} \tag{A.3}$$

where the outer notation 'ddiag' means the tensorization from a vector to a super-diagonal tensor, which is the reverse operation of inner one. Using the property of Khatri-Rao product, we can efficiently calculate $(\boldsymbol{U}^{(s)\odot})^T\boldsymbol{U}^{(s)\odot}$ by

$$\left(\boldsymbol{U}^{(s)\odot}\right)^T \boldsymbol{U}^{(s)\odot} = \left(\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\right)^{\circledast} \tag{A.4}$$

When updating the factor matrix (without coupling information) $\boldsymbol{U}^{(n,s)}$, by keeping all other variables $\boldsymbol{U}^{(m,s)}, m \neq n$ and $\boldsymbol{\mathcal{D}}^{(s)}$ fixed, (7) is represented as follows:

$$F_u = \frac{1}{2}\left\|\boldsymbol{\mathcal{M}}^{(s)}_{(n)} - \boldsymbol{U}^{(n,s)}\boldsymbol{D}^{(s)}\left(\boldsymbol{U}^{(s)\odot_{-n}}\right)^T\right\|_F^2 \tag{A.5}$$

where $\boldsymbol{\mathcal{M}}^{(s)}_{(n)}$ denotes the mode-n matricization of $\boldsymbol{\mathcal{M}}^{(s)}$. $\boldsymbol{D}^{(s)}$ is a diagonal matrix and its diagonal elements correspond to the super-diagonal elements of core tensor $\boldsymbol{\mathcal{D}}^{(s)}$. Similarly, the block-partial gradient $\hat{\boldsymbol{G}}^{(n,s)}$ of (A.5) at $\hat{\boldsymbol{U}}^{(n,s)}$ can be calculated as:

$$\begin{aligned}\hat{\boldsymbol{G}}^{(n,s)} &= \nabla_{\hat{\boldsymbol{U}}^{(n,s)}} F_u \\ &= \hat{\boldsymbol{U}}^{(n,s)}\boldsymbol{D}^{(s)}\left(\boldsymbol{U}^{(s)\odot_{-n}}\right)^T \boldsymbol{U}^{(s)\odot_{-n}}\left(\boldsymbol{D}^{(s)}\right)^T \\ &- \boldsymbol{\mathcal{M}}^{(s)}_{(n)}\boldsymbol{U}^{(s)\odot_{-n}}\left(\boldsymbol{D}^{(s)}\right)^T \\ &= \hat{\boldsymbol{U}}^{(n,s)}\boldsymbol{D}^{(s)}\left(\boldsymbol{U}^{(s)^T}\boldsymbol{U}^{(s)}\right)^{\circledast_{-n}}\boldsymbol{D}^{(s)} \\ &- \boldsymbol{\mathcal{M}}^{(s)}_{(n)}\boldsymbol{U}^{(s)\odot_{-n}}\boldsymbol{D}^{(s)}\end{aligned} \tag{A.6}$$

However, when updating the factor matrix $\boldsymbol{U}^{(n,s)}$ which consists of two parts: $\boldsymbol{U}_C^{(n,s)}$ and $\boldsymbol{U}_I^{(n,s)}$, we first substitute $\boldsymbol{U}^{(n,s)} = \left[\boldsymbol{U}_C^{(n,s)}\ \boldsymbol{U}_I^{(n,s)}\right]$ into (A.5) and have

$$F_{ci} = \frac{1}{2}\left\|\boldsymbol{\mathcal{M}}_{(n)}^{(s)} - \left[\boldsymbol{U}_C^{(n,s)}\ \boldsymbol{U}_I^{(n,s)}\right]\boldsymbol{D}^{(s)}\left(\boldsymbol{U}^{(s)\odot_{-n}}\right)^T\right\|_F^2 \tag{A.7}$$

Let $\boldsymbol{B}^{(n,s)} = \boldsymbol{D}^{(s)}\left(\boldsymbol{U}^{(s)\odot_{-n}}\right)^T \in \mathbb{R}_+^{R^{(s)}\times\prod_{m\neq n}^N I_m}$. Let $\hat{\boldsymbol{G}}_C^{(n,s)}$ denote the block-partial gradient of (A.7) at $\hat{\boldsymbol{U}}_C^{(n,s)}$, which can be calculated as:

$$\begin{aligned}\hat{\boldsymbol{G}}_C^{(n,s)} &= \nabla_{\hat{\boldsymbol{U}}_C^{(n,s)}} F_{ci}\\ &= \hat{\boldsymbol{U}}_C^{(n,s)}\boldsymbol{B}_C^{(n,s)}\left(\boldsymbol{B}_C^{(n,s)}\right)^T\\ &- \left(\boldsymbol{\mathcal{M}}_{(n)}^{(s)} - \hat{\boldsymbol{U}}_I^{(n,s)}\boldsymbol{B}_I^{(n,s)}\right)\left(\boldsymbol{B}_C^{(n,s)}\right)^T\\ &= \left(\hat{\boldsymbol{U}}^{(n,s)}\boldsymbol{B}^{(n,s)} - \boldsymbol{\mathcal{M}}_{(n)}^{(s)}\right)\left(\boldsymbol{B}_C^{(n,s)}\right)^T\end{aligned} \tag{A.8}$$

where $\hat{\boldsymbol{U}}_C^{(n,s)}$ denotes an extrapolated point of $\boldsymbol{U}_C^{(n,s)}$. $\boldsymbol{B}^{(n,s)} = \left[\boldsymbol{B}_C^{(n,s)}; \boldsymbol{B}_I^{(n,s)}\right]$, $\boldsymbol{B}_C^{(n,s)} \in \mathbb{R}_+^{L_n\times\prod_{m\neq n}^N I_m}$ and $\boldsymbol{B}_I^{(n,s)} \in \mathbb{R}_+^{(R^{(s)}-L_n)\times\prod_{m\neq n}^N I_m}$. From (A.6) and (A.8), it can be infered that $\hat{\boldsymbol{G}}_C^{(n,s)}$ is equal to $\hat{\boldsymbol{G}}_{:,1:L_n}^{(n,s)}$. Analogously, the block-partial gradient $\hat{\boldsymbol{G}}_I^{(n,s)}$ at $\hat{\boldsymbol{U}}_I^{(n,s)}$ can be obtained as $\hat{\boldsymbol{G}}_{:,L_n:R^{(s)}}^{(n,s)}$ and $\hat{\boldsymbol{G}}^{(n,s)} = \left[\hat{\boldsymbol{G}}_C^{(n,s)}\ \hat{\boldsymbol{G}}_I^{(n,s)}\right]$.

Appendix A.2. Parameter settings

Consider updating $\boldsymbol{\mathcal{D}}^{(s)}$ and $\boldsymbol{U}^{(n,s)}$ at the kth iteration. Following [-], we set the Lipschitz constants $L_{d,k-1}^{(s)}$ and $L_{u,k-1}^{(n,s)}$ as:

$$L_{d,k-1}^{(s)} = \left\|(\boldsymbol{U}_{k-1}^{(s)\odot})^T\boldsymbol{U}_{k-1}^{(s)\odot}\right\| \tag{A.9}$$

and

$$L_{u,k-1}^{(n,s)} = \left\|\boldsymbol{D}_{k-1}^{(s)}(\boldsymbol{U}_{k-1}^{(s)\odot_{-n}})^T\boldsymbol{U}_{k-1}^{(s)\odot_{-n}}\boldsymbol{D}_{k-1}^{(s)}\right\| \tag{A.10}$$

where $\|\cdot\|$ denotes the spectral norm. Using the property of Khatri-Rao product, we can also efficiently calculate $L_{d,k-1}^{(s)}$ and $L_{u,k-1}^{(n,s)}$ through

$$\begin{aligned}\left(\boldsymbol{U}_{k-1}^{(s)\odot}\right)^T\boldsymbol{U}_{k-1}^{(s)\odot} &= \left[\left(\boldsymbol{U}_{k-1}^{(n,s)}\right)^T\boldsymbol{U}_{k-1}^{(n,s)}\right]^{\circledast}\\ \left(\boldsymbol{U}_{k-1}^{(s)\odot_{-n}}\right)^T\boldsymbol{U}_{k-1}^{(s)\odot_{-n}} &= \left[\left(\boldsymbol{U}_{k-1}^{(m,s)}\right)^T\boldsymbol{U}_{k-1}^{(m,s)}\right]^{\circledast_{-n}}\end{aligned} \tag{A.11}$$

We take the extrapolation weights as

$$w_{d,k-1}^{(s)} = \min\left(\hat{w}_{k-1},\ \delta_w \sqrt{\frac{L_{d,k-2}^{(s)}}{L_{d,k-1}^{(s)}}}\right) \tag{A.12}$$

and

$$w_{u,k-1}^{(n,s)} = \min\left(\hat{w}_{k-1},\ \delta_w \sqrt{\frac{L_{u,k-2}^{(n,s)}}{L_{u,k-1}^{(n,s)}}}\right) \tag{A.13}$$

where $\delta_w < 1$ is predefined (e.g., 0.9999, [42]), and $\hat{w}_{k-1} = \frac{t_{k-1}-1}{t_k}$ with $t_0 = 1$ and $t_k = \frac{1}{2}\left(1 + \sqrt{1 + 4t_{k-1}^2}\right)$. Moreover, we define the extrapolation at points $\boldsymbol{\mathcal{D}}_{k-1}^{(s)}$ and $\boldsymbol{U}_{k-1}^{(n,s)}$ as

$$\hat{\boldsymbol{\mathcal{D}}}_{k-1}^{(s)} = \boldsymbol{\mathcal{D}}_{k-1}^{(s)} + w_{d,k-1}^{(s)}\left(\boldsymbol{\mathcal{D}}_{k-1}^{(s)} - \boldsymbol{\mathcal{D}}_{k-2}^{(s)}\right) \tag{A.14}$$

and

$$\hat{\boldsymbol{U}}_{k-1}^{(n,s)} = \boldsymbol{U}_{k-1}^{(n,s)} + w_{u,k-1}^{(n,s)}\left(\boldsymbol{U}_{k-1}^{(n,s)} - \boldsymbol{U}_{k-2}^{(n,s)}\right) \tag{A.15}$$

References

[1] F. L. Hitchcock, The expression of a tensor or a polyadic as a sum of products, Journal of Mathematics and Physics 6 (1-4) (1927) 164–189.

[2] R. A. Harshman, et al., Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis.

[3] J. D. Carroll, J.-J. Chang, Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition, Psychometrika 35 (3) (1970) 283–319.

[4] A. Cichocki, R. Zdunek, A. H. Phan, S.-i. Amari, Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation, John Wiley & Sons, 2009.

[5] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, T. Ristaniemi, Tensor decomposition of eeg signals: a brief review, Journal of neuroscience methods 248 (2015) 59–69.

[6] F. Cong, A.-H. Phan, P. Astikainen, Q. Zhao, Q. Wu, J. K. Hietanen, T. Ristaniemi, A. Cichocki, Multi-domain feature extraction for small event-related potentials through nonnegative multi-way array decomposition from low dense array eeg, International journal of neural systems 23 (02) (2013) 1350006.

[7] D. Wang, Y. Zhu, T. Ristaniemi, F. Cong, Extracting multi-mode erp features using fifth-order nonnegative tensor decomposition, Journal of neuroscience methods 308 (2018) 240–247.

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

[9] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, H. A. Phan, Tensor decompositions for signal processing applications: From two-way to multiway component analysis, IEEE signal processing magazine 32 (2) (2015) 145–163.

[10] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, C. Faloutsos, Tensor decomposition for signal processing and machine learning, IEEE Transactions on Signal Processing 65 (13) (2017) 3551–3582.

[11] G. Zhou, Q. Zhao, Y. Zhang, T. Adalı, S. Xie, A. Cichocki, Linked component analysis from matrices to high-order tensors: Applications to biomedical data, Proceedings of the IEEE 104 (2) (2016) 310–331.

[12] X. Wang, W. Liu, P. Toiviainen, T. Ristaniemi, F. Cong, Group analysis of ongoing eeg data based on fast double-coupled nonnegative tensor decomposition, Journal of neuroscience methods 330 (2020) 108502.

[13] Y. Jonmohamadi, S. Muthukumaraswamy, J. Chen, J. Roberts, R. Crawford, A. Pandey, Extraction of common task features in eeg-fmri data using coupled tensor-tensor decomposition, bioRxiv (2019) 685941.

[14] X.-F. Gong, Q.-H. Lin, F.-Y. Cong, L. De Lathauwer, Double coupled canonical polyadic decomposition for joint blind source separation, IEEE Transactions on Signal Processing 66 (13) (2018) 3475–3490.

[15] M. Sørensen, I. Domanov, L. De Lathauwer, Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-(l_r,n,l_r,n,1) terms—part ii: Algorithms, SIAM Journal on Matrix Analysis and Applications 36 (3) (2015) 1015–1045.

[16] B. Hunyadi, P. Dupont, W. Van Paesschen, S. Van Huffel, Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 7 (1) (2017) e1197.

[17] Z. Xue, S. Yang, H. Zhang, P. Du, Coupled higher-order tensor factorization for hyperspectral and lidar data fusion and classification, Remote Sensing 11 (17) (2019) 1959.

[18] A. Cichocki, Tensor decompositions: a new concept in brain data analysis?, arXiv preprint arXiv:1305.0395.

[19] V. D. Calhoun, J. Liu, T. Adalı, A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data, Neuroimage 45 (1) (2009) S163–S172.

[20] X.-F. Gong, X.-L. Wang, Q.-H. Lin, Generalized non-orthogonal joint diagonalization with lu decomposition and successive rotations, IEEE Transactions on Signal Processing 63 (5) (2015) 1322–1334.

[21] M. Mørup, Applications of tensor (multiway array) factorizations and decompositions in data mining, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1 (1) (2011) 24–40.

[22] T. Yokota, A. Cichocki, Y. Yamashita, Linked parafac/cp tensor decomposition and its fast implementation for multi-block tensor analysis, in: In-

ternational Conference on Neural Information Processing, Springer, 2012, pp. 84–91.

[23] X. Wang, C. Zhang, T. Ristaniemi, F. Cong, Generalization of linked canonical polyadic tensor decomposition for group analysis, in: International Symposium on Neural Networks, Springer, 2019, pp. 180–189.

[24] G. Zhou, A. Cichocki, Y. Zhang, D. P. Mandic, Group component analysis for multiblock data: Common and individual feature extraction, IEEE transactions on neural networks and learning systems 27 (11) (2015) 2426–2439.

[25] E. Acar, Y. Levin-Schwartz, V. D. Calhoun, T. Adali, Tensor-based fusion of eeg and fmri to understand neurological changes in schizophrenia, in: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2017, pp. 1–4.

[26] E. Karahan, P. A. Rojas-Lopez, M. L. Bringas-Vega, P. A. Valdes-Hernandez, P. A. Valdes-Sosa, Tensor analysis and fusion of multimodal brain images, Proceedings of the IEEE 103 (9) (2015) 1531–1559.

[27] C. Chatzichristos, M. Davies, J. Escudero, E. Kofidis, S. Theodoridis, Fusion of eeg and fmri via soft coupled tensor decompositions, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 56–60.

[28] X. Wang, T. Ristaniemi, F. Cong, Fast implementation of double-coupled nonnegative canonical polyadic decomposition, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 8588–8592.

[29] R. Zdunek, K. Fonał, A. Wołczowski, Linked cp tensor decomposition algorithms for shared and individual feature extraction, Signal Processing: Image Communication 73 (2019) 37–52.

[30] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, W.-K. Ma, Hyperspectral super-resolution: A coupled tensor factorization approach, IEEE Transactions on Signal Processing 66 (24) (2018) 6503–6517.

[31] S. Li, R. Dian, L. Fang, J. M. Bioucas-Dias, Fusing hyperspectral and multispectral images via coupled sparse tensor factorization, IEEE Transactions on Image Processing 27 (8) (2018) 4118–4130.

[32] M. Sørensen, L. De Lathauwer, Coupled tensor decompositions for applications in array signal processing, in: 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2013, pp. 228–231.

[33] B. Ermiş, E. Acar, A. T. Cemgil, Link prediction in heterogeneous data via generalized coupled tensor factorization, Data Mining and Knowledge Discovery 29 (1) (2015) 203–236.

[34] E. Acar, R. Bro, A. K. Smilde, Data fusion in metabolomics using coupled matrix and tensor factorizations, Proceedings of the IEEE 103 (9) (2015) 1602–1620.

[35] G. Zhou, A. Cichocki, S. Xie, Fast nonnegative matrix/tensor factorization based on low-rank approximation, IEEE Transactions on Signal Processing 60 (6) (2012) 2928–2940.

[36] Y. Zhang, G. Zhou, Q. Zhao, A. Cichocki, X. Wang, Fast nonnegative tensor factorization based on accelerated proximal gradient and low-rank approximation, Neurocomputing 198 (2016) 148–154.

[37] N. Parikh, S. Boyd, et al., Proximal algorithms, Foundations and Trends® in Optimization 1 (3) (2014) 127–239.

[38] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate o (1/kˆ 2), in: Dokl. akad. nauk Sssr, Vol. 269, 1983, pp. 543–547.

[39] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM journal on imaging sciences 2 (1) (2009) 183–202.

[40] N. Guan, D. Tao, Z. Luo, B. Yuan, Nenmf: An optimal gradient method for nonnegative matrix factorization, IEEE Transactions on Signal Processing 60 (6) (2012) 2882–2898.

[41] Y. Xu, W. Yin, A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion, SIAM Journal on imaging sciences 6 (3) (2013) 1758–1789.

[42] Y. Xu, Alternating proximal gradient method for sparse nonnegative tucker decomposition, Mathematical Programming Computation 7 (1) (2015) 39–70.

[43] F. Cong, G. Zhou, P. Astikainen, Q. Zhao, Q. Wu, A. K. Nandi, J. K. Hietanen, T. Ristaniemi, A. Cichocki, Low-rank approximation based non-negative multi-way array decomposition on event-related potentials, International journal of neural systems 24 (08) (2014) 1440005.

[44] A. Cichocki, A.-H. Phan, Fast local algorithms for large scale nonnegative matrix and tensor factorizations, IEICE transactions on fundamentals of electronics, communications and computer sciences 92 (3) (2009) 708–721.

[45] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788.

[46] H. Lee, S. Choi, Group nonnegative matrix factorization for EEG classification, in: Artificial Intelligence and Statistics, 2009, pp. 320–327.

[47] A. S. Georghiades, P. N. Belhumeur, D. J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE transactions on pattern analysis and machine intelligence 23 (6) (2001) 643–660.

[48] D. Cai, X. He, J. Han, Spectral regression for efficient regularized subspace learning, in: 2007 IEEE 11th international conference on computer vision, IEEE, 2007, pp. 1–8.

[49] F. Cong, A. H. Phan, Q. Zhao, T. Huttunen-Scott, J. Kaartinen, T. Ristaniemi, H. Lyytinen, A. Cichocki, Benefits of multi-domain feature of mismatch negativity extracted by non-negative tensor factorization from eeg collected by low-density array, International journal of neural systems 22 (06) (2012) 1250025.