

Regularized Statistical Techniques for High Dimensional Medical Imaging Data Processing

by

Jingsai Liang

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy in Computational Sciences

Middle Tennessee State University

May 2017

Dissertation Committee:

Dr. Don Hong, Chair

Dr. Hyrum Carroll

Dr. William Robertson

Dr. Qiang Wu

Dr. John Wallin, Program Director

To my wife Xi Chen.

ACKNOWLEDGMENTS

Time goes so fast. It has been almost five years since I came to MTSU in August 2012. In the past five years, I completed numerous tasks, gained invaluable experiences, and enjoyed my most memorable life. There are so many people on my list to thank. I could not imagine my Ph.D. study and life in the past five years without their generous and great help. First of all, I would like to thank my Ph.D. supervisor, Dr. Don Hong, who brought me from China to MTSU five years ago. For research, he proposed two topics of this dissertation, organized seminars evolving my research, invited related professors giving talks, supported me to present at conferences, supervised the progress of my research, and kept encouraging my confidence. He suggested I earn a master degree in computer science, reviewed my course registrations, and concerned about my course feedback. I could not finish this dissertation without his foresight, plan, and suggestions.

I also thank my committee members Dr. Qiang Wu, Dr. Hyrum Carroll, Dr. William Robertson for their precious suggestions and directions during my dissertation writing.

I appreciate Dr. John Wallin, Dr. Donald Nelson and Dr. Don Hong writing so many recommendation letters for me during my job search. My long and stressful job searching would not be successful without their tremendous help. I also appreciate Dr. Sarah Bleiler-Baxter for teaching me the student-centered learning method and for reviewing my portfolio.

Finally, I especially appreciate my wife, Xi Chen, who came to Murfreesboro because of me. My Ph.D. life definitely would be tedious and painful if she chose to stay in China. She encouraged and believed me when I was in my lowest point of life. I also thank my parents, brother, sisters, and parents-in-law for always trusting my ability and supporting my decision.

ABSTRACT

This dissertation consists of two topics. The first topic is IMSmining: A Tool for Imaging Mass Spectrometry Data Biomarker Selection and Classification. We developed IMSmining, a free software tool combining functions of intuitive visualization of imaging mass spectrometry (IMS) data with advanced analysis algorithms in a single package which is easy to operate. The main functions of IMSmining include data visualization, biomarker selection and classification using advanced multivariate analysis methods such as elastic net, sparse PCA, and wavelets. It can be used to study the correlation and distribution of the IMS data by incorporating the spatial information in the entire image cube and helping to find the distinction of the possible features caused by the biological structure and the potential biomarkers.

The second topic is Non-Gaussian Penalized PARAFAC Analysis for Functional Magnetic Resonance Imaging (fMRI) Data. Independent Component Analysis (ICA) method has been used widely and successfully in fMRI data analysis for both single and group subjects. As an extension of the ICA, Tensorial Probabilistic ICA (TPICA) is used to decompose fMRI group data into three-mode of subject, temporal and spatial. But due to the independent constraint of the spatial components, TPICA is not very efficient in the presence of overlapping of active regions of different spatial components. Parallel Factor Analysis (PARAFAC) is another method to process three-mode data and can be solved by alternating least-squares. PARAFAC may converge into some degenerate solutions if the matrix of one mode is collinear. However, it is reasonable to find significant collinear relationships within subject mode of two similar subjects in group fMRI data. Thus both TPICA and PARAFAC have unavoidable drawbacks. In this topic, we try to alleviate both overlapping and collinear

issues by integrating the characters of PARAFAC and TPICA together, which imposes a non-Gaussian penalty term to each spatial component under the PARAFAC framework. The proposed algorithm can regulate the spatial components, as the high nongaussianity is possible to avoid the degenerate solutions aroused by collinear issue, and get rid of the independent constraint of the spatial components to bypass the overlapping issue. This proposed algorithm outperforms TPICA and PARAFAC on the simulation data. Its performance on real fMRI data is also comparable with other algorithms.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1: INTRODUCTION	1
1.1 Imaging Mass Spectrometry Data	1
1.2 Neuroimaging Data	2
CHAPTER 2: IMSMINING: A TOOL FOR IMAGING MASS SPEC- TROMETRY DATA BIOMARKER SELECTION AND CLASSIFI- CATION	4
2.1 Introduction	4
2.2 Algorithm Content	5
2.2.1 EN4IMS	5
2.2.2 WEN	6
2.2.3 Wavelet4IMS	7
2.3 Software Description	7
2.3.1 Interface	7
2.3.2 Data Visualization	8
2.3.3 Biomarker Selection	9
2.3.4 Classification	9
2.4 Pipeline	10
2.5 Discussion	11

CHAPTER 3: NON-GAUSSIAN PENALIZED PARAFAC ANALYSIS FOR FMRI DATA	12
3.1 Introduction	12
3.2 Model Review	17
3.2.1 Tensorial PICA (TPICA)	18
3.2.2 PARAFAC	20
3.3 Non-Gaussian Penalized PARAFAC	21
3.3.1 Accelerative Method	25
3.3.2 Parameters Selection	26
3.4 Experimental Design	27
3.4.1 Simulation Data	27
3.4.2 Real fMRI Group Data	30
3.5 Results	31
3.5.1 Simulation Data	31
3.5.2 Real fMRI Group Data	34
3.6 Discussion	34
CHAPTER 4: FUTURE WORK	45
REFERENCES	48

LIST OF TABLES

1	Experiment list of simulation.	30
2	Experiment results of the means and variances of ρ in three algorithms.	32
3	Characterization of control and autism populations [1].	45

LIST OF FIGURES

1	Interface of GUI	8
2	Pipeline of GUI	11
3	Simulation data of spatial components (SC) and time courses (TC).	28
4	Paradigm used for fMRI experiment. The checkerboards were showed at the high bar period and hidden at the low bar period. [2]	31
5	Error bars of the experiment results.	36
6	Experiment A in the case of low noise, high overlapping and high collinear.	37
7	Experiment B in the case of low noise, high overlapping and low collinear.	38
8	Experiment C in the case of low noise, low overlapping and high collinear.	40
9	Experiment E in the case of high noise, high overlapping and high collinear.	41
10	Experiment result on the real fMRI data.	41

CHAPTER 1

INTRODUCTION

1.1 Imaging Mass Spectrometry Data

Mass spectrometry (MS) and imaging mass spectrometry (IMS) are both important techniques in proteomics. IMS is a novel technology that is able to incorporate spatial biochemical information in full molecular range [3]. However, there are still many challenges in data processing due to high dimensionality, huge differences between the number of predictors and the sample size, and the incorporation of both spectral and spatial information. All these challenges pose great difficulties in model selection and data processing.

Several software tools are commonly used for IMS/MS data analysis. Biomap and Tissue View are mainly for data visualization. Both software tools lack advanced data analysis functionality such as multivariate analysis methods for biomarker selection and classification. MarkerView and ClinProTools are packages for MS data analysis. Technically, IMS data after using Biomap or Tissue View based on visualization can be exported and then imported to MarkerView or ClinProTools for further data analysis. However, this is not feasible for IMS data processing, especially for those in high resolution. PCA and clustering are most commonly used for IMS data analysis [4]. LDA and multivariate analysis of variance [5] and PCA combined with support vector machine (SVM) [6] were used to process IMS data. However, these methods have their limitations of handling high dimensional IMS cubes and incorporating spatial information.

It is essential to extract the complex/hidden patterns from the IMS data. Modern statistical methods should be used to complete a series of operations for biomarker selection and classification in potential application to disease and cancer diagnosis.

This part will be discussed in Chapter 2.

1.2 Neuroimaging Data

Functional Magnetic Resonance Imaging (fMRI) is a noninvasive technique for studying brain activity. The main principle of fMRI is using the difference of blood oxygenation level dependent (BOLD) to observe the local changes of deoxyhemoglobin concentration in the vascular of the brain. In addition, the area and intensity of neural activity in the brain depend on various tasks.

Usually higher neural activity needs more oxygen to support energy consumption. So in order to measure the neural activity of the brain region, we use a signal named blood oxygen level dependent (BOLD) which reflects the blood oxygenation level [7]. The mechanisms behind the BOLD signal mainly relies on the different magnetic responses between the deoxyhaemoglobin and oxyhaemoglobin. In detail, deoxyhaemoglobin has the effect of suppressing the MR (Magnetic Resonance) signal, while oxyhaemoglobin does not. Neural activity arouses the oxygen demand which speeds up the blood flow carrying lots of oxyhaemoglobin, so that less deoxyhaemoglobin gives rise to an increase of MR signal.

Since fMRI data consists of complex spatial-temporal correlation structure with a

mass of noisy and useless information, statistical analysis plays a key role to analyze and discover the active brain regions and their connections. In general, in the first step, we need to design some specific experiments to acquire the fMRI data. Then some preprocessing tasks are necessarily completed before we use the statistical tools to analyze the data. These preprocessing tasks usually include realignment, coregistration, segmentation, normalization and smoothing. In the data analysis step, powerful statistical tools like parameter estimation, hypothesis testing, and variance analysis will help us understand the fMRI data based on various statistical models.

The aims of fMRI analysis include locating regions of the brain activated by a certain task, determining distributed networks that correspond to brain function and making predictions about psychological or disease status. Popular models in processing the fMRI data include General Linear Model (GLM), Independent Component Analysis (ICA) and Parallel Factor Analysis (PARAFAC). GLM is a hypothesis-driven method without adaptivity. ICA has been challenged recently by a number of studies showing that independence is not adaptive for blind source separation in fMRI [8]. PARAFAC is a method to process three-mode data [9].

This part will be discussed in Chapter 3.

CHAPTER 2

IMSMINING: A TOOL FOR IMAGING MASS SPECTROMETRY DATA BIOMARKER SELECTION AND CLASSIFICATION

2.1 Introduction

IMSmining software package is mainly for IMS data visualization, biomarker selection, model validation, and classification. Visualization functions include the spectrum of a single pixel, the average spectrum of an area, and intensity distribution matrix at a fixed m/z value. The analysis functions include not only PCA, SVM, and LDA methods, but also the most recently developed models SPCA [10, 11], Wavelet4IMS [12], EN4IMS (Elastic Net) [13] and WEN (Weighted Elastic Net) [14] by using the spatial information. The motivation is to provide a convenient and automatic way to analyze and extract useful information from the high dimensional and complex IMS data by not only utilizing the spectrum information within individual pixels but also studying the correlation and distribution using the spatial information.

The remainder of the chapter is organized as follows: In Section 2.2, the main algorithms such as EN4IMS, WEN, Wavelet4IMS are briefly introduced. In Section 2.3, we give the details of the implementation of the software. A summary of the pipeline of this software is given in Section 2.4. Finally, remarks and a brief discussion are presented in Section 2.5.

This work has been published in [15]. This software package can be downloaded from <http://capone.mtsu.edu/dhong/IMSmining.htm>.

2.2 Algorithm Content

2.2.1 EN4IMS

Let us consider the multiple linear regression model with n observations. Suppose that $x_j = (x_{1j}, \dots, x_{nj})^T, j = 1, \dots, p$ are linear independent predictors and $y = (y_1, \dots, y_n)^T$ is the response vector. If we use $X = [x_1, \dots, x_p]$ represent the predictor matrix, the linear regression model can be expressed as

$$y = X\beta + \epsilon \quad (1)$$

where $\beta = (\beta_1, \dots, \beta_p)^T$ and the noise term $\epsilon \sim N(0, \sigma^2 I_n)$. The naive elastic net (EN) criterion is to minimize the following function [16]:

$$L(\lambda_1, \lambda_2, \beta) = \|y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2. \quad (2)$$

There are two penalty terms in equation (2). The ℓ_1 term enforces the model to generate sparse solution and the quadratic term can achieve the group effect. Zou et al. [16] mentioned that the naive EN has some weakness that will result in double amount of shrinkage. Therefore, the EN algorithm modified the naive elastic net as:

$$\hat{\beta}(\text{EN}) = (1 + \lambda_2) \hat{\beta}(\text{naive EN}). \quad (3)$$

where $\hat{\beta}(\text{naive EN})$ is the solution of naive elastic net and $\hat{\beta}(\text{EN})$ is the modified solution of elastic net. $\hat{\beta}(\text{EN})$ is given in [16] by

$$\hat{\beta} = \arg \min_{\beta} \beta^T ((X^T X + \lambda_2 I) / (1 + \lambda_2)) \beta - 2y^T X \beta + \lambda_1 \|\beta\|_1. \quad (4)$$

In the IMSmining software, we apply EN4IMS [13] based on the above EN algorithm to estimate the biomarkers. EN4IMS algorithm incorporates a spatial penalty term into the EN model. IMS information provides huge spatial information located

in each individual pixel. One important fact is that pixels in different locations of the same disease should have similar intensity values, which means the standard deviation of the intensities at the true biomarkers should be small. Conversely, the standard deviation would be very large among the complex tissue structure like bones.

So in EN4IMS, we use a parameter τ to balance two items together. One is the RSS of the linear model and another is the average of spatial standard deviations of the selected ion intensities. In detail, we use 10-fold CV to minimize the following formula:

$$(1 - \tau)\|y - \hat{y}\|_2^2 + \frac{\tau}{M} \sum_{j=1}^M \sqrt{\frac{\sum_{i=1}^N (x_{ij} - \mu_j)^2}{N - 1}}, 0 < \tau < 1. \quad (5)$$

where N is the number of all cancer pixel, x_{ij} is the intensity of a fixed j th m/z value at pixel i , μ_j is the average intensity of all cancer pixels at this fixed j th m/z value and M is the cardinality of active set determined by the EN4IMS model [13].

2.2.2 WEN

In order to consider more precise biomarker selection, Hong and Zhang [14] proposed the following model named Weighted EN (WEN):

$$\arg \min_{\beta} \frac{1}{2} \|y - \sum_{j=1}^p x_j \beta_j\|_2^2 + n \lambda_1 \sum_{j=1}^p \omega_j |\beta_j| + \frac{n}{2} \lambda_2 \sum_{j=1}^p |\omega_j \beta_j|^2. \quad (6)$$

where $\omega_j > 0, j = 1, \dots, p$ are weighted penalty coefficients. In [14], the LARS-WEN algorithm is provided to solve the above WEN model. Experiments show that WEN not only reduces the number of side features but also helps discovering new biomarkers.

2.2.3 Wavelet4IMS

To meet challenges in IMS data processing, an effective and efficient algorithm for IMS data biomarker selection and classification by using methods of multi-resolution analysis are proposed. In [12], L. Xiong and D. Hong proposed Wavelet4IMS algorithm. In addition to apply wavelet transform for IMS data de-noising, measurement for the similarity of wavelet coefficients is introduced, and the idea of wavelet pyramid method for image matching is applied for biomarker selection and the Naive Bayes classifier is used for classification in the wavelet coefficient space. Performance of the algorithm is evaluated with real data and the results of our experiments show that the multiresolution method has higher accuracy in classification.

2.3 Software Description

IMSmining allows users to visualize IMS data, to discover biomarkers, and to perform a pixel level classification for different IMS data sections. This software package is designed to give users a maximum level of convenience together with high flexibility.

2.3.1 Interface

Figure 1 shows the interface of the software based on MATLAB GUI. The first menu is data type options. We can import the data from .mat file or .txt folder or export the biomarker. The next menu contains seven algorithmic options: EN4IMS, WEN, PCA+SVM/LDA, SPCA+SVM/LDA and Wavelet4IMS. We can also use “view menu” to view the spectrum of a single pixel or the average spectrum of a selected area. Toolbar icons can be used to zoom in, zoom out, drag, or rotate the data cube. There are also four figure windows including training, spectrum, testing,

and result. One can use the mouse to drag the squares to select the cancer and noncancer area for training and testing.

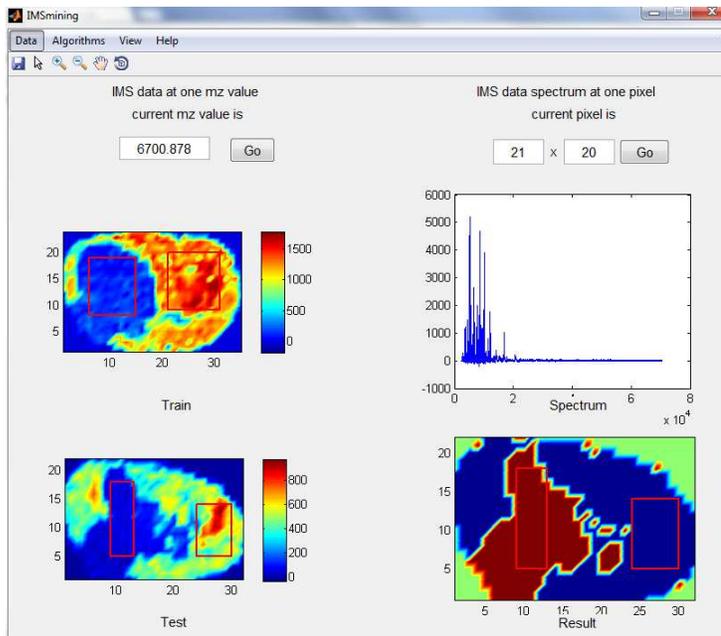


Figure 1: Interface of GUI

2.3.2 Data Visualization

IMSmining provides different methods of visualization for IMS data. Users can see intensity distribution images of different m/z values by clicking on different m/z values on the spectrum image. Users can also see spectrum of different pixels just by clicking on different pixel positions on the distribution images. Users can enlarge the spectrum to see whether the m/z value is corresponding to a true peak. The interactive responses between the intensity images (Upper Left Window) and the spectra (Upper Right Window) are extremely convenient and provide a better understanding of the spatial distribution information for a selected m/z peak. Furthermore, users

can directly select an area of pixels from the upper left window to see the mean spectrum of these selected pixels.

2.3.3 Biomarker Selection

IMSmining provides a series of algorithms, which include very recently developed EN4IMS, WEN models and Wavelet4IMS for IMS data analysis, and other methods such as PCA, SPCA and SVM popularly used in IMS community. Here, m/z values selected by the model are considered as potential biomarkers.

In EN4IMS algorithm, a spatial penalty term is incorporated into the cross validation step of the EN model [16] for IMS data processing [13]. The WEN model associates the weighted coefficients of EN model with intensity spreading information and thus provides a systematic consideration for the spatial information of the IMS data for biomarker selection and classification. Both models inherit good properties from the EN method which produces a sparse model with high prediction accuracy. By taking the spatial information into consideration, these two models help to distinguish the IMS feature peaks caused by biological structure differences from those truly associated with diseases. In Wavelet4IMS algorithm, IMSmining transforms each mass spectrometry to wavelet space and select biomarkers based on multi-resolution analysis.

2.3.4 Classification

IMSmining provides model validation and classifies testing samples. Users can select the training data region directly from the training data figure. After analyzing the training data sets to create the predictive model, validation of models can be done on

the selected cancer and noncancer square area of the testing data sets. To enhance the chance of finding the best model, the tuning parameter λ of EN4IMS and WEN algorithms can be changed accordingly by users. As a result, we can obtain the classification rates of the selected testing area. Besides implementing EN4IMS or WEN algorithm, IMSmining has one method named Wavelet4IMS which uses feature vectors selected from wavelet domain combining with a naive Bayes classifier for classification. IMSmining also can use PCA or SPCA to reduce the dimension of the data and then continue to use SVM or LDA for classification.

2.4 Pipeline

Figure 2 shows the pipeline of IMSmining. After importing the data, users can either view the image of the data or process the data based variety of algorithms. If users only want to view the image, users have two choices: point or area. Then users can import a single pixel or just simply click on the data image. Users also can drag the mouse to select an area to calculate the major statistical value of this specific area.

In another branch, users have three steps to complete the model prediction: algorithm selecting, training image selecting and testing image selecting. Users can stop the algorithm at each step and start over in another algorithm. After users select the images, they need to use the mouse to drag both of the cancer and noncancer area. Once the calculation has been done, IMSmining will show the comparative cancer and noncancer result.

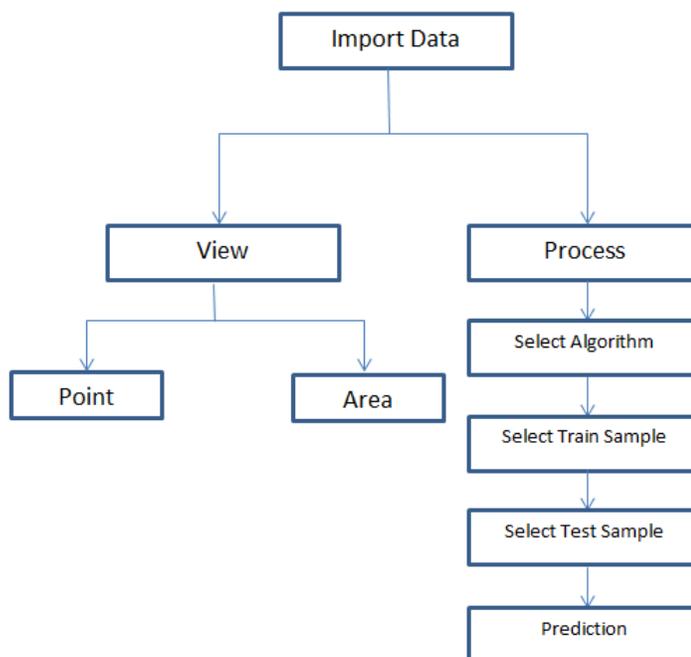


Figure 2: Pipeline of GUI

2.5 Discussion

We developed a software package called IMSmining based on algorithms of EN4IMS, WEN, SPCA and Wavelet4IMS. We have applied this software tool to real IMS data [13, 14]. Compared with other currently popular methods, the models of EN4IMS, WEN and Wavelet4IMS work more efficiently and effectively for IMS data processing in terms of confirming new biomarkers, producing a more accurate feature list including significant peaks, and providing more accurate classification results.

We would like to thank Shannon Cornett, Sara Frappier, and Richard M. Caprioli from the VUMSRC for valuable discussions and providing IMS data sets for the study. This research was partially supported by Middle Tennessee State University Faculty Research Grant #2-21519 and TDCI Grant #33501-12168.

CHAPTER 3
NON-GAUSSIAN PENALIZED PARAFAC ANALYSIS FOR FMRI
DATA

3.1 Introduction

Independent Component Analysis (ICA) is one of the most popular methods to analyze fMRI single or group data especially when the time courses are not available [17, 18, 19, 20], such as the application of autism experiments [21]. Unlike General Linear Model (GLM) [22], ICA is a more general and totally data-driven method to decompose the mixed data into mixing matrix and source signals. One fundamental prerequisite requirement of ICA is that the distributions of the independent source signals must be non-Gaussian [23]. Under this assumption, because Gaussian variable has the largest entropy among all random variables with equal variance, one popular algorithm called fastICA [23] decomposes the data by projecting the data onto the unmixing directions such that the projected data have the maximum nongaussianity. Such unmixing directions of the inverse of the mixing matrix are mutual orthogonal and can be calculated by Newton algorithm. The projected data form the independent source signals. In the application of fMRI data (Temporal \times Spatial), we usually call the source signals as spatial components.

Many variants and improvements of ICA have developed in the past decade [24]. (1) In the application of fMRI, the data usually have temporal or spatial prior information. Such information can be used as the reference or constraint to improve the accuracy and robustness of the spatial components. The temporal information can

be the time courses of the design matrix in the experiment [25, 26]. The spatial information can be the atlas-defined masks [27, 28] or the cortex based information [29]. (2) In some situations, the independent constraints are too strong to hold for some correlated spatial components. Subspace ICA [30, 31, 32] divides the spatial components into several groups and allows certain correlation of the components within the same groups. Multi dimensional ICA [33] or topographic ICA [34] models the dependence through the spatial structure of the fMRI data, such as the neighboring voxels. (3) Considering the sparsity character of the spatial components of the fMRI data, the sparsity constraint can be added to the fastICA to impose this character of the spatial components [35].

Besides the above developments of ICA to improve the accuracy and robustness of spatial components through the characters themselves, it is very natural to consider the presence of the Gaussian noise in the fMRI data. The earlier work in [36, 37] used a joint likelihood or the Gaussian moments to address this issue. Probabilistic ICA (PICA) [38] is another way to generalize the noise free ICA to incorporate the Gaussian noise. PICA assumes that the noise covariance is isotropic so that PICA could use a similar formula comparing with Probabilistic PCA (PPCA) [39] to estimate and remove the noise before decomposing the data.

fMRI data usually contains multiple subjects as a group data. The heuristic way to process the group data is either combining the results of each subject or taking the average of all subjects. Beckmann and Smith [40] developed a three-mode ICA algorithm named Tensorial Probabilistic ICA (TPICA), which is suitable to decompose the group fMRI data into three-mode (subject \times temporal \times spatial). Basically, as

an extension of the PICA, TPICA not only decomposes fMRI data into the mixing matrix and spatial components by PICA, but also continues to decompose the mixing matrix to the tensor product of time courses and subject loadings. In TPICA, the mode of spatial components has the priority to be processed, while the other two modes are estimated by rank one approximations of mixed matrix. Based on TPICA, Y. Guo continued to refine the spatial components modeling by Gaussian Mixture Model and divide the tensor product of mixing matrix into different groups. This work has been completed in [41, 42]. The solution can be approximated by EM [43], but this complex algorithm models every mode deeply and thus is time consuming.

Parallel Factor Analysis (PARAFAC) is another popular method to process three-mode data [44, 9, 45] and is becoming a new approach to process brain information and big data [46, 47]. Unlike TPICA, three modes are equally processed in PARAFAC. By the alternating least square (ALS) approach, one mode is estimated by least-squares from observed fMRI data while fixing other two modes. The three modes are iterated and alternated until the algorithm reaches convergence. PARAFAC does not require the independent constraint of the mode to achieve the solutions. One important property of PARAFAC is that the solution is unique under proper conditions. However, because PARAFAC highly relies on the effective decomposition of least-squares, PARAFAC may converge to some degenerate solutions if one mode of the data cannot meet the full rank requirement of the least-squares.

In fMRI data, it is reasonable to find significant collinear relationships of the subject loadings mode within two similar subjects. PARAFAC also may converge very slowly due to inaccurate estimation of the number of the spatial components. Based

on these reasons and some experiments, the authors of TPICA [40] concluded that TPICA is more robust and accurate than PARAFAC when it comes to estimating the underlying spatial components.

However, TPICA assumes that the spatial components of fMRI data are mutually independent. The contradiction is that it is very common to observe the overlapping of active regions of different spatial components in the spatial mode. In [48, 8], A. Stegeman, N. Helwig and S. Hong admitted that ICA is still the most effective way to deal with two-mode fMRI data, while in the three-mode group fMRI data, they argued that if fMRI data do have trilinear structure, PARAFAC can achieve unique decomposition under proper conditions, such as when the number of spatial components is correctly determined. So without the consideration of the independent constraint among spatial components, PARAFAC can perform much better than TPICA in the presence of overlapping of activate regions in the spatial mode.

So both TPICA and PARAFAC have advantages and disadvantages in processing fMRI data. Although PARAFAC performs better in overlapping issues, it may suffer from collinear issues, which are very common between similar subjects. Similarly, TPICA does not worry about collinear issues, but it is not very effective to solve overlapping issues in three-mode fMRI data. Both of them cannot process three-mode fMRI data perfectly. However, if we can combine the advantages of PARAFAC and TPICA and eliminate their disadvantages, we may figure out a way to process fMRI group data to avoid both collinear and overlapping issues. One typical method is to impose some constraints or penalty terms on three modes of PARAFAC [49, 50] to avoid the degenerate solutions. Alternatively we can combine PARAFAC and ICA

by imposing the independent constraints in PARAFAC [51]. However, in the application of fMRI, one of the main tasks is to locate the active regions of the spatial components. We may only need to impose the constraints on each spatial component of the spatial mode [52]. In [53, 54], the authors proposed a new penalty based on the Information Entropy for the spatial mode of the constrained PARAFAC analysis of resting EEG that allowed the identification in time, frequency and space of those brain networks with minimum spectral entropy.

Inspired by the method of penalty based PARAFAC, via imposing a non-Gaussian penalty term for the spatial components within the PARAFAC, we propose an algorithm that combines advantages and eliminates disadvantages of both TPICA and PARAFAC simultaneously. In the case of collinear issues, the degenerate solution of spatial components can be alleviated because the non-Gaussian penalty term can regulate each spatial component to be as non-Gaussian as possible. Meanwhile, this new algorithm can overcome the overlapping issue because it is still based on the PARAFAC, which does not need the independent constraint of the spatial components. Experiments of simulation data under different situations show that the proposed algorithm improves both accuracy and robustness compared with TPICA and PARAFAC. The results of this proposed algorithm on real fMRI data are also consistent with other methods. This work has been submitted to *Neurocomputing*.

This chapter is organized as follows: In Section 3.2, we will briefly review the related necessary algorithms. Then, we will propose our own non-Gaussian penalized algorithm in Section 3.3. We design a series of experiments in 3.4. The simulation and real data results will be presented in Section 3.5. Finally, in Section 3.6, we will

discuss this proposed algorithm and potential future improvements.

3.2 Model Review

Suppose that the observed fMRI group data Y consists of N subjects. Each subject contains V voxel samplings on T time points. Under the assumption that fMRI data has the trilinear structure, $Y(V \times T \times N)$ can be represented as a combination of three outer products:

$$Y = \sum_{k=1}^K a_k \otimes b_k \otimes c_k. \quad (7)$$

where K is the number of components and \otimes is the outer product. Vector $a_k \in R^V$ is the spatial component. Vector $b_k \in R^T$ is the time course. Vector $c_k \in R^N$ is the subject loading. For this three-way data, PARAFAC is one popular algorithm to decompose the data Y into three modes.

We can reshape the matrix Y to two dimensions via one mode, such as:

$$Y_A = (C \odot B)A^T = MA^T. \quad (8)$$

where $A_{V \times K} = [a_1 a_2 \cdots a_K]$, $B_{T \times K} = [b_1 b_2 \cdots b_K]$, $C_{N \times K} = [c_1 c_2 \cdots c_K]$. $C \odot B = ((Bdiag(c_1))^T, \cdots, (Bdiag(c_K))^T)^T$ denotes Khatri-Rao product of C and B . Using this format, we can deem A as K independent spatial components of spatial mode and M as the mixing matrix. Then this equation turns into a ICA problem which is to decompose fMRI group data into mixing matrix M and spatial mode A .

ICA cannot be generalized to group subjects analysis naturally because different subjects in the group do not share the same independent components. Temporal concatenation of each subject data is the most popular way to organize the group

data. Under this scheme, the group data can be easily processed as a single subject data and solved using same ICA method in each iteration. Then, we can use some back-reconstruction methods [20] to rebuild each subject specific modes.

3.2.1 Tensorial PICA (TPICA)

Negentropy can be used to characterize the nongaussianity of spatial components. For a random vector a , negentropy of a is defined as:

$$J(a) = H(a_{\text{gauss}}) - H(a) \quad (9)$$

where $H(a) = -\int f(a)\log f(a)da$ is the differential entropy of a . One of the most important properties of the entropy is that a Gaussian variable a_{gauss} has the largest entropy than any other variables with the same variance. So negentropy of one random variable a is always negative and an indicator the nongaussianity of a . It is not easy to calculate the value of negentropy using formula (9). In general, we can estimate the negentropy approximately using this formula [23]:

$$J(a) \propto [E\{G(a)\} - E\{G(v)\}]^2 \quad (10)$$

where v is Gaussian variable with zero mean and unit variance, G is any nonquadratic function.

FastICA [23] decomposes the data by projecting the data onto the rows of M^{-1} such that the projected data a_k have the maximum negentropy. M^{-1} is an orthogonal matrix and can be calculated by Newton algorithm.

FastICA is noise free and needs squared mixing matrix. Probabilistic model [38] PICA is one way to incorporate the gaussian noise:

$$Y_A = MA^T + E. \quad (11)$$

where E denotes concatenated isotropic Gaussian noise matrix and time series data at each voxel follows $N(0, \sigma^2 I)$.

Let $\tilde{Y} = U_K(N\Lambda_K)^{\frac{1}{2}}V_K$ be its rank- K SVD. By using the property of equal variance at both sides, we can solve the explicit solution of equation (11).

$$\tilde{M} = U_K(\Lambda_K - \sigma^2 I_K)^{\frac{1}{2}}Q^T \quad (12)$$

$$\tilde{\sigma} = \frac{1}{NT - K} \sum_{i=K+1}^{NT} \lambda_i \quad (13)$$

where λ_i denotes the diagonals of Λ , Q denotes the rotation matrix coming from an ICA algorithm.

The TPICA algorithm continues to model M by the Khatri-Rao product of subject mode C and temporal mode B . At first, TPICA uses PICA to estimate the spatial mode A and mixing matrix M . Then, TPICA rebuilds the temporal components B and subject loadings C from mixing matrix M . If we reshape the i th column M as a $N \times T$ matrix m_i and calculate its SVD decomposition:

$$m_i = U_m \Lambda_m V_m^T \quad (14)$$

$c_i \odot b_i \approx m_i$ can be approximated by the rank-one approximation of m_i . TPICA iterates these two steps by initializing the input of PICA in step one by the results of step two.

3.2.2 PARAFAC

Harshman [9] proposed the PARAFAC method for decomposing the three-way data.

The goal of PARAFAC is to minimize the following error:

$$A, B, C = \operatorname{argmin}_{A, B, C} \left\| Y - \sum_{k=1}^K a_k \otimes b_k \otimes c_k \right\|_2^2. \quad (15)$$

PARAFAC treats all three modes equally and does not incorporate any spatial or temporal information. This means that PARAFAC itself does not restrain the independence of the spatial components, which is one key constraint in TPICA.

The most attractive feature of PARAFAC is the uniqueness of its solution under proper conditions. Regardless of the scaling of values, the decomposition of A, B, C is unique if

$$r_A + r_B + r_C \geq 2K + 2 \quad (16)$$

where r is the rank- r of the matrix. The rank- r is the largest number r such that every subset of r columns of this matrix is independent.

Alternative Least Squares(ALS) is the basic method to solve this problem. ALS iterates least-square estimation for one of A, B, C while fixing the other two matrices. The iterative pseudo code is as following:

$$\begin{aligned} A &= Y_A Z (Z^T Z)^{-1} \text{ where } Z = C \odot B \\ B &= Y_B Z (Z^T Z)^{-1} \text{ where } Z = C \odot A \\ C &= Y_C Z (Z^T Z)^{-1} \text{ where } Z = B \odot A \end{aligned} \quad (17)$$

where Y_\bullet is the reshape of Y according to the modes.

PARAFAC is sensitive to the estimated number of source signals and the rank of the matrices of three modes. Some constraints can be added to PARAFAC model to avoid degenerate solutions. In some applications, imposing the related meaningful constraint can improve the accuracy and interpretation of the solutions. Without loss of generality[52], if we want to add one penalization $P(A)$ on the mode A , we can modify the first equation from formula (17) as:

$$\tilde{A} = \operatorname{argmin}(\|Y_A - ZA^T\|^2 + \lambda P(A)) \quad (18)$$

The nonnegative penalty parameter λ balances the weights of two terms. With the penalty, PARAFAC does not treat the modes equally any more. It can incorporate the related information into one specific mode to improve the quality of the corresponding components.

3.3 Non-Gaussian Penalized PARAFAC

In the formula (18), the constraint or penalty is restricted on the whole matrix A . But in the application of fMRI data, we only want to regulate each spatial component which is the column a of A . So at first, we need to decompose the product of two matrices A and Z to the summation of k products of their corresponding columns.

$$\begin{aligned} & \|Y_A - ZA^T\| \\ = & \|Y_A - \sum_{i=1, i \neq j}^K z_i a_i^T - z_j a_j^T\| \\ = & \|Y_j - z_j a_j^T\| \end{aligned} \quad (19)$$

where z_j is j th column of Z , $Y_j = Y_a - \sum_{i=1, i \neq j}^K z_i a_i^T$. The solution of column a_j by least-squares is $\hat{a}_j = \frac{Y_j^T z_j}{z_j^T z_j}$. We can iterate this procedure to estimate matrix A until it converges.

In the ICA analysis of fMRI data, the fundamental assumption is that the spatial components A follow non-Gaussian distribution. The core method of factICA algorithm is to find a direction to project fMRI data Y so that the projected spatial component a is as non-Gaussian as possible. So in a similar way, inspired by formula (18), we can impose the non-Gaussian penalty on each column a of spatial components A to increase their nongaussianities. Because nongaussianity can be approximated by negentropy, we can set the reciprocal of negentropy of a_j as the penalty term $P(a_j)$.

$$\min \| Y_j - z_j a_j^T \| + \bar{\lambda} P(a_j) \quad (20)$$

Then in the case of collinear issue, the degenerate spatial component can be regulated to approximate the correct component.

In [52], it was proved that the solution of (20) is equivalent to the solution of this following object function:

$$\min \| a_j - \hat{a}_j \| + \lambda P(a_j) \quad (21)$$

where \hat{a}_j is least-squares solution $\frac{Y_j^T z_j}{z_j^T z_j}$ and $\lambda = \frac{\bar{\lambda}}{z_j^T z_j}$.

If the parameter λ is set appropriately, this algorithm will increase the negentropy of each spatial component a_j and keep the regulated component a_j not too far away from their initial value \hat{a}_j . After standardizing the vector a_j and \hat{a}_j and substituting them into the negentropy formula, we have the non-Gaussian penalized formula:

$$\begin{aligned} \min \quad & \| \bar{a}_j - \tilde{a}_j \|^2 + \lambda \frac{1}{(E(G(\bar{a}_j)) - E(G(v)))^2} \\ \text{where} \quad & \bar{a}_j = \frac{a_j - \text{mean}(a_j)}{\text{std}(a_j)}, \tilde{a}_j = \frac{\hat{a}_j - \text{mean}(\hat{a}_j)}{\text{std}(\hat{a}_j)} \end{aligned} \quad (22)$$

where v is Gaussian signal with zero mean and unit variance, G is one nonquadratic function.

The derivative of the object function (22) for \bar{a}_j is the deepest increasing direction d .

$$d = 2(\bar{a}_j - \tilde{a}_j) - \frac{2\lambda g(\bar{a}_j)}{V(E(G(\bar{a}_j)) - E(G(v)))^3} \quad (23)$$

Here, we choose nonquadratic function $G(u) = \frac{1}{p} \log \cosh(pu)$ and its derivative $g(u) = \tanh(pu)$, where $p = 1$.

Choose the \hat{a}_j as the initial value of a_j and move the \bar{a}_j slightly along the negative direction d by step size α one time to decrease the value of object function (22). Then restandardize the new value of \bar{a}_j and move it again along with the current new direction d . Iterate this procedure until it converges. Finally, retrieve the value $a_j = \bar{a}_j * \text{std}(\hat{a}_j) + \text{mean}(\hat{a}_j)$.

This new algorithm can overcome the overlapping issue because it is still based on the PARAFAC framework, which does not need the independent constraint of the spatial components. However the difference is that under this proposed algorithm framework, three modes are not equally processed anymore. Spatial mode will be calculated column by column and each column will be regulated by non-Gaussian penalty. Thus, this new algorithm is composed of three nesting loops. The first loop is PARAFAC framework. The second loop is least-squares solution of each component a_j of matrix A . The third loop is to optimize the current component a_j using the non-Gaussian penalty. So this algorithm needs to calculate three relative errors η_1, η_2, η_3 inside each loop to compare with the predefined terminate parameter $\epsilon_d, \epsilon_A, \epsilon_Y$. If all relative errors $\eta_i, i = 1, 2, 3$ are less than predefined parameter $\epsilon_i, i = 1, 2, 3$, this algorithm converges to a stable decomposition of fMRI data Y .

Initialization:

1. Reshape Y to Y_a, Y_b, Y_c and initialize A, B, C .

Iterative Procedure:

1. $B = Y_b Z (Z^T Z)^{-1}$ where $Z = C \odot A$
2. $C = Y_c Z (Z^T Z)^{-1}$ where $Z = B \odot A$
3. First Sub Iterative Procedure. Set column number $j = 1$. (Calculate columns a_j of A one by one.):
 - 3.1 $\hat{a}_j = Y_j z_j (z_j^T z_j)^{-1}$ where z_j is j th column of $Z = C \odot B$
 - 3.2 Second Sub Iterative Procedure. (Non-Gaussian Penalty of a_j):
 - 3.2.1 $\bar{a}_j = \hat{a}_j + \alpha * \frac{d}{\text{norm}(d)}$.
 - 3.2.2 Calculate $\eta_1 = \frac{\text{norm}(\text{previous } d) - \text{norm}(d)}{\text{norm}(d)}$. If $\eta_1 > \epsilon_d$, return to step 3.2.1. Otherwise continue to step 3.3.
 - 3.3 If $j = K$, continue to step 3.4. Otherwise set $j = j + 1$ and return to step 3.1.
 - 3.4 Calculate $\eta_1 = \frac{\text{norm}(\text{previous } A) - \text{norm}(A)}{\text{norm}(A)}$. If $\eta_2 > \epsilon_A$, set $j = 1$ and return to step 3.1. Otherwise continue to step 4.
4. $err = Y_a - Z A^T$, Calculate $\eta_1 = \frac{\text{norm}(\text{previous } err) - \text{norm}(err)}{\text{norm}(err)}$. If $\eta_3 > \epsilon_Y$, return to step 1. Otherwise, stop this algorithm.

The above pseudocode is the outline of the proposed algorithm.

3.3.1 Accelerative Method

Because this proposed algorithm contains three nesting iterations and the Khatri-Rao product is a time consuming calculation if the dimensions of matrices are very high, it is slower than TPICA, especially when applying it to real fMRI data. We can speed up this algorithm in two ways. (1) Use final result of TPICA as the initial input of A, B, C . The experiments in the next section shows that this initial setting can dramatically reduce the time consuming of this proposed algorithm. Our proposed algorithm can improve the results of TPICA remarkably. No matter how small the stopping criterion of TPICA is, this improvement cannot be achieved by TPICA alone. (2) For multiple subjects fMRI data, during the preprocessing step, we can use subject-specific PCA to reduce the data of each subject individually. Suppose $P_i(M * T, K < M < T)$ is the PCA reducing matrix for i th subject data $Y_i(T * V)$, $i = 1..N$. We have

$$\begin{aligned} Y_i &= (C(i, :) \odot B) * A^T \\ P_i Y_i &= (C(i, :) \odot P_i B) * A^T \\ Y_i^* &= (C(i, :) \odot B_i^*) * A^T \\ Y_i^* &\approx (C(i, :) \odot B^*) * A^T \end{aligned}$$

where $C(i, :)$ is the i th row of C which is also the i th subject loading, and B^* is the mean of B_i^* , $i = 1..N$. After subject-specific PCA preprocessing, the reducing fMRI data Y^* is $[Y_1^{*T} \dots Y_N^{*T}]^T$. So we have

$$Y^* \approx (C \odot B^*) * A^T \quad (24)$$

Based on (24), we can still assume that the reduced fMRI data Y^* has trilinear structure. After the calculation of the proposed algorithm, we can use $P_i, i = 1..N$ to finish the back-reconstruction step.

3.3.2 Parameters Selection

This new algorithm has five parameters to be considered: penalty parameter λ , step size α , three terminate parameters $\epsilon_d, \epsilon_A, \epsilon_Y$. Appropriate values of these parameters are needed to lead to successful decomposition. Inappropriate values of these parameters only result in meaningless and inaccurate decomposition. Here are some suggestions and experiences to choose the efficient values for these parameters.

The most important parameter among them is the penalty parameter λ . The value of λ is the key to balance the non-Gaussian penalized weight and the similarity between initial component \hat{a}_j and improved component a_j . If λ is too big, the nongaussianity of component a_j increases too fast regardless of the similarity, so that a_j diverges to meaningless value and it is very hard to converge. On contrast, if λ is too small, a_j is influenced too slightly by the penalized term. So a_j converges to the nearby of \hat{a}_j very soon. The best choice of λ should be a moderate value and is mainly estimated associated with the value of \hat{a}_j and its negentropy. We can check the chosen λ by if the algorithm increases a bit of negentropy of a_j each time and converges before the predefined maximal loop times. If the value is failed to pass this test, we may decrease the λ until it meets the above rules.

We may also use the step size value α to adjust the changing of a_j in each step. But because this object function is not unimodal, in order to avoid the erratic changing of a_j , we suggest choosing a small value for α .

Three termination parameters ϵ can be set depending on the balance of desired accuracy and running time. Small ϵ can improve the accuracy in expense of the time

consuming. But ϵ_A can be set a little larger than other two ϵ , because all columns of A move in different directions so that A could change a lot in the third nesting loop.

3.4 Experimental Design

We verify our algorithm on both simulation data and real fMRI group data. For simulation data, we compare the results of three algorithms with the ground truth design. For the real fMRI group data, we compare the result of our algorithm with the that in [2].

3.4.1 Simulation Data

We generate four two-dimensional matrices of size 46×56 to represent slices of brain voxels. Each matrix simulates one spatial component (SC) and contains an active region with size 11×16 . In Figure 3, (a)-(d) shows these four simulated spatial components and their active regions. The voxel value of the background is 0. The voxel value at the active region is randomly sampled from uniform distribution $[0,1]$. In order to show the different effects of low and high overlapping SCs, we design two copies of the third SC. The one in (e) is the low overlapping case of SC3 and SC2, while the one in (f) is the high overlapping case of SC3 and SC2. Two overlapping results are demonstrated visually in (e) and (f). If we define the ratio between the shared region and the active region as the overlapping rate, the rate of the low one in (e) is 20% and the rate of the high one in (f) is 70%. SC1, SC2 and one SC3 together form the matrix A with three spatial components.

We simulate time courses (TC) using the convolution of the stimulus functions

with the hemodynamic response function(HRF) which is generated by SPM function [55] `spm_hrf()`. We use block design pattern and a single peaked function as the stimulus functions for TC1 and TC2 respectively. TC3 is a noised version of TC1. Each TC contains 150 time points and is shown in (g)-(i) of Figure 3 respectively. These three TCs together form the matrix B with three times courses.

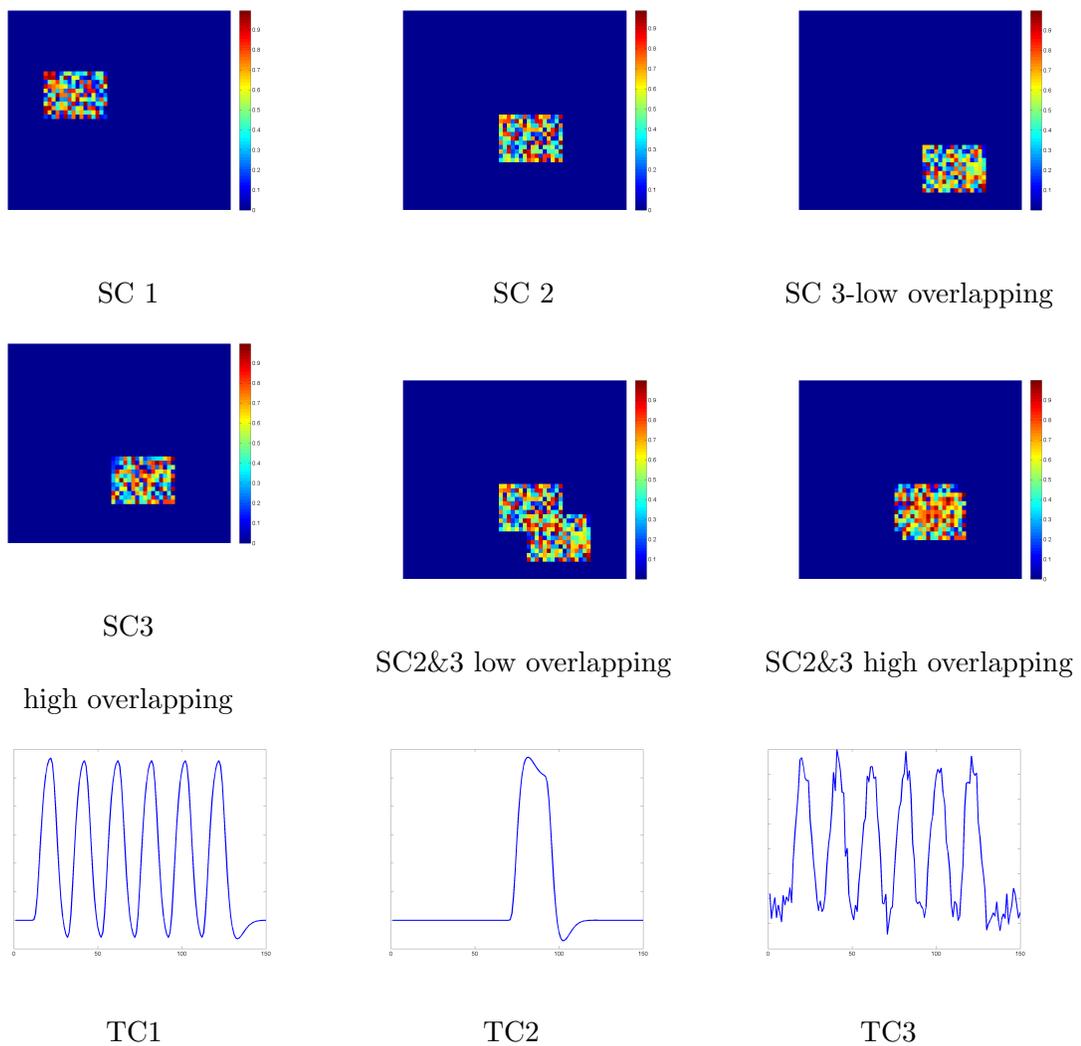


Figure 3: Simulation data of spatial components (SC) and time courses (TC).

In order to verify the advantage of our proposed algorithm in processing collinear issue of the subject loading in group ICA, we provide two subject loading matrices C in equation (25). C_1 is a full rank squared matrix which represents low collinear case, while the rank of C_2 is only 2 which causes the collinear issue.

$$C_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 3 & 1 & 3 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 2 & 2 & 3 \end{pmatrix} \quad (25)$$

We also consider the noise effect to evaluate the robust of three algorithms. We define the SNR of the data Y as

$$\text{SNR} = \frac{(C \odot B)A^T}{E} \quad (26)$$

where E is the Gaussian noise adding on the fMRI data. The mean of E is 0. The standard deviation σ of E has two values: 0.2 and 0.8. Based on the formula (26), the corresponding values of SNR are about 1.5 and 0.35. Note that a low SNR value means high noise level.

In Table 1, we design eight different tests against three indicators: noise level, overlapping level and collinear level. Each indicator has two contrasting values: low and high. Noise level is evaluated by SNR of the data Y , overlapping level is evaluated by overlapping rate of the shared region, and collinear level is evaluated by the rank of subject loading matrix C . In the next section, we will show the comparison method of these three algorithms and the experimental results.

Table 1: Experiment list of simulation.

	Noise Level (SNR)	Overlapping Level (Overlapping Rate)	Collinear Level (Rank of C)
Experiment A	Low (1.50)	High (70%)	High (2)
Experiment B	Low (1.50)	High (70%)	Low (3)
Experiment C	Low (1.50)	Low (20%)	High (2)
Experiment D	Low (1.50)	Low (20%)	Low (3)
Experiment E	High (0.35)	High (70%)	High (2)
Experiment F	High (0.35)	High (70%)	Low (3)
Experiment G	High (0.35)	Low (20%)	High (2)
Experiment H	High (0.35)	Low (20%)	Low (3)

3.4.2 Real fMRI Group Data

We use the data from [2] completed at Johns Hopkins University. Data from nine normal subjects were acquired on a Philips 1.5T Scanner. Functional scans were acquired with an echo planar sequence (64×64 , flip angle = 90, TR = 1 sec, TE = 39 msec) over a 6-min period for a total of 360 time points. A visual paradigm was designed in which an 8Hz reversing black and white checkerboard was presented intermittently in the left and right visual fields. The checkerboards were shown for 30 seconds and were resumed after 60 seconds. Subjects were focusing on a central cross on the checkerboard during the entire 6 minutes. The paradigm is depicted in Figure 4. In this experiment, we only have partial data of three subjects instead of nine subjects used in [2].

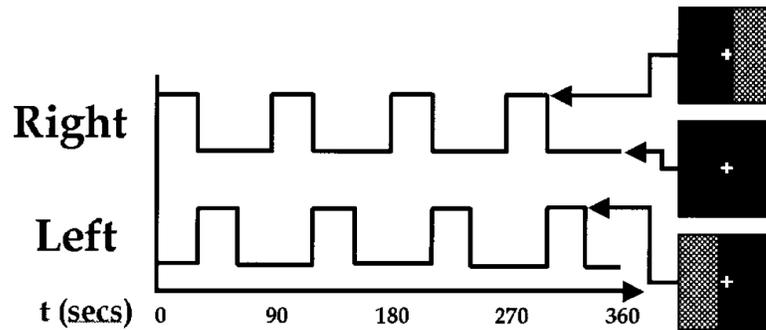


Figure 4: Paradigm used for fMRI experiment. The checkerboards were showed at the high bar period and hidden at the low bar period. [2]

3.5 Results

3.5.1 Simulation Data

In experiments A-H, we run each algorithm ten times and compare results of SC with the ground truth SC in Figure 3 using congruence coefficient ρ , where

$$\rho = \frac{a^T b}{\sqrt{(a^T a)(b^T b)}}, \text{ given column vectors } a, b. \quad (27)$$

Value ρ can be used to measure the similarity between the restored component and the truth component. Table 2 shows the means and variances of ρ of each algorithm in each experiment. Figure 5 draws the content of Table 2 in the format of error bars. Then we can compare these three algorithms intuitively.

In Figure 5, the results of three algorithm in each experiment are showed in one separate subfigure. The horizontal axis lists three algorithms in the order of TPICA, PARAFAC and proposed algorithm from left to right with the color of green, blue and red respectively. The vertical axis represents the value of ρ . The black line seg-

ments depicts three mean values. Three vertical error bars of TPICA, PARAFAC and proposed algorithm represent the intervals of one standard deviation away from the means of ρ .

Table 2: Experiment results of the means and variances of ρ in three algorithms.

	TPICA	PARAFAC	New Algorithm
	Mean(Variance) of Congruence Coefficient		
Experiment A	0.9527(0.0000)	0.9124(0.0041)	0.9777(0.0002)
Experiment B	0.9567(0.0000)	0.9875(0.0000)	0.9896(0.0000)
Experiment C	0.9815(0.0000)	0.9227(0.0033)	0.9861(0.0000)
Experiment D	0.9861(0.0000)	0.9873(0.0000)	0.9897(0.0000)
Experiment E	0.7407(0.0008)	0.6803(0.0056)	0.8680(0.0001)
Experiment F	0.8043(0.0002)	0.8343(0.0002)	0.8958(0.0000)
Experiment G	0.7737(0.0002)	0.7237(0.0036)	0.8722(0.0000)
Experiment H	0.8418(0.0000)	0.8350(0.0001)	0.8940(0.0000)

From Figure 5, it is very clear to see that all the assumptions in the previous sections are verified in these eight experiments. (1) The proposed algorithm is the best one compared with TPICA and PARAFAC in all eight experiments. Not only the mean value of proposed algorithm is the highest, but the standard deviation of the proposed algorithm is also the smallest. A higher mean value means a better decomposition, and a lower standard deviation indicates a more stabilized algorithm. Thus this proposed algorithm overcomes both the collinear and overlapping issue in this simulation. (2) High noise level can lower down the performances of all algorithms. Experiments (a)-(d) with high level of noise are generally better than experiments (e)-(h) with low level of noise. If the noise level is set to be even higher, it would be very hard for any algorithm to decompose the data into meaningful components. (3) High collinear level can dramatically reduce the performance of PARAFAC. Experi-

ments (c) and (g) are typical cases with high collinear level and low overlapping level. We can see that, in these two experiments, PARAFAC is the worst one and is not very stable. Adding a non-Gaussian penalty term on the spatial mode can conquer this intrinsic drawback of PARAFAC. (4) High overlapping level can decrease the performance of TPICA although it is still a quite robust algorithm. Experiments (b) and (f) are testing on high overlapping level and low collinear level. TPICA is clearly the worst one, especially in experiment (d). The independent requirements of ICA hinder any modifications based on itself to deal with the overlapping issue. Thus, we need to modify the algorithm based on PARAFAC in order to avoid this issue.

Figures 6-9 show the decomposition results of the first run of experiments A, B, C and E. These figures again verify the above conclusions visually. The time courses in Figures 6-9 are standardized by subtracting their means and divided by their standard deviations. Figure 6 shows the results in the case of low noise, high overlapping and high collinear. We can see that the active regions of SC2 and SC3 of TPICA can not detach from each other very well. One active region of SC includes a light shaded region from another SC. Figure 7 shows the results in the case of low noise, high overlapping and low collinear. We can see that TPICA performs worst under this situation. The results of time course in TPICA are not very consistent with the truth values and the spatial components in TPICA still have obvious shaded regions around the truth components. Figure 8 shows the results in the case of low noise, low overlapping and high collinear. This time both TPICA and the proposed algorithm outperform PARAFAC. PARAFAC leads to the degenerate solutions. The first spatial component in the results of PARAFAC is totally meaningless. Figure 9 shows the results in the case of high noise, high overlapping and high collinear. We choose

this experiment results to demonstrate the performance of these three algorithms under serious conditions of noise, overlapping and collinear. The recovered spatial components and time courses of our proposed algorithm are also the most consistent with the truth values.

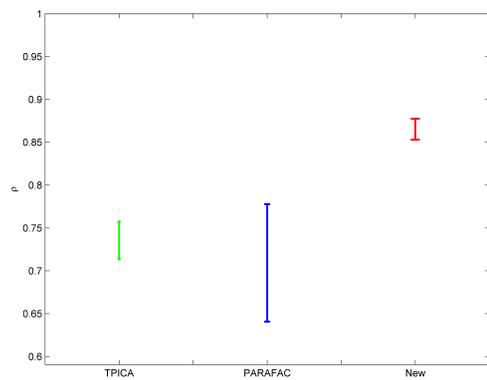
3.5.2 Real fMRI Group Data

We embed our algorithm into the GIFT package [56]. The final results of our algorithm are summarized in Figure 10, which is generated by GIFT. The threshold value is 2.0 and the slice range is -52:4:8. The activated spatial map shows right visual cortex (blue), left visual cortex (red); a transiently task-related component (TTR, green) in bilateral occipital/parietal cortex. These results are consistent with the results in [2], while here we only use three subjects instead of nine.

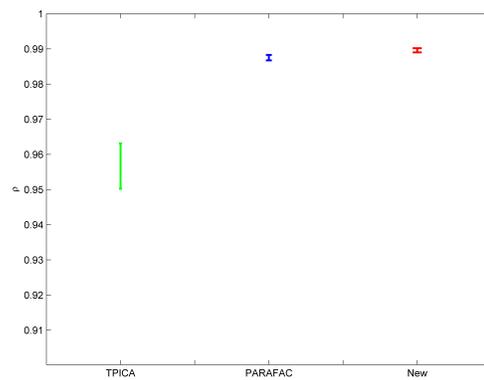
The head part of Figure 10 shows three time courses of the below components. The blue one is the time course of right visual cortex and the red one is the time course of left visual cortex. The peaked and flat periods of two time courses are equivalent to the paradigm in Figure 4 used to control the showing and disappear of the checkerboards in the left and right visual fields.

3.6 Discussion

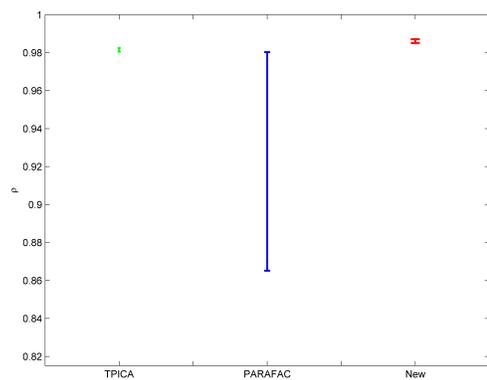
We successfully alleviate both overlapping and collinear issues aroused by ICA and PARAFAC by adding a non-Gaussian penalty term to the spatial mode calculation of the PARAFAC. The advantages of this proposed algorithm have two aspects: get rid of independent constraint to alleviate the overlapping issue and regularize the



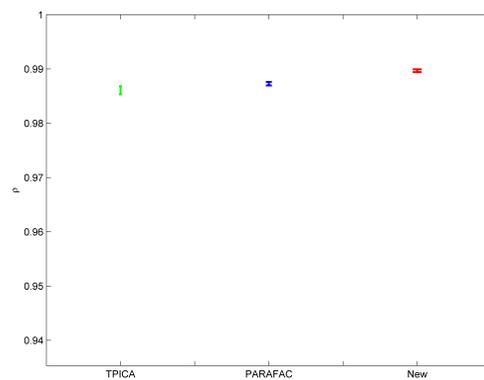
Experiment A



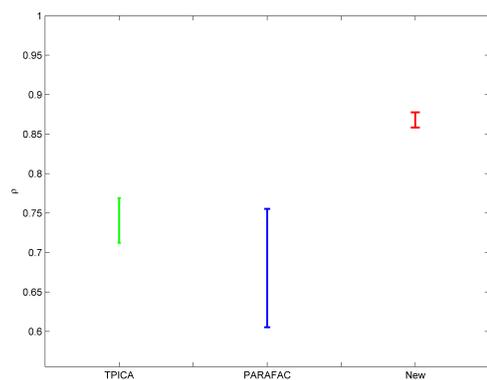
Experiment B



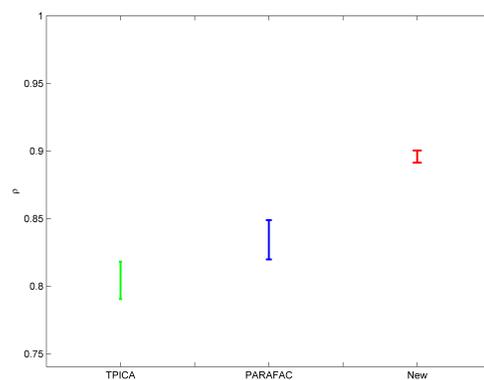
Experiment C



Experiment D



Experiment E



Experiment F

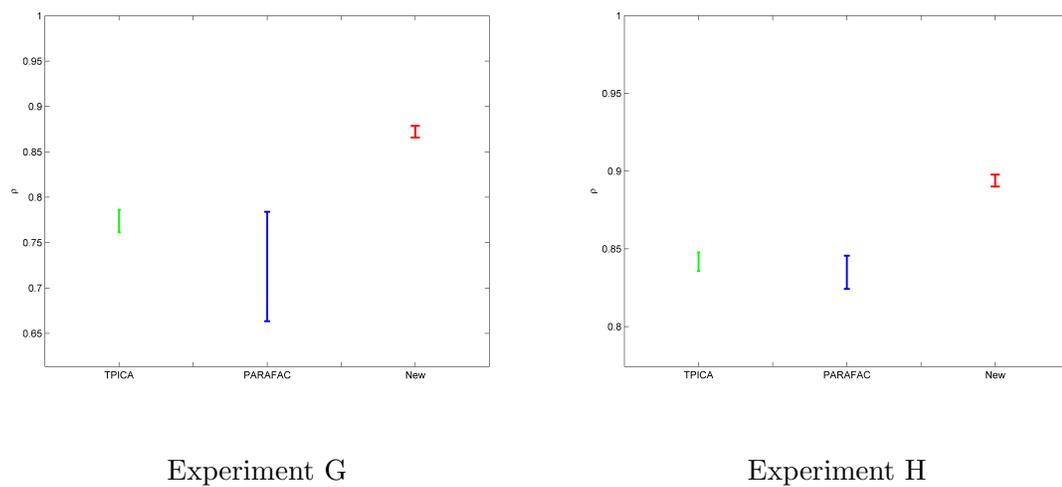
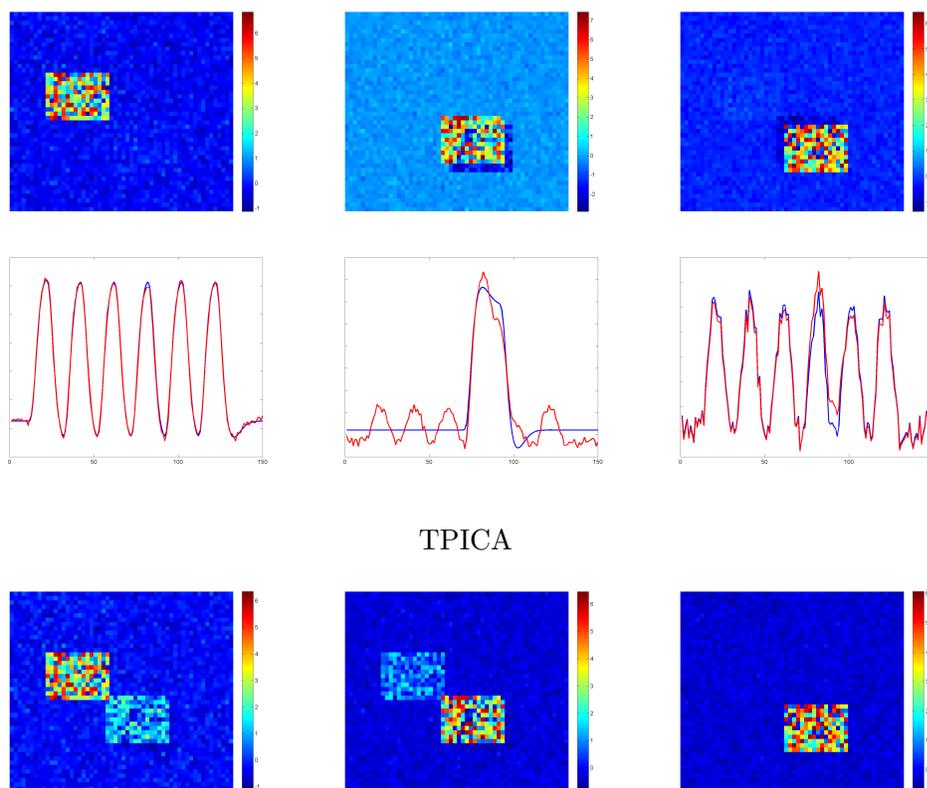


Figure 5: Error bars of the experiment results.



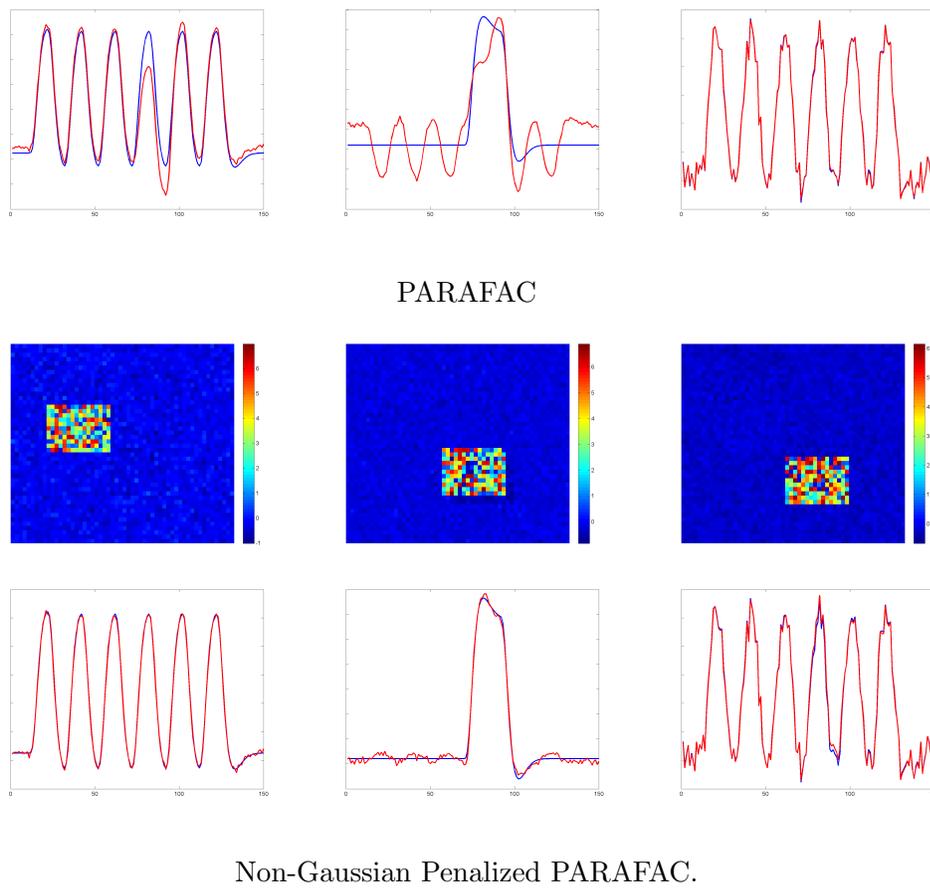
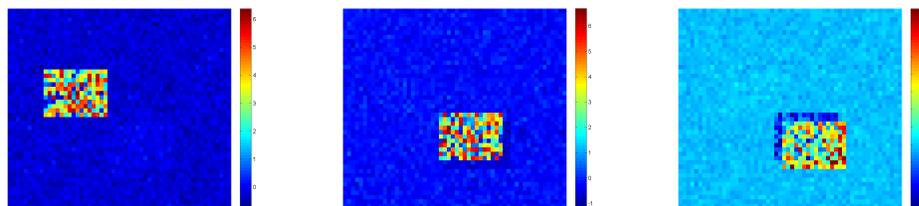
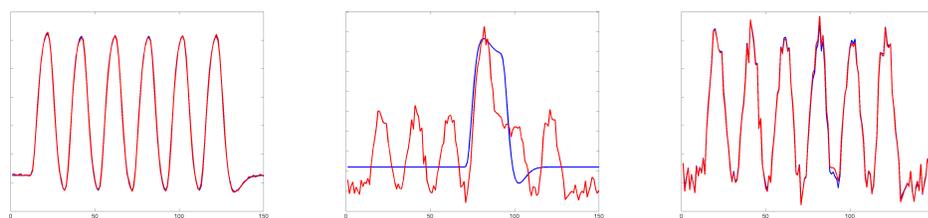
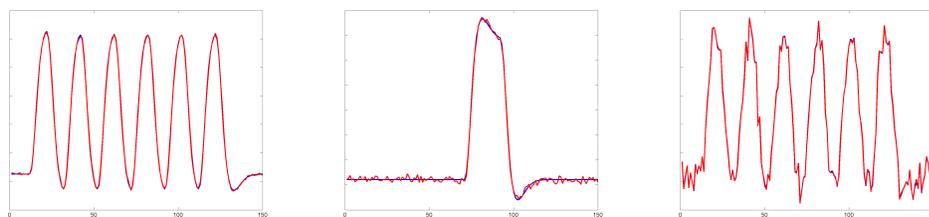
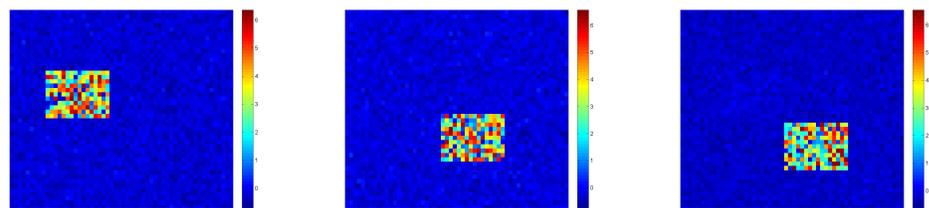


Figure 6: Experiment A in the case of low noise, high overlapping and high collinear.

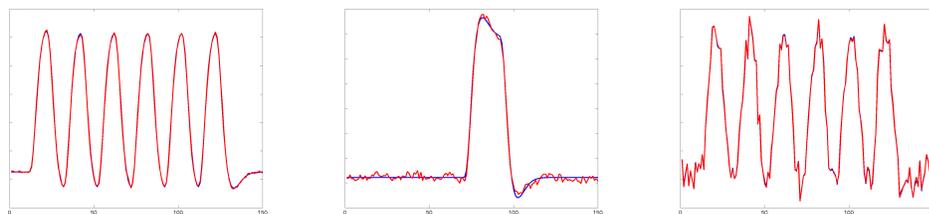
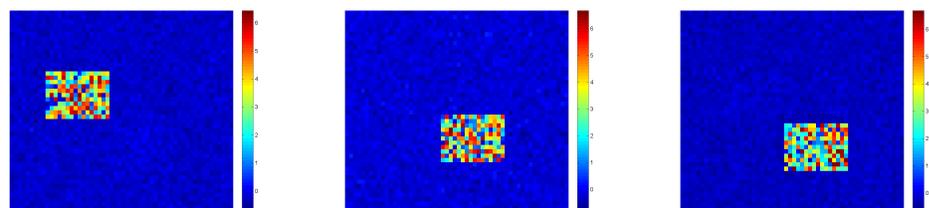




TPICA

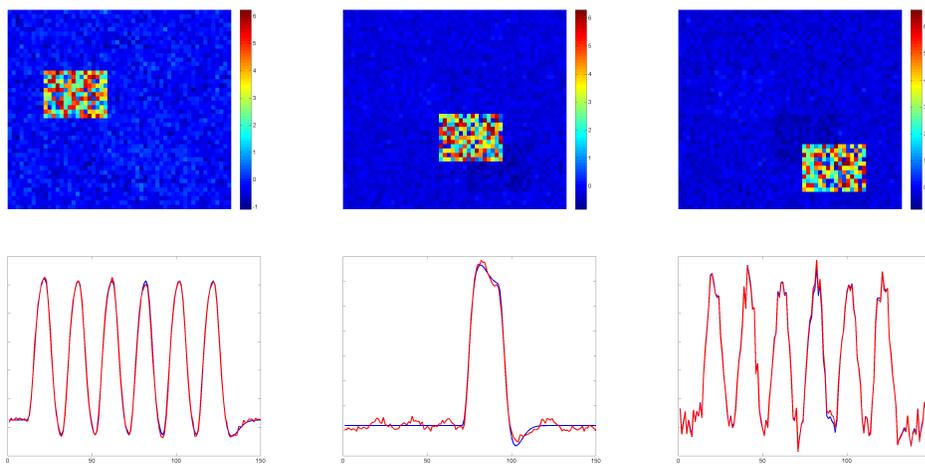


PARAFAC

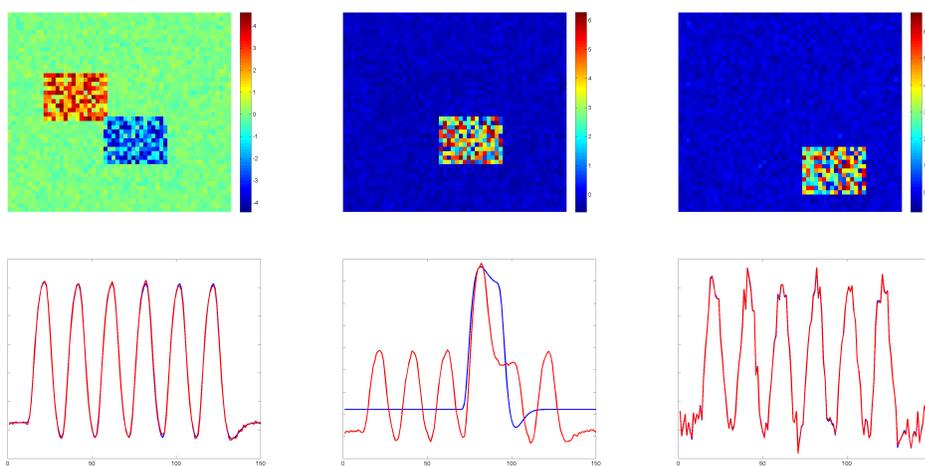


Non-Gaussian Penalized PARAFAC.

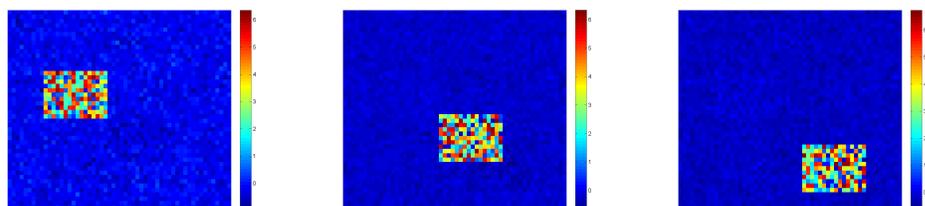
Figure 7: Experiment B in the case of low noise, high overlapping and low collinear.

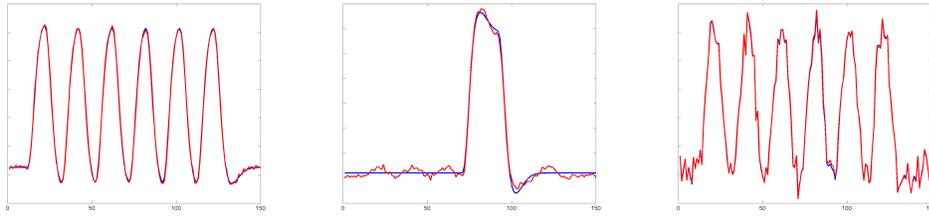


TPICA



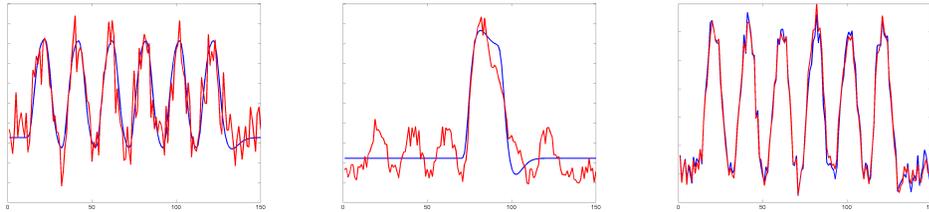
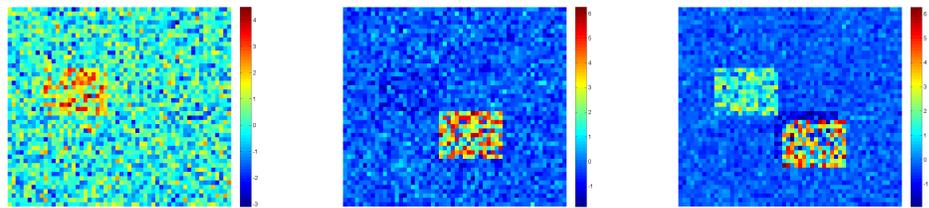
PARAFAC



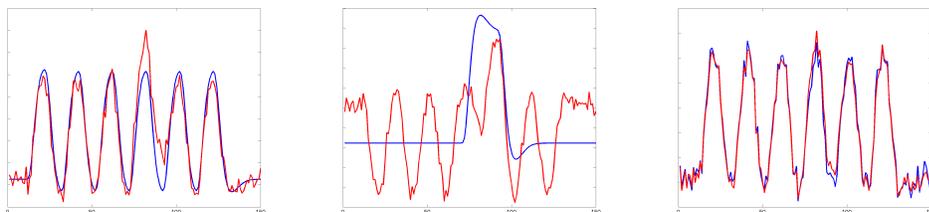
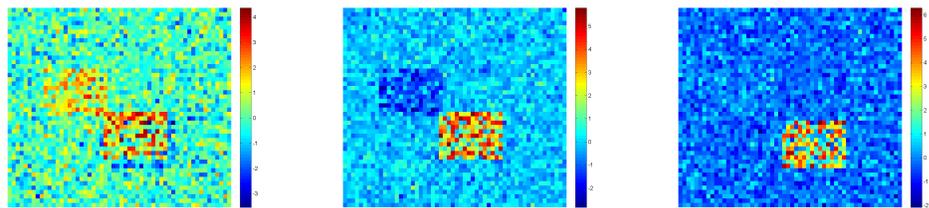


Non-Gaussian Penalized PARAFAC.

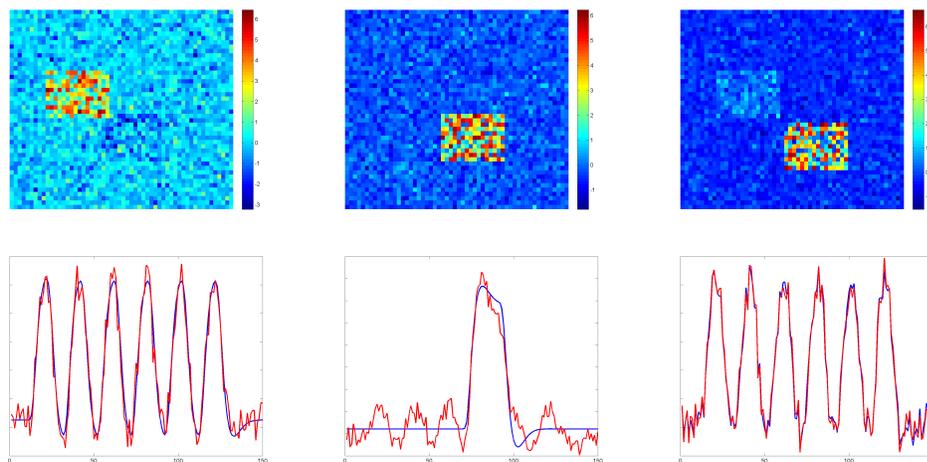
Figure 8: Experiment C in the case of low noise, low overlapping and high collinear.



TPICA



PARAFAC



Non-Gaussian Penalized PARAFAC.

Figure 9: Experiment E in the case of high noise, high overlapping and high collinear.

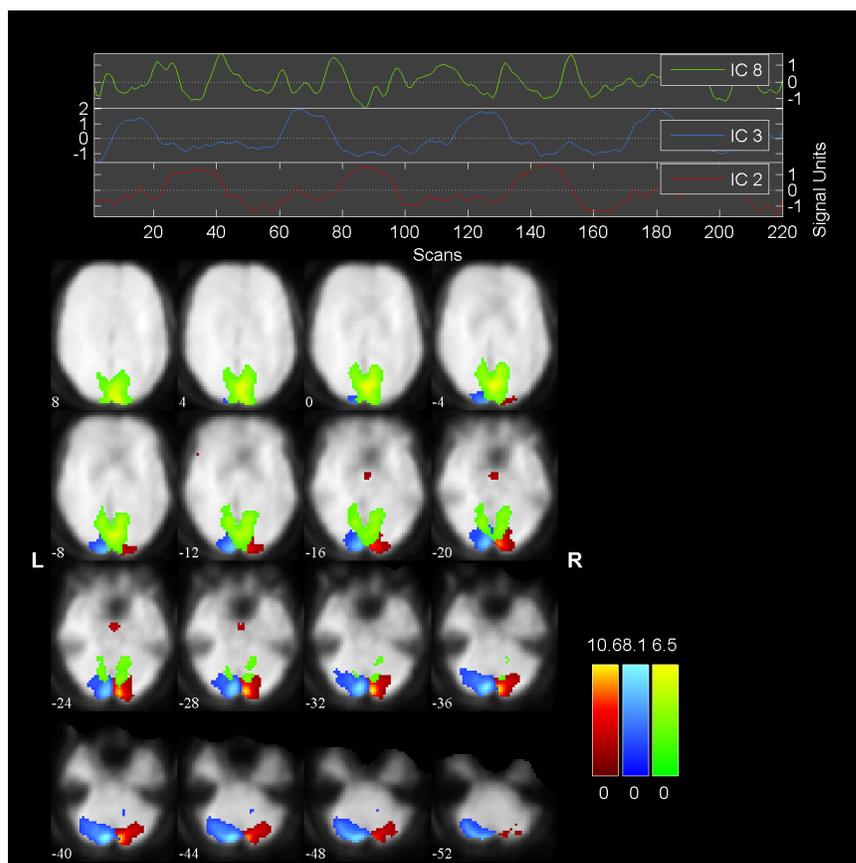


Figure 10: Experiment result on the real fMRI data.

alternative linear regressions to avoid the degenerate solutions. The core method is to add the non-Gaussian term from fastICA to one step of ALS from PARAFAC. Actually, we can extend this kind of idea which combines the ALS and penalty term to solve other similar problems, unless the problems can be modeled by trilinear structures.

This algorithm also can be viewed as the combination of the characters of ICA and PARAFAC. The strategy of the fastICA is to project fMRI data onto several orthogonal directions to maximize the nongaussianity of the projected data, which are the spatial components. So the assumption of nongaussianity is the key character of the spatial components to identify and extract them. The potential degenerate solutions of PARAFAC can be regulated towards the directions (23) of high nongaussianity to maximize the proposed object function (22). By this non-Gaussian penalty term, the degenerate solutions could be improved to good solutions. On the other hand, in the overlapping situation, mutual orthogonal directions calculated by ICA cannot distinguish the correlated spatial components very well. This proposed algorithm is still based on the PARAFAC framework, which does not need the independent constraint of the spatial components. Thus, it can separate the correlated components better than the ICA method.

This proposed algorithm uses the result of fastICA as the initial input. In this way, the algorithm can converge faster and run more robustly than random initial values. Additionally, this better result obtained by this proposed algorithm cannot be achieved by either TPICA or PARAFAC no matter how small the tolerance value of the convergence is, how large the maximum iteration time is, and how exact the

initial value is. In other words, the improvement of our proposed algorithm is because of the intrinsic modification of the penalty term and cannot be achieved by the parameter settings of the other two algorithms. PARAFAC is not robust even when we use the results of TPICA as the initial value.

We want to mention the following three points for future discussion and modification. Firstly, λ is the key parameter to achieve the best performance of this proposed algorithm. An incorrect λ value can lead this algorithm into a meaningless decomposition of the fMRI data. The choice of proper λ should consider the balance of the similarity and nongaussianity. Sometimes we can run this algorithm using the recommended default value at first and adjust the lambda value to rebalance the weights of two parts of the object function. We will think about a better or even automatic way to choose the λ value later.

Secondly, this proposed algorithm is more time consuming than other algorithms due to the calculations of the penalty term and the tensor product. However, this disadvantage could be ignored if the fMRI data does not need real-time computing. Time spent on a better decomposition is deserved because it can save even more time in the next step for screening and explaining the components. We also can use the modified PARAFAC for large-scale problems to overcome this issue [57, 58]. In detail, the new approach includes computing Hadamard products instead of Khatri-Rao products and employing relatively small matrices.

Thirdly, some may argue the assumption that fMRI is strictly trilinear data may not be realistic in applications. However, the running results of our algorithm are

consistent with the results in the original paper. Additionally, this assumption has already been discussed in other papers [40, 8]. We also can apply this non-Gaussian penalty to other applications as long as the data is mixed from non-Gaussian sources and has a trilinear structure. In the future, we may try to apply this non-Gaussian penalty strategy to two-mode data to decompose the correlated source signals instead of independent source signals.

CHAPTER 4

FUTURE WORK

In Chapter 3, both ICA and PARFAC only consider three modes, subject \times temporal \times spatial, directly and intuitively based on the structure of the fMRI data. However, in a typical fMRI experiment, the demographic and symptoms assessment information including the age and gender of the subjects were also collected. For example, in [1], forty high-functioning adolescent and young adult males with autism were compared with 40 male typically developing control volunteers. The range of these 40 patients with autism is from 12 to 42. In many medical applications such as autism and Alzheimer’s disease, age is a vital parameter to evaluate and estimate the severity level of the disease of one subject. In other words, the same disease may show and develop dramatic difference of the fMRI data along the increasing of the age of one subject. However, in [1], fMRI data of all the subjects with autism were analyzed without considering the difference of age. Thus if we could incorporate the age as one parameter to process the fMRI data of different subjects, the results would reflect more accurate details of the disease.

Table 3: Characterization of control and autism populations [1].

	Autism	Control
Characterization	Mean(SD)(range)	Mean(SD)(range)
Age	22.7 (7.4) (12 to 42)	21.6 (7.4) (8 to 39)
VIQ	107.9 (18.9) (63 to 139)	113.5 (12.7) (90 to 140)
PIQ	106.2 (13.6) (81 to 133)	111.8 (12.1) (88 to 135)
...

Some other biological prior information already have been incorporated into fMRI

data analysis, such as the atlas-defined masks [27] and the cortex based information [29]. So similarly, we would like to analyze fMRI data joint with the subject-specific information such as ages.

In [59], Liu, Li, and Wu proposed a new feature screening and variable selection algorithm for varying coefficients. This model is a generalization of the classical linear regression model with different regularization penalty terms. Basically, the coefficients of this model are not a series of numbers, but the functions of a univariate variable. Let y be the response and $x = (x_1, \dots, x_p)^T \in R^p$ be the p -dimensional predictor. We have the following varying coefficient model:

$$y = \beta_0(u) + x^T \beta(u) + \epsilon \quad (28)$$

where ϵ is the error term with zero mean, $\beta_0(u)$ is the intercept function, and $\beta(u) = (\beta_1(u), \dots, \beta_p(u))^T$.

In this model, if we let the variable u be the age of the subjects, y be the fMRI data, and x be the design matrix, we could use two-stage approach [59] along with various penalty functions, such as LASSO and EN, to analyze the fMRI data with age as the parameter.

The remaining issue is that typical medical fMRI experiments are not block design, but are event-related design. Hence, the design matrix x does not exist in the autism experiment [21]. The heuristic method is to combine the ICA and model (3.6). In the first step, we can use ICA to decompose the fMRI data into the product of temporal components x and spatial components β . In the second step, we can use x to approximate the design matrix. Hence, we can use the two-stage algorithm in

[59] to incorporate the age u of all subjects. We can iterate these two steps until the algorithm meets certain criterions. In the future, we will test this model on both simulation and real fMRI data.

REFERENCES

- [1] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher, A. L. Alexander, E. D. Bigler, N. Lange, J. E. Lainhart, Functional connectivity magnetic resonance imaging classification of autism, *Brain* 134 (12) (2011) 3742–3754.
- [2] V. Calhoun, T. Adali, G. Pearlson, J. Pekar, A method for making group inferences from functional MRI data using independent component analysis, *Human Brain Mapping* 14 (3) (2001) 140–151.
- [3] D. Trede, J. H. Kobarg, J. Oetjen, H. Thiele, P. Maass, T. Alexandrov, On the importance of mathematical methods for analysis of maldi imaging mass spectrometry data, *J. Integrative Bioinformatics* 9 (1) (2012) 189.
- [4] R. V. de Plas, F. Ojeda, M. Dewil, L. V. D. Bosch, B. D. Moor, E. Waelkens, Prospective exploration of biochemical tissue composition via imaging mass spectrometry guided by principal component analysis, in: *Pacific Symposium on Biocomputing, 2007*, pp. 458–469.
- [5] E. R. Muir, I. Ndiour, N. A. L. Goasduff, R. Moffitt, Y. Liu, M. C. Sullards, A. Merrill, Y. Chen, M. Wang, Multivariate analysis of imaging mass spectrometry data, *Bioinformatics and Bioengineering* (2007) 472–479.

- [6] M. Gerhard, S.-O. Deininger, F. Schleif, Statistical classification and visualization of maldi-imaging data, *Computer-Based Medical Systems* (2007) 403–405.
- [7] S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle, K. Ugurbil, Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging, *Proceedings of the National Academy of Sciences* 89 (13) (1992) 5951–5955.
- [8] N. E. Helwig, S. Hong, A critique of tensor probabilistic independent component analysis: Implications and recommendations for multi-subject fMRI data analysis, *Journal of Neuroscience Methods* 213 (2) (2013) 263–273.
- [9] R. A. Harshman, M. E. Lundy, PARAFAC: Parallel factor analysis, *Comput. Stat. Data Anal.* 18 (1) (1994) 39–72.
- [10] H. Zou, T. Hastie, R. Tibshirani, Sparse principal component analysis, *Journal of Computational and Graphical Statistics* 15 (2) (2006) 265–286.
- [11] Y. Wang, Q. Wu, Sparse PCA by iterative elimination algorithm, *Adv. Comput. Math.* 36 (1) (2012) 137–151.
- [12] L. Xiong, D. Hong, Multi-resolution analysis method for IMS data biomarker selection and classification, *British Journal of Mathematics and Computer Science* 5 (1) (2015) 64–80.
- [13] F. Zhang, D. Hong, Elastic net based framework for imaging mass spectrometry data biomarker selection and classification, *Statistics in Medicine* 30 (7) (2010) 753–768.

- [14] D. Hong, F. Zhang, Weighted elastic net model for mass spectrometry imaging processing, *Math. Model. Nat. Phenom.* 5 (3) (2010) 115–133.
- [15] J. Liang, D. Hong, F. Zhang, J. Zou, Imsmining: A tool for imaging mass spectrometry data biomarker selection and classification, in: R. N. Mohapatra, D. R. Chowdhury, D. Giri (Eds.), *Mathematics and Computing*, Vol. 139 of Springer Proceedings in Mathematics & Statistics, Springer India, 2015, pp. 155–162.
- [16] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B* 67 (2) (2005) 301–320.
- [17] N. Correa, T. Adali, V. D. Calhoun, Performance of blind source separation algorithms for fMRI analysis using a group ICA method, *Magnetic Resonance Imaging* 25 (5) (2007) 684–694.
- [18] V. D. Calhoun, J. Liu, T. Adali, 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.
- [19] E. A. Allen, E. B. Erhardt, Y. Wei, T. Eichele, V. D. Calhoun, Capturing inter-subject variability with group independent component analysis of fMRI data: A simulation study, *NeuroImage* 59 (4) (2012) 4141–4159.
- [20] V. J. Schmithorst, S. K. Holland, A comparison of three methods for generating group statistical inferences from independent component analysis of functional magnetic resonance imaging data, *Journal of Magnetic Resonance Imaging* 19 (3) (2004) 365–368.
- [21] M. Assaf, K. Jagannathan, V. D. Calhoun, L. Miller, M. C. Stevens, R. Sahl, J. G. O’Boyle, R. T. Schultz, G. D. Pearlson, Abnormal functional connectivity

- of default mode sub-networks in autism spectrum disorder patients, *NeuroImage* 53 (1) (2010) 247–256.
- [22] M. A. Lindquist, The statistical analysis of fmri data, *Statistical Science* 23 (4) (2008) 439–464.
- [23] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transactions on Neural Networks* 10 (3) (1999) 626–634.
- [24] A. Hyvärinen, Independent component analysis: recent advances, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 371 (1984).
- [25] W. Lu, J. C. Rajapakse, ICA with reference, *Neurocomputing* 69 (16-18) (2006) 2244–2257.
- [26] V. Calhoun, T. Adali, M. Stevens, K. Kiehl, J. Pekar, Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis, *NeuroImage* 25 (2) (2005) 527–538.
- [27] Q.-H. Lin, J. Liu, Y.-R. Zheng, H. Liang, V. D. Calhoun, Semiblind spatial ICA of fMRI using spatial constraints, *Human Brain Mapping* 31 (7) (2010) 1076–1088.
- [28] W. Lu, J. Rajapakse, Approach and applications of constrained ICA, *IEEE Transactions on Neural Networks* 16 (1) (2005) 203–212.
- [29] E. Formisano, F. Esposito, F. D. Salle, R. Goebel, Cortex-based independent component analysis of fMRI time series, *Magnetic Resonance Imaging* 22 (10) (2004) 1493–1504.

- [30] S. Ma, X. L. Li, N. M. Correa, T. Adali, V. D. Calhoun, Independent subspace analysis with prior information for fMRI data, in: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2010, pp. 1922–1925.
- [31] A. Sharma, K. K. Paliwal, Subspace independent component analysis using vector kurtosis, *Pattern Recognition* 39 (11) (2006) 2227–2232.
- [32] B. Schölkopf, J. Platt, T. Hofmann, Towards a general independent subspace analysis, MIT Press, 2007, pp. 1361–1368.
- [33] R. Vollgraf, K. Obermayer, Multi dimensional ICA to separate correlated sources, in: T. G. Dietterich, S. Becker, Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems* 14, 2001, pp. 993–1000.
- [34] P. O. H. Aapo Hyvarinen, M. Inki, Topographic independent component analysis, *Neural Computation* 13 (7) (2001) 1527–1558.
- [35] R. Ge, Y. Wang, J. Zhang, L. Yao, H. Zhang, Z. Long, Improved fastICA algorithm in fMRI data analysis using the sparsity property of the sources, *Journal of Neuroscience Methods* 263 (2016) 103–114.
- [36] A. Hyvärinen, Independent component analysis in the presence of gaussian noise by maximizing joint likelihood, *Neurocomputing* 22 (1-3) (1998) 49–67.
- [37] A. Hyvärinen, Fast ICA for noisy data using gaussian moments, in: *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems (ISCAS 99)*, Vol. 5, 1999, pp. 57–61 vol.5.

- [38] C. Beckmann, S. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging, *IEEE Transactions on Medical Imaging* 23 (2) (2004) 137–152.
- [39] M. E. Tipping, C. M. Bishop, Probabilistic principal component analysis, *Journal of the Royal Statistical Society, Series B* 21 (3) (1999) 611–622.
- [40] C. Beckmann, S. Smith, Tensorial extensions of independent component analysis for multisubject fMRI analysis, *NeuroImage* 25 (1) (2005) 294–311.
- [41] Y. Guo, A general probabilistic model for group independent component analysis and its estimation methods, *Biometrics* 67 (4) (2011) 1532–1542.
- [42] Y. Guo, G. Pagnoni, A unified framework for group independent component analysis for multi-subject fMRI data, *NeuroImage* 42 (3) (2008) 1078–1093.
- [43] M. E. Tipping, C. M. Bishop, Mixtures of probabilistic principal component analyzers, *Neural Comput.* 11 (2) (1999) 443–482.
- [44] R. Bro, PARAFAC: Tutorial and applications, *Chemometrics and Intelligent Laboratory Systems* 38 (2) (1997) 149–171.
- [45] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, *SIAM Review* 51 (3) (2009) 455–500.
- [46] A. Cichocki, Tensor decompositions: A new concept in brain data analysis, *CoRR* abs/1305.0395.
URL <http://arxiv.org/abs/1305.0395>

- [47] A. Cichocki, Era of big data processing: A new approach via tensor networks and tensor decompositions, CoRR abs/1403.2048.
URL <http://arxiv.org/abs/1403.2048>
- [48] A. Stegeman, Comparing independent component analysis and the parafac model for multi-subject artificial fMRI data, in: Invited speaker at the IOPS Conference, 2006.
- [49] Y.-Z. Cao, Z.-P. Chen, C.-Y. Mo, H.-L. Wu, R.-Q. Yu, A PARAFAC algorithm using penalty diagonalization error (pde) for three-way data array resolution, *Analyst* 125 (2000) 2303–2310.
- [50] A. Stegeman, Using the simultaneous generalized schur decomposition as a candecom/parafac algorithm for ill-conditioned data, *Journal of Chemometrics* 23 (7-8) (2009) 385–392.
- [51] M. D. Vos, D. Nion, S. V. Huffel, L. D. Lathauwer, A combination of parallel factor and independent component analysis, *Signal Processing* 92 (12) (2012) 2990–2999.
- [52] J. M. S.-B. Eduardo Martínez-Montes, P. A. Valdés-Sosa, Penalized parafac analysis of spontaneous eeg recordings, *Statistica Sinica* 18 (4) (2008) 1449–1464.
- [53] E. Martínez-Montes, M. Vega-Hernández, J. M. Sánchez-Bornot, P. A. Valdés-Sosa, Identifying complex brain networks using penalized regression methods, *J Biol Phys* 34 (2008) 315–323.
- [54] E. Martínez-Montes, R. Sarmiento-Pérez, J. M. Sánchez-Bornot, P. A. Valdés-Sosa, *Advances in Cognitive Neurodynamics ICCN 2007: Proceedings of the*

International Conference on Cognitive Neurodynamics., Springer Netherlands, Dordrecht, 2008, Ch. Information Entropy-Based Penalty for PARAFAC Analysis of Resting EEG, pp. 443–446.

- [55] SPM, <http://www.fil.ion.ucl.ac.uk/spm/>.
- [56] GIFT, <http://mialab.mrn.org/software/gift/index.html>.
- [57] A. H. Phan, P. Tichavsky, A. Cichocki, Fast alternating ls algorithms for high order candecomp/parafac tensor factorizations, *IEEE Transactions on Signal Processing* 61 (19) (2013) 4834–4846.
- [58] A. H. Phan, A. Cichocki, PARAFAC algorithms for large-scale problems, *Neurocomputing* 74 (11) (2011) 1970–1984.
- [59] J. Liu, R. Li, R. Wu, Feature selection for varying coefficient models with ultrahigh-dimensional covariates, *Journal of the American Statistical Association* 109 (505) (2014) 266–274.