Underdetermined Joint Blind Source Separation with Application to Physiological Data

by

Liang Zou

B.Sc., Anhui University, 2010

M.Sc., University of Science and Technology of China, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Electrical and Computer Engineering)

The University of British Columbia

(Vancouver)

September 2017

Abstract

Blind Source Separation (BSS) methods have been attracting increasing attention for their promising applications in signal processing. Despite recent progress on the research of BSS, there are still remaining challenges. Specifically, this dissertation focuses on developing novel Underdetermined Blind Source Separation (UBSS) methods that can deal with several specific challenges in real applications, including limited number of observations, self/cross dependence information and source inference in the underdetermined case. First, by taking advantage of the Noise Assisted Multivariate Empirical Mode Decomposition (NAMEMD) and Multiset Canonical Correlation Analysis (MCCA), we propose a novel BSS framework and apply it to extract the heart beat signal form noisy nano-sensor signals. Furthermore, we generalize the idea of (over)determined joint BSS to that of the underdetermined case. We explore the dependence information between two datasets and propose an underdetermined joint BSS method for two datasets, termed as UJBSS-2. In addition, by exploiting the cross correlation between each pair of datasets, we develop a novel and effective method to jointly estimate the mixing matrices from multiple datasets, referred to as Underdetermined Joint Blind Source Separation for Multiple Datasets (UJBSS-M). In order to improve the time efficiency and relax the sparsity constraint, we recover the latent sources based on subspace representation when the mixing matrices are estimated. As an example application for noise enhanced signal processing, the proposed UJBSS-M method also can be utilized to solve the single-set UBSS problem when suitable noise is added to the observations. Finally, considering the recent increasing need for biomedical signal processing in the ambulatory environment, we propose a novel UBSS method for removing electromyogram (EMG) from Electroencephalography (EEG) signals.

The proposed method for recovering the underlying sources is also applicable to other artifact removal problems. Simulation results demonstrate that the proposed methods yield superior performances over conventional approaches. We also evaluate the proposed methods on real physiological data, and the proposed methods are shown to effectively and efficiently recover the underlying sources.

Lay Summary

Blind source separation aims to extract underlying sources when only the mixed observations are available. In this dissertation, we propose a set of novel underdetermined BSS methods to jointly recover underlying sources from multiple datasets in underdetermined cases (i.e. the number of sources is greater than that of observations). First, we propose a novel BSS framework and apply it to extract the heart beat signal form noisy nano-sensor signals. Taking into account the dependence information between datasets, we further propose two Underdetermined Joint Blind Source Separation (UJBSS) methods, termed as UJBSS-2 and UJBSS-m, which are applicable to two datasets and multiple datasets respectively. Finally, considering the recent increasing need for acquiring EEG signals with limited number of sensors, we propose a novel underdetermined BSS method for removing EMG artifacts from EEG signals. It is shown that our proposed methods outperform conventional BSS methods on both synthetic data and real physiological signals.

Preface

This dissertation is written based on a collection of manuscripts. The majority of the research, including literature study, algorithm development and implementation, numerical studies and report writing, were conducted by the candidate, with suggestions from Prof. Z. Jane Wang. The manuscripts were primarily drafted by the candidate, with helpful revisions and comments from Prof. Z. Jane Wang (papers in Chapter 2–5), Prof. Xun Chen (papers in Chapter 2–3) and Prof. Peyman Servati (paper in Chapter 2).

Chapter 2 is based on the following manuscripts:

- L. Zou, X. Chen, A. Servati, S. Soltanian, P. Servati and Z. J. Wang, "A Blind Source Separation Framework for Monitoring Heart Beat Rate Using Nanofiber-Based Strain Sensors," *IEEE Sensors Journal*, vol.16, pp. 762-772, 2015.
- L. Zou, X. Chen, A. Servati, P. Servati, M. J. McKeown, "A heart beat rate detection framework using multiple nanofiber sensor signals," *2014 IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP)*, pp. 242-246, Xi'an, China, Jul. 2014.

Chapter 3 is based on:

- L. Zou, X. Chen and Z. J. Wang, "Underdetermined Joint Blind Source Separation for Two Datasets Based on Tensor Decomposition," *IEEE Signal Processing Letters*. vol.23, pp. 673-677, 2016

Chapter 4 is mainly based on the following manuscripts:

- L. Zou, X. Chen, X. Ji and Z. J. Wang, "Underdetermined Joint Blind Source Separation of Multiple Datasets," *IEEE Access*, in press, 2017, doi: 10.1109/ACCESS.2017.2695497.
- L. Zou, Z. J. Wang, X. Chen and X. Ji, "Underdetermined joint blind source separation based on tensor decomposition," *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1-4, Vancouver, Canada, May 2014.

And finally, Chapter 5 is based on the following manuscript:

- L. Zou and Z. J. Wang, "Removing Muscle Artifacts From EEG Data Based on a Novel Underdetermined Blind Source Separation Method," in submitting to *IEEE Signal Processing Letters*.

Other publications related to my PhD work are:

- L. Zou, J. Zheng and Z. J. Wang, "3D CNN for Automatic Diagnosis of ADHD Using functional and structural MRI," submitted to *IEEE ACCESS*.
- Y. Hu, L. Zou, X. Huang and X. Lu, "Detection and quantifcation of oal content in ground beef meat using vibrational spectroscopic-based chemometric analysis," submitted to *Scientifc Reports*.
- J. Zheng, L. Zou and Z. J. Wang, "Mid-Level Deep Food Part Mining for Food Image Recognition," submitted to *IET Computer Vision*.

Table of Contents

List of Tables

List of Figures

Glossary

AIC Akaike Information Criterion

ALS Alternating Least Squares

ACC absolute correlation coefficients

ANOVA Analysis of Variance

BSS Blind Source Separation

CANDECOMP Canonical Decomposition

CCA Canonical Correlation Analysis

CPD Canonical Polyadic Decomposition

EEG Electroencephalography

EEMD Ensemble Empirical Mode Decomposing

ELS Enhanced Line Search

EMD Empirical Mode Decomposing

ECG electrocardiogram

EMG electromyogram

EOG electrooculography

FEEL Flexible Electronics and Energy Lab

FMRI Functional Magnetic Resonance Imaging

GGD Generalized Gaussian Distribution

HBR Heart Beat Rate

HOS high order statistics

ICA Independent Component Analysis

IVA Independent Vector Analysis

IMFS Intrinsic Mode Functions

JBSS Joint Blind Source Separation

MAC Mean of Absolute Correlation

MCCA Multiset Canonical Correlation Analysis

MDL Minimum Description Length

MEMD Multivariate Empirical Mode Decomposition

MRI Magnetic Resonance Imaging

NAMEMD Noise Assisted Multivariate Empirical Mode Decomposition

NF nanofiber

NCG nonlinear conjugate gradient

PAN Polyacrylonitrile

PARAFAC Parallel Factor Analysis

PDF probability Density Function

PPG photoplethysmogram

RHBR Resting heart beat rate

RRMSE Relative Root Mean Squared Error

SCA Sparse Component Analysis

SCICA Single-Channel Independent Component Analysis

SNRS signal-to-noise ratios

SSPS single source points

STFT Short-Time Fourier transform

TF Time-Frequency

UBSS Underdetermined Blind Source Separation

UJBSS Underdetermined Joint Blind Source Separation

UJBSS-2 Underdetermined Joint Blind Source Separation for Two Datasets

UJBSS-M Underdetermined Joint Blind Source Separation for Multiple Datasets

WICA Wavelet-Independent Component Analysis

WVD Wigner-Ville distribution

WT Wavelet Transform

Acknowledgments

Doctoral studies are challenging. I would like to thank those who have helped and supported me during these past years.

Special thanks go to my supervisor Prof. Z. Jane Wang. Her guidance and patience inspire me to move on not only in my PhD study but also in other aspects of my life. Without her generous support and wise suggestions, I would not have been able to make it to this point. I also would like to thank Dr. Wang's financial support through Natural Sciences and Engineering Research Council (NSERC) grants.

I own a debt of gratitude to my supervisory committee. Suggestions and inspiration from these great professors are valued most by a student like me, as someone who is new to the research field. Particularly, I would like to thank Prof. Martin McKeown for his valuable suggestions and efforts.

I would like to thank my lab-mates, friends and coauthors, Prof. Yuhu Cheng, Prof. Yi Yang, Dr. Ying Ji, Dr. Chunsheng Zhu, Dr. Tingting Wang, Kunzhong Jian, Jiannan Zheng, Yiming Zhang, Huan Qi, Nandinee Haq, Pegah Kharazmi, Liang Zhao, Jiayue Cai, River Huang. I would also like to express special thanks to my senior lab-mates, including Dr. Xun Chen, Dr. Chen He and Dr. Aiping Liu. Without their professional suggestions and personal encouragement, I would never have grown up as fast as I did.

I also would like to thank the China Scholarship Council for providing me financial support for the last four years.

Lastly, I am forever thankful to my beloved parents. I deeply thank them for their love and support.

Chapter 1

Introduction

One of our most important faculties is our ability to listen to, and follow, one speaker in the presence of others. — Colin Cherry

Blind Source Separation (BSS) refers to a general class of signal processing methods, aiming to recover the underlying source signals from their mixtures without resorting to any prior information about the mixing matrix. In this chapter we will review these state-of-the-art BSS methods. We begin by introducing the basic idea of BSS and its applications. Considering different assumptions of BSS methods, the BSS problems can be categorized in different ways, such as (over)determined vs. underdetermined BSS, single-set vs. joint BSS, and instantaneous vs. convolutional BSS. We further describe the structure of the dissertation at the end of this chapter.

1.1 Introduction to Blind Source Separation and Its Applications

The BSS problem appears in many multi-sensor systems, where each sensor contains the mixture of several underlying sources. For instances, an electrode for collecting Electroencephalography (EEG) signals measures a weighted sum of the

electrical activities of many brain regions as well as a microphone measures sounds of different people/devices in the environment. A fundamental goal of BSS is to recover the underlying sources which usually provide important information but cannot be directly seen in the observed signals. The term 'blind' emphasizes the fact that BSS exploits only the information carried by the observations themselves and does not resort to any prior information about the mixing matrix.

The BSS problem can be mathematically formulated as

$$X(t) = AS(t) + E(t), \tag{1.1}$$

where the observations are noisy instantaneous linear mixtures of the underlying source signals. In this context, $X(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$ denotes the M-dimensional observations and $x_m(t)$ is the m^{th} channel of the observations. $S(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$ denotes the underlying N-dimensional sources and $s_n(t)$ is the n^{th} source. $E(t) = [e_1(t), e_2(t), \ldots, e_M(t)]^T$ represents the additive noise added into each channel of the observations.

A classical example of BSS is the cocktail party problem. At a cocktail party, a group of people talk at the same time over the background music. Human beings are able to selectively focus on and recognize one auditory source in a noisy environment, where their ears capture numerous audio sources, such as the voice of interest, the background music and many others. This refers to the cocktail party effect [10]. The main questions of relevance are: how does the human brain solve the cocktail party problem, and is it possible to build a machine to fulfill this task?

From an engineering perspective, separating the voice of each speaker can be performed by using recordings of several microphones in the room, as illustrated in Fig. 1.1. In order to pick out each voice from the observed recordings, we can perform BSS to recover the sources. Essentially, the BSS problem is a prob-

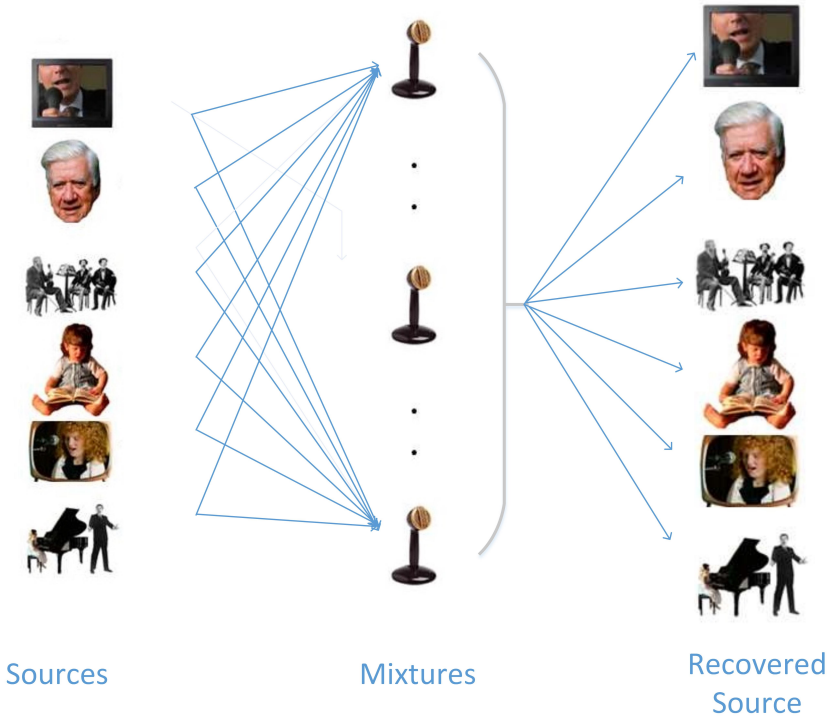

Figure 1.1: Illustration of the cocktail party problem.

lem of parameter estimation. This kind of problem always contains the following three ingredients: a parametric model, a criterion and an optimization technique. The parametric model provides a simple representation of the mixing model and restricts the solution to a particular space. Based on the assumption about the distribution of sources, the criterion is a specific cost function, evaluating the quality of the potential solutions. The optimization technique is the method for obtaining the optimal solution with respect to the criterion function. Generally, for a BSS problem, we can find a demixing matrix W to estimate the source signals as

$$S(t) = WX(t). \tag{1.2}$$

For successful blind source separation, the demixing matrix W satisfies

$$WA = \Phi\Lambda, \tag{1.3}$$

where Φ is an $(N \times N)$-dimensional permutation matrix with exactly one entry of 1 in each row and each column, and 0s elsewhere. Λ is a diagonal matrix with non-zero diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_N$. This equation explicitly shows the typical indeterminacies of permutation and scaling. However, it is clear that, without additional assumptions, the BSS problem is still ill-posed, even if a scale-permutation ambiguity is allowed. A commonly used BSS method is Independent Component Analysis (ICA), which makes use of the following assumptions:

1) The source signals $S(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$ are assumed to be statistically independent. This assumption is critical for ICA and implies that

$$p(S) = p(s_1, s_2, \ldots, s_N) = p(s_1)p(s_2)\ldots p(s_N), \tag{1.4}$$

where $p(S)$ is the joint probability Density Function (PDF) of the sources S, and $p(s_n)(n = 1, 2, \ldots, N)$ is the PDF of the n^{th} channel of sources.

2) At most one source can be Gaussian distributed. The linear mixtures of Gaussian signals are still Gaussian and therefore people cannot separate the source based on the non-Gaussianity.

Besides ICA, many BSS methods based on other properties of sources and mixing models have been proposed. For instance, Sparse Component Analysis (SCA) is a simple but powerful framework to separate source signals from a few sensors [47] when the sources can be represented sparsely at a given basis. In the convolutive BSS, the mixing model is much more complex, which assumes the mixtures are weighted and delayed. Designing a BSS system is a challenging task

despite the fact that human ability to recover sources in noisy environments appears a simple process for humans. The selection of a suitable separation method varies with the assumed mixing process. In order to achieve satisfying results in real-life applications, it is a prerequisite that [25]:

1) The real sources satisfy the assumption of the proposed BSS method, such as independence between sources, sparsity of the sources, non-negative nature of the sources, etc.;

2) The mixing model of the proposed method is correct, which means that the mixing model is identical or closely resembles to the physical model generating the observations.

If the above conditions are not satisfied, e.g., a wrong mixing model or sources' distribution is utilized, the performance of the BSS method can not be guaranteed.

BSS methods have been attracting increasing attention for their promising applications in signal processing. Among the potential applications, four domains have been studied intensively:

1) audio source separation, such as audio signal enhancement by removing unwanted signal components [110];

2) biomedical applications, such as denoising of EEG, Magnetic Resonance Imaging (MRI) and surface electromyogram (EMG) signals [80];

3) communication applications, such as co-channel interference suppression in multiple-antenna systems [25];

4) surveillance applications, such as discovering interesting signals among a set of mixed signals [16].

In addition, BSS methods can also be used in other domains, such as image processing, watermarking, remote sensing, astrophysics and so on [24].

1.2 Related Work

Depending on different assumptions of the existing BSS methods, the BSS problem can be categorized in different ways, such as (over)determined vs. underdetermined BSS, single-set vs. joint BSS, and instantaneous vs. convolutive BSS. This section describes the related work in literature for each of these categories.

1.2.1 (Over)Determined Blind Source Separation vs. Underdetermined Blind source Separation

In order to select an appropriate BSS method, an important concern is the relationship between the number of observations and the number of underlying sources. In the (over)determined case where a sufficient number of observations are available (i.e., the number of observations $M \geq$ the number of sources N), the sources can be effectively recovered without making strong assumptions about the sources and/or the mixing model. A typical way to solve this type of BSS problems is ICA. A wide variety of ICA algorithms have been developed since the original work of Bell etc. Mixing matrices are optimally estimated based on different criteria, such as kurtosis, mutual information, negentropy and log-likelihood [51]. Even though they are presented in different formalisms, these ICA-based BSS methods are shown to be mathematically equivalent [51].

However, in the underdetermined case, where the number of sensors M is smaller than the number of sources N, the BSS problem is much more complex than that in the (over)determined case and ICA-based methods may not work. The main reason is that the mixing matrix is not invertible and hence the estimation of the mixing matrix is not equivalent to the reconstruction of the sources. The performances of most (over)determined BSS algorithms are often degraded adversely or even are no longer applicable to Underdetermined Blind Source Separation (UBSS) problems. Instead, additional prior knowledge about the sources is required to

recover the underlying sources, such as the sparsity of the sources.

Generally speaking, in the underdetermined case, the recovery of the underlying sources consists of two steps: The mixing matrix is first estimated, and then the underlying sources are separated based on the estimated result of the mixing matrix. A large class of algorithms to estimate the mixing matrix in the underdetermined cases starts from the assumption that the sources are sparse. The scatter plot shows high signal values in the direction of the mixing vectors [31]. Some clustering methods, such as K-means and fuzzy c-means, are widely used to estimate the mixing matrix. The clustering-based methods generally perform an exhaustive search in the mixing vector space. Therefore, they are time-consuming, especially when the number of observations is greater than two [31, 37]. In addition, in reality, sources are usually not sufficiently sparse. They can have sparser representations in some transformed domains, such as the Time-Frequency (TF) domain via Short-Time Fourier transform (STFT), Wavelet Transform (WT) and Wigner-Ville distribution (WVD) etc. However, clustering-based methods cannot estimate the mixing matrix precisely if the sources are still insufficiently sparse in the transformed domain. To improve the accuracy of mixing matrix estimation, many algorithms have been proposed for detecting the single source points (SSPS) in the TF domain [37]. Initially, the SSPs denote the regions where only one source is active. However, this condition is noted as too strict and the assumption is relaxed by many researchers. For instance, the SSPS denote the TF regions where only one source is dominant [71]. Several algorithms were proposed to detect the SSPS. Abrard et al. proposed the TIFROM method to detect the single source points by identifying the regions where the complex ratio of the TF transform remains constant [1]. In [7] and [9], the SSPs were detected based on the eigenvalue decomposition of the covariance matrices of the mixtures. After the detection of SSPs, the clustering method can be utilized to estimate the mixing matrix. As an illustrative example,

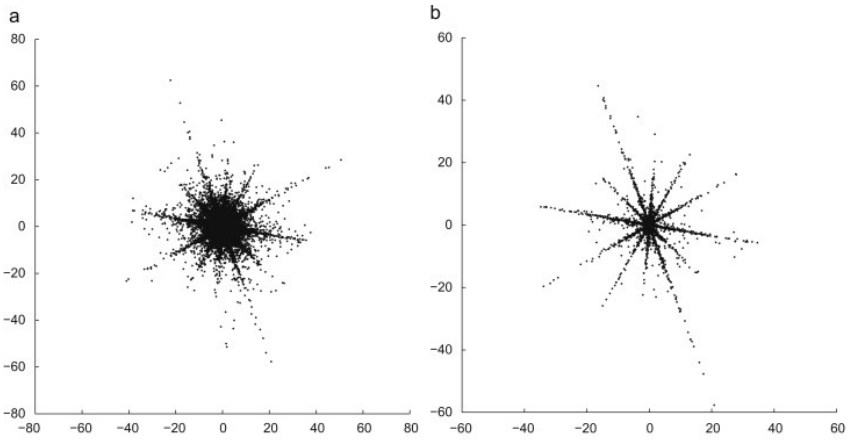

Figure 1.2: Illustrative example of a scatter plot of the mixtures with the number of observations M =2 and the number of sources N =6. (a) is for all DFT coefficients, and (b) is for samples at single source points.

Fig.1.2 shows a scatter diagram of the mixtures taking samples from 40 frequency bins with $M = 2$ and $N = 6$ [91]. Fig.1.2(a) shows the distribution of all the discrete Fourier transform coefficients whereas Fig.1.2(b) shows the result based on the SSPs. It is apparent that the scatter plot in Fig.1.2(b) has a clearer orientation towards the directions of the mixing vectors [91].

Another category of methods to estimate the mixing matrix exploit the statistical properties of the sources. Based on the fourth-order cumulant of the signals, De Lathauwer et al. proposed the FOOBI algorithm [32]. He further exploited the second-order correlation of the sources and applied simultaneous matrix diagonalization-based techniques to estimate the mixing matrix in the underdetermined case [31]. Similar algorithms can also be found in more recent references [84, 111, 124].

In the UBSS problems, the sources are not obtained easily and must be inferred even when the mixing matrix is known or estimated. The sources generally are inferred via the maximum posterior approach or the maximum-likelihood approach.

In order to model the underlying source distributions, Cemgil et al. utilized the student-distribution [15], and Snoussi et al. used the generalized hyperbolic distribution [97], respectively. In order to model a wide class of signals, such as audio and biological signals, Kim modeled the sources' distributions via the Generalized Gaussian Distribution (GGD) and recovered the underlying sources based on subspace representation [59]. Another commonly used method to infer sources is sparse representation. For instance, Li et al. considered the sparse representation of sources via l_1-norm minimization [70, 71]. In order to design overcomplete dictionaries for sparse representation, Aharon et al. proposed a novel algorithm based on K-means clustering and singular value decomposition (K-SVD) [6]. Recently, Zhen et al. exploited the sparse coding of TF representations of observations and proposed a UBSS strategy based on sparse coding [119].

However, all the UBSS algorithms mentioned above have been developed for extracting the sources from a single dataset. To the best of our knowledge, there has been very limited work on UBSS methods specifically proposed for exploring the correlations between multiple datasets and extracting sources simultaneously.

1.2.2 Single Set Blind Source Separation vs. Joint Blind Source Separation

BSS was originally designed for extracting the sources from mixed signals in a single dataset. However, in recent years blind separation of multiple datasets simultaneously (i.e., Joint Blind Source Separation (JBSS)), has been increasingly important in many applications, such as the group data analysis of multi-subject/multi-session Functional Magnetic Resonance Imaging (FMRI) [67]. For the case of K datasets, the mixing process can be modeled as

$$X^{[k]} = A^{[k]}S^{[k]}, \quad k = 1,2,...,K \tag{1.5}$$

where $X^{[k]} \in \mathbb{R}^{M \times T}$ represents the M-dimensional observations, $S^{[k]} \in \mathbb{R}^{N \times T}$ are the N-dimensional sources, $A_k \in \mathbb{R}^{M \times N}$ is the mixing matrix, and it is assumed that $M \geq N$ (i.e., the (over)determined case). T is the number of data samples. In this problem formulation, traditional BSS techniques face the following challenge: even if common sources can be extracted when we apply the BSS to each dataset individually, the mixing matrix of each dataset could have an arbitrary permutation [8]. In addition, this individual strategy also neglects the correlations between the multiple datasets. To address these concerns, JBSS methods have recently been proposed, aiming to extract underlying sources so that the estimated sources are aligned across datasets [8].

There have been a number of methods proposed for JBSS [49, 56, 67, 68, 72]. Among them, Canonical Correlation Analysis (CCA) has been a powerful JBSS tool, whereas originally it was proposed only to analyze the correlation structure between two datasets and not for the joint analysis of multiple datasets [49]. Multiset Canonical Correlation Analysis (MCCA), an extension of CCA, was proposed by Kettenring et al. to analyze the linear relationships between multiple variates by only using second-order statistics [56]. MCCA has been quite effective for the analysis of multi-subject fMRI data [67]. Another extension of CCA, the joint diagonalization using second-order statistics (JDIAG-SOS), provided a computationally efficient way to solve JBSS problems and avoid the deflationary procedure of MCCA [68].

The group ICA and the joint ICA are the extensions of the ICA from one to multiple datasets [14]. The group ICA assumes that different datasets share a common signal subspace and maximizes the statistical independence of the sources within the extracted source subspace [14]. The joint ICA was developed to maximize the independence of joint sources of multiple datasets by assuming that K datasets share the same mixing matrix [14]. Fig. 1.3 shows the difference between

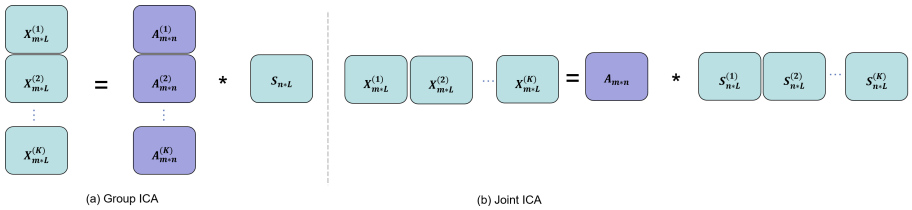

Figure 1.3: Graphical comparison between (a) Group ICA and (b) Joint ICA.

these two methods. Both methods were demonstrated to be powerful in a variety of applications and can provide meaningful results [14]. However, these two approaches rely on stringent assumptions, which may not be satisfied in some studies and applications [19].

It should be noted that the exiting JBSS algorithms mentioned above generally assume that the number of sources equals to or is less than that of the observation signals (i.e., (over)determined case). This assumption may not be true in certain applications. However, to the best of our knowledge, there are no JBSS methods specifically designed for the underdetermined case when the number of sources is greater than that of observation signals.

1.2.3 Instantaneous Blind Source Separation vs. Convolutional Blind Source Separation

In instantaneous BSS, the observations are weighted sums of the individual sources without time delay, as shown in 1.1. However, in many real-life situations, the mixing model is much more complex. For instance, in acoustics, the microphones not only pick up the original sources but also the attenuated and delayed versions of the sources corresponding to the sound waves bouncing back from the wall and ceiling. Therefore, the assumption of the instantaneous mixing model does not hold in this scenario. Then the instantaneous mixing model is converted to the convolutive mixing model, as

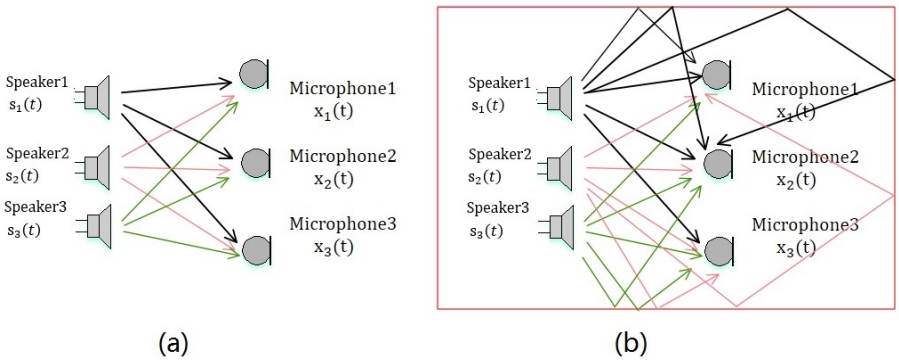

Figure 1.4: Graphical comparison between instantaneous BSS and convolutive BSS: (a) for instantaneous BSS without time delay and (b) for convolutive BSS with reverberation.

$$X(t) = \sum_{k=0}^{K-1} A_k S(t-k), \tag{1.6}$$

where $X(t) = [x_1(t), x_2(t), \ldots, x_M(t)]^T$ denotes the M-dimensional observations, $S(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$ denotes the underlying N-dimensional sources and $[s_1(t-k), s_2(t-k), \ldots, s_N(t-k)]^T$ is the delayed version of the original sources $S(t)$. The observations are the linear mixture of the delayed version of the sources, and the matrix $A_k \in \mathbb{R}^{M \times N}$ denotes the mixing matrix corresponding to the time delay k. A graphical comparison between instantaneous and convolutive models is shown in Fig.1.4. In convolutive BSS, each microphone records filtered versions of the sources, which can be written as,

$$x_m(t) = \sum_{k=0}^{K-1} \sum_{n=1}^{N} A_k(m,n) s_n(t-k). \tag{1.7}$$

Each source contributes to the sum with multiple delays corresponding to multiple paths by which the signal from the corresponding speaker propagates to a microphone. Various attempts have been made to tackle this type of convolutive BSS

problem [95]. A commonly used way is based on frequency-domain [6][13], which converts time-domain observation signals into frequency-domain time-series signals by a STFT.

The instantaneous mixing model is a special case of the convolutive model where the delay is set to be 0. The instantaneous BSS methods can be extended to solve the convolutive BSS problem with more constraints. In this dissertation, the focus is on novel BSS methods for instantaneous mixtures, especially for the underdetermined cases.

1.3 Research Objectives and Methodologies

The goal of this dissertation is to develop a set of novel underdetermined blind source separation methods which are able to cope with several challenges present in real applications. As we introduced previously, the underdetermined BSS problem is generally more difficult than the (over)determined BSS problem where the number of observations is equal to or greater than that of underlying sources. It is a difficult topic with some challenges including limited observations, the convolutive mixing model, an unknown number of sources, dependence information, temporal dynamics, efficiency and so on.

Considering the recent research progress in underdetermined BSS, we are interested in addressing the following concerns and challenges. First, although some recent overcomplete representation methods have been proposed, the problem to recover the underlying sources from limited number of observations remains a difficulty. Traditional underdetermined BSS methods, both the methods based on the sparse properties and the methods based on the statistical properties of source signals, may fail to work when the number of sources is greatly larger than that of the observations. Second, the sources may be mixed with a time delay way rather than in an instantaneous way. The sources corresponding to different observations

may be correlated with each other rather than identical as in the instantaneous BSS mixing model. In that case, traditional BSS methods will fail to solve the BSS problem. Third, the dependence information across multiple datasets should be taken into account in solving the underdetermined BSS problems. Joint analysis in BSS has attracted great attention owing to its ability to simultaneously recover the underlying sources from multiple datasets. However, to the best of our knowledge, existing JBSS methods in the literature consistently assume that the number of observations is equal to or greater than that of the underlying sources. Therefore, these methods are not suitable for solving underdetermined BSS problems. In addition, it is a challenge to obtain the underlying sources in the underdetermined BSS even when the mixing matrix is known. As a result, methods that can effectively and efficiently infer the underlying sources are highly desired.

In this research work, we attempt to develop novel underdetermined BSS methods to address the aforementioned challenges, including limited number of observations, presence of highly correlated sources across observations (rather than identical sources), and self/cross dependence and source inferences. Specifically, the main research contributions of this dissertation are summarized as follows:

- For the case of the underdetermined BSS, we propose a two-step source extraction method termed as NAMEMD-MCCA. Considering the multichannel nature of the measured signals, Noise Assisted Multivariate Empirical Mode Decomposition (NAMEMD) is utilized to decompose each channel of these signals into a series of Intrinsic Mode Functions (IMFS). Then the underdetermined problem is converted into the determined problem, as we have a sufficient number of IMFs. In addition, the sources corresponding to different observations may be highly correlated rather than identical. We decompose each channel into a series of IMFs, which can be regarded as

a dataset. Considering the correlations between the sources across multiple observations, a classical JBSS method is employed to solve this problem. Benefiting from cross-channel information and its increased robustness to artifacts, the proposed method can achieve superior performance in recovering the underlying sources.

- Inspired by the canonical correlation analysis model and the simultaneous matrix diagonalization, we exploit the second order statistics of the observations in two datasets and propose a novel UBSS method which can accurately estimate the mixing matrices in the underdetermined case. In addition, we employ a novel time-frequency analysis method to recover the sources. The proposed method is referred to as the Underdetermined Joint Blind Source Separation for Two Datasets (UJBSS-2).

- Exploiting the cross correlation between each pair of datasets, we propose a novel and effective method to jointly estimate the mixing matrices for multiple datasets in the underdetermined case. We further recover the underlying sources individually based on subspace representation. We extend the idea of (over)determined JBSS to that of the underdetermined case. The proposed BSS method is referred to as Underdetermined Joint Blind Source Separation for Multiple Datasets (UJBSS-M). The basic idea of UJBSS for multiple datasets is shown in Fig.1.5.

- We further explore the second-order statistics of the underlying sources, including autocorrelation and cross correlation information, and propose a novel EMG artifact removal method. The proposed method can recover the latent sources and reconstruct noise-free EEG signals with a limited number of observations.

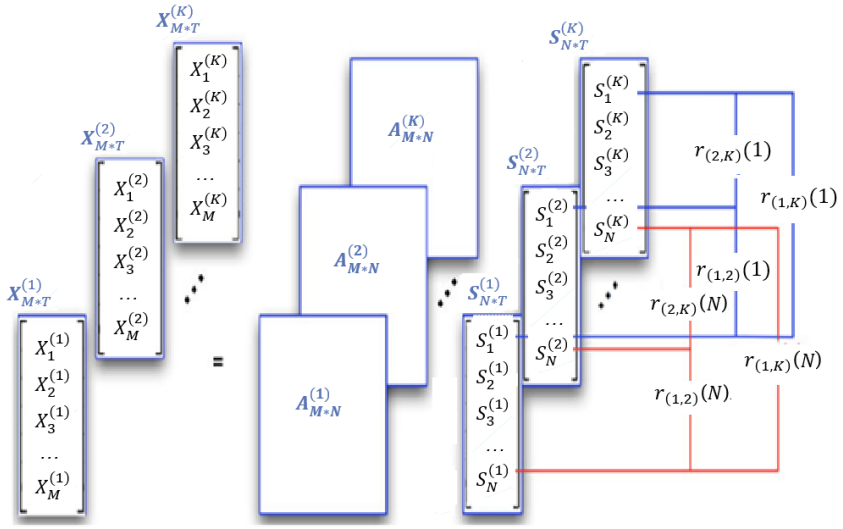

Figure 1.5: The basic idea of UJBSS for multiple datasets. M is the number of observations and N is the number of sources in each dataset. T represents the number of data samples, and K represents the number of datasets. $r_{(k_1,k_2)}(n)$ denotes the correlation between the n^{th} source in the k_1^{th} dataset and that in the k_2^{th} dataset.

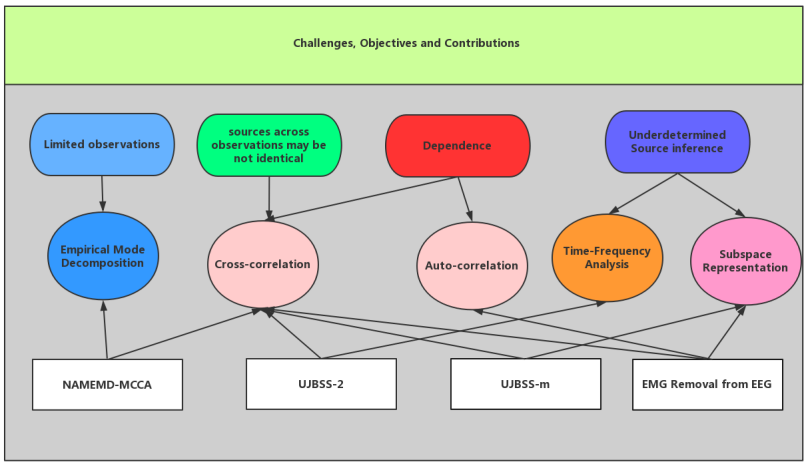

Figure 1.6: The overview of the challenges, objectives and contributions of this research work.

Fig.1.6 illustrates the challenges, objectives and contributions of this research work.

1.4 Thesis Outline

The remainder of this dissertation is structured as follows:

In Chapter 2, in order to address certain concerns of the currently available EMD-BSS based methods, we propose a novel blind source separation framework by combining NAMEMD and MCCA. The proposed method takes advantage of the multivariate data-adaptive nature of the NAMEMD and MCCA, which contributes to accurate extraction of the desired signal. The experimental results on both simulated data and real data demonstrate the superior performance of the proposed method.

In Chapter 3, we extend the idea of (over)determined JBSS to that of the underdetermined case and introduce a novel blind source separation method, termed as UJBSS-2. Considering the dependence information between two datasets, the problem of jointly estimating the mixing matrices is tackled via Canonical Polyadic Decomposition (CPD) of a specialized tensor in which a set of spatial covariance matrices are stacked. Furthermore, the estimated mixing matrices are used to recover sources from each dataset separately.

In Chapter 4, as a generalization of our previous work on two datasets in Chapter 3, we exploit the cross-correlation between each pair of datasets and present another novel blind source separation method, referred to as UJBSS-m. In this chapter, the cross correlation between each pair of datasets is modeled by a third-order tensor in which a set of spatial covariance matrices corresponding to different time delays are stacked. Considering the latent common structure of these constructed tensors, the mixing matrices are jointly estimated via joint canonical polyadic decomposition of these specialized tensors. Furthermore, we recover

the sources from each dataset separately based on the estimated mixing matrices. Simulation results demonstrate that the proposed UJBSS-m method yields superior performances when compared to commonly used single-set UBSS and JBSS methods.

In order to efficiently acquire physiological signals in the ambulatory environment, a small number of sensors is usually desired. Conventional artifact removal methods based on BSS, such as CCA and ICA, could fail to remove the artifacts in this case. In Chapter 5, we further explore the cross correlation and autocorrelation of the underlying sources and propose a novel underdetermined BSS method to remove the EMG artifacts from EEG signals. We conduct a performance comparison through numerical simulations in which EEG recordings are contaminated with muscle artifacts. The results demonstrate that the proposed method can effectively and efficiently remove muscle artifacts while successfully preserve the EEG activity.

Finally, Chapter 6 summarizes the contributions of this dissertation, and discusses future research directions.

Chapter 2

A Blind Source Separation Framework Based on NAMEMD and MCCA

A recently developed novel nanofiber based strain sensor is introduced as a potential alternative to conventional measurement tools for Heart Beat Rate (HBR) monitoring. Since the measured signals in real-life are often contaminated by certain artifacts, in this chapter, to overcome limitations of currently available EMD-BSS based methods and recover the buried heart beat signal accurately, we propose a novel blind source separation framework by combining NAMEMD and MCCA. The proposed method takes advantage of the multivariate data-adaptive nature of the NAMEMD and MCCA, which contributes to accurate extraction of the desired signal. The absolute correlation coefficients (ACC) between the extracted signal and the original source signal are adopted to evaluate the performance of the proposed method in the simulation study. The average of the ACC yielded by the proposed method is 0.902, which is significantly higher than that by state-of-the-art

approaches. We also examine the proposed method on the nano-sensor data collected when the subject performs 11 tasks. It is shown that the proposed method can achieve better performance, especially for preserving the shape of the heart beat signal.

2.1 Motivation and Objectives

As health costs are increasing and the world population is aging [96], monitoring physiological signals, such as body temperature, oxygen concentration, weight and fat levels, has drawn a lot of attention and has been used as an important tool for accessing people's fitness and health [54]. Resting heart beat rate (RHBR), one type of vital physiological signs, indicates the speed of the heart beat in the relaxed condition and can provide an assessment of health status [53]. It has been clinically approved that there is a direct relation between higher RHBR and higher mortality for the elderly and patients with cardiovascular problems, hypertension, and metabolic syndromes [63, 85, 86], whereas a relatively lower RHBR is probably a signature of better fitness and health condition. Therefore, it is of great importance to monitor health condition by measuring and tracking the RHBR.

For real-world applications such as the emerging wearable health monitoring systems [118], there is a need for developing and testing novel, inexpensive, non-invasive and miniature sensors for heart rate monitoring purposes. In this work, we adopt a novel type of nanofiber based strain sensors recently developed by Flexible Electronics and Energy Lab (FEEL). The compliant and conformable nature of this type of sensors, in addition to their sensitivity, stable behavior and low performance power [98], makes them a great choice for home health monitoring systems. Its high sensitivity of such sensors enables us to detect very small deformations such as those induced in the wrist by radial pulses.

To measure the RHBR, 3 nanofiber based strain sensors are deployed on the

wrist. Yet, the signals measured by the nanofiber based sensors are often contaminated by instrumentation noise, motion artifacts and other types of interference. It is challenging to remove artifacts and obtain clean radial pulse signal, especially when the subject might be moving. Although there exists a great variety of artifact reducing or removal techniques based on WT, Empirical Mode Decomposing (EMD) and BSS [29, 66, 93, 108], they have certain limitations in recovering the underlying unknown source signals. Also, it is worth emphasizing that different systems (sensors) will generate different types of bio-signals, such as EEG and EMG, and generally will pose different challenges for signal denoising, artifact removal and data analysis. For a particular problem, the appropriate approach is generally sensor-dependent and application dependent. Since the type of nanofiber-based sensor used in this chapter is novel and its application to heart beat rate estimation is also new, it is not clear which method is more suitable for our specific problem (i.e., RHBR monitoring using a novel type of nano-sensor signals) and whether a new method is desired.

In wavelet-based approaches [13], researchers need to select a set of basis components in advance and then calculate the related coefficients for each component. The performance of these methods heavily depends on the predefined basis components [13]. Given that the ground truth is generally unknown, the predefined basis components can not be guaranteed to match the signals of interest. EMD was first introduced by Huang et al. [50] as a technique for processing nonlinear and non-stationary time series, such as biomedical signals [66]. It decomposes a time series signal into a number of spectrally independent oscillatory components, defined as IMFS. EMD is nonparametric in the sense that all the IMFs are derived empirically from the data. Thus, it is a fully data-driven method [77]. However, this method is very sensitive to noise, making it ineffective in low signal-to-noise ratio (SNR) situations [81]. BSS methods, such as ICA and CCA, are increasingly being used as

artifact removal tools for analyzing multivariate biomedical signals [39]. They can recover the underlying sources of interest when there are sufficient channels. However, these BSS methods commonly assume that the number of sources cannot be greater than that of channels/sensors. In fact, this assumption may not be satisfied in practice, especially when minimal instrumentation is required [19], such as the RHBR measurement case.

To overcome these limitations, WT and Ensemble Empirical Mode Decomposing (EEMD) [114] were generally used to decompose the unidimensional signal from each sensor into multichannel signals before applying BSS methods. Lin et al. proposed a method called wavelet-ICA expanding the 1-D signal into 2-D signal by WT and then extracted independent features from the 2-D signal by FastICA [74]. Considering the advantage of data-driven methods, Mijovic et al. introduced a single-channel technique by combining EEMD and ICA (EEMD-ICA) [81]. It decomposed the single-channel signal into spectrally independent components and then extracted statistically independent source signals. This technique was shown to be more effective compared to wavelet-ICA. Recently, a more efficient method combining EEMD with CCA (EEMD-CCA) was proposed by Sweeney et al. [103]. It provided superior performance for source recovery and artifact suppression. However, these existing methods only focus on one single channel each time and neglect the possible inter-channel information when multiple sensors are available. Two significant obstacles of these methods in dealing with multichannel signals are the problem of uniqueness and mode mixing with the channel-wise EMD, such as EEMD. There is no guarantee that the decomposition results of different channel signals are matched [87], either in the number or their frequency, making the multichannel analysis often very difficult. In addition, it was shown that the multivariate signal processing can provide more insights into complex and nonstationary real-world process [76]. The essences of dynamical

behavior are intrinsic correlations, within a single-channel and more importantly across multiple channels. Therefore, the above methods including EEMD-ICA and EEMD-CCA may not be good choices for multichannel artifacts removal.

Considering the above concerns, in this chapter, we propose a novel framework for extracting heart beat signals by combining NAMEMD [90] with MCCA [49, 72], termed as NAMEMD-MCCA. NAMEMD, which processes the input signal directly in the high dimensional space and considers the inter-channel information, can effectively overcome the problems of uniqueness and mode mixing [77]. MCCA, as an extension of CCA, is developed to find linear transforms that simplify the correlation structure among a group of random vectors [72]. It can jointly extract the sources from each dataset through maximizing the correlations of the extracted sources across datasets. The combination of these two methods benefit from the use of inter-channel information and increased robustness to artefacts. Considering the multichannel nature of the measured signals, NAMEMD is first utilized to decompose each channel of these signals into a series of IMFs. The IMFs with dominant frequencies close to that of typical heart beat signals are selected to form corresponding datasets. These datasets are further employed as the input to MCCA for recovering the heart beat signals, and finally the RHBR can be detected.

We evaluate the performance of the proposed method in both synthetic data and real data. We first validate it on simulated data and illustrate its performance. These simulations are performed under realistic assumptions in which mixed signals are linear mixture of photoplethysmogram (PPG), EMG, motion signals, power frequency signals and Gaussian white noises. We then apply the method to nanofiber sensor signals collected when the subject was performing 11 tasks. RHBRs are estimated by the proposed framework and comparison also suggests its superiority. The main contributions of this chapter are manifold: we explore a novel type

of nanofiber based strain sensor for potential RHBR monitoring; we investigate the state-of-the-art EMD-BSS based methods for exacting RHBR information accurately based on the nano-sensor data; and we further propose the NAMEMD-MCCA for improved RHBR monitoring.

2.2 Methods

In this section, we first briefly introduce the EMD related methods and the BSS techniques. We then present a novel heart beat rate detection method based on the combination of NAMEMD and MCCA. Finally, both synthetic data and real nano-sensor data are described.

2.2.1 Related Methods

Empirical mode decomposition

EMD is a fully data driven method which adaptively decomposes a signal into a residue and a finite set of spectrally independent oscillatory components, called IMFs. As it is adaptive in nature, the IMFs usually offer a physically meaningful representation of the underlying process [113], which render its wide application in the biomedical signal processing.

However, the standard EMD method may suffer from the problem of uniqueness and mode mixing. To overcome these issues, EEMD, a noise assisted version of EMD, was proposed and has been demonstrated to be more robust in real-life applications. It defines each IMF component as the mean of the IMFs extracted by applying standard EMD on the signal corrupted by added white noise of finite amplitude. The noise of each trail is different and its effect in the ensemble mean can be canceled out when there are enough trials. This contributes to establish a uniformly distributed reference scale and then perform a quasi-dyadic structure of

the corresponding filter bank [90]. However, despite being a significant step forward, the EEMD is still a univariate method, which cannot solve the uniqueness problem. In addition, it is further limited by its computational complexity.

Multivariate empirical mode decomposition

The issues of common mode alignment and nonuniqueness have been the major obstacles for applications of these channel-wise EMD methods in multivariate data analysis. The bivariate EMD (BEMD) and rotation-invariant EMD (RI-EMD) were proposed considering the information between two channels whereas these two algorithms can cater only 2-channel signals [11, 92]. To cope with this problem, Multivariate Empirical Mode Decomposition (MEMD) was introduced by Rehman and Mandic in [89], which also demonstrated its ability to do matched-scale decomposition across multichannel signals, preserving the multichannel properties. The input multivariate signals are mapped into multiple real-valued 'projected' signals along directions in m-dimensional spaces. The detail of the MEMD algorithm is described in Algorithm 1.

Similar to the EEMD, Rehman et al., taking the advantage of MEMD, proposed a noise assisted extension of the MEMD, namely NAMEMD. Firstly, based on the MEMD method, it decomposes the original signals along with n-channel multivariate independent Gaussian white noise. At the end of this process, it discards the IMFs corresponding to the n-channel white noise and obtains a set of IMFs for the original signal. It should be noted that it is different from the EEMD, in which the noise is directly added to the observed signals, as the noise here resides in a different subspace. The Gaussian white noise can be used to enforce a filter bank structure and provide better definition of frequency bands inherent to the data. Therefore, the NAMEMD can accurately align the common oscillatory modes in the IMFs from multichannel signals and effectively alleviate the problem of mode

Algorithm 1 The MEMD algorithm

Input: m-dimensional signal $X(t) = [x_1(t), x_2(t), ..., x_m(t)]^T$
Output: the IMFs corresponding to each channel of the $X(t)$

1: Generate a suitable point set based on Hammersley sequences for sampling on (m-1) dimension sphere;
2: Calculate the projections $\{p^k\}_{k=1}^K$ of the original signal along all the K direction vectors $\{v^k\}_{k=1}^K$, where K represents the number of direction vectors;
3: Identify the time points $\{t_i^k\}_{k=1}^K$ correspond to the maxima of each projects;
4: Interpolate each maxima of each projection $[t_i^k, p^k]$ and get the multivariate envelops $\{e^k(t)\}_k^K = 1$;
5: Calculate the mean by $mean(t) = 1/K \sum_{k=1}^K e^k(t)$;
6: Calculate the detail signal $d(t) = x(t) - mean(t)$. If $d(t)$ satisfies the stopping criterion for a multivariate IMF, then obtain an IMF $c(t) = d(t)$ and repeat step 2 to step 6 on $x(t) - d(t)$, otherwise apply them on $d(t)$.

mixing within the resulted IMFs [90].

Independent component analysis

In this section, we illustrate the basic idea of ICA. The observed signal can be denoted as $X(t) = [x_1(t), x_2(t), ..., x_m(t)]^T$ whose elements are linear mixtures of n independent elements of the source component vector $S = [s_1(t), s_2(t), ..., s_n(t)]^T$ and m is greater than or equal to n. We can model the mixing process as follows:

$$X = AS \tag{2.1}$$

where the matrix A represents the linear mixture of the source components. The goal of ICA is to find the unmixing matrix W to recover the original source by assuming that they are independent and no more than one component is Gaussian-signal. The FastICA is a commonly used way to do the independent component analysis and we also evaluate its performance in this work.

Multiset canonical correlations analysis

CCA has been a powerful BSS tool, however, it can only analyze the correlation structure between two datasets [49]. Therefore, it cannot be used in this work, involving three datasets. MCCA, an extension of CCA, was proposed by Kettenring et al. to analyze the linear relationships between multiple variates [56, 72]. Considering the case for K datasets, the observed vector of each dataset contains the linear mixtures of the corresponding l sources. We can model the mixing process as following,

$$X_k = A_k S_k \quad k = 1, 2, ..., K \tag{2.2}$$

Where X_k and $S_k \in R^l$ are l dimensional observed vector and underlying sources respectively, $A_k \in R^{l \times l}$ is the non-singular mixing matrix. Then the source canonical vector can be unmixed based on linear combination of the observed vector by:

$$S_k = W_k X_k \quad k = 1, 2, ..., K, \tag{2.3}$$

where W_k is the canonical transformation unmixing matrix of the k^{th} dataset. It can be obtained by optimizing objective functions related to the overall correlation. The MCCA ensures that the corresponding latent sources are highly correlated across datasets, and meanwhile the latent sources are uncorrelated within each dataset. One thing should be noted is that MCCA reduces to CCA when the number of datasets m equals to 2.

EEMD-ICA

The use of EEMD in combination with ICA for source separation was first proposed by Mijovi et al. in [81] and was employed for the removal of artifacts from single channel EEG and EMG signals. The EEMD was used to decom-

pose a single-channel signal into a multichannel signal, comprised of IMFs. Then these IMFs can be employed as the input to the ICA method with the aim of estimating the underlying true source signal. Finally, the sources determined to be artifacts were discarded and the remaining sources were used to reconstruct the signal-channel signal free of artifacts. Mijovi et al. applied the proposed method EEMD-ICA on simulated data and the real EEG as well as EMG signals and compared it with the other two signal-channel artifacts removing methods, namely Single-Channel Independent Component Analysis (SCICA) and Wavelet-Independent Component Analysis (WICA). It was shown that their proposed method achieved the best performance [81].

EEMD-MCCA

The EEMD-MCCA [123] is an extension of the previously introduced method, namely EEMD-CCA, which was first detailed by Sweeny et al. as a novel single-channel artifact removal technique [103]. Similar to these two signal-channel denoising methods, including EEMD-ICA and EEMD-CCA, it decomposed the single channel signal into multichannel IMFs components and then input the IMFs of interest into the BSS method. However, one thing should be noted is that MCCA does not extract the source signals channel-wisely. Considering the inter-channel information, it can jointly extract the underlying source signals from multiple channels.

2.2.2 Proposed Framework

NAMEMD-MCCA

To extract the heart beat signal from the nanofiber based strain sensor signals, we propose taking advantage of both NAMEMD and MCCA by exploring their com-

bination and denote the proposed method as NAMEMD-MCCA, which is shortly outlined in Algorithm 2. Firstly, NAMEMD is applied to the observed signals. Each channel of the observed signals is decomposed into multiple IMFs components. Utilizing the dyadic filter bank property of MEMD, NAMEMD can effectively improve the performance of frequency localization and reduce mode-mixing. In addition, it is shown that NAMEMD can alleviate the effects of noise in EEMD method [105]. Then the IMFs components determined to be artifacts are discarded and these observed signals are converted into groups of datasets, denoted by X_1, $X_2, \ldots, X_K$. These datasets are selected as input to the MCCA algorithm. MCCA provides an effective way to jointly extract the group of underlying sources while maximizing the correlations among different datasets and retaining the intersubject source variability. Then the source signals are extracted and the potential heart beat signal is selected.

We evaluate the effects of the noise size n and the number of interesting IMFs l. Increasing the size of the independent noise in the NAMEMD generally helps enforce the desired quasi-dyadic structure and reduces the degree of the overlap between the IMFs' spectra. Intuitively a larger noise size n is preferred. However, the computational cost also increases if we add more channels of noise. In our preliminary study, we investigated the performances when the noise size n varies from 1 to 10 and observed that the estimated RHBR is not sensitive to n. We therefore heuristically set n to be 1 for computational efficiency. We also estimated the RHBR when l varies from 3 to 6, and obtained the best performance when we select 4 IMFs for each channel. Therefore, we heuristically set l to be 4 in this chapter when analyzing the nanofiber-based strain sensor signals.

Algorithm 2 The proposed NAMEMD-MCCA algorithm

Input: a m-dimensional signal
$X(t) = [x_1(t), x_2(t), ..., x_m(t)]^T$, the size of the noise n and the number of interesting IMFs for each channel signal l
Output: The extracted k source components for each channel

1: Generate n-channel multivariate independent white noise and combine with the original signal $X(t)$ to obtain a *(m+n)*-channel signal $X^{'}(t)$;
2: Decompose the noise-combined signal $X^{'}(t)$ into IMFs based on the MEMD method and obtain multivariate IMFs;
3: Discard the IMFs corresponding to the n-channel white noise and obtain m set of IMFs for the original signal $X(t)$;
4: Calculate the frequency spectral for each IMF and select l interesting IMFs for each channel based on their dominant frequencies. These IMFs should have the common scale and the same index among channels. Then, combine these IMFs from each channel to be a single dataset and then get m datasets, each of which contains l channel signals;
5: Apply the MCCA method to these m datasets and extract the l source signals for each dataset.

Heart beat rate estimation

We aim to retrieve the heart beat rate from the obtained heart beat source signal, which can be obtained by different BSS methods mentioned above, including EEMD-ICA, EEMD-MCCA and the proposed NAMEMD-MCCA. The heart beat rate estimation method [123] can be divided into 3 steps, including peak detection, peak filtering and calculating the mean of the peak intervals. The details of the detection method are shown in Algorithm 3. We calculate the derivate of the extracted heart beat signal and identify the local extrema. Considering the shape of the typical heart beat signal, the peaks should be larger than the surrounding extrema to some extent. In this work, we measure the difference between some peaks and the corresponding closest valleys and finally set $(2^{nd}\max - 2^{nd}\min)/4$ as the threshold to detect the potential peaks. In addition, we filter the fake peaks and

Algorithm 3 The heart beat rate estimation method

Input: heart beat signal $s_i(t)$ and sampling frequency fs
Output: heart beat rate HBR

1: Peak Detection:
i) calculate the derivate of $s_i(t)$ and detect the local maxima and minima;
ii) considering typical heart beat signals, only the maxima satisfying the following specific condition can be regarded as a peak,

$$\text{Peak} \geqslant \text{Nearest minima} + \frac{2^{nd}\max - 2^{nd}\min}{4},$$

in which the nearest minima represents the left and the right minimum which is closest to the potential peak, the 2^{nd}max and 2^{nd}min mean the second largest and the second smallest magnitude of the heart beat signal.

2: Peak filtering, including:
i) exclude the endpoints from the potential peaks;
ii) exclude the fake peaks which are quite close to each other. The threshold we set here is 800 while the sampling frequency is 1600Hz. This implies that the HBR is no more than 120 bpm.

3: Calculate the mean of the peak intervals:
i) calculate the peak interval and discard the intervals which are larger than 1.2fs, which corresponds to 50 bpm;
ii) calculate the mean of the effective peak interval (MPI) as HBR:

$$\text{HBR} = \frac{fs}{\text{MPI}} \times 60$$

invalid peak intervals according to the normal RHBR ranging from 50-120 bpm. Then we get the estimate of the RHBR based on the mean of the effective peak intervals.

2.2.3 Data Description

Synthetic Data

In this simulation, we apply the proposed NAMEMD-MCCA method to synthetic data and also implement EEMD-MCCA and EEMD-ICA for comparison. In order to simulate the real physiological signal, we generate *m* groups of sources using real PPG from [55], EMG from [52], Motion signals from [82] as well as power frequency signal and Gaussian white noise simulated by MATLAB (version R2014a). The clean PPG signal was obtained from pulse oximeter [55] and the EMG signal was monitored by using VEST(Capintec) [52]. The motion signal, which is usually the result of intermittent mechanical forces, was obtained by using a Holter recorder on an active subject [82]. Considering the difference of their sampling frequencies, we scale them into 300Hz. Each group contains the

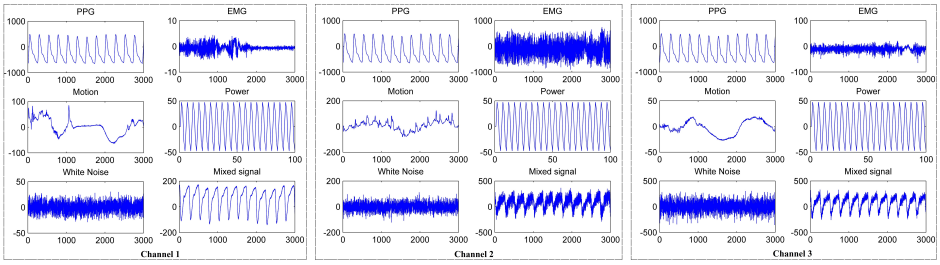

Figure 2.1: Three groups of source signals and mixed signals. In each subfigure, the horizontal axis represents the time index and the vertical axis represents the amplitude of the signal in microvolt.

following 5 sources:

$$\begin{aligned} s_1 &= \text{PPG signal} \\ s_2 &= \text{EMG signal} \\ s_3 &= \text{Motion signal} \\ s_4 &= \text{Power frequency signal (60Hz)} \\ s_5 &= \text{Gaussian white noise} \end{aligned} \tag{2.4}$$

There is 0.01 second time-shift between the PPG signals in successive groups. Considering the different locations of sensors in reality, we make EMG signals and the Motion signals distinct among groups, as shown in Fig. 2.1. The signal recorded by each sensor is simulated as the linear mixture of the corresponding group of sources. m mixed signals are generated as follows,

$$X_i = A_i S_i, \quad i = 1, 2, \ldots, m \tag{2.5}$$

where the $S_i = [s_1; s_2; s_3; s_4; s_5]$ denotes the i^{th} group of sources and the A_i is the mixture matrix randomly generated for the i^{th} group of sources. An example with three channels of the mixed signals is shown in Fig. 2.1. For easy visualization, we just plot the first 100 points of the power frequency signal. As can be seen from Fig. 2.1, the mixed signals in channels 2 and 3 are much noisier than that of channel 1. The related mixture matrices are

$$A_1 = \begin{bmatrix} -0.2435 & 0.2262 & 0.3538 & -0.0721 & -0.1044 \end{bmatrix}, \tag{2.6}$$

$$A_2 = \begin{bmatrix} -0.2433 & -0.2397 & -0.0670 & -0.2155 & 0.2345 \end{bmatrix}, \tag{2.7}$$

$$A_3 = \begin{bmatrix} -0.3305 & -0.1377 & -0.0518 & 0.1027 & -0.3773 \end{bmatrix}. \tag{2.8}$$

Real Data

The fabrication of core-shell nanofiber (NF) mesh strain sensors was explained in detail in our previous paper [98]. In brief, Polyacrylonitrile (PAN) NFs were fabricated using a conventional electrospinning technique. PAN (average molecular weight of 100000 g/mol, Scientific Polymer Products) is dissolved in DMF (99.9%, Fisher Scientific) with concentration of 10 wt% and stirred at 60^oC for 24 h to form a homogenous solution. It is then loaded into a plastic syringe with a blunted G18 needle. An electrospinning unit (KATO TECH CO. LTD.) is used to prepare the nanofibrous samples from the solution. A constant volume of the solution is delivered to the needle at a flow rate of 0.6 ml/h and a high potential of 17 kV is applied to the needle. The non-woven fibers are collected on the collector (aluminum foil, 17 cm far from the needle) connected to the ground to form NF network. The nanofiber mesh is then coated with a thin layer of gold by sputtering, which provides a uniform conformal coating on the surface of NFs. Ribbons of coated NFs with desired dimension of about 2 mm width and 10-50 mm length is then cut and transferred to the surface of a desired elastomer substrate, Polydimethylsiloxane (PDMS), with the thickness of $\approx$1 mm. Electrical contact pads and wirings are prepared by application of silver pint to both ends of the NF ribbon. A PDMS layer is then poured on the top of the NF mesh, followed by degassing and curing process at 90^oC for 45 minutes. It was demonstrated that this novel and comfortable NF-based sensor is highly sensitive and provides exceptional performance for accurate health monitoring [98]. Fig. 2.2(a) and (b) shows the dimensions and the compliant nature of typical NF sensors. The small, soft and comfortable nature highlights the suitability of the NF sensor for wearable health monitoring.

Radial pulse sensing was performed by attaching NF-based sensors to epidermis on the wrist using a double-sided tape. The minute deformation of the skin

caused by radial pulse results in change in resistance of the sensors. Considering the low deformation level, to obtain strong signals, the strain sensor should be attached to a location close to the radial artery. Due to the anatomical variation of different subjects, the location of the radial artery slightly varies in different people. In order to cover a larger area and ensure that at least one sensor is located on the right position, we have used three sensors as depicted in Fig. 2.2 (c) and simultaneously recorded their responses. The measurement set-up is schematically shown in Fig. 2.3. Wires are connected to the end contact pads of the sensors and connected to a costume-made multichannel resistance measurement system, which was designed and developed in our lab. The multichannel resistance measurement system, consisting of an analog circuit, a Wheatstone bridge and an amplifier for each channel, provides the possibility of real-time measurement of multiple sensors simultaneously. It also includes a USB 1608FS-PLUS (Measurement Computing) to convert the analog signal of the circuit to its digital representation for real-time recording of the signals. A LabVIEW program was used for operating and recording the data.

The NF sensor signals were collected when the subject sat or stood steadily. In

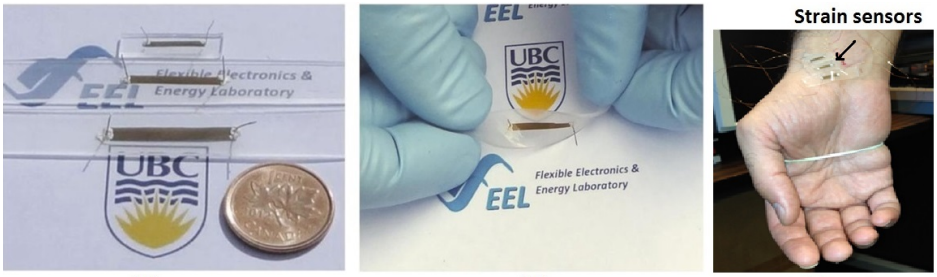

Figure 2.2: Photographs of NF sensors: (a) Dimensions of NF sensors; (b) Flexibility of NF sensors; (c) Three NF sensors used for monitoring the RHBR.

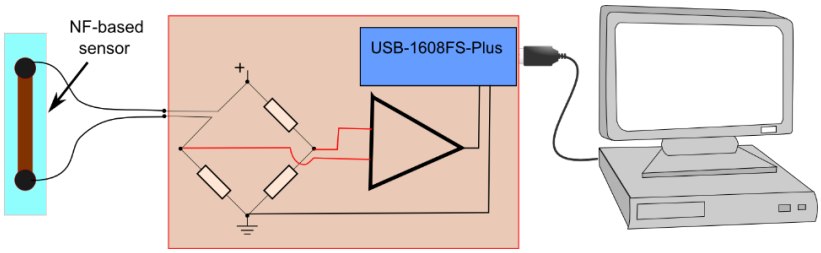

Figure 2.3: Illustration of the measurement system.

addition, to evaluate the robustness of the proposed algorithm, the sensor responses to pulse when the subject was doing the following 9 tasks were also collected. The time duration of each task is 12.5 seconds. The task of 'Fix type1' serves as an example in this chapter.

(1) Finger Flex and Wrist Flex: flexion of all fingers together and flexion of the wrist.

(2) Fix type key1: type key 1 using the index finger while the wrist and forearm was fixed.

(3) Fix type1 and Fix type2: type key 1 and 2 alternatively using the index finger while the forearm was fixed. We repeat this task twice.

(4) Fix type: randomly type while keeping the forearm and wrist fixed.

(5) Fix writing: randomly write using a pen with fixed forearm.

(6) Fix writing1 and Fix writing2: write the number 1 and 2 using a pen while the forearm is fixed respectively.

To better describe the NAMEMD-MCCA-based procedure for monitoring the RHBR from the NF sensor signals, we summarize the entire procedure in Algorithm 4. Since the signals of interest lie in the low frequency band, the original signals are preprocessed by low pass filtering (to 20Hz) and down sampling (by 5) for computational efficiency. Then the NAMEMD-MCCA method is applied to the preprocessed signals. 4 source components for each sensor are extracted by

Algorithm 4 The resting heart beat detection method based on NAMEMD-MCCA

Input: The raw 3-channel nanofiber signals $X(t) = [x_1(t), x_2(t), x_3(t)]^T$, the number of the multivariate independent white noises (n = 1 here), and the number of interesting IMFs for each channel signal (l = 4 here)

Output: The estimated resting heart beat rate RHBR

1: Preprocess the raw nanofiber sensor signals $x(t)$ by low-pass filtering (to 20Hz) and down sampling (with decimation factor = 5) to obtain $g(t)$;
2: Apply the proposed NAMEMD-MCCA algorithm to the preprocessed data $g(t)$ and extract the source signals $s(t)$;
3: Select the heart beat signal $s_i(t)$ from the extracted source signals $s(t)$;
4: Apply the proposed HBR estimation method, based on the peak detection, peak filtering and calculating the mean of the peak intervals, on the extracted heart beat signal. Then the RHBR is detected.

MCCA and the potential heart beat signal is selected from them. Finally, the HBR estimation method is applied to the potential heart beat signal and the RHBR is detected.

2.3 Experimental Results

In this section, the performance of two currently available EMD-BSS methods and the proposed method is evaluated on synthetic datasets. We then examine these three methods on real nano-sensor data to obtain the heart beat signal and estimate the corresponding RHBR.

2.3.1 Simulation Study

We test the performances of three methods, including EEMD-ICA, EEMD-MCCA and the proposed NAMEMD-MCCA, on synthetic data. An example of the synthetic data is shown in Fig.2.1, in which each channel of the mixed signals is simulated as a linear mixture of the corresponding 5 sources. As seen from Fig. 2.1, the signal in channel 2 and 3 are noisier than that in channel 1. It is challenging to

recover the pulse wave signal from these two channels. To this aim, we attempt to incorporate these three channel signals. We test the proposed NAMEMD-MCCA method as well as other two existing methods on this simulated data. The extracted pulse waves corresponding to channel 1 are shown in Fig. 2.4. Compared with the other two methods, NAMEMD-MCCA provides the best performance in preserving the shape of the original signal. The extracted signal by EEMD-ICA conveys more noise and it is nothing like the shape of pulse wave. Although we can get the heart beat from the result of EEMD-MCCA, it would be more difficult or even impossible to get more physical information from this signal.

We also utilize the correlation coefficient (CC) between the original PPG and the extracted pulse waves as an evaluation measure to the effect of these three methods. The CC corresponding to the NAMEMD-MCCA method is 0.978 which is higher than that of EEMD-ICA by 36% and that of EEMD-MCCA by 3%. For definiteness and without loss of generality, we generate 100 synthetic datasets and compare these three methods on them.

As shown in Fig. 2.5, NAMEMD-MCCA and EEMD-MCCA can obtain comparable performance in recovering the pulse wave and they generally performs better than EEMD-ICA, although EEMD-ICA performs better than the other two methods occasionally. As to NAMEMD-MCCA, the average of the absolute CCs is 0.902, whereas that of EEMD-ICA and EEMD-MCCA is just 0.623 and 0.873 respectively. In addition, it should be noted that the extracted source signals obtained by ICA are not unique. Even repeating the analysis on the same datasets, we may get different independent components [20]. In addition, EEMD does not consider the inter-relationship among the signals while MEMD does. Thus, NAMEMD-MCCA performs better than EEMD-MCCA on average.

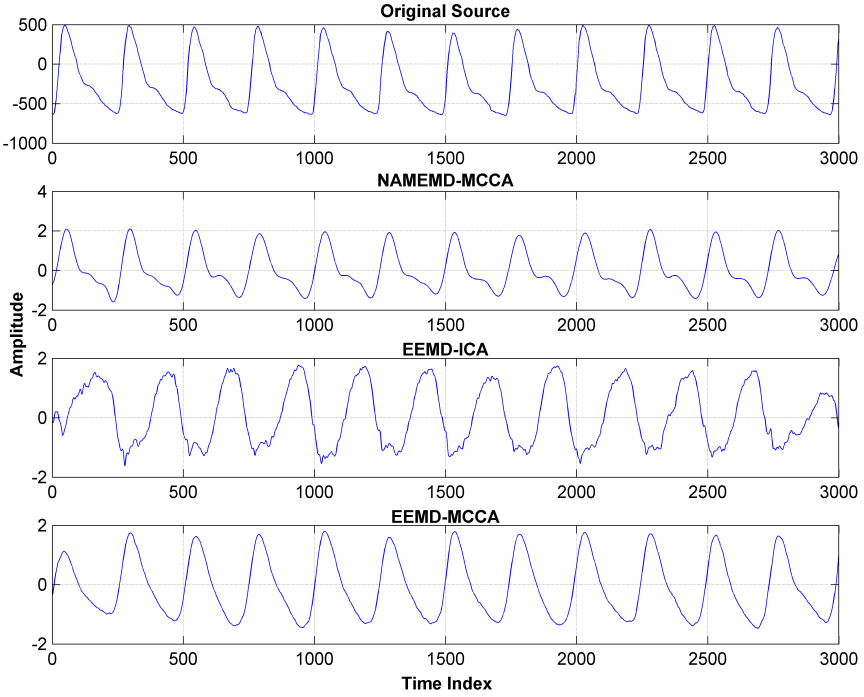

Figure 2.4: Extracted pulse waves based on the synthetic data.

2.3.2 Real Data Study

In this section, we apply the proposed novel NAMEMD-MCCA method to extract the heart beat signal from some noisy nanofiber based strain sensor signals. Then the HBR estimation method is used to detect the RHBR from the obtained heart beat signal.

Empirical mode decomposition of the signal

Before performing dedicated analysis, we preprocess the raw data, shown in the first row of Fig. 2.6, by low-pass filtering and down sampling. Then we decompose the preprocessed signal utilizing the NAMEMD method. For graphical presentation, we just include 4 IMF components for each sensor. The number of these zero-crossings is a rough indication of the mean frequency of each IMF component

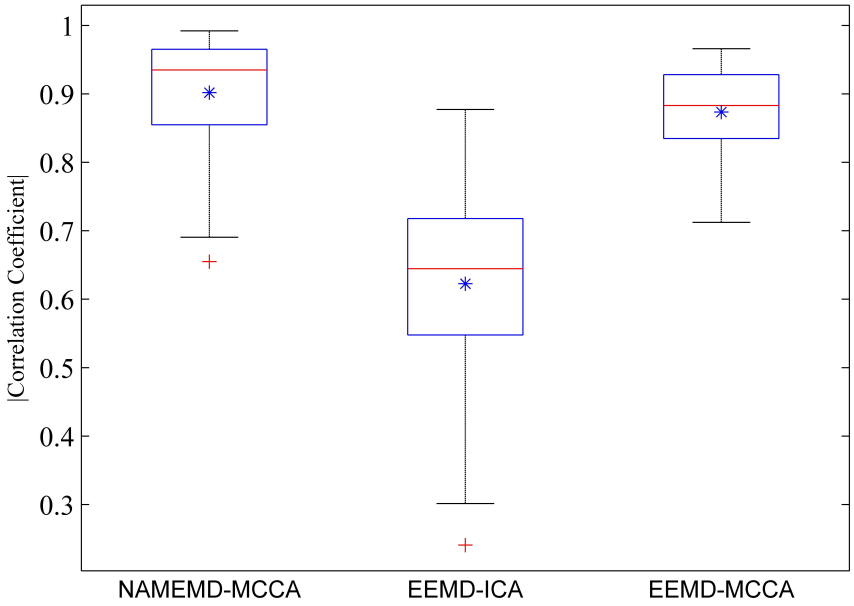

Figure 2.5: The absolute correlation coefficients between the original PPG signals and the extracted pulse waves on 100 synthetic datasets. The blue asterisks represent the averages and the red lines stand for the medians. The edges of the box are the lower and upper quartiles. Outliers beyond the upper whisker or under the lower whiskers are indicated by the red plus signs.

[42]. As shown in Fig. 2.6, the mean frequencies decrease with the component's indexes increase and it is a further indication of the hierarchical structure of the equivalent filter bank. It is suspected that the high frequency noise is appeared in the first several components whereas relative low frequency noise resides in the last several components. We eliminate those components whose frequency contents are irrelevant to that of the heat beat signal. It acts as the first step of denoising and leads to the major advantages of high efficiency in the next step.

The NAMEMD has the ability to align 'common scales' present within multivariate data [90] and it is the main difference with the classical channel-wise EMD methods. To highlight this, we also decompose the preprocessed signal using the

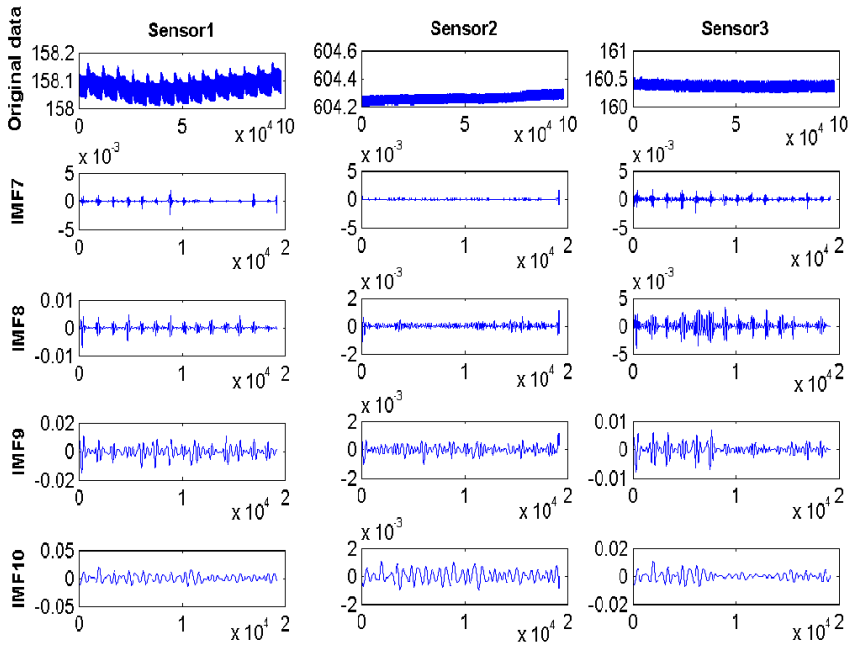

Figure 2.6: Time domain analysis of the NAMEMD results. In each subfigure, the horizontal axis represents the time index and the vertical axis represents the amplitude of the signal.

powerful EEMD method and compare its results with the results of NAMEMD in the frequency domain, shown in Fig.2.7 and Fig. 2.8. As can be seen from Fig.2.7, the common scales are shown in the IMF components with the same indexes. The alleviation of mode mixing mainly benefits from the use of inter-channel information and the Gaussian white noise which residues in a separated subspace from the useful data. Such mode alignment property helps to utilize the similar scales from multiple sources. However, for the EEMD method, the IMF components with the same index always have different scales, as shown in Fig. 2.8. Take the IMF7 as an example, the dominant frequency of IMF7 from sensor 1 is about 3.5 Hz, whereas that from sensor 2 and 3 is about 10 Hz and 7 Hz respectively. Therefore, it is arbitrary to select the IMF components with the same index from the EEMD results

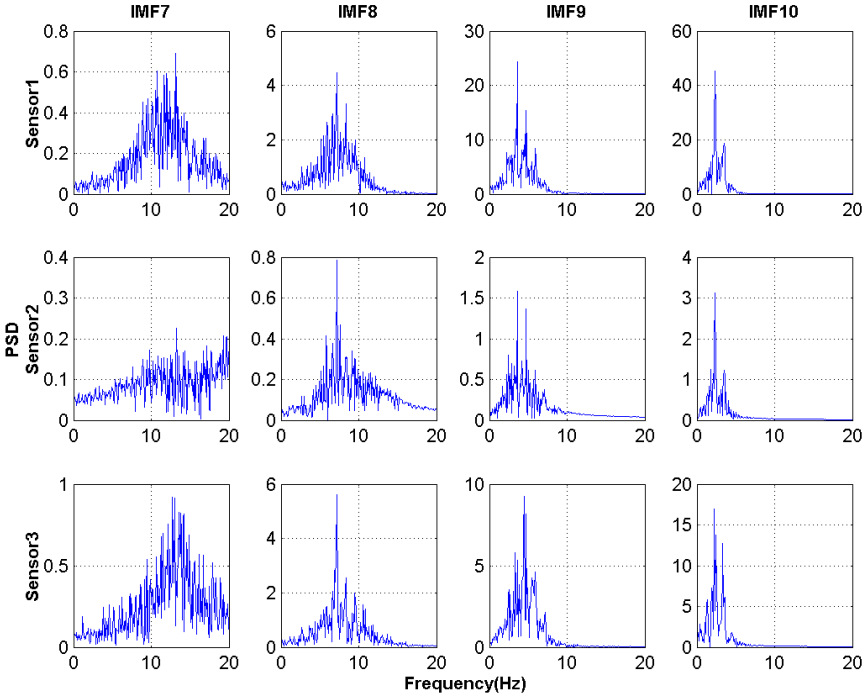

Figure 2.7: Frequency domain analysis of the NAMEMD results.

as the input of the subsequent analysis, such as BSS methods.

Another interesting finding is that we can observe some repetitive signals whose periods are similar to the heart beat signals in the decomposition results of sensor 2 and sensor 3 by NAMEMD, such as IMF10 from sensor 2 and IMF 7 from sensor 3 shown in Fig. 2.6, even though these two sensor signals are polluted by the noise heavily. However, we could not find that similarity in the decomposition results by EEMD. This maybe benefit from the inter-channel information considered by NAMEMD whereas EEMD is channel-wise method. In addition, it also suggests that we must put the sensor at the ideal location if we want to get the heart signal from only one sensor, otherwise it would be much more difficult or even impossible to obtain that signal.

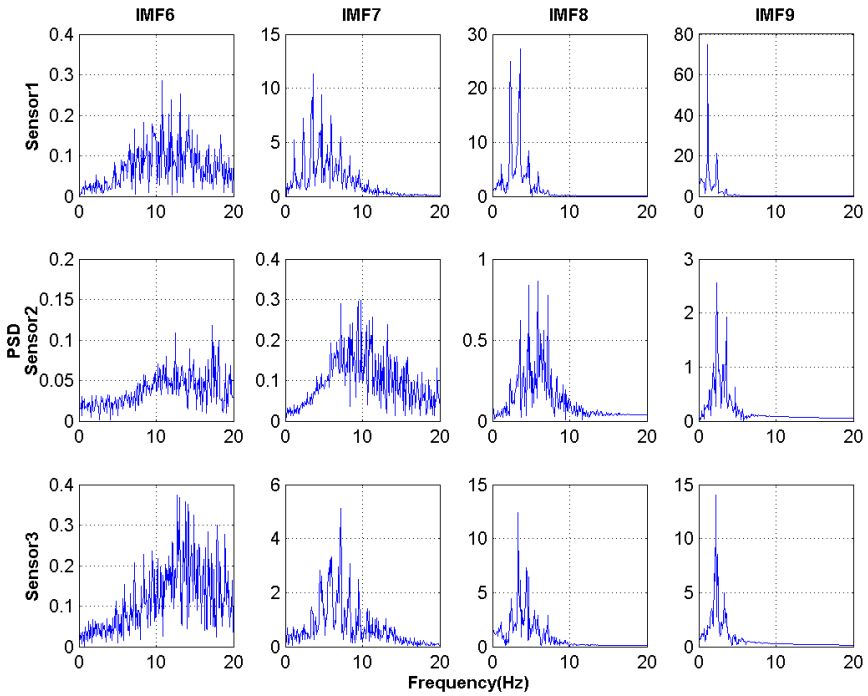

Figure 2.8: Frequency domain analysis of the EEMD results.

Extracted Heart Beat Signals Based on the MCCA

The artifacts components can be selected and removed from the obtained IMF components. Then 12 IMF components (4 from each sensor) are used to be the inputs of the blind source separation techniques, including MCCA and ICA. For the case of MCCA, those IMF components from 3 sensors form 3 datasets, acting as the input of the MCCA. However, for the case of ICA, we combine these 12 IMF components together and try to extract the heart beat signal from the noisy signals by the FastICA. Then the underlying source signals resembling the heart beat signals are selected and displayed. The separation results of NAMEMD-MCCA and some previous methods including EEMD-ICA and EEMD-MCCA are compared in Fig.2.9. Generally, these three techniques can recover the heart beat signals and the heart

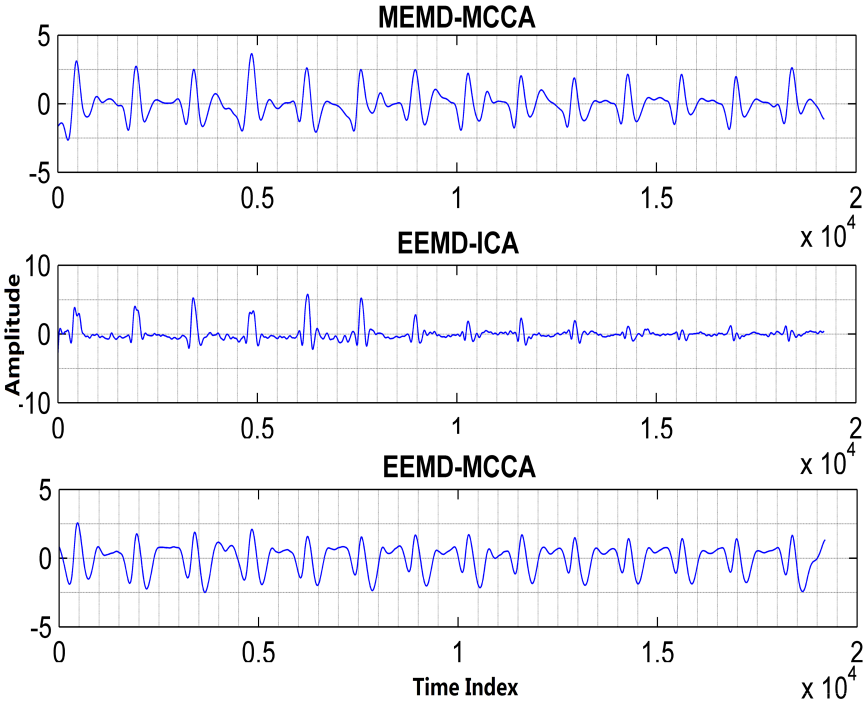

Figure 2.9: Extracted heart beat signals when employing three different methods.

beat rates can be retrieved from them. However, the results of EEMD-ICA and EEMD-MCCA seem to convey more noise. As we can see from the 2^{nd} subfigure of Fig.2.9, there are stark disparities between the magnitudes of different cycles, which may obscure the physical meaning of the heart beat signal. As to the result of EEMD-MCCA method, the gaps between peaks and base line are too small and this may increase the difficulty to detect the peaks sometimes. In contrast, the extracted signal of NAMEMD-MCCA highly resembles the heart beat signal. The similarity between them also suggests the effectiveness of the proposed two-step strategy to recover the underlying source signal.

Table 2.1: The estimated RHBR results when employing three different methods.

Task	NAMEMD-MCCA	EEMD-ICA	EEMD-MCCA
Finger Flex	67.6	68.5	65.4
Fix type	66.1	62.5	65.1
Fix type key1	70.7	60.3	60.9
Fix type1	69.6	69.2	69.7
Fix type2	68.2	71.1	68.3
Fix writing	69.9	68.7	71.3
Fix writing1	68.5	65.7	66.1
Fix writing2	63.6	64.6	63.8
Wrist Flex	70.9	72.1	71.4
Fix Sit	72.1	70.1	71.1
Fix Stand	70.8	70.1	68.8
AVG*	68.9	67.5	67.4
STD*	2.47	3.75	3.44

* AVG and STD represent the average and the standard deviation respectively.

Heart Beat Estimation

We apply the heart beat rate estimation method mentioned in Algorithm 3 to the extracted heart beat signals of 11 tasks and the estimated results are shown in Table 1. To evaluate the reliability of the results, we also asked an experienced physician to measure the pulse rate of the subject during each task manually, from which the average resting pulse rate of the subject derived to be 70 bpm. For the proposed method, the average of the RHBR results is closer to the ground truth and the related standard variation is 2.47, which is the smallest one among the three variations.

2.4 Discussions

In this chapter, we explore the potential of using novel nanofiber based strain sensors for monitoring heart beat signals. Three NF sensors are attached to different

locations on the wrist. These sensors can cover a large area to ensure getting good signals from the right position. Even though the acquired signals may be contaminated, it is still possible to recover the heart beat signal from multi-sensors by exploring the inter-channel information. The nano-sensor signals are recorded under different conditions. Similar to other types of biomedical measurement tools, detecting and extracting source signals from back ground noises is challenging for wearable technologies. The proposed NAMEMD-MCCA method provides superior performances for detecting RHBR from the measured nano-sensor data, as it mainly benefits from the appropriate combination of the NAMEMD and MCCA.

At the first step, the NAMEMD works as adaptive filter banks with scale alignment of channels and can overcome the limitations of classical channel-wise EMD methods, such as standard EMD and the EEMD which can only cater for univariate signals, when dealing with data from multiple sources. If multiple signals are processed separately, the EEMD fails to align the frequency responses from IMFs with same index of multichannel signals, making the subsequent analysis meaningless [90]. However, the NAMEMD method is able to accurately align the common oscillatory modes in corresponding IMFs from multivariate signals, and therefore providing an intuitive and effective way for the analysis of narrowband but non-stationary rhythms from the multichannel nanofiber based strain sensor signals. In addition, unlike the EEMD, the noise in NAMEMD is not added directly to the original signal. Instead it is kept in separate channels of a multivariate signal. In this way, the physically disjointed input signal and the added noise can alleviate the effects of noise interference, a unique property of NAMEMD. Compared with the other two existing algorithms including EEMD-ICA and EEMD-MCCA, the better performance of the proposed method is credited to the use of the inter-channel information, enhanced localization properties and increased robustness to artifacts of the NAMEMD method.

At the second step, the IMFs from each sensor whose dominant frequencies are close to that of typical heart beat signals are analyzed further by the following BSS methods. MCCA can be used to efficiently separate the source signals. Under certain assumptions, the group of corresponding sources from each dataset can be jointly extracted by MCCA through maximizing the correlations among the extracted sources [20, 72]. Yet, ICA always aims to find a linear projection of the multivariate signals that maximize the mutual independence. Comparing with ICA which is based on high order statistics (HOS), MCCA uses the second order statistics and thus yields higher computational efficiency. For the health monitoring system, people always desire a direct feedback, and thus ensuring a lower computational complexity is often of great importance. In addition, the ICA method always fails to converge which may render that some source signals still are mixed with noisy components. A few comparisons have shown that MCCA are more competitive in solving multivariate blind source separation problems [72].

After obtaining the heart beat signal, the HBR estimation method can be applied to detect the subject's RHBR. It should be noted that the selection of the heart beat signal from the separation results of BSS methods is based on visual inspection. In real-world application, a selection criteria or an automated selection method is needed. We are currently working on this issue based on some machine learning strategies.

2.5 Conclusions

The ability to accurately detect flexions and extensions is critical for wearable health monitoring. In this chapter, we propose performing radial pulse sensing using three nanofiber based strain sensors, and we aim to extract the heart beat signal and estimate the corresponding RHBR from the recorded NF-based Sensor data. Since the type of nanofiber-based sensors used in this chapter is novel

and its application to heart beat rate estimation is also new, we need to investigate which approach is more appropriate to analyze this new type of nano sensor signals for estimating the heart beat rate. In addition to investigating two state-of-the-art EMD-BSS based methods, to better address the artifact removal challenge for our specific application, a novel framework based on the NAMEMD and MCCA is proposed to extract the heart beat signal from the multi-channel NF-based sensor signals. It is shown that the proposed method provides the best performance, especially for preserving the shape of the heart beat signal.

This chapter is the first work during my PhD study. It was motivated by the particular application to extract the heart beat signal from the nanofiber sensor signals. We specifically designed this method for this application. However, it is still applicable to the general cases when: 1) the number of observations is smaller than that of sources; 2) there is cross correlation between each pair of observations.

Chapter 3

Underdetermined Joint Blind Source Separation for 2 datasets based on Tensor Decomposition

In this chapter, we aim to jointly separate the underdetermined mixtures of latent sources from two datasets, where the number of sources exceeds the number of observations in each dataset. Currently available BSS methods, including JBSS and UBSS, cannot address this underdetermined problem effectively. We exploit the second-order statistics of observations and introduce a novel blind source separation method, termed as UJBSS-2. Considering the dependence information between two datasets, the problem of jointly estimating the mixing matrices is tackled via CPD of a specialized tensor in which a set of spatial covariance matrices are stacked. Furthermore, the estimated mixing matrices are used to recover the sources from each dataset separately. Numerical results demonstrate the competitive performance of the proposed method when compared to a commonly used JBSS method, multiset canonical correlation analysis (MCCA), and the single-set

UBSS method, UBSS-FAS.

3.1 Motivation and Objectives

With the advances of sensor technologies, we now have access to a large amount of multiset or multimodal data which needs to be jointly analyzed to extract the latent and physiologically meaningful components [5, 17, 121, 122]. Hence, JBSS algorithms which can jointly retrieve the source components are of great interest. JBSS exploits the dependence information across datasets and generally could yield better performances than that of single-set blind source separation (BSS) methods applied to each dataset separately [69]. JBSS also can keep the extracted components aligned across different datasets, an important feature that is not provided by single-set BSS methods.

Numerous models have been introduced to generalize the idea of single-set BSS to JBSS. For example, ICA has been extended to handle multiple datasets [5]. Group ICA and joint ICA attempt to concatenate multiple datasets into one dataset in the vertical and horizontal dimension respectively. Then the standard ICA can be performed on the concatenated single dataset. Independent Vector Analysis (IVA) generalizes ICA to multiple datasets by exploring the statistical dependence across datasets [5]. In addition, CCA has been popular to analyze relationships between two sets of variables [56]. It seeks a linear transformation of the observations such that the obtained corresponding source components across two datasets are maximally correlated. A generalization of CCA from two datasets to multiple datasets, MCCA, is shown to be flexible and powerful for discovering associations across multiple datasets [72]. Another recent extension of CCA is the joint diagonalization of many cross-cumulate matrices [69], which is especially effective when there is no explicit source distribution known in advance [69] .

It is worth noting that the above mentioned JBSS algorithms were originally

proposed for the *determined case*, since they generally assume that the number of sources is equal to or less than that of observations. This assumption may not be true in some practical applications, due to concerns such as the cost or time issues [60]. However, to our best knowledge, in the current literature there is no JBSS method specifically designed for the *underdetermined case* (i.e., the number of sources is greater than that of observations), even though there are some single-set underdetermined BSS (UBSS) methods [31, 62, 116] which can be used to separate the mixtures from each dataset separately. To fill this gap, in this chapter, we plan to extend the idea of JBSS to the underdetermined case for two datasets.

More specifically, inspired by the CCA model and the simultaneous matrix diagonalization [31], we exploit the second-order statistics of the observations in two datasets and propose a novel BSS method, referred to as the UJBSS-2. It is based on tensor decomposition which is attractive for our problem due to its ability to estimate the mixing matrix in UBSS for a single dataset[122]. Unlike the traditional (over)determined JBSS methods, the proposed UJBSS consists of two steps. Firstly, the mixing matrices are jointly estimated through a specialized tensor decomposition of the set of spatial covariance matrices of the observations. This step is the main emphasis of this chapter. Then the estimated mixing matrices are used to recover the sources from each dataset. In this work, we employ a novel time-frequency (TF) analysis method [116] to recover the sources.

3.2 Problem Formulation and Proposed Method

3.2.1 Signal Generation Model and Problem Statement

The problem of interest here is the underdetermined JBSS problem for 2 datasets. For the case of UJBSS for 2 datasets, the observed vector of each dataset contains the linear mixtures of the corresponding N sources. We can model the mixing

process as follows,

$$X^{[k]} = A^{[k]}S^{[k]} + e^{[k]}, \quad k = 1,2 \tag{3.1}$$

where $X^{[k]} = [x_1^{[k]}, x_2^{[k]}, \ldots, x_M^{[k]}]^T$ represents the M-dimensional observations, $s^{[k]} = [s_1^{[k]}, s_2^{[k]}, \ldots, s_N^{[k]}]^T$ means the underlying N-dimensional sources which are assumed to be mutually uncorrelated, $A^{[k]} \in R^{M\times N}$ with $M < N$ (i.e., the underdetermined case) represents the unknown mixing matrix for dataset k, $E^{[k]}$ means the possible additive noise which is generally assumed to be zero mean, temporally white and uncorrelated with the source signals.

The problem under consideration is to estimate the two mixing matrices $A^{[1]}$ and $A^{[2]}$ jointly up to permutation and scaling. In order to demonstrate the performance of the proposed mixing matrices estimation method, we also implement an approach for further source recovering based on the estimated $A^{[1]}$ and $A^{[2]}$.

3.2.2 Mixing Matrices Estimation

We have the following assumptions concerning the sources:

(1) The sources are uncorrelated within each dataset.

$$\begin{aligned} E\{s_i^{[k]}(t)(s_j^{[k]}(t+\tau))^H\} = 0 \\ \forall \tau, \quad 1 \le i \ne j \le N, k = 1,2, \end{aligned} \tag{I}$$

where τ represents the time delay and $\cdot^H$ denotes the complex conjugate transpose.

(2) The corresponding sources from two different datasets have non-zero correlations and the sources with different indices across datasets are not correlated.

$$\begin{aligned} D^{[\tau]} =& E\{S^{[1]}(t)(S^{[2]}(t+\tau))^H\} \\ =& diag[\rho_1(\tau), ..., \rho_N(\tau)]. \end{aligned} \tag{II}$$

Considering the correlations within and between the two datasets, we have the

spatial covariance matrices of the observations with delays, e.g., $C^{[l]}$, satisfy,

$$\begin{aligned} C^{[1]} &= E\{X^{[1]}(t)X^{[2]}(t+\tau_1)^H\} = (A^{[1]})D^{[\tau_1]}(A^{[2]})^H, \\ C^{[2]} &= E\{X^{[1]}(t)X^{[2]}(t+\tau_2)^H\} = (A^{[1]})D^{[\tau_2]}(A^{[2]})^H, \\ &\cdots \\ C^{[L]} &= E\{X^{[1]}(t)X^{[2]}(t+\tau_L)^H\} = (A^{[1]})D^{[\tau_L]}(A^{[2]})^H, \end{aligned} \tag{3.2}$$

in which τ_l means the time delay and the matrix $D^{[\tau_l]} = E\{S^{[1]}(t)S^{[2]}(t+\tau_l)^H\}$ is diagonal, for $l = 1,2,...,L$.

We stack the matrices $C^{[1]}, C^{[2]}, ..., C^{[L]}$ in a tensor $\mathscr{C} \in R^{M\times M\times L}$ as follows: $(\mathscr{C})_{i,j,l} = (C^{[l]})_{i,j}$, i = 1,2,...,M, j = 1,2,...,M, l = 1,2,...,L. We define the matrix D of size $L \times N$ with the element $D_{l,n} = (D^{[\tau_l]})_{n,n}$, for l = 1,2,...,L, n = 1,2,...,N. Then we can represent $\mathscr{C}$ as:

$$\mathscr{C} = \sum_{n=1}^{N} a_n^{[1]} \circ a_n^{[2]} \circ d_n, \tag{3.3}$$

in which the symbol $\circ$ denotes the tensor outer product operation, and $a_n^{[1]}$ and $a_n^{[2]}$ are the n^{th} column of the mixing matrices $A^{[1]}$ and $A^{[2]}$ respectively. With this tensor format, now the problem of estimating the mixing matrices $A^{[1]}$ and $A^{[2]}$ can be reformulated as a problem of CPD, which aims to decompose a higher-order tensor as a linear combination of a minimal number of rank-one tensors [61, 120]. In (3.3), each rank-one tensor is associated with a single source. Therefore, the mixing matrices $A^{[1]}$ and $A^{[2]}$ can be estimated via the unique CPD of $\mathscr{C}$. Actually, the power of CPD mainly stems from its uniqueness property. The uniqueness means that the decomposition is the only possible combination of rank-one tensors which sum to the objective tensor with the exception of permutation and scaling. The uniqueness is based on the rank of tensors. De Lathauwer et al. have studied different methods to determine the rank of a tensor and concluded that the de-

composition of a three-order tensor is unique [30, 61] if the number of sources N satisfies

$$N \leq L \quad and \quad N(N-1) \leq M^2(M-1)^2/2. \tag{3.4}$$

The CPD of the tensor can be calculated by minimizing the Frobenius norm of the difference between the original tensor and its estimated results using the Alternating Least Squares (ALS) algorithm due to its programming simplicity and popularity. In this chapter, the Enhanced Line Search (ELS) is used to enhance the convergence of the ALS [31, 88]. It should be mentioned that the choices of τ_1, τ_2, ..., τ_L may affect the estimation precision of the mixing matrices. Here, we heuristically choose the time delay as $\tau_l \in [0, 60]$ samples. If τ_l is too large, the correlation between two related sources with delay will be close to 0 and then the matrix D might be ill conditioned. It is desired to select τ_1, τ_2, ..., τ_L such that D is well conditioned. In addition, if the time delay τ is too large, the covariance matrix of the sources in two datasets (i.e., $D^{[\tau]}$) will be close to a null matrix and thus the assumption (II) may not hold.

3.2.3 Source Extraction Based on The Estimated Mixing Matrices

A complete JBSS approach consists of both mixing matrix estimation and source extraction, even though our main focus in this chapter is the estimation of mixing matrices. In order to demonstrate the performance of the proposed mixing matrices estimation method, we adopt a recently-developed TF analysis method [116] to recover the latent sources based on the estimated mixing matrices. For any underdetermined and instantaneous mixing model, priori knowledge is essential to restore the latent sources. In this chapter, we assume that the number of sources satisfies $N \leq 2M-1$. WVD [12, 112] is firstly applied to the observations and sparser representation in the TF domain can be achieved. For any time series sig-

nal with finite-energy, such as $y_1(t)$ and $y_2(t)$, the auto WVD is defined as

$$W_{y_1}(t,f) = \int_{-\infty}^{\infty} y_1(t+\tau/2)y_1^*(t-\tau/2)e^{-j2\pi f\tau}d\tau \tag{3.5}$$

and the cross WVD between the two signals is defined as

$$W_{y_1,y_2}(t,f) = \int_{-\infty}^{\infty} y_1(t+\tau/2)y_2^*(t-\tau/2)e^{-j2\pi f\tau}d\tau. \tag{3.6}$$

For the linear mixing model shown in (3.1), we can have

$$W_{x^{[k]}}(t,f) = A^{[k]}W_{s^{[k]}}(t,f)(A^{[k]})^H, \quad k = 1,2. \tag{3.7}$$

In order to extract the auto WVDs of the sources from the known mixtures, it is vital to identify the auto-term TF points where $W_{s_n^{[k]}}$ shows energy concentration. The high TF concentration property of WVD and the sparsity of the sources guarantee that the auto-term TF points of the sources are dominant among all the TF points [116]. Therefore, we can collect enough information and then recover the sources based on the WVD values of all auto-term TF points. According to the definition of auto-term TF points, the spatial time frequency distribution (STFD) matrix of the sources at any auto-term TF point (t_a, f_a) is diagonal. The WVD value of the sources α at auto-term point (t_a, f_a) can be written as $[W_{s_1^{[k]},s_1^{[k]}}(t_a,f_a), W_{s_2^{[k]},s_2^{[k]}}(t_a,f_a), \ldots, W_{s_N^{[k]},s_N^{[k]}}(t_a,f_a)]^T$. Here α satisfies

$$(A^{[k]} \odot A^{[k]})\alpha = rvec(W_{X^{[k]}}(t_a,f_a)), \tag{3.8}$$

where $\odot$ represents the Khatri-Rao product and *rvec* means the vectorization of a matrix. The auto WVD value of the sources at each auto-term TF point, denoted as α, can be uniquely determined based on (3.8) under the following assumptions:

any M columns of the mixing matrix $A^{[k]}$ are linearly independent; the number of sources satisfies the condition $N \leq 2M-1$ as in [116], which is more stringent than (3.4). Here, we adopt an auto-term TF points searching strategy proposed in [116] and reconstruct the sources in the time domain based on the auto WVD values at all auto-term TF points.

The major steps of the proposed UJBSS algorithm are summarized in Algorithm 5.

Algorithm 5 The UJBSS algorithm for two datasets

Input: M-dimensional observations $X^{[1]}$ and $X^{[2]}$.
Output: the estimated mixing matrices $A^{[1]}$, $A^{[2]}$ and the recovered N-dimensional sources $s^{[1]}$, $s^{[2]}$.

1: Construct a three-order tensor $\mathscr{C}$ as in (3.2);
2: Calculate the CPD with N components for the tensor $\mathscr{C}$ and estimate the mixing matrices $A^{[1]}$ and $A^{[2]}$ as in (3.3);
3: Calculate the WVD values of the observations $W_{X^{[1]}}(t,f)$ and $W_{X^{[2]}}(t,f)$ as in (3.5) and (3.6);
4: Calculate the auto WVD values of the sources at auto-term TF points $W_{S^{[1]}}(t,f)$ and $W_{S^{[2]}}(t,f)$ based on the estimated mixing matrices as in (3.8);
5: Recover the sources in the time domain.

3.3 Numerical Study

In this section, we present two simulation studies, where different types of sources are considered, to demonstrate the separation performance of the proposed UJBSS method for two datasets. Two performance indices are used to evaluate the performance of the proposed method. One is the estimation error, defined as:

$$Error = 10log_{10}\{mean(\frac{||A-\hat{A}||}{||A||})\} \tag{3.9}$$

where $\hat{A}$ denotes the optimally ordered estimate of A. The other one measures the Pearson product-moment correlation coefficient (PPMCC) between the estimated sources and the original sources, which is defined as

$$PPMCC = \frac{cov(s_n, \hat{s}_n)}{\sigma_{s_n}\sigma_{\hat{s}_n}} \tag{3.10}$$

where $\hat{s}_n$ means the estimate of the source s_n, cov means the covariance between two variables and σ means the standard deviation. In order to ensure the dependence between the sources of the two datasets, the sources are synthesized as follows,

$$\begin{aligned} S^{[1]} =& [s_1^{[1]}, s_2^{[1]}, \ldots, s_N^{[1]}]^T; \\ S^{[2]} =& [s_1^{[2]}, s_2^{[2]}, \ldots, s_N^{[2]}]^T \\ =& S^{[1]} .* (unifrnd(0,1,S^{[1]}) \end{aligned} \tag{3.11}$$

where $unifrnd(0,1,S^{[1]})$ generates a vector with the same size of $S^{[1]}$ and each element of the vector is randomly drawn from the continuous uniform distribution on the interval (0,1). The average correlation between the source $s_n^{[1]}$ in the first dataset and the corresponding source $s_n^{[2]}$ in the second dataset is about 0.85, which can be regarded as highly correlated.

3.3.1 Simulation 1: Audio Signals

The sources used in this simulation include 8 audio signals, such as a piece of sound from the cable news network (CNN) and a piece of sound of an anonymous singer. The mixing matrices are generated randomly with elements following the uniform distribution $U[-1,1]$. For simplicity, each column of the mixing matrices is normalized into a unit vector. In our first setting, 6 sources are mixed into 4 observations in each dataset and the corresponding sources in the two datasets are correlated. With different signal-to-noise ratios (SNRs), we compare the proposed

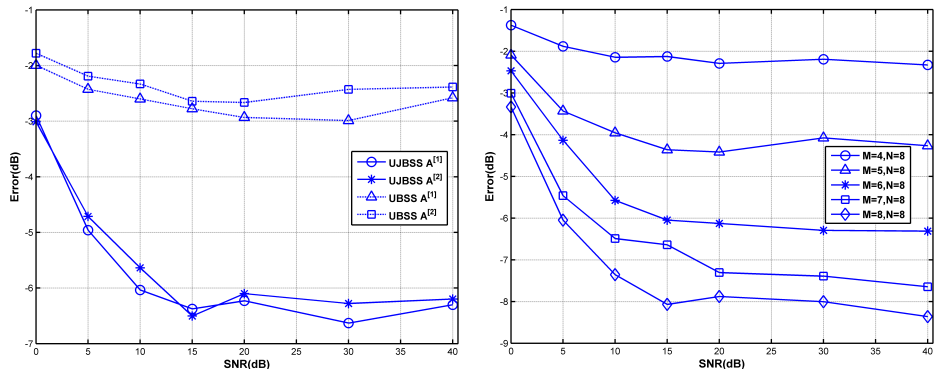

Figure 3.1: Performance of the proposed UJBSS method. (a) Performance comparisons between the proposed UJBSS method and the single-set SOBIUM method. Here the number of sources $N = 6$ and the number of observations $M = 4$. (b) Estimation error of $A^{[1]}$ when employing the proposed UJBSS method. Here the number of sources $N = 8$ and the number of observations M varies from 4 to 8. Similar results are observed for $A^{[2]}$.

UJBSS method with a commonly-used single-set UBSS method, SOBIUM [31], when it is applied to each dataset separately. We repeat the simulation 500 times and the performance is shown in Fig.3.1(a). Benefiting from the dependence information between two different datasets, the proposed UJBSS can provide more accurate estimation of the two mixing matrices, while SOBIUM neglects the possible inter-dataset information. We also note that the *Error* measure from either the UJBSS or the UBSS decreases with the increase of the SNR.

We also test the performance of the proposed method as the number of the observations increases from 4 to 8, while the number of sources is set to be 8. As noted in Fig.3.1(b), the estimation performance is getting better when more observations are available. Furthermore, we also test the performance when the condition (3.4) is not satisfied. For instance, with $M = 3$ and $N = 8$, as expected, the proposed method cannot estimate the mixing matrices correctly. We observe that the *Error* can be close to 0dB even when the SNR is larger than 40dB.

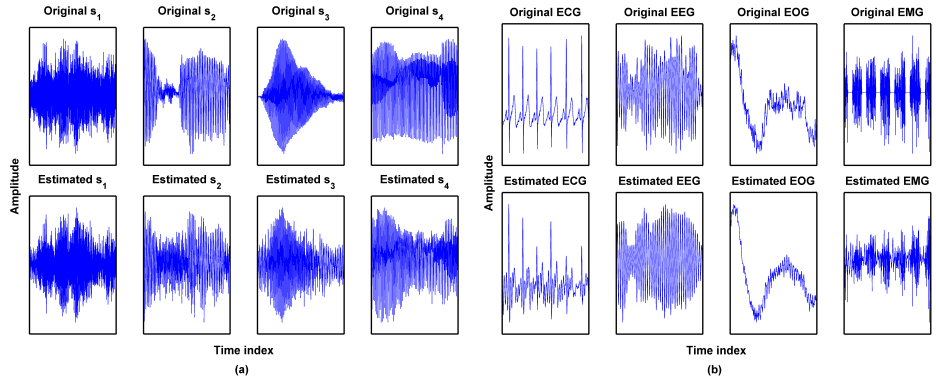

Figure 3.2: The original sources and extracted sources from the first dataset. (a) Simulation 1. (b) Simulation 2. The top subfigures are the original sources and the bottom ones are the extracted sources. Similar results are observed for the second dataset.

We recover the latent sources from each dataset based on the estimated mixing matrices. In an illustrative example, we linearly mix 4 sources into 3 observations in each dataset. Fig.3.2(a) shows the separation results of the first dataset in the time domain. The top four subfigures of Fig.3.2(a) represent the original sources and the bottom four subfigures are the recovered sources from the proposed UJBSS method. Considering the fact that SOBIUM only can estimate the mixing matrices, we further compare the proposed method with a single-set UBSS method, the UBSS-FAS [116], and the JBSS method MCCA [72], in term of the PPMCC between the original sources and the recovered sources. The performance results of these three methods are reported in Table 3.1. Although adopting the same technology in extracting sources, the performance of the proposed method is significantly better than that of the single-set UBSS-FAS method. This observation supports the importance of estimating mixing matrices accurately. MCCA, which has been successfully used in many fields [27], assumes that the number of sources is equal to the number of observations in each dataset and it could not be used to sepa-

rate sources in the underdetermined case directly. We add one observation in each dataset so that MCCA can be applied. Therefore it is not really a fair setting and comparison to the proposed method. However, we note that the performance of MCCA is not as good as that of the proposed method, even with an additional observation signal. The following reasons could contribute to the worse performance of MCCA: it is mainly due to the fact that the correlation coefficients between sources in two datasets are quite close [20, 72]; the performance of MCCA may suffer from the error accumulation of the deflation-based separation methods [69].

Table 3.1: PPMCC performance results in Simulation 1.

	Methods	s_1	s_2	s_3	s_4
Dataset 1	UJBSS	0.935	-0.809	0.840	0.924
	UBSS-FAS[116]	-0.478	0.892	-0.224	0.943
	MCCA*[72]	0.717	0.622	0.774	0.649
Dataset 2	UJBSS	0.866	-0.913	0.919	0.727
	UBSS-FAS[116]	-0.384	-0.867	-0.003	-0.629
	MCCA*[72]	0.711	-0.623	0.767	0.650

*We add one additional observation in each dataset when we evaluate the MCCA.

3.3.2 Simulation 2: Physiological Signals

In this experiment, we employ four physiological signals as sources, including electrocardiogram (ECG), EEG, electrooculography (EOG) and EMG from a publicly available database [45]. The sources corresponding to the second dataset are generated following (3.11). We get similar result as in the first simulation. Here, we just show the performances in term of the estimated sources in the time domain and the PPMCC results between the original sources and the estimated ones. Fig.3.2(b) shows the separation results of the four physiological signals. Although they still seem noisy, the estimated sources from the proposed UJBSS are highly

correlated with the original ones. As shown in Table 3.2, the proposed method yields promising results when it is used to separate the latent and underdetermined mixtures from two datasets. Compared to the classical JBSS method MCCA, the proposed UJBSS approach needs fewer number of observations in each dataset, while it still can provide a competitive performance.

Table 3.2: PPMCC performance results in Simulation 2.

	Methods	ECG	EEG	EOG	EMG
Dataset 1	UJBSS	0.778	-0.920	0.950	0.807
	UBSS-FAS[116]	-0.544	-0.908	0.222	0.450
	MCCA*[72]	0.752	0.885	-0.798	-0.723
Dataset 2	UJBSS	0.793	0.874	0.857	0.784
	UBSS-FAS[116]	-0.494	-0.751	0.393	-0.534
	MCCA*[72]	0.761	-0.885	-0.807	-0.714

*We add one additional observation in each dataset when we evaluate the MCCA.

3.4 Conclusion

In this work, we exploit the spatial covariance of the observations in two datasets and present a novel UJBSS method to jointly estimate the mixing matrices from two datasets when the number of observations is smaller than that of the sources (i.e., the underdetermined case). The mixing matrices are accurately estimated through CPD of a specialized tensor in which a set of covariance matrices are stacked. Further the sources are recovered based on the estimated mixing matrices. Numerical results demonstrate the competitive performances of the proposed method when compared to the commonly used JBSS method and the single-set UBSS method.

This Chapter represents the second work in my PhD study. To the best of our knowledge, there is no any existing JBSS method which can work in the underde-

termined case. Inspired by CCA, we explore the dependence information between two datasets and propose a novel strategy named as UJBSS for two datasets. We utilize the physiological signals, including audio signals and biomedical signals, as case studies. However, it is a general method for recovering the underlying sources from two datasets in the underdetermined case.

Chapter 4

Underdetermined Joint Blind Source Separation of Multiple Datasets

In this chapter, we tackle the problem of jointly separating instantaneous linear underdetermined mixtures of latent sources from multiple datasets, where the number of sources exceeds that of observations in each dataset. Currently available BSS methods, including JBSS and UBSS, cannot address this underdetermined problem effectively. We exploit second-order statistics of observations and present a novel blind source separation method, referred to as UJBSS-M, as a generalization of our previous work on two datasets [124]. In this work, the cross correlation between each pair of datasets is modeled by a third-order tensor in which a set of spatial covariance matrices corresponding to different time delays are stacked. Considering the latent common structure of these constructed tensors, the mixing matrices are jointly estimated via joint canonical polyadic decomposition of these specialized tensors. Furthermore, we recover the sources from each dataset separately based on

the estimated mixing matrices. Simulation results demonstrate that the proposed UJBSS-m method yields superior performances when compared to commonly used single-set UBSS and JBSS methods.

4.1 Motivation and Objectives

The increasing availability of multiset and multimodal signals has posed new challenges for conventional blind source separation (BSS) methods which are originally designed to analyze one data set at a time. There are many applications involving multiple datasets which have dependence relationships between them and need to be jointly analyzed [4, 44, 122], such as EEG, ECG, and MRI data. Hence, JBSS algorithms have attracted great interest in the fields of signal processing owing to their ability to simultaneously recover the underlying and physiologically meaningful components from multiple datasets. The crucial difference between BSS and JBSS is reflected by the fact that BSS only examines each dataset separately, whereas JBSS generalizes BSS to consider the dependence across multiple datasets [22]. Compared with conventional single-set BSS methods, JBSS generally could yield better performances. In addition, JBSS can keep the extracted components aligned across different datasets, an important feature that is not provided by single-set BSS methods.

The most original JBSS method was likely CCA, which has been popular to analyze relationships between two sets of variables [56]. It seeks a linear transformation of the observations such that the obtained corresponding source components across two datasets are maximally correlated. A generalization of CCA from two datasets to multiple datasets, MCCA was shown to be flexible and powerful for discovering associations across multiple datasets [72]. Another recent extension of CCA is the joint diagonalization of many cross-cumulate matrices [69], which is especially effective when there is no explicit source distribution known in advance

[69]. In addition to CCA-type algorithms, numerous models have been introduced to generalize the idea of single-set BSS to JBSS. For example, ICA has been extended to handle multiple datasets [5]. The group ICA and the joint ICA attempt to concatenate multiple datasets into one dataset in the vertical and horizontal dimension respectively, and then the standard ICA can be performed on the concatenated single dataset [5]. Based on the modular Bayesian framework, Groves et al. proposed a novel Linked ICA [48], encapsulating the ideas from both the group ICA and joint ICA. IVA generalizes ICA to multiple datasets by exploring statistical dependences across datasets [5].

It is worth noting that the above mentioned JBSS algorithms were originally proposed for the *determined case*, since they generally assume that the number of sources is equal to or less than that of the observations. This assumption may not be true in some practical applications, due to concerns such as the cost or time issues [60]. However, to our best knowledge, in the current literature there is only very limited work on JBSS methods specifically designed for the *underdetermined case* (i.e., the number of sources is greater than that of observations), even though there have been more single-set underdetermined BSS (UBSS) methods [31, 59, 62, 91, 115, 116, 119] which can be used to unmix the mixtures from each dataset separately. These UBSS methods can be divided mainly into two categories [104]. Most UBSS methods rely upon the sparsity of source signals in a specific domain, e.g., the time-frequency domain [91, 115, 116]. This category of methods usually require exhaustive computation, especially when the number of sources is large. Many algebraic methods were also proposed for unmixing the mixtures in the underdetermined case, most of which are based on decomposition of different data structures, e.g., covariance matrices[31].

In our previous paper [124], we proposed an underdetermined joint blind source separation method for two datasets (UJBSS-2) based on the decomposition of a

specialized tensor. However, it can only jointly estimate the mixing matrices from two datasets and cannot be extended directly to unmix the mixtures from multiple (larger than two) datasets [125]. To fill this gap in the literature, in this chapter, we plan to extend the idea of JBSS to the underdetermined case and generalize the idea of underdetermined joint blind source separation (UJBSS) for two datasets to that for multiple datasets.

More specifically, inspired by the MCCA model and the simultaneous diagonalization of covariance matrices [26, 31, 46], we exploit second-order statistics of the observations in each pair of datasets and propose a novel BSS method, termed as the underdetermined joint blind source separation for multiple datasets (UJBSS-m). Unlike the traditional (over)determined JBSS methods, the proposed UJBSS-m consists of two steps: 1) jointly estimate the mixing matrices from multiple datasets, and 2) recover the underlying sources individually based on each mixing matrix estimated in step 1). The most challenging task is to estimate the unknown mixing matrices precisely, which is the main concern of this chapter. In this work, this problem is tackled via joint canonical polyadic decomposition of specialized tensors. The dependence information between each pair of datasets is modeled by a third-order tensor where a set of spatial covariance matrices related to different time delays are stacked. Considering the possible combinations of two datasets, the pairs of the corresponding tensors share a common factor and then the mixing matrices (i.e., factor matrices of those tensors) can be jointly estimated by optimization-based methods. The estimated mixing matrices are further used to recover the sources from each dataset. In this work, we explore a novel subspace representation based method [116] to recover the sources.

Our main contributions are summarized as follows:

1) This chapter extends the idea of (over)determined JBSS to that of the underdetermined case.

2) Exploiting the cross correlation between each pair of datasets, we propose a novel and effective method to jointly estimate the mixing matrices for multiple datasets. More precise estimates of the mixing matrices can be achieved via the proposed UJBSS-m method compared to several classical single-set UBSS methods and JBSS methods.

3) The proposed UJBSS-m method can be used to solve single set UBSS problems and could achieve better performance in some cases, as demonstrated in the promising application of noise enhanced signal processing.

4) The proposed UJBSS-m method does not rely upon the sparsity of signals and therefore it can be applied to a wide class of signals, e.g., audio/speech and biomedical signals.

4.2 Notations and Preliminaries

In this chapter, we generally use the notation of [2], which was adapted by [57, 61]. A tensor can be interpreted as multi-index numerical array, whereby the order of a tensor is the dimensionality of the array. Scalars, denoted as lowercase letters, e.g., x, are said to be tensors of zero order. Vectors (first-order tensors) are denoted by boldface lowercase letters, e.g., $\boldsymbol{x}$. Matrices (second-order tensors) are denoted by boldface capital letters $\boldsymbol{X}$. Third-order or higher-order tensors are denoted by boldface Euler script letters, e.g., $\boldsymbol{\mathscr{X}}$. The transpose, inverse, Moore-Penrose pseudo inverse, norm are denoted by $(\cdot)^T$, $(\cdot)^{-1}$, $(\cdot)^\dagger$, $\|\cdot\|$.

The operation of matricization reorders the elements of a higher-order tensor into a matrix. For example, mode-n matricization of a Nth-order tensor $\boldsymbol{\mathscr{X}} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ yields a matrix $\boldsymbol{X}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times I_2 \times \cdots \times I_{n-1} \times I_{n+1} \times \cdots \times I_N)}$ whose columns are all mode-n fibers arranged in a specifically predefined order. In this chapter,

tensor element $x_{i_1,i_2,\ldots,i_N}$ corresponds to matrix elements $x_{(n)_{(i_n,j)}}$, where

$$j = 1 + \sum_{\substack{l=1 \\ l\neq n}}^{N} (i_l - 1)(\prod_{\substack{k=1 \\ k\neq n}}^{l-1} I_k). \tag{4.1}$$

The inner product of two same-sized tensors $\mathscr{X}, \mathscr{Y} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is the sum of the products of their elements, i.e.,

$$< \mathscr{X}, \mathscr{Y} >= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1,i_2,\ldots,i_N} y_{i_1,i_2,\ldots,i_N}. \tag{4.2}$$

The Frobenius norm of a tensor $\mathscr{X}$ is the square root of its inner product with itself, i.e.,

$$\| \mathscr{X} \| = \sqrt{< \mathscr{X}, \mathscr{X} >}. \tag{4.3}$$

The outer product of vectors $\{a^{(n)}\} \in \mathbb{R}^{I_n}, n = 1, 2, \ldots, N$ yields a rank-one tensor $\mathscr{X} = a^{(1)} \circ a^{(2)} \circ \ldots a^{(N)}$ with entries $x_{i_1,i_2,\ldots,i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \ldots a^{(N)}_{i_N}$, where the $\circ$ represents the outer product operation. The superscript in parentheses represents one element in a sequence, e.g., $a^{(n)}$ represents the nth vector in a sequence of vectors.

In order to demonstrate multi-way models, the usual matrix product, such as Kronecker product and Khatri-Rao product, is not sufficient. A frequently used operation is the mode-n product, denoted by $\times_n$. The mode-n product of a tensor $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ with a matrix $A \in \mathbb{R}^{J_n \times I_n}$ amounts to the product of all mode-n fibers with A and yields a tensor with the size of $(I_1 \times I_2 \times \cdots \times I_{n-1} \times J_n \times I_{n+1} \cdots \times I_N)$, whose entries are given by

$$\begin{aligned} &(\mathscr{X} \times_n A)_{i_1,i_2,\ldots,i_{n-1},j_n,i_{n+1}\ldots,i_N} \\ &= \sum_{i_n=1}^{I_n} x_{i_1,i_2,\ldots,i_{n-1},i_n,i_{n+1}\ldots i_N} a_{j_n,i_n}. \end{aligned} \tag{4.4}$$

The mode-n product of a tensor and a vector is a special case of the mode-n product of a tensor and a matrix with the size of $(1 \times I_n)$. Note that the order of the result is $(N-1)$, one less than the order of the original tensor. It is often useful to calculate the product of a tensor with a sequence of vectors. Let $\mathscr{X}$ denote a tensor with the size of $I_1 \times I_2 \times \cdots \times I_N$, and let $\{a^{(n)}\}$ $(n = 1,2,\ldots,N)$, be a sequence of vectors, each with the length of I_n. Then the product of $\mathscr{X}$ with a sequence of vectors in all modes yields a scalar, i.e.,

$$\begin{aligned} y &= \mathscr{X} \times_1 a^{(1)} \times_2 a^{(2)} \times_3 \cdots \times_N a^{(N)} \\ &= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N}^{I_N} x_{i_1,i_2,\ldots,i_N} a_{i_1}^{(1)} a_{i_2}^{(2)} \ldots a_{i_N}^{(N)}. \end{aligned} \tag{4.5}$$

We refer the readers to [57, 61] for further details and discussions about various tensor operations.

4.3 Problem Formulation

The problem of interest here is the underdetermined JBSS for multiple datasets, e.g., K datasets. The M observations of each dataset contain the linear mixtures of the corresponding N sources. We can model the mixing process as follows,

$$X^{(k)} = A^{(k)} S^{(k)} + E^{(k)}, \quad k = 1,2,\ldots,K. \tag{4.6}$$

$X^{(k)} = [x_1^{(k)}, x_2^{(k)}, \ldots, x_M^{(k)}]^T$ denotes the M-dimensional observations with real values and $x_m^{(k)}$ is the m^{th} channel of the observations in dataset k. $S^{(k)} = [s_1^{(k)}, s_2^{(k)}, \ldots, s_N^{(k)}]^T$ means the underlying N-dimensional sources with real values and $s_n^{(k)}$ is the n^{th} source for dataset k. $A^{(k)} = [a_1^{(k)}, a_2^{(k)}, \ldots, a_N^{(k)}] \in \mathbb{R}^{M \times N}$ with $M < N$ (i.e., the underdetermined case) denotes the unknown mixing matrix, whose n^{th} column $a_n^{(k)}$ corresponds to the source $s_n^{(k)}$ for dataset k. $E^{(k)}$ means the possible additive noise

which is generally assumed to be zero mean, temporally white and uncorrelated with the source signals.

Similar to several existing JBSS methods, e.g. MCCA [72] and JDAIG-SOS [69], we have the following assumptions regarding the sources:

(1) The sources are uncorrelated within each dataset:

$$\begin{aligned} E\{s_i^{(k)}(t)(s_j^{(k)}(t+\tau))^T\} &= 0 \\ \forall \tau,\ 1 \le i \ne j \le N,\ k &= 1,2,\ldots,K, \end{aligned} \tag{I}$$

where $s_i^{(k)}(t)$ is the i-th source in dataset k and $s_j^{(k)}(t+\tau)$ represents the j-th source with the time delay τ in dataset k.

(2) The corresponding sources from two different datasets have non-zero correlations and the sources with different indices across datasets are not correlated:

$$\begin{aligned} D^{(\tau)} =& E\{S^{(k_1)}(t)(S^{(k_2)}(t+\tau))^T\} \\ =& Diag(\rho_1(\tau),\rho_2(\tau),\ldots,\rho_N(\tau)), \end{aligned} \tag{II}$$

where $Diag(\cdot)$ represents the diagonal matrix, the $\rho_n(\tau) = E\{s_n^{(k_1)}(t)(s_n^{(k_2)}(t+\tau))^T\}$ denotes the covariance between $s_n^{(k_1)}(t)$ and $s_n^{(k_2)}(t+\tau)$. This assumption means that the corresponding sources in multiple datasets are second-order correlated with each other. In addition, the sources within $[s_i^{(1)}, s_i^{(2)}, \ldots, s_i^{(K)}]$ are uncorrelated with the sources within $[s_j^{(1)}, s_j^{(2)}, \ldots, s_j^{(K)}]$ for $1 \le i \ne j \le N$.

The task of estimating the mixing matrices $\{A^{(k)}\}$ and retrieving the underlying sources are not equivalent in the underdetermined case. Therefore, most UBSS methods consist of two stages: estimate the mixing matrices first and then retrieve the underlying sources. The major problem under consideration is to estimate $\{A^{(k)}\}$ jointly up to permutation and scaling. In this chapter, this problem is addressed via a specially designed joint tensor decomposition. In addition, retriev-

ing the underlying sources when the mixing matrices are estimated or known is a classic inverse problem [62]. In order to further demonstrate the performance of the proposed mixing matrices estimation method, we also implement an approach for source recovering based on the estimated $A^{(k)}$.

4.4 Canonical Polyadic Decomposition of Tensor

A polyadic decomposition aims to decompose a higher-order tensor as a linear combination of rank-one tensors [61, 120]. For the case of a third-order tensor $\mathscr{X} \in \mathbb{R}^{I\times J\times K}$, it can be written in the form

$$\mathscr{X} = \sum_{n=1}^{N} a_n \circ b_n \circ c_n, \tag{4.7}$$

where N is a positive integer and $a_n \in R^I$, $b_n \in R^J$, $c_n \in R^K$. Equivalently, it can be written element wisely as

$$x_{i,j,k} = \sum_{n=1}^{N} a_{i,n} b_{j,n} c_{k,n}, \tag{4.8}$$

where $i = 1,2,\ldots,I,\ j = 1,2,\ldots,J$ and $k = 1,2,\ldots,K$. The rank of a tensor is the smallest number of rank-one tensors that yield the tensor in the way as (4.7). If $rank(\mathscr{X}) = N$, (4.7) is the CPD of $\mathscr{X}$, which is also known as the Canonical Decomposition (CANDECOMP) or Parallel Factor Analysis (PARAFAC) [2]. The canonical polyadic approximation means that

$$\begin{aligned} \mathscr{X} &\approx [\![A,B,C]\!] \\ &\equiv \sum_{n=1}^{N} a_n \circ b_n \circ c_n, \end{aligned} \tag{4.9}$$

where $N = rank(\mathscr{X})$.

The factor matrices refer to the combination of the vectors corresponding to each rank-one tensor and can be written as

$$\begin{aligned} A &= [a_1, a_2, \ldots, a_N] \in R^{I \times N} \\ B &= [b_1, b_2, \ldots, b_N] \in R^{J \times N} \\ C &= [c_1, c_2, \ldots, c_N] \in R^{K \times N}. \end{aligned} \tag{4.10}$$

To a large extent, the power of CPD mainly stems from its uniqueness property. The uniqueness of CPD means that the decomposition is the only possible combination of rank-one tensors which sum to the objective tensor with the exception of the indeterminacies of column permutation and scaling. The permutation indeterminacy refers to the fact that we can permute the rank-one terms arbitrarily. The scaling indeterminacy means that we can scale the individual column of the factor matrices as long as their product remains the same, i.e.,

$$\mathscr{X} = \sum_{n=1}^{N} (\alpha_n^1 a_n) \circ (\alpha_n^2 b_n) \circ (\alpha_n^3 c_n) \text{ if } \alpha_n^1 \alpha_n^2 \alpha_n^3 = 1. \tag{4.11}$$

The uniqueness condition is based on the rank of tensors. The most famous result on uniqueness of CPD was reported by J. Kruskal [64]. Kruskal's theorem states that the CPD of a third-order tensor $\mathscr{X} \in \mathbb{R}^{I \times J \times K}$ is deterministically unique if N (where $N = rank(\mathscr{X})$) satisfies

$$N \leq \frac{k_A + k_B + k_C - 2}{2}, \tag{4.12}$$

where $k_{\cdot}$ denotes the k-rank of a given matrix $(\cdot)$, meaning that $k_{\cdot}$ is the largest integer that any $k_{\cdot}$ columns of the matrix $(\cdot)$ are linearly independent. Checking deterministic conditions can be cumbersome. De Lathauwer et al. have studied different methods to determine the rank of a tensor and concluded that the de-

composition of a third-order tensor $\mathscr{X} \in \mathbb{R}^{I\times J\times K}$ is generically unique (i.e., with probability one) [30] provided that N satisfies

$$N \leq K \quad \textit{and} \quad N(N-1) \leq IJ(I-1)(J-1)/2. \tag{4.13}$$

Domanov et al. further complemented the existing bounds for generic uniqueness of the CPD [36] and concluded that the CPD of a third-order tensor $\mathscr{X} \in \mathbb{R}^{I\times J\times K}$ of rank N is generically unique if

$$\begin{aligned} &2 \leq I \leq J \leq K \leq N \\ &N \leq \frac{I+J+2K-2-\sqrt{(I-J)^2+4K}}{2}, \end{aligned} \tag{4.14}$$

or

$$\begin{aligned} &3 \leq I \leq J \leq N \leq K \\ &N \leq (I-1)(J-1). \end{aligned} \tag{4.15}$$

There are two main approaches to compute the CPD of a tensor, namely the linear algebra [35] and optimization based methods [2, 99]. Both types of methods have their own strengths and weaknesses. For a thorough study of the uniqueness conditions and computation, we refer to [30, 34, 61] and the references therein.

4.5 Algorithm for Estimating the Mixing Matrices in UJBSS

How to estimate the mixing matrix is still a challenging problem, even for underdetermined case of single dataset. In this chapter, we propose a novel and effective algorithm to jointly estimate the mixing matrices from multiple dataset, which can be regarded as an extension of the method based on statistical property of signals, e.g., simultaneous diagonalization of the second order autocovariance and CPD of

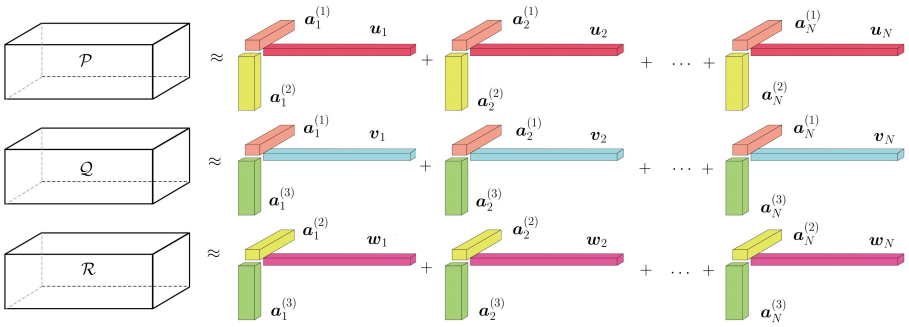

Figure 4.1: Illustration of how to generate tensors by incorporating the dependence information between each pair of datasets.

a specialized tensor [31, 46, 124]. For ease of presentation, we take the case of 3 datasets as an example, e.g., $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$, and it can be easily generalized to the case of more than 3 datasets. The problem is reformulated as joint canonical polyadic decomposition of a sequence of third-order tensors, which share common factor matrices. It should be mentioned that the proposed method is limited to real-valued problems and cannot be directly generalized to complex-valued cases.

4.5.1 Tensor Construction

The cross covariance of the observations with time delay τ, such as the observations in dataset k_1, $X^{(k_1)}(t)$, and the observations in dataset k_2 with time delay τ, $X^{(k_2)}(t+\tau)$, can be formulated as

$$\begin{aligned} &E\{X^{(k_1)}(t)X^{(k_2)}(t+\tau)^T\} \\ =&(A^{(k_1)})E\{S^{(k_1)}(t)S^{(k_2)}(t+\tau)^T\}(A^{(k_2)})^T, \end{aligned} \tag{4.16}$$

where k_1 and k_2 represent the index of each dataset and range from 1 to 3. Considering the correlations within and between each pair of datasets, the covariance

matrices between $X^{(1)}$ and $X^{(2)}$ with time delay τ satisfy

$$\begin{aligned} P^{(1)} &= E\{X^{(1)}(t)X^{(2)}(t+\tau_1)^T\} = (A^{(1)})U^{(\tau_1)}(A^{(2)})^T, \\ P^{(2)} &= E\{X^{(1)}(t)X^{(2)}(t+\tau_2)^T\} = (A^{(1)})U^{(\tau_2)}(A^{(2)})^T, \\ &\vdots \\ P^{(L)} &= E\{X^{(1)}(t)X^{(2)}(t+\tau_L)^T\} = (A^{(1)})U^{(\tau_L)}(A^{(2)})^T, \end{aligned} \tag{4.17}$$

in which τ_l means the time delay and the matrix $U^{(\tau_l)} = E\{S^{(1)}(t)S^{(2)}(t+\tau_l)^T\}$ is diagonal, for $l = 1,2,...,L$.

We stack the sequence of covariance matrices $P^{(1)}, P^{(2)}, \ldots, P^{(L)}$, denoted as $\{P^{(l)}\}$, in a tensor $\mathscr{P} \in \mathbb{R}^{M\times M\times L}$ as follows: $(\mathscr{P})_{i,j,l} = (P^{(l)})_{i,j}$, $i = 1,2,\ldots,M$, $j = 1,2,\ldots,M$, $l = 1,2,\ldots,L$. We define the matrix U of size $L \times N$ with the element $U_{l,n} = (U^{(\tau_l)})_{n,n}$, for l = 1, 2, ..., L, n = 1, 2, ..., N. Then we can represent $\mathscr{P}$ as (see Fig. 4.1):

$$\mathscr{P} = \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(2)} \circ u_n, \tag{4.18}$$

in which $a_n^{(1)}$ and $a_n^{(2)}$ are the n^{th} column of the mixing matrices $A^{(1)}$ and $A^{(2)}$ respectively, and u_n is the n^{th} column of the matrix U.

Similarly, the covariance matrix between the other two pairs of observations with time delay τ_l, denoted as $Q^{(l)}$ and $R^{(l)}$ satisfy

$$\begin{aligned} Q^{(l)} &= E\{X^{(1)}(t)X^{(3)}(t+\tau_l)^T\} = (A^{(1)})V^{(\tau_l)}(A^{(3)})^T, \\ R^{(l)} &= E\{X^{(2)}(t)X^{(3)}(t+\tau_l)^T\} = (A^{(2)})W^{(\tau_l)}(A^{(3)})^T, \end{aligned} \tag{4.19}$$

where $V^{(\tau_l)} = E\{S^{(1)}(t)S^{(3)}(t+\tau_l)^T\}$ and $W^{(\tau_l)} = E\{S^{(2)}(t)S^{(3)}(t+\tau_l)^T\}$ for l = 1, 2, ..., L. Stack these two sequence of covariance matrices $\{Q\}$ and $\{R\}$ in tensors $\mathscr{Q} \in \mathbb{R}^{M\times M\times L}$ and $\mathscr{R} \in \mathbb{R}^{M\times M\times L}$ as follows: $(\mathscr{Q})_{i,j,l} = (Q^{(l)})_{i,j}$, $(\mathscr{R})_{i,j,l} = (R^{(l)})_{i,j}$,

i = 1, 2, ..., M, j = 1, 2, ..., M, l = 1, 2, ..., L. To simplify the notation, we further define the matrix $V \in \mathbb{R}^{L\times N}$ and $W \in \mathbb{R}^{L\times N}$ with the element $V_{l,n} = (V^{(\tau_l)})_{n,n}$ and $W_{l,n} = (W^{(\tau_l)})_{n,n}$, for l = 1, 2, ..., L, n = 1, 2, ..., N. Then these two tensors can be represented as (see Fig. 4.1):

$$\begin{aligned} \mathscr{Q} &= \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(3)} \circ v_n, \\ \mathscr{R} &= \sum_{n=1}^{N} a_n^{(2)} \circ a_n^{(3)} \circ w_n, \end{aligned} \tag{4.20}$$

in which $a_n^{(k)}$ is the n^{th} column of the mixing matrices $A^{(k)}$ for k = 1, 2, 3, v_n and w_n are the n^{th} column of the matrix V and W respectively.

It should be mentioned that the choices of τ_1, τ_2, ..., τ_L may affect the estimation precision of the mixing matrices. If τ_l is too large, the correlation between two related sources with the delay will be close to 0 and then the covariance matrix might be ill conditioned. It is desired to select τ_1, τ_2, ..., τ_L such that U, V and W are well conditioned. In addition, if the time delay τ is too large, the covariance matrix of the sources in two datasets (e.g., $U^{(\tau)}$) will be close to a null matrix and thus the assumption (II) may not hold. Here, we heuristically choose the time delay as $\tau_l \in [0, 200]$ data samples.

Fig. 4.1 illustrates how to generate these tensors by incorporating the dependence information between each pair of datasets. It is worth noting that each pair of tensors share a common factor matrix, e.g., $\mathscr{P}$ and $\mathscr{Q}$ are coupled in the mode of $A^{(1)}$.

4.5.2 Joint Tensor Polyadic Decomposition

Considering the common latent structure, now the problem of estimating the mixing matrices $A^{(k)}$ can be reformulated as a problem of joint CPD of a collection of

tensors, e.g., $\mathscr{P}$, $\mathscr{Q}$ and $\mathscr{R}$ for the case of three datasets. There are two main approaches to jointly decompose a sequence of tensors, i.e., linear algebra [101] and optimization based methods [3, 100, 109]. Sørensen et al. took into account the coupling between multiple tensors and developed a linear algebra based algorithm [101]. This method can provide an explicit solution for exact tensor decomposition. However, in practice data are noisy and consequently the estimation may be not accurate. In addition, it is notable that the linear algebra based method requires the full column rank of the common factor matrices whereas the common factors in our problem are rank deficient [101]. In this chapter, we generalize the idea of coupled matrix and tensor factorization (CMTF) and jointly decompose a sequence of tensors via gradient-based optimization method [3, 100, 109].

The uniqueness condition of the joint CPD is important in practice. Simply said, the solution of the joint CPD will be generic unique if all the individual CPDs are unique. In this chapter, we can get the unique solution of each mixing matrix generically, providing the number of sources satisfies the condition (4.14) or (4.15). It is worth mentioning that this uniqueness condition of the joint CPD might be further relaxed, but the topic itself deserves a stand-alone theoretical paper and it is out of scope of the current paper.

The aim is to find the factor matrices $\{A^{(k)}\} \in \mathbb{R}^{M\times N}$ and the covariance of sources between different datasets U, V and $W \in \mathbb{R}^{L\times N}$ which can minimize the following objective function, a variant of Frobenius norm of the difference between the given tensors and their canonical polyadic approximation, written as

$$
\begin{aligned}
& f(A^{(1)},A^{(2)},A^{(3)},U,V,W) \\
=& \frac{1}{2}\underbrace{\|\mathscr{P}-[\![A^{(1)},A^{(2)},U]\!]\|^2}_{f^{(1)}(A^{(1)},A^{(2)},U)}+\frac{1}{2}\underbrace{\|\mathscr{Q}-[\![A^{(1)},A^{(3)},V]\!]\|^2}_{f^{(2)}(A^{(1)},A^{(3)},V)} \\
+& \frac{1}{2}\underbrace{\|\mathscr{R}-[\![A^{(2)},A^{(3)},W]\!]\|^2}_{f^{(3)}(A^{(2)},A^{(3)},W)}.
\end{aligned}
\tag{4.21}
$$

where $[\![\cdot]\!]$ denotes the canonical polyadic approximation of a given tensor. This equation simultaneously takes the coupling information between different tensors into account. We propose to solve this problem via a gradient-based optimization method. Proposition 1 elaborates the partial derivative of the objective function f with respect to each column of the desired matrices, i.e. $\{a_n^{(k)}\}$, u_n, v_n and w_n for $n = 1,2,\ldots,N$.

The equations in Proposition 1 is proved in the Appendix. Then the gradient of f can be assembled via stacking the partial derivatives with respect to each column of the factor matrices, as

$$
\nabla f = \left[\frac{\partial f}{\partial a_1^{(1)}};\frac{\partial f}{\partial a_2^{(1)}};\ldots;\frac{\partial f}{\partial a_N^{(1)}};\ldots;\frac{\partial f}{\partial w_1};\ldots;\frac{\partial f}{\partial w_N}\right]^T. \tag{4.22}
$$

Once we get this gradient, we can calculate the factor matrices, including the mixing matrices and the covariance matrices, based on any first-order optimization method. In this chapter, we employ the nonlinear conjugate gradient (NCG) algorithm implemented in [38] to solve the unconstrained optimization problem and estimate the mixing matrices of multiple datasets simultaneously. Compared with second-order optimization methods, such as Newton-based methods, NCG always requires less computation and memory[109].

Proposition 1. *The partial derivative of the objective function f with respect to each column of the desired matrices , i.e., $\{a_n^{(k)}\}$, u_n, v_r and w_n, are given by*

$$\frac{\partial f}{\partial a_n^{(1)}} = -\mathscr{P} \times_2 a_n^{(2)} \times_3 u_n - \mathscr{Q} \times_2 a_n^{(3)} \times_3 v_n + \sum_{c=1}^{N} [(a_n^{(2)})^T a_c^{(2)} (u_n)^T u_c + (a_n^{(3)})^T a_c^{(3)} (v_n)^T v_c] a_c^{(1)}$$

$$\frac{\partial f}{\partial a_n^{(2)}} = -\mathscr{P} \times_1 a_n^{(1)} \times_3 u_n - \mathscr{R} \times_2 a_n^{(3)} \times_3 w_n + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (u_n)^T u_c + (a_n^{(3)})^T a_c^{(3)} (w_n)^T w_c] a_c^{(2)}$$

$$\frac{\partial f}{\partial a_n^{(3)}} = -\mathscr{Q} \times_1 a_n^{(1)} \times_3 v_n - \mathscr{R} \times_1 a_n^{(2)} \times_3 w_n + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (v_n)^T v_c + (a_n^{(2)})^T a_c^{(2)} (w_n)^T w_c] a_c^{(3)}$$

$$\frac{\partial f}{\partial u_n} = -\mathscr{P} \times_1 a_n^{(1)} \times_2 a_n^{(2)} + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (a_n^{(2)})^T a_c^{(2)}] u_c$$

$$\frac{\partial f}{\partial v_n} = -\mathscr{Q} \times_1 a_n^{(1)} \times_2 a_n^{(3)} + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (a_n^{(3)})^T a_c^{(3)}] v_c$$

$$\frac{\partial f}{\partial w_n} = -\mathscr{R} \times_1 a_n^{(2)} \times_2 a_n^{(3)} + \sum_{c=1}^{N} [(a_n^{(2)})^T a_c^{(2)} (a_n^{(2)})^T a_c^{(3)}] w_c.$$

4.6 Source Extraction Based on the Estimated Mixing Matrices

Unlike the (over)determined case, the estimation of the mixing matrix is not equivalent to recovering the underlying sources in UBSS. A complete UBSS approach always consists of both mixing matrix estimation and source extraction, even though our main focus in this chapter is the estimation of mixing matrices. Extracting the sources when the mixing matrix is estimated is a classic inverse problem. Many techniques have already been proposed in the literature, including array processing techniques [107] and methods exploiting the sparsity of sources in a domain, e.g., the TF domain [116]. In order to demonstrate the performance of the proposed mixing matrices estimation method, we adopt a recently-developed subspace representation method [59] to recover the latent sources based on the estimated mixing matrices. For simplicity, the proposed method for extracting sources is derived

without considering the background noise. However, it was shown to be robust to the background noise [59].

For any underdetermined non-homogeneous linear equation, the complete solution can be represented as the sum of its particular solution and a general solution of the corresponding homogeneous equation. As to the case in this chapter, $A^{(k)}S^{(k)} = X^{(k)}$, the general solution of source $S^{(k)}$ can be written as

$$S^{(k)} = S_p^{(k)} + S_h^{(k)}, \tag{4.23}$$

where the $S_p^{(k)}$ denotes its particular solution and $S_h^{(k)}$ denotes a general solution of the corresponding homogeneous equation $A^{(k)}S^{(k)} = 0$. One particular solution of the above mentioned non-homogeneous equation is

$$S_p^{(k)} = (A^{(k)})^{\dagger} X^{(k)}, \tag{4.24}$$

where $(A^{(k)})^{\dagger}$ denotes the pseudo-inverse of the mixing matrix $A^{(k)}$. In addition, the general solution of the homogeneous equation $A^{(k)}S^{(k)} = 0$ can be expressed as

$$S_h^{(k)} = VZ^{(k)}, \tag{4.25}$$

where V is an $N * (N - M)$ matrix whose columns are bases of the nullspace of $A^{(k)}$ and $Z^{(k)}$ is an arbitrary matrix with the size of $(N - M) * T$ (T represents the total number of samples in each channel) [102]. The basis matrix V can be obtained from the mixing matrix $A^{(k)}$ and then the problem which aims to estimate the N dimensional observations boils down to the problem of estimating $N - M$ dimensional latent variable $Z^{(k)}$.

In order to be applicable to a wide class of signals, such as audio and biological signals EEG, EMG, the Generalized Gaussian Distribution (GGD) [58] is utilized

to model the source distributions. Mathematically, it is expressed in the following equation

$$p_y(y;\sigma,\beta) = \frac{v(\beta)}{\sigma} exp\{-c(\beta)|\frac{y-\mu}{\sigma}|^{2/(1+\beta)}\}, \tag{4.26}$$

where

$$\begin{aligned} c(\beta) &= (\frac{\Gamma(3/2(1+\beta))}{\Gamma(1/2(1+\beta))})^{1/(1+\beta)} \\ v(\beta) &= \frac{\Gamma(3/2(1+\beta))^{1/2}}{(1+\beta)\Gamma(1/2(1+\beta))^{3/2}}, \end{aligned} \tag{4.27}$$

in which $\Gamma(\cdot)$ is the Gamma function. σ is the standard derivation and μ is the mean of a continuous random variable y. In this chapter, the mean of source is assumed to be 0. We define the parameter set $\theta = \{\beta, \sigma\}$ for simplicity, where each component of $\beta = [\beta_1, \ldots, \beta_N]$ and $\sigma = [\sigma_1, \ldots, \sigma_N]$ correspond to each channel of the sources. The parameters of the GGD θ can be estimated to maximize the likelihood of the observed mixtures $X^{(k)}$ based on Expectation-maximization (EM) algorithm. Then Z can be obtained by sampling from $p(Z^{(k)}|X^{(k)},\theta)$ as

$$\hat{Z}^{(k)} = \frac{1}{G}\sum_{g=1}^{G} Z_g^{(k)}, \tag{4.28}$$

where $\{Z_1^{(k)}, \ldots, Z_G^{(k)}\}$ are the G samples drawn from $p(Z^{(k)}|X^{(k)},\theta)$ using the Markov Chain Monte Carlo (MCMC) method. Then we recover the underlying sources based on

$$\hat{S}^{(k)} = (A^{(k)})^{\dagger} X^{(k)} + V Z^{(k)}. \tag{4.29}$$

The major steps of the proposed UJBSS-m algorithm are summarized in Algorithm 6. The number of time delays is 20 in default. The step size of time delays, i.e. $\tau_{l+1} - \tau_l$, is suggested to be 2 samples (corresponding to 0.25ms) and 5 samples (corresponding to 5ms) for audio signals and physiological signals respectively.

Algorithm 6 The UJBSS-m algorithm based on joint tensor decomposition

Input: M-dimensional observations $\{X^{(k)}\}$ and the number of sources N in each datasets, for k= 1, 2,...,K.

Output: the estimated mixing matrices $\{A^{(k)}\}$ and the recovered N-dimensional sources $\{S^{(k)}\}$, for k= 1, 2,...,K.

STEP 1: For each pair of datasets, e.g., $X^{(k_1)}$ and $X^{(k_2)}(k_1 \neq k_2)$, we calculate the cross covariance matrices as (5.3) and stack them to construct a third-order tensor as in Section 4.5.1. Considering the combination of datasets, we get $\binom{K}{2}$ tensors where each pair of tensors share a common factor matrix, as shown in Fig. 4.1;

STEP 2: Calculate the joint polyadic decomposition of the tensors constructed in step 1 via optimization based method and estimate the mixing matrices $\{A^{(k)}\}$ as in Section 4.5.2;

STEP 3: Estimate the parameters of the Generalized Gaussian distribution based on the EM algorithm.

Initialize: initialize the parameter θ to some random values.

E-step: calculate the expected value of the log likelihood function with respect to the conditional distribution of $Z^{(k)}$ given the observation $X^{(k)}$ under the current estimate of θ. It can be expressed as $E_{p(Z^{(k)}|X^{(k)},\theta^*)}(log(p(Z^{(k)}|X^{(k)},\theta)))$, where θ^* means the parameter value got in the initialization or the previous M-step.

M-step: update the parameter set θ to maximize the above expected value. The updated value is

$$\begin{aligned}\theta &= \operatorname*{argmax}_{\theta} E_{p(Z^{(k)}|X^{(k)},\theta^*)}(log(p(Z^{(k)}|X^{(k)},\theta))) \\ &\approx \operatorname*{argmax}_{\theta} \frac{1}{G}\sum_{g=1}^{G} log(p(Z_g^{(k)}|X^{(k)},\theta)),\end{aligned}$$

where $\{Z_1^{(k)},\ldots,Z_G^{(k)}\}$ are G samples drawn from $p(Z^{(k)}|X^{(k)},\theta)$ based on the MCMC method.

Iterate: iterate the E-step and M-step until convergence.

STEP 4: Recover the sources $S^{(k)}$ based on the minimum mean-square error criterion as in Equation (4.29).

4.7 Numerical Study for the Multiple Dataset Case

To demonstrate the joint separation performance for multiple datasets, simulations are performed on both audio and biological signals when applying the proposed UJBSS-m and several commonly used BSS methods. Two performance indices are used to evaluate the separation performances. One is the estimation error of the mixing matrices, defined as:

$$Error = 10log_{10}\{mean(\frac{||A-\hat{A}||}{||A||})\}, \tag{4.30}$$

where $\hat{A}$ denotes the optimally ordered estimate of A. The other measures the Pearson correlation coefficient (PCC) between the estimated sources and the original ones, which is defined as

$$PCC(s_n^{(k)}, \hat{s}_n^{(k)}) = \frac{cov(s_n^{(k)}, \hat{s}_n^{(k)})}{\sigma_{s_n^{(k)}} \sigma_{\hat{s}_n^{(k)}}}, \tag{4.31}$$

where $\hat{s}_n^{(k)}$ means the estimate of the source $s_n^{(k)}$ in the k_{th} dataset, $cov(\cdot,\cdot)$ means the covariance between two variables and σ means the standard deviation. In order to ensure the dependence between the sources of each pair of datasets, the sources are synthesized as follows,

$$\begin{aligned} S^{(1)} =& [s_1^{(1)}, s_2^{(1)}, \ldots, s_N^{(1)}]^T; \\ S^{(2)} =& S^{(1)} .* (unifrnd(0,1,S^{(1)}) \\ S^{(3)} =& S^{(1)} .* (unifrnd(0,1,S^{(1)}), \end{aligned} \tag{4.32}$$

where $unifrnd(0,1,S^{(1)})$ generates a matrix with the same size of $S^{(1)}$ and each element of the matrix is randomly drawn from the continuous uniform distribution on the interval (0,1). The average correlation between the source $s_n^{(1)}$ and the cor-

responding source $s_n^{(k)}(k=2,3)$ is about 0.85; the average correlation between the source $s_n^{(2)}$ and the corresponding source $s_n^{(3)}$ is about 0.7, both of which can be regarded as highly correlated.

4.7.1 Simulation 1: Audio Signals

The sources used in this simulation include 8 audio signals, such as two pieces of sound from the cable news network (CNN) news and a piece of sound of an anonymous singer, all of which are publicly available[1]. The sampling rate is 8000Hz. The mixing matrices are generated randomly with elements following the uniform distribution $U[-1,1]$. For simplicity, each column of the mixing matrices is normalized into a unit vector. Three datasets are generated following (4.32). In our first setting, 5 sources are mixed into 4 observations in each dataset and the corresponding sources in the different datasets are highly correlated. With different signal-to-noise ratios (SNRS)), we compare the proposed UJBSS-m method with a commonly-used single-set UBSS method, SOBIUM [31], when it is applied to each dataset separately. We also test the performance of our recent work on UJBSS for two datasets, UJBSS-2 [124], when two datasets are available, e.g., $X^{(1)}$ and $X^{(2)}$. We repeat the simulation 1000 times and the performance is shown in Fig. 4.2. Results are given according to the SNR level in the range of -5dB - 40dB. Benefiting from dependence information between different datasets, the proposed UJBSS can provide more accurate estimation of the mixing matrices, while SOBIUM neglects the possible inter-dataset information. Compared with UJBSS-2, the proposed UJBSS-m takes into account more dependence information, among three datasets rather than between two datasets, and yields better performance. We also note that the *Error* measure from the single set UBSS and UJBSS methods decreases with the increase of the SNR. The proposed UJBSS-m consistently

[1] http://research.ics.aalto.fi/ica/cocktail/sounds.html

provides the best results over the whole SNR range, suggesting the performance stability of the proposed algorithm.

We also examine the performance of the proposed method with the decrement of the under-determinacy level, i.e., the number of the observations increases from 4 to 7 while the number of sources is set to be 8. As noted in Fig. 4.3, the estimation performance is getting better when more observations are available. Besides the degraded estimation precision, a higher under-determinacy level also requires higher computational complexity. The performance is getting better when the SNR is increased from -5dB to 20dB. The change of estimation error is not obvious when the SNR is greater than 20dB even there are some fluctuations. It is shown that the estimation performance relies upon several factors such as the noise level SNR, under-determinacy level (i.e., the number of sources for a given number of sensors) and the correlation between each pair of datasets.

We recover the latent sources from each dataset based on the estimated mixing matrices. In an illustrative example, we linearly mix 4 audio sources into 3 observations in each dataset. Fig. 4.4 shows the separation results of the first dataset in the time domain. The top four subfigures of Fig. 4.4 represent the original sources, the middle three subfigures are the mixed observations and the bottom four subfigures are the recovered sources via the proposed UJBSS-m method. In addition, we compare the proposed method with other three single-set UBSS methods, including SOBIUM, UBSS based on subspace representation (UBSS-SR for short) [59] and UBSS based on sparse coding (UBSS-SC for short) [119], as well as the JBSS method MCCA [72] and UJBSS for two datasets UJBSS-2 [124], in term of the PCC between the original sources and the recovered ones. Both UBSS-SR and UBSS-SC are based on single source detection, which assumes that the TF points are occupied by a single source or the corresponding single source possesses dominant energy. However, the performance of estimating the mixing matrix de-

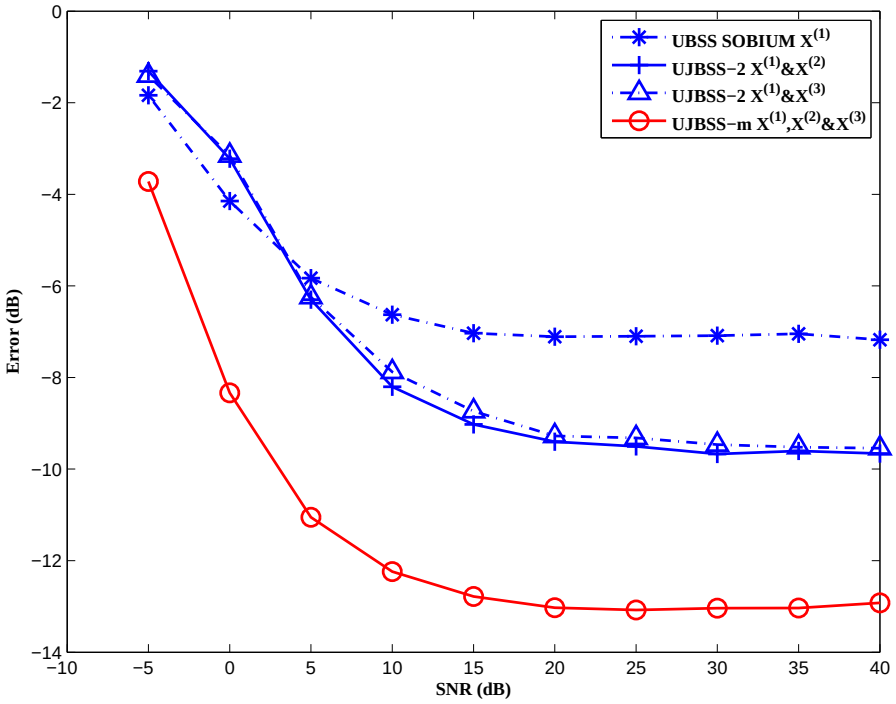

Figure 4.2: Simulation 1: performance comparisons on audio signals when using the proposed UJBSS-m method and other UBSS methods, including the single-set UBSS method SOBIUM [31] and the UJBSS method for two dataset, i.e., UJBSS-2 [124]. Here the number of sources N = 5 and the number of observations M = 4. The number of time delays L = 20 and the step size of time delays (i.e., $\tau_l - \tau_{l-1}$) is 2 data samples, corresponding to 0.25ms. Similar results are observed for $A^{(2)}$ and $A^{(3)}$.

teriorates when this assumption is not satisfied. Furthermore, given that the time-frequency analysis method [116, 124] is memory-intensive and time-consuming, we estimate the mixing matrices via UJBSS-2 and SOBIUM respectively, and then extract the sources using the same method as in UBSS-SR [59]. The performance results of these six methods are reported in Table 4.1.

Despite adopting the same technology in extracting sources, the performance of the proposed method is significantly better than that of the single-set SOBIUM

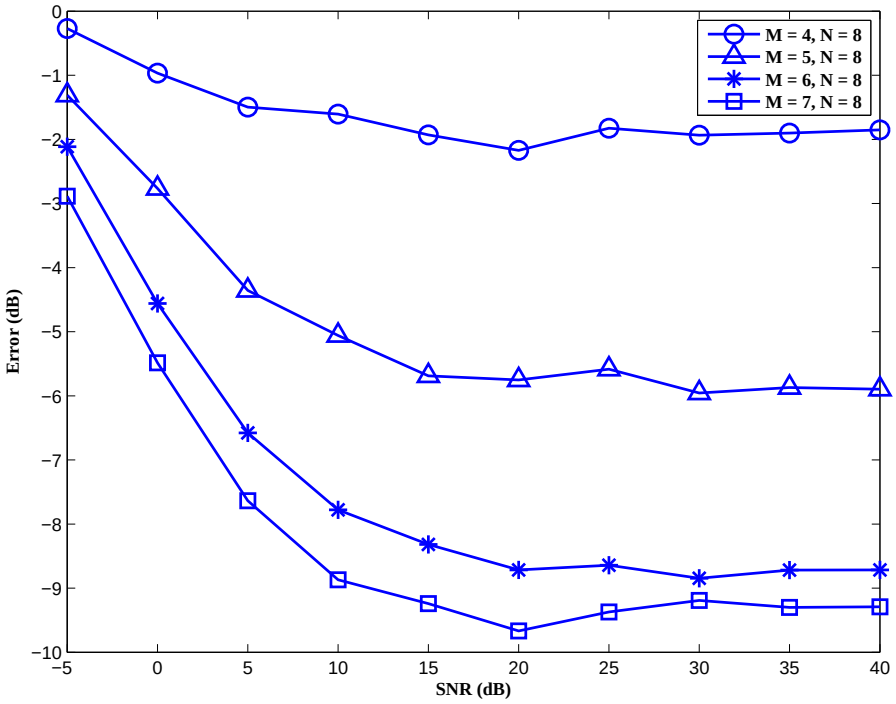

Figure 4.3: Simulation 1: estimation error of $A^{(1)}$ when employing the proposed UJBSS method. Here the number of sources N = 8 and the number of observations M varies from 4 to 7. The number of time delays L = 20 and the step size of time delays (i.e., $\tau_l - \tau_{l-1}$) is 2 data samples, corresponding to 0.25ms.

and UBSS-SR method. This observation confirms the importance of estimating mixing matrices accurately. In addition, the proposed method also outperforms a recently proposed UBSS method UBSS-SC. The main reason is that such UBSS methods always require the sparsity of the sources to some extent, while the assumption may not be satisfied in reality. MCCA, which has been successfully used in many fields [27], assumes that the number of sources is equal to the number of observations in each dataset and it could not be used to separate sources in the underdetermined case directly. We add one observation in each dataset so that

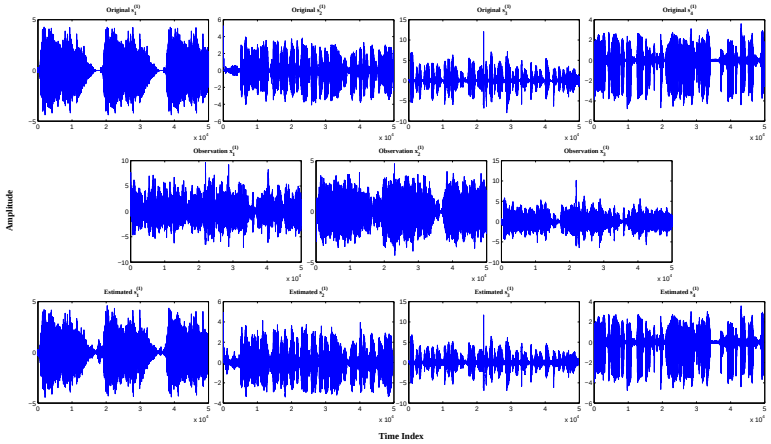

Figure 4.4: Simulation 1: an illustrative example from the proposed UJBSS-m method. First row: The original 4 sources; Second row: 3 channels of the mixed observations; Third row: the recovered 4 sources from the first dataset.

MCCA can be applied. Therefore it is not really a fair setting and comparison to the proposed method. However, we note that the performance of MCCA is not as good as that of the proposed method, even with an additional observation signal. The following reasons could contribute to the worse performance of MCCA: it is mainly due to the fact that the correlation coefficients between sources in two datasets are quite close [20, 72]; the performance of MCCA may suffer from error accumulation of the deflation-based separation methods [69].

4.7.2 Simulation 2: Physiological Signals

In this experiment, we employ four physiological signals as sources, including ECG, EEG, EOG and EMG from a publicly available database [45]. The sampling rate is 1000 Hz. The sources corresponding to the other two datasets are generated following (4.32). We get similar results as that in the simulation 1. As can be seen from Fig. 4.5, the estimation performance is getting better with the increase of the

Table 4.1: PCC performance results in Simulation 1.

	Methods	s_1	s_2	s_3	s_4
Dataset 1	UJBSS-m	0.993	1.000	0.881	0.992
	UJBSS-2[124]	0.989	1.000	0.859	0.990
	SOBIUM[31]	0.980	1.000	0.512	0.967
	UBSS-SC[119]	0.951	0.965	-0.066	0.934
	UBSS-SR[59]	0.901	0.999	0.226	0.967
	MCCA*[72]	0.571	0.909	0.675	0.732
Dataset 2	UJBSS-m	0.910	0.998	0.740	0.944
	UJBSS-2[124]	0.897	0.998	0.714	0.892
	SOBIUM[31]	0.877	0.998	0.686	0.962
	UBSS-SC[119]	0.795	0.940	-0.401	0.607
	UBSS-SR[59]	0.885	0.981	0.730	0.954
	MCCA*[72]	-0.574	0.911	-0.678	-0.733
Dataset 3	UJBSS-m	0.899	1.000	0.783	0.870
	UJBSS-2[124]	0.720	1.000	0.515	0.869
	SOBIUM[31]	0.756	-0.951	0.078	0.880
	UBSS-SC[119]	0.381	-0.434	-0.614	-0.630
	UBSS-SR[59]	-0.557	0.987	0.620	-0.561
	MCCA*[72]	0.577	-0.911	-0.679	0.732

(1) *We add one additional observation in each dataset when we evaluate the MCCA.

SNR of observations. In the whole SNR range, the proposed UJBSS-m method estimates the mixing matrices with higher accuracy than the single-set UBSS method SOBIUM and UJBSS method for two datasets UJBSS-2.

We also investigate the effect of the time delays, as shown in Fig. 4.6. At high SNR level, e.g. SNR = 20dB, the average Error of the proposed UJBSS-m is -11.57dB when the step size of the time delays is 5 data samples,corresponding to 5ms. However, the average Error corresponding to 3 data samples is -6.67dB, significantly larger than that of 5 data samples. The main reason is that the change of the covariance matrices is not obvious for the small step size and the covariance

matrices related to these delays could not provide enough information to estimate the common factors, i.e. the mixing matrices. If the time delay is too large, such as more than 500 data samples (corresponding to 500ms), the covariance between two datasets will be close to 0. Here, we select 5 data samples as the step size of the time delays. In practice, we should select the time delays empirically based on the characters of the sources, e.g., we suggest the time delays smaller than 100ms for physiological signals. In addition, we evaluate the role of the number of time delays and find that it has less impact on the performance. In this chapter, we set the number of time delays to 20.

We further show the performances in term of the PCC results between the original sources and the estimated ones. As shown in Table 4.2, the proposed method yields promising results when it is used to separate the latent and underdetermined mixtures. Compared to the classical JBSS method MCCA, the proposed UJBSS approach needs fewer number of observations in each dataset, while it yields a better performance.

4.8 A Case Study: Solve A Single Set UBSS Problem Based on UJBSS-m

In this section, we show that the proposed UJBSS-m method can be employed to solve a single set UBSS problem, the noise enhanced signal processing problem, with a superior performance. As in Simulation 1, we employ 5 real audio signals as the sources. These 5 audio signals are mixed into 4 observations with a mixing matrix $A^{(1)}$ whose elements follow the uniform distribution $U[-1,1]$. We generate three datasets as

$$\begin{aligned} X^{(1)} &= A^{(1)}S^{(1)} \\ X^{(2)} &= awgn(X^{(1)}, 20 \text{ dB}) \\ X^{(3)} &= awgn(X^{(1)}, 20 \text{ dB}), \end{aligned} \tag{4.33}$$

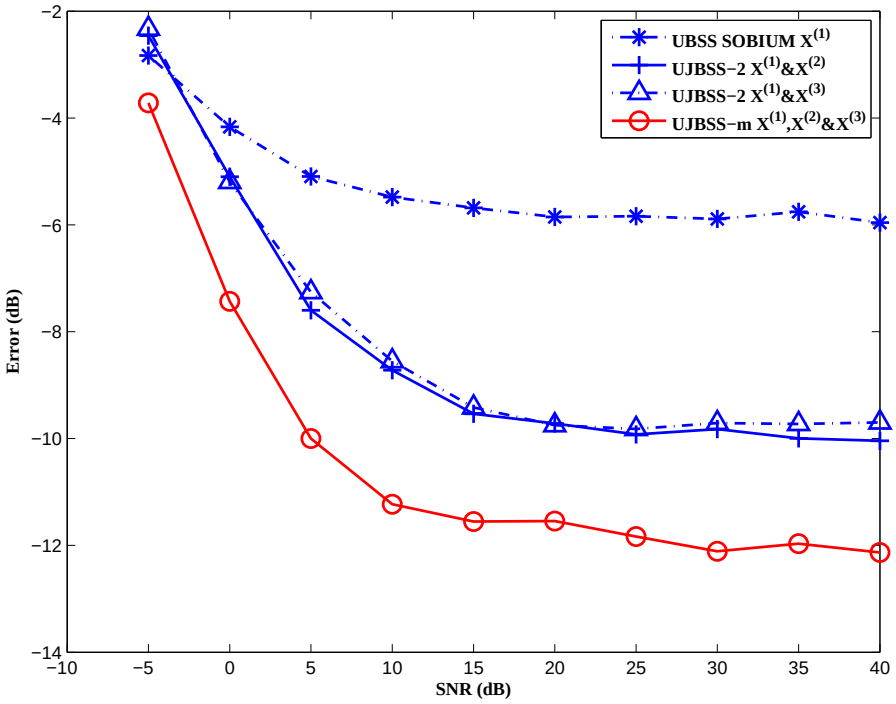

Figure 4.5: Simulation 2: performance comparisons on physiological signals between the proposed UJBSS-m method and two other methods (i.e., the single-set UBSS method SOBIUM [31] and UJBSS method for two dataset, UJBSS-2 [124]). Here the number of sources $N = 4$ and the number of observations $M = 3$. The number of time delays $L = 20$ and the step size of time delays (i.e., $\tau_l - \tau_{l-1}$) is 2 data samples. Similar results are observed for $A^{(2)}$ and $A^{(3)}$.

where $awgn(X^{(1)}, 20\text{ dB})$ represents adding white Gaussian noise to the signals $X^{(1)}$ (i.e., the real observations) with SNR of 20dB. Noise, traditionally regarded as the unwanted signal, can play a very important constructive role in estimation problems, which is known as noise enhanced signal processing. $X^{(2)}$ and $X^{(3)}$ are random noise added signals based on $X^{(1)}$.

The problem of interest here is to estimate the mixing matrix $A^{(1)}$. Traditionally, we can estimate $A^{(1)}$ from the dataset $X^{(1)}$ based on the single set UBSS

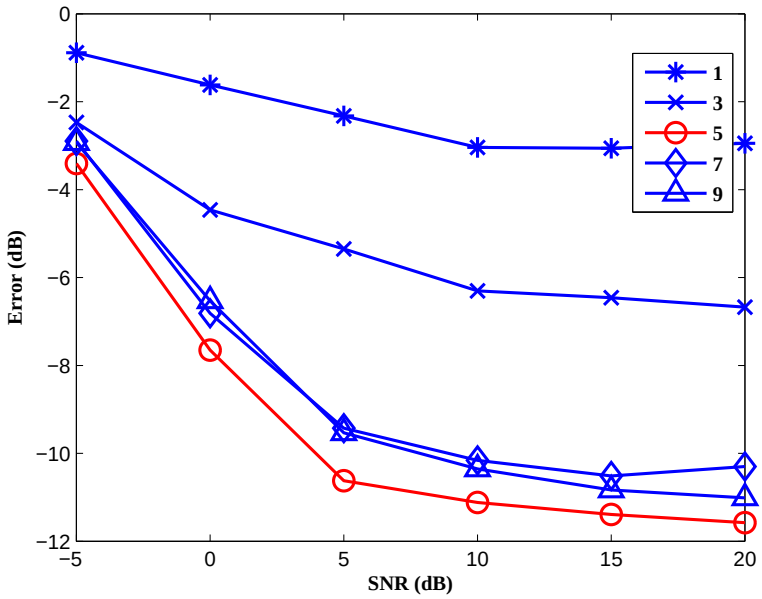

Figure 4.6: Simulation 2: performance of the proposed UJBSS-m method when the step size of time delays (i.e., $\tau_l - \tau_{l-1}$) varies from 1 to 9. Here the number of sources $N = 4$ and the number of observations $M = 3$. The number of time delays $L = 20$. Similar results are observed for $A^{(2)}$ and $A^{(3)}$.

method SOBIUM. Here we also can apply the proposed algorithm as mentioned in Section 4.5. Then we recover the sources via the method mentioned in Section 4.6 based on the estimated mixing matrix $A^{(1)}$. We repeat the experiment for 1000 times and calculate the sum of the absolute PCC (SAPCC) between the recovered sources and the original ones, which is calculated as

$$SAPCC = \sum_{n=1}^{5} abs(PCC(s_n^{(1)}, \hat{s}_n^{(1)})), \tag{4.34}$$

where $abs(\cdot)$ represents the absolute value function. Fig. 4.7 shows the distribution of the performance for 1000 repeats of the experiments. The average SAPCC for

Table 4.2: PCC performance results in Simulation 2.

	Methods	ECG	EOG	EEG	EMG
Dataset 1	UJBSS-m	0.985	0.994	0.910	0.831
	UJBSS-2[124]	0.975	0.986	0.872	0.781
	SOBIUM [31]	-0.815	-0.985	0.800	-0.382
	UBSS-SC[119]	0.090	0.949	-0.707	-0.679
	UBSS-SR[59]	0.504	-0.906	0.115	0.503
	MCCA*[72]	0.613	-0.754	0.695	0.674
Dataset 2	UJBSS-m	0.956	0.821	1.000	0.832
	UJBSS-2[124]	0.883	0.705	0.942	0.308
	SOBIUM [31]	0.851	-0.467	0.999	0.726
	UBSS-SC[119]	-0.420	0.544	0.776	0.726
	UBSS-SR[59]	-0.894	0.288	-0.967	-0.203
	MCCA*[72]	0.634	-0.754	0.678	-0.676
Dataset 3	UJBSS-m	0.779	0.738	0.998	0.997
	UJBSS-2[124]	0.777	0.738	0.998	0.997
	SOBIUM [31]	0.443	0.697	-0.993	-0.983
	UBSS-SC[119]	0.394	0.638	0.722	0.848
	UBSS-SR[59]	0.582	0.609	0.426	0.871
	MCCA*[72]	-0.624	-0.756	0.689	0.676

(1) *We add one additional observation in each dataset when we evaluate the MCCA.

UBSS is 4.53 while that for the proposed UJBSS-m is 4.76, even with the same source extraction technology. The one-way Analysis of Variance (ANOVA) is performed on the results provided by these 1000 repeats. The obtained p value is 1.5677e-11, which means that the results of the proposed UJBSS-m method and single set UBSS method are significantly different. The proposed UJBSS-m algorithm demonstrates more robust and better performance. This example illustrates that the estimation accuracy could be improved by adding suitable noises to the input signals.

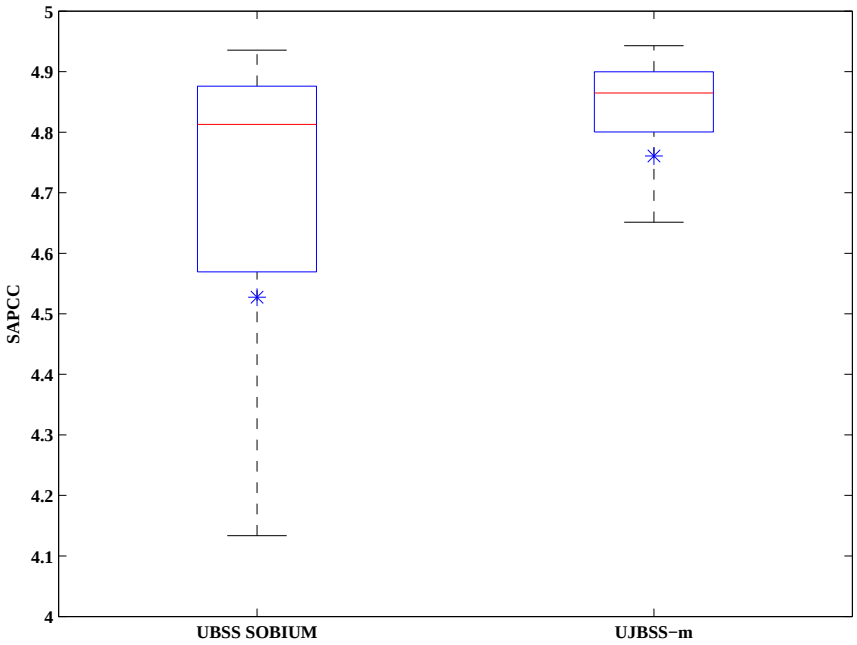

Figure 4.7: The sum of the absolute correlation coefficients between the recovered sources and the original ones. The blue asterisks represent the averages and the red lines stand for the medians. The edges of the box are the lower and upper quartiles.

4.9 Conclusions and Discussion

This chapter is the third work in my PhD study. Benefiting from the dependence information between two datasets, the UJBSS method proposed in chapter 3 gained promising performance. It is nature to generalize the idea for two datasets into multiple datasets. However, the UJBSS-2 for two datasets cannot be directly utilized to solve the problem in multiple datasets. In this chapter, we generalize the UJBSS for two datasets (i.e., UJBSS-2) to the case of multiple datasets. The basic idea is similar to that in UJBSS-2, which estimate the mixing matrices jointly first and then restore the source signals. In this chapter, we exploit the cross correlation of the observations between each pair of datasets and present a novel underdeter-

mined joint blind source separation method, namely UJBSS-m, to jointly estimate the mixing matrices from multiple datasets when the number of observations is smaller than that of the sources. The mixing matrices are accurately estimated through joint canonical polyadic decomposition of a sequence of specialized tensors in which a set of covariance matrices are stacked. Further the sources are recovered based on the estimated mixing matrices. Numerical results on multiple datasets demonstrate the superior performances of the proposed method when compared to the commonly used JBSS and single-set UBSS methods.

As an example application for noise enhanced signal processing, we also show that the proposed UJBSS-m method can be utilized to solve the single-set UBSS problem when suitable noise is added to the observations. In addition, the proposed UJBSS-m method does not rely upon sparsity of signals and therefore it can be applied to a wide class of signals when: 1) the sources within each dataset are uncorrelated and 2) the sources across different datasets are correlated only on corresponding indices.

Chapter 5

Removing Muscle Artifacts from EEG Data via Underdetermined Blind Source Separation

EEG recordings are often contaminated by artifacts from EMG. These artifacts reduce the quality of the EEG signals and disturb further analysis of EEG, such as in brain connectivity modeling. If enough number of EEG recordings are available, then there exists a considerable range of BSS methods which can suppress or remove the distorting effect of such artifacts. However, for many practical applications, such as the ambulatory health-care monitoring, the number of sensors used to collect EEG is limited. As a result, the conventional BSS methods, such as CCA and ICA, do not work in such cases. Considering the increasing need for biomedical signal processing in ambulatory environment, this chapter proposes a novel underdetermined BSS method exploring the cross correlation and autocorrelation of underlying sources. We evaluate the performance of our proposed method through numerical simulations in which EEG recordings are contaminated

with muscle artifacts. The results demonstrate that the proposed method can effectively and efficiently remove muscle artifacts meanwhile preserving the EEG activity successfully. This is a promising tool for real-world biomedical signal processing applications.

5.1 Motivation and Objectives

EEG is extensively used in brain science research, such as neuroscience and cognitive science [106]. However, it is susceptible to various physiological factors other than neural activities. ECG from cardiac activities, EOG from ocular movements, and EMG from muscular activities are the most common artifacts. These undesired artifacts interferer with the signal of interest and disturb subsequent analysis of EEG signals. Compared to other types of artifacts, it is generally more challenging to remove artifacts from the contracting head muscles (i.e. EMG signals) [79]. The main reasons for this difficulty are four folds: 1) EMG signals always have higher amplitude than the smaller EEG signals; 2) EMG signals have a wide spectral distribution, and especially overlaps with the beta activity in 15-30 Hz of EEG signals; 3) EMG signals have a broad anatomical distribution and can be detected across the entire scalp; 4) EMG signals also exhibit less repetition and consequently are harder to stereotype [18, 21, 106].

A artifacts removal is clearly an important issue and is a prerequisite step for our subsequent analysis. In order to remove the EMG artifacts, a number of approaches have been proposed, such as filtering, regression and EMD. An alternative strategy is based on BSS, which is more commonly used and demonstrated to be effective for removing artifacts from EEG. As one of the most popular BSS methods, ICA has been extensively explored for this purpose [21, 28], aiming to separate multichannel EEG into statistical independent components (ICs, i.e. underlying sources). Then the ICs determined to be artifacts can be discarded and the

remaining ICs can be used to reconstruct the artifact-free EEG. However, due to the issue of crosstalk between brain and muscle activity, ICs containing EEG signals are still contaminated by EMG [65, 83]. Therefore, ICA itself may not perform effectively in removing EMG from EEG.

Second order blind identification makes use of the temporal correlation and is shown as an effective alternative to ICA in removing EMG artifacts. However, this method is designed for stationary signal and it may suffer when the underlying source is nonstationary, such as in the case of transient muscular activities [23]. More recently, CCA has been explored as a more reliable method to remove EMG from scalp EEG. It aims to find mutually uncorrelated sources which are maximally autocorrelated. Compared with EEG, EMG has relatively low autocorrelation. Taking advantage of this distinguishable feature, sources with high autocorrelation should correspond to EEG while the source with relatively low autocorrelation is regarded as the EMG artifact. Then the underlying EMG components are discarded to reconstruct the EEG signals. As recently suggested in [21, 29, 43], CCA can achieve superior performance over ICA.

It is worth noting that the above mentioned BSS algorithms generally assume that the number of sources is equal to or less than that of the observations. However, for many practical applications, such as ambulatory health-care monitoring, it is desirable to collect the mixed signals using fewer sensors. In these cases, the above assumption does not hold and UBSS is required. Again, considering the increasing need for biomedical signal processing in ambulatory environment, this chapter proposes a novel UBSS method that investigates second-order statistics of the observations. Same as the existing UBSS methods, this proposed method consists of two steps: the mixing matrix is estimated first, followed by the separation of underlying sources based on the estimated result of the mixing matrix.

More specifically, inspired by stochastic resonance [78], we add tiny random

noise to the EEG recordings and construct multiple datasets across which the underlying sources are highly correlated. We further explore the cross correlation and autocorrelation of underlying sources, rather than solely the partial cross-correlation as in our previous paper [111, 124]. The mixing matrix is estimated accurately via joint polyadic tensor decompositions of a set of tensors where spatial covariance matrices corresponding to different time delays are stacked. Furthermore, the underlying sources, including EEG and EMG, are inferred based on the estimated mixing matrix from the EEG observations. Sources related to muscle activity are identified and removed during EEG reconstruction. We evaluate the performance of the proposed method through numerical simulations in which EEG recordings are contaminated with muscle artifacts. The results demonstrate that the proposed method can effectively and efficiently remove muscle artifacts while preserving the EEG successfully.

5.2 Problem Formulation

The problem of interest here is to recover the underlying N sources with a limited number of observations. In this study, EEG observation signals are denoted by a matrix $X(t) = [x_1(t); x_2(t); \ldots; x_M(t)] \in \mathbb{R}^{M \times T}$, where M represents the number of EEG observations and T is the number of data samples. It is assumed that the underlying N sources S, including the signal of interest (i.e. EEG) and undesired artifacts (i.e. EMG), are linearly mixed into the M observations X. The mixing process is modeled as follows,

$$X = AS + E. \tag{5.1}$$

$A \in \mathbb{R}^{M \times N}$ with $M < N$ (i.e., the underdetermined case) denotes the unknown mixing matrix. E is the possible additive noise which is generally assumed to be the

zero mean, temporally white and uncorrelated with the source signals.

5.3 Proposed Method

It is suggested in [111] that the UJBSS method can be used to solve the single-set UBSS problem. However, this only utilizes part of the cross-correlation [111] and neglects the autocorrelation and the other parts of cross-correlation. In this chapter, we fully exploit the cross-correlation and autocorrelation of the sources and propose a novel UBSS algorithm. We add tiny noise to the observations, X (which is also expressed as $X^{(1)}$), and construct the other two datasets as,

$$\begin{aligned} X^{(1)} =& X = [x_1^{(1)}, x_2^{(1)}, \ldots, x_N^{(1)}]^T \\ X^{(2)} =& awgn(X^{(1)}, 20dB) \\ X^{(3)} =& awgn(X^{(1)}, 20dB), \end{aligned} \tag{5.2}$$

in which $awgn(X^{(1)}, 20dB)$ adds white Gaussian noise to the measured EEG observation signals $X^{(1)}$, with the signal-to-noise ratio equaling to 20dB. (5.2) ensures dependence between each pair of datasets. Similar to CCA, it is reasonable to assume that:

(1) The sources in each dataset are uncorrelated.

$$\begin{aligned} E\{s_i^{(k)}(t)(s_j^{(k)}(t+\tau))^T\} = 0 \\ \forall \tau,\ 1 \le i \ne j \le N,\ k = 1,2,3, \end{aligned} \tag{I}$$

where $s_i^{(k)}(t)$ is the i-th source in dataset k and $s_j^{(k)}(t+\tau)$ represents the j-th source with the time delay τ in dataset k.

(2) The corresponding sources from two different datasets have non-zero cor-

relations and sources with different indices across datasets are not correlated.

$$\begin{aligned} D^{(\tau)} =& E\{S^{(k_1)}(t)(S^{(k_2)}(t+\tau))^T\} \\ =& Diag(\rho_1(\tau), \rho_2(\tau), \ldots, \rho_N(\tau)), \end{aligned} \tag{II}$$

where $Diag(\cdot)$ represents the diagonal matrix, and the $\rho_n(\tau) = E\{s_n^{(k_1)}(t)(s_n^{(k_2)}(t+\tau))^T\}$ denotes the covariance between $s_n^{(k_1)}(t)$ and $s_n^{(k_2)}(t+\tau)$. This assumption suggests that the corresponding sources in multiple datasets are second-order correlated with each other.

In this chapter, we fully exploit the second order auto covariance and cross covariance of EEG signals and propose a novel and effective algorithm to estimate the mixing matrix. The problem is reformulated as a joint canonical polyadic decomposition of a sequence of third-order tensors, which share the common factor matrix $A^{(1)}$. The auto covariance of the EEG signals can be formulated as

$$\begin{aligned} & E\{X^{(1)}(t)X^{(1)}(t+\tau)^T\} \\ =& (A^{(1)})E\{S^{(1)}(t)S^{(1)}(t+\tau)^T\}(A^{(1)})^T, \end{aligned} \tag{5.3}$$

where τ represents the time delay. The covariance matrices corresponding to different time delays, τ_1 to τ_L, satisfy,

$$\begin{aligned} B^{(1)} = \quad & E\{X^{(1)}(t)X^{(1)}(t+\tau_1)^T\} = (A^{(1)})C^{(\tau_1)}(A^{(1)})^T \\ & \vdots \\ B^{(L)} = \quad & E\{X^{(1)}(t)X^{(1)}(t+\tau_L)^T\} = (A^{(1)})C^{(\tau_L)}(A^{(1)})^T, \end{aligned} \tag{5.4}$$

in which $C^{(\tau_1)} = E\{S^{(1)}(t)S^{(1)}(t+\tau_l)^T\}$ is diagonal, $l = 1 \ldots L$. We stack the auto

covariance matrices $\{B^{(l)}\}$ in a tensor $\mathscr{B} \in \mathbb{R}^{M\times M\times L}$ as follows:

$$(\mathscr{B})_{i,j,l} = (B^{(l)})_{i,j}, \tag{5.5}$$

in which $i = 1,2,\ldots,M$, $j = 1,2,\ldots,M$, $l = 1,2,\ldots,L$. We define the matrix C of size $L \times N$ with the element $C_{l,n} = (C^{(\tau_l)})_{n,n}$, for l = 1, 2, ..., L, n = 1, 2, ..., N. Then we have:

$$\mathscr{B} = \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(1)} \circ c_n, \tag{5.6}$$

in which the $\circ$ denotes the outer product operation, $a_n^{(1)}$ is the n^{th} column of the mixing matrix $A^{(1)}$, and c_n is the n^{th} column of the matrix C.

The cross covariance between the EEG signals $X^{(1)}$ and the noise-added signals (i.e. $X^{(2)}$ and $X^{(3)}$) with time delay τ, can be formulated as:

$$\begin{aligned} &E\{X^{(1)}(t)X^{(k)}(t+\tau)^T\} \\ =&(A^{(1)})E\{S^{(1)}(t)S^{(k)}(t+\tau)^T\}(A^{(k)})^T \\ &E\{X^{(1)}(t+\tau)X^{(k)}(t)^T\} \\ =&(A^{(1)})E\{S^{(1)}(t+\tau)S^{(k)}(t)^T\}(A^{(k)})^T \end{aligned} \tag{5.7}$$

in which k = 2 or 3. Considering the correlations within and between each pair of datasets, the cross covariance matrices corresponding to time delay τ_l satisfy:

$$\begin{aligned} F^{(l)} &= E\{X^{(1)}(t)X^{(2)}(t+\tau_l)^T\} = (A^{(1)})G^{(\tau_l)}(A^{(2)})^T \\ H^{(l)} &= E\{X^{(1)}(t+\tau_l)X^{(2)}(t)^T\} = (A^{(1)})I^{(\tau_l)}(A^{(2)})^T \\ J^{(l)} &= E\{X^{(1)}(t)X^{(3)}(t+\tau_l)^T\} = (A^{(1)})K^{(\tau_l)}(A^{(3)})^T \\ P^{(l)} &= E\{X^{(1)}(t+\tau_l)X^{(3)}(t)^T\} = (A^{(1)})Q^{(\tau_l)}(A^{(3)})^T, \end{aligned} \tag{5.8}$$

in which the cross variance between the sources across each pair of datasets, $G^{(\tau_l)}, I^{(\tau_l)}, K^{(\tau_l)}, Q^{(\tau_l)}$,

are diagonal. Similar to (5.5), these sets of cross variance matrices are stacked into tensors $\mathscr{F}, \mathscr{H}, \mathscr{J}, \mathscr{P}$, which can be represented as,

$$\begin{aligned}
\mathscr{F} &= \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(2)} \circ g_n \\
\mathscr{H} &= \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(2)} \circ i_n \\
\mathscr{J} &= \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(3)} \circ k_n \\
\mathscr{P} &= \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(3)} \circ q_n.
\end{aligned} \tag{5.9}$$

Considering the common latent structure where each pair of tensors share the factor matrix $A^{(1)}$, the mixing matrix can be estimated via joint CPD of a collection of tensors $\mathscr{B}, \mathscr{F}, \mathscr{H}, \mathscr{J}, \mathscr{P}$. In this chapter, we generalize the idea of coupled matrix and tensor factorization (CMTF) and jointly decompose these tensors via the gradient-based optimization method [3, 100, 109]. The objective function can be expressed as:

$$\begin{aligned}
& f(A^{(1)}, A^{(2)}, A^{(3)}, C, G, I, K, Q) \\
=& \frac{1}{2}\|\mathscr{B} - [\![A^{(1)}, A^{(1)}, C]\!]\|^2 + \frac{1}{2}\|\mathscr{F} - [\![A^{(1)}, A^{(2)}, G]\!]\|^2 \\
+& \frac{1}{2}\|\mathscr{H} - [\![A^{(1)}, A^{(2)}, I]\!]\|^2 + \frac{1}{2}\|\mathscr{J} - [\![A^{(1)}, A^{(3)}, K]\!]\|^2 \\
+& \frac{1}{2}\|\mathscr{P} - [\![A^{(1)}, A^{(3)}, Q]\!]\|^2
\end{aligned} \tag{5.10}$$

where $[\![\cdot]\!]$ denotes the canonical polyadic approximation of a given tensor. This equation simultaneously takes the coupling information between different tensors into account. We propose to solve this problem via a gradient-based optimization method. The partial derivative of the objective function f with respect to each

column of $A^{(1)}$ is:

$$\begin{aligned}\frac{\partial f}{\partial a_n^{(1)}} =& -2\sum_{n=1}^{N}(\mathscr{B}\times_3 c_n)a_n^{(1)} + 2\sum_{d=1}^{N}(c_n^T c_d)((a_n^{(1)})^T a_d^{(1)})a_d^{(1)} \\ & -\mathscr{F}\times_2 a_n^{(2)}\times_3 g_n - \mathscr{H}\times_2 a_n^{(2)}\times_3 i_n - \mathscr{J}\times_2 a_n^{(3)}\times_3 k_n - \mathscr{P}\times_2 a_n^{(3)}\times_3 q_n \\ & +\sum_{d=1}^{N}[(a_n^{(2)})^T a_d^{(2)}(g_n)^T g_d + (a_n^{(2)})^T a_d^{(2)}(i_n)^T i_d + (a_n^{(3)})^T a_d^{(3)}(k_n)^T k_d \\ & +(a_n^{(3)})^T a_d^{(3)}(q_n)^T q_d]a_d^{(1)}.\end{aligned} \tag{5.11}$$

Similarly, we can calculate the partial derivative of f with respect to other factor matrices and obtain the gradient. Then the mixing matrix $A^{(1)}$ can be calculated based on any first-order optimization method. In this chapter, considering the competitive advantage of 'efficiency' and 'requiring less memory', we employ the nonlinear conjugate gradient algorithm (NCG) implemented in [38] to solve the unconstrained optimization problem and further estimate the mixing matrix $A^{(1)}$.

Once the mixing matrix is estimated, extracting the sources is a classic inverse problem. Here, we adopt a recently-developed subspace representation method [59] to recover the latent sources based on the estimated mixing matrix. The details of this method can be found in Chapter 4. Next, the recovered sources are sorted in terms of their autocorrelations. Due to the relatively low autocorrelation of EMG signals, muscle artifacts are isolated and set to 0 during reconstruction. Subsequently, the cleaned signals X_{eeg} can be obtained. The major steps of the proposed method are summarized in Algorithm 7.

5.4 Data Generation and Performance Indices

In order to evaluate the performance of the proposed method, obtaining the ground truth, i.e. the pure EEG and EMG signals, is quite necessary. In previous studies [29], to obtain the ground truth EEG signals, experienced neurophysiologist in-

Algorithm 7 The proposed method for removing muscle artifact from EEG signals

Input: M-dimensional observations X

Output: The artifact-free EEG data X_{eeg}.

1: Create the other two data via adding Gaussian white noise to the EEG observations X;
2: Calculate the auto covariance and the cross covariance as in (5.4) and (5.8) with different time delays, and construct a sequence of third-order tensors;
3: Calculate the joint CPD of the tensors constructed in step 2 and estimate the mixing matrix A (is also expressed as $A^{(1)}$);
4: Recover the underlying sources $S^{(1)}$, including EEG and EMG artifacts;
5: Sort the recovered sources $S^{(1)}$ in term of their autocorrelations and recognize the EMG artifacts from them;
6: Set the rows of $S^{(1)}$ corresponding to muscle artifacts to be zero and get $S^{(1)}_{new}$;
7: Reconstruct the artifact-free EEG signals X_{eeg} via $X_{eeg} = A^{(1)} S^{(1)}_{new}$.

spected many EEG recordings and select the clean EEG signals from them. However, frequent difficulties have surfaced in acquiring artifact-free EEG signals in reality, and it is even more difficult to ensure that the selected signals are completely free of muscle activities. In this section, we generate synthetic EEG and EMG signals, and examine the performance of the proposed method when the ground truth is available.

The simulated EEG sources are generated according to the phase-resetting theory proposed by Markinen et al. [75]. As in [117] and [21], we generate each EEG source by adding 4 sinusoids with frequencies randomly chosen from the range of 4 Hz to 30 Hz. To illustrate the performance of the proposed method, N EEG sources, S_{EEG}, are produced independently. Here, we set N to be 4. Analogous to the work of Delorme et al. [33], an EMG source, S_{EMG}, is simulated using random noise band-pass filtered between 20 and 60 Hz. The sampling rate here is 250 Hz and each channel is 40 seconds long. In addition, the 4-channel EEG observations are modeled as,

$$\begin{aligned} X &= AS \\ &= A[S_{EEG}; S_{EMG}], \end{aligned} \tag{5.12}$$

where S includes 4 EEG sources and 1 EMG source, and A is the mixing matrix generated randomly with elements following the uniform distribution $U[-1,1]$. For simplicity, each column of the mixing matrix is normalized into a unit vector.

To fairly compare the proposed method with existing EMG artifacts removal methods, 1000 independent simulations are implemented and three performance indices are employed. The first performance measurement is the mean relative estimation error of the mixing matrix A, which is defined as:

$$Error = 10log_{10}\{mean(\frac{||A-\hat{A}||}{||A||})\}, \tag{5.13}$$

where $\hat{A}$ denotes the optimally ordered estimate of A. The second measure is the Mean of Absolute Correlation (MAC) between the estimated sources and the original ones, which is defined as

$$MAC = mean(\frac{1}{N}\sum_{n=1}^{n=N}|\frac{cov(s_n,\hat{s}_n)}{\sigma_{s_n}\sigma_{\hat{s}_n}}|), \tag{5.14}$$

where $\hat{s}_n$ represents the estimate of the source s_n, $cov(\cdot,\cdot)$ represents the covariance between two variables and σ denotes the standard deviation. The Relative Root Mean Squared Error (RRMSE) is the third measure used to evaluate the effect of muscle artifact removal, which is defined as:

$$RRMSE = \frac{RMS(X_{EEG}-\hat{X}_{EEG})}{RMS(X_{EEG})}, \tag{5.15}$$

where $RMS(\cdot)$ denotes the root mean squared (RMS) value of a matrix/vector. For

instance, the RMS value of the EEG observations X is expressed as,

$$RMS(X) = \sqrt{\frac{1}{M \cdot T} \sum_{m=1}^{M} \sum_{t=1}^{T} X_{m,t}}, \tag{5.16}$$

where M is the number of EEG observations (which is 4 in this chapter), and T represents the number of data samples.

5.5 Numerical Study for the Synthetic EEG Data

The original sources in our study include 4 EEG sources (represented by S_{EEG}) and 1 EMG source (S_{EMG}), which are linearly mixed into 4 observations X following (5.1). The other two datasets, $X^{(2)}$ and $X^{(3)}$, are generated following (5.2). In our previous paper [111], we discussed the effect of the step size and the number of time delays. Considering the difference in the sampling rate, here we select 1 data sample corresponding to 4ms as the step size of time delays. Compared to the step size, the number of time delays has less impact on the performance. To further enhance the time efficiency, we set the number of time delays to 10.

Fig.5.1 shows the estimation error as a function of SNRS. We compare the proposed method with a commonly-used single-set UBSS method (SOBIUM) and the previously developed underdetermined joint BSS method (UJBSS-m) [31, 111]. SOBIUM exploits the autocorrelation of the sources and reformulate the problem of estimating the mixing matrix as decomposing a higher-order tensor. UJBSS-m models the cross correlation between each pair of two datasets and it is considered as a great alternative to SOBIUM in solving the single-set underdetermined BSS problem [111]. Compared with the SOBIUM and UJBSS-m, which only utilized autocorrelation or part of cross correlation, the proposed method fully exploits the second-order statistics of the observations. Benefiting from this, the proposed method consistently yields the best results over the entire SNRS range, which also

suggests the stability of the proposed method.

We estimate the mixing matrix A via the single-set UBSS method SOBIUM, the underdetermined joint BSS method UJBSS-m and our proposed method, and further recover the latent sources using the subspace representation method [59]. The source with the lowest autocorrelation is then recognized as the EMG source. Next the EMG source is set to 0 and the EEG sources are used to reconstruct the artifact-free EEG signals X_{EEG}. As an illustrative example, Fig.5.2 demonstrates the original X_{EEG} and the corresponding reconstruction results. It is shown that our proposed method is able to remove the EMG artifact perfectly and the reconstructed EEG signals are highly correlated with the original artifact-free EEG signals. In addition, we also compared the proposed method with a recently developed EMG artifact removal method, EEMD-CCA [21]. EEMD-CCA is a single-channel technique for muscle activity removal and is therefore suitable for removing artifacts with a limited number of observations. It was shown that this EEMD-CCA technique outperforms the multichannel technique based on CCA for removing muscle artifacts from EEG signals [18, 21]. In this chapter, we apply this EEMD-CCA method to each channel of the EEG observations X and decompose each channel into multiple IMFS, which are the input of the CCA method. Given the relative lower autocorrelation, the last canonical variate (CV, i.e. output of the CCA) is set to 0 during EEG reconstruction. We also test the effect of the number of canonical variates which are selected as the EMG artifacts and discarded during EEG reconstruction. The best performance is gained when the number of CV resmbling EMG is set to 1.

In this chapter, we repeat the experiments 1000 times and the resulting average performance is shown in Table 5.1. We further compare the proposed method with three other artifact removal methods, including SOBIUM, UJBSS-m and EEMD-CCA, in terms of MAC, RRMSE and time efficiency. All three of the compared

BSS methods utilize the same technology to recover the sources when the mixing matrix is estimated separately. Despite this, the performance of the proposed method is significantly better than that of SOBIUM and UJBSS-m. This also suggests the importance of estimating mixing matrices accurately. In addition, we test the computational cost of all four of these respective artifact removal methods. To remove the muscle artifact from 10000-datapoint 4-channel EEG observations, the average computational time for the proposed method is 52.810s while that of SOBIUM, UJBSS-m and EEMD-CCA are 11.195s, 26.907s and 52.910s respectively. The implementation is completed in MATLAB on a computer with Intel Core i7-4770 3.40 GHz CPU and 8.00G RAM. All the MATLAB codes used in this chapter are available upon request from the authors via email liangzou@ece.ubc.ca.

Table 5.1: Performance comparison between the proposed method and the other three methods (SOBIUM, UJBSS-m, EEMD-CCA)

	MAC	RRMSE	Average time cost (second)
SOBIUM [31]	0.863	0.147	11.195
UJBSS-m [111]	0.930	0.099	26.907
EEMD-CCA [21]	NA	0.168	52.910
The proposed method	0.940	0.085	52.810

5.6 Conclusions and Discussion

In this chapter, we propose an effective and novel method to remove muscle artifacts from EEG signals. Compared with SOBIUM and UJBSS-m, which only utilize autocorrelation or a portion of cross-correlation, our proposed method fully exploits the second-order statistics of observations. The mixing matrices are accurately estimated through joint CPD of a set of specialized tensors in which covariance matrices corresponding to different time delays are stacked. Subse-

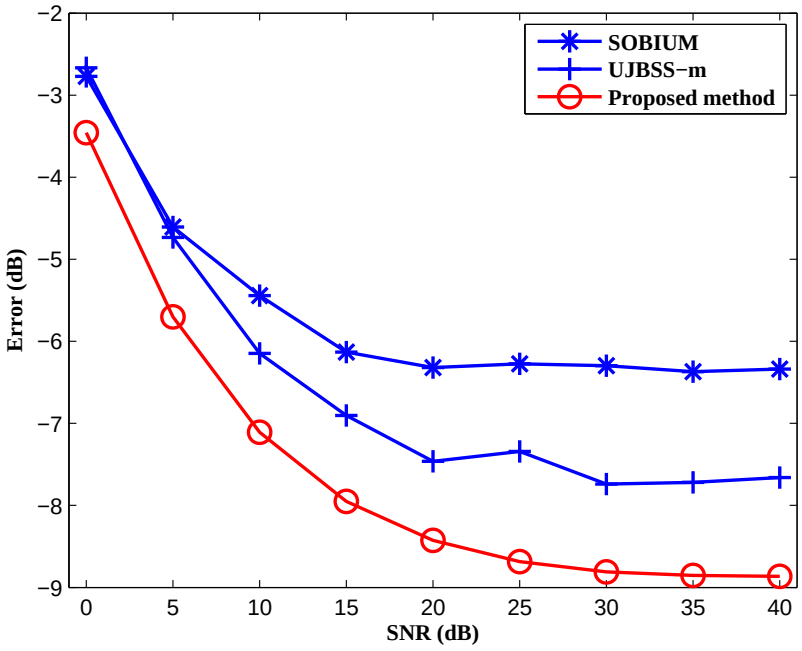

Figure 5.1: Estimation error of A with the change of signal-to-noise ratios. Here, the number of time delays L equals to 10, and the step size of time delays (i.e. $\tau_l - \tau_{l-1}$) is 1 data sample corresponding to 4 ms.

quently, sources are recovered based on these estimated mixing matrices. Compared with EEMD-CCA whose performance relies on the artifact level of the contaminated EEG signals (ratio between the power of pure EEG and EMG), the proposed method is based on the statistical properties of the underlying sources, and therefore is more robust. We evaluate the performance of the proposed method through numerical simulations in which EEG recordings are contaminated with muscle artifact. Our results demonstrate that the proposed method can effectively and efficiently remove muscle artifacts while preserving the EEG activity successfully. Therefore, it is a promising tool for real-world biomedical signal processing applications.

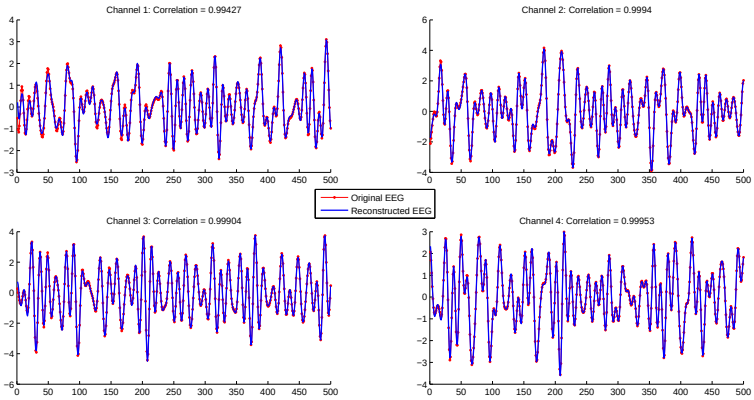

Figure 5.2: An illustrative example of the reconstructed EEG signals based on the proposed EMG removal method.

This is the last piece of my PhD work. It is a new and interesting application of UJBSS. To the best of our knowledge, we are the first to apply the underdetermined BSS method to remove EMG artifacts from a limited number of EEG observations. Further, we note that it is also applicable to remove other kinds of artifacts, such as ECG and EOG.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this dissertation, we have developed a set of novel underdetermined blind source separation approaches for recovering underlying sources when the number of sources is greater than that of the observations. The proposed algorithms aim to address several challenges in real applications, including limited number of observations, self/cross dependence information and source inference. The proposed methods were evaluated on synthetic data and/or real physiological signals. It should be noted that since the underlying ground truth is unavailable in real physiological studies, the performance evaluation in such cases basically depends on visual inspection by experts in the field. The main contributions and findings of this dissertation are summarized as follows.

In Chapter 2, a novel UBSS framework, termed NAMEMD-MCCA, is proposed to extract the heart beat signal from multi-channel NF-based sensor signals. Considering various potential artifacts residing in the measured signals, recovering the underlying heart beat signal is an underdetermined problem. We inves-

tigate state-of-the-art EMD-BSS based methods for exacting RHBR information accurately based on the nano-sensor data and further propose the NAMEMD-MCCA for improved RHBR monitoring. Considering inter-channel information, NAMEMD processes the input NF-based signals in high dimensional space and can effectively overcome the problems of uniqueness and mode mixing [77]. As an extension of CCA, MCCA is able to jointly extract sources through maximizing the correlations of the extracted sources across datasets. The combination of these two methods (NAMEMD-MCCA) benefits from the use of cross-channel information and increased robustness to artifacts. We first apply the proposed methods to synthetic data to illustrate their performance where the underlying truth is known. We then apply the proposed method to real nano-sensor data collected when the subject performs 11 tasks and it is shown that the proposed NAMEMD-MCCA method can achieve superior performance.

Another challenging question that we have posed in the underdetermined BSS field is the underdetermined joint BSS problem, which aims to jointly estimate the mixing matrices and/or extract the underlying source from multiple datasets when the number of sources is greater than that of observations in each dataset. Traditional joint BSS methods are designed for the determined case, which assume that the number of sources is equal to or less than that of the observations. As mentioned, this assumption may not be true in certain practical applications due to concerns including cost or time [60]. However, to the best of our knowledge, in the current literature only very limited work has been done on JBSS methods specifically designed for the underdetermined case. In order to address this concern, in Chapter 3, we propose an underdetermined JBSS for 2 datasets, termed as UJBSS-2. Considering the dependence information between two datasets, we exploit second-order statistics of the observations. The problem of jointly estimating mixing matrices is tackled via CPD of a specialized tensor in which a set of spatial covariance

matrices are stacked. Numerical results demonstrate the competitive performance of UJBSS-2 when compared to a commonly used JBSS method, MCCA, and the single-set UBSS method, UBSS-FAS.

In Chapter 4, we generalize the idea of UJBSS-2 for two datasets into multiple datasets [124]. In this work, we propose a novel and effective method to jointly estimate the mixing matrices for multiple datasets. Moving from our work in Chapter 3, here the dependence information is modeled in a set of three-order tensors, rather than one single tensor. Considering the latent common structure of these constructed tensors, we jointly estimate the mixing matrices via joint canonical polyadic decomposition of these specialized tensors. In order to accurately infer the source signals, we recover them by further utilizing a novel subspace representation based method. This proposed UJBSS-m method does not rely upon the sparsity of signals and therefore it can be applied to a wide class of signals. In addition, we also show that UJBSS-m can be utilized to solve the single-set UBSS problem when suitable noise is added to the observations. Numerical results on both audio and physiological signals demonstrate the superior performances of this proposed method.

In Chapter 5, we propose a novel underdetermined blind source separation method for removing muscle artifacts from EEG signals. EEG recordings are often contaminated by various artifacts, among which the artifact from EMG is particularly difficult to eliminate. Such EMG artifacts reduce the quality of EEG signals and disturb further analysis of EEG, as in brain connectivity modeling. If a high enough number of EEG recordings are available, we can remove or to some extent suppress the distortion effect of such artifacts via a considerable range of BSS methods, such as ICA and CCA. However, for many practical applications, such as ambulatory health-care monitoring, a small number of sensors used to collect EEG is preferred and conventional BSS methods like CCA and ICA, will fail. Consider-

ing the recent increasing need for biomedical signal processing in the ambulatory environment, we explore cross correlation and autocorrelation of the underlying sources and propose a novel underdetermined BSS method. We conduct a performance comparison through numerical simulations in which 4 EEG recordings are contaminated with 1 muscle artifact. It is demonstrated that the proposed method can effectively and efficiently remove the muscle artifact meanwhile successfully preserving the EEG activity.

6.2 Future Work

6.2.1 Multiple Datasets Generation

In Chapter 4 and Chapter 5, we demonstrate that underdetermined joint BSS methods can be utilized to solve single set underdetermined BSS problems. In order to ensure the relative high correlation between sources, we construct multiple datasets through adding a certain amount of weak Gaussian white noise (e.g., with SNR = 20dB) to the observations. Then we apply the proposed underdetermined joint BSS method to these noise assisted datasets. To make a fair comparison, we repeat the simulation 1000 times in both studies. It is demonstrated that the proposed methods gain better performance on average when we use these noise assisted datasets. For instance, the performance of the proposed method in Chapter 4 is better than that of SOBIUM with high confidence (80 percent probability). However, there is not currently a strict theoretical analysis on how to add noise to ensure obtaining better performance with a higher probability. Thus, one possible research direction is to investigate a more rigorous method to add assisting noise adaptively to the given observations.

6.2.2 Estimate the Number of Source Signals in Determined and Underdetermined BSS

The classical BSS problem includes two aspects of research: an estimation of the number of sources and source separation. The 'blind' aspect of BSS refers to the fact that there is generally no prior information available on the number of sources or on the mixing model. However, for conceptual and computational simplicity, most BSS algorithms usually require that the number of sources is specified in advance. ICA makes the assumption that the number of sources is not greater than that of observations. MCCA assumes that the number of sources equals to that of observations [56]. Further, the matrix diagonalization-based technique for the underdetermined BSS has an upper limit for the number of sources. We see that we can obtain a unique solution only when the number of sources satisfies certain conditions [31]. However, these assumptions may not be the case in reality. One classic example of the source separation problem is the cocktail party problem, which has been previously explained, where a number of people are talking simultaneously. Without prior knowledge of the number of sources, we do not know which BSS method is most suitable for solving this problem given the recording signals. In addition, results can vary if the number of sources is set differently. In different scenarios, we may need to choose different types of BSS methods.

An accurate estimation of the number of sources is shown to be of high importance, and it is generally estimated before further source separation. As a result, several approaches have been proposed for estimating the number of sources. Wax and Kailath introduced the eigenvalue-based estimation method and investigated the observation that the number of dominant eigenvalues of the correlation matrix is equal to the number of sources in determined cases [41]. This method was improved by introducing Akaike Information Criterion (AIC) and Minimum Description Length (MDL) and some other measures in the estimation [40, 73, 94].

However, it is challenging to estimate the true number of sources when the observations are noisy. To address this concern, sources number estimation methods which are robust to noisy observations are highly desired.

In addition, to the best of our knowledge, there are only a few researchers discussing different ways to estimate the number of sources in UBSS, and they generally make use of the sparsity of source signals. For instance, SSPS on the time domain or frequency domains are detected and similarity-based clustering methods are used to estimate the number of sources. However, signals may not be as sparse as assumed in the existing methods. Therefore, it is necessary to relax the sparsity constraint. Furthermore, the existing methods for estimating the number of sources for UBSS cases are limited to instantaneous mixing models. We plan to develop advanced methods to estimate the number of sources when the sources are mixed in a convolutional model. Lastly, the number of sources may be related to certain physiological processes, such as depth of sleep. It would thus be of interest in some applications to monitor the dynamic change of the number of sources and the corresponding mixing structure.

6.2.3 Online Underdetermined BSS

For many practical applications, such as ambulatory health-care monitoring, it is desirable to collect mixed signals using fewer sensors. In order to recover sources or remove unwanted noise, underdetermined BSS is preferred in these situations. Generally underdetermined BSS is more difficult to implement due to the lower number of available observations. Underdetermined BSS methods generally consist of two separate steps: mixing matrix estimation and underlying source inference, which are very time consuming. To better serve such practical applications, real-time underdetermined BSS methods with light computational complexity are preferred.

In addition, most existing BSS algorithms assume that the sources are physically stationary, i.e., mixing filters are fixed. However, this assumption does not always hold in real applications. For instance, in the cocktail party problem, it is highly possible that both sources and sensors are not stationary in the room and therefore the mixing model may be time varying. In these situations, it is necessary to model the mixing matrix as time-varying and develop UBSS methods accordingly.

Bibliography

[1] F. Abrard and Y. Deville. A time–frequency blind signal separation method applicable to underdetermined mixtures of dependent sources. *Signal Processing*, 85(7):1389–1403, 2005. → pages 7

[2] E. Acar, D. M. Dunlavy, and T. G. Kolda. A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics*, 25(2):67–86, 2011. → pages 67, 71, 73

[3] E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*, 2011. → pages 77, 103

[4] T. Adali, M. Anderson, and G.-S. Fu. Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging. *IEEE Signal Processing Magazine*, 31(3):18–33, 2014. → pages 64

[5] T. Adali, Y. Levin-Schwartz, and V. D. Calhoun. Multimodal data fusion using source separation: Two effective models based on ica and iva and their properties. *Proceedings of the IEEE*, 103(9):1478–1493, 2015. → pages 50, 65

[6] M. Aharon, M. Elad, and A. Bruckstein. *rmk*-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. → pages 9

[7] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier. Underdetermined blind separation of nondisjoint sources in the time-frequency domain. *IEEE Transactions on Signal Processing*, 55(3): 897–907, 2007. → pages 7

[8] M. Anderson, T. Adali, and X.-L. Li. Joint blind source separation with multivariate gaussian model: Algorithms and performance analysis. *IEEE Transactions on Signal Processing*, 60(4):1672–1683, 2012. → pages 10

[9] S. Arberet, R. Gribonval, and F. Bimbot. A robust method to count and locate audio sources in a stereophonic linear anechoic mixture. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 3, pages III–745. IEEE, 2007. → pages 7

[10] B. Arons. A review of the cocktail party effect. *Journal of the American Voice I/O Society*, 12(7):35–50, 1992. → pages 2

[11] M. Bin Altaf, T. Gautama, T. Tanaka, and D. P. Mandic. Rotation invariant complex empirical mode decomposition. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 3, pages III–1009. IEEE, 2007. → pages 25

[12] B. Bouachache and P. Flandrin. Wigner-ville analysis of time-varying signals. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, volume 7, pages 1329–1332, May 1982. → pages 54

[13] L. Brechet, M.-F. Lucas, C. Doncarli, and D. Farina. Compression of biomedical signals with mother wavelet optimization and best-basis wavelet packet selection. *Biomedical Engineering, IEEE Transactions on*, 54(12):2186–2192, 2007. → pages 21

[14] V. D. Calhoun, J. Liu, and T. Adalı. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage*, 45 (1):S163–S172, 2009. → pages 10, 11

[15] A. T. Cemgil, C. Févotte, and S. J. Godsill. Variational and stochastic inference for bayesian source separation. *Digital Signal Processing*, 17(5): 891–913, 2007. → pages 9

[16] J.-C. Chao. *On the design of robust criteria and algorithms for blind source separation*. PhD thesis, Southern Methodist University, 2007. → pages 5

[17] X. Chen, C. He, Z. J. Wang, and M. J. McKeown. An ic-pls framework for group corticomuscular coupling analysis. *Biomedical Engineering, IEEE Transactions on*, 60(7):2022–2033, 2013. → pages 50

[18] X. Chen, C. He, and H. Peng. Removal of muscle artifacts from single-channel eeg based on ensemble empirical mode decomposition and multiset canonical correlation analysis. *Journal of Applied Mathematics*, 2014, 2014. → pages 97, 108

[19] X. Chen, A. Liu, M. J. McKeown, H. Poizner, and Z. J. Wang. An eemd-iva framework for concurrent multidimensional eeg and unidimensional kinematic data analysis. *IEEE Transactions on Biomedical Engineering*, 61(7):2187–2198, 2014. → pages 11, 22

[20] X. Chen, Z. J. Wang, and M. J. McKeown. A three-step multimodal analysis framework for modeling corticomuscular activity with application to parkinsons disease. *Biomedical and Health Informatics, IEEE Journal of*, 18(4):1232–1241, 2014. → pages 38, 47, 60, 88

[21] X. Chen, A. Liu, J. Chiang, Z. J. Wang, M. J. McKeown, and R. K. Ward. Removing muscle artifacts from eeg data: Multichannel or single-channel techniques? *IEEE Sensors Journal*, 16(7):1986–1997, 2016. → pages 97, 98, 105, 108, 109

[22] X. Chen, Z. J. Wang, and M. McKeown. Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods. *IEEE Signal Processing Magazine*, 33(3):86–107, 2016. → pages 64

[23] S. Choi, A. Cichocki, and A. Beloucharni. Second order nonstationary source separation. *The Journal of VLSI Signal Processing*, 32(1):93–104, 2002. → pages 98

[24] A. Cichocki and S.-i. Amari. *Adaptive blind signal and image processing: learning algorithms and applications*, volume 1. John Wiley & Sons, 2002. → pages 5

[25] P. Comon and C. Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010. → pages 5

[26] M. Congedo, R. Phlypo, and J. Chatel-Goldman. Orthogonal and non-orthogonal joint blind source separation in the least-squares sense. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1885–1889. IEEE, 2012. → pages 66

[27] N. M. Correa, T. Adali, Y.-O. Li, and V. D. Calhoun. Canonical correlation analysis for data fusion and group inferences. *Signal Processing Magazine, IEEE*, 27(4):39–50, 2010. → pages 59, 87

[28] M. Crespo-Garcia, M. Atienza, and J. L. Cantero. Muscle artifact removal from human sleep eeg by using independent component analysis. *Annals of biomedical engineering*, 36(3):467–475, 2008. → pages 97

[29] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel. Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE transactions on Biomedical Engineering*, 53(12):2583–2587, 2006. → pages 21, 98, 104

[30] L. De Lathauwer. A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 28(3):642–666, 2006. → pages 54, 73

[31] L. De Lathauwer and J. Castaing. Blind identification of underdetermined mixtures by simultaneous matrix diagonalization. *Signal Processing, IEEE Transactions on*, 56(3):1096–1105, 2008. → pages xiii, 7, 8, 51, 54, 58, 65, 66, 74, 84, 86, 89, 91, 93, 107, 109, 116

[32] L. De Lathauwer, J. Castaing, and J.-F. Cardoso. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing*, 55(6):2965–2973, 2007. → pages 8

[33] A. Delorme, T. Sejnowski, and S. Makeig. Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. *Neuroimage*, 34(4):1443–1449, 2007. → pages 105

[34] I. Domanov and L. De Lathauwer. Canonical polyadic decomposition of third-order tensors: relaxed uniqueness conditions and algebraic algorithm. *Linear Algebra and its Applications*, 513:342–375, 2017. → pages 73

[35] I. Domanov and L. D. Lathauwer. Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition. *SIAM Journal on Matrix Analysis and Applications*, 35(2):636–660, 2014. → pages 73

[36] I. Domanov and L. D. Lathauwer. Generic uniqueness conditions for the canonical polyadic decomposition and indscal. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1567–1589, 2015. → pages 73

[37] T. Dong, Y. Lei, and J. Yang. An algorithm for underdetermined mixing matrix estimation. *Neurocomputing*, 104:26–34, 2013. → pages 7

[38] D. M. Dunlavy, T. G. Kolda, and E. Acar. Poblano v1. 0: A matlab toolbox for gradient-based optimization. *Sandia National Laboratories, Albuquerque, NM and Livermore, CA, Tech. Rep. SAND2010-1422*, 2010. → pages 78, 104

[39] J. Escudero, R. Hornero, D. Abásolo, and A. Fernández. Quantitative evaluation of artifact removal in real magnetoencephalogram signals with blind source separation. *Annals of biomedical engineering*, 39(8): 2274–2286, 2011. → pages 22

[40] E. Fishler and H. V. Poor. Estimation of the number of sources in unbalanced arrays via information theoretic criteria. *IEEE Transactions on Signal Processing*, 53(9):3543–3553, 2005. → pages 116

[41] E. Fishler, M. Grosmann, and H. Messer. Detection of signals by information theoretic criteria: General asymptotic performance analysis. *IEEE Transactions on Signal Processing*, 50(5):1027–1036, 2002. → pages 116

[42] P. Flandrin, G. Rilling, and P. Goncalves. Empirical mode decomposition as a filter bank. *Signal Processing Letters, IEEE*, 11(2):112–114, 2004. → pages 40

[43] J. Gao, C. Zheng, and P. Wang. Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis. *Clinical EEG and neuroscience*, 41(1):53–59, 2010. → pages 98

[44] S. Ge, Q. Yang, R. Wang, P. Lin, J. Gao, Y. Leng, Y. Yang, and H. Wang. A brain-computer interface based on a few-channel eeg-fnirs bimodal system. *IEEE Access*, 5:208–218, 2017. → pages 64

[45] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley. Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. → pages 60, 88

[46] X.-F. Gong, X.-L. Wang, and Q.-H. Lin. Generalized non-orthogonal joint diagonalization with lu decomposition and successive rotations. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 63(5), 2015. → pages 66, 74

[47] R. Gribonval and S. Lesage. A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges. In *ESANN'06 proceedings-14th European Symposium on Artificial Neural Networks*, pages 323–330. d-side publi., 2006. → pages 4

[48] A. R. Groves, C. F. Beckmann, S. M. Smith, and M. W. Woolrich. Linked independent component analysis for multimodal data fusion. *Neuroimage*, 54(3):2198–2217, 2011. → pages 65

[49] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4): 321–377, 1936. → pages 10, 23, 27

[50] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 454, pages 903–995. The Royal Society, 1998. → pages 21

[51] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000. → pages 6

[52] Y. Ichimaru and G. Moody. Development of the polysomnographic database on cd-rom. *Psychiatry and Clinical Neurosciences*, 53(2): 175–177, 1999. → pages 32

[53] M. T. Jensen, J. L. Marott, P. Lange, J. Vestbo, P. Schnohr, O. W. Nielsen, J. S. Jensen, and G. B. Jensen. Resting heart rate is a predictor of mortality in copd. *European Respiratory Journal*, 42(2):341–349, 2013. → pages 20

[54] S. Junnila, H. Kailanto, J. Merilahti, A.-M. Vainio, A. Vehkaoja, M. Zakrzewski, and J. Hyttinen. Wireless, multipurpose in-home health monitoring platform: Two case trials. *Information Technology in Biomedicine, IEEE Transactions on*, 14(2):447–455, 2010. → pages 20

[55] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont. Multiparameter respiratory rate estimation from the photoplethysmogram. *Biomedical Engineering, IEEE Transactions on*, 60(7):1946–1953, 2013. → pages 32

[56] J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, pages 433–451, 1971. → pages 10, 27, 50, 64, 116

[57] H. A. Kiers. Towards a standardized notation and terminology in multiway analysis. *Journal of chemometrics*, 14(3):105–122, 2000. → pages 67, 69

[58] S. Kim and C. D. Yoo. Underdetermined blind source separation based on generalized gaussian distribution. In *2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pages 103–108. IEEE, 2006. → pages 80

[59] S. Kim and C. D. Yoo. Underdetermined blind source separation based on subspace representation. *IEEE Transactions on Signal processing*, 57(7): 2604–2614, 2009. → pages 9, 65, 79, 80, 85, 86, 89, 93, 104, 108

[60] M. Kleinsteuber and H. Shen. Blind source separation with compressively sensed linear mixtures. *Signal Processing Letters, IEEE*, 19(2):107–110, 2012. → pages 51, 65, 113

[61] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009. → pages 53, 54, 67, 69, 71, 73

[62] Z. Koldovsky, P. Tichavsk, A. H. Phan, and A. Cichocki. A two-stage mmse beamformer for underdetermined signal separation. *IEEE Signal Processing Letters*, 20(12):1227–1230, 2013. ISSN 1070-9908. doi:10.1109/LSP.2013.2285932. → pages 51, 65, 71

[63] E. Kristal-Boneh, H. Silber, G. Harari, and P. Froom. The association of resting heart rate with cardiovascular, cancer and all-cause mortality. eight year follow-up of 3527 male israeli employees (the cordis study). *European heart journal*, 21(2):116–124, 2000. → pages 20

[64] J. B. Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977. → pages 72

[65] D. Labate, F. La Foresta, G. Morabito, I. Palamara, and F. C. Morabito. Entropic measures of eeg complexity in alzheimer's disease through a multivariate multiscale approach. *IEEE Sensors Journal*, 13(9):3284–3292, 2013. → pages 98

[66] J. Lee, D. D. McManus, S. Merchant, and K. H. Chon. Automatic motion and noise artifact detection in holter ecg data using empirical mode decomposition and statistical approaches. *Biomedical Engineering, IEEE Transactions on*, 59(6):1499–1506, 2012. → pages 21

[67] J.-H. Lee, T.-W. Lee, F. A. Jolesz, and S.-S. Yoo. Independent vector analysis (iva): multivariate approach for fmri group study. *Neuroimage*, 40 (1):86–109, 2008. → pages 9, 10

[68] X.-L. Li, M. Anderson, and T. Adalı. Second and higher-order correlation analysis of multiple multidimensional variables by joint diagonalization. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 197–204. Springer, 2010. → pages 10

[69] X.-L. Li, T. Adalı, and M. Anderson. Joint blind source separation by generalized joint diagonalization of cumulant matrices. *Signal Processing*, 91(10):2314–2322, 2011. → pages 50, 60, 64, 65, 70, 88

[70] Y. Li, A. Cichocki, and S.-i. Amari. Analysis of sparse representation and blind source separation. *Neural computation*, 16(6):1193–1234, 2004. → pages 9

[71] Y. Li, S.-I. Amari, A. Cichocki, D. W. Ho, and S. Xie. Underdetermined blind source separation based on sparse representation. *IEEE Transactions on signal processing*, 54(2):423–437, 2006. → pages 7, 9

[72] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun. Joint blind source separation by multiset canonical correlation analysis. *Signal Processing, IEEE Transactions on*, 57(10):3918–3929, 2009. → pages 10, 23, 27, 47, 50, 59, 60, 61, 64, 70, 85, 88, 89, 93

[73] A. P. Liavas and P. A. Regalia. On the behavior of information theoretic criteria for model order selection. *IEEE Transactions on Signal Processing*, 49(8):1689–1695, 2001. → pages 116

[74] J. Lin and A. Zhang. Fault feature separation using wavelet-ica filter. *NDT & E International*, 38(6):421–427, 2005. → pages 22

[75] V. Mäkinen, H. Tiitinen, and P. May. Auditory event-related responses are generated independently of ongoing brain activity. *Neuroimage*, 24(4): 961–968, 2005. → pages 105

[76] D. Mandic et al. Filter bank property of multivariate empirical mode decomposition. *Signal Processing, IEEE Transactions on*, 59(5): 2421–2426, 2011. → pages 22

[77] D. P. Mandic, N. U. Rehman, Z. Wu, and N. E. Huang. Empirical mode decomposition-based time-frequency analysis of multivariate signals: the power of adaptive data analysis. *Signal Processing Magazine, IEEE*, 30(6): 74–86, 2013. → pages 21, 23, 113

[78] M. D. McDonnell and D. Abbott. What is stochastic resonance? definitions, misconceptions, debates, and its relevance to biology. *PLoS Comput Biol*, 5(5):e1000348, 2009. → pages 98

[79] B. W. McMenamin, A. J. Shackman, L. L. Greischar, and R. J. Davidson. Electromyogenic artifacts and electroencephalographic inferences revisited. *Neuroimage*, 54(1):4–9, 2011. → pages 97

[80] L. Mesin, A. Holobar, and R. Merletti. Blind source separation: Application to biomedical signals. 2011. → pages 5

[81] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel. Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis. *Biomedical Engineering, IEEE Transactions on*, 57(9):2188–2196, 2010. → pages 21, 22, 27, 28

[82] G. B. Moody and R. G. Mark. The mit-bih arrhythmia database on cd-rom and software for use with it. In *Computers in Cardiology 1990, Proceedings.*, pages 185–188. IEEE, 1990. → pages 32

[83] H. Nam, T.-G. Yim, S. K. Han, J.-B. Oh, and S. K. Lee. Independent component analysis of ictal eeg in medial temporal lobe epilepsy. *Epilepsia*, 43(2):160–164, 2002. → pages 98

[84] D. Nion, K. N. Mokios, N. D. Sidiropoulos, and A. Potamianos. Batch and adaptive parafac-based blind separation of convolutive speech mixtures. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6): 1193–1207, 2010. → pages 8

[85] P. Palatini, E. Casiglia, P. Pauletto, J. Staessen, N. Kaciroti, and S. Julius. Relationship of tachycardia with high blood pressure and metabolic abnormalities a study with mixture analysis in three populations. *Hypertension*, 30(5):1267–1273, 1997. → pages 20

[86] P. Palatini, E. Casiglia, S. Julius, and A. C. Pessina. High heart rate: a risk factor for cardiovascular death in elderly men. *Archives of internal medicine*, 159(6):585–592, 1999. → pages 20

[87] C. Park, D. Looney, A. Ahrabian, D. Mandic, et al. Classification of motor imagery bci using multivariate empirical mode decomposition. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 21(1): 10–22, 2013. → pages 22

[88] M. Rajih, P. Comon, and R. A. Harshman. Enhanced line search: A novel method to accelerate parafac. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1128–1147, 2008. → pages 54

[89] N. Rehman and D. P. Mandic. Multivariate empirical mode decomposition. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, page rspa20090502. The Royal Society, 2009. → pages 25

[90] N. U. Rehman and D. P. Mandic. Filter bank property of multivariate empirical mode decomposition. *Signal Processing, IEEE Transactions on*, 59(5):2421–2426, 2011. → pages 23, 25, 26, 40, 46

[91] V. G. Reju, S. N. Koh, and Y. Soon. An algorithm for mixing matrix estimation in instantaneous blind source separation. *Signal Processing*, 89 (9):1762–1773, 2009. → pages 8, 65

[92] G. Rilling, P. Flandrin, P. Gonçalves, and J. M. Lilly. Bivariate empirical mode decomposition. *Signal Processing Letters, IEEE*, 14(12):936–939, 2007. → pages 25

[93] D. Safieddine, A. Kachenoura, L. Albera, G. Birot, A. Karfoul, A. Pasnicu, A. Biraben, F. Wendling, L. Senhadji, and I. Merlet. Removal of muscle artifact from eeg data: comparison between stochastic (ica and cca) and deterministic (emd and wavelet-based) approaches. *EURASIP Journal on Advances in Signal Processing*, 2012(1):1–15, 2012. → pages 21

[94] H. Sawada, R. Mukai, S. Araki, and S. Makino. Estimating the number of sources using independent component analysis. *Acoustical science and technology*, 26(5):450–452, 2005. → pages 116

[95] H. Sawada, S. Araki, and S. Makino. Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(3):516–527, 2011. → pages 13

[96] T. Shany, S. J. Redmond, M. R. Narayanan, and N. H. Lovell. Sensors-based wearable systems for monitoring of human movement and falls. *Sensors Journal, IEEE*, 12(3):658–670, 2012. → pages 20

[97] H. Snoussi and J. Idier. Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures. *IEEE Transactions on Signal Processing*, 54(9):3257–3269, 2006. → pages 9

[98] S. Soltanian, A. Servati, R. Rahmanian, F. Ko, and P. Servati. Highly piezoresistive compliant nanofibrous sensors for tactile and epidermal electronic applications. *Journal of Materials Research*, 30(01):121–129, 2015. → pages 20, 34

[99] L. Sorber, M. Van Barel, and L. De Lathauwer. Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-(l_r,l_r,1) terms, and a new generalization. *SIAM Journal on Optimization*, 23(2):695–720, 2013. → pages 73

[100] L. Sorber, M. Van Barel, and L. De Lathauwer. Structured data fusion. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):586–600, 2015. → pages 77, 103

[101] M. Sørensen, I. Domanov, and 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):1015–1045, 2015. → pages 77

[102] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2003. ISBN 9780961408893. → pages 80

[103] K. T. Sweeney, S. F. McLoone, and T. E. Ward. The use of ensemble empirical mode decomposition with canonical correlation analysis as a novel artifact removal technique. *Biomedical Engineering, IEEE Transactions on*, 60(1):97–105, 2013. → pages 22, 28

[104] P. Tichavsky and Z. Koldovsky. Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources. *IEEE Transactions on Signal Processing*, 59(3):1037–1047, 2011. → pages 65

[105] N. ur Rehman, C. Park, N. E. Huang, and D. P. Mandic. Emd via memd: multivariate noise-aided computation of standard emd. *Advances in Adaptive Data Analysis*, 5(02), 2013. → pages 29

[106] J. A. Urigüen and B. Garcia-Zapirain. Eeg artifact removalstate-of-the-art and guidelines. *Journal of neural engineering*, 12(3):031001, 2015. → pages 97

[107] H. L. Van Trees. *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons, 2004. → pages 79

[108] R. R. Vázquez, H. Velez-Perez, R. Ranta, V. L. Dorr, D. Maquin, and L. Maillard. Blind source separation, wavelet denoising and discriminant analysis for eeg artefacts and noise cancelling. *Biomedical Signal Processing and Control*, 7(4):389–400, 2012. → pages 21

[109] N. Vervliet, O. Debals, and L. De Lathauwer. Tensorlab 3.0-numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization. In *2016 Conference Record of the 50th Asilomar Conference on Signals, Systems and Computers. IEEE*, 2016. → pages 77, 78, 103

[110] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE transactions on audio, speech, and language processing*, 14(4):1462–1469, 2006. → pages 5

[111] L. Z. Wang, X. Chen, X. Ji, and Z. J. Underdetermined joint blind source separation of multiple datasets. *IEEE Access*, 5:7474–7487, 2017. → pages 8, 99, 100, 107, 109

[112] E. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40(5):749–759, 1932. → pages 54

[113] Z. Wu and N. E. Huang. A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611, 2004. → pages 24

[114] Z. Wu and N. E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1 (01):1–41, 2009. → pages 22

[115] G. Wunder, H. Boche, T. Strohmer, and P. Jung. Sparse signal processing concepts for efficient 5g system design. *IEEE Access*, 3:195–208, 2015. → pages 65

[116] S. Xie, L. Yang, J.-M. Yang, G. Zhou, and Y. Xiang. Time-frequency approach to underdetermined blind source separation. *Neural Networks and Learning Systems, IEEE Transactions on*, 23(2):306–316, 2012. → pages 51, 54, 55, 56, 59, 60, 61, 65, 66, 79, 86

[117] N. Yeung, R. Bogacz, C. B. Holroyd, S. Nieuwenhuis, and J. D. Cohen. Theta phase resetting and the error-related negativity. *Psychophysiology*, 44(1):39–49, 2007. → pages 105

[118] T. Yilmaz, R. Foster, and Y. Hao. Detecting vital signs with wearable wireless sensors. *Sensors*, 10(12):10837–10862, 2010. → pages 20

[119] L. Zhen, D. Peng, Z. Yi, Y. Xiang, and P. Chen. Underdetermined blind source separation using sparse coding. *IEEE Transactions on Neural Networks and Learning Systems*, 2016. → pages 9, 65, 85, 89, 93

[120] G. Zhou and A. Cichocki. Canonical polyadic decomposition based on a single mode blind source separation. *Signal Processing Letters, IEEE*, 19 (8):523–526, 2012. → pages 53, 71

[121] G. 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*, PP(99): 1–14, 2015. ISSN 2162-237X. doi:10.1109/TNNLS.2015.2487364. → pages 50

[122] G. Zhou, Q. Zhao, Y. Zhang, T. Adali, S. Xie, and A. Cichocki. Linked component analysis from matrices to high-order tensors: Applications to biomedical data. *Proceedings of the IEEE*, 104(2):310–331, 2016. ISSN 0018-9219. doi:10.1109/JPROC.2015.2474704. → pages 50, 51, 64

[123] L. Zou, X. Chen, A. Servati, P. Servati, and M. J. McKeown. A heart beat rate detection framework using multiple nanofiber sensor signals. In *Signal and Information Processing (ChinaSIP), 2014 IEEE China Summit & International Conference on*, pages 242–246. IEEE, 2014. → pages 28, 30

[124] L. Zou, X. Chen, and Z. J. Wang. Underdetermined joint blind source separation for two datasets based on tensor decomposition. *IEEE Signal Processing Letters*, 23(5):673–677, 2016. → pages xiii, 8, 63, 65, 74, 84, 85, 86, 89, 91, 93, 99, 114

[125] L. Zou, Z. J. Wang, X. Chen, and X. Ji. Underdetermined joint blind source separation based on tensor decomposition. In *Electrical and Computer Engineering (CCECE), 2016 IEEE Canadian Conference on*, pages 1–4. IEEE, 2016. → pages 66

Appendix A

Derivations

The Appendix is the proof of **Proposition** 1 in Chapter 4, as

The partial derivative of the objective function f with respect to each column of the desired matrices , i.e., $\{a_n^{(k)}\}$, u_n, v_r and w_n, are given by

$$\frac{\partial f}{\partial a_n^{(1)}} = -\mathscr{P} \times_2 a_n^{(2)} \times_3 u_n - \mathscr{Q} \times_2 a_n^{(3)} \times_3 v_n + \sum_{c=1}^{N} [(a_n^{(2)})^T a_c^{(2)} (u_n)^T u_c + (a_n^{(3)})^T a_c^{(3)} (v_n)^T v_c] a_c^{(1)}$$

$$\frac{\partial f}{\partial a_n^{(2)}} = -\mathscr{P} \times_1 a_n^{(1)} \times_3 u_n - \mathscr{R} \times_2 a_n^{(3)} \times_3 w_n + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (u_n)^T u_c + (a_n^{(3)})^T a_c^{(3)} (w_n)^T w_c] a_c^{(2)}$$

$$\frac{\partial f}{\partial a_n^{(3)}} = -\mathscr{Q} \times_1 a_n^{(1)} \times_3 v_n - \mathscr{R} \times_1 a_n^{(2)} \times_3 w_n + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (v_n)^T v_c + (a_n^{(2)})^T a_c^{(2)} (w_n)^T w_c] a_c^{(3)}$$

$$\frac{\partial f}{\partial u_n} = -\mathscr{P} \times_1 a_n^{(1)} \times_2 a_n^{(2)} + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (a_n^{(2)})^T a_c^{(2)}] u_c$$

$$\frac{\partial f}{\partial v_n} = -\mathscr{Q} \times_1 a_n^{(1)} \times_2 a_n^{(3)} + \sum_{c=1}^{N} [(a_n^{(1)})^T a_c^{(1)} (a_n^{(3)})^T a_c^{(3)}] v_c$$

$$\frac{\partial f}{\partial w_n} = -\mathscr{R} \times_1 a_n^{(2)} \times_2 a_n^{(3)} + \sum_{c=1}^{N} [(a_n^{(2)})^T a_c^{(2)} (a_n^{(2)})^T a_c^{(3)}] w_c.$$

A.1 Proof for Proposition 1

Proof. The three components of the objective function in (4.21), i.e., $f^{(1)}(A^{(1)},A^{(2)},U)$, $f^{(2)}(A^{(1)},A^{(3)},V)$ and $f^{(3)}(A^{(2)},A^{(3)},W)$, share similar structure, which is the difference between one tensor and the corresponding estimated results. Therefore, we take $f^{(1)}(A^{(1)},A^{(2)},U)$ and its partial derivative with respect to $a_n^{(1)}$ for further analysis. It can be rewritten as

$$\begin{aligned} & f^{(1)}(A^{(1)},A^{(2)},U) \\ =& \|\mathscr{P}-[\![A^{(1)},A^{(2)},U]\!]\|^2 \\ =& \underbrace{\|\mathscr{P}\|^2}_{f_1^{(1)}} - 2\underbrace{<\mathscr{P},[\![A^{(1)},A^{(2)},U]\!]>}_{f_2^{(1)}} + \underbrace{\|[\![A^{(1)},A^{(2)},U]\!]\|^2}_{f_3^{(1)}}. \end{aligned} \tag{A.1}$$

The first summand $f_1^{(1)}$ does not involve any variable and therefore

$$\frac{\partial f_1^{(1)}}{\partial a_n^{(1)}} = 0, \tag{A.2}$$

where 0 is the zero vector with the same length as $a_n^{(1)}$. The second summand $f_2^{(1)}$ is the inner product of the tensor $\mathscr{P}$ with its its polyadic decomposition, and it can be computed as

$$\begin{aligned} f_2^{(1)} &= <\mathscr{P},[\![A^{(1)},A^{(2)},U]\!]> \\ &= <\mathscr{P},\sum_{n=1}^{N} a_n^{(1)}\circ a_n^{(2)}\circ u_n> \\ &= \sum_{n=1}^{N}\sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{L} p_{i_1,i_2,i_3} a_{i_1,n}^{(1)} a_{i_2,n}^{(2)} u_{i_3,n} \\ &= \sum_{n=1}^{N}(\mathscr{P}\times_1 a_n^{(1)}\times_2 a_n^{(2)}\times_3 u_n) \\ &= \sum_{n=1}^{N}(\mathscr{P}\times_2 a_n^{(2)}\times_3 u_n)^T a_n^{(1)}. \end{aligned} \tag{A.3}$$

The partial derivative of $f_2^{(1)}$ with respect to each column of $A^{(1)}$ is

$$\frac{\partial f_2^{(1)}}{\partial a_n^{(1)}} = \mathscr{P} \times_2 a_n^{(2)} \times_3 u_n. \tag{A.4}$$

The third summand is the square of the Frobenius norm of $\mathscr{P}$'s polyadic decomposition, and it can be computed as

$$\begin{aligned} f_3^{(1)} =& \|[\![A^{(1)}, A^{(2)}, U]\!]\|^2 \\ =& < \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(2)} \circ u_n, \sum_{n=1}^{N} a_n^{(1)} \circ a_n^{(2)} \circ u_n > \\ =& \sum_{b=1}^{N} \sum_{c=1}^{N} \underbrace{((a_b^{(1)})^T (a_c^{(1)})(a_b^{(2)})^T (a_c^{(2)})(u_b)^T (u_c))}_{F(b,c)} \\ =& F(n,n) + \sum_{\substack{b=1 \\ b\neq n}}^{N} \sum_{\substack{c=1 \\ c\neq n}}^{N} F(b,c) + 2 \sum_{\substack{c=1 \\ c\neq n}}^{N} F(n,c), \end{aligned} \tag{A.5}$$

where b and c denote the indices of the factor matrices. The first summand of $f_3^{(1)}$ is

$$F(n,n) = (a_n^{(1)})^T (a_n^{(1)})(a_n^{(2)})^T (a_n^{(2)})(u_n)^T (u_n), \tag{A.6}$$

and its partial derivative with respect to the n^{th} column of the factor matrix $A^{(1)}$ is

$$\frac{\partial F(n,n)}{\partial a_n^{(1)}} = 2((a_n^{(2)})^T a_n^{(2)} u_n^T u_n) a_n^{(1)}. \tag{A.7}$$

The second summand of $f_3^{(1)}$ does not involve the variable $a_n^{(1)}$ and therefore the corresponding partial derivative with respect to $a_n^{(1)}$ is the zero vector with the same length as $a_n^{(1)}$. The third summand of $f_3^{(1)}$ is

$$2 \sum_{\substack{c=1 \\ c\neq n}}^{N} F(n,c) = 2 \sum_{\substack{c=1 \\ c\neq n}}^{N} (a_n^{(1)})^T (a_c^{(1)})(a_n^{(2)})^T (a_c^{(2)})(u_n)^T (u_c), \tag{A.8}$$

and its partial derivative with respect to the $a_n^{(1)}$ can be computed as $2\sum_{\substack{c=1\\c\neq n}}^{N}[(a_n^{(2)})^T a_c^{(2)}(u_n)^T u_c]a_c^{(1)}$. Therefore,

$$\begin{aligned}\frac{\partial f_3^{(1)}}{\partial a_n^{(1)}} =& 2((a_n^{(2)})^T a_n^{(2)} u_n^T u_n)a_n^{(1)} \\ &+ 2\sum_{\substack{c=1\\c\neq n}}^{N}[(a_n^{(2)})^T a_c^{(2)}(u_n)^T u_c]a_c^{(1)} \\ =& 2\sum_{c=1}^{N}[(a_n^{(2)})^T a_c^{(2)}(u_n)^T u_c]a_c^{(1)}.\end{aligned} \tag{A.9}$$

Combining all the above results, i.e., equation (A.2), (A.4) and (A.9), the partial derivative of $f^{(1)}(A^{(1)},A^{(2)},U)$ with respect to the $a_n^{(1)}$ can be computed as

$$\begin{aligned}&\frac{\partial f^{(1)}(A^{(1)},A^{(2)},U)}{\partial a_n^{(1)}} \\ =&\frac{\partial f_1^{(1)}}{\partial a_n^{(1)}} - 2\frac{\partial f_2^{(1)}}{\partial a_n^{(1)}} + \frac{\partial f_3^{(1)}}{\partial a_n^{(1)}} \\ =&-2\mathscr{P}\times_2 a_n^{(2)}\times_3 u_n + 2\sum_{c=1}^{N}[(a_n^{(2)})^T a_c^{(2)}(u_n)^T u_c]a_c^{(1)}.\end{aligned} \tag{A.10}$$

Similarly, we can calculate the partial derivative of $f^{(2)}(A^{(1)},A^{(3)},V)$ with respect to the $a_n^{(1)}$ as

$$\begin{aligned}&\frac{\partial f^{(2)}(A^{(1)},A^{(3)},V)}{\partial a_n^{(1)}} \\ =&-2\mathscr{Q}\times_2 a_n^{(3)}\times_3 v_n + 2\sum_{c=1}^{N}[(a_n^{(3)})^T a_c^{(3)}(v_n)^T v_c]a_c^{(1)}.\end{aligned} \tag{A.11}$$

$f^{(3)}(A^{(2)},A^{(3)},W)$ does not involve the variable $a_n^{(1)}$ and therefore

$$\frac{\partial f^{(3)}(A^{(2)},A^{(3)},W)}{\partial a_n^{(1)}} = 0. \tag{A.12}$$

Consequently, the partial derivative of the objective function with respect to $a_n^{(1)}$ is

$$\begin{aligned}
&\frac{\partial f(A^{(1)},A^{(2)},A^{(3)},U,V,W)}{\partial a_n^{(1)}} \\
=&\frac{1}{2}\frac{\partial f^{(1)}}{\partial a_n^{(1)}}+\frac{1}{2}\frac{\partial f^{(2)}}{\partial a_n^{(1)}}+\frac{1}{2}\frac{\partial f^{(3)}}{\partial a_n^{(1)}} \\
=&-\mathscr{P}\times_2 a_n^{(2)}\times_3 u_n-\mathscr{Q}\times_2 a_n^{(2)}\times_3 v_n \\
&+\sum_{c=1}^{N}[(a_n^{(2)})^T a_c^{(2)}(u_n)^T u_c+(a_n^{(3)})^T a_c^{(3)}(v_n)^T v_c]a_c^{(1)}.
\end{aligned} \tag{A.13}$$

This completes the proof of the first equation in Proposition 1. The proof of other equations is similar to that of (A.13) and thus omitted here.∎