Title
Machine learning algorithms for independent vector analysis and blind source separation

Permalink
https://escholarship.org/uc/item/6bw2065v

Author
Lee, In Tae

Publication Date
2009

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Machine Learning Algorithms for Independent Vector Analysis and Blind Source Separation

A Dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in

Electrical Engineering (Communication Theory and Systems)

by

In Tae Lee

Committee in charge:

Dr. Bhaskar Rao, Chair
Dr. Te-Won Lee, Co-Chair
Dr. Alon Orlitsky
Dr. Kenneth Kreutz-Delgado
Dr. Lawrence Saul
Dr. Sanjoy Dasgupta

2009
Copyright
In Tae Lee, 2009
All rights reserved.
The Dissertation of In Tae Lee is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2009
## TABLE OF CONTENTS

Signature Page ....................................................... iii

Table of Contents .................................................... iv

List of Figures ....................................................... vi

List of Tables ....................................................... viii

Acknowledgements ................................................... ix

Vita ................................................................. xi

Abstract of the Dissertation ...................................... xiii

1 Introduction .......................................................... 1
  1.1 Notations ....................................................... 1
  1.2 Background and Motivations .................................. 4
    1.2.1 Bin-Wise Separation by Independent Component Analysis . 6
    1.2.2 Permutation Problem ..................................... 7
    1.2.3 Scaling Problem ........................................... 8
  1.3 Organization ................................................... 9

2 Contrast Functions of Independent Component Analysis and Independent Vector Analysis .......................................... 11
  2.1 Contrast Functions of Independent Component Analysis ...... 11
    2.1.1 Mutual Information ....................................... 12
    2.1.2 Likelihood Function ...................................... 12
    2.1.3 Pictorial View in Information Geometry ................. 13
  2.2 Independent Vector Analysis .................................. 16
  2.3 Contrast Functions of IVA (or MICA): Spatially White Data .... 18
    2.3.1 Entropic Contrasts ........................................ 18
    2.3.2 Likelihood Function ...................................... 22
  2.4 Acknowledgement ............................................... 22

3 Independent Vector Analysis Algorithms for Convolutive Blind Speech Separation ........................................... 23
  3.1 Modeling the STFT Components of Speech ...................... 23
    3.1.1 $L_p$-Norm-Invariant Joint Densities .................... 24
    3.1.2 Sparse and Spherically Symmetric Joint Densities ...... 25
  3.2 Contrast Optimization .......................................... 27
    3.2.1 Quadratic Taylor Expansion in Complex Notations ....... 27
    3.2.2 Newton’s Method .......................................... 29
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The structure of a STFT-domain BSS system.</td>
<td>6</td>
</tr>
<tr>
<td>1.2</td>
<td>The structure of a STFT-domain BSS system where the approach by independent vector analysis is added.</td>
<td>8</td>
</tr>
<tr>
<td>2.1</td>
<td>Mutual information is depicted in information geometry. Mutual information is the KL distance from the joint PDF of the output signals to a hyperplane, i.e. the product manifold.</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Likelihood function is depicted in information geometry. Likelihood can be regarded as the KL distance from the joint PDF of the output signals to a fixed point, i.e. the source target.</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>The mixture model of ICA.</td>
<td>16</td>
</tr>
<tr>
<td>2.4</td>
<td>The mixture model of IVA. ICA is extended to a formulation with multivariate variables, or vectors, where the components of each vector are dependent. The difference of IVA from multidimensional ICA or subspace ICA is that the mixing process is restricted to the source components on the same horizontal layer to form a standard ICA mixture.</td>
<td>17</td>
</tr>
<tr>
<td>2.5</td>
<td>The mixture model of MICA or ISA. The model consists of a single mixture layer where all source components are mixed together.</td>
<td>18</td>
</tr>
<tr>
<td>2.6</td>
<td>Entropic contrasts of multidimensional ICA or IVA in the information geometry of zero-mean white PDFs.</td>
<td>21</td>
</tr>
<tr>
<td>3.1</td>
<td>The two-dimensional density plots and the contour plots of (a) independent Laplace distribution and (b) spherically symmetric Laplace (SSL) distribution. When being compared with independent Laplace distribution, it can be seen that SSL has some variance dependency.</td>
<td>26</td>
</tr>
<tr>
<td>3.2</td>
<td>The real recording environment of the experiment with real acoustic recordings.</td>
<td>33</td>
</tr>
<tr>
<td>3.3</td>
<td>The recorded observation signals.</td>
<td>34</td>
</tr>
<tr>
<td>3.4</td>
<td>The separation results of the Fast fixed-point IVA algorithm with ML_SEND contrast.</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>Undirected graphs for IVA dependence models. (a) A global clique to represent the super-Gaussian and radially symmetric dependence. (b) Chain-like series of overlapped cliques to represent the proposed dependence.</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>Geometric configuration of the simulated room environment. For each experiment, two source locations were chosen from A to K and the microphone locations were fixed as shown in the figure.</td>
<td>39</td>
</tr>
</tbody>
</table>
Figure 5.1: Left: the original sources. Middle: mixed signal corrupted by noise. Right: separated signal. ........................................... 46
Figure 5.2: The Kurtosis of each frequency bin for speech and music. . . . . 47

Figure 6.1: The approach of independent vector analysis (IVA) for solving the permutation problem. ........................................... 50
Figure 6.2: The covariance matrices of (a) a male speech, (b) a female speech, and (c) a piece of music where the acoustic signals were 12 seconds long. ........................................... 51
Figure 6.3: A closer look of Fig. 6.2. ........................................... 52
Figure 6.4: The experiments: (a) geometric configuration, (b) experiment numbers and source locations, and (c) separation performances in signal to interference ratio (SIR). ........................................... 56
LIST OF TABLES

4.1 Separation performances in SIR (dB). Proposed methods are denoted by "the number of bins in each clique" and "the number of bins of each shift". ............................................. 39
ACKNOWLEDGEMENTS

I would like to convey my thanks to the following individuals for helping me accomplish my dissertation. My deepest gratitude goes to Dr. Te-Won Lee who trained, directed, and supported me throughout my Ph.D. studies. I was very fortunate to have been supervised by him since the time I spent with him in San Diego was so invaluable that it will remain as unforgettable memory to me. Also, I would like to deeply thank my committee members, Prof. Bhaskar Rao, Prof. Alon Orlitsky, Prof. Kenneth Kreutz-Delgado, Prof. Lawrence Saul, and Prof. Sanjoy Dasgupta, for providing me with the highest insight and knowledge through classes and discussions. Also, my colleagues and friends, Dr. Hyun-Jin Park, Dr. Taesu Kim, Dr. Jiucang Hao, Dr. Gil-Jin Jang, Duangmanee Putthividhya, and Wonseok Shin, deserve many thanks from me for fruitful discussions and help. I would like to extend my appreciation to the staff members in Institute for Neural Computation and the department of Electrical and Computer Engineering, who friendly assisted me with my paperworks. Finally, I would like to present all my love and thanks to my family members for supporting me with the greatest love.

The results included in this dissertation have been published in the following articles:

(Chapter 2 – 3)

(Chapter 3)

Intae Lee, Taesu Kim, and Te-Won Lee, “Independent Vector Analysis for Blind Speech Separation”, in *Blind Speech Separation (Signals and Communication Tech-


(Chapter 6)

The dissertation author was the primary investigator and author of these publications.
VITA

2000 B.S. Electrical Engineering
Seoul National University, Seoul, Republic of Korea

2000-2005 Doctoral Study Abroad Scholarship
Ministry of Information and Communication, Republic of Korea

2002-2004 Teaching Assistant
Department of Electrical & Computer Engineering and
Department of Computer Science & Engineering
University of California, San Diego, CA

2003-2009 Research Assistant
Department of Electrical & Computer Engineering
University of California, San Diego, CA

2005 M.S. Electrical Engineering
University of California, San Diego, CA

2007 Graduate Intern
HP Labs.
Palo Alto, CA

2007 Korea-U.S. Science Cooperation Center – Korean-American Scientist Engineer Association (KUSCO-KSEA) Scholarship for Graduate Students

2009 Ph.D. Electrical Engineering
University of California, San Diego, CA

PUBLICATIONS

Journal Papers


Book Chapter


Proceedings


ABSTRACT OF THE DISSERTATION

Machine Learning Algorithms for Independent Vector Analysis and Blind Source Separation

by

In Tae Lee

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California San Diego, 2009

Dr. Bhaskar Rao, Chair
Dr. Te-Won Lee, Co-Chair

Blind signal separation (BSS) aims at recovering unknown source signals from the observed sensor signals where the mixing process is also unknown. As a popular method to solve this problem, independent component analysis (ICA) maximizes the mutual independence among, or equivalently the non-Gaussianity of, the signals and has been very successful especially when the unknown mixing process is instantaneous. In most realistic situations, however, there are time delay and reverberations which involve long filter lengths in the time domain.

Such convolutive BSS problems are often tackled in the frequency domain, or short-time Fourier transform (STFT) domain, mainly because the convolutive mixture model can be approximated to bin-wise instantaneous mixtures given the frame size is long enough to cover the main part of the convolved impulse responses. While the bin-wise instantaneous mixtures can be separated by the ICA algorithms for complex-valued variables, there are several factors that have significant influence on the final separation performance, which are the permutation problem, incomplete bin-wise separation, and noise.

Permutation problem refers to the random alignment of the STFT components that are separated by ICA. It is due to the permutation indeterminacy of ICA
and it hinders proper reconstruction of the original time-domain signals. To solve this problem, a multidimensional ICA framework that is called independent vector analysis (IVA) has been proposed. IVA exploits the mutual dependence among the STFT components originating from the same source and employs a multivariate dependence model. In this thesis, various dependence models and methods are proposed in the framework of IVA to solve the convolutive BSS problem, which include $L_p$-norm invariant joint densities, density functions represented by overlapped cliques in graphical models, Newton’s update optimization, and an EM algorithm using a mixture of multivariate Gaussians prior where Gaussian noise is added in the model.

While IVA is an effective framework to solve the convolutive BSS, the high dimensionality in the STFT domain makes it difficult to model the joint probability density function (PDF) of the fullband STFT components. On the other hand, bin-wise separation is a simpler task for which a permutation correction algorithm has to follow. For permutation correction, overall measures of magnitude correlation have been popular. However, the positive correlation is stronger between STFT components that are close to each other and correlation is a measure computed pair-wise. Thus, in this thesis, subband likelihood functions are proposed for the permutation correction which is fast to obtain and robust in solving the permutation problem.
Introduction

1.1 Notations

In this chapter, the notations are defined. Scalars (including complex-valued ones), vectors, and matrices will be denoted with normal lowercase or normal uppercase, bold lowercase, and bold uppercase letters, respectively.

- $z$ or $z_i$: a dummy variable.
- $z$: the multivariate dummy variable whose $k$-th component is $z_i$.
- $(\cdot)^T$: the transpose operator.
- $(\cdot)^H$: the Hermitian transpose operator.
- $(\cdot)^{-1}$: the matrix inverse operator.
- $|\cdot|$: either the absolute value of a scalar or the cardinality of a set.
- $\|\cdot\|_p$: the $L^p$-norm of a vector, i.e. $\|z\|_p = (|z_1|^p + |z_2|^p + \cdots)^{\frac{1}{p}}$.
- $N$: the number of frames in the short-time Fourier transform (STFT) domain.
- $f_v(\cdot)$ or $f_v(\cdot)$: the probability density function (PDF) of a multivariate random variable (RV) $v$ or a uni-variate RV $v$.
- $\text{KL}(\cdot||\cdot)$ or $\text{D}(\cdot||\cdot)$: The Kullback-Leibler divergence of two PDFs.
• \( \mathbb{E} [\cdot] \): the expectation with respect to the empirical distribution, i.e. \( \frac{1}{N} \sum_{n=1}^{N} \cdot \).
• \( s_k(t) \) or \( s_{kt} \): the \( k \)-th source signal in the time domain.
• \( s(t) \) or \( s_t \): the array of source signals whose \( k \)-th component is \( s_k(t) \).
• \( x_l(t) \) or \( x_{lt} \): the \( l \)-th microphone signal in the time domain.
• \( x(t) \) or \( x_t \): the array of microphone signals whose \( l \)-th component is \( x_l(t) \).
• \( y_k(t) \) or \( y_{kt} \): the estimate of \( s_k(t) \).
• \( y(t) \) or \( y_t \): the array of estimated signals whose \( k \)-th component is \( y_k(t) \).
• \( u_l(t) \) or \( u_{lt} \): the \( l \)-th microphone noise signal in the time domain.
• \( u(t) \) or \( u_t \): the array of noise signals whose \( l \)-th component is \( n_l(t) \).
• \( a_{lk}(\tau) \): the impulse response from the \( k \)-th source signal to the \( l \)-th microphone in the time domain which are assumed to be time-invariant.
• \( A(\tau) \): the impulse response matrix in the time domain whose \((l,k)\)-th component is \( a_{lk}(\tau) \).
• \( K \): the number of source signals.
• \( L \): the number of microphone signals.

The relations among the time-domain signals and the filters are

\[
x(t) = \sum_{\tau} A(\tau) s(t - \tau) + u(t),
\]

\[
x_l(t) = \sum_{k=1}^{K} \sum_{\tau} a_{lk}(\tau) s_k(t - \tau) + u_l(t), \quad l = 1, \cdots, L.
\]

• \( s^b_k[n] \) or \( s^b_{kn} \): the STFT’ed \( k \)-th source signal in the \( b \)-th bin.
• \( s_k[n] \) or \( s_{kn} \): the STFT’ed \( k \)-th source signal whose \( b \)-th component is \( s^b_k[n] \).
• \( s^b[n] \) or \( s^b_n \): the column vector whose \( k \)-th component is \( s^b_k[n] \).
• \( x^b_l[n] \): the STFT’ed \( l \)-th microphone signal in the \( b \)-th bin.
• \( x_i[n] \): the STFT’ed \( l \)-th microphone signal whose \( b \)-th component is \( x_i^b[n] \).
• \( x^b[n] \): the column vector whose \( l \)-th component is \( x_i^b[n] \).
• \( y_k^b[n] \) or \( y_{kn}^b \): the estimate of \( s_k^b[n] \).
• \( y_k[n] \) or \( y_{kn} \): the estimated signal whose \( b \)-th component is \( y_k^b[n] \).
• \( y^b[n] \) or \( y^n_b \): the column vector whose \( k \)-th component is \( s_k^b[n] \).
• \( u_l^b[n] \): the STFT’ed \( l \)-th noise signal in the \( b \)-th bin.
• \( u_l[n] \): the STFT’ed \( l \)-th noise signal whose \( b \)-th component is \( u_l^b[n] \).
• \( u_b[n] \): the column vector whose \( l \)-th component is \( u_l^b[n] \).
• \( a_{bk}^l \): the STFT coefficient of the impulse response \( a_{bk}(\tau) \) in the \( b \)-th bin. Here we assume that the time-domain filter is time-invariant.
• \( A^b \): the mixing matrix in the \( b \)-th bin. Here we assume that the impulse responses are time-invariant.
• \( Q^b \): the spatial pre-whitening matrix in the \( b \)-th bin.
• \( W^b \): the separating matrix in the noiseless case which is the estimate of \((A^b)^{-1}\).
• \( w_k^b \): the \( k \)-th row vector of \( W^b \).
• \( w_{kl}^b \): the \( l \)-th component of \( w_k^b \).
• \( B \): the number of frequency bins.

Hence, the process of spatial pre-whitening is

\[
x^b[n] \leftarrow Q^b x^b[n], \quad b = 1, \ldots, B, \quad (1.3)
\]

and the relations are

\[
x^b[n] = A^b s^b[n] + u^b[n], \quad b = 1, \ldots, B, \quad (1.4)
\]
\[
x_i^b[n] = \sum_{k=1}^{K} a_{ik}^b s_k[n] + u_i^b[n],
\]
\[
= a_{il}^b s_l[n] + u_i^b[n], \quad l = 1, \ldots, L, \quad b = 1, \ldots, B. \quad (1.5)
\]
In the noiseless case,
\[
\begin{align*}
\mathbf{y}_b[n] &= \mathbf{W}_b \mathbf{x}_b[n], \quad b = 1, \cdots, B, \\
\mathbf{y}_k[n] &= \sum_{l=1}^{L} w_{kl}^b x_l^b[n], \\
&= \mathbf{w}_k^b \mathbf{x}_k[n], \quad k = 1, \cdots, K, \quad b = 1, \cdots, B.
\end{align*}
\]

Note that in the approximated STFT-domain expressions, there is some error that results from the difference between the actual linear convolution and the circular convolution in discrete Fourier transform (DFT). This problem will be ignored throughout this dissertation by assuming that the frame size is big enough to keep this error small.

### 1.2 Background and Motivations

Blind signal separation (BSS), or blind source separation, refers to a set of problems that aim to separate individual source signals from their mixtures. In those problems, source signals are mixed in the recording of the sensors, and sensor noise that is not considered as an individual source signal can be added to the mixtures in the model but is ignored in many cases. BSS can be categorized by the relative number of sensors to the number of sources: under-determined BSS if the number of sensors is less than the number of sources, determined BSS if the numbers equal, and over-determined BSS otherwise. For sure, (over-)determined BSS is the easier type of problem and there are many algorithms that can solve the (over-)determined problems especially when the mixture is instantaneously linear and noise is negligible, i.e. when the mixing and the separating processes can be denoted respectively as
\[
\begin{align*}
\mathbf{x}(t) &= \mathbf{A}\mathbf{s}(t), \\
\mathbf{y}(t) &= \mathbf{W}\mathbf{x}(t),
\end{align*}
\]
where the mixing matrix \( \mathbf{A} \) is assumed to be full rank and invariant over time. Since over-determined BSS can be reduced to determined BSS using dimensionality
reduction techniques, for convenience we will assume throughout this dissertation that the number of observation signals $L$ equals the number of source signals $K$ and that the mixing matrices are invertible.

When the source signals are acoustic, the mixture is convolutive rather than instantaneous, i.e. there are propagation time delay and reverberation in the signal recording, and thus the mixing process can be denoted approximately as

$$x(t) = \sum_{\tau} A(\tau)s(t - \tau), \quad (1.10)$$

$$x_l(t) = \sum_{k=1}^{K} \sum_{\tau} a_{lk}(\tau)s_k(t - \tau), \quad l = 1, \ldots, L. \quad (1.11)$$

By applying short-time Fourier transform (STFT) to the time-domain signals, we can convert the convolutive mixture problem in the time domain to a set of instantaneous linear mixture problems in the STFT domain, which is also called the frequency domain or the time-frequency (T-F) domain. In each frequency bin, the mixture and its separation can be denoted as

$$x^b[n] = A^b s^b[n], \quad (1.12)$$

$$y^b[n] = W^b x^b[n], \quad (1.13)$$

for $b = 1, \ldots, B$. Note that, here, $n$ is the frame number and the mixing matrices, as well as the signals, are complex-valued.

While the number of data samples largely decreases in this approach, dealing with the signals in the STFT domain is still beneficial since learning each filter coefficient can be performed separately in each bin and the separation of instantaneous linear mixtures is much easier. Thus it can better handle longer filter lengths and complex-valued independent component analysis (ICA) algorithms designed for instantaneous linear mixtures can be employed.

After bin-wise separations, however, the permutation indeterminacy and scaling indeterminacy of ICA yield problems that are called the permutation problem and the scaling problem. These problems hinder proper reconstruction of the original time-domain signals and need to be fixed after, or, if possible, during, the bin-wise separations. Thus there are three major tasks involved in the STFT-domain BSS
system which are the bin-wise separation, the permutation correction, and the scaling correction. The structure of the BSS system in the STFT domain is shown in Fig. 1.1.

![Diagram of STFT-domain BSS system]

Figure 1.1: The structure of a STFT-domain BSS system.

1.2.1 Bin-Wise Separation by Independent Component Analysis

One of the most popular methods that solve the (over-)determined BSS problem of instantaneous linear mixture is ICA [Herault and Jutten, 1986, Comon, 1994]. A major assumption in ICA is that the source signals are statistically independent and thus the original signals can be recovered by exploiting the independence among them. In the simplest form of ICA, the mixing process is assumed instantaneously linear, the noise signals are neglected, the number of source signals is at most the number of sensor signals (over-determined), and there are at most one Gaussian-distributed source signal. In this setting, which is equivalent to the approximated bin-wise mixture of STFT-domain BSS, ICA can separate the original source signals from their mixtures up to permutation and scaling.

To learn the separating matrix, each source signal is dealt with as random samples that follow certain probability distribution. In the case when each source distribution is known in advance, simple and robust ICA algorithms can be derived
from the maximum likelihood (ML) perspective. The prior information about the source distribution is brought into the objective function, a.k.a. contrast, by some nonlinear functions where the corresponding source distribution is the source prior, or also called the source target. In many cases, however, strong characterization of the source distribution is unavailable and an incorrect source prior can result in poor separation results.

A number of ICA algorithms avoid likelihood contrasts and take flexible approaches in order to be blindly applicable [Hyvärinen and Oja, 1997, Lee et al., 1999, Hyvärinen, 1999, Cardoso, 1999, Pham, 2000, Bach and Jordan, 2002, Learned-Miller and Fisher III, 2003, Boscolo et al., 2004]. However, for complex-valued signals mixed by a complex-valued mixing matrix which is the case of the bin-wise separation in STFT-domain BSS, a universally applicable ICA algorithm has not been proposed yet.

### 1.2.2 Permutation Problem

The permutation indeterminacy of ICA results in permutation disorder along the frequency bins. In order to solve the permutation problem, positive inter-frequency correlation of the magnitudes [Anemueller and Kollmeier, 2000, Murata et al., 2001] or dominance measures [Sawada et al., 2007] and direction of arrival (DOA) estimation [Kurita et al., 2000, Ikram and Morgan, 2002, Sawada et al., 2003] have been used. Also, as another approach to solve the permutation problem, a generalized ICA framework termed independent vector analysis (IVA) has been used which is an extension of ICA from uni-variate components to multivariate components and utilizes the dependence inside the independent groups of signals [Kim et al., 2007, Lee et al., 2007a, Hiroe, 2006]. Using the framework of IVA, the permutation problem can be avoided during the separation. Thus, as another option to the bin-wise separation and the permutation correction together, IVA can be added to the structure of the STFT-domain BSS system in Fig. 1.1 and is depicted in Fig. 1.2.

Permutation problem is still regarded as the most important and most difficult task in STFT-domain BSS. Although there have been successful algorithms
1.2.3 Scaling Problem

In STFT-domain BSS, other than the permutation problem, the scaling problem of assigning proper variance to each STFT component with respect to the other STFT components has to be solved. This arises because of the scaling indeterminacy of ICA, and also IVA. Please note that the scaling indeterminacy in the complex field includes phase ambiguity. Here, we employ the well-known method of minimal distortion principle [Matsuoka and Nakashima, 2001] to fix the scaling problem. After all unmixing matrices, i.e. $W^b$’s, are learned, we further multiply the diagonal of $(W^b)^{-1}$ to the left of $W^b$, for $b = 1, \cdots, B$:

$$W^b \leftarrow \text{diag} \left((W^b)^{-1}\right) W^b$$

The reason this can solve the scaling problem is as follows. Presuming that a separating matrix $W^b$ is properly learned up to permutation and scaling satisfying
the following:

\[ W^b A^b = D^b P^b, \quad b = 1, \ldots, B, \]  

(1.15)

where \( D^b \) is a diagonal scaling matrix and \( P^b \) is a permutation matrix. Note that these two matrices \( D^b \) and \( P^b \) represent respectively the scaling indeterminacy and the permutation indeterminacy of ICA. Then,

\[
\text{diag}\left((W^b)^{-1}\right) = \text{diag}\left(A^b (P^b)^T (D^b)^{-1}\right) \\
= \text{diag}\left(A^b (P^b)^T (D^b)^{-1}\right). 
\]

(1.16)

Hence,

\[
\text{diag}\left((W^b)^{-1}\right) W^b A^b = \text{diag}\left(A^b (P^b)^T (D^b)^{-1}\right) D^b P^b \\
= \text{diag}\left(A^b (P^b)^T\right) P^b, 
\]

(1.19)

where the scaling matrix \( D^b \) has disappeared.

1.3 Organization

This dissertation focuses on solving the STFT-domain BSS problems using statistical approaches. The main tasks of BSS that are discussed in this dissertation are the separation of instantaneously mixed STFT coefficients and permutation correction.

The organization of the thesis is as follows. Chap. 2 is an introductory chapter of ICA and IVA discussing their available objective functions. In Chap. 3, various multivariate probability density models of STFT-domain speech signals that are to be used in the IVA framework are proposed and fast fixed-point IVA algorithms are derived. In Chap. 4, a multivariate density model that can be represented by chain-like overlapped cliques in graphical models is proposed for the separation of convolutedly mixed speech signals in the IVA framework. In Chap. 5, a flexible IVA approach using mixture of Gaussians as the source prior is derived where Gaussian noise is added in the model. Other than using the IVA framework for BSS, a permutation correction scheme following the bin-wise separation is also proposed in
Chap. 6. The new scheme uses a sliding sub-band likelihood function for permutation correction. Lastly, Chap. 7 wraps up the dissertation.
Contrast Functions of Independent Component Analysis and Independent Vector Analysis

ICA takes advantage of the assumption that the source signals are statistically independent and thus learns the separating matrix by maximizing the mutual independence among, or by maximizing the non-Gaussianity of, the output signals. In this chapter, for better understanding of ICA and IVA, the objective functions are discussed.

2.1 Contrast Functions of Independent Component Analysis

Most ICA algorithms ignore the time structure of the signals and deal with them as i.i.d. samples of random variables (RVs) and thus for convenience we will mostly omit the time structures “[t]” and “[n]” and express the signals as RVs.
2.1.1 Mutual Information

Since ICA aims at maximizing the statistical independence among output signals, mutual information has been employed as one of the general objective functions of ICA. The mutual information of the output signal \( y_k \)'s is denoted by

\[
I(y_1; \cdots ; y_K) = KL \left( f_y \mid \mid \prod_{k=1}^{K} f_{y_k} \right)
\]

(2.1)

\[
= \int_{\mathbb{R}^{\text{dim}(z)}} f_y(z) \log \frac{f_y(z)}{\prod_{k=1}^{K} f_{y_k}(z_k)} \, dz.
\]

(2.2)

Mutual information can be regarded as a metric that measures the distance from a given joint PDF to the product of its marginal PDFs. However, note that KL divergence has no symmetry property: \( KL(p \mid \mid q) \neq KL(q \mid \mid p) \). Indeed mutual information as a contrast function is minimized to zero if and only if the output component \( y_k \)'s are mutually independent and measures how much independent the components of the output data are.

2.1.2 Likelihood Function

Likelihood function is also a very popular contrast function that has been employed in ICA algorithms such as infomax [Bell and Sejnowski, 1995, Cardoso, 1997]. In case the outcome is given while the model parameter is unknown, maximum likelihood (ML) approach learns the model parameter such that the likelihood function, i.e. the probability that the corresponding model yields the specific outcome, is maximized. In ICA formulation, the likelihood function can be denoted as

\[
\Pr(D_x | A) = \prod_n \Pr(x[n] | A, s[n] = z) f_s(z) \, dz
\]

(2.3)

\[
= \prod_n \int \Pr(x[n] | A, s[n] = z) f_s(z) \, dz
\]

(2.4)

\[
= \prod_n \int \delta(x[n] - A z) f_s(z) \, dz
\]

(2.5)

\[
= \prod_n |\det(W)| f_s(W x[n])
\]

(2.6)

\[
= |\det(W)|^N \prod_n f_s(y[n]),
\]

(2.7)
where $D_x$ and $\delta(\cdot)$ denote the given observation data and the Dirac delta function, respectively. Note that the matrices $A$ and $W$ are the model parameters and are related to each other by $W = (A)^{-1}$. Also note that, in the likelihood function of ICA, a source prior $f_s(\cdot)$ is needed. The normalized log-likelihood can be modified as

$$\frac{1}{N} \log \Pr(D_x|A) = \log \left( |\det(W)| \right) + \tilde{E} \left[ \log f_s(y) \right] \tag{2.8}$$

$$= \log \left( |\det(W)| \right) + H(f_y) - KL(f_y||f_s) \tag{2.9}$$

$$= H(f_x) - KL(f_y||f_s), \tag{2.10}$$

where $\tilde{E}[\cdots]$ denotes $\frac{1}{N} \sum_n \cdots$ and $H(\cdot)$ denotes the entropy function;

$$H(f_y) \left( = H(y) \right) = - \int_{\mathbb{R}^{\dim(z)}} f_y(z) \log f_y(z) dz. \tag{2.11}$$

Note that in (2.10), $H(f_x)$ is a constant term and thus the likelihood function is equivalent to $KL(f_y||\prod_{k=1}^K f_{yk})$, i.e. ML is equivalent to minimizing the KL divergence from the output joint density $f_y(\cdot)$ to the source prior $f_s(\cdot)$. For this reason, $f_s(\cdot)$ is also called the source target (that $f_y(\cdot)$ is aimed at). Since the source prior, or source target, is usually assumed a priori, to distinguish it from the real source distribution we will denote it as $\hat{f}_s(\cdot)$ from here on. Since ICA assumes statistical independence among source signals, the source targets are products of marginal PDFs:

$$\hat{f}_s(\cdot) = \prod_{k=1}^K \hat{f}_{sk}(\cdot). \tag{2.12}$$

### 2.1.3 Pictorial View in Information Geometry

By expressing mutual information and likelihood contrasts of ICA using KL divergence, it can be seen that they are very similar: $KL(f_y||\prod_{k=1}^K f_{yk})$ and $KL(f_y||\prod_{k=1}^K \hat{f}_{sk})$, respectively. Yet, it should be noted that they are not the same, i.e. the PDF $\prod_{k=1}^K f_{yk}(\cdot)$ in mutual information is the product of the output signals’ marginal PDFs while the PDF $\prod_{k=1}^K \hat{f}_{sk}(\cdot)$ in likelihood is a fixed source target.

This can be better understood when they are, although roughly, illustrated in information geometry as it was done in [Cardoso, 2000]. Information geometry is the space of PDFs where each point in the space corresponds to a PDF. Let’s
define product manifold as the set of exponential family distributions that are the products of their marginal PDFs and draw it as a hyperplane in the space. Then the original source PDF $\prod_{k=1}^{K} f_{s_k}(\cdot)$ will lie somewhere on it and also the source target $\prod_{k=1}^{K} \hat{f}_{s_k}(\cdot)$ can be chosen to be a point on it.

Figure 2.1: Mutual information is depicted in information geometry. Mutual information is the KL distance from the joint PDF of the output signals to a hyperplane, i.e. the product manifold.

In Fig. 2.1, the points that correspond to the PDF of the source array $s$, the observation array $x$, and the output array $y$, product manifold, and mutual information are depicted. $\prod_{k=1}^{K} f_{s_k}(\cdot)$ and $f_{x}(\cdot)$ are fixed points in the space and $f_{y}(\cdot)$ moves around in the space with respect to the separating matrix $W$. In the figure, two different $f_{y}(\cdot)$’s are depicted respectively to $W = W_1$ and $W = W_2$. Note that $x$ and $y$ are constrained to be linearly transformed array signals of $s$ and thus the
points that $f_y(\cdot)$ can occupy are restricted, which is depicted as a solid curve in the figure. Since $\prod_{k=1}^K f_{y_k}(\cdot)$ is the closest point on the product manifold from $f_y(\cdot)$ in terms of KL divergence, mutual information can be regarded as a distance from a point, i.e. the joint PDF of the output signals, to a hyperplane, i.e. the product manifold.

Figure 2.2: Likelihood function is depicted in information geometry. Likelihood can be regarded as the KL distance from the joint PDF of the output signals to a fixed point, i.e. the source target.

In Fig. 2.2, instead of mutual information, the likelihood function is depicted. Here, likelihood measures the distance from a point, i.e. the joint PDF of the output signals, to another fixed point, i.e. the source target.

The goal of employing mutual information or likelihood function as a contrast function of ICA is to put $f_y(\cdot)$ as close as possible to the product manifold. Mutual
information can directly maximize the independence among the output components regardless of the source types unless there are more than one Gaussian sources. However, there is a need to estimate the density of each output signal $y_k$ from the data. This results in heavy computation and slow learning speed. On the other hand, in order to use likelihood it is important to choose a proper source target. The importance becomes more significant as the dimensionality of individual source signals rises.

### 2.2 Independent Vector Analysis

The model of IVA consists of a set of the standard ICA models (Fig. 2.3) where the univariate sources across different layers are dependent such that they can be aligned and grouped together as a multivariate variable, or vector. In Fig. 2.4, a $2 \times 2$ IVA mixture model is depicted where $\mathbf{s}_1 \left( = [s_{11}^1, s_{12}^1, \ldots, s_{1B}^1]^T \right)$ and $\mathbf{s}_2 \left( = [s_{21}^2, s_{22}^2, \ldots, s_{2B}^2]^T \right)$ denote the multivariate sources and $\mathbf{x}_1 \left( = [x_{11}^1, x_{12}^1, \ldots, x_{1B}^1]^T \right)$ and $\mathbf{x}_2 \left( = [x_{21}^2, x_{22}^2, \ldots, x_{2B}^2]^T \right)$ denote the observed multivariate mixtures. As it can be seen, the mixing of the multivariate sources is constrained component-wise forming ICA mixture models in each layer. The difference of applying IVA from applying multiple ICA’s to such a scenario is that IVA groups the dependent sources as a multivariate variable and learns each group as a whole.

![Figure 2.3: The mixture model of ICA.](image)

The idea of grouping dependent sources together as a multidimensional source in ICA was first proposed by J.-F. Cardoso [Cardoso, 1998] with the name of multidimensional ICA (MICA). Also, A. Hyvärinen [Hyvärinen and Hoyer, 2000] proposed
Figure 2.4: The mixture model of IVA. ICA is extended to a formulation with multivariate variables, or vectors, where the components of each vector are dependent. The difference of IVA from multidimensional ICA or subspace ICA is that the mixing process is restricted to the source components on the same horizontal layer to form a standard ICA mixture.

a maximum likelihood algorithm of MICA, which is subspace ICA, or independent subspace analysis (ISA). While IVA closely resembles those models and while MICA (including ISA) and IVA share the same contrasts, they differ in their mixture models. For comparison, the mixture model of MICA is depicted in Fig. 2.5. While IVA consists of multiple standard ICA layers where the component-wise mixing keeps the dependent source components unmixed, MICA consists of a single mixture layer which cannot be decomposed into multiple ICA mixtures, since it is assumed that the dependent sources are also mixed.

The IVA model fits the model of the frequency-domain (over-)determined BSS problem. The individual mixture layer in IVA corresponds to the instantaneous mixture in each frequency bin and the dependent sources grouped together as a multivariate variable correspond to the (short-time) frequency components of a time
2.3 Contrast Functions of IVA (or MICA): Spatially White Data

Many ICA algorithms keep the output data zero-mean and spatially white since uncorrelatedness is a necessary condition for independence and since it increases the learning speed. While it can also be done in MICA algorithms including ISA, in IVA, however, it is only feasible for the output data in each layer, \( y^b \left( = [y^b_1, y^b_2, \cdots]^T \right) \), since mixing and separating in IVA are restricted to each layer. Here, we keep \( y^b \) zero-mean and white by preprocessing the observation data \( x^b \left( = [x^b_1, x^b_2, \cdots]^T \right) \) to be zero-mean and white, and by constraining the separating matrix \( W^b \)'s to be orthogonal,

\[
E[x^b(x^b)^H] = I, \quad (2.13)
\]
\[
W^b(W^b)^H = I, \quad (2.14)
\]
for \( b = 1, 2, \cdots, B \). For convenience, we will assume that \( x^b \) is already preprocessed to be zero-mean and white from here on.

2.3.1 Entropic Contrasts

Since the basic idea of separating mutually independent multivariate sources is the same for MICA and IVA, they share the same contrast functions. Similar to
ICA, the contrast functions of IVA can be represented by the mutual information among the multidimensional variable \( y_k \)'s,

\[
\text{KL}(f_y \| \prod_{k=1}^{K} f_{y_k}) = \sum_{k=1}^{K} H(y_k) - H(y). \tag{2.15}
\]

Please note that the output signal \( y_k \left( = [y_1^k, y_2^k, \cdots, y_B^k]^T \right) \) is a vector now. This contrast reduces to the minimum, zero, if and only if \( y_k \)'s are mutually independent. Note that the term \( H(y) \) in (2.15) is constant since \( \log|\text{det}(W^b)| = 0 \) for \( b = 1, \cdots, B \), which follows from (2.14), and hence, minimizing the KL divergence term in (2.15) is equivalent to minimizing the sum of the entropies of the multidimensional variables, i.e.

\[
\text{arg min}_{\{W^b\}} \text{KL}(f_y \| \prod_{k=1}^{K} f_{y_k}) = \text{arg min}_{\{W^b\}} \sum_{k=1}^{K} H(y_k). \tag{2.16}
\]

with the constraint of orthogonal \( W^b \)'s.

In order to make the discussion that will follow simpler, we assume, in addition to the mutual independence of \( s_k \left( = [s_1^k, \cdots, s_B^k] \right) \)'s,

\[
\mathbb{E}[s_k(s_k)^H] = \mathbf{I}, \tag{2.17}
\]

i.e. the components of each multidimensional source are, although dependent, uncorrelated and the variance of each component is set to one which can be assumed owing to the scaling indeterminacy in ICA. Then, by the constraint of zero-mean and white \( y^b \), the whole output signal \( y \left( = [y^1; y^2; \cdots; y^B] \right) \) is also kept zero-mean and white. (The notation \([a; b; \cdots]\) is meant to denote \([a^T, b^T, \cdots]^T\) for column vectors \( a \) and \( b \));

\[
\mathbb{E}[yy^H] = \mathbf{I}. \tag{2.18}
\]

As in ICA, negentropy can be employed and another entropic contrast of IVA that is equivalent to the other entropic contrasts can be obtained. Negentropy is defined as follows.

\[
N(y) = \text{KL}(f_y \| f_{y^0}), \tag{2.19}
\]
where $y^G$ denotes the (multidimensional) Gaussian RV that has the same mean vector and the same covariance matrix with RV $y$. From the following Pythagorean relation in information geometry,

$$N(y_k) = H(y^G_k) - H(y_k),$$

(2.20)

note that $H(y^G_k)$ in (2.20) is a constant term since, from (2.18), $y_k$ is zero-mean and white such that $y^G_k$ is always fixed to be the unique zero-mean and white multidimensional Gaussian RV. Hence, by plugging in the relation of (2.20) into the right-hand-side of (2.16), we can easily see that,

$$\arg\min_{\{W^b\}} \sum_{k=1}^K H(y_k) = \arg\max_{\{W^b\}} \sum_{k=1}^K N(y_k),$$

(2.21)

with the constraint of orthogonal $W^b$’s. It is clear that the entropic contrasts of IVA in (2.16) and (2.21) are extensions of the classic entropic contrasts of ICA from univariate variables to multivariate variables, since by replacing $y_k$ for the $y_k$ terms, we come up with the entropic contrasts of ICA.

As it was done for ICA (J.-F. Cardoso, [Cardoso, 2000]), the relation between the mutual information contrast (2.16) and the sum of negentropies (2.21) can be visually seen in information Geometry. By defining the Gaussian manifold and independent vector manifold as the set of multivariate Gaussian PDFs and the set of PDFs where the multivariate $y_k$’s are mutually independent, respectively, the mutual information (2.16) and the sum of negentropies (2.21) can be regarded as the distances from a point $f_y(\cdot)$ to the Gaussian manifold and to the independent vector manifold, respectively.

In Fig. 2.6, the points, manifolds, and the entropic contrasts are depicted and indicated where the PDFs are for zero-mean and white RVs. Note that the Gaussian manifold is drawn as a single point since a zero-mean and white multivariate Gaussian PDF is unique. When updating the separating matrix $W^b$’s, the point $f_y(\cdot)$ moves around correspondingly while keeping the distance

$$N(y) = KL(f_y \| \prod_{k=1}^K f_{y_k}) + \sum_{k=1}^K N(y_k)$$

(2.22)
constant. Hence our goal is either to minimize the distance from $f_y(\cdot)$ to the independent vector manifold, $\text{KL}(f_y \parallel \prod_{k=1}^{K} f_{y_k})$, or to maximize the distance from $\prod_{k=1}^{K} f_{y_k}(\cdot)$ to the Gaussian manifold, $\sum_{k=1}^{K} N(y_k)$.

Please note that the assumption on the multivariate source components in (2.17) is not essential for the derivation of the negentropy contrast in (2.21) or the Pythagorean relation in (2.22). However, the motivation of the assumption is that the source distributions we later assume are spherically symmetric multivariate distributions which imply uncorrelatedness in the multivariate source components, and also that the explanation becomes much simpler with the assumption given.

Figure 2.6: Entropic contrasts of multidimensional ICA or IVA in the information geometry of zero-mean white PDFs.
2.3.2 Likelihood Function

As $y$ is $[y^1; \cdots; y^B]$, we will let $s\ (=[s^1; \cdots; s^B])$ and $x\ (=[x^1; \cdots; x^B])$ denote the whole source signal and the whole observation signal, respectively. The (normalized) log-likelihood contrast of IVA is in the form of

$$\tilde{E}\left[\log\left(\hat{f}_s(y)\right)\right] + \sum_b \log |\det(W^b)| = \sum_{k=1}^K \tilde{E}\left[\log\left(\hat{f}_{s_k}(y_k)\right)\right],$$

(2.23)
since we keep the separating matrices orthogonal, such that

$$\log |\det(W^b)| = 0, \quad b = 1, \cdots, B.$$  

(2.24)

2.4 Acknowledgement

This material is based upon work partially supported by the national science foundation (NSF) under grant No. 0535251. Also, I would like to acknowledge Taesu Kim and Te-Won Lee as the co-authors of the works included in this chapter.
Independent Vector Analysis Algorithms for Convolutive Blind Speech Separation

3.1 Modeling the STFT Components of Speech

In applying IVA to frequency-domain BSS problems, we come to deal with complex-valued variables. Because real-valued contrast functions of complex-valued variables are not analytic, in a number of signal processing applications there arises the question of how to deal with complex variables. There are standard ways to handle the problem. The first is to write every complex variable in terms of two real variables, i.e. replace each complex variable $z$ with two real variables which are the real and imaginary parts of $z$, and then use the tools for real-valued variables. An alternative is to follow the lines of [Brandwood, 1983, van den Bos, 1994], which is equivalent to the first method described, but cleaner in notations. These methods basically carry out the optimization in the real field with respect to the real and imaginary parts of the complex variables.

In most scenarios it is known that complex RVs show circular symmetry around the origin. Hence, as the relation between the separated real variables, we assume circularity in the source variables such that

$$\hat{E}[ss^T] = 0.$$  \hspace{1cm} (3.1)
And hence,

\[ \tilde{E}[xx^T] = O. \] (3.2)

With the circularity assumption on the STFT components of speech, they need to be modeled with a proper multivariate PDF. Some researchers observed that speech has the property of spherical invariance and they modeled bandlimited speech with spherically invariant random process (SIRP). For details and relevant topics, see [Brehm and Stammler, 1987, Buchner et al., 2004] and the references therein. The frequency models of speech that have been proposed in IVA formulation are closely related to SIRP.

### 3.1.1 Lp-Norm-Invariant Joint Densities

A group of PDFs which can be represented in the form of

\[
\hat{f}_{sk}(z) \propto e^{-\frac{1}{\sigma}(\|z\|_p)^m}
= e^{-\frac{1}{\sigma} \left( \sum |z^k|^p \right)^{\frac{1}{m}}} (3.4)
\]

where \( \sigma \) is the constant that controls the variance, has been proposed in [Lee and Lee, 2007]. The densities are Lp-norm invariant for the given value of \( p \) and the sparseness can be controlled by the parameter \( m \). Also the contour shape of the PDF varies by choosing different \( p \) values. Note that each complex variable in (3.4) is always circular (i.e. circularly symmetric in its PDF) even though the PDF of the whole multivariate variable is not spherically symmetric for \( p \neq 2 \). For convenience, we will denote each PDF with the given \( p \) and \( m \) values in (3.4) by \( N-I(p,m) \). For example, independent Laplace and corresponds to \( N-I(1,1) \). When \( N-I(p,m) \) is used as the source target in the likelihood contrast function (2.23), we will denote the contrast function by \( ML_{N-I(p,m)} \).

In speech separation, the best performances were obtained for \( (p, m) \) being around \((1.9, 7)\) and for \( p \) being from 1.8 to 2.0 the performances were similar, which justifies the efficiency of using sparse and spherical joint densities for modeling speech in the frequency domain.
3.1.2 Sparse and Spherically Symmetric Joint Densities

In [Kim et al., 2007, Lee et al., 2007a, Hiroe, 2006], several sparse and spherically symmetric joint PDFs were proposed for modeling the STFT components of speech. One, which we will call spherically symmetric Laplace distribution (SSL) is

\[
\hat{f}_{s_k}(z) \propto e^{-\frac{1}{\sigma}||z||_2}.
\]

SSL was first used in ISA [Hyvärinen and Hoyer, 2000] in order to extract groups of features. Note N-I(2, 1) equals SSL. Also a spherically symmetric exponential (L2-)norm distribution (SEND) was proposed:

\[
\hat{f}_{s_k}(z) \propto e^{-\frac{1}{\sigma}||z||_2} \frac{1}{||z||_2^{d-1}},
\]

where \(d\) is the total dimension in real field. It should be noted that the PDFs of complex variables are in fact functions of real variables which are the real and imaginary parts of the complex variables and thus the PDFs are functions of \(2B\) real variables \((d = 2B)\). SEND was derived such that it has spherical contours and the L2-norm of the variables \(||s_k||_2\) follows exponential distribution.

Each distribution can be regarded as an extension of Laplace distribution, a.k.a. double exponential distribution. These distributions shrink to the Laplace distribution if \(d = 1\). As \(ML_{N-I(p,m)}\), we will denote the likelihood contrast function with the source target of SSL and SEND by \(ML_{SSL}\) and \(ML_{SEND}\), respectively.

For better understanding of spherically symmetric sparse PDFs, the density plots and the the contour plots of the two-dimensional independent Laplace distribution and SSL are compared in Fig. 3.1 where the variance of each marginal density is set to be one. By comparing the density of the points around the diagonals in the contour plots, it can be seen that the ones in SSL have higher density than the ones in independent Laplace distribution which implies that the variables of SSL have some variance dependency.

When we replace \(\hat{f}_{s_k}(\cdot)\) in the likelihood contrast (2.23) with either SEND or SSL, the likelihood contrast can be written in the form of

\[
\sum_k \hat{E}[G\left(\sum_b |y_k^b|^2\right)]
\]
Figure 3.1: The two-dimensional density plots and the contour plots of (a) independent Laplace distribution and (b) spherically symmetric Laplace (SSL) distribution. When being compared with independent Laplace distribution, it can be seen that SSL has some variance dependency.

with the relation of

$$ G\left(\sum_b |y_b^i|^2\right) = -\log \hat{f}_{sk}(y_k), $$

(3.8)

where $y_k^i = (w_k^i)^H x^b$ and $(w_k^i)^H$ denotes the $i$-th row of the separating matrix $W^b$ with the normalization constraint of $(w_k^i)^H w_k^i = 1$. Note that the contrast has
changed its sign to be negative likelihood.

3.2 Contrast Optimization

ICA algorithms mainly consist of two parts. One is the choice of contrast function and the other is the choice of the optimization method. For optimization, a number of ICA algorithms use gradient descent update rule including natural gradient [Amari et al., 1996], or relative gradient [Cardoso, 1995]. Also fixed-point iteration [Hyvärinen and Oja, 1997], Newton’s method update rule [Hyvärinen, 1999], or exhaustive search by rotation [Learned-Miller and Fisher III, 2003] have been applied. Since the contrast functions were selected, proper optimization methods need to be chosen for the IVA algorithms to be derived. Since gradient descent methods are straightforward, here we employ the Newton’s method update rule. Newton’s method update, when compared to other gradient descent methods, converges fast and is free from selecting the learning rate. In addition, Newton’s method would find rather stationary points than specified maxima or minima, and thus it has more flexibility in separating sources; for instance, a FastICA algorithm which is derived from ML perspective and is designed for super-Gaussian sources can also find sub-Gaussian sources.

3.2.1 Quadratic Taylor Expansion in Complex Notations

In order to use complex notations while applying Newton’s method to the contrasts, we employ the definitions of complex gradient and complex Hessian in [Brandwood, 1983] and [van den Bos, 1994]. It was shown that a complex variable and its complex conjugate must be considered as separate variables and that the complex gradient and Hessian correspond to their real counterparts one-to-one by simple linear transformations [van den Bos, 1994]. That is, for

\[ \mathbf{w} \ (= [w_1, w_2, \cdots]^T) = \mathbf{u} + j\mathbf{v} \ (= [u_1, u_2, \cdots]^T + j[v_1, v_2, \cdots]^T) \]  \hspace{1cm} (3.9)

\[ \mathbf{z} \ (= [u_1, v_1, u_2, v_2, \cdots]^T) \]  \hspace{1cm} (3.10)

\[ \mathbf{w}_D \ (= [w_1, w_1^*, w_2, w_2^*, \cdots]^T) \]  \hspace{1cm} (3.11)
\[ g(w) = g_D(w, w^*) \]
\[ = h(u, v) \]

where \( u \) and \( v \) are real-valued vectors, it was shown that the quadratic Taylor polynomial of \( h(u, v) \) with respect to \( z \) around the point \( z = 0 \) is equivalent to a Taylor polynomial form in complex notations as the following:
\[
\begin{align*}
\frac{\partial h(O, O)}{\partial z}z &+ \frac{1}{2}z^T \frac{\partial^2 h(O, O)}{\partial z \partial z^T}z = g_D(O, O) + \frac{\partial g_D(O, O)}{\partial w^T D} w_D + \frac{1}{2} w_D^T \frac{\partial^2 g_D(O, O)}{\partial w_D \partial w_D^T} w_D 
\end{align*}
\]

if (and only if),
\[
\begin{align*}
\frac{\partial}{\partial w_k} &= \frac{1}{2} \left( \frac{\partial}{\partial u_k} - j \frac{\partial}{\partial v_k} \right), & \frac{\partial}{\partial w_k^*} &= \frac{1}{2} \left( \frac{\partial}{\partial u_k} + j \frac{\partial}{\partial v_k} \right), \\
\frac{\partial^2}{\partial a \partial b} &= \frac{\partial}{\partial a} \left( \frac{\partial}{\partial b} \right) = \frac{\partial}{\partial b} \left( \frac{\partial}{\partial a} \right) = \frac{\partial^2}{\partial a \partial b} \quad a, b \in \{w_1, w_2^*\}. 
\end{align*}
\]

However, although in complex notations, it is still cumbersome to apply Newton’s method to a contrast using the quadratic Taylor polynomial form on the right-hand-side of (3.15). Here, we introduce a quadratic Taylor polynomial in notations of the original complex variables \( w \) and \( w^* \) instead of \( w_D \). By changing the order of the components in \( w_D \) such that \( w_D = [w; w^*] \) and plugging it into the right-hand-side of (3.15), it can be seen that the expression decomposes into
\[
\begin{align*}
g(w) &\approx g(w_o) + \frac{\partial g(w_o)}{\partial w^T} (w - w_o) + \frac{\partial g(w_o)}{\partial w^H} (w - w_o)^* \\
&+ \frac{1}{2} (w - w_o)^T \frac{\partial^2 g(w_o)}{\partial w \partial w^T} (w - w_o) \\
&+ \frac{1}{2} (w - w_o)^H \frac{\partial^2 g(w_o)}{\partial w^* \partial w^H} (w - w_o)^* \\
&+ (w - w_o)^H \frac{\partial^2 g(w_o)}{\partial w^* \partial w^T} (w - w_o)
\end{align*}
\]

where the point of Taylor expansion has changed from \( w = O \) to \( w = w_o \). The above notations we used for gradient and Hessian with vectors are defined as
\[
\begin{align*}
\frac{\partial}{\partial w} &= \left[ \frac{\partial}{\partial w_1}, \frac{\partial}{\partial w_2}, \ldots \right]^T, & \frac{\partial}{\partial w^T} &= \left( \frac{\partial}{\partial w} \right)^T, \\
\frac{\partial}{\partial w^*} &= \left[ \frac{\partial}{\partial w_1^*}, \frac{\partial}{\partial w_2^*}, \ldots \right]^T, & \frac{\partial}{\partial w^H} &= \left( \frac{\partial}{\partial w^*} \right)^T,
\end{align*}
\]
\[ \frac{\partial^2}{\partial a \partial b^T} = \frac{\partial}{\partial a} \left( \frac{\partial}{\partial b^T} \right) = \frac{\partial}{\partial b^T} \left( \frac{\partial}{\partial a} \right) = \frac{\partial^2}{\partial b^T \partial a^T} \quad \text{for} \quad a, b \in \{ w, w^* \}. \quad (3.21) \]

Although this form is enough for the derivation of a Newton’s method update as it was done in [Lee et al., 2006, 2007a], to be strict we can employ the quadratic Taylor expansion form with respect to the whole set of \( w_k^b \)'s instead of a single \( w_k^b \) since the contrast of IVA is a function of the whole \( w_k^b \)'s:

\[
g(\{ w_i^b \}) \approx \sum_{b_1} \frac{\partial g(\{ w_i^b \})}{\partial (w_i^b)_{T}} (w_i^b - w_{k,o}^b) + \sum_{b_1} \frac{\partial^2 g(\{ w_i^b \})}{\partial (w_i^b)^H} (w_i^b - w_{k,o}^b)^* \\
+ \sum_{b_1, b_2} \frac{1}{2} (w_i^b - w_{k,o}^b)^T \frac{\partial^2 g(\{ w_i^b \})}{\partial w_k^b \partial (w_k^b)^T} (w_i^b - w_{k,o}^b) \\
+ \sum_{b_1, b_2} \frac{1}{2} (w_i^b - w_{k,o}^b)^H \frac{\partial^2 g(\{ w_i^b \})}{\partial w_k^b \partial (w_k^b)^T} (w_i^b - w_{k,o}^b)^* \\
+ \sum_{b_1, b_2} (w_i^b - w_{k,o}^b)^H \frac{\partial^2 g(\{ w_i^b \})}{\partial (w_k^b)^* \partial (w_k^b)^T} (w_i^b - w_{k,o}^b). \quad (3.22) \]

### 3.2.2 Newton’s Method

Let’s set \( g(\{ w_k^b \}) \) to be the summation term of the contrast in (3.7) as

\[
g(\{ w_k^b \}) = \tilde{E} \left[ G \left( \sum_f (w_k^b)^H x^b \right)^2 \right] - \sum_f \lambda_k^b ((w_k^b)^H w_k^b - 1), \quad (3.23) \]

where the normalization constraint is added with Lagrange multiplier \( \lambda_k^b \)'s. Please note that the other contrast function, \( \text{ML}_{N-1}(p, m) \), can also be optimized by Newton’s method in the same manner. The \( w_i^{b_1} \) that optimizes the function \( g(\{ w_k^b \}) \) will set the gradient \( \frac{\partial g(\{ w_k^b \})}{\partial (w_k^{b_1})^*} \) to be zero and hence from (3.22),

\[
\begin{align*}
\frac{\partial g(\{ w_k^{b_1} \})}{\partial (w_k^{b_1})^*} + \sum_{b_2} \frac{\partial^2 g(\{ w_k^b \})}{\partial (w_k^{b_1})^* \partial (w_k^{b_2})^T} (w_i^{b_2} - w_{k,o}^{b_2}) \\
+ \sum_{b_2} \frac{\partial^2 g(\{ w_k^{b_1} \})}{\partial (w_k^{b_1})^* \partial (w_k^{b_2})^H} (w_i^{b_2} - w_{k,o}^{b_2})^* \\
\approx 0. \quad (3.24)
\end{align*}
\]
Then the derivative terms in (3.24) become

\[
\frac{\partial g(w_{k,o}^b)}{\partial (w_{k}^{b_1})^*} = \tilde{E} \left[ (y_{k,o}^b)^* G' \left( \sum_b |y_{k,o}^b|^2 \right) x_{k}^{b_1} \right] - \lambda_{k}^b w_{k,o}^{b_1},
\]

(3.25)

\[
\frac{\partial^2 g\{w_{k,o}^b\}}{\partial (w_{k}^{b_1})^* \partial (w_{k}^{b_2})^T} = \tilde{E} \left[ (G' \left( \sum_b |y_{k,o}^b|^2 \right) + |y_{k,o}^b|^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) x_{k}^{b_1} (x_k^{b_1})^H \right] - \lambda_{k}^b I
\]

(3.26)

\[
= \tilde{E} \left[ G' \left( \sum_b |y_{k,o}^b|^2 \right) + |y_{k,o}^b|^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] \tilde{E} [x_{k}^{b_1} (x_k^{b_1})^H] - \lambda_{k}^b I
\]

(3.27)

\[
\approx \tilde{E} \left[ G' \left( \sum_b |y_{k,o}^b|^2 \right) + |y_{k,o}^b|^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] I - \lambda_{k}^b I
\]

(3.28)

\[
= \left( \tilde{E} \left[ G' \left( \sum_b |y_{k,o}^b|^2 \right) + |y_{k,o}^b|^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] - \lambda_{k}^b \right) I,
\]

(3.29)

\[
\frac{\partial^2 g\{w_{k,o}^b\}}{\partial (w_{k}^{b_1})^* \partial (w_{k}^{b_2})^T} = \tilde{E} \left[ (y_{k,o}^b)^* y_{k,o}^{b_2} G'' \left( \sum_b |y_{k,o}^b|^2 \right) x_{k}^{b_1} (x_k^{b_2})^H \right]
\]

(3.30)

\[
= \tilde{E} \left[ (y_{k,o}^b)^* y_{k,o}^{b_2} G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] \tilde{E} [x_{k}^{b_1} (x_k^{b_2})^H]
\]

(3.31)

\[
\approx \tilde{E} \left[ (y_{k,o}^b)^* y_{k,o}^{b_2} G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] I
\]

(3.32)

\[
= \tilde{E} \left[ (y_{k,o}^b)^* y_{k,o}^{b_2} G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] O,
\]

(3.33)

\[
= O,
\]

(3.34)

\[
\frac{\partial^2 g\{w_{k,o}^b\}}{\partial (w_{k}^{b_1})^* \partial (w_{k}^{b_1})^H} = \tilde{E} \left[ ((y_{k,o}^b)^*)^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) x_{k}^{b_1} (x_k^{b_1})^T \right]
\]

(3.35)

\[
= \tilde{E} \left[ ((y_{k,o}^b)^*)^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] \tilde{E} [x_{k}^{b_1} (x_k^{b_1})^T]
\]

(3.36)

\[
= O,
\]

(3.37)

\[
\frac{\partial^2 g\{w_{k,o}^b\}}{\partial (w_{k}^{b_1})^* \partial (w_{k}^{b_2})^H} = \tilde{E} \left[ ((y_{k,o}^b)^*)^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) x_{k}^{b_1} (x_k^{b_2})^T \right]
\]

(3.38)

\[
= \tilde{E} \left[ ((y_{k,o}^b)^*)^2 G'' \left( \sum_b |y_{k,o}^b|^2 \right) \right] \tilde{E} [x_{k}^{b_1} (x_k^{b_2})^T]
\]

(3.39)

\[
= O,
\]

(3.40)
where \( y_{k,o}^b = (w_{k,o}^b)^H x^b \) and some approximations by separation of expectations were done in (3.28), (3.32), (3.37), and (3.41) as was done in [Bingham and Hyvärinen, 2000]. (3.29), (3.33), (3.38), and (3.42) follow from (2.13) and (3.2), respectively.

By putting all results into the Newton’s method in (3.24), our corresponding iterative algorithm becomes as follows,

\[
\begin{align*}
\mathbf{w}_{k,o}^{b_1} &\leftarrow \mathbf{w}_{k,o}^{b_2} - \frac{\tilde{E}\left[(y_{k,o}^{b_1})^* G'(\sum_b |y_{k,o}^b|^2) x^{b_1}\right] - \lambda_k^{b_1} \mathbf{w}_{k,o}^{b_2}}{\tilde{E}\left[G'(\sum_b |y_{k,o}^b|^2) + |y_{k,o}^{b_1}|^2 G''\left(\sum_b |y_{k,o}^b|^2\right)\right] - \lambda_k^{b_1}},
\end{align*}
\]

where it can be easily evaluated that the Lagrange multiplier \( \lambda_k^{b_1} \) should be

\[
\lambda_k^{b_1} = \tilde{E}\left[|y_{k,o}^{b_1}|^2 G'(\sum_b |y_{k,o}^b|^2)\right].
\]

Although we derived the Newton step with respect to the whole set of \( \mathbf{w}_{k,o}^b \)'s, it is to be seen that the result is the same as the one in [Lee et al., 2006, 2007a] where the Newton step was derived with respect to a single \( \mathbf{w}_{k,o}^b \). As it can be seen from the derivation, it resulted from spatially white source data.

Also, instead of evaluating \( \lambda_k^{b_1} \), we can remove it by multiplying the numerator in (3.43) on both sides of the equation. Hence, with the need of normalization, the learning rule becomes

\[
\mathbf{w}_{k,o}^{b_1} \leftarrow \tilde{E}\left[G'(\sum_b |y_{k,o}^b|^2) + |y_{k,o}^{b_1}|^2 G''\left(\sum_b |y_{k,o}^b|^2\right)\right] \mathbf{w}_{k,o}^{b_1}
- \tilde{E}\left[(y_{k,o}^{b_1})^* G'(\sum_b |y_{k,o}^b|^2) x^{b_1}\right].
\]

In addition to normalization, the rows of the separating matrix \( \mathbf{W}^b \)'s need to be decorrelated. Since the contrasts are not one-unit contrast functions, the symmetric decorrelation scheme

\[
\mathbf{W}^b \leftarrow (\mathbf{W}^b (\mathbf{W}^b)^H)^{-\frac{1}{2}} \mathbf{W}^b
\]

fits better than the deflationary decorrelation scheme.

It is known that many local optima exist in a cost function framework which results from wrong permutation [Ikram and Morgan, 2000, Davies, 2002]. In our algorithms we employ identity matrix as the initial separating matrix in all frequency bins in order to be helped avoiding those local optima.
3.3 Experiments

Here we show the speech separation result of IVA algorithms when applied to real data that are recorded in a real office environment. Detailed experimental results with simulated data can be seen in [Lee and Lee, 2007, Lee et al., 2007a, Jang et al., 2007]. Three human voices and a piece of hip-hop music played from a stationary speaker were captured by four equidistantly aligned microphones. The sources were located approximately $1m \sim 2m$ away from the microphones. 8-second long data were used. Also, 2048-point FFT and a 2048-tab Hanning window with the shift size of 512 samples were chosen.

Fig. 3.2 shows the real room environment where the source mixtures in Fig. 3.3 were recorded. Fig. 3.4 shows the mixed separated sources by an ML IVA algorithm that has SEND as its source prior, respectively. Note that the first plot in Fig.3.4 is the separated piece of music, and that, IVA is successful and robust even in a very challenging scenario.

3.4 Conclusions

In this chapter, various multivariate probability density models for speech signals in the IVA framework have been studied and various fast fixed-point algorithms have been derived. In ML approach, IVA uses a multivariate source prior that captures the inter-bin dependence. The multivariate source prior is a vector representation of the ICA where the multivariate sources are assumed independent while the elements in the source are dependent. This adds a source constraint to the ICA learning so that each source is learned with the inter-bin relation and hence the permutation problem can be solved. In IVA, similar to ICA, we have certain flexibilities in defining or modeling the multidimensional source prior of speech. For speech, while modeling the source prior in a more accurate manner will lead to better performance, joint PDFs that are sparse and spherically symmetric or alike have proven to be proper and effective models.
3.5 Acknowledgement

This material is based upon work partially supported by the national science foundation (NSF) under grant No. 0535251. Also, I would like to acknowledge Taesu Kim and Te-Won Lee as the co-authors of the works included in this chapter.
Figure 3.3: The recorded observation signals.
Figure 3.4: The separation results of the Fast fixed-point IVA algorithm with \textit{ML}_{\text{SEND}} contrast.
Independent Vector Analysis Using Densities Represented by Chain-like Overlapped Cliques in Graphical Models for Separation of Convolutedly Mixed Speech Signals

The radially symmetric multivariate source priors originally employed in IVA for the separation of acoustic signals is

\[ f(z) \propto e^{-\frac{1}{2} \sqrt{\sum_{b=1}^{B} |z_b|^2}}, \]

where \( z = [z_1, \ldots, z_B]^T \) is a multivariate dummy variable and \( \sigma \) adjusts the variance of the variables. Such a super-Gaussian, hyper-spherical PDF can capture the positive correlation of the magnitudes of the variables. The overall hyper-spherical structure of the density, however, deviates from the real case since the STFT components of an acoustic signal exhibit a dependence pattern where the positive correlation of the magnitudes becomes weaker as the frequency difference increases unless they are in harmonic relation. Such discrepancy suggests that the performance of IVA for convolutive BSS could be improved by employing a multivariate source prior that models the inter-bin dependence type better.
4.1 Proposed Model and Method

As an undirected graph in graphical models, the overall hyper-spherical densities in previous methods can be represented by a global clique and is depicted in Fig. 4.1(a). Another undirected graph that better resembles the dependence pattern of the STFT components would be a chain-like series of overlapped cliques which is depicted in Fig. 4.1(b). In this model, the dependence between the components propagate through the chained overlaps and it weakens as the difference between the component indices increases.

Figure 4.1: Undirected graphs for IVA dependence models. (a) A global clique to represent the super-Gaussian and radially symmetric dependence. (b) Chain-like series of overlapped cliques to represent the proposed dependence.

By combining the dependence pattern in (4.1) and the graphical model in Fig. 4.1(b), we propose a dependence model as the multivariate PDF that is represented by the undirected graph of overlapped cliques:

\[
    f(z) \propto e^{-\sqrt{\sum_{b=f_1}^{l_1} \frac{z_b}{\sigma_{zb}}^2} - \sqrt{\sum_{b=f_2}^{l_2} \frac{z_b}{\sigma_{zb}}^2} - \cdots - \sqrt{\sum_{b=f_C}^{l_C} \frac{z_b}{\sigma_{zb}}^2}},
\]

where \( \sigma_{zb} \) adjusts the variance of the variables and \( C \) is the number of cliques. \( f_c \) and \( l_c \) are, respectively, the first and the last indices of the \( c \)-th clique satisfying

\[
    f_c < l_{c-1}, \quad c = 2, 3, \ldots, C,
\]

such that the series of cliques have chain-like overlaps.
When employing the new dependence model in the framework of IVA for convolutive BSS, the log-likelihood function becomes as follows:

\[
L \left( \{W^b\} \right) = \sum_{b=1}^{B} \log |\det(W^b)| - \sum_{k=1}^{K} \tilde{E} \left[ \sum_{c=1}^{C} \left| \sum_{b=fc}^{l_c} \frac{y_k^b}{\sigma_{cb}} \right|^2 \right]. \tag{4.4}
\]

We apply natural gradient learning rule [Amari et al., 1996] for its optimization:

\[
\Delta W^b = \frac{\partial L}{\partial (W^b)^*} (W^b)^H W^b. \tag{4.5}
\]

### 4.2 Experiments

In order to evaluate the performance of the IVA algorithm with the new dependence model, we simulated various 2×2 speech separation problems in the time domain using 8-second long real speech signals sampled at 8 kHz and room impulse responses obtained by an image method [Allen and Berkley, 1979]. The geometric configuration of the simulated room environment is depicted in Fig. 4.2. We set the room size to be 7 m \times 5 m \times 2.75 m and set all heights of the microphone and source locations to be 1.5 m. 100 ms was chosen as the reverberation time and the corresponding reflection coefficients were set to be 0.57 for every wall, floor, and ceiling.

When applying the algorithm to the simulated BSS problems in the STFT domain, 2048-point FFT, 2048-tab Hanning window, and a shift size of 512 samples were used. In the proposed multivariate PDF, local cliques were chosen to be of the same size to include either 50% or 25% of all frequency bins and the cliques are shifted by either one bin or half the clique size. For comparison, the original IVA using the joint PDF in (4.1) has also been evaluated. The separation performances were measured by the signal to interference ratio (SIR) in dB defined as

\[
\text{SIR} = 10 \log \left( \frac{\sum_{n,b} \sum_{k=1}^{K} \left| h_{kq(k)}^b s_{q(k)}^b[n] \right|^2}{\sum_{n,b} \sum_{k \neq l} \left| h_{kq(l)}^b s_{q(l)}^b[n] \right|^2} \right), \tag{4.6}
\]

where \(q(k)\) indicates the separated source index that the \(k\)-th source appears and \(h_{kq(l)}\) is the overall impulse response defined as \(h_{kq(l)} = \sum_{i=1}^{K} w_{ki}^b a_{q(l)}^b\), where \(a_{kl}^b\) and
Figure 4.2: Geometric configuration of the simulated room environment. For each experiment, two source locations were chosen from A to K and the microphone locations were fixed as shown in the figure.

$w_{kl}^b$ denote the $(k, l)$-th component of $A^b$ and $W^b$, respectively. The results are shown in Tab. 4.1. The proposed algorithm consistently outperformed the original IVA in terms of SIR.

Table 4.1: Separation performances in SIR (dB). Proposed methods are denoted by “the number of bins in each clique” and “the number of bins of each shift”.

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source locations</td>
<td>A,I</td>
<td>B,G</td>
<td>E,G</td>
<td>I,K</td>
<td>C,D</td>
<td>E,F</td>
<td>I,J</td>
</tr>
<tr>
<td>Original method</td>
<td>17.1</td>
<td>18.2</td>
<td>17.3</td>
<td>12.4</td>
<td>15.8</td>
<td>15.6</td>
<td>15.1</td>
</tr>
<tr>
<td>512-bin cliques, 256-bin shift</td>
<td>22.9</td>
<td>20.1</td>
<td>19.6</td>
<td>15.0</td>
<td>17.5</td>
<td>19.4</td>
<td>18.6</td>
</tr>
<tr>
<td>512-bin cliques, 1-bin shift</td>
<td>23.1</td>
<td>20.3</td>
<td>19.9</td>
<td>15.0</td>
<td>17.6</td>
<td>19.7</td>
<td>18.6</td>
</tr>
<tr>
<td>256-bin cliques, 128-bin shift</td>
<td>24.1</td>
<td>19.7</td>
<td>19.9</td>
<td>15.1</td>
<td>17.6</td>
<td>19.5</td>
<td>18.3</td>
</tr>
<tr>
<td>256-bin cliques, 1-bin shift</td>
<td>24.2</td>
<td>20.6</td>
<td>20.0</td>
<td>15.0</td>
<td>17.7</td>
<td>19.5</td>
<td>18.5</td>
</tr>
</tbody>
</table>
4.3 Conclusions

Modeling the inter-frequency dependence of speech signals in a more accurate manner leads to a more appropriate representation. This representation is captured by the vector representation of the multivariate source by means of an undirected graph with chain-like series of overlapped cliques in graphical models. By the graphical representation it is possible to extend this approach to other forms of PDFs. The impact of this approach could be far more significant for natural signals where complex multidimensional signal dependence is essential.

4.4 Acknowledgement

This material is based upon work partially supported by the national science foundation (NSF) under grant No. 0535251. Also, I would like to acknowledge Gil-Jin Jang and Te-Won Lee as the co-authors of the work presented in this chapter.
Adaptive Independent Vector Analysis for The Separation of Convoluted Mixtures using EM Algorithm

Most IVA algorithms have been derived for the separation of speech signals in the maximum likelihood (ML) framework. The source priors were pre-specified by products of simple multidimensional super-Gaussian densities and only the separating matrices were estimated. Generally, however, it is difficult to model the collective frequency components of various sources and the true source models are mostly highly complicated and unknown. Hence, it can be easily expected that flexible and more accurate source priors will lead to better separation performance.

In this chapter, we propose a novel adaptive IVA algorithm to separate the mixture of convolutedly mixed signals in the presence of sensor noise. Motivated by independent factor analysis (IFA) [Attias, 1999], we model the joint probability density function (PDF) of the collective frequency components by multivariate Gaussian mixture model (GMM) and allow sensor noise which has not been considered in the previous ML approaches of IVA. An efficient expectation maximization (EM) algorithm is derived to estimate the mixing matrices and the parameters of an unknown source prior together. Signal estimation is achieved through Bayesian inference by computing the minimum mean squared error (MMSE) of the signal
5.1 The Adaptive Independent Vector Analysis Model

In this section, we define the acoustic model for convoluted mixing in both the time domain and the short-time Fourier transform (STFT) domain and introduce the multivariate GMM which will serve as the source priors.

5.1.1 Acoustic Model for Convoluted Mixing

The acoustic model for convoluted mixing can be described as,

\[ X_l[n] = \sum_k \sum_i a_{lk}[i] S_k[n-i] + u_l[n], \]

where \( a_{lk} \) is the time-domain transfer function from the \( k \)-th source to the \( l \)-th observation, \( S_k[n] \) is the \( k \)-th source signal at time \( n \), \( u_l[n] \) is the noise. Here in this paper, we will only consider the situation of 2 sources and 2 microphones. A generalization to multiple sources is straightforward.

The separation of convoluted mixtures can be tackled more conveniently when (approximately) converted into a instantaneous linear mixing model in the frequency domain by STFT. Here we have

\[ x^b_l = A^b s^b_l + u^b_l. \]

5.1.2 Multivariate Density Models of a Signal

In contrast with those IVA algorithms where the source signals were modeled by identical multivariate densities, we employ the flexible multivariate GMM as the source prior:

\[ p(s_{kt}) = \sum_{s_{kt}} p(s_{kt}) p(s_{kt}|s_{kt}) = \sum_{s_{kt}} p(s_{kt}) \prod_b N(s^b_{kt}|0, \nu^b_{s_{kt}}). \]
where $s_{kt} = [s_{k1t}^1, \cdots, s_{k1t}^B]^T$ and $\varsigma_{kt}$ denotes the state of the mixture model. Here, $\mathcal{N}(s_{kt}^b|0, \nu_{\varsigma kt}^b)$ is the Gaussian density for complex variables with precision $\nu_{\varsigma kt}^b$ (inverse of covariance), i.e.

$$\mathcal{N}(s_{kt}^b|0, \nu_{\varsigma kt}^b) = \frac{\nu_{\varsigma kt}^b}{\pi} e^{-\nu_{\varsigma kt}^b |s_{kt}^b|^2}. \quad (5.5)$$

Although we assume diagonal precision matrix for each (conditional) multivariate Gaussian density $p(s_{kt}^b|\varsigma_{kt})$, their mixture $p(s_{kt})$ imposes dependency on the components, which is essential to prevent the permutation problem in IVA. In addition, of course, we assume independence among multivariate sources. We also assume Gaussian noise,

$$p(u_{kt}^b) = \mathcal{N}(u_{kt}^b|0, \lambda^b) = \frac{\lambda^b}{\pi} e^{-\frac{|u_{kt}^b|^2}{\lambda^b}}. \quad (5.6)$$

In the limit where $\lambda^b$ goes to infinity, the acoustic model in (5.2) reduces to noiseless model.

Although the parameters of both source priors $p(s_{1t})$ and $p(s_{2t})$ can be learned blindly, in this paper we deal with the case when one source type, e.g. speech, is known such that $p(s_{1t})$ can be trained a priori by the same type of sources and be fixed. Please note that there are many cases when the type of the signal to be cleaned is specified in advance, e.g. speech enhancement, and thus by using its pre-trained source prior the number of data required for proper learning can be reduced significantly. The parameters of the source prior $p(s_{2t})$, as well as the mixing matrices $\{A^b\}$, are estimated from the data.

### 5.2 EM Algorithm for Parameter Estimation

The unknown parameters $\theta = \{A^b, p(\varsigma_{2t}), \nu_{\varsigma_{2t}}^b, \lambda^b\}$ can be estimated via EM algorithm.

**E-step:** The posteriors of the source signal can be obtained by

$$\log q(s_{1t}^b, s_{2t}^b|s_{1t}^b, s_{2t}^b) \propto \log p(x_{1t}^b, x_{2t}^b|s_{1t}^b, s_{2t}^b) + \log p(s_{1t}^b|\varsigma_{1t}) + \log p(s_{2t}^b|\varsigma_{2t}) + c. \quad (5.7)$$
Because of the GMM source prior, the right-hand-side (RHS) of the above equation is quadratic in \( x_{1t}^b \) and \( x_{2t}^b \). Thus the signal posterior conditioned on the states \( \varsigma_1 \) and \( \varsigma_2 \) is Gaussian:

\[
q(s_{1t}^b, s_{2t}^b | \varsigma_1 t, \varsigma_2 t) = \mathcal{N}(s_{1t}^b, s_{2t}^b | \mu_{\varsigma_1 \varsigma_2 t}^b, \Phi_{\varsigma_1 \varsigma_2 t}^b), \tag{5.8}
\]

whose precision and mean are, respectively,

\[
\Phi_{\varsigma_1 \varsigma_2 t}^b = \lambda^b (A^b)^T A^b + \left( \begin{array}{cc} \nu_{\varsigma_1 t}^b & 0 \\ 0 & \nu_{\varsigma_2 t}^b \end{array} \right) \tag{5.9}
\]

and

\[
\mu_{\varsigma_1 \varsigma_2 t}^b = \lambda^b (\Phi_{\varsigma_1 \varsigma_2 t}^b)^{-1} (A^b)^T y_t^b. \tag{5.10}
\]

To compute the posterior state probability, we need to evaluate \( p(x_{1t}^b, x_{2t}^b | \varsigma_1 t, \varsigma_2 t) \), which is Gaussian with zero mean and precision matrix \( \Sigma_{\varsigma_1 \varsigma_2 t}^b \) given by

\[
(\Sigma_{\varsigma_1 \varsigma_2 t}^b)^{-1} = A^b \left( \begin{array}{cc} \frac{1}{\nu_{\varsigma_1 t}^b} & 0 \\ 0 & \frac{1}{\nu_{\varsigma_2 t}^b} \end{array} \right) (A^b)^T
+ \left( \begin{array}{cc} \frac{1}{\bar{x}} & 0 \\ 0 & \frac{1}{\bar{x}} \end{array} \right). \tag{5.11}
\]

Let’s define \( f_{\varsigma_1 \varsigma_2 t}(t) \) as the following:

\[
f_{\varsigma_1 \varsigma_2 t}(t) \\
= \log p(x_{1t}, x_{2t} | \varsigma_1 t, \varsigma_2 t) + \log p(\varsigma_1 t) + \log p(\varsigma_2 t) \tag{5.12}
\]

\[
= \sum_b \log p(x_{1t}^b, x_{2t}^b | \varsigma_1 t, \varsigma_2 t) + \log p(\varsigma_1 t) + \log p(\varsigma_2 t), \tag{5.13}
\]

where \( x_{kt} = [x_{kt}^1, \cdots, x_{kt}^B]^T \). Then, since

\[
\log q(\varsigma_1 t, \varsigma_2 t | x_{1t}, x_{2t}) \propto f_{\varsigma_1 \varsigma_2 t}(t), \tag{5.14}
\]

the posterior state probability can be computed as

\[
q(\varsigma_1 t, \varsigma_2 t | x_{1t}, x_{2t}) = \frac{1}{Z_t} e^{f_{\varsigma_1 \varsigma_2 t}(t)}, \tag{5.15}
\]

where

\[
Z_t = \sum_{\varsigma_1 \varsigma_2 t} e^{f_{\varsigma_1 \varsigma_2 t}(t)}. \tag{5.16}
\]
**M-step:** The update rules for mixing matrices \( \{A^b\} \) are

\[
A^b = \left( \sum_t < x_t^b (s_t^b)^T >_q \right) \left( \sum_t < s_t^b (s_t^b)^T >_q \right)^{-1},
\]

(5.17)

where \(< \cdot >_q\) denotes expectation over \(q\).

The update rules for the precisions of the source prior are

\[
\frac{1}{\nu^{b_{\varsigma_t}}} = \frac{1}{N} \sum_{t,c_1,c_2} q(s_{c_1},s_{c_2}|x_{1t},x_{2t}) ((\Phi_{s_{c_1},s_{c_2}})^{-1}_{22}) \sum_{t,c_1,c_2} q(s_{c_1},s_{c_2}|x_{1t},x_{2t}) \mu_{s_{c_1},s_{c_2}}^b + \frac{1}{N} \sum_t q(s_2|x_{1t},x_{2t}).
\]

(5.18)

where \((M)^{2,2}\) denotes the \((2, 2)\)-th element of the matrix \(M\). The state probability of source prior is computed as

\[
p(s_2) = \frac{1}{N} \sum_t q(s_2|x_{1t},x_{2t}).
\]

(5.19)

The update rules for noise precisions \(\lambda^b\) are given by

\[
\frac{2N}{\lambda^b} = \sum_t x_t^b(x_t^b)^T - \text{Tr}(A^b < s_t^b (s_t^b)^T >_q) - \text{Tr}\left((A^b)^T < x_t^b (s_t^b)^T >_q\right) + \text{Tr}\left((A^b)^T A^b < s_t^b (s_t^b)^T >_q\right).
\]

(5.20)

where \(\text{Tr}(\cdot)\) stands for the trace operation.

**Signal Estimation and Scaling** For noiseless ICA, original sources can be estimated by applying the inverse of the mixing matrices, \(\{(A^b)^{-1}\}\), to mixed observation. However, this approach is not optimal if sensor noise is considered. We use MMSE estimator by computing the mean of posterior distribution \(q(s_t^b|x_t^b)\),

\[
\bar{s}_t = < s_t >_q = \sum_{s_{c_1},s_{c_2}} q(s_{c_1},s_{c_2}|x_{1t},x_{2t}) \mu_{s_{c_1},s_{c_2}}^b.
\]

(5.22)

where \(\mu_{s_{c_1},s_{c_2}}^b\) is given in (5.10).

Since ICA and IVA also suffer from scaling problem, for proper signal reconstruction the well-known minimal distortion principle [Matsuoka and Nakashima, 2001] is applied to \(A^b\) at the end of the learning as

\[
A^b \leftarrow A^b (\text{diag}(A^b)^{-1})
\]

(5.23)
5.3 Experiments

We applied the algorithm to the mixture of speech and music, under noisy condition. 8-second-long clean male speech and a piece of music sampled at 8 kHz were convolved with room impulse responses generated by an image method [Allen and Berkley, 1979] and were mixed together. The mixed signals were then corrupted with white Gaussian noise at the signal to noise ratio (SNR) of 10 dB. In this experiment, we used Hanning window of length 512 samples and shift size of length 128 samples to analyze the signal. A 512-point fast Fourier transform (FFT) was used to obtain frequency domain coefficients. Source one (speech) is modeled by GMM with 10 components trained with standard EM algorithm. The FFT coefficients for each frequency bin are preprocessed by whitening matrices,

\[ Q^b = \left( \langle x^b (x^b)^T \rangle \right)^{-1/2}. \] (5.24)

We initialized \( A_k \)'s to be identity matrices and ran the EM algorithm for 400 iterations. The time domain signal is reconstructed by overlap-adding after applying inverse FFT. The performance was compared with the separation results of IVA algorithm that uses the fixed source prior in [Kim et al., 2007] for both sources (speech and music).

![Figure 5.1: Left: the original sources. Middle: mixed signal corrupted by noise. Right: separated signal.](image)

We used the signal to interference ratio (SIR) defined in (4.6) as the performance measure. The SIR result of our adaptive IVA algorithm was 11.73 dB. (When
Figure 5.2: The Kurtosis of each frequency bin for speech and music.

also the multivariate source prior for music was fixed after learning its parameters from the clean music data, the SIR result slightly increased to be 12.44 dB.) The acoustic wave is shown in Fig. 5.1. The IVA algorithm with the fixed source priors in [Kim et al., 2007] resulted in SIR = 5.89 dB and even for noiseless case the result was as low as 9.01. Perceptually, the separated speech is very clear with almost no noticeable noise and distortion. For the separated music signal, the noise level is significantly reduced, although slight distortion is noticed. The separated signals using IVA with fixed sources in [Kim et al., 2007] has higher interference and contains obvious noise because IVA is unable to denoise.

The kurtosis of the two sources is shown in Fig. 5.2. The music has high kurtosis for low frequency components, while speech has high kurtosis for high frequency components. The figure shows that musical signal is closer to Gaussian (whose kurtosis is 0), because music is the mixture of various instruments. The difference in statistical properties explains that previous IVA approaches work sub-optimally because it assumes identical source prior for both sources.
5.4 Conclusions

We proposed a novel adaptive approach to IVA in order to make up for its weaknesses. In the approach, where multivariate GMM is used as the source priors and sensor noise is allowed, learning the parameters of the source prior, separating the sources, and denoising are achieved simultaneously. We applied the new algorithm to $2 \times 2$ mixture problems where one source type was assumed to be known and thus its source prior could be trained in advance. The new algorithm successfully separated a speech signal from a music signal (which was assumed to be unknown) even in the presence of sensor noise, while the IVA approaches that use multivariate super-Gaussian densities as fixed source priors performed sub-optimally, or poorly with noise.

5.5 Acknowledgement

This material is based upon work partially supported by the national science foundation (NSF) under grant No. 0535251. Also, I would like to acknowledge Jiucang Hao and Te-Won Lee as the co-authors of the work presented in this chapter.
Permutation Correction in Blind Source Separation Using Sliding Subband Likelihood Function

Proportional variance dependency among the frequency components is characteristic of natural signals and has been utilized in frequency-domain blind source separation to solve the permutation problem. In order to increase robustness in such methods, overall measures have been preferred to the measures between directly neighboring frequency components. The overall variance dependency pattern in the fullband, however, can vary by signals and is difficult to be modeled, whereas in smaller subbands the proportional variance dependency is more definite. Here, a novel permutation correction method that utilizes the proportional variance dependency in small subbands is proposed. A windowed likelihood function that uses source priors with internal variance dependency is employed as the measure of permutation correction. This method not only shows robust separation performance but also is computation-wise very efficient.
6.1 Background

The IVA approach is illustrated in Fig. 6.1 in a $2 \times 2$ mixing case. In the figure, the unknown bin-wise mixing process which includes the source signals and the mixing matrices is bracketed and the frequency components of each estimated source signal are grouped together representing a multivariate signal. Instead of applying ICA algorithms bin-wise and correcting the permutation disorder afterwards, IVA deals with the frequency components together as a multivariate signal where there is internal dependency among the components. Thus, using probabilistic approach, multivariate probability density functions (PDFs) are employed as the source priors instead of uni-variate PDFs.

Figure 6.1: The approach of independent vector analysis (IVA) for solving the permutation problem.

As many ICA algorithms employ fixed representative source priors, e.g. Laplace distribution for time-domain speech and other super-Gaussian-distributed signals, IVA algorithms have employed certain super-Gaussian-distributed multivariate PDFs that show proportional variance dependency of the complex-valued
components [Lee et al., 2007b]. However, it is usually difficult to choose a proper source prior for IVA because the complicated dependency among the components has to be captured efficiently in the source prior. And that, since the fast Fourier transform (FFT) frame size needs to be large enough to cover the lengths of the impulse responses, it results in high dimensionality of the source signals. In Fig. 6.2, the normalized covariance matrix of the magnitudes of the frequency components are shown for two pieces of speech and a piece of music. As it can be seen, the dependency types are complicated and they even differ among speeches. Hence, when the source signal is unknown, errors in the fixed source priors are inevitable and such errors can degrade the separation performance of IVA.

Figure 6.2: The covariance matrices of (a) a male speech, (b) a female speech, and (c) a piece of music where the acoustic signals were 12 seconds long.

Instead of using fixed source priors, we can also use ones that are able to adapt to various types of probability densities as the mixture of multivariate Gaussians [Lee et al., 2008]. In this case, however, huge number of data points is essential for learning the unmixing matrices and the parameters of the source priors.

The permutation correction methods also provided reasonable solutions to the permutation problem. However, they also suffer from drawbacks. The DOA method depends much on the robustness of the DOA estimator and its performance easily degrades when the source signals are located in close directions or when the recording environment is reverberant such that the reflections hinder the DOA estimation. In the magnitude covariance or dominance measure methods, the base measure needs pair-wise computation. When using this measure for adjoining pairs only, the permutations can easily flip especially in such bins where the signals are
ill separated or where the signal data is statistically ill behaved. Once flipped in the middle, wrong permutations are likely to follow. Thus, as an overall measure, the pair-wise measures are often evaluated and combined for all possible pairs of frequency signals. However, the overall computation is heavy and inefficient. In addition and more importantly, not all pairs of frequency components show variance dependency or similar dominance patterns. In Fig. 6.3, which is a closer look of Fig. 6.2, it can be seen that significant number of pairs of frequency components do not show proportional variance dependency. A number of pairs even show inversely proportional variance dependency.

6.2 Proposed Permutation Correction Method

Usually, in acoustic signals, there are several types of inter-bin dependency patterns to be seen. Among those, proportional variance dependency is the key factor that has been modeled in the source priors of the previously proposed IVA methods. With respect to proportional variance dependency there are roughly two types. One is the dependency among harmonic frequency components, and the other is the one among neighboring frequency components.

Since the number of harmonic patterns can be infinitely large, it becomes intractable to employ a representative model that captures the harmonic structure of an unknown acoustic signal. On the meanwhile, the dependency pattern in small subbands (as opposed to fullband or large subbands) is relatively less random. In Fig. 6.3, it can be seen that most of the normalized covariance components that are
sufficiently close to the diagonal have positive values for all three acoustic signals. This implies, in sufficiently small subbands, the proportional inter-bin variance-dependency is more definite and stronger.

By taking this into account in the permutation problem, it will be advantageous to apply likelihood functions over sliding subbands. The following are the fullband source priors (of normalized signals) that have been proposed in IVA framework [Kim et al., 2007, Lee et al., 2007b, 2008]:

\[
f(z) = \frac{1}{Z} e^{-\frac{1}{\sigma} \|z\|_p^p}, \tag{6.1}
\]

\[
f(z) = \frac{1}{Z} e^{-\left(\sum_k \frac{1}{\sigma_k} \|z_k\|^2\right)}, \tag{6.2}
\]

\[
f(z) = \sum_k \pi_k \frac{1}{Z_k} e^{-z^T D_k z}. \tag{6.3}
\]

where \( z = [z_1, z_2, \ldots, z_B]^T \) is a multivariate dummy variable denoting the frequency components of a source, \( Z \) and \( Z_k \) are normalization factors, \( \sigma \) and \( \sigma_k \) are the coefficients that control the variance in the PDFs, \( D_k \) is the inverse of the \( k \)-th covariance matrix in a multivariate Gaussian mixture model, and \( \pi_k \) is the state probability of the \( k \)-th multivariate distribution. As mentioned earlier, these source priors are characterized by capturing the proportional variance dependency among the frequency components, and they can be employed as the subband source priors, too. Here in this paper, the results by the spherically symmetric Laplace distribution (\( (6.1) \) when \( p = 2 \) and \( m = 1 \)) will be shown.

### 6.3 Algorithm Description

In large, the algorithm is composed of three parts.

#### 6.3.1 Bin-Wise ICA Separation

For the bin-wise ICA separation, the FastICA algorithm for complex-valued signals is applied [Bingham and Hyvärinen, 2000]. FastICA algorithm spatially pre-whitens the signals and constrains the unmixing matrices \( W^b \)'s to be orthogonal such
that the output signals are uncorrelated. Assuming $x^b$'s are spatially pre-whitened by the following equation

$$x^b = Q^b x^b_{old}, \quad (6.4)$$

the algorithm is written as

$$w^b_k \leftarrow \tilde{E} \left[ G'( |y^b_k|^2 ) + |y^b_k|^2 G'' ( |y^b_k|^2 ) \right] w^b_k - \mu \tilde{E} \left[ x^b_k ( y^b_k )^* G' ( |y^b_k|^2 ) \right] \quad (6.5)$$

with the symmetric decorrelation of

$$W^b \leftarrow (W^b (W^b)^H)^{-1} W^b. \quad (6.6)$$

For the nonlinearity function $G(\cdot)$, the following was chosen:

$$G(z) = \sqrt{z}. \quad (6.7)$$

### 6.3.2 Permutation Correction

As from the previous discussion, the (normalized) subband log-likelihood function with the source prior in (6.1) is employed as the measure of permutation correction:

$$\mathcal{L}_S = -\frac{1}{\sigma} \sum_k \tilde{E} \left[ \sqrt{\sum_{b \in S} (y^b_k)^2} \right] + \sum_{b \in S} \log \left| \det(W^b) \right| \quad (6.8)$$

where $S$ denotes the set of the bins in a sliding subband. Note that the second term in the right-hand-side of the likelihood function can be ignored since the FastICA algorithm that is used for the bin-wise separation constrains the unmixing matrices to be orthogonal. The size of the subband that is chosen in this draft is fixed around 40 Hz. The permutation in the $b$-th bin is determined such that it maximizes the measure $\mathcal{L}_S$ in (6.8) where $S = \{ b - |S| + 1, \cdots, b \}$.

### 6.3.3 Scaling Correction

The minimal distortion principle [Matsuoka and Nakashima, 2001] is applied to fix the scaling problem that arises because of the scaling indeterminacy of ICA.
After learning the unmixing matrices $W^b$'s, the diagonal of $(W^b Q^b)^{-1}$ is multiplied to the left of $W^b Q^b$ [Lee et al., 2007b]:

$$W^b_{\text{final}} \leftarrow \text{diag}((W^b Q^b)^{-1}) W^b Q^b.$$  

(6.9)

### 6.4 Experiments

The BSS algorithm has been applied to $2 \times 2$ speech separation problems. The mixed speech signals were synthetic signals generated in a simulated room environment. In generating synthetic data, 12-second-long clean speech signals were used. Also, 4096-point FFT and a 4096-tab long Hanning window with the shift size of 512 samples were chosen.

The geometric configuration of the simulated room environment is depicted in Fig. 6.4(a). The room size was 7 m $\times$ 5 m $\times$ 2.75 m and the heights of the microphones and sources were 1.5 m. A reverberation time of 200 ms was chosen and the corresponding reflection coefficients were set to 0.57 for every wall, floor, and ceiling. Clean speech signals were convolved with room impulse responses that were obtained by an image method [Allen and Berkley, 1979] using this room configuration. Various $2 \times 2$ case simulations (Fig. 6.4(b)) were carried out. The separation performance was measured by the signal to interference ratio (SIR) in dB. The separation performance of the proposed algorithm was compared with the separation performance of an IVA algorithm that uses the same kind of source prior except that the source prior is overall. Thus the objective function of the compared method is equal to the measure in (6.8) where $S = \{1, \cdots, B\}$. In order to have similar conditions in the algorithms, the FastIVA algorithm [Lee et al., 2007b] that also keeps the output data uncorrelated and uses approximated Newton update optimization method is employed for comparison. The results are shown in Fig. 6.4(c).

### 6.5 Discussion

The idea of using a sliding subband likelihood function is advantageous in that it is computation-wise very efficient. Also, it can be adopted in the IVA frame-
Figure 6.4: The experiments: (a) geometric configuration, (b) experiment numbers and source locations, and (c) separation performances in signal to interference ratio (SIR).

work to yield subband-wise IVA algorithms. Furthermore, in the sliding subbands, not only the maximum likelihood approach but also other machine learning or decision making techniques can be employed.

6.6 Acknowledgement

This material is based upon work partially supported by the national science foundation (NSF) under grant No. 0535251.
Conclusions

In this dissertation, a number of techniques have been proposed and applied to solve the blind source separation (BSS) problem in the short-time Fourier transform (STFT) domain. The bin-wise separation and the permutation correction in STFT-domain BSS can be performed at the same time when using the framework of independent vector analysis (IVA). In IVA, it is essential to employ an effective multivariate source prior that captures the characteristic of, and the dependence among, the STFT components. By testing $L_p$-norm invariant joint densities it was to be seen that super-Gaussian and spherically symmetric multivariate priors meet the purpose. Also, it was further shown that using such multivariate source priors as joint densities represented by chain-like overlapped cliques in graphical models or mixture of multivariate Gaussians can increase the performance of IVA separation. Other than the common natural gradient method, Newton-method-like fixed point method was easily employed in the IVA algorithms when utilizing a Taylor expansion form in complex notations. The number of iterations needed for successful signal separations was to be reduced remarkably. A flexible adaptive method that adds sensor noise in the mixing model and employs mixture of multivariate Gaussians and expectation maximization as the source prior and the optimization method, respectively, was also proposed. It showed to be effective in separating a wider range of signals in noisy conditions. Instead of using the framework of IVA, permutation correction after bin-wise separation is also advantageous in certain aspects. A permutation correction method using a sliding sub-band likelihood function not only
shows robust separation performance in highly reverberant mixing conditions but also is computation-wise very efficient.
Bibliography


Index

Bayesian inference, 41
blind signal/source separation (BSS), 4
complex variables, 27
convolutive mixture, 5
determined BSS, 4
dominance measure, 51
expectation maximization (EM), 41, 43
frequency domain, 5
Gaussian manifold, 20
graphical models, 37
independent component analysis (ICA), 5, 6, 11
independent factor analysis (IFA), 41
independent subspace analysis (ISA), 17
independent vector analysis (IVA), 7, 16, 50
independent vector manifold, 20
information geometry, 13
instantaneous mixture, 4
Kullback-Leibler (KL) divergence, 1
likelihood function, 12
minimal distortion principle, 8, 45, 54
minimum mean squared error (MMSE), 41
multidimensional independent component analysis (MICA), 16
multivariate Gaussian mixture model, 41, 42
multivariate source prior, 23
mutual information, 12
natural gradient, 38
negentropy, 19
Newton’s method, 29
orthogonal matrix, 18
over-determined BSS, 4
permutation indeterminacy, 7, 9
permutation problem, 5, 7
phase ambiguity, 8
Pythagorean relation in information geometry, 20

64
scaling indeterminacy, 9
scaling problem, 5, 8, 45
short-time Fourier transform (STFT), 42
short-time Fourier transform (STFT) domain, 5
signal to interference ratio (SIR), 38, 46
source prior, 13
speech model, 23
spherically invariant random process (SIRP), 24
symmetric decorrelation, 31
Taylor expansion, 27
time-frequency domain, 5
under-determined BSS, 4
undirected graph, 37