DOCTORAL DISSERTATION

BLIND SOURCE SEPARATION BASED ON INDEPENDENT COMPONENT ANALYSIS AND ITS EXTENSION

by

TAKASHI ITAHASHI

March, 2012

Department of Brain Science and Engineering
Graduate School of Life Science and Systems Engineering
Kyushu Institute of Technology

One sees clearly only with the heart.
What is essential is invisible to the eye.

— *The Little Prince, Antonie de Saint–Exupery, 1943.*

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to Professor Kiyotoshi Matsuoka, my supervisor, for his guidance and continuous support. I could not find adequate words to express my gratitude to him. This work could not have been accomplished without his direction and great help. I am also most grateful to the examining committee members evaluating my work and giving me valuable suggestions.

I am deeply grateful to all current and former members of my laboratory for their kind support and valuable comments. I would especially like to thank Mr. Tsubasa Yoshihara, Mr. Akinari Onishi, and Miss Neda Kosibaty.

I would like to express my gratitude to my friends for their advices, encouragements and for giving me energy to continue my work.

Finally, I am most grateful to my family, especially my grandmother, for their support, help and for educating me.

CONTENTS

1 INTRODUCTION

The modern society produces enormous amounts of data on a daily basis. Since we have access to large amounts of data containing relatively small amounts of useful information, we suffer from data overload both in our daily lives, and in diverse scientific disciplines such as neuroscience, neuroinformatics, bioinformatics, telecommunications and economics. Thus, it is very important to extract underlying informational components, or simply sources, from such data.

Independent component analysis (ICA) is a well-known method for extracting the sources from given data [18, 19, 25, 68]; it assumes that the sources are statistically independent of each other. This is not an unrealistic assumption in many cases, and ICA has been used with success in many applications such as data mining, brain imaging (e.g., functional magnetic resonance imaging: fMRI), and speech enhancement [9, 14, 50, 54, 64].

Imagine, for example, that multiple people speak at the same time in a room and there are a number of microphones arranged in different locations. Each of microphones records a mixture of their speech signals, and we want to estimate the speech signals only using a set of measured signals. This example is known as the cocktail party problem. In this case, it is these source signals that are of primary interest, but they are buried within a large set of measured signals. ICA can be used to solve this problem, because it comes from the fact that these speech signals are generated by unrelated physical processes (i.e., by different persons).

ICA belongs to a class of blind source separation (BSS) methods for separating given data into latent sources without any prior knowledge about the mixing process, where the data can take the form of sounds, images, and telecommunication records. The term "blind" implies that no explicit knowledge of sources and mixing process is available. Depending on the nature of the sources of interest (e.g., independency, non-negativity, and sparseness), the BSS problem is solved by several approaches such as ICA, non-negative matrix factorization (NMF), and sparse component analysis (SCA) [16, 17, 40, 55].

In the general case of the BSS problem, the number of sources and that of measured signals are unknown. Let N and M be the number of the sources and of the measured signals, respectively. In view of the relation between N and M, the BSS problem can be classified into three cases:

1. The case that the number of sources is equivalent to that of measured signals is termed *determined BSS* ($M = N$). Most

commonly used BSS algorithms belong to this case. This is the case we will focus on.

2. The case that the measured signals outnumber the sources is called *overdetermined BSS* ($M > N$). In this case, overdetermined BSS algorithms usually comprise two stages; the first stage is to reduce the sensor dimensionality using an appropriate preprocessing, e.g. principal component analysis (PCA), to obtain the case $M = N$, and the second one consists of ICA for extracting independent components [31, 71].

3. The more difficult case is known as *underdetermined* (i.e., with more sources than sensors) *BSS*. The signals are irreversibly mixed and it is thus impossible to exactly invert the transformation. Mostly the sparseness of the sources in the time-frequency domain is used to determine clusters which correspond to the separated sources [73].

In the simplest form of the determined BSS, a mixing process is assumed to be instantaneous; time delay is not taken into account. Many algorithms have been proposed based on this model. In practical applications such as speech separation, however, the mixing process must be considered as a convolutive one, for which the demixing process must take a form of multiple-input multiple-output (MIMO) filter. In this thesis, we deal with BSS of convolutive one.

For BSS of convoluted mixtures, two approaches are known: the time-domain approach and the frequency-domain one [1, 13, 15, 20, 21, 33, 58, 61, 67, 69]. The latter is usually preferred, because of its computational efficiency and robustness for separation of signals mixed in a high reverberant environment. In the approach, each of the N observed signals is transformed into a set of frequency-domain components; let the number of them be K. Then, a convolutive mixing process can be treated as K instantaneous mixing processes, each of which involves N frequency-domain signals. Thus, a convolutive BSS problem can be simplified down to K instantaneous BSS problems.

This simplification comes at the cost of a cumbersome issue called the permutation problem [28]. The order of the source signals recovered by an BSS algorithm is essentially indeterminate. The permutations in the K demixing matrices become thus different from each other in general. In order to finally obtain the original sources, some additional processing is therefore required to align those permutations:

1. *Smoothness constraint*. A popular method for this end is to impose a smoothness constraint on the separators of adjacent frequencies [7, 61, 67].

2. *Direction-of-arrival (DoA) estimation*. The permutation is performed on the separator at each frequency to achieve phase consistency

across frequency as inferred from the sound propagation model for individual point sources [37, 53, 66].

3. *Correlations of spectral envelops*. When the source signals are nonstationary ones such as speech, envelopes of the frequency components originated from the same source have strong correlations. Utilizing this fact, the permutation is aligned [6, 52].

However, all methods including these ideas require some complicated and time-consuming procedures.

On the other hand, some algorithms evaluate statistical independence among the sources in the time domain [1, 13, 15, 21, 33, 47]. The time-domain approach is often thought to be free from the permutation problem, but it is not necessarily the case, as reported in [47]. Moreover, the approach requires computational time and sometimes shows numerical instability.

In order to overcome the problems that occur in both of two approaches, several algorithms were proposed:

1. *Frequency-domain implementation of a time-domain algorithm*. The computational cost of the time-domain approach mainly comes from convolution operation with long demixing filter. In order to reduce the computational cost, a frequency-domain implementation of a time-domain algorithm was proposed; it evaluates independence among the signals in *the time domain*, but updates the demixing filter in *the frequency domain* [8, 32, 38, 39, 44, 57]. Because of transformation of demixing filter into the frequency domain, the approach can relax the computational cost of counterpart of the time-domain algorithm. However, the algorithm sometimes shows numerical instability as in the case of a time-domain one.

2. *Minimization of an integrated frequency-domain function in the time-domain*. Some approaches define demixing process in the time domain, but optimize the parameters based on a frequency-domain objective function [34, 51]. Since these approaches minimize the cost function in the time domain, they can avoid the permutation problem.

3. *Independent vector analysis*. Recently, a new extension of ICA to the permutation problem was proposed, which is called *independent vector analysis* (IVA). IVA treats the K components originated from the same source as a vector whose components are mutually *dependent*. Due to the inner dependency, the approach can solve the permutation problem. Although the effectiveness of IVA algorithms is reported in many papers [4, 5, 23, 24, 30, 35, 36, 41, 42, 43, 56, 59, 70], there are very few mathematical analyses of IVA algorithms [29].

The objective of this thesis is to propose an effective algorithm for the convolutive BSS problem and to prove the effectiveness of IVA mathematically.

This thesis is organized as follows.

Chapter 2 introduces mathematical definitions that will appear in subsequent chapters. And also, we overview the basic principle behind ICA methods.

Chapter 3 formulates a convolutive BSS problem, and describes ambiguities that is inherent in BSS. Then, we propose a frequency-domain implementation of a time-domain algorithm. This chapter also introduces two additional techniques for solving the permutation problem and for guaranteeing the boundedness of the separated signals.

The basic mixing and demixing models dealt with in IVA are described in Chapter 4. Then, we derive a necessary and sufficient condition for the desired separator to be obtained by minimizing a certain measure representing the difference between a prior source probability density function (pdf) and an actual output pdf of the separator. Also, we prove that, if a desired separator is a stable equilibrium of the measure, any permuted separator never becomes a stable equilibrium.

Chapter 5 revisits an IVA algorithm, and derives some IVA algorithms incorporating different constraints.

In Chapter 6, we conducted some simulations with synthetic data and speech signals to confirm the effectiveness of the algorithms proposed in this thesis.

Finally, Chapter 7 concludes this thesis.

2 PRELIMINARIES

2.1 INTRODUCTION

This chapter introduces mathematical definitions and overviews the basic concept of independent component analysis (ICA).

This chapter is organized as follows. Firstly, mathematical notations are given in Section 2.2. Complex derivatives will be appeared in subsequent chapters. Secondly, we review the fundamental principle of ICA in Section 2.3.

2.2 MATHEMATICAL PRELIMINARIES

Throughout this thesis, $\mathbb{R}$ stands for a set of real numbers, and also $\mathbb{C}$ denotes a set of complex numbers. In general we shall denote scalars by normal letters, e.g. a, vectors by bold-faced, lower-case letters, e.g. $\mathbf{a} \in \mathbb{R}^N$, matrices by bold-faced, upper-case letters, e.g. $\mathbf{A} \in \mathbb{R}^{N \times M}$. As is usual, all vectors are dealt with as column vectors for purpose of matrix multiplication. If $\mathbf{a}$ is an N-dimensional vector, then $\|\mathbf{a}\|_2$ stands for the l_2-norm of $\mathbf{a}$. For the sake of simplicity of notation, we use $\|\cdot\|$ as notation for the l_2-norm. If $\mathbf{A}$ is an $N \times N$ square matrix, then $\det \mathbf{A}$ stands for determinant of $\mathbf{A}$, and $\mathrm{tr}\mathbf{A}$ stands for its trace (i.e., the sum of the diagonal entries of $\mathbf{A}$).

Let $\mathbf{A}$ and $\mathbf{B}$ be $N \times M$ matrices. The Hadamard product of two equal-size matrices is the element-wise multiplication denoted by $\circledast$ (or $.*$ for MATLAB notation) and defined as

$$\mathbf{A} \circledast \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1M}b_{1M} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2M}b_{2M} \\ \vdots & \vdots & & \vdots \\ a_{N1}b_{N1} & a_{N2}b_{N2} & \cdots & a_{NM}b_{NM} \end{bmatrix}.$$

A block matrix $\mathbf{A}$ partitioned into NM blocks is denoted by

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1M} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2M} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{N1} & \mathbf{A}_{N2} & \cdots & \mathbf{A}_{NM} \end{bmatrix} = \left[\mathbf{A}_{ij}\right]. \tag{2.1}$$

The i-th block of vector $\mathbf{a}$ is denoted by $\mathbf{a}_i$, and the k-th element of $\mathbf{a}_i$ is written as $a_i^{(k)}$.

2.2.1 *Real derivatives*

Let $f : \mathbb{R}^N \to \mathbb{R}$ be a scalar-valued function of $\mathbf{x} = [x_1, \ldots, x_N]^\top \in \mathbb{R}^N$, and the i-th partial derivative of f at the point $\mathbf{x}$ is denoted by $(\partial f / \partial x_i)(\mathbf{x})$ or $\partial f(\mathbf{x}) / \partial x_i$. Assuming all of these partial derivatives exist, the gradient of f at $\mathbf{x}$ is defined as the column vector

$$\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) \triangleq \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_N}(\mathbf{x}) \end{bmatrix}. \tag{2.2}$$

Similarly, if $f : \mathbb{R}^{N\times M} \to \mathbb{R}$ is a scalar-valued function of $n \times m$ matrix $\mathbf{X} = [x_{ij}]$, then the gradient matrix of f is given by

$$\frac{\partial f}{\partial \mathbf{X}}(\mathbf{X}) \triangleq \begin{bmatrix} \frac{\partial f}{\partial x_{11}}(\mathbf{X}) & \cdots & \frac{\partial f}{\partial x_{1M}}(\mathbf{X}) \\ \vdots & & \vdots \\ \frac{\partial f}{\partial x_{N1}}(\mathbf{X}) & \cdots & \frac{\partial f}{\partial x_{NM}}(\mathbf{X}) \end{bmatrix}. \tag{2.3}$$

If $f : \mathbb{R}^N \to \mathbb{R}^M$ is a vector-valued function, it is called differentiable if each component f_i of f is a continuously differentiable. The gradient matrix of f is the $N \times M$ matrix whose i-th column is the gradient $\partial f_i(\mathbf{x}) / \partial \mathbf{x}$ of f_i. Hence,

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) &\triangleq \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}}(\mathbf{x}) & \cdots & \frac{\partial f_M}{\partial \mathbf{x}}(\mathbf{x}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_M}{\partial x_1}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_N}(\mathbf{x}) & \cdots & \frac{\partial f_M}{\partial x_N}(\mathbf{x}) \end{bmatrix} \end{aligned} \tag{2.4}$$

The transpose of $\partial f(\mathbf{x}) / \partial \mathbf{x}$ is known as the *Jacobian* of f and is a matrix whose (i, j)-th entry is equivalent to the partial derivative $\partial f_i / \partial x_j$.

Now suppose that each one of the partial derivatives of a function $f : \mathbb{R}^N \to \mathbb{R}$ is a continuously differentiable function of $\mathbf{x} \in \mathbb{R}^N$. The i-th partial derivative of $\partial f(\mathbf{x}) / \partial x_j$ at a point $\mathbf{x}$ is denoted by$\partial^2 f(\mathbf{x}) / \partial x_i \partial x_j$; it can be written in matrix form

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^\top}(\mathbf{x}) \triangleq \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_N \partial x_1}(\mathbf{x}) & \cdots & \cdots & \frac{\partial^2 f}{\partial x_N^2}(\mathbf{x}) \end{bmatrix}, \tag{2.5}$$

which is known as the Hessian of f.

2.2.2 *Complex derivatives*

The optimization of parameters is a common problem in many disciplines such as signal processing and neural networks. The problem is often formulated as a minimization (or maximization) of a real-valued function of complex parameters. For solving such optimization problems, derivation with respect to a complex parameter is required. However, such function is not complex-differentiable and hence we cannot use the standard tool of complex analysis.

A common approach to optimization with respect to complex variables is to replace these with their real and imaginary parts and then to minimize (or maximize) with respect to the *real* variables. Hence, the gradient of the function is taken with respect to the real and imaginary parts. In general, the approach increases the dimensionality of the problem and usually leads to a complicated form.

In order to obtain a compact representation, Brandwood defined a partial complex gradient operator which solves this kind of optimization problem [11], and van Bos described the simple relation between real and complex gradient operators and defined the complex Hessian operator [10]. The complex differential calculus is called Wirtinger Calculus, and has been used in several papers [42, 45, 60].

To start with, we consider a simple case. Let $f : \mathbb{C} \to \mathbb{R}$ be a function of complex variable x; it comprises the complex conjugation of x to be real. Hence, f can be rewritten as $f(x, x^*)$, i.e., $f : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$. Assume that f is holomorphic with respect to each variable (x and x^*) and $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a function of two real variables ($\overline{x}$ and $\underline{x}$) such that $f(x, x^*) = g(\overline{x}, \underline{x})$. Then, the partial derivatives of f are defined by

$$\begin{cases} \frac{\partial f}{\partial x}(x, x^*) = \frac{1}{2}\left(\frac{\partial g}{\partial \overline{x}}(\overline{x}, \underline{x}) - j\frac{\partial g}{\partial \underline{x}}(\overline{x}, \underline{x})\right) \\ \frac{\partial f}{\partial x^*}(x, x^*) = \frac{1}{2}\left(\frac{\partial g}{\partial \overline{x}}(\overline{x}, \underline{x}) + j\frac{\partial g}{\partial \underline{x}}(\overline{x}, \underline{x})\right) \end{cases} \tag{2.6}$$

or in a vector-matrix form

$$\begin{bmatrix} \frac{\partial f}{\partial x}(x, x^*) \\ \frac{\partial f}{\partial x^*}(x, x^*) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix}\begin{bmatrix} \frac{\partial g}{\partial \overline{x}}(\overline{x}, \underline{x}) \\ \frac{\partial g}{\partial \underline{x}}(\overline{x}, \underline{x}) \end{bmatrix} \tag{2.7}$$

where $(\partial f/\partial x)(x, x^*)$ (or $(\partial f/\partial x^*)(x, x^*)$) is evaluated while treating x^* (or x) as a constant in f.

Equation (2.6) can be derived as follows. Complex variables x and x^* can be represented by their real and imaginary parts:

$$\begin{cases} x = \overline{x} + j\underline{x} \\ x^* = \overline{x} - j\underline{x} \end{cases} \tag{2.8}$$

where $\overline{x}, \underline{x} \in \mathbb{R}$ and $j^2 = -1$. In a vector-matrix form, Eq. (2.8) can be taken a form of

$$\begin{bmatrix} x \\ x^* \end{bmatrix} = \begin{bmatrix} 1 & j \\ 1 & -j \end{bmatrix} \begin{bmatrix} \overline{x} \\ \underline{x} \end{bmatrix} \tag{2.9}$$

If f is analytic with respect to each variable, then we could define the partial derivatives of f by

$$\frac{\partial f}{\partial x}(x, x^*) \text{ and } \frac{\partial f}{\partial x^*}(x, x^*)$$

using the standard complex differentiation. Differentiating g with respect to each variable and using the chain rule, we obtain

$$\begin{cases} \frac{\partial g}{\partial \overline{x}}(\overline{x}, \underline{x}) = \frac{\partial f}{\partial x}(x, x^*)\frac{\partial x}{\partial \overline{x}} + \frac{\partial f}{\partial x^*}(x, x^*)\frac{\partial x^*}{\partial \overline{x}} \\ \frac{\partial g}{\partial \underline{x}}(\overline{x}, \underline{x}) = \frac{\partial f}{\partial x}(x, x^*)\frac{\partial x}{\partial \underline{x}} + \frac{\partial f}{\partial x^*}(x, x^*)\frac{\partial x^*}{\partial \underline{x}} \end{cases} \tag{2.10}$$

From Eq. (2.8), we have

$$\frac{\partial x}{\partial \overline{x}} = \frac{\partial x^*}{\partial \overline{x}} = 1$$

and

$$\frac{\partial x}{\partial \underline{x}} = -\frac{\partial x^*}{\partial \underline{x}} = j.$$

Substituting them into Eq. (2.10) yields

$$\begin{cases} \frac{\partial g}{\partial \overline{x}}(\overline{x}, \underline{x}) = \frac{\partial f}{\partial x}(x, x^*) + \frac{\partial f}{\partial x^*}(x, x^*) \\ \frac{\partial g}{\partial \underline{x}}(\overline{x}, \underline{x}) = j\frac{\partial f}{\partial x}(x, x^*) - j\frac{\partial f}{\partial x^*}(x, x^*) \end{cases} \tag{2.11}$$

Solving Eq. (2.11) with respect to $(\partial f/\partial x)(x, x^*)$ and $(\partial f/\partial x^*)(x, x^*)$ leads to Eq. (2.6). A necessary and sufficient condition for f to have a stationary point is that $(\partial f/\partial x)(x, x^*) = 0$, where the partial derivative with respect to x deals with x^* as a constant in f. The partial derivatives satisfy the following relation

$$\frac{\partial f}{\partial x^*}(x, x^*) = \left(\frac{\partial f}{\partial x}(x, x^*)\right)^*.$$

Next, let $f : \mathbb{C}^2 \to \mathbb{R}$ be a function of two variables x_1 and x_2. It comprises the complex conjugations of x_1 and x_2 as in the case described above. Hence, it can be written as $f(x_1, x_2, x_1^*, x_2^*)$, i.e., $f : \mathbb{C}^2 \times \mathbb{C}^2 \to \mathbb{R}$. For the sake of simplicity of notation, a set of complex variables x_1 and x_2 is denoted by vector $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$, and a set of complex conjugations is denoted by $\mathbf{x}^* = \begin{bmatrix} x_1^* & x_2^* \end{bmatrix}^\top$ Assume that f is analytic with respect to each variable (x_1, x_2, x_1^* and x_2^*) and $g : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is a function of four real variables ($\overline{x}_1, \overline{x}_2, \underline{x}_1$ and $\underline{x}_2$) such that $f(\mathbf{x}, \mathbf{x}^*) = g(\overline{\mathbf{x}}, \underline{\mathbf{x}})$. Note that $\overline{\mathbf{x}}$ and $\underline{\mathbf{x}}$ correspond to a set of real parts of x_1 and x_2 and that of their imaginary parts, respectively. Then, the partial derivatives of f are defined by

$$\begin{cases} \frac{\partial f}{\partial x_1}(\mathbf{x},\mathbf{x}^*) = \frac{1}{2}\left(\frac{\partial g}{\partial \overline{x}_1}(\overline{\mathbf{x}},\underline{\mathbf{x}}) - j\frac{\partial g}{\partial \underline{x}_1}(\overline{\mathbf{x}},\underline{\mathbf{x}})\right) \\ \frac{\partial f}{\partial x_2}(\mathbf{x},\mathbf{x}^*) = \frac{1}{2}\left(\frac{\partial g}{\partial \overline{x}_2}(\overline{\mathbf{x}},\underline{\mathbf{x}}) - j\frac{\partial g}{\partial \underline{x}_2}(\overline{\mathbf{x}},\underline{\mathbf{x}})\right) \\ \frac{\partial f}{\partial x_1^*}(\mathbf{x},\mathbf{x}^*) = \frac{1}{2}\left(\frac{\partial g}{\partial \overline{x}_1}(\overline{\mathbf{x}},\underline{\mathbf{x}}) + j\frac{\partial g}{\partial \underline{x}_1}(\overline{\mathbf{x}},\underline{\mathbf{x}})\right) \\ \frac{\partial f}{\partial x_2^*}(\mathbf{x},\mathbf{x}^*) = \frac{1}{2}\left(\frac{\partial g}{\partial \overline{x}_2}(\overline{\mathbf{x}},\underline{\mathbf{x}}) + j\frac{\partial g}{\partial \underline{x}_2}(\overline{\mathbf{x}},\underline{\mathbf{x}})\right) \end{cases}$$

It can be expressed as a simple vector-matrix form such that

$$\begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x},\mathbf{x}^*) \\ \frac{\partial f}{\partial x_2}(\mathbf{x},\mathbf{x}^*) \\ \frac{\partial f}{\partial x_1^*}(\mathbf{x},\mathbf{x}^*) \\ \frac{\partial f}{\partial x_2^*}(\mathbf{x},\mathbf{x}^*) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 0 & -j & 0 \\ 0 & 1 & 0 & -j \\ 1 & 0 & j & 0 \\ 0 & 1 & 0 & j \end{bmatrix}\begin{bmatrix} \frac{\partial g}{\partial \overline{x}_1}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \\ \frac{\partial g}{\partial \overline{x}_2}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \\ \frac{\partial g}{\partial \underline{x}_1}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \\ \frac{\partial g}{\partial \underline{x}_2}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \end{bmatrix}$$

or

$$\begin{bmatrix} \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x},\mathbf{x}^*) \\ \frac{\partial f}{\partial \mathbf{x}^*}(\mathbf{x},\mathbf{x}^*) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \mathbf{I}_2 & -j\mathbf{I}_2 \\ \mathbf{I}_2 & j\mathbf{I}_2 \end{bmatrix}\begin{bmatrix} \frac{\partial g}{\partial \overline{\mathbf{x}}}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \\ \frac{\partial g}{\partial \underline{\mathbf{x}}}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \end{bmatrix},$$

where $\mathbf{I}_2$ denotes the 2×2 identity matrix.

Similarly, it can be extended to a real-valued function of an N-dimensional complex vector. Let $f : \mathbb{C}^N \to \mathbb{R}$ be a function of $\mathbf{x} = [x_1, \ldots, x_N]^\top \in \mathbb{C}^N$ and it is equivalent to $f(\mathbf{x},\mathbf{x}^*)$, where $\mathbf{x}^* = [x_1^*, \ldots, x_N^*]^\top$. As in the case of (2.8), the vectors ($\mathbf{x}$ and $\mathbf{x}^*$) can be represented by their real and imaginary parts:

$$\begin{cases} \mathbf{x} = \overline{\mathbf{x}} + j\underline{\mathbf{x}} \\ \mathbf{x}^* = \overline{\mathbf{x}} - j\underline{\mathbf{x}} \end{cases} \tag{2.12}$$

or in a vector-matrix form

$$\mathbf{x}_\mathbb{C} \triangleq \begin{bmatrix} \mathbf{x} \\ \mathbf{x}^* \end{bmatrix} = \begin{bmatrix} \mathbf{I}_N & j\mathbf{I}_N \\ \mathbf{I}_N & -j\mathbf{I}_N \end{bmatrix}\begin{bmatrix} \overline{\mathbf{x}} \\ \underline{\mathbf{x}} \end{bmatrix} = \mathbf{J}\mathbf{x}_\mathbb{R}, \tag{2.13}$$

where $\overline{\mathbf{x}} = [\overline{x}_1, \ldots, \overline{x}_N]^\top \in \mathbb{R}^N$ and $\underline{\mathbf{x}} = [\underline{x}_1, \ldots, \underline{x}_N]^\top \in \mathbb{R}^N$ and $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. For the sake of simplicity of notation, we omit the subscript N of the identity matrix hereafter. If $g : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is a function of two N-dimentional real vector ($\overline{\mathbf{x}}$ and $\underline{\mathbf{x}}$) such that $f(\mathbf{x},\mathbf{x}^*) = g(\overline{\mathbf{x}},\underline{\mathbf{x}})$, then the complex gradient of f at $(\mathbf{x},\mathbf{x}^*)$ is given by

$$\frac{\partial f}{\partial \mathbf{x}_\mathbb{C}}(\mathbf{x},\mathbf{x}^*) = \frac{1}{2}\mathbf{J}^*\frac{\partial g}{\partial \mathbf{x}_R}(\overline{\mathbf{x}},\underline{\mathbf{x}}) = \mathbf{J}^{-\top}\frac{\partial g}{\partial \mathbf{x}_R}(\overline{\mathbf{x}},\underline{\mathbf{x}}) \tag{2.14}$$

or

$$\frac{\partial f}{\partial \mathbf{x}_\mathbb{C}^*}(\mathbf{x},\mathbf{x}^*) = \frac{1}{2}\mathbf{J}\frac{\partial g}{\partial \mathbf{x}_R}(\overline{\mathbf{x}},\underline{\mathbf{x}}). \tag{2.15}$$

It should be noted that the inverse of $\mathbf{J}$ takes a form of

$$\mathbf{J}^{-1} = \frac{1}{2}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -j\mathbf{I} & j\mathbf{I} \end{bmatrix} = \frac{1}{2}\mathbf{J}^{H}, \tag{2.16}$$

where the superscript H stands for the Hermitian transpose operator.

In optimization problem, one matter for concern is that which of the complex gradients (i.e., $\partial f / \partial \mathbf{x}$ or $\partial f / \partial \mathbf{x}^*$) provides the steepest descent direction for a function. If $f : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ is complex-differentiable with respect to each variable, then the first-order Taylor expansion of f gives

$$\begin{aligned} \delta f &= \sum_{n=1}^{N}\left(\frac{\partial f}{\partial x_n}\delta x_n + \frac{\partial f}{\partial x_n^*}\delta x_n^*\right) = \frac{\partial f}{\partial \mathbf{x}^\top}\delta \mathbf{x} + \frac{\partial f}{\partial \mathbf{x}^H}\delta \mathbf{x}^* \\ &= 2\mathrm{Re}\left\{\frac{\partial f}{\partial \mathbf{x}^\top}\delta \mathbf{x}\right\}, \end{aligned} \tag{2.17}$$

where $\mathrm{Re}\{x\}$ denotes the real part of x. The right-hand side of the above equation can be regarded as the inner product of two vectors, i.e.,

$$\frac{\partial f}{\partial \mathbf{x}^\top}\delta \mathbf{x} = \left\langle \frac{\partial f}{\partial \mathbf{x}^*}, \delta \mathbf{x} \right\rangle,$$

where $\langle \mathbf{a}, \mathbf{b} \rangle$ stands for the inner product of $\mathbf{a}$ and $\mathbf{b}$. Hence, we apply the Cauchy-Schwarz inequality to the inner product

$$\left| \left\langle \frac{\partial f}{\partial \mathbf{x}^*}, \delta \mathbf{x} \right\rangle \right| \leq \left\| \frac{\partial f}{\partial \mathbf{x}^*} \right\| \|\delta \mathbf{x}\|.$$

Equality holds only when $\delta \mathbf{x} = \eta\,(\partial f / \partial \mathbf{x}^*)$, where η is a scalar, which may be complex. The maximum value of $\mathrm{Re}\{\langle \partial f / \partial \mathbf{x}^*, \delta \mathbf{x} \rangle\}$ occurs only when η is real and positive. In other words, it takes a maximum value only when $\delta \mathbf{x}$ and $\partial f / \partial \mathbf{x}^*$ are in the same direction in the vector space $\mathbb{C}^N$.

As is analogous to the Hessian of f defined in the real domain, the complex Hessian of f can be defined. Let $f : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ be a function of $\mathbf{x}$ and $\mathbf{x}^*$. Assume that f is complex-differentiable with respect to each variable and each one of the partial derivatives of f is also complex-differentiable with respect to each variable. Let $g : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$ is a function of $\overline{\mathbf{x}}$ and $\underline{\mathbf{x}}$ such that $f(\mathbf{x}, \mathbf{x}^*) = g(\overline{\mathbf{x}}, \underline{\mathbf{x}})$, and then g is also continuously differentiable. Using the relation (2.14), the Hessian of f is given by

$$\begin{aligned} \frac{\partial^2 f}{\partial \mathbf{x}_{\mathbb{C}}^* \partial \mathbf{x}_{\mathbb{C}}^\top}(\mathbf{x}, \mathbf{x}^*) &= \frac{1}{4}\mathbf{J}\frac{\partial^2 g}{\partial \mathbf{x}_{\mathbb{R}} \partial \mathbf{x}_{\mathbb{R}}^\top}(\overline{\mathbf{x}}, \underline{\mathbf{x}})\,\mathbf{J}^H \\ &= \frac{1}{2}\mathbf{J}\frac{\partial^2 g}{\partial \mathbf{x}_{\mathbb{R}} \partial \mathbf{x}_{\mathbb{R}}^\top}(\overline{\mathbf{x}}, \underline{\mathbf{x}})\,\mathbf{J}^{-1}. \end{aligned} \tag{2.18}$$

This matrix is Hermitian, because it satisfies

$$
\begin{aligned}
\left\{ \frac{\partial^2 f}{\partial \mathbf{x}_{\mathbb{C}}^* \partial \mathbf{x}_{\mathbb{C}}^\top} \left(\mathbf{x}, \mathbf{x}^* \right) \right\}^H &= \frac{1}{4} \left(\mathbf{J} \frac{\partial^2 g}{\partial \mathbf{x}_{\mathbb{R}} \partial \mathbf{x}_{\mathbb{R}}^\top} \left(\overline{\mathbf{x}}, \underline{\mathbf{x}} \right) \mathbf{J}^H \right)^H \\
&= \frac{1}{4} \mathbf{J} \frac{\partial^2 g}{\partial \mathbf{x}_{\mathbb{R}} \partial \mathbf{x}_{\mathbb{R}}^\top} \left(\overline{\mathbf{x}}, \underline{\mathbf{x}} \right) \mathbf{J}^H = \frac{\partial^2 f}{\partial \mathbf{x}_{\mathbb{C}}^* \partial \mathbf{x}_{\mathbb{C}}^\top} \left(\mathbf{x}, \mathbf{x}^* \right).
\end{aligned} \tag{2.19}
$$

2.2.3 *Uncorrelatedness and statistical independence*

Two random variables X and Y are *uncorrelated* if their covariance is zero:

$$
E \left(\left(X - \mu_X \right) \left(Y - \mu_Y \right) \right) = 0, \tag{2.20}
$$

where μ_X denotes the mean of X and $E(\cdot)$ stands for expectation calculation.

Let $\mathbf{x}$ and $\mathbf{y}$ be random *vectors*. These vectors are uncorrelated if their cross-covariance matrix is a zero matrix

$$
E \left(\left(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}} \right) \left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}} \right)^\top \right) = \mathbf{O}, \tag{2.21}
$$

where $\boldsymbol{\mu}_{\mathbf{x}}$ is the mean vector of $\mathbf{x}$. This is equivalent to

$$
E \left(\mathbf{x} \mathbf{y}^\top \right) = \boldsymbol{\mu}_{\mathbf{x}} \boldsymbol{\mu}_{\mathbf{y}}^\top. \tag{2.22}
$$

Hence, in the case of zero-mean random vectors, zero covariance is equivalent to zero correlation.

Statistical independence is defined in terms of probability density functions (pdfs). The random variables X and Y are *statistically independent* if and only if the joint density $p_{X,Y}(x, y)$ of X and Y must factorize into the product of their marginal densities $p_X(x)$ and $p_Y(y)$, i.e.,

$$
p_{X,Y}(x, y) = p_X(x) \, p_Y(y). \tag{2.23}
$$

Statistical independence is a key concept of ICA.

Independent random variables satisfy the following property:

$$
E \left(g(X) \, h(Y) \right) = E \left(g(X) \right) E \left(h(Y) \right), \tag{2.24}
$$

where g and h are any integrable functions. In words, any nonlinear correlation becomes the product of $E(g(X))$ and $E(h(Y))$ if and only if the random variables are statistically independent of each other. This is because

$$
\begin{aligned}
E \left(g(X) \, h(Y) \right) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u_1) \, h(u_2) \, p_{X,Y}(u_1, u_2) \, du_1 du_2 \\
&= \int_{-\infty}^{\infty} g(u_1) \, p_X(u_1) \, du_1 \int_{-\infty}^{\infty} h(u_2) \, p_Y(u_2) \, du_2 \\
&= E \left(g(X) \right) E \left(h(Y) \right).
\end{aligned}
$$

Equation (2.24) suggests that statistical independence is a stronger property than uncorrelatedness. Equation (2.22) is obtained from the independence property (2.24) as a special case where both g and h are linear functions, and takes into account second-order statistics (correlations or covariances) only.

2.3 OVERVIEW OF INDEPENDENT COMPONENT ANALYSIS

Recently, there has been an increasing interest in statistical models for learning data representations. A well-known approach for this is independent component analysis (ICA), the concept of which was initially proposed by [18]. So far, many ICA algorithms have been proposed and been used in many applications [14, 25]. So far, many ICA algorithms have been proposed and have been used in various applications [14, 25]. In order to understand

For the sake of clarification, let's look at the case where two sources generate signals and two weighted mixtures of sources are observed by two sensors. This can be represented by

$$\begin{cases} x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \\ x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \end{cases}$$

or in a vector-matrix form

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \tag{2.25}$$

where the matrix $\mathbf{A}$, which is referred to as the mixing matrix, is a 2×2 coefficient matrix. Figure 1 shows the effect of the mixing matrix in the case that it takes the following form:

$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}.$$

The estimation of sources with knowledge of observed signals is known as the signal separation problem. We usually assume that the mixing matrix is invertible (i.e., full rank matrix). There thus exists a 2×2 matrix $\mathbf{W}$ with coefficients w_{ij} such that we can separate them as:

$$\begin{cases} y_1(t) = w_{11}x_1(t) + w_{12}x_2(t) \\ y_2(t) = w_{21}x_1(t) + w_{22}x_2(t) \end{cases} \tag{2.26}$$

or in a vector-matrix form

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t), \tag{2.27}$$

where $\mathbf{y}(t)$ is estimated source vector, and matrix $\mathbf{W}$ is called the demixing matrix or the separator. If the mixing matrix $\mathbf{A}$ is given, then the separator obviously takes the form of the inverse of matrix

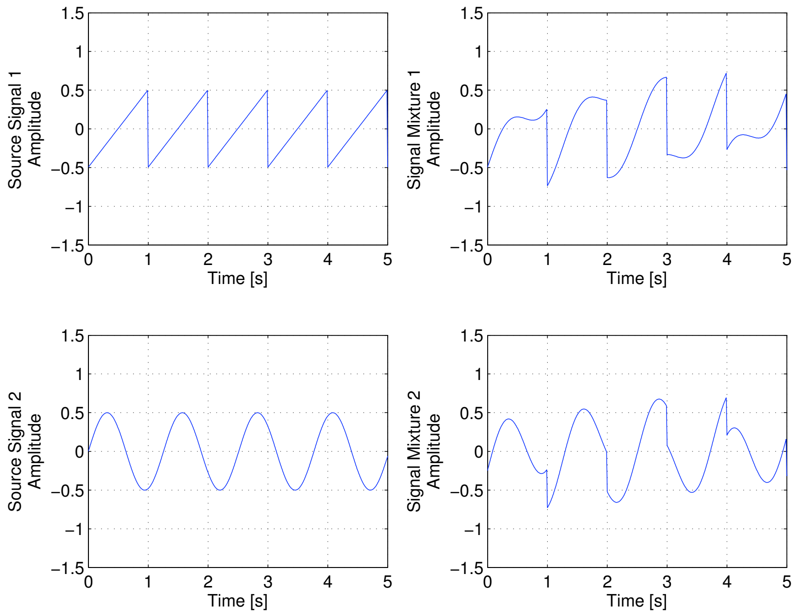

Figure 1: The original source signals and the observed signals. *Left*: Each of two sources generates a statistically independent *source signal*. In this case, source signals are saw-tooth wave and sinusoidal wave. *Right*: If there are two sensors, then the output of each sensor is a mixture of source signals or signal mixture. The unknown source-sensor distances are represented by the mixing matrix labeled **A**. In this case, the mixing matrix **A** takes the form of symmetric matrix whose diagonal elements are one and non-diagonal elements are 0.5.

A: i.e., $\mathbf{W} = \mathbf{A}^{-1}$. When the matrix **A** and source signals $\mathbf{s}(t)$ are unknown, this signal separation problem is called the blind signal separation (BSS) problem. Our task is to determine the separator from the observed signals so that the separator's outputs are equivalent to the source signals.

ICA determines the separator so that the estimated source signals $y_1(t)$ and $y_2(t)$ are statistically independent. As example, the comparison of waveforms between original signals and signals estimated by ICA is shown in Fig. 2. As can be seen, these are very close to the original source signals. Hence, ICA is a promising approach to the BSS problem. Here we consider the case where the number of sources and that of sensors are two, but it can be generalized to the case where the number of sources is N and that of sensors is $M \, (\neq N)$, of course.

As can be seen from Fig. 2, ICA has certain indeterminacy in respect to the word "source". One is called scaling indeterminacy and the other is called permutation indeterminacy. The former one means that the amplitude of recovered signals cannot be determined. Since the mixing matrix and source signals are unknown, the effect of

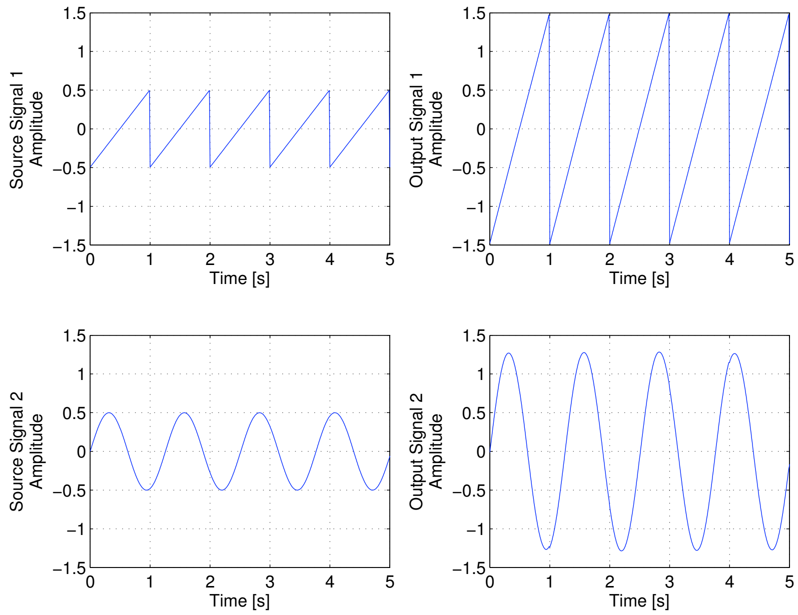

Figure 2: The original source signals and the recovered signals using ICA. *Left*: two source signals shown in 1. *Right*: the recovered signals using an ICA algorithm. ICA extracts two independent signals from the set of signal mixtures. Note that ICA rearranges signals so that an extracted signal $y_i(t)$ and its source signal $s_i(t)$ are not necessarily on the same row.

multiplication of one of the source estimates with a scalar constant k is canceled by dividing its corresponding column in the mixing matrix by k. Namely, Eq. (2.25) can be rewritten as

$$\begin{aligned}\mathbf{x}(t) &= \left(\frac{1}{k_1}\mathbf{a}_1\right)(k_1 s_1(t)) + \left(\frac{1}{k_2}\mathbf{a}_2\right)(k_2 s_2(t)) \\ &= \mathbf{A}\mathbf{D}^{-1}\mathbf{D}\mathbf{s}(t),\end{aligned} \tag{2.28}$$

where $\mathbf{a}_i$ stands for the i-th column vector in the mixing matrix $\mathbf{A}$ and $\mathbf{D}$ is a diagonal matrix whose elements are k_1 and k_2. Define $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{D}^{-1}$ and $\tilde{\mathbf{s}}(t) = \mathbf{D}\mathbf{s}(t)$, then Eq. (2.28) becomes

$$\mathbf{x}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(t). \tag{2.29}$$

Since we do not know the mixing matrix and source signals, we can not distinguish the above equation and Eq. (2.25). A common strategy to this is to assume that each of the source signals is unit variance, i.e., $E\left(s_i^2(t)\right) = 1$.

With permutation indeterminacy it is meant that the order of the independent components cannot be determined. This is because it is free to decide what independent components that will be the first one. The details about these ambiguities can be found in [25].

3 CONVOLUTIVE BSS ALGORITHM BASED ON NONLINEAR CORRELATION

3.1 INTRODUCTION

In simplest blind source separation (BSS) problem, the mixing process is assumed to be instantaneous; time delay is not taken into account. In practical applications such as speech separation, however, the mixing process must be considered as convolutive. A prominent example is the so-called cocktail party problem, where we want to recover the speech signals of multiple speakers who are simultaneously talking in a room. The room will generally be reverberant due to reflections on the walls, i.e., the source signals are filtered by a linear multiple-input multiple-output (MIMO) system before they are picked up by the sensors. This chapter deals with the convolutive BSS problem.

Most of existing convolutive BSS methods can be classified into two groups in terms of the processing domain, i. e., the time-domain approach and the frequency-domain one. The former one incorporates spatiotemporal structure into the demixing processs. Hence, the time-domain approach becomes more complicated and requires high computational cost as reverberation increases. On the other hand, the latter one transforms the observations into the frequency domain and deals with a convolutive BSS problem as multiple instantaneous BSS problems. From the experimental study using the conventional frequency-domain approach, the source-separation performance is saturated because of the permutation problem.

In order to reduce the computational cost, a frequency-domain implementation of a time-domain algorithm was proposed [8, 13, 32, 39, 44, 57]. The approach evaluates independence among signals in the time domain, but updates the demixing process in the frequency domain. Hence, it can reduce the computational cost came from the convolution operation with long demixing filter while avoiding the permutation problem.

This chapter proposes a frequency-domain implementation algorithm that vanishes nonlinear cross-correlation defined in the time-domain. And also, we propose additional techniques for enhancing the robustness of the proposed algorithm.

This chapter is organized as follows. In Section 3.2, we formulate the convolutive BSS problem and describe two kinds of indeterminacy in convolutive BSS. We propose a convolutive BSS algorithm and additional techniques for enhancing the robustness of the proposed method in Sections 3.3 and 3.4. Section 3.5 describes actual implementation of the proposed algorithm.

3.2 CONVOLUTIVE BSS PROBLEM

3.2.1 *Mixing and demixing processes*

Let us consider a situation where N sources generates statistically independent signals $s_i(t)\ \ (i = 1, \dots N; t = 1, \dots, T)$ and N sensors observe convoluted mixtures of source signals. As is usual, each source is assumed to be a random process with zero mean, i.e., $E(s_i(t)) = 0\ \ (i = 1, \dots, N)$. In order for the BSS problem to be solvable, at most one source is allowed to be Gaussian.

In an element-wise representation, the mixing process is defined as

$$x_i(t) = \sum_{j=1}^{N} \sum_{\tau=0}^{\infty} a_{ij}(\tau)\, s_j(t-\tau)\ \ (i = 1, \dots N) \tag{3.1}$$

This can be expressed in a vector-matrix form:

$$\mathbf{x}(t) = \sum_{\tau=0}^{\infty} \mathbf{A}(\tau)\, \mathbf{s}(t-\tau), \tag{3.2}$$

where $\mathbf{s}(t) \triangleq [s_1(t), \dots, s_N(t)]^\top$ and $\mathbf{x}(t) \triangleq [x_1(t), \dots, x_N(t)]^\top$. If there is no time delay, then Eq. (3.2) becomes

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{3.3}$$

This is actually the same mixing process shown in Sect. 2.3. Hence, an instantaneous mixing process can be considered as the particular case of convolutive one.

Using z-transform, Eq. (3.2) can be rewritten as

$$\mathbf{x}(z) = \mathbf{A}(z)\, \mathbf{s}(z), \tag{3.4}$$

where $\mathbf{s}(z) \triangleq [s_1(z), \dots, s_N(z)]^\top$, $\mathbf{x}(z) \triangleq [x_1(z), \dots, x_N(z)]^\top$ and $\mathbf{A}(z) = [a_{ij}(z)]$ $(a_{ij}(z) \triangleq \sum a_{ij}(\tau) z^{-\tau})$. In this thesis, we assume that the mixing process $\mathbf{A}(z)$ is nonsingular, i.e., the inverse of $\mathbf{A}^{(f)}$ exists for every f, where f stands for frequency.

The demixing process, which is referred to as the separating filter or the separator, takes the following form

$$\mathbf{y}(t) = \sum_{l=-\infty}^{\infty} \mathbf{W}(l)\, \mathbf{x}(t-l), \tag{3.5}$$

where $\mathbf{y}(t) \triangleq [y_1(t), \dots, y_N(t)]^\top$. It should be noted that the impulse response $\{\mathbf{W}(l)\}$ generally takes a non-causal form, i.e., $\mathbf{W}(l) = \mathbf{O}\ \ (l < 0)$, where $\{\mathbf{W}(l)\} \triangleq \{\dots, \mathbf{W}(-1), \mathbf{W}(0), \mathbf{W}(1), \dots\}$ and $\mathbf{O}$ is the zero matrix.

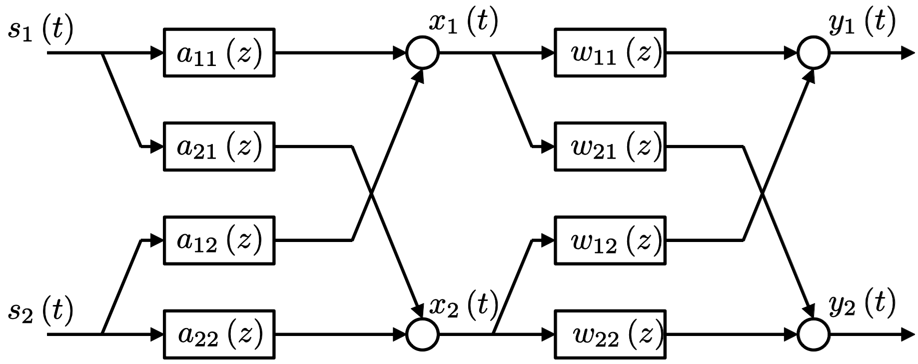

Figure 3: The scheme of blind source separation in the case of $N = 2$. Each sensor observes convoluted mixture of source signals. The recovered signals are obtained through the demixing process. The demixing process is determined so that the recovered signals become as statistically independent as possible.

If we know the mixing process $\mathbf{A}(z)$ beforehand to retrieve source signals from observed signals, then the separator should become $\mathbf{A}^{-1}(z)$. The main difficulty in BSS is that the separator must be determined only from the observations $\mathbf{x}(t)$ because there is no prior information about the mixing process. The scheme of BSS is depicted in Fig. 3.

3.2.2 *Two kinds of indeterminacy in convolutive BSS*

In BSS the term "source" has two kinds of indeterminacy. One is permutation indeterminacy: if $s_1(t), \ldots, s_N(t)$ are source signals, then $s_{\sigma(1)}(t), \ldots s_{\sigma(N)}(t)$, in which $\{\sigma(1), \ldots, \sigma(N)\}$ is an arbitrary permutation of $\{1, \ldots, N\}$, can be considered as another set of source signals. The other is filtering indeterminacy (or scaling indeterminacy in the particular case of instantaneous mixtures). Given the source signals, their arbitrarily linear filtered signals $d_1(z) s_1(z), \ldots, d_N(z) s_N(z)$ can also be considered as another set of source signals.

The two kinds of indeterminacy are summarizes as follows. If $\mathbf{s}(z)$ is a source vector, then $\tilde{\mathbf{s}}(z) = \mathbf{D}(z) \mathbf{P}\mathbf{s}(z)$ can be considered as a source vector, where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}(z)$ is a diagonal transfer function matrix with arbitrary analytic functions $d_1(z), \ldots, d_N(z)$ on the diagonal. There is no way to distinguish $\mathbf{s}(t)$ from $\tilde{\mathbf{s}}(t)$, because the information given beforehand is only that the sources are statistically independent.

Corresponding to the indeterminacy in the definition of sources, the design of the separator has a certain freedom. This thesis calls any separator of the following form a *valid* separator

$$\mathbf{W}(z) = \mathbf{D}(z) \mathbf{P}\mathbf{A}^{-1}(z). \tag{3.6}$$

If the separator is valid, then the overall transfer function matrix becomes a diagonal transfer function matrix with a column permutation, i.e., $\mathbf{W}(z)\,\mathbf{A}(z) = \mathbf{D}(z)\,\mathbf{P}$.

3.2.3 *Elimination of filtering indeterminacy*

Although two kinds of indeterminacy mentioned above are inherent in BSS, in actual applications one anyhow need to eliminate them to obtain the separator specifically. As for filtering indeterminacy $\mathbf{D}(z)$, there are several approaches.

A conventional way is to determine the separator $\mathbf{W}(z)$ so that the separator's outputs $y_i(t)$ will be independent and identically distributed (iid) [2]. For example, the separator's outputs satisfy

$$\mathbb{E}\left(f_i\left(y_i(t)\right) y_i(t-\tau)\right) = \begin{cases} 1 & \tau = 0 \\ 0 & \text{otherwise} \end{cases}$$

where f_i is arbitrary nonlinear function depending on a probability density function (pdf) of the i-th source. This kind of normalization approach involves several serious problems:

1. The quality of signals is deteriorated by the whitening operation. The signal quality is important in practical applications such as speech recognition, because it affects to recognition rate.

2. For non-stationary signals such as speech signals, the separator fluctuates with time, depending on the properties of the source signals.

3. When a source signal at a sensor becomes very small, some of entries in the separator may tend to be very large value, which can possibly induce some instability.

Another well-known approach is obtained by the minimal distortion principle (MDP) [46]. The MDP constraint for the separator is given as

$$\operatorname{diag}\left\{\mathbf{W}^{-1}(z)\right\} = \mathbf{I} \tag{3.7}$$

where $\operatorname{diag}\{\mathbf{W}(z)\}$ denotes a diagonal matrix whose diagonal elements are $w_{11}(z), \ldots, w_{NN}(z)$. Hence, a valid separator satisfying the MDP constraint retrieves the source signals observed in each sensor. It can be confirmed as follows.

If a separator is valid, then it takes a form of $\mathbf{W}(z) = \mathbf{D}(z)\,\mathbf{P}\mathbf{A}^{-1}(z)$. Substituting it into the left-hand side of Eq. (3.7) leads to

$$\begin{aligned} \operatorname{diag}\left\{\mathbf{W}^{-1}(z)\right\} &= \operatorname{diag}\left\{\mathbf{A}(z)\,\mathbf{P}\mathbf{D}^{-1}(z)\right\} \\ &= \operatorname{diag}\left\{\mathbf{A}(z)\,\mathbf{P}\right\}\mathbf{D}^{-1}(z) \end{aligned} \tag{3.8}$$

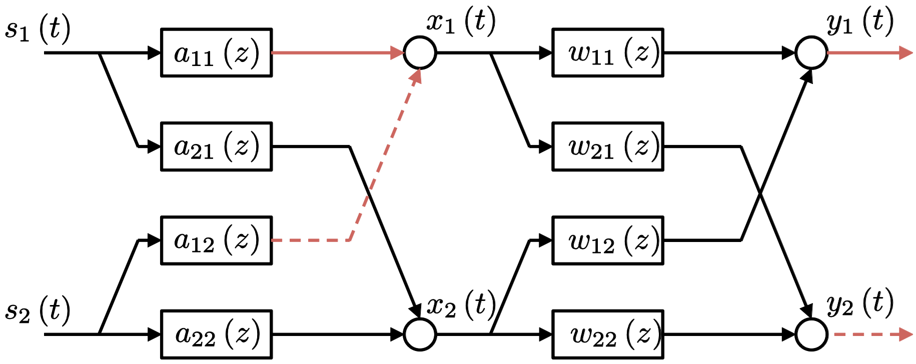

Figure 4: The desired outputs in the case of linear constraint ($N = 2$). First sensor observes a convoluted mixture of two source signals. If a proper separator is obtained, then the outputs of the separator becomes source signals observed at the first sensor.

Hence, filtering indeterminacy $\mathbf{D}(z)$ becomes

$$\mathbf{D}(z) = \operatorname{diag}\{\mathbf{A}(z)\,\mathbf{P}\} = \operatorname{diag}\left\{a_{1\sigma(1)}(z), \ldots, a_{N\sigma(N)}(z)\right\} \tag{3.9}$$

That implies the i-th output signal $y_i(t)$ is the $\sigma(i)$-th source signal observed at the i-th sensor:

$$y_i(t) = \sum_{\tau=0}^{\infty} a_{i\sigma(i)}(\tau)\, s_{\sigma(i)}(t-\tau).$$

As is described above, the MDP constraint has the useful property for applications. Due to its nonlinearity, however, it is difficult to incorporate the constraint into a convolutive BSS algorithm [46, 47].

Recently, another constraint based on the MDP was proposed by Matsuoka [48]. The constraint, which is called the linear constraint, is given by

$$\mathbf{e}^{\top}\mathbf{W}(z) = \mathbf{f}^{\top}, \tag{3.10}$$

where $\mathbf{e} \triangleq \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}^{\top} \in \mathbb{R}^{N}$ and $\mathbf{f} \triangleq \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^{\top} \in \mathbb{R}^{N}$. It a valid separator satisfies the constraint, it becomes

$$\begin{aligned} \mathbf{W}(z) &= \operatorname{diag}\left\{\mathbf{f}^{\top}\mathbf{A}(z)\,\mathbf{P}^{\top}\right\}\mathbf{P}\mathbf{A}^{-1}(z) \\ &= \operatorname{diag}\left\{a_{1\sigma(1)}(z), \ldots, a_{1\sigma(N)}(z)\right\}\mathbf{P}\mathbf{A}^{-1}(z). \end{aligned} \tag{3.11}$$

As is depicted in Fig. 4, this implies that the $\sigma(i)$-th source signal observed at the first sensor will appear in $y_i(t)$:

$$y_i(t) = \sum_{\tau} a_{1\sigma(i)}(\tau)\, s_{\sigma(i)}(t-\tau).$$

In contrast to the MDP constraint, the linear constraint can be easily incorporated into BSS algorithms [49, 72]. In this chapter we use this constraint for eliminating the filtering indeterminacy.

3.3 CONVOLUTIVE BSS ALGORITHM BASED ON NONLINEAR CORRELATION

3.3.1 *Measure for evaluating independence among signals*

If output signals are statistically independent, then any nonlinear correlation becomes zero. Hence, the objective is to find the separator $\mathbf{W}(z)$ that vanishes a nonlinear cross-correlation:

$$\mathrm{off}\left\{E\left(\boldsymbol{\varphi}(\mathbf{y}(t))\,\mathbf{y}^{\top}(t-\tau)\right)\right\} = \mathbf{O}\ \ (\tau = \ldots, -1, 0, 1, \ldots), \tag{3.12}$$

where $\boldsymbol{\varphi}(\mathbf{y}(t)) \triangleq [\varphi(y_1(t)), \ldots, \varphi(y_N(t))]^{\top}$ and $\varphi(\cdot) : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable nonlinear function. The operator $\mathrm{off}\{\mathbf{A}\}$ sets every diagonal elements of matrix $\mathbf{A}$ to be zero, i.e., $\mathbf{A} = \mathrm{diag}\{\mathbf{A}\} + \mathrm{off}\{\mathbf{A}\}$. Equation (3.12) gives $N(N-1)$ conditions for every τ and the remaining degrees of freedom in $\mathbf{W}(z)$, N, are eliminated by the linear constraint described in Section3.2.3. In this chapter, we call a valid separator satisfying the linear constraint as a *desired* separator. If a separator is desired, then it takes the form of (3.11).

If f_τ is an $N \times N$ vector-valued function of $N \times N$ impulse response $\{\mathbf{W}(l)\}$, then, Eq. (3.12) can be regarded as

$$\mathrm{off}\{f_\tau(\{\mathbf{W}(l)\})\} = \mathbf{O}\ \ (\tau = \ldots, -1, 0, 1, \ldots), \tag{3.13}$$

where $f_{ij,\tau}(\{\mathbf{W}(l)\})$ corresponds to $E\left(\varphi(y_i(t))\,y_j(t-\tau)\right)$. A basic strategy for derivation of the algorithm is to adopt the Newton method into (3.13) (i.e., (3.12)) then solve the simultaneous linear equation with respect to $\{\triangle\mathbf{W}(l)\}$.

The later subsection is devoted to the mathematical derivation of the algorithm. The derivation consists of two parts: a linear approximation of Eq. (3.12) and a transformation of the linear equation into the frequency domain. The actual implementation is described in the next Section 3.5.

3.3.2 *Derivation of algorithm*

In order to obtain the desired separator, we adopt the Newton method into Eq. (3.12). Suppose that, at a step, we have an approximate solution $\mathbf{W}(z)$ for the desired separator and move it to the desired one: $\mathbf{W}(z) \leftarrow \mathbf{W}(z) + \triangle\mathbf{W}(z)$. Namely, what we should do is to find $\mathbf{W}(z) + \triangle\mathbf{W}(z)$ so that it may satisfy

$$\mathrm{off}\left\{E\left(\boldsymbol{\varphi}(\mathbf{y}(t) + \triangle\mathbf{y}(t))\,(\mathbf{y}(t-\tau) + \triangle\mathbf{y}(t-\tau))^{\top}\right)\right\} = \mathbf{O}, \tag{3.14}$$

where

$$\triangle\mathbf{y}(t) = \sum_{l=-\infty}^{\infty} \triangle\mathbf{W}(l)\,\mathbf{x}(t-l). \tag{3.15}$$

The first-order Taylor expansion of the left-hand side of Eq. (3.14) leads to

$$\begin{aligned}
&\sum_{l=-\infty}^{\infty} \mathrm{off}\left\{E\left(\mathbf{\Phi}\left(\mathbf{y}\left(t\right)\right) \triangle \mathbf{W}\left(l\right) \mathbf{x}\left(t-l\right) \mathbf{y}^{\top}\left(t-\tau\right)\right)\right.\\
&\left.+E\left(\boldsymbol{\varphi}\left(\mathbf{y}\left(t\right)\right) \mathbf{x}^{\top}\left(t-l-\tau\right)\right) \triangle \mathbf{W}^{\top}\left(l\right)\right\}\\
&= -\mathrm{off}\left\{E\left(\boldsymbol{\varphi}\left(\mathbf{y}\left(t\right)\right) \mathbf{y}^{\top}\left(t-\tau\right)\right)\right\} \ \left(\tau = \ldots, -1, 0, 1, \ldots\right),
\end{aligned} \tag{3.16}$$

where $\mathbf{\Phi}\left(\mathbf{y}\left(t\right)\right) \triangleq \mathrm{diag}\left\{\varphi'\left(y_1\left(t\right)\right), \ldots, \varphi'\left(y_N\left(t\right)\right)\right\}$ and $\varphi'\left(y_i\left(t\right)\right)$ denotes the derivative of $\varphi\left(y_i\left(t\right)\right)$. Although solving Eq. (3.16) with respect to $\triangle \mathbf{W}\left(z\right)$ and updating $\mathbf{W}\left(z\right) \leftarrow \mathbf{W}\left(z\right) + \triangle \mathbf{W}\left(z\right)$ would give more accurate separator, it is difficult to directly solve the equation.

In order to derive an easier equation we introduce a variable $\triangle \mathbf{G}\left(z\right)$ based on the concept of the natural gradient. Let the inverse of the impulse response $\left\{\mathbf{W}\left(l\right)\right\}$ be $\left\{\bar{\mathbf{W}}\left(l\right)\right\}$; then, the observed signals can be expressed as

$$\mathbf{x}\left(t\right) = \sum_{k=-\infty}^{\infty} \bar{\mathbf{W}}\left(k\right) \mathbf{y}\left(t-k\right). \tag{3.17}$$

Substituting (3.17) into (3.16) leads

$$\begin{aligned}
&\sum_{l,k} \mathrm{off}\left\{E\left(\mathbf{\Phi}\left(\mathbf{y}\left(t\right)\right) \triangle \mathbf{W}\left(l\right) \bar{\mathbf{W}}\left(k-l\right) \mathbf{y}\left(t-k\right) \mathbf{y}^{\top}\left(t-\tau\right)\right)\right.\\
&\left.+E\left(\boldsymbol{\varphi}\left(\mathbf{y}\left(t\right)\right) \mathbf{y}^{\top}\left(t-k-\tau\right)\right) \bar{\mathbf{W}}^{\top}\left(k-l\right) \mathbf{W}^{\top}\left(l\right)\right\}\\
&= -\mathrm{off}\left\{E\left(\boldsymbol{\varphi}\left(\mathbf{y}\left(t\right)\right) \mathbf{y}^{\top}\left(t-\tau\right)\right)\right\} \ \left(\tau = \ldots, -1, 0, 1, \ldots\right).
\end{aligned} \tag{3.18}$$

Define

$$\triangle \mathbf{G}\left(k\right) = \left[\triangle g_{ij}\left(k\right)\right] \triangleq \sum_{l=-\infty}^{\infty} \triangle \mathbf{W}\left(l\right) \bar{\mathbf{W}}\left(k-l\right), \tag{3.19}$$

then Eq. (3.18) becomes

$$\begin{aligned}
&\sum_{k=-\infty}^{\infty} \mathrm{off}\left\{E\left(\mathbf{\Phi}\left(\mathbf{y}\left(t\right)\right) \triangle \mathbf{G}\left(k\right) \mathbf{y}\left(t-k\right) \mathbf{y}^{\top}\left(t-\tau\right)\right)\right.\\
&\left.+E\left(\boldsymbol{\varphi}\left(\mathbf{y}\left(t\right)\right) \mathbf{y}^{\top}\left(t-k-\tau\right)\right) \triangle \mathbf{W}^{\top}\left(k\right)\right\}\\
&= -\mathrm{off}\left\{E\left(\boldsymbol{\varphi}\left(\mathbf{y}\left(t\right)\right) \mathbf{y}^{\top}\left(t-\tau\right)\right)\right\} \ \left(\tau = \ldots, -1, 0, 1, \ldots\right).
\end{aligned} \tag{3.20}$$

The above-mentioned equation can be decomposed into a set of pairs of equations with respect to $\triangle g_{ij}\left(z\right)$ and $\triangle g_{ji}\left(z\right)$ $\left(i \neq j\right)$ as follows. Since we have assumed that $\mathbf{W}\left(z\right)$ approximately attains separation, variables $y_i\left(t\right)$ and $y_j\left(t-\tau\right)$ $\left(j \neq i\right)$ are almost mutually independent. That leads to

$$\begin{aligned}
E\left(\varphi'\left(y_i\left(t\right)\right) y_j\left(t-k\right) y_q\left(t-\tau\right)\right) &\approx 0 \ \left(j \neq q\right),\\
E\left(\varphi\left(y_i\left(t\right)\right) y_j\left(t-\tau-k\right)\right) &\approx 0 \ \left(i \neq j\right).
\end{aligned}$$

Incorporating them and picking out the (i,j)-th entry in (3.20) yield

$$\begin{aligned}&\sum_{k=-\infty}^{\infty}\left\{E\left(\varphi'\left(y_i\left(t\right)\right)y_j\left(t-k\right)y_j\left(t-\tau\right)\right)\triangle g_{ij}\left(k\right)\right.\\&\left.+E\left(\varphi\left(y_i\left(t\right)\right)y_i\left(t-k-\tau\right)\right)\triangle g_{ji}\left(k\right)\right\}\\&=-E\left(\varphi\left(y_i\left(t\right)\right)y_j\left(t-\tau\right)\right)\quad\left(i\neq j\right).\end{aligned}\tag{3.21}$$

Taking into account independence between $\varphi'(y_i(t))$ and $y_j(t-k)\,y_j(t-\tau)$, Eq. (3.4) can be approximated as

$$\sum_{k=-\infty}^{\infty}c_i\left(t\right)r_{jj}\left(t,\tau-k\right)\triangle g_{ij}\left(k\right)+\gamma_{ii}\left(t,\tau+k\right)\triangle g_{ji}\left(k\right)=-\gamma_{ij}\left(t,\tau\right),\tag{3.22}$$

where

$$\begin{aligned}c_i\left(t\right)&\triangleq E\left(\varphi'\left(y_i\left(t\right)\right)\right),\\r_{ij}\left(t,\tau\right)&\triangleq E\left(y_i\left(t\right)y_j\left(t-\tau\right)\right),\\\gamma_{ij}\left(t,\tau\right)&\triangleq E\left(\varphi\left(y_i\left(t\right)\right)y_j\left(t-\tau\right)\right).\end{aligned}$$

Exchanging i and j in (3.22) leads to another equation with respect to $\triangle g_{ij}(z)$ and $\triangle g_{ji}(z)$:

$$\sum_{k=-\infty}^{\infty}c_j\left(t\right)r_{ii}\left(t,\tau-k\right)\triangle g_{ji}\left(k\right)+\gamma_{jj}\left(t,\tau+k\right)\triangle g_{ij}\left(k\right)=-\gamma_{ji}\left(t,\tau\right).\tag{3.23}$$

Thus, our task comes down to solving Eqs. (3.22) and (3.23) with respect to every pair of $\triangle g_{ij}(z)$ and $\triangle g_{ji}(z)$ $(i\neq j)$. It should be noted that, we should not omit the time index t in $c_i(t)$, $r_{ij}(t,\tau)$ and $\gamma_{ij}(t,\tau)$ because a random process of source is assumed to be a non-stationary one. This is a very important point in separation of speech signals.

If the separating filter is very large, then solving Eqs. (3.22) and (3.23) with respect to $\triangle g_{ij}(z)$ and $\triangle g_{ji}(z)$ $(i\neq j)$ becomes time-consuming. It is because solving them requires the inverse matrix calculation, the size of which depends on the length of the separating filter.

In order to reduce the computational load, Eqs. (3.22) and (3.23) are transformed into the frequency-domain:

$$\begin{cases}c_i\left(t\right)r_{jj}^{(f)}\left(t\right)\triangle g_{ij}^{(f)}+\gamma_{ii}^{(f)}\left(t\right)\triangle g_{ji}^{(f)*}=-\gamma_{ij}^{(f)}\left(t\right)\\c_j\left(t\right)r_{ii}^{(f)}\left(t\right)\triangle g_{ji}^{(f)}+\gamma_{jj}^{(f)}\left(t\right)\triangle g_{ij}^{(f)*}=-\gamma_{ji}^{(f)}\left(t\right)\end{cases}\tag{3.24}$$

where

$$\begin{aligned}r_{ij}^{(f)}\left(t\right)&\triangleq\sum_{\tau=-\infty}^{\infty}r_{ij}\left(t,\tau\right)e^{-j2\pi f\tau},\\\gamma_{ij}^{(f)}\left(t\right)&\triangleq\sum_{\tau=-\infty}^{\infty}\gamma_{ij}\left(t,\tau\right)e^{-j2\pi f\tau},\\\triangle g_{ij}^{(f)}&\triangleq\sum_{\tau=-\infty}^{\infty}\triangle g_{ij}\left(\tau\right)e^{-j2\pi f\tau},\end{aligned}$$

and $\triangle g_{ji}^{(f)*}$ stands for the complex conjugation of $\triangle g_{ji}^{(f)}$. In a vector-matrix form, the above-mentioned equation can be expressed as

$$\begin{bmatrix} c_i(t)\, r_{jj}^{(f)}(t) & \gamma_{ii}^{(f)}(t) \\ \gamma_{jj}^{(f)*}(t) & c_j(t)\, r_{ii}^{(f)}(t) \end{bmatrix} \begin{bmatrix} \triangle g_{ij}^{(f)} \\ \triangle g_{ji}^{(f)*} \end{bmatrix} = - \begin{bmatrix} \gamma_{ij}^{(f)}(t) \\ \gamma_{ji}^{(f)*}(t) \end{bmatrix}$$

or

$$\boldsymbol{\Gamma}_{ij}^{(f)}(t)\, \triangle \mathbf{g}_{ij}^{(f)} = -\boldsymbol{\gamma}_{ij}^{(f)}(t). \tag{3.25}$$

To find $\Delta \mathbf{g}_{ij}^{(f)}$ that approximates Eq. (3.25) for every t, we minimize the following function

$$Q_{ij}\left(\triangle \mathbf{g}_{ij}^{(f)}\right) \triangleq \left\langle \left\| \boldsymbol{\Gamma}_{ij}^{(f)} \triangle \mathbf{g}_{ij}^{(f)} + \boldsymbol{\gamma}_{ij}^{(f)} \right\|^2 \right\rangle_t, \tag{3.26}$$

where $\langle \cdot \rangle_t$ denotes the time average. Since Q_{ij} is a real-valued function of two complex variables $\triangle g_{ij}^{(f)}$ and $\triangle g_{ji}^{(f)}$ (i.e., $Q_{ij} : \mathbb{C}^2 \times \mathbb{C}^2 \to \mathbb{R}$), we use the complex derivatives described in Chap. 2 in order to minimize the function. Applying the complex derivatives to Q_{ij} and solving $\partial Q_{ij} / \partial \triangle \mathbf{g}_{ij}^{(f)*} = \mathbf{0}$ with respect to $\triangle \mathbf{g}_{ij}^{(f)}$ leads

$$\triangle \mathbf{g}_{ij}^{(f)} = -\left\langle \boldsymbol{\Gamma}_{ij}^{(f)H}(t)\, \boldsymbol{\Gamma}_{ij}^{(f)}(t) \right\rangle_t^{-1} \left\langle \boldsymbol{\Gamma}_{ij}^{(f)H}(t)\, \boldsymbol{\gamma}_{ij}^{(f)}(t) \right\rangle_t \quad (i \neq j). \tag{3.27}$$

This algorithm cannot determine the diagonal entries of $\triangle \mathbf{G}^{(f)}$, because $\triangle g_{ii}^{(f)}$ does not appear in the (i, j)-th entry in Eq. (3.20). Hence, the remaining entries are determined as follows. Equation (3.10) implies $\mathbf{e}^\top \triangle \mathbf{W}^{(f)} = \mathbf{0}^\top$, leading to $\mathbf{e}^\top \triangle \mathbf{G}^{(f)} = \mathbf{0}^\top$. Thus, the diagonal entries are determined by

$$\operatorname{diag}\left\{\triangle \mathbf{G}^{(f)}\right\} = -\operatorname{diag}\left\{ \sum_{i \neq 1} \triangle g_{i1}^{(f)}, \ldots, \sum_{i \neq N} \triangle g_{iN}^{(f)} \right\}. \tag{3.28}$$

The matrix $\triangle \mathbf{W}^{(f)}$ can be calculated as

$$\triangle \mathbf{W}^{(f)} = \triangle \mathbf{G}^{(f)} \mathbf{W}^{(f)} \text{ for every } f, \tag{3.29}$$

then $\mathbf{W}^{(f)}$ is updated as

$$\mathbf{W}^{(f)} \leftarrow \mathbf{W}^{(f)} + \triangle \mathbf{W}^{(f)}. \tag{3.30}$$

The impulse response $\{\mathbf{W}(l)\}$ can be found by applying the inverse Fourier transform to $\mathbf{W}^{(f)}$:

$$\mathbf{W}(l) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{W}^{(f)} e^{j2\pi f l} df. \tag{3.31}$$

3.3.3 *What is the difference between the frequency-domain approach and the frequency-domain implementation one?*

In the formulation of the proposed algorithm, we begin with aiming to determine the separator so that the nonlinear cross correlation $E\left(\varphi\left(y_i\left(t\right)\right)y_j\left(t-\tau\right)\right)$ will be eliminated. In the final stage of the formulation, however, we transformed the time-domain correlation into the frequency-domain one, $\gamma_{ij}^{(f)}\left(t\right)$, which is actually estimated as a time average of $F\left[\varphi\left(y_i\left(m,t\right)\right)\right]F^*\left[y_j\left(m,t\right)\right]$, where $F\left[\cdot\right]$ denotes the fast Fourier transform (FFT) operator. Concrete operations will be described in later section. In contrast, in the conventional frequency-domain approach, signals $y_i\left(t\right)$ are first transformed into the frequency-domain ones, $F\left[y_i\left(m,t\right)\right]$, and then their nonlinear correlation, i.e., a time average of $\varphi\left(F\left[y_i\left(m,t\right)\right]\right)F^*\left[y_j\left(m,t\right)\right]$, is evaluated. Since both the methods are eventually formulated in the frequency domain, they do not appear to be so much different from each other, but that is not the case.

In order to see what happens in the two approaches, let us consider the case where each source signal $\left(N=2\right)$ comprises only two frequency components of f and $3f$:

$$\begin{aligned} s_1\left(t\right) &= c_1^{(f)}\left(t\right)+c_1^{(3f)}\left(t\right), \\ s_2\left(t\right) &= c_2^{(f)}\left(t\right)+c_2^{(3f)}\left(t\right), \end{aligned}$$

where $c_i^{(f)}\left(t\right)$ represents a sinusoidal wave of frequency f. In addition, we assume that $c_i^{(3f)}\left(t\right)$ is a higher harmonic of $c_i^{(f)}\left(t\right)$, i.e., $c_i^{(3f)}\left(t\right)$ is synchronized with $c_i^{(f)}\left(t\right)$. If the separator reproduces the sources as

$$\begin{aligned} y_1\left(t\right) &= c_1^{(f)}\left(t\right)+c_1^{(3f)}\left(t\right), \\ y_2\left(t\right) &= c_2^{(f)}\left(t\right)+c_2^{(3f)}\left(t\right) \end{aligned} \tag{3.32}$$

then the separation is successful, of course. In the conventional frequency-domain approach, however, the separator can possibly produce undesirable outputs as

$$\begin{aligned} y_1\left(t\right) &= c_1^{(f)}\left(t\right)+c_2^{(3f)}\left(t\right), \\ y_2\left(t\right) &= c_2^{(f)}\left(t\right)+c_1^{(3f)}\left(t\right) \end{aligned} \tag{3.33}$$

Since the frequency-domain approach evaluates only independence between $c_1^{(f)}\left(t\right)$ and $c_2^{(f)}\left(t\right)$ or that between $c_1^{(3f)}\left(t\right)$ and $c_2^{(3f)}\left(t\right)$, there is no way to decide which of Eqs. (3.32) and (3.33) is the correct answer.

If a nonlinear operation is applied to the signals before converting them into the frequency domain, the situation becomes very different. If we adopt, for example, the cubic function for φ, we have in the case of Eq. (3.33)

$$y_1\left(t\right)^3 = c_1^{(f)}\left(t\right)^3+3c_1^{(f)}\left(t\right)^2c_2^{(3f)}\left(t\right)+\cdots.$$

$c_1^{(f)}(t)^3$ includes a component of frequency $3f$ and it is synchronized with or completely *dependent* on $c_1^{(3f)}(t)$ in $y_2(t)$. That implies that, by checking independence between the frequency components of $3f$ in $y_1(t)^3$ and $y_2(t)$, one can conclude that Eq. (3.33) is not a desired result. Generally speaking, the nonlinear operation φ for $y_i(t)$ in the time domain produces some higher-order harmonics of each frequency component and that enables us to check independence among *different* frequency components in $y_1(t), \ldots, y_N(t)$.

Thus we emphasize that the proposed approach is essentially a time-domain one. The frequency-domain treatment given there is just to relax the computational load.

3.4 ADDITIONAL TECHNIQUES

3.4.1 *Additional constraint on the separator*

For sound data obtained in practical environments, the length of the separator (an FIR filter) must be very long, typically one thousand taps. For such a high-order filter, the algorithm proposed in previous section does not work well by itself. We often find incomprehensible instability in executing the algorithm. In the early stage of the iteration, the separator appears to be updated in the desired direction, but as the iteration proceeds, the norm of the separator and the amplitude of the output suddenly grow large and finally reach the incomputable level.

To cope with this instability, the algorithm needs to introduce another constraint than just the condition that $y_i(t)$ be (nonlinearly) uncorrelated to each other. Below we show a simple but very effective way for solving the problem.

The additional constraint is

$$\mathrm{tr}\left\{E\left(\mathbf{y}(t)\,\mathbf{y}^\top(t-\tau)\right)\right\} = \mathbf{f}^\top E\left(\mathbf{x}(t)\,\mathbf{x}^\top(t-\tau)\right)\mathbf{f} \tag{3.34}$$

or

$$\sum_{i=1}^{N} E\left(y_i(t)\,y_i(t-\tau)\right) = E\left(x_1(t)\,x_1(t-\tau)\right). \tag{3.35}$$

For $\tau = 0$, the above relation reduces to $\sum_{i=1}^{N} E\left(y_i(t)^2\right) = E\left(x_1(t)^2\right)$. This implies that the variance of output $y_i(t)$ never exceeds that of first sensor signal $x_1(t)$. Namely, if the desired separator is searched within the set of separators satisfying (3.34) or (3.35), it never diverges to infinity.

The validity of the above condition can easily be understood. Equation (3.10) leads to $\mathbf{e}^\top \mathbf{y}(t) = \mathbf{f}^\top \mathbf{x}(t)$ and hence

$$\mathbf{e}^\top E\left(\mathbf{y}(t)\,\mathbf{y}^\top(t-\tau)\right)\mathbf{e} = \mathbf{f}^\top E\left(\mathbf{x}(t)\,\mathbf{x}^\top(t-\tau)\right)\mathbf{f} \tag{3.36}$$

or

$$E\left(\left(\sum_{i=1}^{N} y_i(t)\right)\left(\sum_{i=1}^{N} y_i(t-\tau)\right)\right) = E\left(x_1(t)\, x_1(t-\tau)\right). \tag{3.37}$$

For $\tau = 0$, it implies $E\left(\left(\sum_{i=1}^{N} y_i(t)\right)^2\right) = E\left(x_1(t)^2\right)$. This relation by itself, however, does not guarantee the boundedness of the variance of $y_i(t)$; it only demands that the variance of $\mathbf{e}^\top \mathbf{y}(t)$ be bounded. To cope with this issue, we utilize the following simple fact: when the separator achieves perfect separation, the (linear) correlation matrix $E\left(\mathbf{y}(t)\,\mathbf{y}^\top(t-\tau)\right)$ must be diagonal; in other words, $\mathbf{e}^\top E\left(\mathbf{y}(t)\,\mathbf{y}^\top(t-\tau)\right)\mathbf{e}$ must be equal to $\mathrm{tr}\left\{E\left(\mathbf{y}(t)\,\mathbf{y}^\top(t-\tau)\right)\right\}$. Combining this fact and Eq. (3.36) leads to the necessary condition (3.34) for the desired separator.

Equation (3.36) can be rewritten as

$$\mathrm{tr}\left\{\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} \mathbf{W}(k)\,\mathbf{R}(\tau+l-k)\,\mathbf{W}^\top(l)\right\} = \mathbf{f}^\top \mathbf{R}(\tau)\,\mathbf{f} \tag{3.38}$$

or in the frequency domain

$$\mathrm{tr}\left\{\mathbf{W}^{(f)}\mathbf{R}^{(f)}\mathbf{W}^{(f)H}\right\} = \mathbf{f}^\top \mathbf{R}^{(f)}\mathbf{f}, \tag{3.39}$$

where

$$\mathbf{R}(\tau) \triangleq E\left(\mathbf{x}(t)\,\mathbf{x}^\top(t-\tau)\right),$$
$$\mathbf{R}^{(f)} \triangleq \sum_{\tau=-\infty}^{\infty} \mathbf{R}(\tau)\, e^{-j2\pi f\tau}.$$

The proposed algorithm incorporates the additional constraint as follows. Let $\mathbf{W}^{(f)}$ be a separator obtained at a step of the iteration. The algorithm modifies $\mathbf{W}^{(f)}$ so that it will satisfy the above constraint with the least modification. This can be done in the following way. The matrix that satisfies the linear constraint with the minimum norm proves to be $\mathbf{e}\mathbf{f}^\top/N$. In the space of the separator, we consider a straight line that connects $\mathbf{e}\mathbf{f}^\top/N$ and $\mathbf{W}^{(f)}$; every separator on the line, which is represented by $\beta_f \mathbf{W}^{(f)} + (1-\beta_f)\,\mathbf{e}\mathbf{f}^\top/N \;\; (\beta_f > 0)$, satisfies the linear constraint. The parameter β_f for which relation (3.39) with $\beta_f \mathbf{W}^{(f)} + (1-\beta_f)\,\mathbf{e}\mathbf{f}^\top/N$ and solving with respect to β_f. Thus, solution is

$$\beta_f = \left(\frac{(N-1)\,\mathbf{f}^\top \mathbf{R}^{(f)}\mathbf{f}}{N\cdot \mathrm{tr}\left\{\mathbf{W}^{(f)}\mathbf{R}^{(f)}\mathbf{W}^{(f)H}\right\} - \mathbf{f}^\top \mathbf{R}^{(f)}\mathbf{f}}\right)^{\frac{1}{2}}. \tag{3.40}$$

Using this, the separator is modified as

$$\mathbf{W}^{(f)} \leftarrow \beta_f \mathbf{W}^{(f)} + (1-\beta_f)\,\frac{\mathbf{e}\mathbf{f}^\top}{N}. \tag{3.41}$$

3.4.2 *Varying-window-size technique*

The time-domain approach is thought to be free from the permutation problem. The reason was described in subsection 3.3.3. In the actual experiments, however, we sometimes face undesired permutation even when a time-domain approach is taken.

In [47], an example is shown in which an undesirable permutation occurs even though the algorithm adopted there is totally a time-domain approach. The authors observed the following phenomenon in an experiment of two sources. According to what kind of permutation occurred, the frequency axis could be divided into a small number of regions or frequency bands. In one region, an s_1-component appears in $y_1(t)$ and an s_2-component in $y_2(t)$; in another region, the relation is reversed. That example suggests that just employing the time-domain approach does not solve the permutation problem as we wish.

The algorithm used in [47] is the natural gradient method proposed by Amari et al. [15]. Since the final configuration of the frequency regions was dependent on the initial setting of the separator parameters, the above phenomenon seems due to the fact that the algorithm only makes a *local* search for the desired separator. A way to solve the problem is to suppress the frequency resolution of the separator; it is equivalent to shorten the substantial length of the separating filter. This idea is proposed in [61].

While the suppression of the frequency resolution alleviates the permutation problem, it may lead to poor separation accuracy, as reported in [26]. In [27, 47], an effective way to solve this trade-off is proposed. In the early steps of the repetitive modification, the frequency resolution of the separator is made very low. As the learning proceeds, the resolution is raised gradually, and finally reaches the level of the inverse of the full length of the separating filter. Also in the present algorithm, we incorporate this procedure, i.e., we alter the substantial width l_2 of the separating filter as $l_2(q+1) = 2l_2(q)$ until it reaches the full length of the separating filter, where q stands for the step number of iteration. The short filter length in the early stage yields a low frequency resolution, and the long filter length in the final stage provides a high frequency resolution. Some examples in later section and chapter will show that this varying-window-size technique is very effective and indispensable to accurate separation.

3.5 ACTUAL IMPLEMENTATION

In the actual implementation, the time series data are decomposed into frames with interval $2L$. Time t and frequency f in $\gamma_{ij}^{(f)}(t)$ are replaced by the frame number m and discrete frequencies $k/2L$. Cross spectrum $\gamma_{ij}^{(f)}(t)$, for example, is calculated by applying the FFT to the data in the m-th frame:

$$\gamma_{ij}^{(k)}(m) = F\left[\varphi\left(y_i(m,t)\right)\right] F^*\left[y_j(m,t)\right] \tag{3.42}$$

for $m = 1, \ldots, M$ and $k = -L+1, \ldots, L$, where $F\left[y_i(m,t)\right]$ is the FFT of $y_i(t)$ in the m-th frame. Other spectra are also calculated in the same operation. $c_i(t)$ is defined as the time average of $\varphi'(y_i(t))$ in the m-th frame.

According to the operations, $\Delta \mathbf{g}_{ij}^{(k)}$ is calculated by

$$\triangle \mathbf{g}_{ij}^{(k)} = -\left\langle \mathbf{\Gamma}_{ij}^{(k)H}(m)\, \mathbf{\Gamma}_{ij}^{(k)}(m) \right\rangle_m^{-1} \left\langle \mathbf{\Gamma}_{ij}^{(k)H}(m)\, \gamma_{ij}^{(k)}(m) \right\rangle_m \quad (i \neq j), \tag{3.43}$$

where $\Delta g_{ij}^{(k)}$ denotes $\Delta g_{ij}^{(f)}$ for $f = k/2L$ and $\langle \cdot \rangle_m$ represents average with respect to m. The separator is expressed as an FIR filter:

$$\mathbf{W}(z) = \sum_{l=-L+1}^{L} \mathbf{W}(l)\, z^{-l}.$$

In order to understand how to implement the frequency-domain implemented nonlinear de-correlation (FDI-ND) algorithm, we summarize the whole procedure. Let q be the step number of the iteration and $u(q,l)$ be a window function that corresponds to the varying-window-size technique.

(P1) Initialize $\mathbf{W}(0,z)$ and $u(0,l)$.

(P2) Calculate $\mathbf{R}^{(k)}$ by $\frac{1}{M}\sum_{m=1}^{M} F\left[\mathbf{x}(m,t)\right] F\left[\mathbf{x}(m,t)\right]^H$.

(P3) Calculate the output signals $\mathbf{y}(t)$ using Eq. (3.5), then apply a nonlinear function φ to the outputs in *the time domain*.

(P4) Obtain $F\left[\mathbf{y}(m,t)\right]$and $F\left[\varphi\left(\mathbf{y}(m,t)\right)\right]$ in each frame by applying the short-time Fourier transform (STFT). Then, calculate $c_i(m), r_{ii}^{(k)}(m)$ and $\gamma_{ij}^{(k)}(m)$ for every pair of i and j in the same operation described above.

(P5) Obtain $\triangle \mathbf{G}^{(k)}(q)$ using Eqs. (3.28) and (3.43).

(P6) Update the separator in each frequency bin: $\mathbf{W}^{(k)}(q+1) \leftarrow \mathbf{W}^{(k)}(q) + \triangle \mathbf{G}^{(k)}(q)\, \mathbf{W}^{(k)}(q)$.

(P7) Calculate β_k by Eq. (3.40), and then modify the separator as $\mathbf{W}^{(k)}(q+1) \leftarrow \beta_k \mathbf{W}^{(k)}(q+1) + (1-\beta_k)\, \mathbf{e}\mathbf{f}^\top / N$.

(P8) After modification, calculate the inverse FFT of the separator.

(P9) Apply the varying-window-size technique to the separator: $\mathbf{W}(q+1,l) \leftarrow u(q,l)\, \mathbf{W}(q+1,l)\ \ (l = -L+1, \ldots, L)$.

(P10) Update the length of $u(q,l)$ to be long; $l_2(q+1) = 2l_2(q)$ until it reaches the full length L.

(P11) Go back to P3 until a stopping criteria is met.

3.6 summary

In this chapter, we propose a convolutive BSS algorithm based on nonlinear correlation. It evaluates statistical independence among the signals by a nonlinear cross-correlations defined in the time domain and handles the separating filter in the frequency domain in order to relax the computational cost. In addition, the algorithm incorporates two additional techniques: one is a constraint that guarantees the boundedness of the output signals, and the other is a varying-window-size technique for solving the permutation problem [27, 46, 72].

4 STABILITY ON INDEPENDENT VECTOR ANALYSIS

4.1 INTRODUCTION

As is described in Chapter 2, ICA is a method for finding statistically independent sources from observed data. The model used in ICA assumes simple linear mixtures of source components. If the source signals are recorded in an acoustic environment, the mixtures of sources should be considered as convoluted mixtures, i.e., there are propagation time delay and reverberation in the observed signals.

The convolutive BSS problem can be simplified down into multiple, instantaneous BSS problem by transforming the observed data into the frequency domain. One therefore can apply ICA to data in each frequency bin. This simplification, however, comes from at the cost of a cumbersome issue called the permutation problem. Although several approaches, such as smoothness constraint on the demixing matrices of adjacent frequencies, and direction-of-arrival estimation, have been proposed, they require some complicated and time-consuming procedures [7, 37, 53, 61, 66, 67].

Recently, a new approach to the permutation problem was proposed, which is termed independent vector analysis (IVA). IVA is a method for recovering source vectors while preserving the inner-dependency of elements in each vector. Hence, it is a prospective tool for solving the convolutive BSS problem without the permutation problem. The framework of IVA is as follows:

(I1) The prior probability density function (pdf) q_i of source vector i is introduced; then the prior joint pdf of all the sources is $\Pi_{i=1}^{N} q_i$.

(I2) Let the actual joint pdf of the output of the separator be p, which is a function of K instantaneous separators.

(I3) Define a measure of difference between $\Pi_{i=1}^{N} q_i$ and p.

(I4) Minimizing the measure gives a desired separator.

A key point in IVA is that the prior pdf q_i is chosen such that a set of the frequency components in each source vector are statistically dependent.

The effectiveness of IVA approaches has been demonstrated by many applications [41, 42, 43]. However, there are very few mathematical analyses of the algorithm. In order for an IVA algorithm to work as intended, at least the following propositions must hold:

Proposition 1. *The desired set of separators must be a local minimum of a measure for evaluating independence among the source vectors;*

This proposition is very basic proposition in order to obtain the desired separator by minimizing the measure.

Proposition 2. *The set of separators that includes undesired permutations must not be a local minimum of the measure.*

The objective of this chapter is to show a necessary and sufficient condition for these propositions under some assumptions and to prove that the separator satisfying the first proposition 1 satisfies the second proposition 2 as a necessity. The latter result strongly supports the validity of the IVA approach.

This chapter is organized as follows. In Section 4.2, we describe the mixing and demixing models dealt with in IVA. Section 4.3 derives a stability condition for the desired set of the separators, then we investigate a condition for permuted separator not to be positive semi-definite in Section 4.4. Stability analysis for complex-valued IVA is performed in Sections 4.5 and 4.6.

4.2 SEPARATION OF REAL-VALUED SOURCE VECTORS

In this and next sections, variables and constants are all real-valued; a complex-valued signals will be shown in Section 4.5.

4.2.1 *Mixing and demixing models*

In this section, we define the mixing and demixing models dealt with in IVA. To start with, an element-wise representation is used, where $M \times K$ observation $x_i^{(k)}$ is assumed to be given as a scaled sum of $N \times K$ latent sources $s_j^{(k)}$:

$$x_i^{(k)} = \sum_{j=1}^{N} a_{ij}^{(k)} s_j^{(k)} \ \ (i = 1, \ldots, M; k = 1, \ldots, K) \,. \tag{4.1}$$

An important point in IVA is that the mixing is made only among the components $s_1^{(k)}, \ldots, s_N^{(k)}$ with a common superscript k, while no mixing occurs between $s_i^{(k)}$ and $s_j^{(l)}$ with $k \neq l$. Each of observed mixtures $x_i^{(k)}$ and of sources $s_j^{(k)}$ are random variables.

In a vector-matrix representation, Eq. (4.1) can be written as

$$\mathbf{x}^{(k)} = \mathbf{A}^{(k)} \mathbf{s}^{(k)} \ \ (k = 1, \ldots, K) \,, \tag{4.2}$$

where $\mathbf{s}^{(k)} \triangleq \left[s_1^{(k)}, \ldots, s_N^{(k)}\right]^\top$ and $\mathbf{x}^{(k)} \triangleq \left[x_1^{(k)}, \ldots, x_M^{(k)}\right]^\top$. One can tell at a glance that the mixing model used in IVA is modeled as a set of models in ICA. The mixing model of IVA in the case of two sources is depicted in Fig. 5.

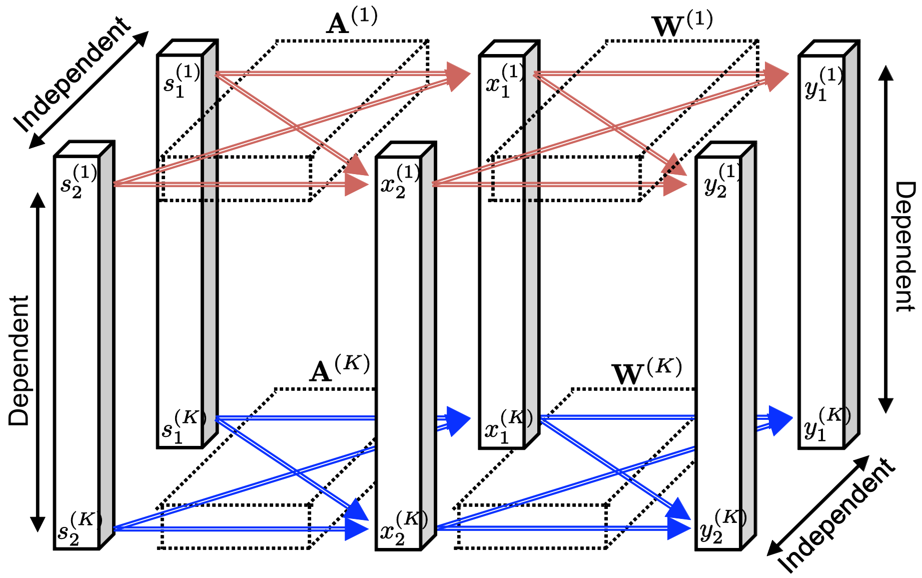

Figure 5: The mixing and demixing models of IVA in the case of $N = 2$. ICA is extended to a formulation with vectors, where the components in each vector are dependent. The difference between IVA and multidimensional ICA or independent subspace analysis (ISA) is that the mixing in IVA is made only the components on the same subspace, while the mixing in ISA occurs whole components using the same linear mixing model of ICA.

A different point of view leads another vector-matrix representation such that

$$\mathbf{x}_i = \sum_{j=1}^{N} \mathbf{a}_{ij} \circledast \mathbf{s}_j \tag{4.3}$$

or

$$\mathbf{x}_i = \sum_{j=1}^{N} \mathbf{A}_{ij}\mathbf{s}_j, \tag{4.4}$$

where $\mathbf{s}_j \triangleq \left[s_j^{(1)}, \ldots, s_j^{(K)}\right]^\top$, $\mathbf{x}_i \triangleq \left[x_i^{(1)}, \ldots, x_i^{(K)}\right]^\top$, and symbol $\circledast$ denotes the Hadamard product. Matrices $\mathbf{A}_{ij}$ are $K \times K$ diagonal matrices whose elements are $a_{ij}^{(1)}, \ldots, a_{ij}^{(K)}$, i.e., $\mathbf{A}_{ij} = \mathrm{diag}\left\{\mathbf{a}_{ij}\right\}$.

IVA has assumptions so that underlying sources can be estimated in a correct way: the first assumption is that source vectors $\mathbf{s}_1, \ldots, \mathbf{s}_N$ are statistically independent of each other, and the other is that components in a vectors are mutually dependent. In order for the problem to be simple, this and subsequent chapters takes into account additional assumptions. One is that the mixing matrices $\mathbf{A}^{(k)}$ are invertible and the other is that the number of source vectors is equivalent to that of observed vectors, i.e., $M = N$.

The demixing model for recovering those vectors is given by

$$\mathbf{y}^{(k)} = \mathbf{W}^{(k)}\mathbf{x}^{(k)}, \tag{4.5}$$

where $\mathbf{y}^{(k)} \triangleq \left[y_1^{(k)}, \ldots .y_N^{(k)}\right]^\top$. This is conceptually the same as in the case of the frequency-domain ICA. Corresponding to Eqs. (4.3) and (4.4), the demixing model can be rewritten as

$$\mathbf{y}_i = \left[y_i^{(1)}, \ldots, y_i^{(K)}\right]^\top = \sum_{j=1}^{M} \mathbf{w}_{ij} \circledast \mathbf{x}_j$$

and

$$\mathbf{y}_i = \sum_{j=1}^{M} \mathbf{W}_{ij} \mathbf{x}_j. \tag{4.6}$$

Define block vectors $\mathbf{x} = \left[\mathbf{x}_1^\top, \ldots, \mathbf{x}_M^\top\right]^\top$, $\mathbf{y} = \left[\mathbf{y}_1^\top, \ldots, \mathbf{y}_N^\top\right]^\top$ and block matrix $\mathbf{W} = \left[\mathbf{W}_{ij}\right]$, then the demixing model can simply be written as $\mathbf{y} = \mathbf{W}\mathbf{x}$. Henceforth, the demixing model or matrix $\mathbf{W}$ will be referred to as a separator.

4.2.2 *Permutation problem in IVA*

ICA inherently has two kinds of indeterminacy: scaling indeterminacy and permutation one. In the case of IVA, we face a more complicated situation: *partial* permutation. Let us look at it in the case of two sources. If a desired separator is successfully obtained, its output will be

$$\begin{aligned} \mathbf{y}_1 &= \left[b_1^{(1)} s_1^{(1)} \quad b_1^{(2)} s_1^{(2)} \quad \cdots \quad b_1^{(K)} s_1^{(K)}\right]^\top, \\ \mathbf{y}_2 &= \left[b_2^{(1)} s_2^{(1)} \quad b_2^{(2)} s_2^{(2)} \quad \cdots \quad b_2^{(K)} s_2^{(K)}\right]^\top \end{aligned} \tag{4.7}$$

or their exchange

$$\begin{aligned} \mathbf{y}_1 &= \left[b_2^{(1)} s_2^{(1)} \quad b_2^{(2)} s_2^{(2)} \quad \cdots \quad b_2^{(K)} s_2^{(K)}\right]^\top, \\ \mathbf{y}_2 &= \left[b_1^{(1)} s_1^{(1)} \quad b_1^{(2)} s_1^{(2)} \quad \cdots \quad b_1^{(K)} s_1^{(K)}\right]^\top. \end{aligned}$$

Since the labeling of the source vectors is essentially arbitrary, we may use the natural expression (4.7) without loss of generality. Because of indeterminacy in the scaling of source signals, constants $b_i^{(k)}$ depend upon the IVA algorithm adopted. In order to make simple the representation of the output of the desired separator, we redefine $b_i^{(k)} s_i^{(k)}$ by $s_i^{(k)}$ and write Eq. (4.7) simply as

$$\begin{aligned} \mathbf{y}_1 &= \left[s_1^{(1)} \quad s_1^{(2)} \quad \cdots \quad s_1^{(K)}\right]^\top \triangleq \mathbf{s}_1, \\ \mathbf{y}_2 &= \left[s_2^{(1)} \quad s_2^{(2)} \quad \cdots \quad s_2^{(K)}\right]^\top \triangleq \mathbf{s}_2. \end{aligned} \tag{4.8}$$

In this chapter, however, we investigate what will happen when some *partial* permutations occur among $\mathbf{y}_1, \ldots, \mathbf{y}_N$. Let us consider the case that permutation occurs solely between $y_1^{(1)}$ and $y_2^{(1)}$; then we have

$$\begin{aligned} \mathbf{y}_1 &= \left[c_1^{(1)} \underline{s_2^{(1)}} \quad c_1^{(2)} s_1^{(2)} \quad \cdots \quad c_1^{(K)} s_1^{(K)} \right]^\top, \\ \mathbf{y}_2 &= \left[c_2^{(1)} \underline{s_1^{(1)}} \quad c_2^{(2)} s_2^{(2)} \quad \cdots \quad c_2^{(K)} s_2^{(K)} \right]^\top. \end{aligned} \tag{4.9}$$

Henceforth, we refer to a separator bringing about any kind of permutation as a *permuted separator*. A cumbersome problem associated with a permuted separator is that the magnitudes of the source signals in $y_i^{(k)}$ $(k = 1, \ldots, K)$ become all different from those of the desired separator; namely, $c_i^{(k)} \neq 1$ $(k = 1, \ldots, K)$. This issue makes the representation of the output of the separator quite complicated. In ordinary IVA algorithms, however, if the number of permuted components in $\mathbf{y}_i$ is relatively small compared to the dimension K of the vectors, the coefficients $c_i^{(k)}$ of the *unpermuted* entries do not change so much, i.e. $c_i^{(k)} \approx 1$ $(k \neq 1)$. Thus, we can write Eq. (4.9) as

$$\begin{aligned} \mathbf{y}_1 &= \left[\tilde{s}_1^{(1)} \quad s_1^{(2)} \quad \cdots \quad s_1^{(K)} \right]^\top, \\ \mathbf{y}_2 &= \left[\tilde{s}_2^{(1)} \quad s_2^{(2)} \quad \cdots \quad s_2^{(K)} \right]^\top, \end{aligned} \tag{4.10}$$

where $\tilde{s}_1^{(1)} = c_1^{(1)} s_2^{(1)}$ and $\tilde{s}_2^{(1)} = c_2^{(1)} s_1^{(1)}$. In general, if $y_i^{(k)}$ is not originated from source $\mathbf{s}_i$, then we write it as $\tilde{s}_i^{(k)}$. It will be convenient to consider an imaginary output $s_i^{(k)}$ even in the case of $y_i^{(k)} = \tilde{s}_i^{(k)}$; it is the output $y_i^{(k)}$ that would be provided by the desired separator. Obviously, $\tilde{s}_i^{(k)}$ is statistically independent of $s_i^{(l)}$ $(l = 1, \ldots, K)$.

Later analyses will evaluate such a statistical expectation as $E\left(\|\mathbf{y}_i\|^\alpha y_i^{(k)2} \right)$. For the desired separator, $E\left(\|\mathbf{y}_i\|^\alpha y_i^{(k)2} \right)$ becomes $E\left(\|\mathbf{s}_i\|^\alpha s_i^{(k)2} \right)$ because of their statistical independence. If some permutation occurs and sources $s_i^{(m_1)}, \ldots, s_i^{(m_L)}$ in $\mathbf{y}_i$ are replaced by $\tilde{s}_i^{(m_1)}, \ldots, \tilde{s}_i^{(m_L)}$, then we will have

$$E\left(\|\mathbf{y}_i\|^\alpha y_i^{(k)2} \right) = \begin{cases} E\left(\|\mathbf{y}_i\|^\alpha \tilde{s}_i^{(k)2} \right) & \text{for } k = m_1, \ldots, m_L, \\ E\left(\|\mathbf{y}_i\|^\alpha s_i^{(k)2} \right) & \text{for } k \neq m_1, \ldots, m_L, \end{cases} \tag{4.11}$$

where

$$\|\mathbf{y}_i\| = \left(\sum_{l \neq m_1, \ldots, m_L} s_i^{(l)2} + \sum_{l = m_1, \ldots, m_L} \tilde{s}_i^{(l)2} \right)^{\frac{1}{2}}.$$

In this chapter, we only consider the case where the number L of the permuted entries is very small compared to the dimension K of the source vectors. Then, the contribution of $\tilde{s}_i^{(m_1)}, \ldots, \tilde{s}_i^{(m_L)}$ to $\|\mathbf{y}_i\|$ is very small and hence $\|\mathbf{y}_i\|$ is almost equal to $\|\mathbf{s}_i\|$. Thus, Eq. (4.11) can be approximated as

$$E\left(\|\mathbf{y}_i\|^{\alpha} y_i^{(k)2}\right) = \begin{cases} E\left(\|\mathbf{s}_i\|^{\alpha}\right) E\left(\tilde{s}_i^{(k)2}\right) & \text{for } k = m_1, \ldots, m_L, \\ E\left(\|\mathbf{s}_i\|^{\alpha} s_i^{(k)2}\right) & \text{for } k \neq m_1, \ldots, m_L. \end{cases} \tag{4.12}$$

This kind of approximation will be made repeatedly throughout this chapter.

4.2.3 *Function for evaluating independence among the sources*

Let a *prior* pdf of source $\mathbf{s}_i$ be $q(\mathbf{u}) = q\left(u^{(1)}, \ldots, u^{(K)}\right)$, where $\mathbf{u}$ is dummy vector; although, in general, $q(\mathbf{u})$ should be written as $q_{\mathbf{s}_i}(\mathbf{u})$, we omit the subscript for the sake of simplicity of notation and also because a common model is usually used for all sources in actual applications. On the other hand, let the *actual* pdf of $\mathbf{y}$ be $p_{\mathbf{y}}(\mathbf{u}_1, \ldots, \mathbf{u}_N)$. A most popular approach to find the desired separator is to minimize the Kullback-Leibler divergence between $p_{\mathbf{y}}(\mathbf{u}_1, \ldots, \mathbf{u}_N)$ and $\Pi_{i=1}^{N} q(\mathbf{u}_i)$, which is

$$\begin{aligned} Q(\mathbf{W}) &= \int_{\mathbb{R}^{NK}} p_{\mathbf{y}}(\mathbf{u}_1, \ldots, \mathbf{u}_N) \log \frac{p_{\mathbf{y}}(\mathbf{u}_1, \ldots, \mathbf{u}_N)}{\Pi_{i=1}^{N} q(\mathbf{u}_i)} d\mathbf{u}_1 \ldots d\mathbf{u}_N \\ &= -\log|\det \mathbf{W}| + \sum_{i=1}^{N} E\left(-\log q(\mathbf{y}_i)\right) + const. \end{aligned} \tag{4.13}$$

The choice of a prior pdf depends on the prior knowledge about the nature of sources of interest (e.g., non-Gaussianity). While ICA usually consider only the non-Gaussianity of sources, IVA must consider not only the non-Gaussianity of sources but also the inner dependency of vectors. Some researches observed that speech has the property of spherically invariance, and then they introduced spherically invariant random process for band-limited speech [12, 13]. The frequency components models of speech that have been used in IVA have similar property, i.e., spherically symmetric [35, 41]. In this thesis, we adopt spherically-symmertic, generalized-Gaussian model:

$$q(\mathbf{u}) = \alpha \cdot \exp\left(-\frac{1}{\beta}\|\mathbf{u}\|^{\beta}\right) \quad (0 < \beta < 2), \tag{4.14}$$

where α is a normalization term. The restriction of $0 < \beta < 2$ comes from the fact that we are solely interested in separation of the so-called super-Gaussian signals such as speech signals. For sub-Gaussian signals the spherical function might not be appropriate, but the issue is beyond the scope of this thesis.

For better understanding of the property of spherically symmetric, the density and contour plots of two-dimensional spherically-symmetric, generalized Gaussian model and independent Laplace

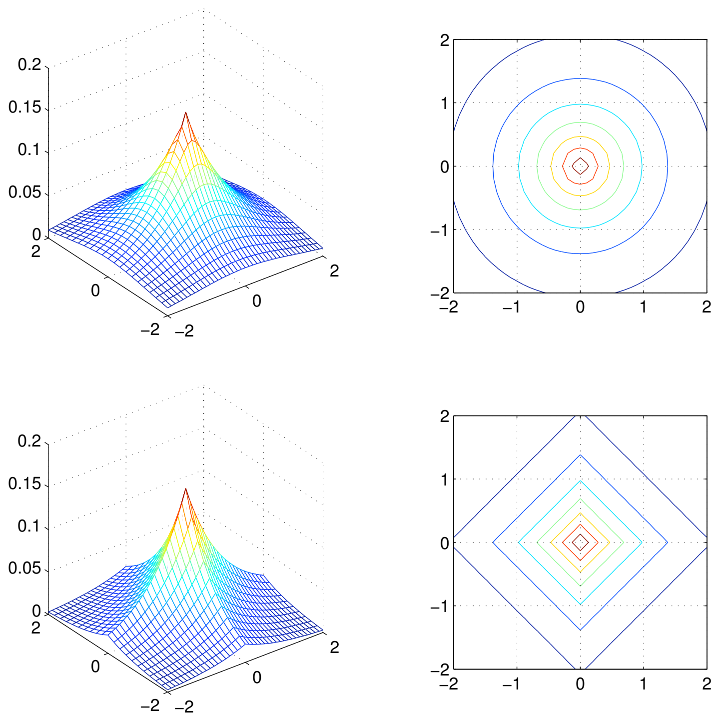

Figure 6: Two-dimensional density plot and the contour plot of spherically-symmetric, generalized-Gaussian model and independent Laplace distribution. *Top*: two-dimensional density and contour plots of spherically-symmetric, generalized-Gaussian model ($\beta = 1$). If $\beta = 1$, the model is known as spherically symmetric Laplace (SSL) distribution [22]. *Bottom*: two-dimensional density and contour plots of *independent* Laplace distribution.

distribtuion are illustrated in Fig. 6. The variance of each marginal density is set to be one.

If the model $q(\mathbf{u})$ is appropriately chosen, then minimization of function $Q(\mathbf{W})$ will provides a desired separator $\mathbf{W}_{\text{opt}}$. For any form of $q(\mathbf{u})$, $\mathbf{W}_{\text{opt}}$ is a stationary point of $Q(\mathbf{W})$. Therefore, what we have to do is to derive some conditions for β so that the second-order differential $d^2Q(\mathbf{W}_{\text{opt}})$ be positive definite; the condition depends on the pdf of the source vectors, of course. On the other hand, we will also investigate instability of a permuted separator $\tilde{\mathbf{W}}$. We can show that $\tilde{\mathbf{W}}$ is also a stationary point of $Q(\mathbf{W})$. Hence, in order to prevent the undesirable permutation, we need to determine β such that $d^2Q(\tilde{\mathbf{W}})$ be *not* positive definite. The main objective is to prove the following proposition:

Proposition 3. *if $q(\mathbf{u})$ or β is chosen such that the desired separator can be obtained by minimizing $Q(\mathbf{W})$, then any permuted separator cannot be a local minimum of $Q(\mathbf{W})$.*

This proposition shows the effectiveness of the IVA approach in a definite manner.

4.2.4 *First and second-order differentials of* $-\log q(\mathbf{y}_i)$

In later analyses we will need the first and second-order differentials of $-\log q(\mathbf{y}_i)$. They are

$$-d\log q(\mathbf{y}_i) = d\mathbf{y}_i^\top \frac{1}{q(\mathbf{y}_i)}\frac{\partial q}{\partial \mathbf{y}_i}(\mathbf{y}_i) = d\mathbf{y}_i^\top \boldsymbol{\varphi}(\mathbf{y}_i), \tag{4.15}$$

and

$$\begin{aligned} -d^2\log q(\mathbf{y}_i) &= d^2\mathbf{y}_i^\top \boldsymbol{\varphi}(\mathbf{y}_i) + d\mathbf{y}_i^\top \frac{\partial \boldsymbol{\varphi}}{\partial \mathbf{y}_i^\top}(\mathbf{y}_i)\, d\mathbf{y}_i \\ &= d^2\mathbf{y}_i^\top \boldsymbol{\varphi}(\mathbf{y}_i) + d\mathbf{y}_i^\top \boldsymbol{\Psi}(\mathbf{y}_i)\, d\mathbf{y}_i. \end{aligned} \tag{4.16}$$

The differentials are induced by the perturbation $d\mathbf{W}$ of $\mathbf{W}$. We put $d\mathbf{V} = [d\mathbf{V}_{ij}] = d\mathbf{W}\mathbf{W}^{-1}$, where $d\mathbf{V}_{ij} = \mathrm{diag}\{dv_{ij}^{(1)}, \ldots, dv_{ij}^{(K)}\}$. Then, $d\mathbf{y}_i$ and $d^2\mathbf{y}_i$ are expressed as

$$d\mathbf{y}_i = \sum_{j=1}^{N} d\mathbf{V}_{ij}\mathbf{y}_j$$

and

$$d^2\mathbf{y}_i = \sum_{j=1}^{N}\sum_{k=1}^{N} d\mathbf{V}_{ij} d\mathbf{V}_{jk}\mathbf{y}_k,$$

respectively. Substituting them into (4.15) and (4.16) leads to

$$-d\log q(\mathbf{y}_i) \;=\; \sum_{j=1}^{N} \mathbf{y}_j^\top d\mathbf{V}_{ij}\boldsymbol{\varphi}(\mathbf{y}_i), \tag{4.17}$$

and

$$\begin{aligned} -d^2\log q(\mathbf{y}_i) = &\sum_{k=1}^{N}\sum_{j=1}^{N} \mathbf{y}_k^\top d\mathbf{V}_{jk} d\mathbf{V}_{ij}\boldsymbol{\varphi}(\mathbf{y}_i) \\ &+ \sum_{j=1}^{N}\sum_{k=1}^{N} \mathbf{y}_j^\top d\mathbf{V}_{ij}\boldsymbol{\Psi}(\mathbf{y}_i)\, d\mathbf{V}_{ik}\mathbf{y}_k. \end{aligned} \tag{4.18}$$

Here we introduce another expression of the above equations, which will be useful when evaluating their expectation. Corresponding to *vectors* $\mathbf{y}_i$, $\boldsymbol{\varphi}(\mathbf{y}_i)$ and *diagonal matrix* $d\mathbf{V}_{ij}$, define *diagonal matrices* $\mathbf{Y}_i = \mathrm{diag}\{y_i^{(1)}, \ldots, y_i^{(K)}\}$, $\boldsymbol{\Phi}(\mathbf{y}_i) = \mathrm{diag}\left\{\varphi^{(1)}(\mathbf{y}_i), \ldots, \varphi^{(K)}(\mathbf{y}_i)\right\}$ and *vector* $d\mathbf{v}_{ij} = \left[dv_{ij}^{(1)}, \ldots, dv_{ij}^{(K)}\right]^\top$, then we have $d\mathbf{V}_{ij}\mathbf{y}_j = \mathbf{Y}_j d\mathbf{v}_{ij}$. Equations (4.17) and (4.18) can then be rewritten as

$$-d\log q(\mathbf{y}_i) = \sum_{j=1}^{N} d\mathbf{v}_{ij}^\top \mathbf{Y}_j \boldsymbol{\varphi}(\mathbf{y}_i) \tag{4.19}$$

and

$$
\begin{aligned}
-d^2 \log q\left(\mathbf{y}_i\right) = & \sum_{k=1}^{N} \sum_{j=1}^{N} d\mathbf{v}_{jk}^{\top} \mathbf{Y}_k \mathbf{\Phi}\left(\mathbf{y}_i\right) d\mathbf{v}_{ij} \\
& + \sum_{j=1}^{N} \sum_{k=1}^{N} d\mathbf{v}_{ij}^{\top} \mathbf{Y}_j \mathbf{\Psi}\left(\mathbf{y}_i\right) \mathbf{Y}_k d\mathbf{v}_{ik}.
\end{aligned} \tag{4.20}
$$

The key point of the above rewriting from (4.17) and (4.18) to (4.19) and (4.20) is that, while the random variables are separated by $d\mathbf{V}_{ij}$ in the former expression, they are aggregated in the latter expression. This rewriting will enable an easier expectation calculation of a product of some random variables.

4.3 STABILITY ANALYSES OF THE DESIRED SEPARATOR

4.3.1 *The first-order differential of $Q(\mathbf{W})$ at $\mathbf{W} = \mathbf{W}_{opt}$*

The differential of the first term in the right-hand side of (4.13) proves to be

$$
\begin{aligned}
-d \log |\det \mathbf{W}| &= -\mathrm{tr} d\mathbf{W}\mathbf{W}^{-1} = \mathrm{tr} d\mathbf{V} \\
&= -\sum_{i=1}^{N} \mathrm{tr} d\mathbf{V}_{ii} = -\sum_{i=1}^{N} d\mathbf{v}_{ii}^{\top} \mathbf{e},
\end{aligned} \tag{4.21}
$$

where $\mathbf{e} = [1, \ldots, 1]^{\top}$. Obviously the second-order differential of $-\log|\det \mathbf{W}|$ with respect to $d\mathbf{V}$ vanishes.

According to Eq. (4.19), the first-order differential of the second term of (4.13) is

$$
dE\left(-\log q\left(\mathbf{y}_i\right)\right) = \sum_{j=1}^{N} d\mathbf{v}_{ij}^{\top} E\left(\mathbf{Y}_j \boldsymbol{\varphi}\left(\mathbf{y}_i\right)\right). \tag{4.22}
$$

When $\mathbf{W} = \mathbf{W}_{\text{opt}}$, i.e., $\mathbf{y}_i = \mathbf{s}_i$ and hence $\mathbf{Y}_i = \mathbf{S}_i \triangleq \mathrm{diag}\left\{s_i^{(1)}, \ldots, s_i^{(K)}\right\}$, the above-mentioned equation becomes

$$
dE\left(-\log q\left(\mathbf{y}_i\right)\right) = d\mathbf{v}_{ii}^{\top} E\left(\mathbf{S}_i \boldsymbol{\varphi}\left(\mathbf{s}_i\right)\right). \tag{4.23}
$$

In the above derivation we have used relation $E\left(\mathbf{S}_j \boldsymbol{\varphi}\left(\mathbf{s}_i\right)\right) = E\left(\mathbf{S}_j\right) E\left(\boldsymbol{\varphi}\left(\mathbf{s}_i\right)\right) = \mathbf{0}\ (j \neq i)$, where $\mathbf{0}$ stands for the zero column vector; this kind of explanation associated with independence among $\mathbf{s}_1, \ldots, \mathbf{s}_N$ will be omitted hereafter. Thus, combining (4.21) and (4.23) leads to the first-order total differential of $Q(\mathbf{W})$ at $\mathbf{W}_{\text{opt}}$:

$$
\begin{aligned}
dQ\left(\mathbf{W}_{\text{opt}}\right) &= \sum_{i=1}^{N} d\mathbf{v}_{ii}^{\top} \left\{E\left(\mathbf{S}_i \boldsymbol{\varphi}\left(\mathbf{s}_i\right)\right) - \mathbf{e}\right\} \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} dv_{ii}^{(k)} \left\{E\left(\varphi\left(\mathbf{s}_i\right) s_i^{(k)}\right) - 1\right\}.
\end{aligned} \tag{4.24}
$$

The fact that $dQ\left(\mathbf{W}_{\text{opt}}\right) = 0$ is for any values of $dv_{ii}^{(k)}$ implies that the desired separator obtained by minimizing $Q\left(\mathbf{W}\right)$ determines the scaling of the source signals as

$$E\left(\varphi^{(k)}\left(\mathbf{s}_i\right) s_i^{(k)}\right) = 1 \;\; (i = 1, \ldots, N; k = 1, \ldots, K) \tag{4.25}$$

or equivalently

$$E\left(\mathbf{\Phi}\left(\mathbf{s}_i\right) \mathbf{S}_i\right) = \mathbf{I} \;\; (i = 1, \ldots, N). \tag{4.26}$$

There are important equations which will be used repeatedly.

4.3.2 *The second-order differential of $Q\left(\mathbf{W}\right)$ at $\mathbf{W} = \mathbf{W}_{opt}$*

The expectation of $d^2\left(-\log q\left(\mathbf{y}_i\right)\right)$ is

$$\begin{aligned} d^2 E\left(-\log q\left(\mathbf{y}_i\right)\right) = & \sum_{j=1}^{N} \sum_{k=1}^{N} d\mathbf{v}_{ij}^{\top} E\left(\mathbf{Y}_j \mathbf{\Psi}\left(\mathbf{y}_i\right) \mathbf{Y}_k\right) d\mathbf{v}_{ik} \\ & + \sum_{j=1}^{N} \sum_{k=1}^{N} d\mathbf{v}_{jk}^{\top} E\left(\mathbf{Y}_k \mathbf{\Phi}\left(\mathbf{y}_i\right)\right) d\mathbf{v}_{ij}, \end{aligned} \tag{4.27}$$

and when $\mathbf{W} = \mathbf{W}_{\text{opt}}$ or $\mathbf{y}_i = \mathbf{s}_i$ (i.e., $\mathbf{Y}_i = \mathbf{S}_i$), it is

$$\begin{aligned} d^2 E\left(-\log q\left(\mathbf{y}_i\right)\right) &= \sum_{j=1}^{N} \left\{ d\mathbf{v}_{ij}^{\top} E\left(\mathbf{S}_j \mathbf{\Psi}\left(\mathbf{s}_i\right) \mathbf{S}_j\right) d\mathbf{v}_{ij} + d\mathbf{v}_{ji}^{\top} d\mathbf{v}_{ij} \right\} \\ &= \sum_{j=1}^{N} \left\{ d\mathbf{v}_{ij}^{\top} E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right) \circledast \left(\mathbf{s}_j \mathbf{s}_j^{\top}\right)\right) d\mathbf{v}_{ij} + d\mathbf{v}_{ji}^{\top} d\mathbf{v}_{ij} \right\}, \end{aligned} \tag{4.28}$$

where symbol $\circledast$ stands for the Hadamard product. We derive the above summation into terms with $j \neq i$ and those with $j = i$:

$$\sum_{\substack{j=1 \\ j \neq i}}^{N} \left\{ d\mathbf{v}_{ij}^{\top} E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right)\right) \circledast E\left(\mathbf{s}_j \mathbf{s}_j^{\top}\right) d\mathbf{v}_{ij} + d\mathbf{v}_{ji}^{\top} d\mathbf{v}_{ij} \right\} \tag{4.29}$$

and

$$d\mathbf{v}_{ii}^{\top} \left\{ \mathbf{I} + E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right) \circledast \left(\mathbf{s}_i \mathbf{s}_i^{\top}\right)\right) \right\} d\mathbf{v}_{ii}. \tag{4.30}$$

Combining (4.29) and (4.30) and summing them with respect to i, we obtain

$$\begin{aligned} d^2 Q\left(\mathbf{W}_{\text{opt}}\right) = & \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left\{ d\mathbf{v}_{ij}^{\top} \mathbf{\Gamma}_{ij} d\mathbf{v}_{ij} + d\mathbf{v}_{ji}^{\top} d\mathbf{v}_{ij} \right\} \\ & + \sum_{i=1}^{N} d\mathbf{v}_{ii}^{\top} \left\{ \mathbf{I} + \mathbf{M}_i \right\} d\mathbf{v}_{ii}, \end{aligned} \tag{4.31}$$

where

$$\mathbf{\Gamma}_{ij} \triangleq \left[\gamma_{ij}^{(k,l)}\right] = E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right)\right) \circledast E\left(\mathbf{s}_j \mathbf{s}_j^{\top}\right), \tag{4.32}$$

$$\mathbf{M}_i \triangleq E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right) \circledast \left(\mathbf{s}_i \mathbf{s}_i^{\top}\right)\right). \tag{4.33}$$

Equation (4.31) implies that $d^2Q\left(\mathbf{W}_{\text{opt}}\right)$ can be decomposed into $N\left(N+1\right)/2$ independent sets of quadratic forms composed of two or one vector: $N\left(N-1\right)/2$ sets which each comprises a pair of $d\mathbf{v}_{ij}$ and $d\mathbf{v}_{ji}$ $(i \neq j)$, and N sets which each involves $d\mathbf{v}_{ii}$.

4.3.3 *Stability of the desired separator*

In order to obtain the desired separator $\mathbf{W}_{\text{opt}}$ by minimizing $Q\left(\mathbf{W}\right)$, the second-order differential $d^2Q\left(\mathbf{W}\right)$ must be positive definite at $\mathbf{W}_{\text{opt}}$. A *necessary and sufficient condition* for that is

(A) Matrix $\begin{bmatrix} \mathbf{\Gamma}_{ij} & \mathbf{I} \\ \mathbf{I} & \mathbf{\Gamma}_{ji} \end{bmatrix}$ is positive definite for every pair of i and $j\left(\neq i\right)$, and

(B) matrix $\mathbf{I}+\mathbf{M}_i$ is positive definite for every i.

We refer to this pair of conditions as the *stability* condition for the desired separator.

We here also describe a *necessary* condition for the stability, which will play an important role in the *instability* analysis for a permuted separator. From (4.31), we extract three terms associated with $dv_{ij}^{(k)}dv_{ji}^{(k)}$, $dv_{ij}^{(k)2}$ and $dv_{ji}^{(k)2}$ $(i \neq j)$:

$$\begin{aligned}&\gamma_{ij}^{(k,k)}dv_{ij}^{(k)2}+2dv_{ij}^{(k)}dv_{ji}^{(k)}+\gamma_{ji}^{(k,k)}dv_{ji}^{(k)2}\\&=E\left(\psi^{(k,k)}\left(\mathbf{s}_i\right)\right)E\left(s_j^{(k)2}\right)dv_{ij}^{(k)2}+2dv_{ij}^{(k)}dv_{ji}^{(k)}\\&+E\left(\psi^{(k,k)}\left(\mathbf{s}_j\right)\right)E\left(s_i^{(k)2}\right)dv_{ji}^{(k)2}.\end{aligned} \tag{4.34}$$

In order for $d^2Q\left(\mathbf{W}_{\text{opt}}\right)$ to be positive definite, the above quadratic form must be positive definite. Thus, we obtain a *necessary* condition for the stability:

(A1) $\zeta_i^{(k,k)} > 0$ and $\zeta_i^{(k,k)}\sigma_i^{(k)2}\cdot\zeta_j^{(k,k)}\sigma_j^{(k)2} > 1$ for every $i, j\left(\neq i\right)$ and k, where

$$\zeta_i^{(k,k)} \triangleq E\left(\psi^{(k,k)}\left(\mathbf{s}_i\right)\right), \tag{4.35}$$

$$\sigma_i^{(k)2} \triangleq E\left(s_i^{(k)2}\right). \tag{4.36}$$

Since $q\left(\mathbf{u}\right)$ is spherically symmetric generalized Gaussian model, $\boldsymbol{\varphi}\left(\mathbf{y}_i\right)$ and $\mathbf{\Psi}\left(\mathbf{y}_i\right)$ become

$$\boldsymbol{\varphi}\left(\mathbf{y}_i\right)=\mathbf{y}_i\|\mathbf{y}_i\|^{\beta_2}, \tag{4.37}$$

$$\mathbf{\Psi}\left(\mathbf{y}_i\right)=\|\mathbf{y}_i\|^{\beta_2}\mathbf{I}+\|\mathbf{y}_i\|^{\beta_4}\mathbf{y}_i\mathbf{y}_i^\top, \tag{4.38}$$

where $\beta_2 \triangleq \beta - 2$ and $\beta_4 \triangleq \beta - 4$. Hence, we have

$$\begin{aligned} \zeta_i^{(k,k)} &= E\left(\|\mathbf{s}_i\|^{\beta_2}\right) + E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2}\right) \\ &= \left\{1 + \frac{E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)}\right\} E\left(\|\mathbf{s}_i\|_2^{\beta_2}\right), \end{aligned} \tag{4.39}$$

Hence, a necessary condition for the stability can be rewritten as

(A2) $\eta_i^{(k)} > 0$ and $\eta_i^{(k)} \xi_i^{(k)} \cdot \eta_j^{(k)} \xi_j^{(k)} > 1$ for every $i, j (\neq i)$ and k, where

$$\eta_i^{(k)} \triangleq 1 + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)}, \tag{4.40}$$

$$\xi_i^{(k)} \triangleq E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(s_i^{(k)2}\right). \tag{4.41}$$

Index $\xi_i^{(k)}$ can be rewritten as

$$\xi_i^{(k)} = \frac{E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(s_i^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2} s_i^{(k)2}\right)}, \tag{4.42}$$

because

$$E\left(\|\mathbf{s}_i\|^{\beta_2} s_i^{(k)2}\right) = 1 \tag{4.43}$$

holds due to $E\left(\varphi^{(k)}(\mathbf{s}_i)\, s_i^{(k)}\right) = 1$. $\xi_i^{(k)}$ represented in the above form is scale-invariant in the sense that its value does not change if vector $\mathbf{s}_i$ is arbitrarily scaled. Since $\eta_i^{(k)}$ defined by (4.40) is also scale-invariant, condition (A2) with (4.40) and (4.42) (not (4.41)) is scale-invariant. Note that parameters $\eta_i^{(k)}$ and $\xi_i^{(k)}$ $(k = 1, \ldots, K)$ are determined independently of other sources than $\mathbf{s}_i$.

According to definition (4.40), inequality $\eta_i^{(k)} < 1$ holds, because $\beta_2 < 0$. Combining this and condition (A2), we have

$$0 < \eta_i^{(k)} < 1 \ (i = 1, \ldots, N; k = 1, \ldots, K). \tag{4.44}$$

Also note that, since (A2) must be satisfied for every pair of i and $j (\neq i)$, inequality

$$\eta_i^{(k)} \xi_i^{(k)} > 1 \tag{4.45}$$

must hold at least for $N - 1$ components in $s_1^{(k)}, \ldots, s_N^{(k)}$; in other words, the opposite inequality $\eta_i^{(k)} \xi_i^{(k)} \leq 1$ is allowed only for one of them. Inequalities (4.44) and (4.45) will play an important role in the *instability* analysis for permuted separators.

4.3.4 *Uncorrelatedness of each source vector*

If, moreover, $\mathbf{s}_1^{(1)}, \ldots, \mathbf{s}_N^{(K)}$ are uncorrelated with each other, i.e.,

$$E\left(\mathbf{s}_i \mathbf{s}_i^\top\right) = E\left(\mathbf{S}_i^2\right), \tag{4.46}$$

then condition (A1) with (B) becomes also a *sufficient* condition for the stability, as shown below. In this case, $\mathbf{\Gamma}_{ij}$ becomes a diagonal matrix as

$$\begin{aligned}\mathbf{\Gamma}_{ij} &= E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right)\right) \circledast E\left(\mathbf{S}_j^2\right) \\ &= \mathrm{diag}\left\{\zeta_i^{(1,1)}, \ldots, \zeta_i^{(K,K)}\right\} E\left(\mathbf{S}_j^2\right) \\ &= \mathrm{diag}\left\{\zeta_i^{(1,1)} \sigma_j^{(1)2}, \ldots, \zeta_i^{(K,K)} \sigma_j^{(K)2}\right\}.\end{aligned} \tag{4.47}$$

Note that $E\left(\mathbf{S}_j^2\right)$ is a diagonal matrix and $\circledast$ denotes the Hadamard product; only the diagonal part of $E\left(\mathbf{\Psi}\left(\mathbf{s}_i\right)\right)$ is therefore significant. Thus, we have

$$\begin{aligned}d^2 Q\left(\mathbf{W}_{\mathrm{opt}}\right) = &\sum_{i=1}^{N} \sum_{j \neq i} d\mathbf{v}_{ij}^\top \mathrm{diag}\left\{\zeta_i^{(1,1)} \sigma_j^{(1)2}, \ldots, \zeta_i^{(K,K)} \sigma_j^{(K)2}\right\} d\mathbf{v}_{ij} \\ &+ \sum_{i=1}^{N} \sum_{j \neq i} d\mathbf{v}_{ji}^\top d\mathbf{v}_{ij} + \sum_{i=1}^{N} d\mathbf{v}_{ii}^\top \left\{\mathbf{I}_K + \mathbf{M}_i\right\} d\mathbf{v}_{ii}.\end{aligned} \tag{4.48}$$

Since $\mathrm{diag}\left\{\zeta_i^{(1,1)} \sigma_j^{(1)2}, \ldots, \zeta_i^{(K,K)} \sigma_j^{(K)2}\right\}$ is diagonal, we can easily obtain a *necessary and sufficient* condition for stability, which is (A1) together with (B).

Although the uncorrelatedness among the source components appears to be a rather strong assumption, it is quite valid in the case of frequency-domain signal separation of stationary sources, in which each sound is treated as a set of uncorrelated frequency components [52]. However, instability of permuted separators will provide without the assumption of uncorrelatedness.

4.3.5 *Evenness among the components in each source*

In usual ICA (i.e. the case of $K = 1$), whether an algorithm adopted works well or not depends crucially on the degree of non-Gaussianity of the sources. In IVA (i.e. the case of $K \geq 2$), there arises another important aspect that should be considered: a certain uniformity among the components in each source. We write the joint pdf of $s_i^{(1)2}, \ldots s_i^{(K)2}$ as $p_{\mathbf{s}_i^2}\left(u^{(1)}, \ldots, u^{(K)}\right)$. If function $p_{\mathbf{s}_i^2}$ is invariant to any permutation of the arguments, then we say that source $\mathbf{s}_i$ is *evenly* distributed.

Note that, in our framework, the evenness among $s_i^{(1)2}, \ldots s_i^{(K)2}$ should be defined for a vector whose components are normalized as $E\left(\|\mathbf{s}_i\|^{\beta_2} s_i^{(k)2}\right) = 1$ (or $E\left(\varphi^{(k)}(\mathbf{s}_i)\, s_i^{(k)}\right) = 1$). Under the evenness condition, we have

$$E\left(s_i^{(k)2}\right) = \frac{E\left(\|\mathbf{s}_i\|^2\right)}{K}, \tag{4.49}$$

$$E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2}\right) = \frac{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)}{K} \tag{4.50}$$

for every k, because $\sum_{k=1}^{K} E\left(s_i^{(k)2}\right) = E\left(\|\mathbf{s}_i\|^2\right)$ and $\sum_{k=1}^{K} E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2}\right) = E\left(\|\mathbf{s}_i\|^{\beta_2}\right)$. Note that evenness does not require that the source vector be *spherically* distributed.

For evenly distributed sources, the stability condition derived in the last subsection can be expressed in a considerably simple form. Indices $\eta_i^{(k)}$ and $\xi_i^{(k)}$ are then

$$\eta = 1 + \frac{\beta}{K}, \tag{4.51}$$

$$\xi_i = \frac{E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(\|\mathbf{s}_i\|^2\right)}{E\left(\|\mathbf{s}_i\|^{\beta}\right)}. \tag{4.52}$$

Since $-2 < \beta_2 < 0$ and $K \geq 2$, index η takes a value between 0 and 1. Stability condition (A2) thus reduces to

(A3) $\eta^2 \xi_i \xi_j > 1$ for every pair of i and j ($\neq i$).

Here, we define the inverse of $\eta\xi_i$:

$$\kappa_i \triangleq \frac{1}{\eta\xi_i} = \frac{E\left(\|\mathbf{s}_i\|^{\beta}\right)}{\left(1 + \frac{\beta_2}{K}\right) E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(\|\mathbf{s}_i\|^2\right)}. \tag{4.53}$$

Then, (A3) is rewritten as

(A4) $\kappa_i \kappa_j < 1$ for every pair of i and j ($\neq i$).

If source $\mathbf{s}_i$ is Gaussian, then index κ_i takes the value of unity (see below). It can be considered a measure of non-Gaussianity of a source: a generalized kurtosis of a random *vector*.

Next, let us investigate condition (B): positive definiteness of $\mathbf{I} + \mathbf{M}_i$. Since $q(\mathbf{u})$ is given as (4.14), $\mathbf{M}_i$ becomes

$$\mathbf{M}_i = \mathbf{I} + \beta_2 E\left(\|\mathbf{s}_i\|^{\beta_4}\left(\mathbf{s}_i\mathbf{s}_i^\top\right) \circledast \left(\mathbf{s}_i\mathbf{s}_i^\top\right)\right). \tag{4.54}$$

Define

$$\mathbf{C}_i = E\left(\|\mathbf{s}_i\|^{\beta_4}\left(\mathbf{s}_i\mathbf{s}_i^\top\right) \circledast \left(\mathbf{s}_i\mathbf{s}_i^\top\right)\right) = \left[E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2} s_i^{(l)2}\right)\right]. \tag{4.55}$$

It is easy to show that $\mathbf{C}_i\mathbf{e} = \mathbf{e}$ holds, implying that an eigenvalue of $\mathbf{C}_i$ is 1. Under the evenness condition, we can moreover show that the *maximum* eigenvalues $\mathbf{C}_i$ is 1, because $\mathbf{C}_i$ takes a simple form as

$$\mathbf{C}_i = \begin{bmatrix} 1-(K-1)\,c & c & \cdots & c \\ c & 1-(K-1)\,c & \ddots & \vdots \\ \vdots & \ddots & \ddots & c \\ c & \cdots & c & 1-(K-1)\,c \end{bmatrix}. \qquad (4.56)$$

The *minimum* eigenvalue of $\mathbf{I} + \mathbf{M}_i = 2\mathbf{I} + \beta_2\mathbf{C}_i$ is therefore $2 + \beta_2 = \beta\ (> 0)$ and hence $\mathbf{I} + \mathbf{M}_i$ proves to be positive definite. This implies that, if every source vector is evenly distributed, condition (B) holds automatically and hence it is unnecessary.

The *target* model $q(\mathbf{u})$ dealt with in this thesis is spherical and hence even. Because it is difficult to know the distribution of a source vector beforehand, we usually have no choice but to adopt an evenly distributed model for $q(\mathbf{u})$. The degree of evenness of each source is therefore very important. If the distribution deviates far from evenness, it might become impossible to choose β so as to satisfy the stability condition.

Let us again investigate positive definiteness of $\mathbf{I} + \mathbf{M}_i$, but without the assumption of evenness. The maximum eigenvalue of $\mathbf{C}_i$ can be evaluated by

$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^\top \mathbf{C}_i \mathbf{x} = \max_{\|\mathbf{x}\|=1} E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2} s_i^{(l)2}\right) x_k x_l. \qquad (4.57)$$

The summation in the right-hand side of (4.57) can be written as

$$\sum_{k=1}^{K}\sum_{l=1}^{K} \|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2} s_i^{(l)2} x_k x_l = \left(\sum_{k=1}^{K} \|\mathbf{s}_i\|^{-2} s_i^{(k)2} x_k\right)\left(\sum_{l=1}^{K} \|\mathbf{s}_i\|^{\beta_2} s_i^{(l)2} x_l\right).$$

Since $\left|\sum_{k=1}^{K} \|\mathbf{s}_i\|^{-2} s_i^{(k)2} x_k\right| \le \max_{k=1}^{K} |x_k|$, we have

$$\sum_{k=1}^{K}\sum_{l=1}^{K} E\left(\|\mathbf{s}_i\|^{\beta_4} s_i^{(k)2} s_i^{(l)2}\right) x_k x_l \le E\left(\left|\sum_{k=1}^{K} \|\mathbf{s}_i\|^{-2} s_i^{(k)2} x_k\right| \left|\sum_{l=1}^{K} \|\mathbf{s}_i\|^{\beta_2} s_i^{(l)2} x_l\right|\right)$$
$$\le E\left(\left(\max_{k=1}^{K} |x_k|\right)\left(\left|\sum_{l=1}^{K} \|\mathbf{s}_i\|^{\beta_2} s_i^{(l)2} x_l\right|\right)\right) = \left(\max_{k=1}^{K} |x_k| \left(\sum_{l=1}^{K} |x_l|\right)\right).$$

Hence, we obtain

$$\max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{C}_i \mathbf{x} \le \max_{\|\mathbf{x}\|=1} \left\{ \left(\max_{k=1}^{K} |x_k|\right)\left(\sum_{l=1}^{K} |x_l|\right)\right\} = \frac{1+\sqrt{K}}{2}. \qquad (4.58)$$

A proof of the equality in (4.58) is given in the Appendix. Thus, what we can say (at present) is only that the eigenvalues of $\mathbf{I} + \mathbf{M}_i$ is not smaller than $2 - (2-\beta)\frac{1+\sqrt{K}}{2}$, which may take a negative value for large K. This result seems to suggest that parameter β should be sufficiently close to 2 for safety.

4.4 INSTABILITY ANALYSES OF PERMUTED SEPARATORS

4.4.1 *Stationary condition for the permuted separator*

Let us investigate what will happen for the permuted separator $\tilde{\mathbf{W}}$. Let the permuted components in $\mathbf{y}_i$ be $y_i^{(k)}$ $(k = m_1, \ldots, m_L)$: i.e., $y_i^{(k)} = \tilde{s}_i^{(k)}$ $(k = m_1, \ldots, m_L)$. Then, (4.22) together with spherically symmetric generalized Gaussian model leads

$$dE\left(-\log q\left(\mathbf{y}_i\right)\right) = \sum_{k=m_1,\ldots,m_L} dv_{ii}^{(k)} E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(\tilde{s}_i^{(k)2}\right) + \sum_{k\neq m_1,\ldots,m_L} dv_{ii}^{(k)} E\left(\|\mathbf{s}_i\|^{\beta_2} s_i^{(k)2}\right). \tag{4.59}$$

Here we have used the approximation $\|\mathbf{y}_i\| \approx \|\mathbf{s}_i\|$. In the above equation, which components are permuted (i.e., $m_1, \ldots, m_L$) depends on i. Combining the above equation and (4.21), we obtain

$$dQ\left(\tilde{\mathbf{W}}\right) = \sum_{i=1}^{N}\left\{\sum_{k\neq m_1,\ldots,m_L} dv_{ii}^{(k)}\left(E\left(\|\mathbf{s}_i\|^{\beta_2} s_i^{(k)2}\right) - 1\right) + \sum_{k=m_1,\ldots,m_L} dv_{ii}^{(k)}\left(E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(\tilde{s}_i^{(k)2}\right) - 1\right)\right\}. \tag{4.60}$$

Thus, we find that $\tilde{s}_i^{(k)}$ is scaled as

$$E\left(\tilde{s}_i^{(k)2}\right) = E\left(\|\mathbf{s}_i\|^{\beta_2}\right)^{-1} \quad (k = m_1, \ldots, m_L). \tag{4.61}$$

This is a most important equation in this section.

In the next subsection, we need to evaluate expectations $E\left(\|\mathbf{s}_j\|^{\beta_2}\tilde{s}_i^{(k)2}\right)$ and $E\left(\|\mathbf{s}_j\|^{\beta_4}\tilde{s}_i^{(k)2}\right)$ in which $\tilde{s}_i^{(k)}$ is originated from source $\mathbf{s}_j$ and hence $\tilde{s}_i^{(k)2}$ is correlated with $\|\mathbf{s}_j\|^{\beta_2}$ and $\|\mathbf{s}_j\|^{\beta_4}$. Consider $\tilde{s}_i^{(k)} = a s_j^{(k)}$, then we have

$$a^2 = \frac{E\left(\tilde{s}_i^{(k)2}\right)}{E\left(s_j^{(k)2}\right)} = \frac{1}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(s_j^{(k)2}\right)},$$

and hence

$$\tilde{s}_i^{(k)} = \frac{s_j^{(k)}}{\sqrt{E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(s_j^{(k)2}\right)}}. \tag{4.62}$$

By substituting (4.62) into $E\left(\|\mathbf{s}_j\|^{\beta_2}\tilde{s}_i^{(k)2}\right)$ and $E\left(\|\mathbf{s}_j\|^{\beta_4}\tilde{s}_i^{(k)2}\right)$, we obtain

$$E\left(\|\mathbf{s}_j\|^{\beta_2}\tilde{s}_i^{(k)2}\right) = \frac{1}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(s_j^{(k)2}\right)}, \tag{4.63}$$

$$E\left(\|\mathbf{s}_j\|^{\beta_4}\tilde{s}_i^{(k)2}\right) = \frac{E\left(\|\mathbf{s}_j\|^{\beta_4} s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right) E\left(s_j^{(k)2}\right)}. \tag{4.64}$$

The key point of these equations, including (4.61), is that such an expectation as $E\left(\|\mathbf{s}_j\|^{\beta_2}\tilde{s}_i^{(k)2}\right)$ can be expressed by other expectations not including $\tilde{s}_i^{(k)}$.

4.4.2 *Instability of the permuted separator*

We now show that the permuted separator cannot be a local minimum of Q. The outline of the proof is as follows. When permutation occurs among $y_1^{(k)},\ldots,y_N^{(k)}$, a cyclic permutation is included such as $\tilde{s}_{i_1}^{(k)} \propto s_{i_2}^{(k)}$, $\tilde{s}_{i_2}^{(k)} \propto s_{i_3}^{(k)},\ldots,\ \tilde{s}_{i_M}^{(k)} \propto s_{i_1}^{(k)}$ $(M \geq 2)$. Out of $d^2Q\left(\tilde{\mathbf{W}}\right) = \cdots + d^2E\left(-\log q(\mathbf{y}_{i_1})\right)\cdots + d^2E\left(-\log q(\mathbf{y}_{i_2})\right)\cdots$, we pick three terms of $dv_{i_1i_2}^{(k)}dv_{i_2i_1}^{(k)}$, $dv_{i_1i_2}^{(k)2}$, and $dv_{i_2i_1}^{(k)2}$:

$$\begin{aligned}&\left\{E\left(\|\mathbf{y}_{i_1}\|^{\beta_2}y_{i_1}^{(k)2}\right)+E\left(\|\mathbf{y}_{i_2}\|^{\beta_2}y_{i_2}^{(k)2}\right)\right\}dv_{i_1i_2}^{(k)}dv_{i_2i_1}^{(k)}\\&+E\left(\|\mathbf{y}_{i_1}\|^{\beta_2}y_{i_2}^{(k)2}+\beta_2\|\mathbf{y}_{i_1}\|^{\beta_4}y_{i_1}^{(k)2}y_{i_2}^{(k)2}\right)dv_{i_1i_2}^{(k)2}\\&+E\left(\|\mathbf{y}_{i_2}\|^{\beta_2}y_{i_1}^{(k)2}+\beta_2\|\mathbf{y}_{i_2}\|^{\beta_4}y_{i_2}^{(k)2}y_{i_1}^{(k)2}\right)dv_{i_2i_1}^{(k)2}.\end{aligned}\tag{4.65}$$

Then, we prove that this small quadratic form cannot be positive semi-definite if stability condition (A2) for the desired separator is satisfied. This implies that $d^2Q\left(\tilde{\mathbf{W}}\right)$ cannot be positive semi-definite and hence the separator cannot be a (local) minimum of $Q\left(\mathbf{W}\right)$.

The quadratic form with respect to $dv_{i_1i_2}^{(k)}$, and $dv_{i_2i_1}^{(k)}$ can take two different forms, depending on the length of the cycle: the case of $M=2$ and the case of $M\geq 3$. Below, for simplicity of notation, we write i_1 and i_2 as i and j, respectively.

4.4.2.1 *The case of $M=2$*

$y_i^{(k)} = \tilde{s}_i^{(k)} \propto s_j^{(k)}$ and $y_j^{(k)} = \tilde{s}_j^{(k)} \propto s_i^{(k)}$, i.e., permutation occurs between $y_i^{(k)}$ and $y_j^{(k)}$. The quadratic form related to $dv_{ij}^{(k)}$ and $dv_{ji}^{(k)}$ then becomes

$$\begin{aligned}&\left\{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(\tilde{s}_i^{(k)2}\right)+E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(\tilde{s}_j^{(k)2}\right)\right\}dv_{ij}^{(k)}dv_{ji}^{(k)}\\&+\left\{E\left(\|\mathbf{s}_i\|^{\beta_2}\tilde{s}_j^{(k)2}\right)+\beta_2E\left(\|\mathbf{s}_i\|^{\beta_4}\tilde{s}_j^{(k)2}\right)E\left(\tilde{s}_i^{(k)2}\right)\right\}dv_{ij}^{(k)2}\\&+\left\{E\left(\|\mathbf{s}_j\|^{\beta_2}\tilde{s}_i^{(k)2}\right)+\beta_2E\left(\|\mathbf{s}_j\|^{\beta_4}\tilde{s}_i^{(k)2}\right)E\left(\tilde{s}_j^{(k)2}\right)\right\}dv_{ji}^{(k)2}.\end{aligned}\tag{4.66}$$

Note that $\tilde{s}_j^{(k)2}$ is correlated with $\|\mathbf{s}_i\|^{\beta_2}$ and $\|\mathbf{s}_i\|^{\beta_4}$, but $\tilde{s}_i^{(k)2}$ is not. Substituting (4.61), (4.63), and (4.64) into this, we obtain

$$\begin{aligned} & 2dv_{ji}^{(k)}dv_{ij}^{(k)} \\ & + \left\{ \frac{1}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(s_i^{(k)2}\right)} + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}s_i^{(k)2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(s_i^{(k)2}\right)E\left(\|\mathbf{s}_i\|^{\beta_2}\right)} \right\} dv_{ij}^{(k)2} \\ & + \left\{ \frac{1}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} + \beta_2 \frac{E\left(\|\mathbf{s}_j\|^{\beta_4}s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \right\} dv_{ji}^{(k)2} \\ & = 2dv_{ji}^{(k)}dv_{ij}^{(k)} + \frac{\eta_i^{(k)}}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(s_i^{(k)2}\right)}dv_{ij}^{(k)2} + \frac{\eta_j^{(k)}}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)}dv_{ji}^{(k)2}. \end{aligned} \tag{4.67}$$

A sufficient condition for the above expression *not* to be positive semi-definite is

$$\frac{\eta_i^{(k)}}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(s_i^{(k)2}\right)} \frac{\eta_j^{(k)}}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} < 1$$

or

$$\frac{\eta_i^{(k)}\eta_j^{(k)}}{\xi_i^{(k)}\xi_j^{(k)}} < 1. \tag{4.68}$$

This inequality is automatically satisfied under the stability condition (A2), because

$$\frac{\eta_i^{(k)}\eta_j^{(k)}}{\xi_i^{(k)}\xi_j^{(k)}} = \frac{\eta_i^{(k)2}\eta_j^{(k)2}}{\left(\eta_i^{(k)}\xi_i^{(k)} \cdot \eta_j^{(k)}\xi_j^{(k)}\right)} < \eta_i^{(k)2}\eta_j^{(k)2} < 1.$$

4.4.2.2 *The case of $M \geq 3$*

$y_i^{(k)} = \tilde{s}_i^{(k)} \propto s_j^{(k)}$ and $y_j^{(k)} = \tilde{s}_j^{(k)} \propto s_m^{(k)}$ $(m \neq i, j)$. In this case, $\tilde{s}_j^{(k)}$ is not only independent of $\mathbf{s}_j$, but also of $\mathbf{s}_i$. As was described under Eq. (4.45), we may assume $\eta_j^{(k)}\xi_j^{(k)} > 1$ without loss of generality. The quadratic form related with $dv_{ij}^{(k)}$ and $dv_{ji}^{(k)}$ is then

$$\begin{aligned} & \left\{ E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(\tilde{s}_i^{(k)2}\right) + E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(\tilde{s}_j^{(k)2}\right) \right\} dv_{ij}^{(k)}dv_{ji}^{(k)} \\ & + \left\{ E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(\tilde{s}_j^{(k)2}\right) + \beta_2 E\left(\|\mathbf{s}_i\|^{\beta_4}\right)E\left(\tilde{s}_j^{(k)2}\right)E\left(\tilde{s}_i^{(k)2}\right) \right\} dv_{ij}^{(k)2} \\ & + \left\{ E\left(\|\mathbf{s}_j\|^{\beta_2}\tilde{s}_i^{(k)2}\right) + \beta_2 E\left(\|\mathbf{s}_j\|^{\beta_4}\tilde{s}_i^{(k)2}\right)E\left(\tilde{s}_j^{(k)2}\right) \right\} dv_{ji}^{(k)2}. \end{aligned} \tag{4.69}$$

Substituting (4.61), (4.63), and (4.64) into this, we obtain

$$
\begin{aligned}
&2dv_{ji}^{(k)}dv_{ij}^{(k)} + \left\{ \frac{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(\|\mathbf{s}_i\|^{\beta_2}\right)} \right\} dv_{ij}^{(k)2} \\
&+ \left\{ \frac{1}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} + \beta_2 \frac{E\left(\|\mathbf{s}_j\|^{\beta_4} s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \right\} dv_{ji}^{(k)2}
\end{aligned}
\tag{4.70}
$$

A sufficient condition so that the above quadratic form be *not* positive semi-definite is

$$
\left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)^2} \right\} \frac{1}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} \left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_j\|^{\beta_4} s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \right\} < 1
$$

or

$$
\left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)^2} \right\} \frac{\eta_j^{(k)}}{\xi_j^{(k)}} < 1, \tag{4.71}
$$

where

$$
\begin{aligned}
&\left\{ \frac{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(\|\mathbf{s}_i\|^{\beta_2}\right)} \right\} \\
&\times \left\{ \frac{1}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} + \beta_2 \frac{E\left(\|\mathbf{s}_j\|^{\beta_4} s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \right\} \\
&= \left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)^2} \right\} \left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_j\|^{\beta_4} s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \right\} \\
&\times \frac{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \frac{1}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} \\
&= \left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)^2} \right\} \frac{1}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)E\left(s_j^{(k)2}\right)} \left\{ 1 + \beta_2 \frac{E\left(\|\mathbf{s}_j\|^{\beta_4} s_j^{(k)2}\right)}{E\left(\|\mathbf{s}_j\|^{\beta_2}\right)} \right\}.
\end{aligned}
$$

Since

$$
1 + \beta_2 \frac{E\left(\|\mathbf{s}_i\|^{\beta_4}\right)}{E\left(\|\mathbf{s}_i\|^{\beta_2}\right)^2} < 1
$$

and

$$
0 < \frac{\eta_j^{(k)}}{\xi_j^{(k)}} = \frac{\eta_j^{(k)2}}{\left(\eta_j^{(k)}\xi_j^{(k)}\right)} < \eta_j^{(k)2} < 1
$$

hold under the stability condition (A1) and condition of (4.44), the above inequality is satisfied as a necessity.

Example. In order to confirm the validity of the result, let us consider an example. Two independent, K-dimensional vector signals are generated by the spherically symmetric Laplace distribution such that

$$p\left(\mathbf{s}_i\right)=\frac{1}{2\alpha^K\pi^{\frac{K}{2}}}\frac{\Gamma\left(\frac{K}{2}\right)}{\Gamma\left(K\right)}\exp\left\{-\frac{\|\mathbf{s}_i\|}{\alpha}\right\}\ (i=1,2), \tag{4.72}$$

where α stands for the scale parameter, and $\Gamma\left(\cdot\right)$ denotes the Gamma function. The output vectors are obtained through the overall matrix

$$\mathbf{y}=\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix}=\mathbf{WA}\begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{bmatrix}=\mathbf{Gs}, \tag{4.73}$$

where $\mathbf{s}_i=\left[s_i^{(1)},\ldots s_i^{(K)}\right]^T$, and $\mathbf{y}_i=\left[y_i^{(1)},\ldots,y_i^{(K)}\right]^T$.

Let us consider the case that the overall matrix $\mathbf{G}$ takes the form of

$$\mathbf{G}=\begin{bmatrix}\cos\theta & 0 & \cdots & 0 & -\sin\theta & 0 & \cdots & 0\\ 0 & 1 & \ddots & \vdots & 0 & 0 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & 0\\ \sin\theta & 0 & \cdots & 0 & \cos\theta & 0 & \cdots & 0\\ 0 & 0 & \ddots & \vdots & 0 & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1\end{bmatrix}, \tag{4.74}$$

where angle θ varies from $-\pi$ to π. That implies that separation is completely achieved only when $\theta=0$, and permutation between $y_1^{(1)}$ and $y_2^{(1)}$ when $\theta=\pm\pi/2$ is considered.

For the prior pdf of the sources, we adopted a spherically-symmetric, generalized-Gaussian model:

$$q\left(\mathbf{u}\right)=\frac{1}{2\pi^{\frac{K}{2}}}\frac{\Gamma\left(\frac{K}{2}\right)}{\Gamma\left(\frac{K}{\beta}\right)}\beta^{-\frac{K}{\beta}+1}\exp\left\{-\frac{\|\mathbf{u}\|^\beta}{\beta}\right\}\ (0<\beta<2), \tag{4.75}$$

where β denotes the shape parameter. The Kullback-Leibler divergence given in (11), therefore, becomes

$$\begin{aligned}Q\left(\mathbf{G}\right)=&\frac{1}{\beta}\sum_{i=1}^{2}E\left(\|\mathbf{y}_i\|^\beta\right)-\frac{1}{\alpha}\sum_{i=1}^{2}E\left(\|\mathbf{s}_i\|\right)\\ &+2\left\{\log\Gamma\left(\frac{K}{\beta}\right)-\log\Gamma\left(K\right)+\left(\frac{K}{\beta}-1\right)\log\beta-K\log\alpha\right\}.\end{aligned} \tag{4.76}$$

The scaling of the sources is adjusted to coincide with $E\left(\|\mathbf{s}_i\|^\beta\right)=K$, and hence the scale parameter α is given by

$$\alpha=\left(\frac{\Gamma\left(K+1\right)}{\Gamma\left(K+\beta\right)}\right)^{\frac{1}{\beta}}. \tag{4.77}$$

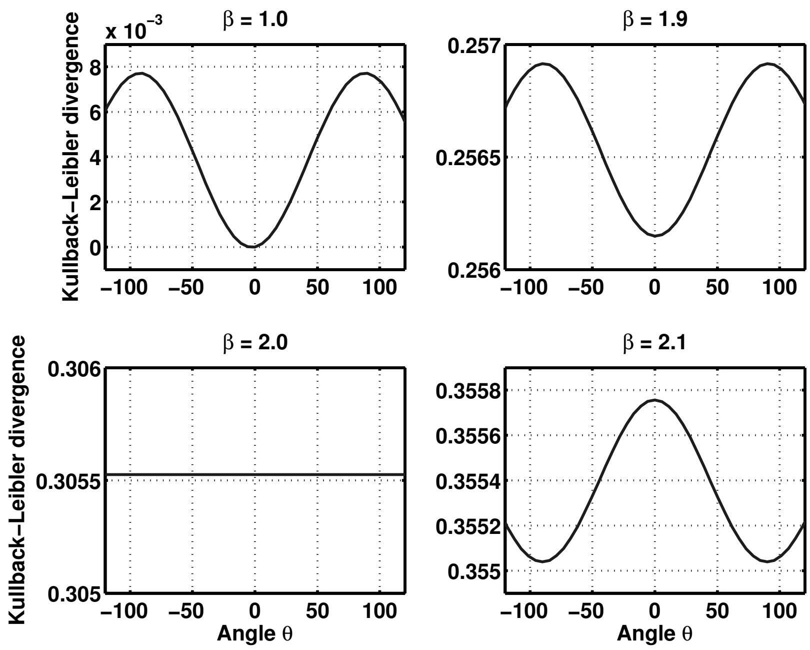

Figure 7: The Kullback-Leibler divergence Q as a function of angle θ with different β. The number of components in each vector was 128, i.e., $K = 128$. For sample average, 100000 samples were used. When β was less than 2.0, the minimum of Q clearly lay at $\theta = 0$. When β exceeded 2.0, the minima of Q lay at $-\pi/2$ and $\pi/2$.

These expectations were computed as the sample average, i.e.,

$$E\left(\|\mathbf{y}_i\|^\beta\right) \approx \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{y}_i(t)\|^\beta \text{ and } E\left(\|\mathbf{s}_i\|\right) \approx \frac{1}{T}\sum_{t=1}^{T}\|\mathbf{s}_i(t)\|.$$

Figure 7 shows the Kullback-Leibler divergence Q as a function of angle θ in the case of $K = 128$ and $T = 100000$ with different β. When β was less than 2.0, any permuted solution never became a local minimum of the function. When, however, β exceeded two, the permuted solution became a local minimum. Since the same pdf is used for generating $\mathbf{s}_1(t)$ and $\mathbf{s}_2(t)$, the stability condition (A3) is reduced to $\beta < 2.0$. That implies that the optimal solution is always a local minimum of Q if β is less than 2.0.

4.5 SEPARATION OF COMPLEX-VALUED SOURCE VECTORS

4.5.1 *Complex-valued sources and the demixing process*

An important application of IVA is one to sound signal separation, where the observed data are decomposed into a number of sets of frequency components and some ICA algorithm is applied to each set of components with the same frequency. In this approach the frequency components of the observed signals are complex-valued. In this section we roughly describe IVA for complex-valued sources.

There are two approaches to handle a real-valued function of a complex variable. One is to replace every complex number with its real and imaginary parts, and then deal with a real-valued function of two real variables. The other is to use the Wirtinger calculus, in which a complex variable and its complex conjugate can be treated independently as described in Chap. 2. Since the latter one provides a convenient form, the Wirtinger calculus is used to elucidate the stability condition for the complex-valued IVA.

In the complex domain, the demixing process takes the form

$$\overline{\mathbf{y}}_i + j\underline{\mathbf{y}}_j = \sum_{k=1}^{N} \left(\overline{\mathbf{W}}_{ik} + j\underline{\mathbf{W}}_{ik}\right)\left(\overline{\mathbf{x}}_k + j\underline{\mathbf{x}}_k\right) \;\; (i = 1, \ldots, N) \tag{4.78}$$

Its equivalent expression is

$$\mathbf{y}_{i,\mathbb{R}} = \sum_{k=1}^{N} \mathbf{W}_{ik,\mathbb{R}} \mathbf{x}_{k,\mathbb{R}}, \tag{4.79}$$

where

$$\mathbf{x}_{k,\mathbb{R}} = \begin{bmatrix} \overline{\mathbf{x}}_k \\ \underline{\mathbf{x}}_k \end{bmatrix}, \quad \mathbf{y}_{i,\mathbb{R}} = \begin{bmatrix} \overline{\mathbf{y}}_i \\ \underline{\mathbf{y}}_i \end{bmatrix}, \quad \text{and } \mathbf{W}_{ij,\mathbb{R}} = \begin{bmatrix} \overline{\mathbf{W}}_{ij} & -\underline{\mathbf{W}}_{ij} \\ \underline{\mathbf{W}}_{ij} & \overline{\mathbf{W}}_{ij} \end{bmatrix};$$

matrices $\overline{\mathbf{W}}_{ij}$ and $\underline{\mathbf{W}}_{ij}$ are diagonal. The desired separator is obtained by minimizing function

$$Q(\mathbf{W}_{\mathbb{R}}) = -\log|\det \mathbf{W}_{\mathbb{R}}| + \sum_{i=1}^{N} E\left(-\log q(\mathbf{y}_{i,\mathbb{R}})\right) + const, \tag{4.80}$$

where the prior pdf $q(\mathbf{u}_{\mathbb{R}}) = q(\overline{\mathbf{u}}, \underline{\mathbf{u}}) = q(\overline{u}^{(1)}, \ldots, \overline{u}^{(K)}, \underline{u}^{(1)}, \ldots, \underline{u}^{(K)})$ for $\mathbf{s}_{i,\mathbb{R}}$ is

$$q(\mathbf{u}) = \alpha \exp\left(-\frac{1}{\beta}\|\mathbf{u}_{\mathbb{R}}\|^{\beta}\right) = \alpha \exp\left\{-\frac{1}{\beta}\left(\|\overline{\mathbf{u}}\|^2 + \|\underline{\mathbf{u}}\|^2\right)^{\frac{\beta}{2}}\right\}. \tag{4.81}$$

4.5.2 *The first and second-order differentials of* $-\log q(\mathbf{y}_{i,\mathbb{C}})$

In the same way as the real-valued IVA, we can derive

$$-d\log q(\mathbf{y}_{i,\mathbb{R}}) = \sum_{j=1}^{N} \mathbf{y}_{j,\mathbb{R}}^{\top} d\mathbf{V}_{ij,\mathbb{R}}^{\top} \boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{R}}\right) \tag{4.82}$$

and

$$\begin{aligned} -d^2 \log q\left(\mathbf{y}_{i,\mathbb{R}}\right) = &\sum_{k=1}^{N}\sum_{j=1}^{N} \mathbf{y}_{k,\mathbb{R}}^{\top} d\mathbf{V}_{jk,\mathbb{R}}^{\top} d\mathbf{V}_{ij,\mathbb{R}}^{\top} \boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{R}}\right) \\ &+ \sum_{j=1}^{N}\sum_{k=1}^{N} \mathbf{y}_{j,\mathbb{R}}^{\top} d\mathbf{V}_{ij,\mathbb{R}}^{\top} \boldsymbol{\Psi}\left(\mathbf{y}_{i,\mathbb{R}}\right) d\mathbf{V}_{ik,\mathbb{R}} \mathbf{y}_{k,\mathbb{R}}. \end{aligned} \tag{4.83}$$

where

$$d\mathbf{V}_{ij,\mathbb{R}} = \begin{bmatrix} d\overline{\mathbf{V}}_{ij} & -d\underline{\mathbf{V}}_{ij} \\ d\underline{\mathbf{V}}_{ij} & d\overline{\mathbf{V}}_{ij} \end{bmatrix},$$

$$\boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{R}}\left(t\right)\right) = -\frac{\partial \log q\left(\mathbf{y}_{i,\mathbb{R}}\right)}{\partial \mathbf{y}_{i,\mathbb{R}}},$$

$$\boldsymbol{\Psi}\left(\mathbf{y}_{i,\mathbb{R}}\right) = \frac{\partial^2 \log q\left(\mathbf{y}_{i,\mathbb{R}}\right)}{\partial \mathbf{y}_{i,\mathbb{R}} \partial \mathbf{y}_{i,\mathbb{R}}^\top}.$$

As is described in Chap. 2, vectors $\mathbf{y}_{i,\mathbb{R}}$, $\boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{R}}\right)$ and matrices $d\mathbf{V}_{ij,\mathbb{R}}$, $\boldsymbol{\Psi}\left(\mathbf{y}_{i,\mathbb{R}}\right)$ can be transformed into the complex domain:

$$\mathbf{y}_{i,\mathbb{C}} = \begin{bmatrix} \mathbf{y}_i \\ \mathbf{y}_i^* \end{bmatrix} = \begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ \mathbf{I} & -j\mathbf{I} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{y}}_i \\ \underline{\mathbf{y}}_i \end{bmatrix} = \mathbf{J}\mathbf{y}_{i,\mathbb{R}},$$

$$\boldsymbol{\varphi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right) = \begin{bmatrix} \boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{C}}\right) \\ \boldsymbol{\varphi}^*\left(\mathbf{y}_{i,\mathbb{C}}\right) \end{bmatrix} = \begin{bmatrix} -\frac{\partial \log q(\mathbf{y}_{i,\mathbb{C}})}{\partial \mathbf{y}_i} \\ -\frac{\partial \log q(\mathbf{y}_{i,\mathbb{C}})}{\partial \mathbf{y}_i^*} \end{bmatrix} = \frac{1}{2}\mathbf{J}^*\boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{R}}\right)$$

and

$$d\mathbf{V}_{ij,\mathbb{C}} = \begin{bmatrix} d\mathbf{V}_{ij} & \mathbf{O} \\ \mathbf{O} & d\mathbf{V}_{ij}^* \end{bmatrix} = \mathbf{J}d\mathbf{V}_{ij,\mathbb{R}}\mathbf{J}^{-1},$$

$$\boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right) = \begin{bmatrix} -\frac{\partial^2 \log q(\mathbf{y}_{i,\mathbb{C}})}{\partial \mathbf{y}_i^* \partial \mathbf{y}_i^\top} & -\frac{\partial^2 \log q(\mathbf{y}_{i,\mathbb{C}})}{\partial \mathbf{y}_i^* \partial \mathbf{y}_i^H} \\ -\frac{\partial^2 \log q(\mathbf{y}_{i,\mathbb{C}})}{\partial \mathbf{y}_i \partial \mathbf{y}_i^\top} & -\frac{\partial^2 \log q(\mathbf{y}_{i,\mathbb{C}})}{\partial \mathbf{y}_i \partial \mathbf{y}_i^H} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}\left(\mathbf{y}_{i,\mathbb{C}}\right) & \mathbf{H}^*\left(\mathbf{y}_{i,\mathbb{C}}\right) \\ \mathbf{H}\left(\mathbf{y}_{i,\mathbb{C}}\right) & \mathbf{Z}^*\left(\mathbf{y}_{i,\mathbb{C}}\right) \end{bmatrix} = \frac{1}{2}\mathbf{J}\boldsymbol{\Psi}\left(\mathbf{y}_{i,\mathbb{R}}\right)\mathbf{J}^{-1},$$

where

$$d\mathbf{V}_{ij} = \begin{bmatrix} dv_{ij}^{(1)} & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & dv_{ij}^{(K)} \end{bmatrix} \in \mathbb{C}^{K\times K}.$$

Note that matrix $\boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right)$ is Hermitian matrix. Using the above transformations, $-d\log q\left(\mathbf{y}_{i,\mathbb{R}}\right)$ and $-d^2\log q\left(\mathbf{y}_{i,\mathbb{R}}\right)$ can be rewritten as

$$\begin{aligned} -d\log q\left(\mathbf{y}_{i,\mathbb{C}}\right) &= \sum_{j=1}^{N} \mathbf{y}_{j,\mathbb{R}}^\top \mathbf{J}^\top \mathbf{J}^{-\top} d\mathbf{V}_{ij,\mathbb{R}}^\top \mathbf{J}^\top \mathbf{J}^{-\top} \boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{R}}\right) \\ &= \sum_{j=1}^{N} \mathbf{y}_{j,\mathbb{C}}^\top d\mathbf{V}_{ij,\mathbb{C}}^\top \boldsymbol{\varphi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right) \qquad (4.84) \\ &= \sum_{j=1}^{N} \mathbf{y}_j^\top d\mathbf{V}_{ij} \boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{C}}\right) + \sum_{j=1}^{N} \mathbf{y}_j^H d\mathbf{V}_{ij}^H \boldsymbol{\varphi}^*\left(\mathbf{y}_{i,\mathbb{C}}\right) \end{aligned}$$

and

$$-d^2 \log q\left(\mathbf{y}_{i,\mathbb{C}}\right) = \sum_{k=1}^{N}\sum_{j=1}^{N} \mathbf{y}_{k,\mathbb{C}}^{\top} d\mathbf{V}_{jk,\mathbb{C}}^{\top} d\mathbf{V}_{ij,\mathbb{C}}^{\top} \boldsymbol{\varphi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right) + \sum_{j=1}^{N}\sum_{k=1}^{N} \mathbf{y}_{j,\mathbb{C}}^{H} d\mathbf{V}_{ij,\mathbb{C}}^{H} \boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{R}}\right) d\mathbf{V}_{ik,\mathbb{C}} \mathbf{y}_{k,\mathbb{C}}. \tag{4.85}$$

Define

$$d\mathbf{v}_{ij} = \begin{bmatrix} dv_{ij}^{(1)} \\ \vdots \\ dv_{ij}^{(K)} \end{bmatrix} \in \mathbb{C}^K,\ \mathbf{Y}_j = \begin{bmatrix} y_j^{(1)} & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & y_j^{(K)} \end{bmatrix} \in \mathbb{C}^{K\times K},$$

and

$$\boldsymbol{\Phi}\left(\mathbf{y}_{i,\mathbb{C}}\right) = \begin{bmatrix} \varphi^{(1)}\left(\mathbf{y}_{i,\mathbb{C}}\right) & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \varphi^{(K)}\left(\mathbf{y}_{i,\mathbb{C}}\right) \end{bmatrix} \in \mathbb{C}^{K\times K}.$$

Then, Eqs. (4.84) and (4.85) can be rewritten as

$$-d \log q\left(\mathbf{y}_{i,\mathbb{C}}\right) = \sum_{j=1}^{N} d\mathbf{v}_{ij}^{\top} \mathbf{Y}_j \boldsymbol{\varphi}\left(\mathbf{y}_{i,\mathbb{C}}\right) + \sum_{j=1}^{N} d\mathbf{v}_{ij}^{H} \mathbf{Y}_j^{H} \boldsymbol{\varphi}^*\left(\mathbf{y}_{i,\mathbb{C}}\right) \tag{4.86}$$

and

$$-d^2 \log q\left(\mathbf{y}_{i,\mathbb{C}}\right) = \sum_{k=1}^{N}\sum_{j=1}^{N} d\mathbf{v}_{jk,\mathbb{C}}^{\top} \mathbf{Y}_{k,\mathbb{C}} \boldsymbol{\Phi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right) d\mathbf{v}_{ij,\mathbb{C}} + \sum_{j=1}^{N}\sum_{k=1}^{N} d\mathbf{v}_{ij,\mathbb{C}}^{H} \mathbf{Y}_{j,\mathbb{C}}^{H} \boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{y}_{i,\mathbb{C}}\right) \mathbf{Y}_{k,\mathbb{C}} d\mathbf{v}_{ik,\mathbb{C}}. \tag{4.87}$$

4.5.3 *The first-order differential of $Q(\mathbf{W}_{\mathbb{C}})$ at $\mathbf{W}_{\mathbb{C}} = \mathbf{W}_{\mathbb{C},opt}$*

The first-order differentials of the first and second terms of (4.80) at $\mathbf{W}_{\mathbb{R}} = \mathbf{W}_{\mathbb{R},\mathrm{opt}}$ are, respectively,

$$\begin{aligned} -d \log\left|\det \mathbf{W}_{\mathbb{R}}\right| &= -\mathrm{tr} d\mathbf{V}_{\mathbb{R}} = -\mathrm{tr}\begin{bmatrix} d\overline{\mathbf{V}} & -d\underline{\mathbf{V}} \\ d\underline{\mathbf{V}} & d\overline{\mathbf{V}} \end{bmatrix} = -2\sum_{i=1}^{N} \mathrm{tr} d\overline{\mathbf{V}}_{ii} \\ &= -2\sum_{i=1}^{N}\sum_{k=1}^{K} d\overline{v}_{ii}^{(k)} = -\sum_{i=1}^{N}\sum_{k=1}^{K}\left(dv_{ii}^{(k)} + dv_{ii}^{(k)*}\right) \\ &= -\sum_{i=1}^{N}\left(d\mathbf{v}_{ii}^{\top} + d\mathbf{v}_{ii}^{H}\right)\mathbf{e}, \end{aligned} \tag{4.88}$$

$$dE\left(-\log q\left(\mathbf{y}_{i,\mathbb{C}}\right)\right) = d\mathbf{v}_{ii}^{\top} E\left(\mathbf{S}_i \boldsymbol{\varphi}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right) + d\mathbf{v}_{ii}^{H} E\left(\mathbf{S}_i^{H} \boldsymbol{\varphi}^*\left(\mathbf{s}_{i,\mathbb{C}}\right)\right) \tag{4.89}$$

because

$$\mathbf{W}_{\mathbb{C},\mathrm{opt}} = \begin{bmatrix} \mathbf{W}_{\mathrm{opt}} & \mathbf{O} \\ \mathbf{O} & \mathbf{W}_{\mathrm{opt}}^* \end{bmatrix} = \mathbf{J}\mathbf{W}_{\mathbb{R},\mathrm{opt}}\mathbf{J}^{-1}.$$

Thus, the first-order differential of $Q\left(\mathbf{W}_{\mathbb{R}}\right)$ at $\mathbf{W}_{\mathbb{R}} = \mathbf{W}_{\mathbb{R},\mathrm{opt}}$ is

$$\begin{aligned} dQ\left(\mathbf{W}_{\mathbb{R},\mathrm{opt}}\right) &= \sum_{i=1}^{N} d\mathbf{v}_{ii}^{\top}\left\{E\left(\mathbf{S}_i\boldsymbol{\varphi}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right)-\mathbf{e}\right\} \\ &+ \sum_{i=1}^{N} d\mathbf{v}_{ii}^{H}\left\{E\left(\mathbf{S}_i^{H}\boldsymbol{\varphi}^{*}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right)-\mathbf{e}\right\} \\ &= 2\sum_{i=1}^{N}\mathrm{Re}\left\{\left\{E\left(\mathbf{S}_i\boldsymbol{\varphi}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right)-\mathbf{e}\right\}^{\top} d\mathbf{v}_{ii}\right\} \\ &= 2\sum_{i=1}^{N}\mathrm{Re}\left\{\left\langle E\left(\mathbf{S}_i^{*}\boldsymbol{\varphi}^{*}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right)-\mathbf{e}, d\mathbf{v}_{ii}\right\rangle\right\}, \end{aligned} \tag{4.90}$$

where $\langle \mathbf{a}, \mathbf{b}\rangle$ stands for the inner product of two vectors $\mathbf{a}$ and $\mathbf{b}$. From this we find that, when the desired separator is obtained as a minimum of $Q(\mathbf{W})$, the output $\mathbf{y}_i = \mathbf{s}_i$ is scaled as

$$E\left(\varphi^{(k)*}\left(\mathbf{s}_i,\mathbf{s}_i^{*}\right)s_i^{(k)*}\right) = 1 \;\; (i=1,\ldots,N; k=1,\ldots,K) \tag{4.91}$$

or

$$E\left(\boldsymbol{\Phi}^{*}\left(\mathbf{s}_i,\mathbf{s}_i^{*}\right)\mathbf{S}_i^{*}\right) = \mathbf{I} \;\; (i=1,\ldots,N). \tag{4.92}$$

It should be noted that, $E\left(\varphi^{(k)}\left(\mathbf{s}_i,\mathbf{s}_i^{*}\right)s_i^{(k)}\right) = 1 \;\; (i=1,\ldots,N; k=1,\ldots,K)$ is also a stationary point of $Q\left(\mathbf{W}\right)$.

4.5.4 *The second-order differential of $Q(\mathbf{W}_{\mathbb{C}})$ at $\mathbf{W}_{\mathbb{C}} = \mathbf{W}_{\mathbb{C},opt}$*

The second-order differential of $E\left(-\log q\left(\mathbf{y}_{i,\mathbb{C}}\right)\right)$ for $\mathbf{W}_{\mathbb{C}} = \mathbf{W}_{\mathbb{C},\mathrm{opt}}$ is

$$\begin{aligned} d^2E\left(-\log q\left(\mathbf{y}_{i,\mathbb{C}}\right)\right) &= \sum_{j=1}^{N} d\mathbf{v}_{ji,\mathbb{C}}^{\top}E\left(\mathbf{S}_{i,\mathbb{C}}\boldsymbol{\Phi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right)d\mathbf{v}_{ij,\mathbb{C}} \\ &+ \sum_{j=1}^{N} d\mathbf{v}_{ij,\mathbb{C}}^{H}E\left(\mathbf{S}_{j,\mathbb{C}}^{H}\boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{C}}\right)\mathbf{S}_{j,\mathbb{C}}\right)d\mathbf{v}_{ij,\mathbb{C}} \\ &= \sum_{j=1}^{N} d\mathbf{v}_{ji,\mathbb{C}}^{\top}d\mathbf{v}_{ij,\mathbb{C}} \\ &+ \sum_{j=1}^{N} d\mathbf{v}_{ij,\mathbb{C}}^{H}E\left(\boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{C}}\right)\circledast\left(\mathbf{s}_{j,\mathbb{C}}^{*}\mathbf{s}_{j,\mathbb{C}}^{\top}\right)\right)d\mathbf{v}_{ij,\mathbb{C}}. \end{aligned} \tag{4.93}$$

Divide this summation into the terms with $j \neq i$ and those with $j = i$:

$$\sum_{j\neq i}\left\{d\mathbf{v}_{ji,\mathbb{C}}^{\top}d\mathbf{v}_{ij,\mathbb{C}} + d\mathbf{v}_{ij,\mathbb{C}}^{H}E\left(\boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{C}}\right)\right)\circledast E\left(\mathbf{s}_{j,\mathbb{C}}^{*}\mathbf{s}_{j,\mathbb{C}}^{\top}\right)d\mathbf{v}_{ij,\mathbb{C}}\right\} \tag{4.94}$$

and

$$d\mathbf{v}_{ii,\mathbb{C}}^{\top}d\mathbf{v}_{ii,\mathbb{C}} + d\mathbf{v}_{ii,\mathbb{C}}^{H}E\left(\boldsymbol{\Psi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{C}}\right)\circledast\left(\mathbf{s}_{i,\mathbb{C}}^{*}\mathbf{s}_{i,\mathbb{C}}^{\top}\right)\right)d\mathbf{v}_{ii,\mathbb{C}} \tag{4.95}$$

Thus, the second-order differential of $Q(\mathbf{W})$ at $\mathbf{W}_{\mathbb{R}} = \mathbf{W}_{\mathbb{R},\mathrm{opt}}$ is

$$\begin{aligned} d^2 Q\left(\mathbf{W}_{\mathbb{R},\mathrm{opt}}\right) = & \sum_{i=1}^{N} d\mathbf{v}_{ii,\mathbb{C}}^{H} \left\{\mathbf{P} + \mathbf{M}_{i,\mathbb{C}}\right\} d\mathbf{v}_{ii,\mathbb{C}} \\ & + \sum_{i=1}^{N} \sum_{j \neq i} d\mathbf{v}_{ji,\mathbb{C}}^{H} \mathbf{P} d\mathbf{v}_{ij,\mathbb{C}} + d\mathbf{v}_{ij,\mathbb{C}}^{H} \mathbf{\Gamma}_{ij,\mathbb{C}} d\mathbf{v}_{ij,\mathbb{C}}, \end{aligned} \tag{4.96}$$

where

$$\mathbf{P} = \begin{bmatrix} \mathbf{O} & \mathbf{I} \\ \mathbf{I} & \mathbf{O} \end{bmatrix}, \tag{4.97}$$

$$\mathbf{M}_{i,\mathbb{C}} = E\left(\mathbf{\Psi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{C}}\right) \circledast \left(\mathbf{s}_{i,\mathbb{C}}^{*} \mathbf{s}_{i,\mathbb{C}}^{\top}\right)\right), \tag{4.98}$$

$$\mathbf{\Gamma}_{ij,\mathbb{C}} = \begin{bmatrix} \mathbf{\Gamma}_{ij} & \mathbf{Y}_{ij}^{*} \\ \mathbf{Y}_{ij} & \mathbf{\Gamma}_{ij}^{*} \end{bmatrix} = E\left(\mathbf{\Psi}_{\mathbb{C}}\left(\mathbf{s}_{i,\mathbb{R}}\right)\right) \circledast E\left(\mathbf{s}_{j,\mathbb{C}}^{*} \mathbf{s}_{j,\mathbb{C}}^{\top}\right). \tag{4.99}$$

It should be noted that matrices $\mathbf{M}_{i,\mathbb{C}}$ and $\mathbf{\Gamma}_{ij,\mathbb{C}}$ are Hermitian matrices. This corresponds to (4.31) of the real-valued case.

4.6 STABILITY CONDITIONS FOR THE DESIRED SEPARATOR

We can write a *necessary and sufficient* condition for the above quadratic form to be positive definite in the same way as (A) and (B) in subsection 4.3.3. A necessary and sufficient condition for that is

(A) Matrix $\begin{bmatrix} \mathbf{\Gamma}_{ij,\mathbb{C}} & \mathbf{P} \\ \mathbf{P} & \mathbf{\Gamma}_{ji,\mathbb{C}} \end{bmatrix}$ is positive definite for every pair of i and $j\,(\neq i)$;

(B) Matrix $\mathbf{P} + \mathbf{M}_{i,\mathbb{C}}$ is positive definite for every i.

As in the case of the real-valued IVA, we also derive a necessary condition for the stability. Picking out terms associated with $\left|dv_{ij}^{(k)}\right|^2$, $\left|dv_{ji}^{(k)}\right|^2$, $dv_{ij}^{(k)} dv_{ji}^{(k)}$ and $dv_{ij}^{(k)*} dv_{ji}^{(k)*}$ from (4.96) leads

$$\begin{aligned} & 2dv_{ji}^{(k)} dv_{ij}^{(k)} + 2dv_{ji}^{(k)*} dv_{ij}^{(k)*} \\ & + \left(\gamma_{ij}^{(k,k)} + \gamma_{ij}^{(k,k)*}\right) \left|dv_{ij}^{(k)}\right|^2 + \left(\gamma_{ji}^{(k,k)} + \gamma_{ji}^{(k,k)*}\right) \left|dv_{ji}^{(k)}\right|^2 \\ & + v_{ij}^{(k,k)} dv_{ij}^{(k)2} + v_{ij}^{(k,k)*} dv_{ij}^{(k)*2} + v_{ji}^{(k,k)*} dv_{ji}^{(k)*2} + v_{ji}^{(k,k)} dv_{ji}^{(k)2} \\ & = 2dv_{ji}^{(k)} dv_{ij}^{(k)} + 2dv_{ji}^{(k)*} dv_{ij}^{(k)*} \\ & + 2\zeta_{i}^{(k,k)} \sigma_{j}^{(k,k)} \left|dv_{ij}^{(k)}\right|^2 + 2\zeta_{j}^{(k,k)} \sigma_{i}^{(k,k)} \left|dv_{ji}^{(k)}\right|^2 \\ & + \rho_{i}^{(k,k)} \theta_{j}^{(k,k)} dv_{ij}^{(k)2} + \rho_{i}^{(k,k)*} \theta_{j}^{(k,k)*} dv_{ij}^{(k)*2} \\ & + \rho_{j}^{(k,k)} \theta_{i}^{(k,k)} dv_{ji}^{(k)2} + \rho_{j}^{(k,k)*} \theta_{i}^{(k,k)*} dv_{ji}^{(k)*2}, \end{aligned} \tag{4.100}$$

where

$$\begin{aligned}\zeta_i^{(k,k)} &= E\left(-\frac{\partial^2 \log q\left(\mathbf{s}_{i,\mathbb{C}}\right)}{\partial s_i^{(k)*}\partial s_i^{(k)}}\right),\\ \sigma_j^{(k,k)} &= E\left(\left|s_j^{(k)}\right|^2\right),\\ \rho_i^{(k,k)} &= E\left(-\frac{\partial^2 \log q\left(\mathbf{s}_{i,\mathbb{C}}\right)}{\partial s_i^{(k)2}}\right),\\ \theta_j^{(k,k)} &= E\left(s_j^{(k)2}\right).\end{aligned}$$

The above quadratic form can be represented in a vector-matrix form such that

$$\begin{bmatrix} dv_{ij}^{(k)*} & dv_{ij}^{(k)} & dv_{ji}^{(k)*} & dv_{ji}^{(k)} \end{bmatrix} \boldsymbol{\mathcal{H}}_{ij}^{(k,k)} \begin{bmatrix} dv_{ij}^{(k)} \\ dv_{ij}^{(k)*} \\ dv_{ji}^{(k)} \\ dv_{ji}^{(k)*} \end{bmatrix}, \tag{4.101}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{H}}_{ij}^{(k,k)} &= \begin{bmatrix} \mathbf{H}_{ij}^{(k,k)} & \mathbf{P} \\ \mathbf{P} & \mathbf{H}_{ji}^{(k,k)} \end{bmatrix},\\ \mathbf{H}_{ij}^{(k,k)} &= \begin{bmatrix} \zeta_i^{(k,k)}\sigma_j^{(k,k)} & \rho_i^{(k,k)*}\theta_j^{(k,k)*} \\ \rho_i^{(k,k)}\theta_j^{(k,k)} & \zeta_i^{(k,k)}\sigma_j^{(k,k)} \end{bmatrix}.\end{aligned}$$

In order for $d^2 Q\left(\mathbf{W}_{\mathbb{C},\mathrm{opt}}\right)$ to be positive definite, the Hermitian matrix $\boldsymbol{\mathcal{H}}_{ij}^{(k,k)}$ must be positive definite for every $i, j\,(\neq i)$ and k.

For any 2×2 matrix, two eigenvalues of the matrix are positive if and only if both trace and determinant of the matrix are positive. $\boldsymbol{\mathcal{H}}_{ij}^{(k,k)}$ is positive definite if and only if $\mathbf{H}_{ji}^{(k,k)}$ and $\mathbf{H}_{ij}^{(k,k)}$ are positive definite and $\mathbf{C}_{ij}^{(k,k)} = \mathbf{H}_{ji}^{(k,k)} - \mathbf{P}\mathbf{H}_{ij}^{(k,k)-1}\mathbf{P}$ (or $\mathbf{C}_{ji}^{(k,k)}$) is positive definite. Thus, we obtain a necessary condition for the stability:

(A1) If $d^2Q\left(\mathbf{W}_{\mathbb{C},\mathrm{opt}}\right)$ is positive definite, then the following inequalities must hold for every $i, j\,(\neq i)$ and k:

$$\zeta_i^{(k,k)} > 0, \tag{4.102}$$

$$\zeta_i^{(k,k)2}\sigma_j^{(k,k)2} > \left|\rho_i^{(k,k)}\theta_j^{(k,k)}\right|^2, \tag{4.103}$$

$$\zeta_j^{(k,k)}\sigma_i^{(k,k)} - \frac{\zeta_i^{(k,k)}\sigma_j^{(k,k)}}{\zeta_i^{(k,k)2}\sigma_j^{(k,k)2} - \left|\rho_i^{(k,k)}\theta_j^{(k,k)}\right|^2} > 0, \tag{4.104}$$

$$\begin{aligned}&\left(\zeta_i^{(k,k)2}\sigma_j^{(k,k)2} - \left|\rho_i^{(k,k)}\theta_j^{(k,k)}\right|^2\right)\\ &\times\left(\zeta_j^{(k,k)2}\sigma_i^{(k,k)2} - \left|\rho_j^{(k,k)}\theta_i^{(k,k)}\right|^2\right) + 1\\ &- 2\zeta_i^{(k,k)}\zeta_j^{(k,k)}\sigma_i^{(k,k)}\sigma_j^{(k,k)} - 2\mathrm{Re}\left\{\rho_i^{(k,k)}\rho_j^{(k,k)}\theta_i^{(k,k)}\theta_j^{(k,k)}\right\} > 0.\end{aligned} \tag{4.105}$$

This condition does not seem to be so useful. Thus, we introduce some assumptions.

Firstly, we assume circularity of $s_i^{(k)}$ such that

$$\theta_{ii}^{(k,k)} = E\left(s_i^{(k)2}\right) = 0 \ \ (i = 1, \ldots, N; k = 1, \ldots, N)\,. \tag{4.106}$$

In the strict-sense, a complex random variable $X = \overline{X} + j\underline{X}$ is circular, if X and $Xe^{j\alpha}$ have the same density function, where α is arbitrary angle [45, 63]. Incorporating this assumption, the necessary condition (A1) can be further simplified down to

(A2) $\zeta_i^{(k,k)} > 0$ and $\zeta_i^{(k,k)}\zeta_j^{(k,k)}\sigma_i^{(k,k)}\sigma_j^{(k,k)} > 1$ for every $i, j\,(\neq i)$ and k.

This is the same necessary condition in the case of the real-valued case.

In addition to circularity, we introduce uncorrelatedness of each source vector described in subsection 4.3.4:

$$E\left(\mathbf{s}_i^{(k)}\mathbf{s}_i^{(k)H}\right) = E\left(\left|\mathbf{S}_i^{(k)}\right|^2\right) = \mathrm{diag}\left\{\sigma_i^{(1,1)}, \ldots, \sigma_i^{(K,K)}\right\}. \tag{4.107}$$

The necessary condition (A2) together with (B) becomes a necessary and sufficient condition for the stability under circularity and uncorrelatedness. It is because$\boldsymbol{\Gamma}_{ij,\mathbb{C}}$ becomes a $2K \times 2K$ diagonal matrix such that

$$\boldsymbol{\Gamma}_{ij,\mathbb{C}} = \begin{bmatrix}\boldsymbol{\Gamma}_{ij} & \mathbf{O}\\ \mathbf{O} & \boldsymbol{\Gamma}_{ij}\end{bmatrix} = \mathrm{diag}\left\{\gamma_{ij}^{(1,1)}, \ldots, \gamma_{ij}^{(K,K)}, \gamma_{ij}^{(1,1)}, \ldots, \gamma_{ij}^{(K,K)}\right\},$$

and then large quadratic form can be decomposed into a set of small quadratic forms. We can also discuss the instability of a permuted separator, but we do not describe it because the derivation is essentially the same as in the case of real-valued signals.

4.7 SUMMARY

We have derived a necessary and sufficient condition for the desired separator to be obtained by minimizing a measure representing the difference between the prior source pdf and the actual output pdf of the separator. Also we have proven that, if a desired separator minimizes the measure, then any permuted separator never becomes a local minimum. This remarkable property strongly supports the validity of the IVA approach.

In addition we have investigated a necessary and sufficient condition under different assumptions such as uncorrelatedness, evenness and complex-valued case. If the components within each source are uncorrelated, a necessary condition (A1) together with (B) becomes a necessary and sufficient condition for the desired separator. In the case of evenly distributed sources, condition (B) holds as a necessity; the stability condition takes a simple form. These assumptions, such as uncorrelatedness and evenness of the source vector, are very important to realize a perfect separation by an IVA algorithm assuming a spherically symmetric source prior. The stability condition for the complex-valued IVA is essentially the same as the condition for the real-valued IVA, under circularity of each source component and uncorrelatedness of each source vector. In this chapter, however, we have not discussed at all the case that the sources are sub-Gaussian. This case might require a quite different theoretical framework from that of the present chapter.

APPENDIX A. PROOF OF EQUALITY IN (4. 58)

Define $p_k \triangleq |x_k| \geq 0$. Then, $\left(\max_{k=1}^K |x_k|\right)\left(\sum_{k=1}^K |x_k|\right)$ is written as

$$\left(\max_{k=1}^{K} p_k\right)\left(\sum_{k=1}^{K} p_k\right) \triangleq f(\mathbf{p}) = f(p_1, \ldots, p_K), \tag{4.108}$$

where $p_k \geq 0$. Thus, we have

$$\max_{\|\mathbf{x}\|=1}\left\{\left(\max_{k=1}^{K} |x_k|\right)\left(\sum_{k=1}^{K} |x_k|\right)\right\} = \max_{\substack{\sum_{k=1}^K p_k^2=1 \\ p_1,\ldots,p_K \geq 0}} f(\mathbf{p}). \tag{4.109}$$

Let $\mathbf{p}^* = [p_1^*, \ldots, p_K^*]$ minimize $f(\mathbf{p})$ with the constraints. Without loss of generality, we can consider that $\max_{k=1}^K p_k^* = p_1^*$. Then, we have

$$\max_{\substack{\sum_{k=1}^K p_k^2=1 \\ p_1,\ldots,p_K \geq 0}} f(\mathbf{p}) = \max_{\substack{\sum_{k=2}^K p_k^2=1-p_1^{*2} \\ p_2,\ldots,p_K \geq 0}} f(p_1^*, p_2, \ldots, p_K). \tag{4.110}$$

It is easy to show that maximization of the right-hand side of the above equation is attained when $p_2, \ldots, p_K$ are all equal: $p_2^* = \cdots = p_K^* \triangleq p^*$. From this and the constraint $\sum_{l=2}^{K} p_l^{*2} = 1 - p_1^{*2}$, we find that

$$p^* = \sqrt{\frac{1 - p_1^{*2}}{K-1}}. \tag{4.111}$$

Since $p_1^* \geq p^*$, inequality $p_1^* \geq 1/\sqrt{K}$ must hold. Substituting (4.111) into (4.110), we have

$$f(\mathbf{p}^*) = p_1^{*2} + \sqrt{K-1} p_1^* \sqrt{1 - p_1^{*2}}. \tag{4.112}$$

This function must take the maximum value as a function of $p_1^* \left(\geq 1/\sqrt{K} \right)$. Solving $df/dp_1^* = 0$, we obtain

$$p_1^* = \sqrt{\frac{1}{2}\left(1 + \frac{1}{\sqrt{K}}\right)}. \tag{4.113}$$

Substituting (4.113) into (4.112), we obtain

$$f(\mathbf{p}^*) = \frac{1}{2}\left(1 + \sqrt{K}\right). \tag{4.114}$$

Note that $df/dp_1^* = 0$ leads to another solution

$$p_1^* = \sqrt{\frac{1}{2}\left(1 - \frac{1}{\sqrt{K}}\right)},$$

but it is not the maximum solution.

5 VARIANTS OF INDEPENDENT VECTOR ANALYSIS ALGORITHMS

5.1 INTRODUCTION

This chapter proposes several IVA algorithms based on the gradient descent method. These algrithms incoporate different constraint into the separator, e.g. orthogonal, nonholonomic, and the linear constraints. The effectiveness of the algorithms will be confirmed in next chapter.

5.2 VARIANTS OF IVA ALGORITHMS

In this section, we revisit some IVA algorithms based on maximization of maximum likelihood or minimization of the Kullback-Leibler divergence even though their derivations have been already presented in papers [24, 35, 36]. Then, we propose an IVA algorithm incorporating the linear constraint.

5.2.1 *Contrast function*

In order to write a loss function, we initially use the mapping $\mathbb{C}^K \to \mathbb{R}^{2K}$ such that

$$\begin{aligned}\mathbf{y}_{i,\mathbb{R}} = \begin{bmatrix} \overline{\mathbf{y}}_i \\ \underline{\mathbf{y}}_i \end{bmatrix} &= \sum_{j=1}^{N} \begin{bmatrix} \overline{\mathbf{W}}_{ij} & -\underline{\mathbf{W}}_{ij} \\ \underline{\mathbf{W}}_{ij} & \overline{\mathbf{W}}_{ij} \end{bmatrix} \begin{bmatrix} \overline{\mathbf{x}}_j \\ \underline{\mathbf{x}}_j \end{bmatrix} \\ &= \sum_{j=1}^{N} \mathbf{W}_{ij,\mathbb{R}} \mathbf{x}_{j,\mathbb{R}} \ \ (i = 1, \ldots, N),\end{aligned}$$

where $\overline{\mathbf{y}}_i = \left[\overline{y}_i^{(1)}, \ldots, \overline{y}_i^{(K)}\right]^\top \in \mathbb{R}^K$, $\underline{\mathbf{y}}_i = \left[\underline{y}_i^{(1)}, \ldots, \underline{y}_i^{(K)}\right]^\top \in \mathbb{R}^K$ and $\mathbf{W}_{ij,\mathbb{R}}$ is a $2K \times 2K$ matrix.

Most learning algorithms are based on minimization or maximization of a loss function. Maximization of log-likelihood or minimization of the Kullback-Leibler divergence lead to the same type of loss function,

$$l\left(\mathbf{y}\left(t\right), \mathbf{W}\right) = -\log\left|\det \mathbf{W}_{\mathbb{R}}\right| - \sum_{i=1}^{N} \log\left(p_i\left(\mathbf{y}_{i,\mathbb{R}}\left(t\right)\right)\right) \qquad (5.1)$$

where p_i is probability density function (pdf) of output vector $\mathbf{y}_{i,\mathbb{R}}(t)$ and $|\det \mathbf{W}_{\mathbb{R}}|$ means the absolute value of determinant of matrix $\mathbf{W}_{\mathbb{R}}$. We can then apply the stochastic gradient descent method to derive the algorithm.

The first-order total differential of Eq. (5.1) yields

$$dl = -\text{tr}\left(d\mathbf{W}_{\mathbb{R}}\mathbf{W}_{\mathbb{R}}^{-1}\right) + \sum_{i=1}^{N} \boldsymbol{\psi}_i^{\top}\left(\mathbf{y}_{i,\mathbb{R}}(t)\right) d\mathbf{y}_{i,\mathbb{R}}(t), \tag{5.2}$$

where

$$\boldsymbol{\psi}_i\left(\mathbf{y}_{i,\mathbb{R}}(t)\right) = -\begin{bmatrix} \frac{\partial \log p_i(\mathbf{y}_{i,R}(t))}{\partial \overline{\mathbf{y}}_i} \\ \frac{\partial \log p_i(\mathbf{y}_{i,R}(t))}{\partial \underline{\mathbf{y}}_i} \end{bmatrix} \in \mathbb{R}^{2K}. \tag{5.3}$$

To express the second term on the right-hand side of Eq. (5.2) by simple notations, we write $\log p_i\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) = \log p_i\left(\mathbf{y}_{i,R}(t)\right)$, and using the relation described in Chap. 2, Eq. (5.3) becomes

$$\begin{aligned} \begin{bmatrix} \boldsymbol{\psi}_i\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) \\ \boldsymbol{\psi}_i^*\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) \end{bmatrix} &= -\begin{bmatrix} \frac{\partial \log p_i\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right)}{\partial \mathbf{y}} \\ \frac{\partial \log p_i\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right)}{\partial \mathbf{y}^*} \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} \mathbf{I} & -j\mathbf{I} \\ \mathbf{I} & j\mathbf{I} \end{bmatrix}\begin{bmatrix} \frac{\partial \log p_i(\mathbf{y}_{i,R}(t))}{\partial \overline{\mathbf{y}}_i} \\ \frac{\partial \log p_i(\mathbf{y}_{i,R}(t))}{\partial \underline{\mathbf{y}}_i} \end{bmatrix} \end{aligned} \tag{5.4}$$

As a result, the second term on the right-hand sider of Eq. (5.2) can be written as

$$\begin{aligned} &\boldsymbol{\psi}_i^{\top}\left(\mathbf{y}_{i,\mathbb{R}}(t)\right) d\mathbf{y}_{i,\mathbb{R}}(t) \\ &= \boldsymbol{\psi}_i^{\top}\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) d\mathbf{y}_i(t) + \boldsymbol{\psi}_i^{H}\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) d\mathbf{y}_i^*(t). \end{aligned} \tag{5.5}$$

From $\mathbf{y}_{\mathbb{R}} = \mathbf{W}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}}$ and $d\mathbf{y}_{\mathbb{R}} = d\mathbf{W}_{\mathbb{R}}\mathbf{x}_{\mathbb{R}} = d\mathbf{W}_{\mathbb{R}}\mathbf{W}_{\mathbb{R}}^{-1}\mathbf{y}_{\mathbb{R}}$, we put $d\mathbf{V}_{\mathbb{R}} = d\mathbf{W}_{\mathbb{R}}\mathbf{W}_{\mathbb{R}}^{-1}$, where

$$d\mathbf{V}_{ij,\mathbb{R}} = \begin{bmatrix} d\overline{\mathbf{V}}_{ij} & -d\underline{\mathbf{V}}_{ij} \\ d\underline{\mathbf{V}}_{ij} & d\overline{\mathbf{V}}_{ij} \end{bmatrix} \in \mathbb{R}^{2K\times 2K}.$$

The first term on the right-hand side of Eq. (5.2) can be rewritten as

$$\begin{aligned} \text{tr}\left\{d\mathbf{V}_{\mathbb{R}}\right\} &= \sum_{i=1}^{N} \text{tr}\left\{d\mathbf{V}_{ii,\mathbb{R}}\right\} = 2\sum_{i,k} d\overline{v}_{ii}^{(k)} \\ &= \sum_{i,k} dv_{ii}^{(k)} + dv_{ii}^{(k)*} \\ &= \text{tr}\left\{d\mathbf{V}\right\} + \text{tr}\left\{d\mathbf{V}^*\right\}. \end{aligned} \tag{5.6}$$

The first-order total differential of Eq. (5.1) is given by

$$\begin{aligned} dl = &-\text{tr}d\mathbf{V} - \text{tr}d\mathbf{V}^* + \sum_{i,j} \boldsymbol{\psi}_i^{\top}\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) d\mathbf{V}_{ij}\mathbf{y}_j(t) \\ &+ \sum_{i,j} \boldsymbol{\psi}_i^{H}\left(\mathbf{y}_i(t), \mathbf{y}_i^*(t)\right) d\mathbf{V}_{ij}^*\mathbf{y}_j^*(t). \end{aligned} \tag{5.7}$$

5.2.2 *Unconstrained IVA*

As is described in Chap. 2.2, the gradient for $d\mathbf{V}^{(k)}$ are given by

$$\triangle\mathbf{V}^{(k)}(t) = -\eta\frac{\partial dl}{\partial d\mathbf{V}^{(k)*}} = \eta\left\{\mathbf{I} - \boldsymbol{\varphi}^{(k)*}(\mathbf{y}(t), \mathbf{y}^*(t))\,\mathbf{y}^H(t)\right\} \quad (5.8)$$

where η is a learning rate and

$$\boldsymbol{\varphi}^{(k)}(\mathbf{y}(t), \mathbf{y}^*(t)) = -\begin{bmatrix} \frac{\partial \log p_1(\mathbf{y}_1(t), \mathbf{y}_1^*(t))}{\partial y_1^{(k)}} \\ \frac{\partial \log p_2(\mathbf{y}_2(t), \mathbf{y}_2^*(t))}{\partial y_2^{(k)}} \\ \vdots \\ \frac{\partial \log p_N(\mathbf{y}_N(t), \mathbf{y}_N^*(t))}{\partial y_N^{(k)}} \end{bmatrix}.$$

For the sake of simplicity of notation, define

$$\mathbf{y}_{\mathbb{C}}^{(k)}(t) = \begin{bmatrix} \mathbf{y}^{(k)}(t) \\ \mathbf{y}^{(k)*}(t) \end{bmatrix} \in \mathbb{C}^{2N},$$

then $\boldsymbol{\varphi}^{(k)}(\mathbf{y}(t), \mathbf{y}^*(t))$ can be rewritten as $\boldsymbol{\varphi}^{(k)}\left(\mathbf{y}_{\mathbb{C}(t)}\right)$ Using $\triangle\mathbf{V}^{(k)}(t) = \triangle\mathbf{W}^{(k)}(t)\,\mathbf{W}^{(k)-1}(t)$ leads to

$$\triangle\mathbf{W}^{(k)}(t) = \eta\left\{\mathbf{I} - \boldsymbol{\varphi}^{(k)*}(\mathbf{y}_{\mathbb{C}}(t))\,\mathbf{y}^{(k)H}(t)\right\}\mathbf{W}^{(k)} \quad (k = 1, \ldots, K). \tag{5.9}$$

The batch version of the algorithm is already proposed in [24, 35].

The stationary points of the unconstrained IVA algorithm satisfy

$$E\left[\boldsymbol{\varphi}^{(k)*}(\mathbf{y}_{\mathbb{C}}(t))\,\mathbf{y}^{(k)H}(t)\right] = \mathbf{I} \quad (k = 1, \ldots, K). \tag{5.10}$$

Namely, this algorithm forces the outputs to have constant magnitude. This normalization is not suitable for non-stationary sources such as speech, and it might be a problem such that, if a source component at a sensor becomes very small, some entries in the separator may tend to be very large, which can induce some instability.

5.2.3 *IVA with orthogonal constraint*

As in the case of ICA, IVA often applies the whitening operation to the observed signals and impose orthogonal (or unitary) constraint on the separator in order to eliminate scaling indeterminacy. The separator with orthogonal constraint satisfies

$$\mathbf{W}^{(k)H}\mathbf{W}^{(k)} = \mathbf{W}^{(k)}\mathbf{W}^{(k)H} = \mathbf{I} \;\; (k = 1, \ldots, K). \tag{5.11}$$

If the separators $\mathbf{W}^{(k)}$ are orthogonal, then their updates $\triangle\mathbf{W}^{(k)}\mathbf{W}^{(k)H}$ becomes skew-Hermitian matrices in order to preserve the orthogonal constraint. It is because Eq. (5.11) and using $\mathbf{W}^{(k)} + \triangle\mathbf{W}^{(k)}$ leads to

$$\triangle\mathbf{W}^{(k)}\mathbf{W}^{(k)H} + \left(\triangle\mathbf{W}^{(k)}\mathbf{W}^{(k)H}\right)^H = \mathbf{O} \;\; (k = 1, \ldots, K), \tag{5.12}$$

or

$$\triangle \mathbf{V}^{(k)} + \triangle \mathbf{V}^{(k)H} = \mathbf{O} \;\; (k = 1, \ldots, K) . \tag{5.13}$$

Thus, the steepest direction preserving the orthogonal constraint is obtained by projecting the gradient onto the space of skew-symmetric matrices. The unconstrained IVA algorithm given in (5.8) (or 5.9) changes to

$$\triangle \mathbf{V}^{(k)} = \eta \left\{ \frac{\mathbf{y}(t)\, \boldsymbol{\varphi}^{(k)\top}(\mathbf{y}_{\mathbb{C}}(t)) - \boldsymbol{\varphi}^{(k)*}(\mathbf{y}_{\mathbb{C}}(t))\, \mathbf{y}^{H}(t)}{2} \right\} \tag{5.14}$$

or

$$\triangle \mathbf{W}^{(k)} = \eta \left\{ \frac{\mathbf{y}(t)\, \boldsymbol{\varphi}^{(k)\top}(\mathbf{y}_{\mathbb{C}}(t)) - \boldsymbol{\varphi}^{(k)*}(\mathbf{y}_{\mathbb{C}}(t))\, \mathbf{y}^{H}(t)}{2} \right\} \mathbf{W}^{(k)} . \tag{5.15}$$

After updating $\mathbf{W}^{(k)}$, we should further rectify the separator so as to satisfy the orthogonal constraint. This manipulation may apparently seem unnecessary, because the algorithm (5.15) have already satisfied the orthogonal constraint. However, it should be incorporated to avoid the accumulation of small round-off errors. For each k, the separator is rectified by

$$\mathbf{W}^{(k)} \leftarrow \left(\mathbf{W}^{(k)} \mathbf{W}^{(k)H} \right)^{-\frac{1}{2}} \mathbf{W}^{(k)} \;\; (k = 1, \ldots, K) .$$

5.2.4 *IVA with nonholonomic constraint*

As is described in subsection 5.2.2, the unconstrained IVA algorithm forces the magnitude of the outputs of the separator. Hence, it is desirable to add a constraint to the algorithm so that the resulting algorithm does not control the magnitude of output signals. In order to avoid the drawback of the unconstrained IVA, we apply nonholonomic constraint to the unconstrained IVA as in the case of ICA [3, 15]. Thus, the unconstrained IVA given in (5.9) changes to

$$\triangle \mathbf{W}^{(k)} = \eta \left\{ \boldsymbol{\Lambda}^{(k)}(t) - \boldsymbol{\varphi}^{(k)*}(\mathbf{y}_{\mathbb{C}}(t))\, \mathbf{y}^{H}(t) \right\} \mathbf{W}^{(k)} \tag{5.16}$$

where

$$\boldsymbol{\Lambda}^{(k)} = \mathrm{diag} \left\{ \varphi^{(k)*}(\mathbf{y}_{1,\mathbb{C}}(t))\, y_1^{(k)*}(t), \ldots, \varphi^{(k)*}(\mathbf{y}_{N,\mathbb{C}}(t))\, y_N^{(k)*}(t) \right\} .$$

This algorithm was proposed in [36].

5.2.5 *IVA with linear constraint*

The algorithms described in subsections 5.2.2 and 5.2.3 force the magnitude of output signals to be constant. In this subsection, we propose

IVA algorithm incorporating the linear constraint described in Chapter 3 to preserve the signal quality without any post-processing.

Taking into account the linear constraint $\mathbf{e}^\top \mathbf{W}^{(k)} = \mathbf{f}^\top \ (k = 1, \ldots, K)$, the nonholonomic IVA becomes

$$\triangle \mathbf{W}^{(k)} = \eta \left\{ \boldsymbol{\Theta}^{(k)} - \boldsymbol{\varphi}^{(k)*} \left(\mathbf{y}_{\mathbb{C}} \left(t \right) \right) \mathbf{y}^{H} \left(t \right) \right\} \mathbf{W}^{(k)}, \qquad (5.17)$$

where

$$\boldsymbol{\Theta}^{(k)} = \operatorname{diag} \left\{ \mathbf{e}^\top \varphi^{(k)*} \left(\mathbf{y}_{\mathbb{C}} \left(t \right) \right) \mathbf{y}^{H} \left(t \right) \right\}.$$

Due to the linear constraint, the outputs of the separator become the source signals observed at first sensor, and hence the quality of signals is preserved[48, 72].

As in the case of IVA with orthogonal constraint, IVA with linear constraint also requires an additional manipulation in order to avoid the round-off error on actual computational calculation such that

$$\mathbf{W}^{(k)} \left(t + 1 \right) \leftarrow \left\{ \mathbf{I} - \frac{1}{N} \mathbf{e} \mathbf{e}^\top \right\} \mathbf{W}^{(k)} \left(t + 1 \right) + \frac{\mathbf{e} \mathbf{f}^\top}{N}.$$

5.3 SUMMARY

In this chapter, we have revisited IVA algorithm proposed in [24, 35, 36], and have proposed variants of the algorithm using different constraints such as orthogonal, and the linear constraints. The separation performance of those algorithms are confirmed in the next chapter.

6 SIMULATIONS

This chapter confirms the validity of the algorithms proposed in this thesis using synthetic data and speech signals.

6.1 SIMULATIONS WITH SYNTHETIC DATA

We conducted some simulations in which source signals are stationary and non-stationary signals (source signals are non-Gaussian, of course). Non-stationary source signals are generated as follows. If $u_i(t)$ is a non-Gaussian stationary signal, then the i-th non-stationary source signal is given by

$$s_i(t) = \eta_i(t)\, u_i(t) \;\; (i = 1, \ldots, N), \tag{6.1}$$

where

$$\eta_i(t) = 2\sin\left(\frac{\pi}{b_i} t\right) \;\; (i = 1, \ldots, N) \tag{6.2}$$

and b_i is a constant.

For the sake of simplicity, we used the following unimodular transfer functions as the mixing processes:

$$\mathbf{A}(z) = \begin{bmatrix} 1 + a^2 z^{-2} & a z^{-1} \\ a z^{-1} & 1 \end{bmatrix}$$

and

$$\mathbf{A}(z) = \begin{bmatrix} 1 + 2a^2 z^{-2} & a z^{-1} + a^2 z^{-2} & a z^{-1} \\ a z^{-1} + a^2 z^{-2} & 1 + a^2 z^{-2} & a z^{-1} \\ a z^{-1} & a z^{-1} & 1 \end{bmatrix}.$$

The separation accuracy is evaluated in terms of signal-to-interference ratio (SIR) improvement. In order to calculate SIR improvement, we need the individual source observations $s_{ki}(t)$ defined as

$$s_{ki}(t) = \sum_{\tau=0}^{\infty} a_{ki}(\tau)\, s_i(t - \tau), \tag{6.3}$$

where $s_{ki}(t)$ stands for the i-th source signal observed at the k-th sensor. Those observations are not available in the actual BSS procedure. The SIR improvement for the i-th output is calculated by $\text{SIR improvement}_i = \text{OSIR}_i - \text{ISIR}_i$ [dB]. These two types of SIRs are defined by the power ratio between the components related to the target sources and interference sources, at first sensor (because of the linear constraint) and the i-th output:

$$\mathrm{ISIR}_i = 10\log_{10}\frac{\sum_p \sum_k \left|s_{1i}^{(k)}(p)\right|^2}{\sum_p \sum_k \left|\sum_{j\neq i} s_{1j}^{(k)}(p)\right|^2}\ [\mathrm{dB}]\,, \tag{6.4}$$

$$\mathrm{OSIR}_i = 10\log_{10}\frac{\sum_p \sum_k \left|y_{ii}^{(k)}(p)\right|^2}{\sum_p \sum_k \left|\sum_{j\neq i} y_{ij}^{(k)}(p)\right|^2}\ [\mathrm{dB}]\,, \tag{6.5}$$

where $s_{1i}^{(k)}(p)$ is obtained by applying the short time Fourier transform (STFT) to $s_{1i}(t)$ in the p-th frame and k stands for frequency. $y_{ij}^{(k)}(p)$ is an s_j-component of $y_i^{(k)}(p)$ and is calculated by

$$y_{ij}^{(k)}(p) = \sum_{m=1}^{N} w_{im}^{(k)} s_{mj}^{(k)}(p) \quad (i = 1, \dots, N; j = 1, \dots, N)\,. \tag{6.6}$$

Since some algorithms do not satisfy the linear constraint, the obtained separator were modified to satisfy the linear constraint so that the results can be compared with the others methods. The modified separator can be obtained as

$$\mathbf{W}^{(k)} \leftarrow \mathrm{diag}\left(\mathbf{e}\mathbf{f}^\top \mathbf{W}^{(k)-1}\right)\mathbf{W}^{(k)} \quad (k = 1, \dots, K)\,. \tag{6.7}$$

The separation results are shown in Table 1 and 2. FDI-ND did not work well in separation of stationary signals, while it performed well in that of non-stationary signals. It is because the algorithm is designed for the non-stationary signals. IVA worked well for every case. The batch version of IVA with linear constraint showed relatively low performance due to the slow convergence.

In order to show convergence speed and stability of the iterative calculation, the convergence plots of algorithms in the case of two and three sources are shown in Figure 8 and 9. IVA achieved relatively fast convergence, while IVA with linear constraint showed slow convergence.

6.2 SIMULATIONS WITH SPEECH SIGNALS

We conducted some experiments on speech separation. In the center of a sound proof room, four microphones and some loudspeakers were arranged as shown in Fig. 10 (the reverberation time was approximately 100 ms). The combinations of sources and microphones used in the experiments were shown in Table 3. Four kinds of speeches consisting of two males and two females were used as source signals.

To evaluate separation peformance of an obtained separator, the following recording was made for each speech. Only the i-th loudspeaker was turned on, and the voice from it was recorded by the

Table 1: Separation performance of the algorithms in the case of $N = 2$. The averaged SIR improvements shown below are the maximum value of the averaged SIR improvement in iterative calculations.

Algorithm		Averaged SIRI [dB]	Iterations	Averaged SIRI [dB]	Iterations
		$a = 0.5$ (stationary)		$a = 0.5$ (non-stationary)	
FDI-ND		12.94	1e+4 (56)	21.80	1e+4 (1988)
IVA[35]	batch	14.18	1e+4 (1e+4)	**24.29**	1e+4 (9425)
	online	**18.55**	1e+4 **(30)**	23.93	1e+4 (6)
$\text{IVA}_{\text{unitary}}$	batch	12.00	1e+4 (1e+4)	16.54	1e+4 (1e+4)
	online	18.47	1e+4 (122)	24.22	1e+4 (34)
$\text{IVA}_{\text{nonholonomic}}$[36]	batch	12.62	1e+4 (1e+4)	24.26	1e+4 (1e+4)
	online	18.43	1e+4 (49)	24.09	1e+4 (428)
$\text{IVA}_{\text{linear}}$	batch	5.96	1e+4 (1e+4)	9.40	1e+4 (1e+4)
	online	15.56	1e+4 (1e+4)	20.50	1e+4 (1e+4)

Table 2: Separation performance of the algorithms in the case of $N = 3$. The averaged SIR improvements shown below are the maximum value of the averaged SIR improvement in iterative calculations.

Algorithm		Averaged SIRI [dB]	Iterations	Averaged SIRI [dB]	Iterations
		$a = 0.5$ (stationary)		$a = 0.5$ (non-stationary)	
FDI-ND		10.54	1e+4 (1)	**19.76**	1e+4 (82)
IVA[35]	batch	15.8	1e+4 (1e+4)	12.28	1e+4 (1e+4)
	online	20.90	1e+4 (33)	12.42	1e+4 (1744)
$\text{IVA}_{\text{unitary}}$	batch	12.85	1e+4 (1e+4)	11.88	1e+4 (1e+4)
	online	**20.92**	1e+4 (140)	12.50	1e+4 (1548)
$\text{IVA}_{\text{nonholonomic}}$[36]	batch	13.12	1e+4 (1e+4)	12.19	1e+4 (1e+4)
	online	20.90	1e+4 (55)	12.52	1e+4 (15)
$\text{IVA}_{\text{linear}}$	batch	10.09	1e+4 (1e+4)	10.52	1e+4 (1e+4)
	online	19.64	1e+4 (5990)	12.45	1e+4 (1e+4)

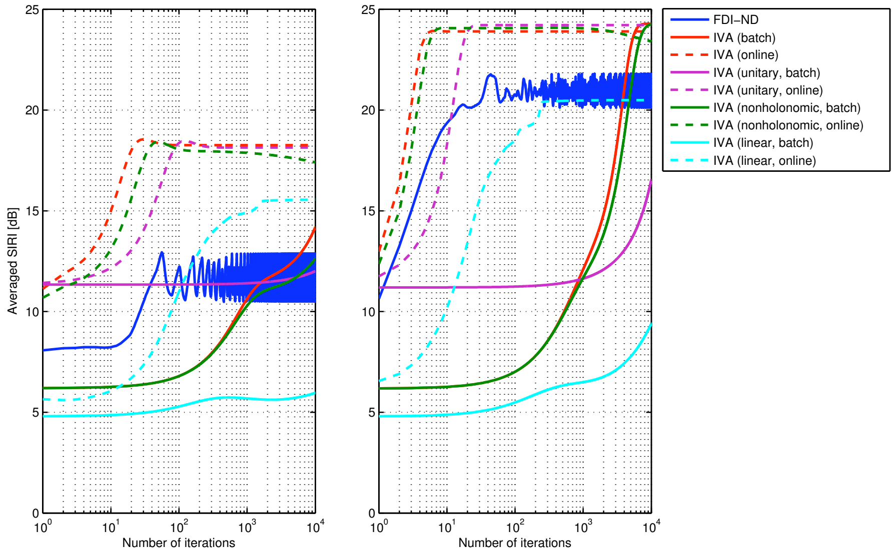

Figure 8: Convergence plot of the algorithms in the case of $N = 2$. Left: Source signals are stationary signals picked out from the Laplace distribution. Right: Source signals are non-stationary signals. Although the online version of IVA algorithms performed well in both cases, the batch version of IVA with linear constraint showed very slow convergence compared with the others.

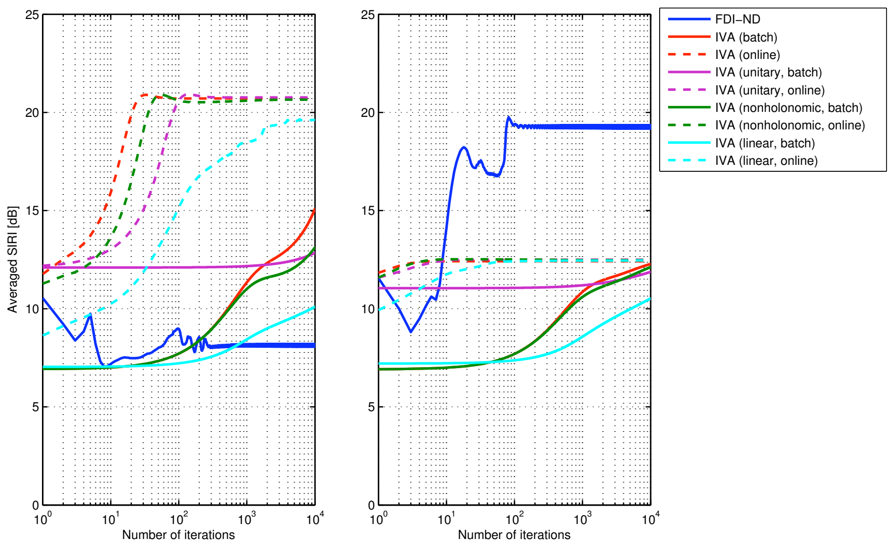

Figure 9: Convergence plot of the algorithms in the case of three sources. Left: Source signals are stationary signals picked out from the Laplace distribution. Right: Source signals are non-stationary signals. Although FDI-ND did not work well in the case of stationary signals, it showed accurate performance in the case of non-stationary signals.

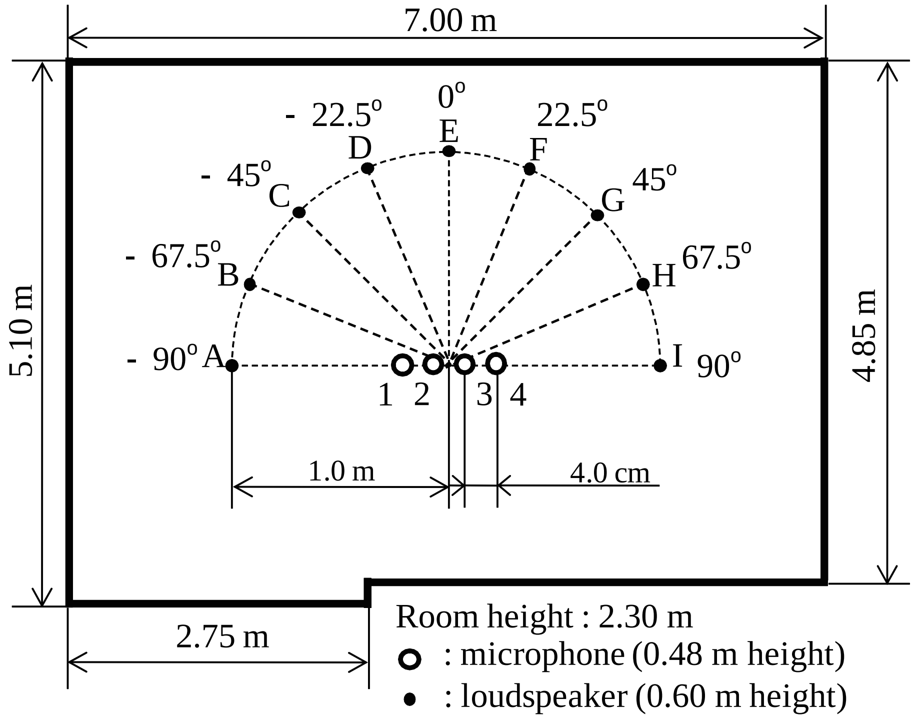

Figure 10: The experimental setup. Loudspeakers were arranged with the same radius 1.0 m and microphones are directly aligned at intervals of 4.0 cm in a sound proof room (the reverberation time was approximately 100 ms).

four microphones. $s_{ki}(t)$ denotes the i-th speaker's sound recorded by the k-th microphone. The summation of $s_{ki}(t)$ with respect to i was regarded as the sound that would be observed by the k-th microphone in an actual situation:

$$x_k(t) = \sum_{i=1}^{N} s_{ki}(t) \quad (k = 1, \ldots, N).$$

The recording was made for ten seconds at a sampling rate of 10 kHz. We chose the 2048-point FFT and 2048-point Hanning window with shift size of 512 samples. As in the case of synthetic data, the performance of the proposed algorithms were evaluated by the averaged SIR improvement.

We comoared the performance of the proposed methods with that of some algorithms: the frequency-domain implementation algorithm (FDIBSS) [57], and the natural gradient algorithm in the frequency domain (FDICA) [3] that uses a polar coordinate nonlinear function as an activation function [65]. FDICA solved the permutation problem using the inter-frequency correlation of signal envelopes[52]. In FDICA and IVA wtih unitary constraint, we applied the whitening operation to the observed data in each frequency bin, and the separating matrix was constrained to be an unitary matrix in each frequency bin. As

Table 3: The source and microphone locations in experiment. In experiment 1, for example, two sources located in A and I were recorded by microphones 2 and 3.

Exp. #	Source locations	Mics. used
1	A, I	2, 3
2	B, H	2, 3
3	C, G	2, 3
4	D, F	2, 3
5	A, C, G, I	all
6	B, D, F, H	all
7	C, D, E, F	all

Table 4: Separation performance of the algorithms. The averaged SIR improvements shown below are the maximum value of the averaged SIR improvement in iterative calculations. The number of iterations is 1000.

Algorithm	Exp. #	1	2	3	4	5	6	7
FDI-ND		16.26	**14.81**	7.85	8.55	7.24	12.77	7.85
FDIBSS[57]		**16.50**	13.97	9.35	12.17	4.41	5.56	6.16
FDICA[3]		4.38	3.13	3.42	4.37	5.11	5.56	4.29
IVA[35]	batch	1.25	2.59	1.56	1.82	2.22	1.70	1.40
	online	14.62	13.62	13.87	13.49	**11.69**	**13.81**	**11.41**
$\text{IVA}_{\text{unitary}}$	batch	1.23	1.35	1.10	1.26	2.4	2.21	1.94
	online	5.74	5.11	5.45	6.28	3.05	2.90	2.69
$\text{IVA}_{\text{nonholonomic}}$[36]	batch	2.27	2.59	1.56	1.82	2.33	1.68	1.29
	online	7.27	7.91	**17.12**	**16.05**	4.99	9.46	5.28
$\text{IVA}_{\text{linear}}$	batch	0.82	0.18	0.59	1.17	1.61	1.64	0.9
	online	8.17	11.74	13.23	12.73	4.51	7.33	9.0

in the case of synthetic data, the obtained separating matrices were modified by Eq. (6.7).

The separation results are shown in Table 4. While FDI-ND, FDIBSS and IVA performed well in the case of 2×2 convoluted mixtures, the other methods, especially batch version of IVA algorithms, showed poor separation performance. In the case of 4×4 convoluted mixtures, the online version of IVA algorithm and FDI-ND worked well compared with the other approaches.

7 CONCLUSIONS

The main objective of this thesis is to propose efficient algorithms for convolutive BSS. Time-domain convolutive BSS algorithms have high computational cost and sometimes shows numerical instability. The algorithm proposed in chapter 3 can relax the computational load compared with a time-domain algorithm, and it incorporates additional techniques to enhance the robustness of the proposed method. It works well in separation of non-stationary signals.

We have derived a necessary and sufficient condition for the desired separator to be obtained by minimizing a measure representing the difference between the prior source pdf and the actual output pdf of the separator. We also have proven that, if a desired separator minimizes the measure, any permuted separator never becomes a local minimum. This remarkable property strongly supports the validity of the IVA approach.

In addition we have investigated a necessary and sufficient condition under different assumptions such as uncorrelatedness, evenness and complex-valued case. If components in each source are uncorrelated, a necessary condition together with (B) becomes a necessary and sufficient condition for the desired separator. In the case of evenly distributed sources, condition (B) holds as a necessity; then the stability condition becomes a simple form. These assumptions are very important to realize a perfect separation by an IVA algorithm assuming a spherically symmetric source prior. The stability condition for the complex-valued IVA takes a more complicated form, but it is essentially the same as the condition for the real-valued IVA, under circularity and uncorrelatedness.

We have proposed IVA algorithms with several constraints such as unitary, nonholonomic, and the linear constraints. IVA with linear constraint can preserves the signal quality, and hence it is suitable for separation of speech signals. Since those algorithms are based on the natural gradient descent method, they show slow convergence. In order to realize real-time processing, it is necessary to apply the Newton method or another methods that can converge more faster.

BIBLIOGRAPHY

[1] R. Aichner, S. Araki, S. Makino, T. Nishikawa, H. Saruwatari, Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming, in: Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing, 2002, pp. 445–454.

[2] S. Amari, S. C. Douglas, A. Cichocki, H. H. Yang, Multichannel blind deconvolution and equalization using the natural gradient, in: Proceedings of IEEE International Workshop on Wireless Communication, 1997, pp. 101–104.

[3] S. Amari, T.–P. Chen, A. Cichocki, Nonholonomic orthogonal learning algorithms for blind source separation, Neural Computation 12 (6) (2000) 1463–1484.

[4] M. Anderson, X.–L. Li, T. Adali, Nonorthogonal independent vector analysis using multivariate Gaussian model, in: Proceedings of the 9th international conference on Latent variable analysis and signal separation, 2010, pp. 354–361.

[5] M. Anderson, X.–L. Li, T. Adali, Complex-valued independent vector analysis: Application to multivariate Gaussian model, Signal Processing (in press).

[6] J. Anemuller, B. Lollmeier, Amplitude modulation decorrelation for convolutive blind source separation, in: Proceedings of the Second International Conference on Independent Component Analysis and Blind Signal Separation, 2000, pp. 215–220.

[7] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, N. Kitawaki, Combined approach of array processing and independent component analysis for blind separation of acoustic signals, IEEE Trans. Audio and Speech Processing 11 (3) (2003) 204–215.

[8] A. D. Back, A. C. Tsoi, Blind deconvolution of signals using a complex recurrent network, in: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, 1994, pp. 565–574.

[9] E. Bingham, Advances in independent component analysis with applications to data mining, Ph. D. thesis, Helsinki University of Technology, 2003.

[10] A. van den Bos, Complex gradient and Hessian, IEE Proc. Vision Image and Signal Process. 141 (1994) 380–382.

[11] D. H. Brandwood, A complex gradient operator and its application in adaptive array theory, in: Proceedings of the IEE Communications, Radar and Signal Processing 130 (1) (1983) 11–16.

[12] H. Brehm, W. Stammler, Description and generation of spherically invariant speech-model signals, Signal Processing (12) (1987) 119–141.

[13] H. Buchner, R. Aichner, W. Kallermann, Blind source separation for convolutive mixtures: A unified treatment, In Y. Huang and J. Benesty (eds.), Audio Signal Processing for Next-Generation Multimedia Communication Systems, Kluwer Academic Publishers, 2004, pp. 255–293.

[14] V. D. Calhoun, T. Adali, G. D. Pearlson, J. J. Pekar, A method for making group inferences from functional MRI data using independent component analysis, Human Brain Mapping 14 (2001) 140–151.

[15] S. Choi, S. Amari, A. Cichocki, R. Liu, Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels, in: Proceedings of International Workshop on Independent Component Analysis and Blind Source Separation, 1999, pp. 371–376.

[16] A. Cichocki, Blind source separation: New tools for extraction of source signals and denoising, in: Proceedings of SPIE 2005, 2005, pp. 11–24.

[17] A. Cichocki, R. Zdunek, A.–H. Phan, S. Amari, Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis, John Wiley & Sons, 2009.

[18] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (3) (1994) 287–314.

[19] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Academic Press, 2010.

[20] S. Deligne, R. Gopinath, An EM algorithm for convolutive independent component analysis, Neurocomputing 49 (2002) 187–211.

[21] S. C. Douglas, M. Gupta, H. Sawada, S. Makino, Spatio-temporal FastICA algorithms for the blind separation of convolutive mixtures, IEEE Trans. Audio, Speech, and Language Processing 15 (5) (2007) 1511–1520.

[22] T. Eltoft, T. Kim, T.–W. Lee, On the multivariate Laplace distribution, IEEE Signal Processing Letters 13 (5) (2006) 300–303.

[23] J. Hao, I. Lee, T.–W. Lee, T. J. Sejnowski, Independent vector analysis for source separation using a mixture of Gaussians prior, Neural Computation 22 (6) (2010) 1646–1673.

[24] A. Hiroe, Solution of permutation problem in frequency domain ICA, using multivariate probability density functions, in: Lecture Notes in Computer Science, 2006, pp. 601–608.

[25] A. Hyvarinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, New York, 2001.

[26] M. Z. Ikram, D. R. Morgan, Exploring permutation inconsistency in blind source separation of speech signals in a reverberant environment, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2000, pp. 1041–1044.

[27] M. Z. Ikram, D. R. Morgan, A multiresolution approach to blind separation of speech signals in a reverberant environment, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001, pp. 2757–2760.

[28] M. Z. Ikram, D. R. Morgan, Permutation inconsistency in blind speech separation: investigation and solutions, IEEE Trans. Speech and Audio Processing 13 (1) (2005) 1–13.

[29] T. Itahashi, K. Matsuoka, Stability analysis on independent vector analysis, in: Proceedings of the 11th WSEAS international conference on Multimedia systems & signal processing, 2011, pp. 98–103.

[30] G.–J. Jang, I. Lee, T.–W. Lee, Independent vector analysis using non-spherical joint densities for the separa-

tion of speech signals, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, pp. 629–632.

[31] M. Joho, H. Mathis, R. H. Lambert, Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture, in: Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation, 2000, pp. 81–86.

[32] M. Joho, P. Schniter, Frequency domain realization of a multichannel blind deconvolution algorithm based on the natural gradient, in: Proceedings of ICA, 2003, pp. 543–548.

[33] M. Kawamoto, K. Matsuoka, N. Ohnishi, A method of blind separation for convolved non-stationary signals, Neurocomputing 22 (1998) 157–171.

[34] M. Kawamoto, Y. Inouye, Blind deconvolution of MIMO-FIR systems with colored inputs using second-order statistics, IEICE Trans. fundamentals E86-A (3) (2003).

[35] T. Kim, H. T. Attias, S.–Y. Lee, T.–W. Lee, Blind source separation exploiting higher-order frequency dependencies, IEEE Trans. Audio, Speech, and Language Processing 15 (1) (2007) 70–79.

[36] T. Kim, Real-time independent vector analysis for convolutive blind source separation, IEEE Trans. Circuits and Systems 57 (7) (2010) 1431–1438.

[37] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, F. Itakura, Evaluation of blind signal separation method using directivity pattern under reverberant conditions, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2000, pp. 3140–3143.

[38] R. H. Lambert, Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures, Ph.D. thesis, University of Southern California, 1996.

[39] R. H. Lambert, A. J. Bell, Blind separation of multiple speakers in a multipath environment, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1997, pp. 423–426.

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

[41] I. Lee, T.–W. Lee, On the assumption of spherical symmetry and sparseness for the frequency-domain speech model, IEEE Trans. Audio, Speech, and Language Processing 15 (5) (2007) 1521–1528.

[42] I. Lee, T. Kim, T.–W. Lee, Fast fixed-point independent vector analysis algorithms for convolutive blind source separation, Signal Processing 87 (2007) 1859–1871.

[43] J.–H. Lee, T.–W. Lee, F. A. Jolesz, S.–S. Yoo, Independent vector analysis: Multivariate approach for fMRI group study, NeuroImage 40 (2008) 86–109.

[44] T.–W. Lee, A. J. Bell, R. Orglmeister, Blind source separation of real world signals, in: Proceedings of IEEE International Conference on Neural Networks, 1997, pp. 2129–2135.

[45] H. Li, T. Adali, Algorithms for complex ML ICA and their stability analysis using Wirtinger calculus, IEEE Trans. Signal Processing 58 (12) (2010) 6156–6167.

[46] K. Matsuoka, S. Nakashima, Minimal distortion principle for blind source separation, in: Proceedings of the 41st SICE Annual Conference, 2002, pp. 2138–2143.

[47] K. Matsuoka, Y. Ohba, S. Isogai, Blind separation of sound sources in real-world situations, in: Proceedings of the 18th International Congress on Acoustics, 2004, pp. 81–84.

[48] K. Matsuoka, Elimination of filtering indeterminacy in blind source separation, Neurocomputing 71 (2008) 2113–2126.

[49] K. Matsuoka, T. Itahashi, A time-domain blind source separation by the Newton method, in: Proceedings of IEEE International Symposium on Circuits and Systems 2009, 2009, pp. 2790–2793.

[50] N. Mitianoudis, M. Davis, Audio source separation of convolutive mixtures, IEEE Trans. Audio and Speech Processing 11 (5) (2003) 489–497.

[51] T. Mei, F. Yin, J. F. Chicaro, Blind source separation based on time-domain optimization of frequency-domain independence criterion, IEEE Trans. Audio,

Speech, and Language Processing 14 (6) (2006) 2075–2085.

[52] N. Murata, S. Ikeda, A. Ziehe, An approach to blind source separation based on temporal structure of speech signals, Neurocomputing 41 (1–4) (2001) 1–24.

[53] F. Nesta, T. S. Wada, B.–H. Juang, Coherent spectral estimation for a robust solution of the permutation problem, in: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009, pp. 105–108.

[54] T. Nishikawa, Blind source separation based on multi-stage independent component analysis, Ph.D. thesis, Nara Institute of Science and Technology, 2005.

[55] P. D. O'Grady, B. A. Pearlmutter, S. T. Rickard, Survey of sparse and non-sparse methods in source separation, International Journal of Imaging Systems and Technology 15 (2005) 18–33.

[56] M. Oh, H.–M. Park, Blind source separation based on independent vector analysis using feed-forward network, Neuro computing (in press).

[57] M. Ohata, K. Matsuoka, Frequency-domain implementation of a time-domain blind separation algorithm for convolutive mixtures of sources, in: Proceedings of the 7th International Conference on Independent Component Analysis and Signal Separation, 2007, pp. 528–535.

[58] J. A. Palmer, K. Kreutz–Delgado, Q. Wang, S. Makeig, Mixture convolutive independent component analysis, Technical Report, UCSD, 2006.

[59] J. A. Palmer, K. Kreutz–Delgado, S. Makeig, Probabilistic formulation of independent vector analysis using complex Gaussian scale mixtures, in: Proceedings of the 8th International Conference on Independent Component Analysis and Signal Separation, 2009, pp. 90–97.

[60] J. A. Palmer, K. Kreutz–Delgado, S. Makeig, Strong sub- and super-gaussianity, in: Proceedings of the 9th International Conference on Latent Variable Analysis and Signal Separation, 2010, pp. 303–310.

[61] L. Parra, C. Spence, Convolutive blind separation of non-stationary sources, IEEE Trans. Speech and Audio Processing 8 (3) (2000) 320–327.

[62] D. T. Pham, C. Serviere, H. Boumaraf, Blind separation of convolutive audio mixtures using nonstationarity, in: Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation, 2003, pp. 981–986.

[63] B. Picinbono, On circularity, IEEE Trans. Signal Processing 42 (1994) 3474–3482.

[64] S. M. Rolfe, An independent component analysis based tool for exploring functional connections in the brain, master thesis, University of Washington, 2007.

[65] H. Sawada, R. Mukai, S. Araki, S. Makino, Polar coordinate based nonlinear function for frequency domain blind source separation, IEICE Trans. Fundamentals E86-A (3) (2003) 590–596.

[66] H. Sawada, R. Mukai, S. Araki, S. Makino, A robust and precise method for solving the permutation problem of frequency-domain blind source separation, IEEE Trans. Speech Audio Processing 12 (5) (2004) 530–538.

[67] P. Smaragdis, Blind separation of convolved mixtures in the frequency domain, Neurocomputing 22 (1998) 21–34.

[68] J. V. Stone, Independent component analysis: A tutorial introduction, The MIT Press, 2004.

[69] J. Thomas, Y. Deville, S. Hosseini, Time-domain fast fixed-point algorithms for convolutive ICA, IEEE Signal Processing Letters 13 (4) (2006) 228–231.

[70] J. Via, M. Anderson, X.–L. Li, T. Adali, A maximum likelihood approach for independent vector analysis of Gaussian data sets, in: Proceedings of the 2011 IEEE International Workshop on Machine Learning for Signal Processing, 2011, pp. 1–6.

[71] S. Winter, H. Sawada, S. Makino, Geometrical interpretation of the PCA subspace method for overdetermined blind source separation, in: Proceedings of 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), 2003, pp. 775–780.

[72] T. Yoshihara, A study on blind source separation and blind source extraction that preserve signal quality, Ph.D. thesis (in Japanese), Kyushu Institute of Technology, 2010.

[73] M. Ziblevsky, B. A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, Neural Computation 13 (2001) 863–882.