Complex-valued mixing matrix estimation for blind separation of acoustic mixtures

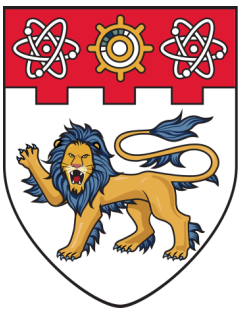
Nguyen, Anh Hai Trieu

2020

Nguyen, A. H. T. (2020). Complex-valued mixing matrix estimation for blind separation of acoustic mixtures. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/143962

https://doi.org/10.32657/10356/143962



NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Complex-valued mixing matrix estimation for blind separation of acoustic mixtures

Nguyen Hai Trieu Anh

School of Electrical & Electronic Engineering

2020

COMPLEX-VALUED MIXING MATRIX ESTIMATION FOR BLIND SEPARATION OF ACOUSTIC MIXTURES

NGUYEN HAI TRIEU ANH

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2020

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarized materials, and has not been submitted for a higher degree to any other University or Institution.

17/12/2019
Date

Nguyen Hai Trieu Anh

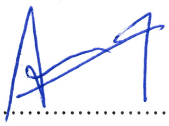
Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

17/12/2019

.....

Date



.....

A/Prof Andy W. H. Khong

Authorship Attribution Statement

This thesis contains material from three papers published in the following peer-reviewed journal / from papers accepted at conferences in which I am listed as the first author.

Chapter 4 is published in the papers *A. H. T. Nguyen, V. G. Reju, A. W. H. Khong, and I. Y. Soon, "Learning complex-valued latent filters with absolute cosine similarity," in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., 2017, pp. 2412–2416* and *A. H. T. Nguyen, V. G. Reju, and A. W. H. Khong, "Directional Sparse Filtering for Blind Estimation of Under-determined Complex-valued Mixing Matrices," submitted to IEEE Trans. Signal Process.*

The contributions of the co-authors are as follows:

- A/Prof Khong and A/Prof Soon provided the initial project direction and edited the manuscript drafts. They have also provided technical feedback through discussions.

- I prepared the manuscript drafts and the manuscripts were revised by Dr. Reju.

- I derived and implemented the proposed algorithm.

- I derived the re-parameterization technique to solve semi-unitary constrained optimization problems.

- I conducted the simulations and analyzed the results.

- I derived the concentration of i.i.d. complex generalized Gaussian sources about the columns of the mixing matrix.
- I derived the concentration of L_p radially symmetric sources about the columns of the mixing matrix.
- I derived the stability condition for learning on directional data.
- I conducted the simulations to verify the stability condition for several directional algorithms.

Chapter 5 is, in part, published in the paper *A. H. T. Nguyen, V. G. Reju, and A. W. H. Khong, "Directional Sparse Filtering for Blind Estimation of Underdetermined Complex-valued Mixing Matrices," submitted to IEEE Trans. Signal Process.*

- A/Prof Khong provided the initial project direction, provided technical feedback during meetings, and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscripts were revised by Dr. Reju.
- I derived the concentration of i. i. d. complex generalized Gaussian sources about columns of the mixing matrix.
- I derived the concentration of L_p radially symmetric sources about columns of the mixing matrix.
- I derived the stability condition for learning on directional data.
- I conducted the simulations to verify the stability condition.

Chapter 6 is published in the paper *A. H. T. Nguyen, V. G. Reju, and A. W. H. Khong, "A Method Based on L-BFGS to Solve Constrained Complex-valued*

ICA," in Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., 2019, pp. 4370–4374.

The contributions of the co-authors are as follows:

- A/Prof Khong provided the initial project direction, provided technical feedback during meetings, and edited the manuscript drafts.
- I prepared the manuscript drafts and they were revised by Dr. Reju.
- I derived and implemented the proposed algorithms.
- I conducted the simulations and analyzed the results.

Chapter 7 contains unpublished results.

17/12/2019 .....

Date

.....

Nguyen Hai Trieu Anh

Acknowledgments

I would like to use this opportunity to express my heartfelt gratitude to A/Prof Andy Khong for his patience and encouragement during my PhD journey. I would never reach this far without his guidance, inspiration, and support. I will always appreciate him. Secondly, I would like to thank Dr. Reju for spending a significant amount of time to revise my papers. I am also grateful to be mentored by him. Thirdly, I would like to thank my previous advisor Dr. Soon Ing Yann for his support. I kindly thank the team members – Hanh, Joshua, Zhiwei, and Karn – for their invaluable assistance. I am delightful and lucky to work with you. I also appreciate all other friends who have shared memorable moments with me during my journey – Wu Kai, Huizhi, Kelvin, Rohith, Phung, Trung Anh, Ngoc Khang, and Nguyen Le. Furthermore, I thank Nguyen Dzung for being the best thing happened to me. Last but not least, I would like to thank my parents and little sister for their unconditional love.

Contents

Statement of Originality iii

Supervisor Declaration Statement v

Authorship Attribution Statement vii

Acknowledgements xi

Summary xvii

List of Figures xix

List of Tables xxiii

List of Abbreviations xxv

List of Notations xxvii

1 Introduction 1

1.1 Motivation 1

1.2 Objectives 4

1.3 Main contribution of the Thesis 4

1.4 Organization of the Thesis 5

2 Mathematical preliminary 7

2.1 Introduction 7

2.2 Notation 7
2.3 Differentiation of complex-valued vector and matrix functions . . 8
2.4 Majorization, Schur-convexity, and Schur-concavity 11
2.5 Power mean inequality 12
2.6 Summary 13

3 Background **15**
3.1 Frequency-domain blind source separation 15
3.2 Complex-valued mixing matrix estimation 18
3.2.1 Whitening 19
3.2.2 Complex-valued mixing matrix estimation 20
3.2.2.1 Complex-valued independent component analysis 20
3.2.2.2 Sparse component analysis 24
3.2.2.3 Non-negative matrix factorization 30
3.2.3 Source separation 33
3.3 Performance criteria 35
3.4 Summary 37

4 Directional Sparse Filtering **39**
4.1 Introduction 39
4.2 The proposed DSF-based method 40
4.2.1 Problem formulation 40
4.2.2 The sparse penalty criteria for directional data 42
4.2.3 The proposed sparse penalty for directional data 44
4.2.4 Unconstrained re-parameterization for semi-unitary constrained optimization problems 47
4.2.5 Implementation of the proposed DSF algorithm 54
4.3 Simulation results 55
4.3.1 Simulation results for synthetic data 55

4.3.2 Simulation results for blind separation of under-determined convolutive speech mixtures . . . 58
4.4 Summary . . . 62

5 Stability Condition for Directional Methods 65
5.1 Introduction . . . 65
5.2 The L_2-normalization of i.i.d. generalized Gaussian mixtures . . . 66
5.3 The L_2-normalization of L_p radially symmetric mixtures . . . 69
5.4 Stability condition . . . 70
5.5 Stability of several directional algorithms . . . 77
5.5.1 The stability of K-hyperlines . . . 78
5.5.2 The stability of sparse filtering . . . 78
5.5.3 The stability of the directional sparse filtering . . . 80
5.6 Summary . . . 82

6 A Method Based on L-BFGS to Solve Complex-valued ICA by Entropy Bound Minimization 83
6.1 Introduction . . . 83
6.2 Complex-valued ICA by entropy bound minimization . . . 84
6.3 CEBM based on L-BFGS . . . 86
6.4 Simulations . . . 89
6.5 Summary . . . 93

7 Simultaneous Estimation of the Temporal Whitening Filters and the Mixing Matrix in Complex-valued ICA by Entropy Rate Bound Minimization 95
7.1 Introduction . . . 95
7.2 The proposed CERBM+ algorithm . . . 100
7.3 Simulations . . . 102
7.3.1 Simulation results for synthetic data . . . 102

7.3.2 Simulation results for speech separation 106
7.4 Summary 110

8 Conclusion and Recommendations **113**
8.1 Conclusion 113
8.2 Recommendations for further research 115

Author's Publications **117**

Bibliography **119**

Summary

This thesis concerns several aspects of complex-valued instantaneous mixing matrix estimation. While the main application of the work is to extract individual speech signal from group-based conversations in real-world environments, the methods developed in this thesis can be applied to other types of complex-valued instantaneous mixtures. The thesis includes several main contributions: improvement of estimation performance of sparse component analysis in under-determined case where the number of sensors is less than the number of sources, the connection between sparse component analysis and independent component analysis in determined case, and improvement of estimation performance of two independent component analysis methods in determined case where the number of sensors is equal to the number of sources.

For under-determined blind source separation, mixture models of directional statistics are often employed to estimate the mixing matrix. These mixture models commonly adopt the Gaussian function to model the distribution of the observed data about a specific direction. Since the Gaussian function may not be well-suited for sparse signals, the performance of mixture models of directional statistics is limited. Furthermore, solving mixture models of directional statistics is computationally costly due to the computation of spatial covariance matrices. These issues motivated the development of directional sparse filtering (DSF) for complex-valued mixing matrix estimation. Using power mean of the magnitude-squared cosine distances, directional sparse filtering searches for a set

of directions such that each data sample is expressed by only a few vectors of the set. Simulations on synthetic data and real mixtures of recorded speeches indicate that the proposed directional sparse filtering outperforms directional clustering baselines while having lower computational complexity.

Next, a stability condition for invertible mixing matrix estimation using multi-variate non-separable contrast functions is derived. The stability condition is verified via Monte-Carlo simulation for several directional clustering methods and directional sparse filtering algorithm, given that the sources are super-Gaussian. The stability condition implies that the minima of the cost functions corresponding to these sparse component analysis methods include the mixing matrix in a similar manner to that of independent component analysis. This result shows that sparse component analysis methods based on directional clustering can be interpreted as independent component analysis using quasi-maximum likelihood estimation.

For determined blind source separation, two re-parameterizations are proposed to convert the constrained independent component analysis by entropy bound minimization into unconstrained optimization problems. After that, these unconstrained optimization problems are solved via quasi-Newton methods. Experiment results show that the proposed method yields higher estimation performance than the baseline algorithm, particularly when the source distributions are close to Gaussian distribution. The re-parameterizations for independent component analysis by entropy bound minimization are further extended to estimate the mixing matrix and the temporal whitening filters simultaneously. This results in a new algorithm for independent component analysis by entropy rate bound minimization. Compared to the conventional independent component analysis by entropy rate bound minimization, the new method shows improvement in terms of mixing matrix estimation and speech separation.

List of Figures

3.1 Approximation of convolutive speech mixtures by instantaneous complex-valued linear mixing model in time-frequency domain. . . 16

3.2 Block diagram of frequency-domain blind source separation. . . . 18

3.3 Block diagram of a complex-valued blind source separation algorithm. 19

3.4 The distribution of horizontal and vertical projection of two female speech signals. (a) Unmixed sources. (b) Mixed signals. . . . 20

3.5 Scatter plot of two under-determined mixtures of three sub-Laplace sources. Red, green, and blue lines indicate the directions of the columns of the mixing matrix. 25

4.1 The values of L_1 norm and the proposed penalty as functions of θ for an arbitrary data vector $\mathbf{x} = [\cos\theta,\ \sin\theta]^T$ when $\mathbf{A} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$. 46

4.2 Variation of the average MER with: a) sample size for the estimation of a 2×4 mixing matrix, b) sparseness for the estimation of a 2×4 mixing matrix, c) number of sources for $M = 2$. Variation of the average CPU time as a function of: d) sample size for the estimation of a 2×4 mixing matrix, e) sparseness for the estimation of a 2×4 mixing matrix, f) number of sources for $M = 2$. 57

4.3 Variation of the average MER with: a) sample size for the estimation of a 16×20 mixing matrix, b) sparseness for the estimation of a 16×20 mixing matrix, c) number of sources for $M = 16$. Variation of the average CPU time as a function of: d) sample size for the estimation of a 16×20 mixing matrix, e) sparseness for the estimation of a 16×20 mixing matrix, f) number of sources for $M = 16$. 58

5.1 Directional term of the marginal distribution of L_2-normalized mixtures. The black arrows denote the columns of the mixing matrix. Left: super-Gaussian CGGD mixtures ($p = 1$). Right: sub-Gaussian CGGD mixtures ($p = 3$). 68

5.2 Stability condition of sparse filtering. 79

5.3 The stability condition for the proposed contrast function. 80

6.1 Average NAI and failure rate as functions of the sample size. . . . 91

6.2 Average NAI and failure rate as functions of the number of sources. 92

6.3 Average NAI and failure rate as functions of the shape parameter ($p = 2$ is Gaussian). 93

7.1 Average NAI and CPU time as functions of the number of sources. 103

7.2 Average NAI and CPU time as functions of the number of sources. 104

7.3 Average NAI and CPU time as functions of the number of sources. 105

7.4 Locations of speech sources and microphones. 107

7.5 Average SDR and ESTOI over all environments with respect to the order of the whitening filters K for blind separation of two speech sources from two convolutive mixtures. 108

7.6 Separation performance in SDR for blind separation of two speech sources from two convolutive mixtures in: a) All environments. b) Lecture theater. c) Office. d) Outdoor. 109

7.7 Separation performance in ESTOI for blind separation of two speech sources from two convolutive mixtures in: a) All environments. b) Lecture theater. c) Office. d) Outdoor. 110

List of Tables

3.1 Computational complexity per iteration for several directional clustering algorithms. 30

4.1 Speech separation performance of DSF on 2mic_4src subset of SiSEC dev1 dataset. 61

6.1 Complex-valued ICA by entropy bound minimization based on L-BFGS. 89

7.1 Conventional complex-valued ICA by entropy rate bound minimization. 99

7.2 Complex-valued ICA by entropy rate bound minimization based on L-BFGS. 102

List of Abbreviations

ACGMM	Angular central Gaussian mixture model
BSS	Blind source separation
CEBM	Complex-valued ICA by entropy bound minimization
CEBM+	The proposed CEBM based on L-BFGS
CERBM	Complex-valued ICA by entropy rate bound minimization
CERBM+	The proposed CERBM based on L-BFGS
CFASTICA	Complex-valued FastICA
CGGD	Complex generalized Gaussian distribution
CMN	Complex-valued ICA by maximization of non-Gaussianity
DSF	The proposed directional sparse filtering
ESTOI	Extended short-time objective intelligibility
FOCUSS	Focal under-determined system solution
GMM	Gaussian mixture model
ICA	Independent component analysis
IVA	Independent vector analysis
ILRMA	Independent low-rank matrix analysis
JADE	Joint approximate diagonalization of eigen-matrices
KHL	K-hyperlines
KSP	K-subspaces

L-BFGS	Limited-memory Broyden-Fletcher-Goldfarb-Shanno
MDP	Minimum distortion principle
MEP	Maximum entropy principle
MER	Mixing error ratio
NAI	Normalized Amari index
NMF	Non-negative matrix factorization
RISS	Rotation invariant and sparse sensitivity
SDR	Signal distortion ratio
SF	Sparse filtering
SIR	Signal interference ratio
SiSEC	Signal separation evaluation campaign
SNR	Signal noise ratio

List of Notations

$\jmath$	The imaginary unit ($\sqrt{-1}$)
$(\cdot)^*$	Conjugation
$(\cdot)^H$	Conjugate transposition
$(\cdot)^T$	Transposition
$\Im\{\cdot\}$	Imaginary part
$\Re\{\cdot\}$	Real part
$\mathcal{E}\{\cdot\}$	Expectation
$\odot$	Element-wise multiplication
$\oslash$	Element-wise division
$\otimes$	Kronecker product
$\mathbb{C}$	Set of complex numbers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of non-negative real numbers
$\angle z$	Phase of z
$\lvert z \rvert$	Modulus of z

z^*	Complex conjugate of z
$\|\mathbf{z}\|_p$	L_p-norm of a vector $\mathbf{z}$
$\mathbf{e}_n$	The nth standard basis vector
$\mathcal{B}_p^N$	Unit L_p-norm ball in an N-dimensional complex vector space
$\mathcal{S}_p^N$	Surface of $\mathcal{B}_p^N$
$\mathbf{I}_N$	Indentity matrix in $\mathbb{C}^{N\times N}$
$\mathbf{K}_{MN}$	The commutation matrix in $\mathbb{C}^{M\times N}$
$\mathcal{D}_{\mathbf{Z}^*}\mathbf{F}$	The Jacobian matrix of $\mathbf{F}$ w.r.t. $\mathbf{Z}^*$
$\mathcal{D}_{\mathbf{Z}}\mathbf{F}$	The Jacobian matrix of $\mathbf{F}$ w.r.t. $\mathbf{Z}$
Ddiag($\mathbf{Z}$)	The matrix given by Diag(diag($\mathbf{Z}$))
Diag($\mathbf{z}$)	The diagonal matrix where diagonal elements are specified by a vector $\mathbf{z}$
diag($\mathbf{Z}$)	The vector formed by diagonal elements of a matrix $\mathbf{Z}$
vec($\mathbf{Z}$)	Vectorization of $\mathbf{Z}$
$\nabla_{\mathbf{Z}^*} g$	The gradient of g
Δ_{N-1}	Standard simplex in an N-dimensional space
$\boldsymbol{\theta} \prec \boldsymbol{\phi}$	The vector $\boldsymbol{\theta}$ is majorized by a vector $\boldsymbol{\phi}$
F	The number of frequency bins
M	The number of input channels
N	The number of sources
T	The number of time frames

τ	Time-frame index
f	Frequency-bin index
t	The time index
$\mathbf{x}$	The observed data vector
x_m	The mth observed signal
$\bar{\mathbf{x}}$	L_2-normalization of $\mathbf{x}$
$\mathbf{s}$	The source vector
s_n	The nth source signal
$\bar{\mathbf{s}}$	L_2-normalization of $\mathbf{s}$
$\bar{s}_n$	The nth element of $\bar{\mathbf{s}}$
$\widehat{\mathbf{s}}$	Estimate of $\mathbf{s}$
$\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$	The magnitude-squared cosine-similarity vector
$\mathbf{u}_0$	Element-wise magnitude-square vector of $\bar{\mathbf{s}}$
$\mathbf{y}$	The demixed vector
y_n	The nth demixed signal
$\mathbf{Q}$	Whitening matrix
$\mathbf{A}$	The mixing matrix
$\mathbf{a}_n$	The nth column of $\mathbf{A}$
$\widehat{\mathbf{A}}$	The estimated mixing matrix
$\widehat{\mathbf{a}}_n$	The nth column of $\widehat{\mathbf{A}}$

$\mathbf{W}$	The demixing matrix
$\mathbf{w}_n$	The nth column of $\mathbf{W}$
$\mathcal{H}_r$	The entropy rate
$\mathcal{I}$	Mutual information
$\mathcal{I}_r$	The mutual information rate
$H(z)$	Entropy of z
ϕ	The hyperparameter for Dirichlet prior of GMM
ρ	Circularity coefficient

Chapter 1

Introduction

1.1 Motivation

Blind source separation (BSS) of convolutive mixtures aims to recover latent sources from observed mixtures without knowledge of the mixing filters. It has a wide range of applications across different fields, particularly in biomedical signal processing [1, 2] and blind separation of convolutive audio mixtures [3–5]. Two of the most popular convolutive BSS methods are complex-valued independent component analysis (ICA) [6–12] and sparse component analysis (SCA) [13–20], where they involve complex-valued mixing matrix estimation followed by source reconstruction.

To perform ICA, one commonly seeks a set of demixing filters such that the filtered signals are independent. This can be achieved by the use of non-linear functions to optimize for one of several related objectives such as the maximization of non-Gaussianity, minimization of negentropy, maximum likelihood, and minimization of mutual information [7, 9, 10, 12, 21, 22]. Under this framework, the performance of an ICA algorithm depends largely on whether its non-linearity matches the probability density function of the sources. Once the mixing matrix is estimated, the latent source signals can be reconstructed

by solving linear matrix equation [23]. Given a fixed temporal dependency in the data, the ICA cost function becomes increasingly flat near the true mixing matrix when the source distributions approach Gaussian distribution. It is therefore attractive to employ second-order optimization methods which exploit the curvature of the cost functions to mitigate issues associated with plateau regions. Indeed, fixed-point iteration, a quasi-Newton method, is the most common implementation of ICA [6]. However, the Hessian approximation in fixed-point ICA techniques is restrictive and may not reflect the true curvature of ICA objectives well. For a better approximation of the Hessian matrix, the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) based ICA have been proposed [24,25]. These methods are, however, not suitable for constrained ICA. In [26], the authors propose a real-valued orthogonally constrained ICA by performing L-BFGS updates on the Stiefel manifold. Nevertheless, this algorithm requires the estimate of the sign of the kurtosis – a highly noisy statistics especially when the sources are close to Gaussian. It must also reset L-BFGS whenever the kurtosis of any demixed signal changes its sign. These issues motivate the development of new quasi-Newton ICA algorithms in Chapter 6 and 7 where there is no restriction on the joint Hessian so that a better estimate can be achieved.

For under-determined scenarios where there are more sources than mixtures, the mixing matrix can still be estimated from mixtures of independent sources by joint decomposition of second-order or higher-order statistics of the observed data [27–32]. However, this is only possible if the mixing matrix satisfies certain rank conditions. These rank conditions limit the maximum number of sources these methods can deal with. On the other hand, SCA methods only require the mixing matrix to have non-collinear columns. The SCA methods exploit the fact that natural signals, when being transformed into a suitable domain, exhibit sparsity [33, 34]. With sufficient degree of source sparseness, it is likely

that one source dominates other sources for a significant number of samples, i.e., the source vectors are approximately disjoint orthogonal [35]. For such sources, the mixed data will concentrate along directions specified by columns of the mixing matrix. This implies that by clustering the mixture samples based on directions, one can estimate the under-determined mixing matrix up to a certain scaling and permutation ambiguity [13–17, 36]. The sources are then recovered via L_p-norm minimization [33, 37], masking [17, 20, 38, 39], or Kalman filtering [40]. Notable SCA algorithms based on directional clustering include the K-hyperlines (KHL) [13, 15, 41] and the Gaussian mixture model (GMM) of line orientations [14, 16, 17]. Since KHL is a hard clustering algorithm, its performance is limited. On the other hand, while GMM can achieve reasonable performance in separating convolutive mixtures, it models the distribution along each direction using a Gaussian function which does not necessarily fit well for sparse signals. In addition, the algorithm employs a computationally intensive expectation-maximization procedure, where each centroid is determined by performing the eigen-decomposition of a weighted covariance matrix. It is also useful to note that, the expectation-maximization process may converge to an undesirable saddle point [42]. More recently, it has been shown that sparse filtering (SF) behaves in a manner that is similar to cosine-metric clustering without the need of expectation-maximization procedure [43, 44]. Unfortunately, the penalty function of sparse filtering is designed for feature learning and not for under-determined mixing matrix estimation. This motivates the development of a new sparse filtering for mixing matrix estimation algorithm for directional data in Chapter 4.

Directional clustering methods for SCA differ from ICA in three aspects: the mixed data is normalized to have unit norm, the contrast function may be a non-separable function, and the sources are assumed to be approximately disjoint, i.e., highly sparse. It has been shown that the centroids of a clustering

algorithm will converge almost surely to the columns of the mixing matrix given that the sources are Laplace distributed, the mixing matrix is invertible, and the clustering algorithm is rotation invariant and sparse sensitivity (RISS) [45]. However, the RISS condition given in [45] is specific for Laplacian-distributed sources. This condition also depends on a conjecture which is confirmed only for a two-dimensional real vector space. The recovery condition is not only hard to verify in practice [46] but also does not take unit-norm normalization into account. Therefore, there is a need to derive a stability condition for invertible mixing matrix estimation using multi-variate non-separable contrast functions and non Laplacian sources. Such stability condition will allow one to know whether a SCA method can recover the unknown mixing matrix. This motivates the derivation of the stability condition in Chapter 5.

1.2 Objectives

The thesis aims to improve the performance of complex-valued mixing matrix estimation which plays a central role in unsupervised blind separation of convolutive mixtures. Both under-determined scenario and determined scenario are considered. The thesis also aims to establish the recovery condition of mixing matrix estimation from normalized directional data.

1.3 Main contribution of the Thesis

The main contributions of this thesis can be summarized as follows:

a) The development of a new mixing matrix estimation algorithm based on the magnitude-squared cosine-similarity for under-determined mixing matrix estimation. The proposed directional sparse filtering algorithm exploits the benefits of both sparse filtering and directional clustering to improve the estimation performance while having low computational complexity.

b) To gain insights into unsupervised algorithms that are based on cosine similarity, a stability condition is derived for the special case of determined mixing-matrix estimation from L_2-normalized data (the L_2-normalization occurs implicitly in cosine-similarity). From the derived stability condition, it can be seen that many Schur-concave contrast functions including the directional sparse filtering, KHL, and sparse filtering are suitable for mixing matrix estimation from normalized super-Gaussian mixtures in a manner similar to that of independent component analysis.

c) A new method to solve complex-valued ICA by entropy bound minimization is proposed by formulating constrained ICA cost functions as unconstrained optimization problems and solving these problems with quasi-Newton optimization method. The proposed ICA method does not apply any restriction on the joint Hessian. This may reflect the true Hessian more accurately than the existing fixed-point iteration method. Consequently, higher estimation performance can be achieved, particularly for source signals that have low non-Gaussianity.

d) Joint learning of the whitening filters and the mixing matrix is proposed for faster and better complex-valued ICA by entropy rate bound minimization. In the conventional complex-valued ICA by entropy rate bound minimization, the temporal whitening filters are optimized for each update of the mixing matrix. Such procedure is not only computationally intensive, but also can reduce the total estimation performance. These two drawbacks can be eliminated by simply estimating both the whitening filters and the mixing matrix simultaneously.

1.4 Organization of the Thesis

The remainder of this thesis is organized as follows: Chapter 2 includes the mathematical preliminary as well as the use of Schur-concave functions as sparsity penalty functions. Chapter 3 reviews the concepts of blind source separation

as well as two most common BSS approaches – independent component analysis and sparse component analysis. In Chapter 4, directional sparse filtering which employs a novel sparsity penalty function is proposed to improve the estimation performance. The stability condition for sparse component analysis based on magnitude-squared cosine similarity is presented in Chapter 5. Chapter 6 and Chapter 7 describe new approaches for solving complex-valued independent component analysis. The thesis is concluded in Chapter 8 with future research direction.

Chapter 2

Mathematical preliminary

2.1 Introduction

This chapter describes the notation that is used throughout this dissertation. Tools for the analysis of complex-valued matrix functions will be briefly reviewed as well. Finally, the underlying concept of Schur-concavity that is often adopted in sparse component analysis will be introduced [47–49].

2.2 Notation

Throughout the thesis, we will use lowercase symbols for scalars, lowercase bold symbols for vectors, and uppercase bold symbols for matrices. The operators $\mathcal{E}\{\cdot\}$, $\Re\{\cdot\}$, $\Im\{\cdot\}$, $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\otimes$, $\odot$, and $\oslash$, respectively, denote the expectation, real part, imaginary part, conjugation, transposition, conjugate transposition, Kronecker product, element-wise multiplication, and element-wise division. Unless otherwise stated, the expectation will be taken along the time dimension. The expectation is also replaced by sample average in practice.

The real field and complex field are, respectively, denoted by $\mathbb{R}$ and $\mathbb{C}$. The modulus, phase, and conjugate of a complex number $z = \Re\{z\} + \jmath\Im\{z\}$ are defined as $|z|$, $\angle z$, and z^*, respectively, where $\jmath = \sqrt{-1}$ is the imaginary unit. We

will use z_n to denote the nth element of a vector $\mathbf{z}$, $\mathbf{z}_n$ to denote the nth column of a matrix $\mathbf{Z}$, $\mathbf{z}_{m:}$ to denote the mth row of a matrix $\mathbf{Z}$, $z_{mn} = [\mathbf{Z}]_{mn}$ to denote the element at the mth row and nth row of $\mathbf{Z}$, and $\text{vec}(\mathbf{Z})$ to denote a long column vector created by stacking columns of $\mathbf{Z}$ in increasing order. The Frobenius norm of a matrix $\mathbf{Z}$ is given by $\|\mathbf{Z}\|_{\text{F}} = (\sum_{m=1}^{M}\sum_{n=1}^{N}|z_{mn}|^2)^{1/2}$. In addition, we define $\mathbf{I}_N \in \mathbb{C}^{N\times N}$ as the identity matrix, $\mathbf{e}_n$ as the nth Euclidean basis vector, and $\mathbf{K}_{MN}$ as the commutation matrix such that $\mathbf{K}_{MN}\text{vec}(\mathbf{Z}) = \text{vec}(\mathbf{Z}^T)$, $\forall \mathbf{Z} \in \mathbb{C}^{M\times N}$.

The L_p-norm of a vector is defined by $\|\mathbf{z}\|_p = (\sum_{n=1}^{N}|z_n|^p)^{1/p}$ for $p > 0$. The corresponding unit ball and its surface in L_p-norm are, respectively, $\mathcal{B}_p^N = \left\{\mathbf{z} \in \mathbb{C}^{N\times 1} : \|\mathbf{z}\|_p \leq 1\right\}$ and $\mathcal{S}_p^N = \left\{\mathbf{z} \in \mathbb{C}^{N\times 1} : \|\mathbf{z}\|_p = 1\right\}$. The L_2-normalization of $\mathbf{z}$ is $\bar{\mathbf{z}} = \mathbf{z}/\|\mathbf{z}\|_2$.

2.3 Differentiation of complex-valued vector and matrix functions

Blind separation of complex-valued signals commonly involves optimizing real-valued functions of complex-valued parameters in the form of a likelihood, a measure of non-Gaussianity, a sparsity penalty, or a clustering objective. Such functions are non-analytic since the classical complex derivative does not exist for $g(z) : \mathbb{C} \to \mathbb{R}$ [50, Sec. 13.4]. To circumvent this issue, we can parameterize a function of complex-valued arguments in two equivalent forms – real-composite form and complex-augmented form [51–53], where the latter is usually known as Wirtinger calculus.

For an arbitrarily defined complex vector $\mathbf{w} \in \mathbb{C}^N$, one can define its real-composite vector and complex-augmented vector, respectively, as: $\mathbf{w}_{\mathcal{R}} = [\Re\{\mathbf{w}\}^T, \Im\{\mathbf{w}\}^T]^T$, and $\mathbf{w}_{\mathcal{C}} = [\mathbf{w}^T, \mathbf{w}^H]^T$. The mapping from $\mathbf{w}_{\mathcal{R}}$ to $\mathbf{w}_{\mathcal{C}}$ is a one-to-one

linear map

$$\mathbf{w}_{\mathcal{C}} = \mathbf{T}_N \mathbf{w}_{\mathcal{R}}, \tag{2.1}$$

where $\mathbf{T}_N = \begin{bmatrix} \mathbf{I}_N & \jmath\mathbf{I}_N \\ \mathbf{I}_N & -\jmath\mathbf{I}_N \end{bmatrix}$ and $\mathbf{T}_N^{-1} = \frac{1}{2}\mathbf{T}_N^H$. Since differentiation is a linear operation, for a generic vector function $J(\mathbf{w}) : \mathbb{C}^N \mapsto \mathbb{C}^M$, we have

$$\frac{\partial J}{\partial \mathbf{w}_{\mathcal{C}}} = \frac{1}{2}\mathbf{T}_N^* \frac{\partial J}{\partial \mathbf{w}_{\mathcal{R}}} = \frac{1}{2}\begin{bmatrix} \frac{\partial J}{\partial \Re\{\mathbf{w}\}} - \jmath\frac{\partial J}{\partial \Im\{\mathbf{w}\}} \\ \frac{\partial J}{\partial \Re\{\mathbf{w}\}} + \jmath\frac{\partial J}{\partial \Im\{\mathbf{w}\}} \end{bmatrix}, \tag{2.2}$$

$$\frac{\partial J}{\partial \mathbf{w}_{\mathcal{R}}} = \mathbf{T}_N^T \frac{\partial J}{\partial \mathbf{w}_{\mathcal{C}}} = 2\begin{bmatrix} \Re\{\frac{\partial J}{\partial \mathbf{w}^*}\} \\ \Im\{\frac{\partial J}{\partial \mathbf{w}^*}\} \end{bmatrix}. \tag{2.3}$$

Since $\mathbf{w} = \Re\{\mathbf{w}\} + \jmath\Im\{\mathbf{w}\}$, (2.2) implies that $\frac{\partial \mathbf{w}}{\partial \mathbf{w}^*} = \frac{\partial \mathbf{w}^*}{\partial \mathbf{w}} = 0$. This also implies that analyzing $J(\mathbf{w}_{\mathcal{C}})$ is more convenient than analyzing $J(\mathbf{w}_{\mathcal{R}})$ since one can treat $\mathbf{w}$ as a constant when finding the derivative w.r.t. $\mathbf{w}^*$ and vice versa. In addition, (2.3) allows us to determine $\frac{\partial J}{\partial \mathbf{w}_{\mathcal{R}}}$ conveniently instead of direct computation (which, quite often, can be complicated [51, 53]).

When the co-domain of $J(\mathbf{w})$ is $\mathbb{R}$, the two equivalent Taylor expansions of $J(\mathbf{w})$ are given by [51]

$$J(\mathbf{w}_{\mathcal{R}} + \Delta\mathbf{w}_{\mathcal{R}}) = J(\mathbf{w}_{\mathcal{R}}) + \Delta\mathbf{w}_{\mathcal{R}}^T \frac{\partial J(\mathbf{w}_{\mathcal{R}})}{\partial \mathbf{w}_{\mathcal{R}}} + \frac{1}{2}\Delta\mathbf{w}_{\mathcal{R}}^T \frac{\partial^2 J(\mathbf{w}_{\mathcal{R}})}{\partial \mathbf{w}_{\mathcal{R}} \partial \mathbf{w}_{\mathcal{R}}^T}\Delta\mathbf{w}_{\mathcal{R}}, \tag{2.4}$$

$$J(\mathbf{w}_{\mathcal{C}} + \Delta\mathbf{w}_{\mathcal{C}}) = J(\mathbf{w}_{\mathcal{C}}) + \Delta\mathbf{w}_{\mathcal{C}}^T \frac{\partial J(\mathbf{w}_{\mathcal{C}})}{\partial \mathbf{w}_{\mathcal{C}}} + \frac{1}{2}\Delta\mathbf{w}_{\mathcal{C}}^H \frac{\partial^2 J(\mathbf{w}_{\mathcal{C}})}{\partial \mathbf{w}_{\mathcal{C}}^* \partial \mathbf{w}_{\mathcal{C}}^T}\Delta\mathbf{w}_{\mathcal{C}}, \tag{2.5}$$

where $\Delta\mathbf{w}_{\mathcal{R}}$ and $\Delta\mathbf{w}_{\mathcal{C}}$ denote small changes in $\mathbf{w}_{\mathcal{R}}$ and $\mathbf{w}_{\mathcal{C}}$, respectively.

By applying Cauchy–Schwarz inequality on the first-order terms of (2.4) and (2.5), the steepest descend directions of $\Delta J(\Delta\mathbf{w}_{\mathcal{R}})$ and $\Delta J(\Delta\mathbf{w}_{\mathcal{C}})$ are, respectively, given by $-\frac{\partial J(\mathbf{w}_{\mathcal{R}})}{\partial \mathbf{w}_{\mathcal{R}}}$ and $-\frac{\partial J(\mathbf{w}_{\mathcal{C}})}{\partial \mathbf{w}_{\mathcal{C}}^*}$. Consequently, the Newton directions are given by $-\left[\frac{\partial^2 J(\mathbf{w}_{\mathcal{R}})}{\partial \mathbf{w}_{\mathcal{R}} \partial \mathbf{w}_{\mathcal{R}}^T}\right]^{-1} \frac{\partial J(\mathbf{w}_{\mathcal{R}})}{\partial \mathbf{w}_{\mathcal{R}}}$ and $-\left[\frac{\partial^2 J(\mathbf{w}_{\mathcal{C}})}{\partial \mathbf{w}_{\mathcal{C}}^* \partial \mathbf{w}_{\mathcal{C}}^T}\right]^{-1} \frac{\partial J(\mathbf{w}_{\mathcal{C}})}{\partial \mathbf{w}_{\mathcal{C}}^*}$, respectively. It is useful to note that for the Newton direction in the space of $\mathbf{w}_{\mathcal{C}} = [\mathbf{w}^T, \mathbf{w}^H]^T$, one only needs to compute the first half corresponding to $\mathbf{w}$ since the second half is the

conjugation of the first half.

Extension of complex vector calculus in the above to matrix cases can be achieved via the vectorization operator [54]. For an arbitrarily chosen matrix $\mathbf{Z}$, let $\text{vec}(\mathbf{Z})$ be the vectorization of $\mathbf{Z}$ created by stacking columns of $\mathbf{Z}$ in their increasing order. The matrix differentiation of a complex-valued matrix function $\mathbf{F}(\mathbf{Z})$ in Wirtinger calculus is given by [53, 55]

$$d\text{vec}(\mathbf{F}) = (\mathcal{D}_{\mathbf{Z}}\mathbf{F})d\text{vec}(\mathbf{Z}) + (\mathcal{D}_{\mathbf{Z}^*}\mathbf{F})d\text{vec}(\mathbf{Z}^*), \tag{2.6}$$

where $\mathcal{D}_{\mathbf{Z}}\mathbf{F} = \partial\text{vec}(\mathbf{F})/\partial\text{vec}(\mathbf{Z})^T$ and $\mathcal{D}_{\mathbf{Z}^*}\mathbf{F} = \partial\text{vec}(\mathbf{F})/\partial\text{vec}(\mathbf{Z}^*)^T$ are the Jacobian matrices of $\mathbf{F}$. By definition, the value at the ith row and the jth column of $\mathcal{D}_{\mathbf{Z}}\mathbf{F}$ is the partial derivative of the ith component of $\text{vec}(\mathbf{F})$ with respect to the jth component of $\text{vec}(\mathbf{Z})$, i.e.,

$$[\mathcal{D}_{\mathbf{Z}}\mathbf{F}]_{ij} = \frac{\partial\,[\text{vec}(\mathbf{F})]_i}{\partial\,[\text{vec}(\mathbf{Z})]_j}. \tag{2.7}$$

For the nested matrix function $g(\mathbf{F}(\mathbf{Z})) : \mathbb{C}^{M\times N} \to \mathbb{R}$, its gradient, defined by $[\nabla_{\mathbf{Z}^*}g]_{mn} = 2\partial g/\partial z^*_{mn}$, is related to the Jacobians as

$$\text{vec}(\nabla_{\mathbf{Z}^*}g) = 2(\mathcal{D}_{\mathbf{Z}^*}g)^T, \tag{2.8}$$

where the constant 2 is due to (2.3). Since g is real-valued, we have $\nabla_{\mathbf{Z}}g = (\nabla_{\mathbf{Z}^*}g)^*$. Therefore, by applying (2.6) on $\mathbf{F}(\mathbf{Z}, \mathbf{Z}^*)$ and $g(\mathbf{F}, \mathbf{F}^*)$ we obtain the gradient of a composite cost function

$$\text{vec}(\nabla_{\mathbf{Z}^*}g) = (\mathcal{D}_{\mathbf{Z}^*}\mathbf{F})^T\text{vec}(\nabla_{\mathbf{F}^*}g)^* + (\mathcal{D}_{\mathbf{Z}}\mathbf{F})^H\text{vec}(\nabla_{\mathbf{F}^*}g). \tag{2.9}$$

More comprehensive topics of complex-valued matrix calculus are covered in [54, 55].

2.4 Majorization, Schur-convexity, and Schur-concavity

The concept of Schur-convexity/Schur-concavity is valuable for blind source separation because Schur-concavity promotes sparsity and the minimizers of a Schur-concave function or a Schur-convex function over a simplex are independent of the choice of the function.

Considering two vectors $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_N]^T \in \mathbb{R}_+^N$ and $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_N]^T \in \mathbb{R}_+^N$, let $\theta_{[n]}$ and $\phi_{[n]}$ be the nth largest component of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ respectively. The vector $\boldsymbol{\theta}$ is majorized by $\boldsymbol{\phi}$ [56] (denoted by $\boldsymbol{\theta} \prec \boldsymbol{\phi}$) if

$$\begin{aligned} \sum_{n=1}^{l} \theta_{[n]} &\leq \sum_{n=1}^{l} \phi_{[n]} \text{ for } l = 1, \ldots, N-1; \\ \text{and } \sum_{n=1}^{N} \theta_{[n]} &= \sum_{n=1}^{N} \phi_{[n]}. \end{aligned} \tag{2.10}$$

We are interested in a class of functions that is monotone with respect to majorization. A function $g(\boldsymbol{\theta}) : \mathbb{R}_+^N \to \mathbb{R}$ is Schur-concave if it is non-increasing regarding majorization, i.e.,

$$\boldsymbol{\theta} \prec \boldsymbol{\phi} \Rightarrow g(\boldsymbol{\theta}) \geq g(\boldsymbol{\phi}). \tag{2.11}$$

Furthermore, we say $g(\boldsymbol{\theta})$ is strictly Schur-concave if the equality of (2.11) occurs only when $\boldsymbol{\theta}$ is a coordinate-permutation of $\boldsymbol{\phi}$. In addition, if $g(\boldsymbol{\theta})$ is symmetric (i.e., invariant to permutation of arguments) and differentiable in $\mathbb{R}_+^N$ (see, e.g., [47, 48, 57]), one can replace condition in (2.11) with

$$(\theta_n - \theta_m)\left(\frac{\partial}{\partial \theta_n} g(\boldsymbol{\theta}) - \frac{\partial}{\partial \theta_m} g(\boldsymbol{\theta})\right) \leq 0, \ \forall n \neq m. \tag{2.12}$$

Similarly, we have a Schur-convex function if the inequalities in (2.11) and (2.12) are reversed.

In the context of sparse filtering and sparse coding, we note that Schur-

concave functions promote sparse solution. For instance, it can be seen that $[\frac{1}{N},\ldots,\frac{1}{N}]^T \prec \boldsymbol{\theta} \prec [1,0,\ldots,0]^T$ for every vector $\boldsymbol{\theta}$ belong to the standard simplex $\Delta_{N-1} = \left\{(\theta_1,\ldots,\theta_N)^T \middle| \theta_n \in [0,1], \sum_{n=1}^N \theta_n = 1\right\}$. Therefore, any strictly Schur-concave function $g(\boldsymbol{\theta})$ defined on Δ_{N-1} will attain its global minimum at N distinct extreme points of Δ_{N-1}, which correspond to sparse solutions $[1,0,\ldots,0]^T$, $[0,1,0\ldots,0]^T$, …, $[0,\ldots,0,1]^T$.

2.5 Power mean inequality

For a sequence of positive number $x_1,\ldots,x_N$, the power mean of order r, also known as the generalized mean, is defined as

$$\mu(\mathbf{x};\ r) = \left[\frac{1}{N}\sum_{n=1}^{N}(x_n)^r\right]^{1/r},\ r \neq 0. \tag{2.13}$$

The power mean becomes the arithmetic mean for $r = 1$, the harmonic mean for $r = -1$, and the geometric mean for $r \to 0$, i.e.,

$$\lim_{r\to 0}\mu(\mathbf{x};\ r) = \left(\prod_{n=1}^{N} x_n\right)^{1/n}. \tag{2.14}$$

Here, the above limits can be shown by utilizing the continuity of logarithm and applying L'Hôpital's rule on $\ln\lim_{r\to 0}\mu(\mathbf{x};\ r)$. For $r \to -\infty$ and $r \to \infty$, the power mean also becomes the minimum and maximum function, respectively,

$$\lim_{r\to-\infty}\mu(\mathbf{x};\ r) = \lim_{r\to-\infty} x_{\min}\left[\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x_n}{x_{\min}}\right)^r\right]^{1/r} = x_{\min}, \tag{2.15}$$

$$\lim_{r\to\infty}\mu(\mathbf{x};\ r) = \lim_{r\to\infty} x_{\max}\left[\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x_n}{x_{\max}}\right)^r\right]^{1/r} = x_{\max}, \tag{2.16}$$

where $x_{\min} = \min_{n=1,\ldots,N} x_n$ and $x_{max} = \max_{n=1,\ldots,N} x_n$. In the above equations, we have used $\lim_{r\to-\infty} y^r = 0$ for $y > 1$ and $\lim_{r\to\infty} y^r = 0$ for $0 \leq y < 1$.

Let $r_a < r_b < 0$, we have $r_a/r_b > 1$ and thus the function y^{r_a/r_b} is convex for

$y > 0$. By applying Jensen inequality, we obtain

$$\frac{1}{N}\sum_{n=1}^{N}(x_n)^{r_a} \geq \left(\frac{1}{N}\sum_{n=1}^{N}(x_n)^{r_b}\right)^{r_a/r_b}. \tag{2.17}$$

Since y^{1/r_a} is a decreasing function for $r_a < 0$, we have the power mean inequality for negative order as follows

$$\left(\frac{1}{N}\sum_{n=1}^{N}(x_n)^{r_a}\right)^{1/r_a} \leq \left(\frac{1}{N}\sum_{n=1}^{N}(x_n)^{r_b}\right)^{1/r_b}, \; r_a < r_b < 0. \tag{2.18}$$

Note that, the power mean inequality is also true for any $r_a < r_b$, $r_a \in \mathbb{R}$, and $r_b \in \mathbb{R}$.

2.6 Summary

In this chapter, the Wirtinger calculus has been reviewed. It is noted as being a convenient tool for analyzing real-valued functions of complex-valued parameters. The real-composite and complex-augmented forms to parameterize such functions have been introduced. While it can be seen that the two forms are completely equivalent, the complex-augmented form allows much simple derivation of the gradient. The chapter also introduces the concept of Schur-concavity, its necessary conditions, and its usage as sparsity penalty. These properties will be employed in the subsequent chapters of this thesis. Finally, the power mean is introduced and the power mean inequality for negative orders is proven.

Chapter 3

Background

3.1 Frequency-domain blind source separation

Blind source separation recovers latent sources from observed mixtures without knowing either the mixing process or the sources [58, 59]. It has many applications ranging from image recognition [60, 61], medical signal analysis [1, 2], to speech separation [3–5]. Due to the high indeterminacy of the BSS problem, the mixing process, which models the relationship between the observed data and unknown sources, is often assumed to be a linear convolution operation

$$x_m(t) = \sum_{n=1}^{N} a_{mn}(t) * s_n(t), \ \ m = 1, \ldots, M, \tag{3.1}$$

where t is the time index, M is the number of observed signals, N is the number of sources, $x_m(t)$ is the mth observed signal, $s_n(t)$ is the nth unknown source, and $a_{mn}(t)$ is the latent time-invariant filter modeling the propagation channel between $s_n(t)$ and $x_m(t)$. This mixing model is a noiseless *convolutive mixing model*. When $a_{mn}(t)$ is a scale of the Dirac delta function, the mixing model becomes instantaneous. Using short-time Fourier transform (STFT) with an appropriate choice of window length, we can approximate the convolutive mixing model in the time domain by an instantaneous mixing process in the time-

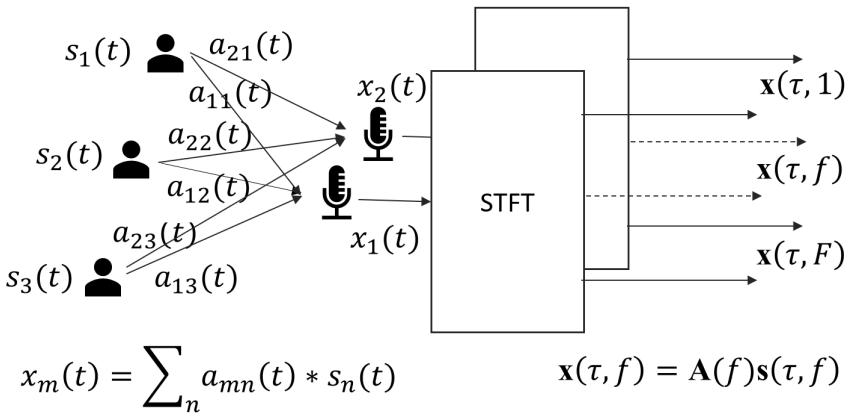


Fig. 3.1: Approximation of convolutive speech mixtures by instantaneous complex-valued linear mixing model in time-frequency domain.

frequency domain [62]

$$\mathbf{x}(\tau, f) = \mathbf{A}(f)\mathbf{s}(\tau, f) \tag{3.2}$$

$$= \sum_{n=1}^{N} \mathbf{a}_n(f) s_n(\tau, f), \tag{3.3}$$

where $\tau \in [1, 2, \ldots, T]$ is the time-frame index, $f \in [1, 2, \ldots, F]$ is the frequency-bin index, $\mathbf{x}(\tau, f) = [x_1(\tau, f), x_2(\tau, f), \ldots, x_M(\tau, f)]^T \in \mathbb{C}^{M\times 1}$ is the observed data, $\mathbf{s}(\tau, f) = [s_1(\tau, f), s_2(\tau, f), \ldots, s_N(\tau, f)]^T \in \mathbb{C}^{N\times 1}$ is the unknown source vector, and $\mathbf{A}(f) = [\mathbf{a}_1(f), \mathbf{a}_2(f), \ldots, \mathbf{a}_N(f)] \in \mathbb{C}^{M\times N}$ is the latent mixing matrix in which $\mathbf{a}_n(f) = [a_{1n}(f), a_{2n}(f), \ldots, a_{Mn}(f)]^T \in \mathbb{C}^{M\times 1}$ denotes its nth column. Here, $s_n(\tau, f)$ is the STFT representation of $s_n(t)$, $x_m(\tau, f)$ is the STFT representation of $x_m(t)$, and $a_{mn}(f)$ is the STFT representation of $a_{mn}(t)$. For illustration, Fig. 3.1 depicts the use of STFT to approximate $M = 2$ convolutive mixtures of $N = 3$ speakers by instantaneous complex-valued mixtures at each frequency bin.

It is useful to note that in *determined BSS* (i.e., $M = N$), we assume $\mathbf{A}^{-1}(f)$ exists. On the other hand, in *under-determined BSS* (i.e., $M < N$), we assume no pair of columns of $\mathbf{A}(f)$ are collinear, all the columns are "well-spread", and no column is a zero vector. *Over-determined BSS* where $M > N$ is not discussed

since it can be converted to the determined case [63].

Although the convolutive BSS problem in (3.1) can be solved directly in the time domain [64–66], frequency-domain BSS based on the narrow-band approximation in (3.2) is more computationally attractive. In frequency-domain BSS, we can estimate sources at each frequency bin independently followed by a permutation alignment procedure as shown in Fig. 3.2 [17, 39, 67–71]. Here, permutation alignment is required since the order of separated sources is inconsistent across all frequency bins. The high correlation of source activity pattern among frequency bins, direction of arrival, or a combination of both have been exploited to solve the permutation problem [17, 39, 71, 72]. On the other hand, since permutation alignment contributes to an increase in computational cost, many permutation-free methods, such as independent vector analysis [11, 73, 74] or full-band directional clustering [75, 76], have also been proposed for frequency-domain BSS. Nevertheless, it is found that the overall separation performance can still be improved by applying permutation alignment at the output of permutation-free methods [41]. We note that the performance of the entire frequency-domain BSS algorithm depends mainly on the quality of blind source separation, and that different permutation alignment procedures can be employed with only a modest variation in separation performance being introduced [17, 70]. In addition to permutation ambiguity, there is inevitably scaling ambiguity after mixing matrix estimation. Depending on the source estimation method, this scaling ambiguity can create undesirable artifacts in the reconstructed sources. Fortunately, the scaling ambiguity can be solved using minimum distortion principle [77] or back-projection [78].

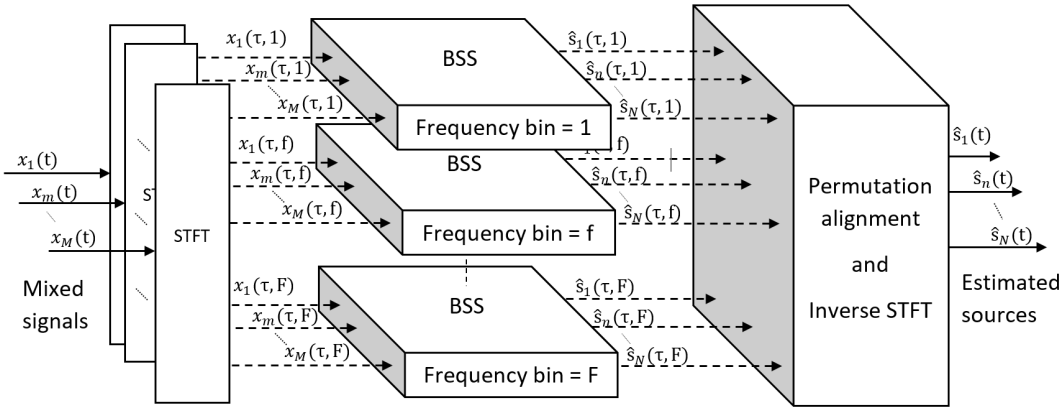


Fig. 3.2: Block diagram of frequency-domain blind source separation.

3.2 Complex-valued mixing matrix estimation

To this end, it can be seen that the complex-valued instantaneous BSS problem at each frequency given in (3.2) is a subset of the matrix factorization problem which commonly has no unique solution. To determine the estimated source $\widehat{\mathbf{s}}(\tau, f)$ from $\mathbf{x}(\tau, f)$, one often assumes that the sources have zero mean and identity covariance matrix, i.e., $E\{\mathbf{s}(\tau, f)\mathbf{s}^H(\tau, f)\} = \mathbf{I}_N$. Nonetheless, additional assumptions such as independence or sparsity are required to produce a provable recovery of the sources [58, 59, 79]. The technique relying on statistical independence is called *independent component analysis* while the technique relying on sparsity is called *sparse component analysis*. They both involve a mixing matrix estimation stage followed by a source reconstruction stage. Figure 3.3 shows the block diagram of a typical complex-valued instantaneous BSS algorithm where there is also an additional whitening step to facilitate the estimation process.

For ease of presentation, we will drop the time-frame index and/or frequency-bin index for the remainder of the thesis. Unless stated otherwise, all expectation operators are replaced by empirical mean in practice and they will be taken along the time-frame dimension.

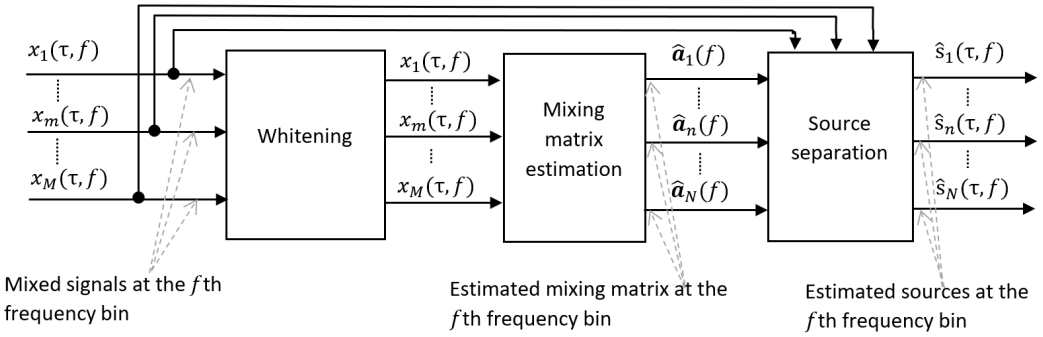
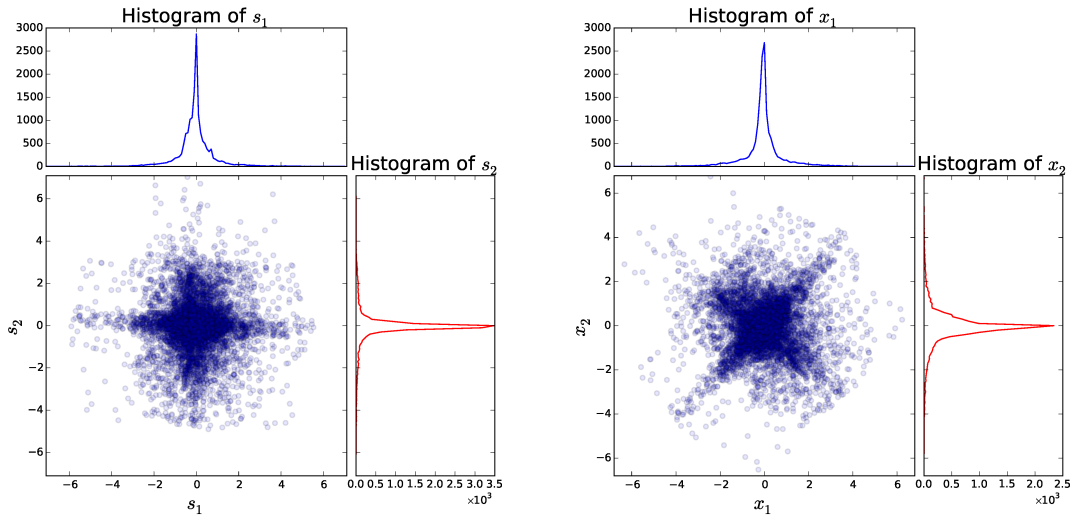


Fig. 3.3: Block diagram of a complex-valued blind source separation algorithm.

3.2.1 Whitening

Whitening is a linear decorrelation technique that finds a transformation such that the covariance matrix of the transformed data is the identity matrix. More specifically, let $\mathbf{Q}$ be the whitening matrix and $\mathbf{x}_W = \mathbf{Q}\mathbf{x} = \mathbf{Q}\mathbf{A}\mathbf{s}$ denote the whitened data, we search for $\mathbf{Q}$ such that $\mathcal{E}\{\mathbf{x}_W\mathbf{x}_W^H\} = \mathbf{Q}\mathcal{E}\{\mathbf{x}\mathbf{x}^H\}\mathbf{Q}^H = \mathbf{I}_M$. A suitable choice of $\mathbf{Q}$ is $\mathbf{Q} = \mathbf{E}\mathbf{D}^{-\frac{1}{2}}\mathbf{E}^H$ where $\mathbf{E}$ and $\mathbf{D}$ constitute the eigen-decomposition of the covariance matrix $\mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$. Since the sources are assumed to be zero-mean and independent with unit variance, $\mathcal{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_N$. Therefore, $\mathcal{E}\{\mathbf{x}_W\mathbf{x}_W^H\} = \mathbf{Q}\mathbf{A}\mathbf{A}^H\mathbf{Q}^H = \mathbf{I}_M$. If we use $\mathbf{A}_W = \mathbf{Q}\mathbf{A}$ to denote the mixing matrix which models the relationship between the whitened observed data and the sources, we can see that $\mathbf{A}_W$ is a unitary matrix in determined case and is a semi-unitary matrix in under-determined case, i.e., $\mathbf{A}_W\mathbf{A}_W^H = \mathbf{I}_M$. This implies that the rows of $\mathbf{A}_W$ are mutually orthogonal and have unit norm. One can show that $\mathbf{A}_W$ has $2MN - M^2$ degrees of freedom compared to $2MN$ degrees of freedom of an arbitrary matrix in $\mathbb{C}^{M\times N}$. Hence, to reduce the search space, it is natural for our mixing matrix estimation algorithm to perform whitening on the observed data as well as constraining our mixing matrix to be semi-unitary. For the remainder of this dissertation, we assume that the data is pre-whitened and that the mixing matrix is semi-unitary. Note that, when expectation is substituted by the empirical mean in practice, the mixing matrix



(a) (b)

Fig. 3.4: The distribution of horizontal and vertical projection of two female speech signals. (a) Unmixed sources. (b) Mixed signals.

is only approximately semi-unitary.

3.2.2 Complex-valued mixing matrix estimation

3.2.2.1 Complex-valued independent component analysis

Assume that $\mathbf{x}$ follows the linear model given in (3.2), let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ be the demixing matrix and let $\mathbf{y} = [y_1, \ldots, y_N]^T = \mathbf{W}^H\mathbf{x}$ be the demixed signals. In complex-valued ICA, we find the demixing matrix $\mathbf{W}$ so that the demixed signals $y_1, \ldots, y_N$ are as independent as possible. This can be achieved by maximization of non-Gaussianity [7, 21], minimization of negentropy [10], or minimization of mutual information [12, 22]. While these criteria are closely related, we will review the most common objective – maximization of non-Gaussianity.

According to central limit theorem, mixtures of non-Gaussian variables are closer to Gaussian than their original distribution. Figure 3.4 shows the scatter plots of two unmixed and two mixed female speeches. One can see that the histograms of the mixed signals exhibit wider tail and are less spiky than the histograms of original sources (i.e., they exhibit more Gaussianity). This char-

acteristic is exploited in ICA where the algorithm searches for the rotation that maximizes the non-Gaussianity. To do so, depending on whether the source is super-Gaussian or sub-Gaussian, we minimize or maximize the following objective [7, 21]

$$\mathcal{J}(\mathbf{w}_n^H\mathbf{x}) = \mathcal{E}\{g(\left|\mathbf{w}_n^H\mathbf{x}\right|^2)\} \text{ s.t. } \|\mathbf{w}_n\|_2 = 1, \tag{3.4}$$

for $n = 1, \ldots, N$ where $g(z)$ is a suitable nonlinear contrast function such as $g(z) = z^\alpha$ for $\alpha \in \mathbb{R}\backslash\{0, 1\}$.

To optimize $\mathcal{J}(\mathbf{w}_n^H\mathbf{x})$, one can use fixed-point iteration which is a projected quasi-Newton algorithm. For the objective in (3.4), the Lagrangian is

$$\mathcal{L}(\mathbf{w}_n, \lambda) = \mathcal{J}(\mathbf{w}_n^H\mathbf{x}) + \lambda(\mathbf{w}_n^H\mathbf{w}_n - 1). \tag{3.5}$$

With reference to (2.5), the Newton update to the above Lagrangian is [10, 21]

$$\mathbf{w}_{n\mathcal{C}} \leftarrow \mathbf{w}_{n\mathcal{C}} - \left(\frac{\partial^2 J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^* \partial \mathbf{w}_{n\mathcal{C}}^T} + \lambda\mathbf{I}\right)^{-1} \left(\frac{\partial J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^*} + \lambda\mathbf{w}_{nC}\right), \tag{3.6}$$

where $\mathbf{w}_{n\mathcal{C}} = [\mathbf{w}_n^T, \mathbf{w}_n^H]^T$ is the complex-augmented form of $\mathbf{w}_n$. Consequently, by multiplying both sides of (3.6) by $\frac{\partial^2 J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^* \partial \mathbf{w}_{n\mathcal{C}}^T} + \lambda\mathbf{I}$, one obtains

$$\left(\frac{\partial^2 J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^* \partial \mathbf{w}_{n\mathcal{C}}^T} + \lambda\mathbf{I}\right)\mathbf{w}_{n\mathcal{C}} \leftarrow -\frac{\partial J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^*} + \frac{\partial^2 J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^* \partial \mathbf{w}_{n\mathcal{C}}^T}\mathbf{w}_{n\mathcal{C}}. \tag{3.7}$$

By definition, the Hessian of $\mathcal{J}(\mathbf{w}_n^H\mathbf{x})$ is given by

$$\frac{\partial^2 \mathcal{J}(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^* \partial \mathbf{w}_{n\mathcal{C}}^T} = \begin{bmatrix} \frac{\partial^2 J}{\partial \mathbf{w}_n^* \partial \mathbf{w}_n^T} & \frac{\partial^2 J}{\partial \mathbf{w}_n^* \partial \mathbf{w}_n^H} \\ \frac{\partial^2 J}{\partial \mathbf{w}_n \partial \mathbf{w}_n^T} & \frac{\partial^2 J}{\partial \mathbf{w}_n \partial \mathbf{w}_n^H} \end{bmatrix} \tag{3.8}$$

$$= \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \\ \mathbf{B}_2^* & \mathbf{B}_1^* \end{bmatrix}, \tag{3.9}$$

where $\mathbf{B}_1 = \mathcal{E}\{(g''(|y_n|^2)\,|y_n|^2 + g'(|y_n|^2))\mathbf{x}\mathbf{x}^H\}$, $\mathbf{B}_2 = \mathcal{E}\{g''(|y_n|^2)y_n^{*2}\mathbf{x}\mathbf{x}^T\}$, $g'(y_n)$

is the first-order derivative of $g(y_n)$, and $g''(y_n)$ is the second-order derivative of $g(y_n)$. To simplify the learning rule, FastICA algorithm [6] and its complex-valued extensions [7, 10, 21] adopt the approximation $\mathcal{E}\{xxf(x)\} \approx \mathcal{E}\{xx\}\mathcal{E}\{f(x)\}$. Following the same approximation, one can approximate $\mathbf{B}_1$ and $\mathbf{B}_2$ as $\mathbf{B}_1 \approx \mathcal{E}\{g''(|y_n|^2)\,|y_n|^2 + g'(|y_n|^2)\}\mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$ and $\mathbf{B}_2 \approx \mathcal{E}\{g''(|y_n|^2)y_n^{*2}\}\mathcal{E}\{\mathbf{x}\mathbf{x}^T\}$. Note that, the term $\frac{\partial^2 J(\mathbf{w}_{n\mathcal{C}})}{\partial \mathbf{w}_{n\mathcal{C}}^* \partial \mathbf{w}_{n\mathcal{C}}^T} + \lambda\mathbf{I}$ on the left-hand side of (3.7) can be ignored near convergence point due to normalization [21]. Since the lower half of the update for $\mathbf{w}_{n\mathcal{C}}$ in (3.7) is just the conjugation of the upper half, substituting the approximation of $\mathbf{B}_1$ and $\mathbf{B}_2$ into (3.7) yields the fixed-point update for nth demixing filter

$$\begin{aligned}\mathbf{w}_n \leftarrow & -\mathcal{E}\{g'(|y_n|^2)y_n^*\mathbf{x}\} + \mathcal{E}\{g''(|y_n|^2)\,|y_n|^2 + g'(|y_n|^2)\}\mathcal{E}\{\mathbf{x}\mathbf{x}^H\}\mathbf{w}_n + \\ & \mathcal{E}\{g''(|y_n|^2)y_n^{*2}\}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{w}_n^*. \end{aligned} \tag{3.10}$$

After $\mathbf{w}_n$ is computed for all n, an orthonormalization process such as Gram–Schmidt process can be applied so that $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ is unitary. In practice, ICA algorithms usually employ the following orthonormalization step,

$$\mathbf{W} \leftarrow \left(\mathbf{W}\mathbf{W}^H\right)^{-\frac{1}{2}}\mathbf{W}. \tag{3.11}$$

Defining vec($\mathbf{W}$) as a vector created by stacking all columns of $\mathbf{W}$, it can be seen that the Hessian of vec($\mathbf{W}$) in fixed point iteration is restricted to be block diagonal. This is because the updates for all $\mathbf{w}_n$ are computed independently before orthonormalization. Furthermore, the Hessian of each column $\mathbf{w}_n$ may also be inaccurate since the independence assumption is not strictly held in practice. These can reduce the rate of convergence of complex-valued ICA. In Chapter 6, we will mitigate this issue using re-parameterization.

The major concern of ICA is the condition imposed on the contrast function and the sources so that one can utilize (3.4) to recover the mixing matrix. The

following stability theorem outlines such condition.

Theorem 3.1 (*[7]*). *Given the mixing model in (3.2) where the observed signals are pre-whitened, the local minimizers of* $\mathcal{E}\left\{g\left(\left|\mathbf{w}_n^H\mathbf{x}\right|^2\right)\right\}$ *subjected to* $\|\mathbf{w}_n\|_2^2 = 1$ *include the* n*th column of the mixing matrix if the corresponding circular complex-valued source* s_n *and the contrast function* $g(z)$ *satisfy*

$$\mathcal{E}\left\{g'\left(|s_n|^2\right) - |s_n|^2 g'\left(|s_n|^2\right) + |s_n|^2 g''\left(|s_n|^2\right)\right\} > 0, \tag{3.12}$$

where $g'(y_n)$ *is the first-order derivative of* $g(y_n)$ *and* $g''(y_n)$ *is the second-order derivative of* $g(y_n)$.

Theorem 3.1 has been extended to non-circular sources and been proven for several classes of non-linear real-valued function with complex arguments [9, 10, 21, 80]. Assuming that the sources are non-Gaussian, Theorem 3.1 is true for almost all non-linear function. For Gaussian sources, it can be seen that one cannot solve the ICA problem because any unitary rotation of Gaussian random vector is also a Gaussian random vector.

The conventional ICA approaches discussed in the above are not suitable for under-determined scenario where $N > M$. However, if the under-determined mixing matrix satisfies certain rank conditions, it can be estimated by tensor decomposition of second-order statistics of the mixtures [30], four-order statistics of the mixtures [28, 29], or higher-order statistics of the mixtures [27, 31, 32, 81]. While the ICA methods based on tensor analysis are able to work in the under-determined scenario, the maximum number of sources N they can process are limited for a given number of sensors M. Generally, the higher-order tensor methods such as [32, 81, 82] can handle more sources than second-order and four-order methods such as [29, 30]. On the other hand, the higher-order methods require stronger rank conditions on the mixing matrix than lower-order methods do. The next section will include the discussion about sparse component

analysis for under-determined mixing matrix estimation. Compared to tensor analysis methods, the sparse methods employ a more restrictive assumption on the sources while having a less restrictive assumption on the mixing matrix.

3.2.2.2 Sparse component analysis

Natural signals such as speech in the time-frequency domain or image in the wavelet domain are sparse in the sense that they have only a few non-zero elements [33, 83]. It is also useful to note that a source $\mathbf{s}$ is k-sparse if it has at most k non-zero elements [84]. With sufficient degree of sparseness a source vector $\mathbf{s}$ may have at most one dominant source at any time index [35, 84] and this property is known as *approximately disjoint orthogonality* [35, 36, 58, 85–88]. Under this situation, every sample will be distributed along the directions of the columns of the mixing matrix. For instance, considering the case where the nth source is dominant at τth time-frame index, we have

$$\mathbf{x}(\tau) \approx \mathbf{a}_n s_n(\tau). \tag{3.13}$$

Figure 3.5 illustrates two mixtures of three real-valued sources together with the directions of the columns of the mixing matrix. The source distribution is assumed to be $f(s) = c\exp(-|s|^{0.2})$ with c being a normalizing factor. One can clearly observe that the majority of data points lie close to the directions specified by the mixing matrix. Therefore, for approximately disjoint orthogonal sources, one can recover the mixing matrix up to a scale and permutation ambiguity by identifying the distinct single source points or clustering the data based on its direction [33, 89].

To identify single-source points, consider the case where the mixing matrix is real-valued, (3.13) implies

$$\Re\{\mathbf{x}(\tau)\} \approx \mathbf{a}_n \Re\{s_n(\tau)\}, \tag{3.14}$$

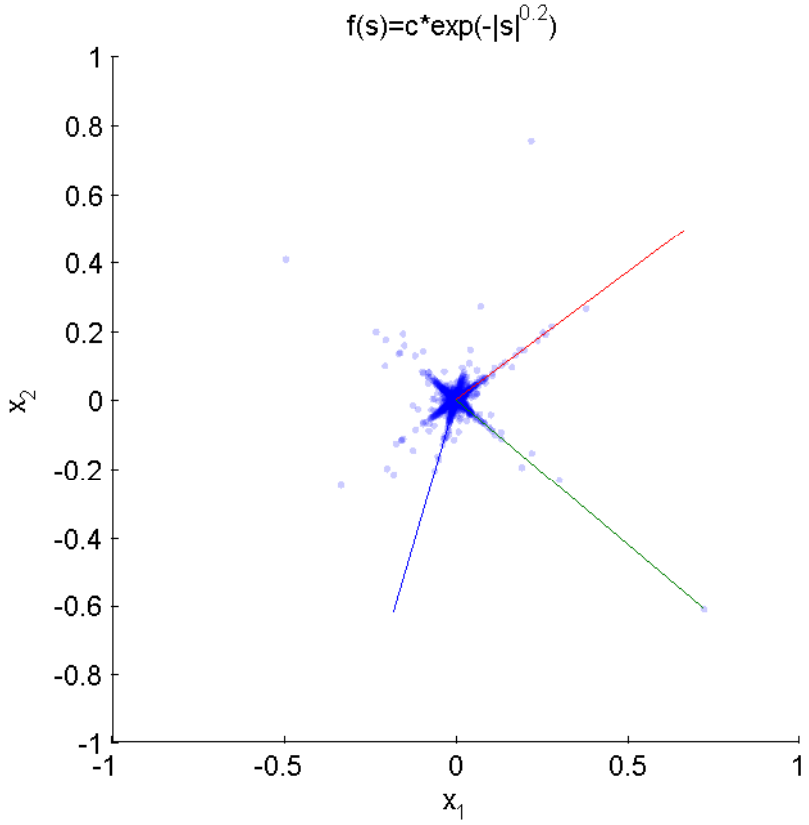


Fig. 3.5: Scatter plot of two under-determined mixtures of three sub-Laplace sources. Red, green, and blue lines indicate the directions of the columns of the mixing matrix.

$$\Im\{\mathbf{x}(\tau)\} \approx \mathbf{a}_n \Im\{s_n(\tau)\}, \tag{3.15}$$

where $\Re\{s_n(\tau)\}$ and $\Im\{s_n(\tau)\}$ respectively denote the real part and imaginary part of $s_n(\tau)$. As a consequence, the angle between the real part and the imaginary part of a single source point is close to zero, thus we can identify single-source points using the following condition [90]

$$\cos^{-1}\left(\frac{\Re\{\mathbf{x}^T(\tau)\}\Im\{\mathbf{x}(\tau)\}}{\|\Re\{\mathbf{x}(\tau)\}\|_2 \, \|\Im\{\mathbf{x}(\tau)\}\|_2}\right) < \Delta\theta, \tag{3.16}$$

where $\Delta\theta$ is a small angle. When the mixing matrix is complex-valued, this condition is not valid since (3.14) and (3.15) are violated.

Instead of identifying the single source points directly, one can perform directional clustering on the observed data to estimate the mixing matrix. In direc-

tional clustering-based SCA methods, the observed data will be L_2-normalized to have unit norm, i.e., $\bar{\mathbf{x}}(\tau) = \mathbf{x}(\tau)/\left\|\mathbf{x}(\tau)\right\|_2$. As there are infinite number of unit vectors having the same direction in the complex vector space, the distance metric used for clustering should be phase-invariant. One of such distance metric is the phase-invariant cosine distance [13, 15, 91]

$$D(\mathbf{x}(\tau), \widehat{\mathbf{a}}_n) = 1 - \frac{\left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}(\tau)\right|^2}{\|\widehat{\mathbf{a}}_n\|_2^2}, \tag{3.17}$$

where $\widehat{\mathbf{a}}_n$ is the estimate of the nth column of the mixing matrix $\mathbf{a}_n$.

K-hyperlines clustering: Replacing Euclidean distance in K-means clustering with the above phase-invariant cosine distance, the optimization problem of K-hyperlines clustering [13, 15, 41, 91] is obtained and is given by

$$\begin{aligned} \min\ & \frac{1}{T}\sum_{\tau=1}^{T}\sum_{n=1}^{N} C_n(\tau)\left(1 - \left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}(\tau)\right|^2\right), \\ \text{s.t.}\ & \sum_{n=1}^{N} C_n(\tau) = 1,\ C_n(\tau) \in \{0,1\},\ \tau = 1, \ldots, T, \\ & \|\widehat{\mathbf{a}}_n\|_2^2 = 1,\ n = 1, \ldots, N. \end{aligned} \tag{3.18}$$

In the above, $\widehat{\mathbf{a}}_n$ is the nth centroid which corresponds to the estimate of nth column of the mixing matrix and $C_n(\tau)$ is the membership between the nth centroids and the τth sample. By solving the corresponding Lagrangian when fixing the centroids, the membership function is updated as follows

$$C_n^{(\mathrm{KHL})}(\tau) \leftarrow \begin{cases} 1, & \text{if } 1 - \left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}(\tau)\right|^2 \leq 1 - \left|\widehat{\mathbf{a}}_l^H \bar{\mathbf{x}}(\tau)\right|^2 \text{ for all } l \neq n, \\ 0, & \text{otherwise.} \end{cases} \tag{3.19}$$

To reduce the memory requirement and computational cost, we can replace the membership matrix by N sets of samples

$$P_n^{(\mathrm{KHL})} = \left\{\tau | C_n^{(\mathrm{KHL})}(\tau) \equiv 1\right\}. \tag{3.20}$$

After that, by fixing the membership $P_n^{(\mathrm{KHL})}$, one can show that the new centroid of the nth cluster $\widehat{\mathbf{a}}_n$ should be chosen as the eigen-vector corresponding to the largest eigen-value of the following matrix

$$\mathbf{R}_n^{(\mathrm{KHL})} \leftarrow \frac{1}{\left|P_n^{(\mathrm{KHL})}\right|} \sum_{\tau \in P_n^{(\mathrm{KHL})}} \bar{\mathbf{x}}(\tau)\bar{\mathbf{x}}^H(\tau), \tag{3.21}$$

where $\left|P_n^{(\mathrm{KHL})}\right|$ is the number of elements in $P_n^{(\mathrm{KHL})}$. The updates of the membership and the centroids will be alternated until convergence. The computational cost per iteration for KHL is $O(N \max_n \left(\left|P_n^{(\mathrm{KHL})}\right|\right) M^2 + NM^3)$ where the first term is due to the computation of N intra-covariance matrix in (3.21) and the second term is due to eigen-decomposition. Let us assume that all sources having the same activity, we have $O(\max_n \left(\left|P_n^{(\mathrm{KHL})}\right|\right)) = O(T/N)$. Therefore, the complexity of KHL is $O(TM^2 + NM^3)$.

Gaussian mixture model of line orientations: While KHL clustering has low computational cost, its performance is limited due to the binary membership $C_n^{(\mathrm{KHL})}(\tau)$. Therefore, directional soft-clustering methods based on Gaussian mixture model have been proposed [14, 16, 17]. This model can be described by

$$f(\bar{\mathbf{x}}(\tau)|\boldsymbol{\theta}) = \sum_{n=1}^{N} \alpha_n p(\bar{\mathbf{x}}(\tau)|m(\tau) = n, \widehat{\mathbf{a}}_n^H, \sigma_n^2), \tag{3.22}$$

$$f(\bar{\mathbf{x}}(\tau)|m(\tau) = n, \widehat{\mathbf{a}}_n^H, \sigma_n^2) = \frac{1}{(\pi\sigma_n^2)^{M-1}} \exp\left(-\frac{1}{\sigma_n^2}(1 - \left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}(\tau)\right|^2)\right), \tag{3.23}$$

where $\boldsymbol{\theta}$ is the set of all unknown parameters of the model and α_n is the mixture weight with $\sum_n \alpha_n = 1$. The latent variable $m(\tau)$ denotes the cluster assigned

for $\bar{\mathbf{x}}(\tau)$, $f(\bar{\mathbf{x}}(\tau)|m(\tau)=n,\hat{\mathbf{a}}_n,\sigma_n^2)$ is the likelihood that $\bar{\mathbf{x}}(\tau)$ belongs to the nth cluster, $\hat{\mathbf{a}}_n$ is the unit-norm vector denoting the estimate of $\mathbf{a}_n$, and σ_n^2 is the dispersion of the nth cluster. Using the expectation-maximization framework, one can find the posterior probability in expectation step via

$$C_n^{(\mathrm{GMM})}(\tau) \leftarrow \frac{\alpha_n f(\bar{\mathbf{x}}(\tau)|m(\tau)=n,\hat{\mathbf{a}}_n^H,\sigma_n^2)}{\sum_{n=1}^N \alpha_n f(\bar{\mathbf{x}}(\tau)|m(\tau)=n,\hat{\mathbf{a}}_n^H,\sigma_n^2)}. \tag{3.24}$$

The maximization step is then

$$\mathbf{R}_n^{(\mathrm{GMM})} \leftarrow \sum_{\tau=1}^{T} C_n^{(\mathrm{GMM})}(\tau)\bar{\mathbf{x}}(\tau)\bar{\mathbf{x}}^H(\tau), \tag{3.25}$$

$$\sigma_n^2 \leftarrow \frac{1}{(M-1)\sum_{\tau=1}^{T} C_n^{(\mathrm{GMM})}(\tau)} \sum_{\tau=1}^{T} C_n^{(\mathrm{GMM})}(\tau)\left(1-\left|\hat{\mathbf{a}}_n^H\bar{\mathbf{x}}(\tau)\right|^2\right), \tag{3.26}$$

$$\alpha_n \leftarrow \frac{1}{T}\sum_{\tau=1}^{T} C_n^{(\mathrm{GMM})}(\tau). \tag{3.27}$$

Similarly to K-hyperlines, the estimate of the nth column of the mixing matrix $\hat{\mathbf{a}}_n$ is the dominant eigen-vector of $\mathbf{R}_n^{(\mathrm{GMM})}$. To control the level of source activity, a Dirichlet prior can be applied on the mixture weights [17]

$$f(\alpha_1,\ldots,\alpha_N) = \frac{\Gamma(N\phi)}{\Gamma(\phi)^N}\prod_{n=1}^{N}\alpha_n^{\phi-1}, \tag{3.28}$$

where Γ is the gamma function and $\phi>0$ is a hyper-parameter. Using the above Dirichlet prior, the update in (3.27) becomes

$$\alpha_n \leftarrow \frac{\phi-1+\sum_{\tau=1}^{T} C_n^{(\mathrm{GMM})}(\tau)}{N(\phi-1)+T}. \tag{3.29}$$

For $0<\phi<1$, it is more likely that one mixture weight is considerably larger than other mixture weights. This has potential application in extracting a dominant source from weaker interferences. For $\phi=1$, all the combinations of mixture weights have a fair chance. On the other hand, for $\phi>1$, it is more

likely that the mixture weights are equal or approximately equal. This implies that a small value of ϕ implies that some sources are more active than the others, whereas a large value of ϕ implies that all sources have similar activity, i.e., $\alpha_1, \ldots, \alpha_N$ are approximately equal. The computational cost per iteration for GMM is $O(NTM^2 + NM^3)$ where the first term is due to the computation of N weighted covariance matrices in (3.25) and the second term is due to eigen-decomposition.

Angular central Gaussian mixture model: The angular central Gaussian mixture model (ACGMM) uses a full-rank spatial covariance matrix $\mathbf{B}_n$ for each cluster as follows [92]

$$f(\bar{\mathbf{x}}(\tau)|\boldsymbol{\theta}) = \sum_{n=1}^{N} \alpha_n f(\bar{\mathbf{x}}(\tau)|m(\tau) = n, \mathbf{B}_n), \tag{3.30}$$

$$f(\bar{\mathbf{x}}(\tau)|m(\tau) = n, \mathbf{B}_n) = \frac{(M-1)!}{2\pi^M \det(\mathbf{B}_n)} \frac{1}{(\bar{\mathbf{x}}^H(\tau)\mathbf{B}_n^{-1}\bar{\mathbf{x}}_n(\tau))^M}. \tag{3.31}$$

As opposed to K-hyperlines and GMM, ACGMM can model the rotationally asymmetric cluster since $\mathbf{B}_n$ is full rank. Consequently, ACGMM can potentially achieve better prediction than GMM. The expectation-maximization updates are given by

$$C_n^{(\text{ACGMM})}(\tau) \leftarrow \frac{\alpha_n f(\bar{\mathbf{x}}(\tau)|m(\tau) = n, \mathbf{B}_n)}{\sum_{n=1}^{N} f(\bar{\mathbf{x}}(\tau)|m(\tau) = n, \mathbf{B}_n)}, \tag{3.32}$$

$$\mathbf{B}_n \leftarrow M \frac{\sum_{\tau=1}^{T} C_n^{(\text{ACGMM})}(\tau) \frac{\bar{\mathbf{x}}_n(\tau)\bar{\mathbf{x}}_n^H(\tau)}{\bar{\mathbf{x}}_n^H(\tau)\mathbf{B}_n^{-1}\bar{\mathbf{x}}_n(\tau)}}{\sum_{\tau=1}^{T} C_n^{(\text{ACGMM})}(\tau)}, \tag{3.33}$$

$$\alpha_n \leftarrow \frac{1}{T} \sum_{\tau=1}^{T} C_n^{(\text{ACGMM})}(\tau). \tag{3.34}$$

Similarly to both K-hyperlines and GMM, the estimate of nth column of the mixing matrix $\widehat{\mathbf{a}}_n$ is the dominant eigen-vector of $\mathbf{B}_n$. However, this is only performed once when the algorithm converges. Therefore, ACGMM requires a computational complexity of $O(NTM^2)$ for each iteration.

We note that one can reduce the computational complexity of K-hyperlines, GMM, and ACGMM by utilizing a fast approximation method of the dominant eigen-vector, e.g. randomized algorithm [93] or power iteration [94]. Furthermore, by assuming that there are much more data samples than number of sources, i.e., $T \gg N$, we can ignore the computational cost of eigen-decomposition in directional clustering algorithms as shown in Table 3.1. It can be seen that KHL is significantly faster than GMM and ACGMM.

Algorithms	Computational complexity per iteration
KHL	$O(TM^2)$
GMM	$O(NTM^2)$
ACGMM	$O(NTM^2)$

Table 3.1: Computational complexity per iteration for several directional clustering algorithms.

In [95], the spatial model based on ACGMM has been integrated with deep neural network for supervised blind speech separation. This approach is highly successful on improving the separation performance over the deep-learning baselines where only spectro-temporal features are exploited.

Departing from directional clustering models which directly partition the observed data, Yang et al. have recently proposed to estimate the mixing matrix by clustering the dominant eigen-vectors of local covariance matrices [96]. This is motivated by local dominance of speech sources, i.e., each speech source is dominant across a few consecutive frames. Simulation results show that this clustering method can produce good estimation of under-determined mixing matrices from convolutive speech mixtures. Note that, the local dominant assumption may not hold for other types of sparse sources.

3.2.2.3 Non-negative matrix factorization

Beside ICA and SCA, there is a distinct family of blind source separation algorithms employing non-negative matrix factorization (NMF) to model the power

spectrum of the sources [97, 98]. Note that, the frequency-bin index is explicitly included in this subsection since NMF operates in both time and frequency dimensions. In NMF model of [97], the source $s_n(\tau, f)$ is assumed to be a time-varying zero-mean Gaussian random variable. Consequently, the source image received at the sensors, $\mathbf{s}_n^{\text{img}}(\tau, f) = \mathbf{a}_n s_n(\tau, f)$, is a time-varying zero-mean complex Gaussian vector with the time-varying covariance matrix given by

$$\mathbf{R}_{\mathbf{s}_n^{\text{img}}}(\tau, f) = v_n(\tau, f)\mathbf{R}_n(f), \tag{3.35}$$

where $v_n(\tau, f) = \mathcal{E}\{|s_n(\tau, f)|\}$ is the time-varying spectro-temporal covariance of the nth source and $\mathbf{R}_n(f) = \mathbf{a}_n\mathbf{a}_n^H$ is the time-invariant rank-1 spatial covariance matrix of the nth source. The spectro-temporal covariance $v_n(\tau, f)$ is further assumed to follow the NMF model,

$$v_n(\tau, f) = \sum_l w_n(f, l)h_n(l, \tau), \tag{3.36}$$

where $w_n(f, l) > 0$ and $h_n(l, \tau) > 0$.

In present of time-invariant zero-mean Gaussian noise $\mathbf{n}(\tau, f)$, the noisy mixtures can be written as $\mathbf{x}(\tau, f) = \sum_n \mathbf{s}_n^{\text{img}}(\tau, f) + \mathbf{n}(\tau, f)$. As a consequence, $\mathbf{x}(\tau, f)$ is also a zero-mean Gaussian vector, i.e.,

$$\mathbf{x}(\tau, f) \sim \frac{1}{\det(\pi\mathbf{R}_{\mathbf{x}}(\tau, f))} \exp\left(-\mathbf{x}^H(\tau, f)\mathbf{R}_{\mathbf{x}}^{-1}(\tau, f)\mathbf{x}(\tau, f)\right), \tag{3.37}$$

where

$$\begin{aligned}\mathbf{R}_{\mathbf{x}}(\tau, f) &= \sum_n \mathbf{R}_{\mathbf{s}_n^{\text{img}}}(\tau, f) + \mathbf{R}_b(f) \\ &= \mathbf{A}(f)\mathbf{R}_s(\tau, f)\mathbf{A}^H(f) + \mathbf{R}_b(f)\end{aligned} \tag{3.38}$$

and $\mathbf{R}_s(\tau, f) = \text{Diag}(v_1(\tau, f), \ldots, v_N(\tau, f))$. Let $\boldsymbol{\theta}^{(\text{NMF})} = \{\widehat{\mathbf{A}}(f), \widehat{\mathbf{R}}_b(f), \widehat{w}_n(f, l),$

$\widehat{h}_n(l,\tau), \forall f,l,\tau\}$ be the estimate of $\{\mathbf{A}(f), \mathbf{R}_b(f), w_n(f,l), h_n(l,\tau), \forall f,l,\tau\}$, it can be seen that $\boldsymbol{\theta}^{(\mathrm{NMF})}$ can be found by maximum likelihood estimation (MLE),

$$\min_{\boldsymbol{\theta}^{(\mathrm{NMF})}} \sum_{\tau,f} \mathbf{x}^H(\tau,f)\mathbf{R}_{\mathbf{x}}^{-1}(\tau,f)\mathbf{x}(\tau,f) + \ln\det\mathbf{R}_{\mathbf{x}}(\tau,f). \tag{3.39}$$

The above optimization problem can be solved using expectation-maximization algorithm or multiplicative updating rules under certain normalization constraints on $\widehat{\mathbf{A}}(f)$, $\widehat{w}_n(f,l)$ and $\widehat{h}_n(l,\tau)$ [97].

Since rank-1 assumption of $\mathbf{R}_n(f)$ can be invalid in reverberant environments, the model in (3.37) has been extended to use full-rank spatial covariance matrices [99,100]. This resulted in better speech separation than the original rank-1 NMF does [97]. Subsequently, the spatial prior information was also incorporated into (3.37), giving the maximum a posteriori estimation in [101]. On the other hand, as the expectation-maximization algorithm is highly sensitive to the initial choice of the mixing matrix, time differences of arrival estimation [102] and tensor decomposition [103,104] have recently been adopted to address this initialization issue. Nonetheless, the methods in [97,99,102–104] are not truly blind since they employed an oracle initialization procedure where the initial values of $\widehat{w}_n(f,l)$ and $\widehat{h}_n(l,\tau)$ were computed from the ground-truth sources $s_n(\tau,f)$. In practice, this can severely affect the separation performance as the optimal initial values of $\widehat{w}_n(f,l)$ and $\widehat{h}_n(l,\tau)$ are rarely available.

Hybrid methods based on NMF have also been studied. In [105], NMF model was combined with independence model, giving the independent low-rank matrix analysis (ILRMA) algorithm for determined source separation. In [106, 107], deep-learning prior has been incorporated into multi-channel NMF model [98]. Simulation results show that deep-learning prior can enhance the separation performance of multi-channel NMF.

3.2.3 Source separation

In the determined scenario, once the mixing matrix (or the demixing matrix) is estimated, the source can be computed by solving the following linear equation

$$\widehat{\mathbf{s}}(\tau) = \widehat{\mathbf{A}}^{-1}\mathbf{x}(\tau), \tag{3.40}$$

where $\widehat{\mathbf{A}}$ is the estimated mixing matrix and $\widehat{\mathbf{A}}^{-1} \equiv \mathbf{W}$ is the demixing matrix. However, there exists a scaling ambiguity problem which can be mitigated using the minimum distortion principle (MDP) [77]. Given the estimated mixing matrix $\widehat{\mathbf{A}}$, the demixing matrix according to MDP is defined by

$$\mathbf{W}^{(\mathrm{MDP})} = \mathrm{Ddiag}(\widehat{\mathbf{A}})\widehat{\mathbf{A}}^{-1}, \tag{3.41}$$

where $\mathrm{Ddiag}(\widehat{\mathbf{A}})$ is the diagonal matrix formed by diagonal elements of $\widehat{\mathbf{A}}$. Suppose that $\widehat{\mathbf{A}}^{-1} = \mathbf{\Psi}\mathbf{A}^{-1}$ where $\mathbf{\Psi}$ is an unknown diagonal scaling matrix, it can be seen that $\mathbf{W}^{(\mathrm{MDP})} = \mathrm{Ddiag}(\mathbf{A}\mathbf{\Psi}^{-1})\mathbf{\Psi}\mathbf{A}^{-1} = \mathrm{Ddiag}(\mathbf{A})\mathbf{A}^{-1}$ which does not depend on $\mathbf{\Psi}$. The scaling ambiguity can also be mitigated using back-projection method [67, 78, 108], i.e., decomposing a reference mixture $x_{\mathrm{ref}}(\tau)$ into a summation of estimated source components,

$$x_{\mathrm{ref}}(\tau) = \sum_n c_n^* \widehat{s}_n(\tau) = \mathbf{c}^H \widehat{\mathbf{s}}(\tau). \tag{3.42}$$

Here, the scaling coefficients $\mathbf{c}$ can be determined via least-square projection,

$$\mathbf{c} = \mathbf{G}_{\hat{s}\hat{s}}^{-1} \left[\langle x(\tau), \widehat{s}_1(\tau)\rangle, \ldots, \langle x(\tau), \widehat{s}_N(\tau)\rangle\right]^H, \tag{3.43}$$

where $[\mathbf{G}_{\hat{s}\hat{s}}]_{nn'} = \langle \widehat{s}_n(\tau), \widehat{s}_{n'}(\tau)\rangle$ is the Gram matrix of the estimated sources and $\langle a(\tau), b(\tau)\rangle = \sum_\tau a(\tau)b^*(\tau)$ denotes the complex inner product between a signal $a(\tau)$ and a signal $b(\tau)$. Once the scaling coefficients are computed, the estimated

sources are normalized as follows,

$$\widehat{\mathbf{s}}(\tau) \leftarrow \mathrm{Diag}(\mathbf{c})\widehat{\mathbf{s}}(\tau). \tag{3.44}$$

Due to the above normalization, the final estimated sources are not effected by scaling ambiguity.

For the under-determined scenario, estimation of the sources is more challenging since the mixing matrix is not invertible. In such a case, one can make use of L_1-norm minimization [33, 36, 84, 88] or L_p-norm minimization [37] to reconstruct the sources, i.e.,

$$\begin{aligned} &\min_{\mathbf{s}(\tau)} \ \|\mathbf{s}(\tau)\|_p^p, \\ &\text{s.t. } \mathbf{x}(\tau) = \widehat{\mathbf{A}}\mathbf{s}(\tau). \end{aligned} \tag{3.45}$$

For separation of convolutive mixtures, it is found that $p \to 0$ yields the best result for the subspace minimization method in [37]. Note that, the algorithm in [37] may not yield a sparse solution for $p_{\text{crit}} < p < 1$ where $p_{\text{crit}} \lesssim 0.75$. Alternatively, one can solve the L_p-norm minimization problem using focal underdetermined system solution (FOCUSS) [49, 109–111],

$$\widehat{\mathbf{s}}(\tau) \leftarrow \mathbf{\Psi}^{-1}(\widehat{\mathbf{s}}(\tau))\widehat{\mathbf{A}}^H(\widehat{\mathbf{A}}\mathbf{\Psi}^{-1}(\widehat{\mathbf{s}}(\tau))\widehat{\mathbf{A}}^H)^{-1}\mathbf{x}(\tau), \tag{3.46}$$

where $\mathbf{\Psi}(\widehat{\mathbf{s}}(\tau)) = \mathrm{Diag}(|\widehat{s}_1(\tau)|^{p-2}, \ldots, |\widehat{s}_N(\tau)|^{p-2})$. For convergence, FOCUSS requires that the initial sources $\widehat{\mathbf{s}}(\tau)$ must all be strictly non-zero [112]. Nevertheless, since the L_p-norm is non-convex for $0 < p < 1$ and FOCUSS is only guaranteed to converge to stationary points, this algorithm may stuck at local minima. Therefore, the initial $\widehat{\mathbf{s}}(\tau)$ must be carefully chosen. In [96], it is shown that FOCUSS with a suitable choice of initial $\widehat{\mathbf{s}}(\tau)$ can yield a significantly better source reconstruction than several prior works. Interestingly, for convolutive

speech separation, as opposed to subspace method in [37], the FOCUSS method in [96] achieves the best performance when $p = 0.8$. Similarly to the determined scenario, once the sources are extracted, the minimum distortion principle or back-projection can be employed to solve the scaling ambiguity issue.

On the other hand, when the sources are sufficiently sparse, one can also use time-frequency masking [17, 38, 39, 87, 113] to extract the sources. This is achieved using

$$\widehat{\mathbf{s}}(\tau) = \mathbf{M}(\tau) \odot \mathbf{x}(\tau), \tag{3.47}$$

where $\mathbf{M}(\tau)$ is the spectral mask and the symbol $\odot$ denotes element-wise matrix multiplication. For directional clustering methods, the membership functions, e.g., $C_n^{(\mathrm{GMM})}(\tau)$ and $C_n^{(\mathrm{ACGMM})}(\tau)$ have been used for masking. Compared to L_p-norm minimization, time-frequency masking is simpler as well as more computationally efficient.

3.3 Performance criteria

To evaluate the performance of mixing matrix estimation in determined scenario, we choose the most common performance index which is the normalized Amari index (NAI) defined by [9, 114]

$$\begin{aligned}\mathrm{NAI} =& \frac{1}{2N(N-1)} \sum_{m=1}^{N} \left(\sum_{n=1}^{N} \frac{p_{mn}}{\max_l |p_{ml}|} - 1 \right) \\ &+ \frac{1}{2N(N-1)} \sum_{m=1}^{N} \left(\sum_{n=1}^{N} \frac{p_{mn}}{\max_l |p_{ln}|} - 1 \right),\end{aligned} \tag{3.48}$$

where $\mathbf{P} = [p_{mn}] = \widehat{\mathbf{A}}^{-1}\mathbf{A}$. The lower NAI qualifies the more accurate estimation of the mixing matrix. It is noted that a poor estimation of the mixing matrix will yield NAI that is higher than -10 dB [10]. Since NAI can not be computed for rectangular matrix, we will use the mixing-error-ration (MER) as

our performance index in under-determined scenario

$$\text{MER} = \frac{20}{N}\sum_{n=1}^{N}\log\left(\left\|\widehat{\mathbf{a}}_n^{\text{coll}}\right\| / \left\|\widehat{\mathbf{a}}_n^{\text{orth}}\right\|\right), \tag{3.49}$$

where $\widehat{\mathbf{a}}_n^{\text{coll}}$ and $\widehat{\mathbf{a}}_n^{\text{orth}}$ are, respectively, the collinear and the orthogonal component of $\widehat{\mathbf{a}}_n$ with respect to $\mathbf{a}_n$. In contrast with NAI, a higher value of MER indicates a better estimation of $\mathbf{A}$.

For the overall performance of a BSS algorithm, we use signal-distortion-ratio (SDR) and signal-interference-ratio (SIR) as our performance criteria [115, 116]. Let $s_{ij}(t)$ be the source image of the jth source at ith sensor, the estimated source $\widehat{s}_{ij}(t)$ can be decomposed into

$$\widehat{s}_{ij}(t) = s_{ij}(t) + e_{ij}^{\text{spat}}(t) + e_{ij}^{\text{interf}}(t) + e_{ij}^{\text{artif}}(t) \tag{3.50}$$

where $e_{ij}^{\text{spat}}(t)$, $e_{ij}^{\text{interf}}(t)$, and $e_{ij}^{\text{artif}}(t)$ are respectively spatial error, interference, and artifact. These errors are evaluated as follows

$$e_{ij}^{\text{spat}}(t) = \text{Proj}_j^{(L)}(\widehat{s}_{ij}(t)) - s_{ij}(t), \tag{3.51}$$

$$e_{ij}^{\text{interf}}(t) = \text{Proj}_{\text{all}}^{(L)}(\widehat{s}_{ij}(t)) - \text{Proj}_j^{(L)}(\widehat{s}_{ij}(t)), \tag{3.52}$$

$$e_{ij}^{\text{artif}}(t) = \widehat{s}_{ij}(t) - \text{Proj}_{\text{all}}^{(L)}(\widehat{s}_{ij}(t)), \tag{3.53}$$

where $\text{Proj}_j^{(L)}(\widehat{s}_{ij}(t))$ is the least-square projection of $\widehat{s}_{ij}(t)$ onto the subspace spanned by $s_{ij}(t-t_0), 0 \le i \le M, 0 \le t_0 \le L$, and $\text{Proj}_{\text{all}}^{(L)}(\widehat{s}_{ij}(t))$ is the least-square projection of $\widehat{s}_{ij}(t)$ onto the subspace spanned by $s_{ij}(t-t_0), 0 \le j \le N, 0 \le i \le M, 0 \le t_0 \le L$. Here, $L = 32$ ms is typically chosen for audio signals [116]. The SDR is then given by

$$\text{SDR} = \frac{1}{N}\sum_{j=1}^{N} 10\log\frac{\sum_{i,t} s_{ij}^2(t)}{\sum_{i,t}(e_{ij}^{\text{spat}}(t) + e_{ij}^{\text{interf}}(t) + e_{ij}^{\text{artif}}(t))^2}. \tag{3.54}$$

The SIR is defined as

$$\text{SIR} = \frac{1}{N}\sum_{j=1}^{N} 10\log\frac{\sum_{i,t} s_{ij}^2(t)}{\sum_{i,t}(e_{ij}^{\text{interf}}(t))^2}. \tag{3.55}$$

In particular, we also use the extended short-time objective intelligibility (ESTOI) [117] to evaluate the performance of speech separation.

The NAI values reported in this dissertation are calculated with "bss_isi" function from IVA toolbox[1] while SDR and SIR are computed using BSS_EVAL toolbox[2].

3.4 Summary

This chapter reviews frequency-domain BSS as an efficient approach for separation of convolutive mixtures. It can be seen that complex-valued mixing matrix estimation plays a central role in the performance of the overall BSS algorithm. Inaccuracy in mixing matrix estimation will affect subsequent steps such as source estimation and permutation alignment. Three main approaches for complex-valued mixing matrix estimation – ICA, SCA, and NMF – are discussed. Existing ICA methods use fixed-point iteration or projected gradient descent to optimize their constrained objective functions. Independence assumption is exploited to simplify the Hessian estimation in fixed-point ICA. The estimation of ICA filters is also carried out independently for each filter. These issues can limit the estimation performance of ICA. SCA methods, on the other hand, are based on directional soft-clustering and these techniques often employ Gaussian functions to model the distribution of the observed data. Since natural signals tend to exhibit sparseness, the use of existing SCA methods may not achieve desirable performance. Next, NMF models are introduced. These

[1] http://mlsp.umbc.edu/codes/
[2] http://bass-db.gforge.inria.fr/bss_eval/

models are highly sensitive to initialization, and therefore prior knowledge of the sources are often utilized to address this issue. The NMF models are also more suitable in supervised source separation whereby the NMF basis are trained on a corpus of signals. Finally, several source separation methods as well as two methods to address scaling ambiguity are discussed.

Chapter 4

Directional Sparse Filtering

4.1 Introduction

In this chapter, we focus on the problem of estimating the complex-valued latent filters from observed data when there are more filters than the data dimension (i.e., under-determined mixing process or over-complete representation). For convenience , we re-write the mixing process in (3.2) given by

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \sum_{n=1}^{N} \mathbf{a}_n s_n, \tag{4.1}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T \in \mathbb{C}^{M\times 1}$ is the observed data, $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T \in \mathbb{C}^{N\times 1}$ is the unknown source vector, and $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N] \in \mathbb{C}^{M\times N}$ is the latent mixing matrix given that $\mathbf{a}_n = [a_{1n}, a_{2n}, \ldots, a_{Mn}]^T \in \mathbb{C}^{M\times 1}$ denotes the nth column of $\mathbf{A}$. Note that, for brevity, we have removed the time-frame and frequency indices in (4.1) and remainder of this chapter.

When $M = N$, provided the mixing matrix is invertible, this learning problem can be addressed by independent component analysis (ICA) in which a set of directions is found so that the projections of data onto these directions are maximally non-Gaussian [23]. However, ICA is not applicable in under-determined cases where $M < N$ due to the non-invertible mixing matrix.

Fortunately, natural signals such as speech in the time-frequency domain or image in the wavelet domain are sparse; they have only a few non-zero elements [33, 34]. With sufficient degree of sparseness, the sources may be approximately disjoint-orthogonal, i.e., it is likely that there is only one dominant source at a particular sample index [35], i.e., the largest component of the source vector is sufficiently larger than all other components so that the mixed data concentrates about the directions given by columns of the mixing matrix. This can be seen as a special case of quasi-sparsity [118]. In this case, most of the observed data will lie along directions defined by columns of the mixing matrix. As an illustrative example, if the lth source is dominant, we have, from (4.1), $\mathbf{x} \approx \mathbf{a}_l s_l$, i.e., $\mathbf{x}$ and $\mathbf{a}_l$ are nearly collinear in the complex domain. We will refer to this type of sources and data as approximately disjoint sources [35] and directional data, respectively. More precisely, we assume that samples of the directional data are highly concentrated about the modes of the joint distribution, and that the mentioned modes coincide with the mixing matrix.

4.2 The proposed DSF-based method

4.2.1 Problem formulation

Let $\widehat{\mathbf{A}}$ be the estimate of the mixing matrix $\mathbf{A}$, $\widehat{\mathbf{s}}$ be the estimate of the source vector $\mathbf{s}$, and $g(\cdot)$ be the sparse-enforcing function. We consider the sparse filtering problem for under-determined mixing matrix estimation being defined as

$$\min_{\widehat{\mathbf{A}}} \ \mathcal{E}\{g(\widehat{\mathbf{A}}^H \mathbf{x})\} \text{ s.t. } \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}, \tag{4.2}$$

where the expectation is taken along the time-frame index, for each frequency bin. We note that, as opposed to the original sparse filtering algorithm in [44], the formulation above includes the semi-unitary constraint to exploit the fact

that the true mixing matrix is semi-unitary.

For conventional sparse penalty such as L_q-norm, where $g(\widehat{\mathbf{s}}) = \|\widehat{\mathbf{s}}\|_q$ given that $q > 0$, it can be seen that (4.2) yields trivial solutions (i.e., one or more columns of $\widehat{\mathbf{A}}$ are zero-vectors). Furthermore, such penalty depends on the norm of $\mathbf{x}$ and therefore yields a biased estimate for directional data. Taking cue from directional clustering and instead of assigning a sparse code for $\widehat{\mathbf{s}}$, a sparse code for the magnitude-squared cosine-similarity vector between $\widehat{\mathbf{A}}$ and $\mathbf{x}$ is assigned, and is given by

$$\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x}) = \left[\frac{\left|\widehat{\mathbf{a}}_1^H\mathbf{x}\right|^2}{\|\widehat{\mathbf{a}}_1\|_2^2\,\|\mathbf{x}\|_2^2}, \ldots, \frac{\left|\widehat{\mathbf{a}}_N^H\mathbf{x}\right|^2}{\|\widehat{\mathbf{a}}_N\|_2^2\,\|\mathbf{x}\|_2^2}\right]. \tag{4.3}$$

For directional data, at any instance, since the observed samples are more likely to be associated with one source than the other sources, the magnitude-squared cosine-similarity vector will exhibit one dominant component and several other non-dominant components. Therefore, striking a balance between increasing the largest component while suppressing the other components will recover the mixing matrix up to some scaling ambiguity.

Due to the normalization factor in each element of $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$ in (4.3), it is useful to note that $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$ is invariant to any non-zero scaling of $\mathbf{x}$ or $\widehat{\mathbf{a}}_i, \forall i$ and the proposed sparse filtering problem becomes

$$\min_{\widehat{\mathbf{A}}} \; \mathcal{E}\{g(\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x}))\} \text{ s.t. } \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}. \tag{4.4}$$

By definition, the elements of $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x}) = [u_1, \ldots, u_N]$ lie within closed interval of $[0, 1]$. Since there are at most M mutually orthogonal vectors in $\mathbb{C}^{M\times 1}$ and that $u_n = 0$ if and only if $\widehat{\mathbf{a}}_n$ is orthogonal to $\mathbf{x}$, the vector $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$ has at least $N-M+1$ non-zero elements. Hence, for the under-determined case, conventional sparse penalties that force $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$ to have as many zero elements as possible

do not take into account cases where $\mathbf{s}$ may have less than $N - M + 1$ active sources of which one source is dominant. It is therefore necessary to find a new class of sparse penalties for mixing matrix estimation from directional data.

4.2.2 The sparse penalty criteria for directional data

For directional data, a zero penalty for $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$ is proposed to be assigned whenever $\mathbf{x} \equiv \mathbf{x}_{\text{SSP}}$ is a single-source point, i.e., $\mathbf{x}_{\text{SSP}} = s_l \mathbf{a}_l$ for an arbitrary source index l, and whenever $\widehat{\mathbf{A}}$ differs from $\mathbf{A}$ only by some scaling and permutation of columns, i.e., $\widehat{\mathbf{A}} = \mathbf{ADP}$, where $\mathbf{D} \in \mathbb{C}^{N \times N}$ is the diagonal scaling matrix with non-zero diagonal entries and $\mathbf{P}$ is the column permutation matrix. From (4.3), since $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$ is invariant to any non-zero scale of $\mathbf{x}$ or any column of $\widehat{\mathbf{A}}$, we have $\mathbf{u}(\mathbf{ADP}; s_l \mathbf{a}_l) = \mathbf{u}(\mathbf{AP}; \mathbf{a}_l) = \mathbf{u}(\mathbf{A}; \mathbf{a}_l)\mathbf{P}$. For illustration, we assume $l = 1$ giving $\mathbf{u}(\mathbf{A}; \mathbf{a}_1) = [1, c_{21}, \ldots, c_{N1}]$, where c_{n1} is defined as the magnitude-squared cosine-similarity between the nth and the first column of $\mathbf{A}$. It is evident that, for the under-determined case, the values of c_{n1} are dependent on the mixing matrix, and they can all be non-zero even when $\mathbf{s}$ has only one non-zero element. As a result, for the proposed algorithm, emphasis is placed on promoting the dominant element of $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$ while suppressing the non-dominant elements. For an accurate estimation of the mixing matrix, it is therefore argued that $g(\mathbf{u})$ shall satisfy the following three criteria:

1. The penalty function $g(\mathbf{u})$ should decrease when the largest element of $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$ increases. This criterion is inspired by directional clustering where the aim is to reinforce the dominant element of $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$.

2. The value of $g(\mathbf{u})$ should be the same and should be minimum for any $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$ of which the largest element is 1, regardless of the values of non-dominant elements. Otherwise, $g(\mathbf{u})$ should be dependent on all elements of $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$. This criterion guarantees that we treat all single-source points,

which contain most valuable information of the mixing matrix, in a fair and favorable way. It also ensures that we will not lose useful directional information in points that are designated as approximate single-source points. Note that we do not identify single-source points nor approximately single-source points directly.

3. The function $g(\mathbf{u})$ should be a Schur-concave function since such a function promotes sparsity of $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$ [47]. This criterion is motivated by sparse filtering where the proposed algorithm attempts to suppress non-dominant elements of $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$. This effectively limits a sub-set of directions to be active within a sample.

Existing contrast functions do not satisfy all these three criteria. In particular, the non-linearities used in ICA or SCA methods that are based on L_1 norm or L_1/L_2 norm ratio (for instance, $g^{(\mathrm{SCA})}(\mathbf{u}) = \sum_n \sqrt{a_1 + u_n}$, where a_1 is a small constant) do not meet the first two criteria. The first criterion is invalid because $g^{(\mathrm{SCA})}(\mathbf{u})$ will increase when the largest components of $\mathbf{u}$ increases while the remainder components of $\mathbf{u}$ are fixed. The second criterion is also invalid since $g^{(\mathrm{SCA})}(\mathbf{u})$ is an increasing function w.r.t. all u_n. On the other hand, while the non-linear of KHL $g^{(\mathrm{KHL})}(\mathbf{u}) = \min_n(1-u_n)$ satisfies the first and third criteria, it completely removes the effect of non-dominant elements in $\mathbf{u}$. Therefore, it is sub-optimal for imperfectly disjoint sources where there are more than one active sources in a data sample. For GMM method [14, 16, 17], the contrast function corresponding to the negative log-likelihood is given by

$$g^{(\mathrm{GMM})}(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\sigma}) = -\ln \sum_n \alpha_n \exp(-\frac{1-u_n}{\sigma_n^2}), \tag{4.5}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ are additional parameters of the model. For finite values of $\boldsymbol{\sigma}$ and non-zero values of $\boldsymbol{\alpha}$, $g^{(\mathrm{GMM})}(\mathbf{u}, \boldsymbol{\alpha}, \boldsymbol{\sigma})$ yields different penalties for $\mathbf{u}(\widehat{\mathbf{A}};\mathbf{x})$ whose largest element is 1, depending on the value of non-dominant elements of

$\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$. This implies that the model may assign different penalties for single-source points corresponding to different sources, which, as a result, does not satisfy the second criterion.

4.2.3 The proposed sparse penalty for directional data

The use of a smooth upper-bound of $g^{(\mathrm{KHL})}(\mathbf{u})$ is proposed in this thesis. This upper-bound is based on the power mean with negative exponent of $1-\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x})$, as a family of contrast functions that satisfies all the three criteria stated above. The proposed contrast function is given by

$$g^{(\mathrm{DSF})}(\mathbf{u};\ r) = \left[\frac{1}{N}\sum_{n}(1-u_n)^r\right]^{1/r},\ r<0, \tag{4.6}$$

where r is a hyper-paremeter and the superscript DSF denotes for the directional sparse filtering. Defining $u_{\max}$ and $n_{\max}$ as the largest element of $\mathbf{u}$ and its corresponding index, respectively, and by factoring out the term $(1-u_{\max})$, (4.6) can be expressed as

$$g^{(\mathrm{DSF})}(\mathbf{u};\ r) = (1-u_{\max})\left[\frac{1}{N}+\frac{1}{N}\sum_{n\neq n_{\max}}\left(\frac{1-u_n}{1-u_{\max}}\right)^r\right]^{1/r}. \tag{4.7}$$

The above equation shows that $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ satisfies the first criterion since its value decreases with increasing $u_{\max}$. To verify the second criterion, we first note that $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ is an increasing function w.r.t. r due to the power mean inequality [119, p. 203], and therefore $g^{(\mathrm{DSF})}(\mathbf{u};\ r) \geq g^{(\mathrm{DSF})}(\mathbf{u};\ -\infty)$. In addition, since only the term $(1-u_{\max})$ in (4.6) may become zero (because $0 \leq u_n \leq 1$ and we assume none of the column pairs of $\mathbf{A}$ are collinear), taking

the limits of $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ given in (4.7) yields

$$\lim_{u_{\max}\to 1} g^{(\mathrm{DSF})}(\mathbf{u};\ r) = 0, \tag{4.8}$$

$$\lim_{r\to-\infty}\ g^{(\mathrm{DSF})}(\mathbf{u};\ r) = \min_n(1-u_n) = 1-u_{\max}. \tag{4.9}$$

Hence, any $\mathbf{u}$ with $u_{\max}=1$ is assigned the same smallest penalty, which, in our case, is zero. For the third criterion, $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ is strictly Schur-concave since it is a composition of the power mean function (which is strictly Schur-concave for $r<1$ and $r\neq 0$) and a linear function [56].

To gain insights into the proposed contrast function, we examine the partial gradient of $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ given by

$$\frac{\partial g^{(\mathrm{DSF})}(\mathbf{u};\ r)}{\partial u_n} = -\frac{1}{N}\left(\frac{g^{(\mathrm{DSF})}(\mathbf{u};\ r)}{1-u_n}\right)^{1-r}, \tag{4.10}$$

$$\lim_{u_{\max}\to 1}\frac{\partial g^{(\mathrm{DSF})}(\mathbf{u};\ r)}{\partial u_n} = \begin{cases} -\frac{1}{N^{1/r}}, & n\equiv n_{\max};\\ 0, & \text{otherwise.}\end{cases} \tag{4.11}$$

The above partial gradient indicates that the proposed contrast function offers a trade-off between $u_{\max}$ and other elements of $\mathbf{u}$; other elements in $\mathbf{u}$ contribute lesser information in the learning process when $u_{\max}\to 1$ or $r\to-\infty$. This is derived from the fact that the absolute value of $\frac{\partial}{\partial u_n}g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ reduces with reducing $\left(\frac{1-u_{\max}}{1-u_n}\right)^{1-r}$. Moreover, it can be seen that $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ satisfies the analytical requirement of Schur-concavity given in (2.12) because

$$\begin{aligned}&(u_n-u_m)\left(\frac{\partial g^{(\mathrm{DSF})}(\mathbf{u};\ r)}{\partial u_n}-\frac{\partial g^{(\mathrm{DSF})}(\mathbf{u};\ r)}{\partial u_m}\right)\\ =&-\frac{(g^{(\mathrm{DSF})}(\mathbf{u};\ r))^{1-r}}{N}(u_n-u_m)\left(\frac{1}{(1-u_n)^{1-r}}-\frac{1}{(1-u_m)^{1-r}}\right)\\ \leq&0,\end{aligned} \tag{4.12}$$

where the last step is due to non-negativity of $g^{(\mathrm{DSF})}(\mathbf{u};\ r)$ and that the product

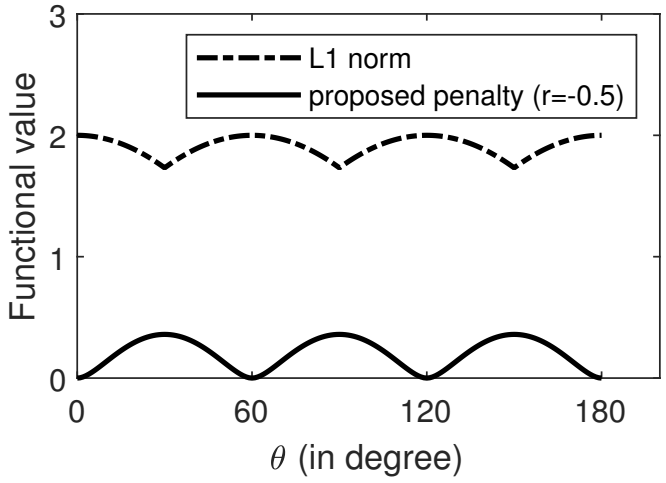


Fig. 4.1: The values of L_1 norm and the proposed penalty as functions of θ for an arbitrary data vector $\mathbf{x} = [\cos\theta,\ \sin\theta]^T$ when $\mathbf{A} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$.

of the last two multiplicative terms is non-negative for $r < 0$.

To illustrate the proposed contrast function compared to a simple sparsity penalty function such as the L_1 norm, let us consider an under-determined matrix $\mathbf{A} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$. We also assume the sample vector to be $\mathbf{x} = [\cos\theta,\ \sin\theta]^T$. Figure 4.1 depicts the value of L_1 norm and the proposed penalty function (for $r = -0.5$) evaluated at $\widehat{\mathbf{A}} \equiv \mathbf{A}$ for θ from 0° to 180°. It can be seen that the proposed contrast function is globally minimized when $\mathbf{x}$ lies in the directions specified by the columns of $\mathbf{A}$, i.e., $\theta = 0^\circ$, 60°, 120°, and 180° (with $\theta = 0^\circ$ and $\theta = 180^\circ$ specifying the same direction). On the other hand, L_1 norm is globally minimized when $\theta = 30^\circ$, 90°, and 150°, i.e., whenever $\mathbf{x}$ is orthogonal to a column of $\mathbf{A}$. The L_1 norm also assigns the highest penalty for points distributed about the columns of $\mathbf{A}$ ($\theta = 0^\circ$, 60°, 120°, and 180°). This shows that the proposed penalty is more suitable for mixing matrix estimation.

With reference from (4.4) and the contrast function in (4.6), the proposed optimization problem for under-determined mixing matrix estimation is then

given by

$$\min_{\widehat{\mathbf{A}}} J^{(\mathrm{DSF})}(\widehat{\mathbf{A}};\ r) \quad \text{s.t. } \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}, \tag{4.13}$$

where

$$J^{(\mathrm{DSF})}(\widehat{\mathbf{A}};\ r) = \mathcal{E}\Big\{\Big[\frac{1}{N}\sum_{n=1}^{N}\Big(1 - \frac{\left|\widehat{\mathbf{a}}_n^H\mathbf{x}\right|^2}{\|\widehat{\mathbf{a}}_n\|_2^2\,\|\mathbf{x}\|_2^2}\Big)^r\Big]^{1/r}\Big\}. \tag{4.14}$$

4.2.4 Unconstrained re-parameterization for semi-unitary constrained optimization problems

It is important to note that the above optimization is not convex since multiple local minima exist for $J^{(\mathrm{DSF})}(\widehat{\mathbf{A}};\ r)$ and that the set $\{\widehat{\mathbf{A}} \in \mathbb{C}^{N\times M} | \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}\}$ is not convex. For example, if $\widehat{\mathbf{A}}_{\mathrm{opt}}$ is a minimizer of $J^{(\mathrm{DSF})}(\widehat{\mathbf{A}};\ r)$ then any column-permutation of $\widehat{\mathbf{A}}$ is also a minimizer. The semi-unitary constraint is also not convex because summation of two semi-unitary matrices is not always semi-unitary. Therefore, we propose to simplify the optimization of this semi-unitary constrained problem by re-parameterizing (4.13).

We first review the concepts related to the optimization on Stiefel manifolds [120–122]. The set $St(M,N) = \{\widehat{\mathbf{A}} \in \mathbb{C}^{M\times N} | \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}\}$ is known as the complex Stiefel manifold[1]. Given a function $f(\widehat{\mathbf{A}}) : St(M,N) \mapsto \mathbb{R}$ and its Euclidean gradient $\nabla_{\widehat{\mathbf{A}}^*} f$, $\widehat{\mathbf{A}}$ is a stationary point of $f(\widehat{\mathbf{A}})$ if

$$\nabla_{\widehat{\mathbf{A}}^*} f - \widehat{\mathbf{A}}(\nabla_{\widehat{\mathbf{A}}^*} f)^H \widehat{\mathbf{A}} = \mathbf{0}. \tag{4.15}$$

For $\widehat{\mathbf{A}} \in St(M,N)$, the tangent space of $St(M,N)$ at $\widehat{\mathbf{A}}$ is found by differenti-

[1] In this thesis, the definition of the Stiefel manifold is different from the definition given in [120] by a Hermitian transpose.

ating $\widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}$, i.e.,

$$\mathcal{T}_{\widehat{\mathbf{A}}}(M,N) = \{\mathbf{T} \in \mathbb{C}^{M\times N} | \mathbf{T}\widehat{\mathbf{A}}^H + \widehat{\mathbf{A}}\mathbf{T}^H = \mathbf{0}\}. \tag{4.16}$$

The normal space of $St(M,N)$ at $\widehat{\mathbf{A}}$ given by

$$\mathcal{N}_{\widehat{\mathbf{A}}}(M,N) = \{\mathbf{W}\widehat{\mathbf{A}} | \mathbf{W} \in \mathbb{C}^{M\times M}, \mathbf{W} = \mathbf{W}^H\} \tag{4.17}$$

is the orthogonal complement of $\mathcal{T}_{\widehat{\mathbf{A}}}(M,N)$ w.r.t. the standard Euclidean inner product, i.e., $\langle \mathbf{T}, \mathbf{N}\rangle = \Re\{\text{trace}\{\mathbf{T}\mathbf{N}^H\}\} = \mathbf{0}$ for any $\mathbf{T} \in \mathcal{T}_{\widehat{\mathbf{A}}}(M,N)$ and $\mathbf{N} \in \mathcal{N}_{\widehat{\mathbf{A}}}(M,N)$. In fact, $\mathcal{N}_{\widehat{\mathbf{A}}}(M,N)$ is constituted by full-row rank matrices in $\mathbb{C}^{M\times N}$ that have $\widehat{\mathbf{A}}$ as their nearest semi-unitary matrix in least-square sense [120, Proposition 7].

Instead of finding a semi-unitary matrix that minimizes a particular cost function $J(\widehat{\mathbf{A}})$, i.e.,

$$\min_{\widehat{\mathbf{A}}} \; J(\widehat{\mathbf{A}}) \text{ s.t. } \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}, \tag{4.18}$$

the proposed re-parameterization technique aims to find a full-row rank matrix such that its nearest semi-unitary matrix will minimize the cost function, i.e.,

$$\min_{\mathbf{B}} \; J(\widehat{\mathbf{A}}) \text{ s.t. } \widehat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}. \tag{4.19}$$

Lemma 4.1. *Let* $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ *be the economy-sized singular value decomposition (eSVD) of a full row-rank matrix* $\mathbf{B}$*, where* $\mathbf{U} \in \mathbb{C}^{M\times M}$ *is unitary,* $\mathbf{\Sigma} \in (0,\infty)^{M\times M}$ *is diagonal,* $\mathbf{V} \in \mathbb{C}^{N\times M}$ *is semi-unitary. The nearest semi-unitary matrix of* $\mathbf{B}$ *in least-square sense is then unique and is given by* $\widehat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B} = \mathbf{U}\mathbf{V}^H$ *[120]. Let* $\boldsymbol{\sigma} = \text{diag}(\mathbf{\Sigma}) \in (0,\infty)^{\text{M}\times 1}$ *be the diagonal vector of* $\mathbf{\Sigma}$*, and* $\mathbf{1}_M \in \mathbb{C}^{M\times 1}$ *be the all-one column vector. For the optimization problem given in (4.19), let* $\nabla_{\widehat{\mathbf{A}}^*} J$ *denote the Euclidean gradient of* $J(\widehat{\mathbf{A}})$ *w.r.t.* $\widehat{\mathbf{A}}^*$*. The Euclidean gradient of the nested cost* $J((\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B})$ *w.r.t.* $\mathbf{B}^*$ *is then*

given by

$$\nabla_{\mathbf{B}^*} J = \mathbf{U}(\mathbf{C}^H + \mathbf{C})\mathbf{\Sigma}\mathbf{V}^H + \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^H \nabla_{\widehat{\mathbf{A}}^*} J, \tag{4.20}$$

where

$$\mathbf{C} = -(\mathbf{\Sigma}^{-1}\mathbf{U}^H(\nabla_{\widehat{\mathbf{A}}^*} J)\mathbf{V}) \oslash (\mathbf{1}_M \boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_M^T). \tag{4.21}$$

Furthermore, we have $\nabla_{\mathbf{B}^*} J \in \mathcal{T}_{\widehat{\mathbf{A}}}(M, N)$ *and*

$$\nabla_{\mathbf{B}^*} J = \mathbf{0} \Leftrightarrow \nabla_{\widehat{\mathbf{A}}^*} J - \widehat{\mathbf{A}}(\nabla_{\widehat{\mathbf{A}}^*} J)^H \widehat{\mathbf{A}} = \mathbf{0}. \tag{4.22}$$

Proof. Let $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{W}$ be some matrices with appropriate shapes. The Kronecker product between $\mathbf{X}$ and $\mathbf{Y}$ is defined by

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} x_{11}\mathbf{Y} & \cdots & x_{1N}\mathbf{Y} \\ \vdots & \ddots & \vdots \\ x_{M1}\mathbf{Y} & \cdots & x_{MN}\mathbf{Y} \end{bmatrix}. \tag{4.23}$$

Two important identities of the Kronecker product will often be used in this proof,

$$(\mathbf{X} \otimes \mathbf{Y})(\mathbf{Z} \otimes \mathbf{W}) = (\mathbf{X}\mathbf{Z}) \otimes (\mathbf{Y}\mathbf{W}), \tag{4.24}$$

$$(\mathbf{X} \otimes \mathbf{Y})^{\mathrm{op}} = \mathbf{X}^{\mathrm{op}} \otimes \mathbf{Y}^{\mathrm{op}}, \tag{4.25}$$

where op is the matrix inversion, or transposition, or Hermitian transposition operator. The two properties of the vectorization operator,

$$\mathrm{vec}(\mathbf{X}\mathbf{Y}\mathbf{Z}) = (\mathbf{Z}^T \otimes \mathbf{X})\mathrm{vec}\mathbf{Y}, \tag{4.26}$$

$$\mathbf{K}_{MN}\mathrm{vec}\mathbf{X} = \mathrm{vec}(\mathbf{X}^T), \tag{4.27}$$

will also be utilized where $\mathbf{K}_{MN} = \mathbf{K}_{NM}^T$ is the commutation matrix.

Consider the matrix function $\widehat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}$ such that $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ is the

economical singular value decomposition of $\mathbf{B} \in \mathbb{C}^{M\times N}$, the following matrices $\mathbf{B}_1 = \mathbf{B}\mathbf{B}^H = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^H$, $\mathbf{B}_2 = \mathbf{B}_1^{-1} = \mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^H$, $\mathbf{B}_3 = \mathbf{B}_2^{1/2} = \mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^H$ and $\widehat{\mathbf{A}} = \mathbf{B}_3\mathbf{B} = \mathbf{U}\mathbf{V}^H$ are defined. For a generic function $\mathbf{F}(\mathbf{Z}_0, \mathbf{Z}_1)$, the matrix differential $d\mathbf{F}(\mathbf{Z}_0, \mathbf{Z}_1)$ is the first order term w.r.t. $d\mathbf{Z}_0$ and $d\mathbf{Z}_1$ of $\mathbf{F}(\mathbf{Z}_0 + d\mathbf{Z}_0, \mathbf{Z}_1 + d\mathbf{Z}_1) - \mathbf{F}(\mathbf{Z}_0, \mathbf{Z}_1)$ [54], thus

$$d(\mathbf{Z}_0\mathbf{Z}_1) = \mathbf{Z}_0 \cdot d\mathbf{Z}_1 + d\mathbf{Z}_0 \cdot \mathbf{Z}_1. \tag{4.28}$$

Applying (4.28) on $\mathbf{B}_3\mathbf{B}_3 = \mathbf{B}_2$, $\mathbf{B}_2\mathbf{B}_1 = \mathbf{I}_M$, and $\mathbf{B}_1 = \mathbf{B}\mathbf{B}^H$ in their respective orders gives

$$d\mathbf{B}_2 = d\mathbf{B}_3 \cdot \mathbf{B}_3 + \mathbf{B}_3 \cdot d\mathbf{B}_3, \tag{4.29}$$

$$d\mathbf{B}_1 = -\mathbf{B}_2^{-1} \cdot d\mathbf{B}_2 \cdot \mathbf{B}_1, \tag{4.30}$$

$$d\mathbf{B}_1 = d\mathbf{B} \cdot \mathbf{B}^H + \mathbf{B} \cdot d\mathbf{B}^H. \tag{4.31}$$

The above equations imply that

$$\mathbf{B}_2^{-1} \cdot d\mathbf{B}_3 \cdot \mathbf{B}_3^{-1} + \mathbf{B}_3^{-1} \cdot d\mathbf{B}_3 \cdot \mathbf{B}_2^{-1} = -d\mathbf{B} \cdot \mathbf{B}^H - \mathbf{B} \cdot d\mathbf{B}^H. \tag{4.32}$$

By applying the vectorization operator on both sides of the above equation, the following relationship

$$(\mathbf{B}_3^{-T} \otimes \mathbf{B}_2^{-1} + \mathbf{B}_2^{-T} \otimes \mathbf{B}_3^{-1}) d\mathrm{vec}\mathbf{B}_3 = -(\mathbf{B}^* \otimes \mathbf{I}_M) d\mathrm{vec}\mathbf{B} - (\mathbf{I}_N \otimes \mathbf{B})\mathbf{K}_{MN} d\mathrm{vec}\mathbf{B}^* \tag{4.33}$$

is obtained. Evidently, $\mathbf{U}^* \otimes \mathbf{U}$ is the eigen-matrix of both $\mathbf{B}_3^{-T} \otimes \mathbf{B}_2^{-1}$ and $\mathbf{B}_2^{-T} \otimes \mathbf{B}_3^{-1}$ since

$$\mathbf{B}_3^{-T} \otimes \mathbf{B}_2^{-1} = (\mathbf{U}^* \otimes \mathbf{U})(\mathbf{\Sigma} \otimes \mathbf{\Sigma}^2)\mathbf{U}^T \otimes \mathbf{U}^H, \tag{4.34}$$

$$\mathbf{B}_2^{-T} \otimes \mathbf{B}_3^{-1} = (\mathbf{U}^* \otimes \mathbf{U})(\mathbf{\Sigma}^2 \otimes \mathbf{\Sigma})\mathbf{U}^T \otimes \mathbf{U}^H. \tag{4.35}$$

Therefore by expanding the left side of (4.33), we have

$$(\mathbf{B}_3^{-T}\otimes\mathbf{B}_2^{-1}+\mathbf{B}_2^{-T}\otimes\mathbf{B}_3^{-1})^{-1} = (\mathbf{U}^*\otimes\mathbf{U})(\mathbf{\Sigma}\otimes\mathbf{\Sigma}^2+\mathbf{\Sigma}^2\otimes\mathbf{\Sigma})^{-1}(\mathbf{U}^T\otimes\mathbf{U}^H). \quad (4.36)$$

Next, since $\widehat{\mathbf{A}} = \mathbf{B}_3\mathbf{B}$, we have

$$d\mathrm{vec}\widehat{\mathbf{A}} = (\mathbf{B}^T\otimes\mathbf{I}_M)d\mathrm{vec}\mathbf{B}_3 + (\mathbf{I}_N\otimes\mathbf{B}_3)d\mathrm{vec}\mathbf{B}. \quad (4.37)$$

Defining the diagonal matrix $\mathbf{\Lambda} = (\mathbf{\Sigma}\otimes\mathbf{I}_M + \mathbf{I}_M\otimes\mathbf{\Sigma})^{-1}$, from (4.37), (4.36), (4.33), (2.6) and the singular value decomposition, the following Jacobians

$$\mathcal{D}_{\mathbf{B}}\widehat{\mathbf{A}} = (\mathbf{I}_N\otimes\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^H) - (\mathbf{V}^*\otimes\mathbf{U}\mathbf{\Sigma}^{-1})\mathbf{\Lambda}(\mathbf{\Sigma}\mathbf{V}^T\otimes\mathbf{U}^H), \quad (4.38)$$

$$\mathcal{D}_{\mathbf{B}^*}\widehat{\mathbf{A}} = -(\mathbf{V}^*\otimes\mathbf{U}\mathbf{\Sigma}^{-1})\mathbf{\Lambda}(\mathbf{U}^T\otimes\mathbf{\Sigma}\mathbf{V}^H)\mathbf{K}_{MN} \quad (4.39)$$

are obtained, where $\mathbf{K}_{MN}$ denotes the commutation matrix in $\mathbb{C}^{M\times N}$ such that $\mathbf{K}_{MN} = \mathbf{K}_{NM}^T$ and $\mathbf{K}_{MN}\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(\mathbf{X}^T), \forall\mathbf{X}$. We note that the unique structure of $\mathbf{\Lambda}$ and the following identity for any matrix $\mathbf{M}\in\mathbb{C}^{M\times M}$,

$$\mathbf{\Lambda}\mathrm{vec}(\mathbf{M}) = \mathrm{vec}\left(\mathbf{M}\oslash(\mathbf{1}_M\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_M^T)\right). \quad (4.40)$$

Using (4.40) and (2.9), the gradient given in (4.21) and (4.20) is finally obtained.

To prove the equivalence between optimization on Stiefel manifold and the proposed re-parameterized objective function, we note that

$$\mathbf{\Sigma}[\mathbf{M}\oslash(\mathbf{1}_M\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_M^T)] + [\mathbf{M}\oslash(\mathbf{1}_M\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_M^T)]\mathbf{\Sigma} = \mathbf{M}, \quad (4.41)$$

$$\mathbf{\Sigma}[(\mathbf{\Sigma}^{-1}\mathbf{M} + \mathbf{M}\mathbf{\Sigma}^{-1})\oslash(\mathbf{1}_M\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_M^T)]\mathbf{\Sigma} = \mathbf{M}. \quad (4.42)$$

The two identities above can be shown by vectorizing both sides of these equations. From (4.16), to show that $\nabla_{\mathbf{B}^*}J \in \mathcal{T}_{\widehat{\mathbf{A}}}(M,N)$, it is sufficient to verify

that

$$(\nabla_{\mathbf{B}^*} J)\widehat{\mathbf{A}}^H + \widehat{\mathbf{A}}(\nabla_{\mathbf{B}^*} J)^H = 0. \tag{4.43}$$

Indeed, due to (4.41) and (4.21), we have

$$\begin{aligned}(\nabla_{\mathbf{B}^*} J)\widehat{\mathbf{A}}^H + \widehat{\mathbf{A}}(\nabla_{\mathbf{B}^*} J)^H =& (\nabla_{\mathbf{B}^*} J)\mathbf{V}\mathbf{U}^H + \mathbf{U}\mathbf{V}^H(\nabla_{\mathbf{B}^*} J)^H \\ =& \mathbf{U}[(\mathbf{C}^H + \mathbf{C})\mathbf{\Sigma} + \mathbf{\Sigma}^{-1}\mathbf{U}^H(\nabla_{\widehat{\mathbf{A}}^*} J)\mathbf{V}]\mathbf{U}^H \\ &+ \mathbf{U}[\mathbf{\Sigma}(\mathbf{C}^H + \mathbf{C}) + \mathbf{V}^H(\nabla_{\widehat{\mathbf{A}}^*} J)^H\mathbf{U}\mathbf{\Sigma}^{-1}]\mathbf{U}^H \\ =& \mathbf{0}. \end{aligned} \tag{4.44}$$

Now suppose $\nabla_{\mathbf{B}^*} J = \mathbf{0}$, multiplying both sides of (4.20) with $\mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$, the gradient $\nabla_{\widehat{\mathbf{A}}^*} J$ is obtained, giving

$$\nabla_{\widehat{\mathbf{A}}^*} J = -\mathbf{U}\mathbf{\Sigma}(\mathbf{C}^H + \mathbf{C})\mathbf{\Sigma}\mathbf{V}^H. \tag{4.45}$$

Therefore, since $\widehat{\mathbf{A}} = \mathbf{U}\mathbf{V}^H$, we have

$$\nabla_{\mathbf{B}^*} J = \mathbf{0} \implies \nabla_{\widehat{\mathbf{A}}^*} J - \widehat{\mathbf{A}}(\nabla_{\widehat{\mathbf{A}}^*} J)^H\widehat{\mathbf{A}} = \mathbf{0}. \tag{4.46}$$

To prove the reverse implication, assuming $\nabla_{\widehat{\mathbf{A}}^*} J - \widehat{\mathbf{A}}(\nabla_{\widehat{\mathbf{A}}^*} J)^H\widehat{\mathbf{A}} = \mathbf{0}$, we have

$$\mathbf{U}^H(\nabla_{\widehat{\mathbf{A}}^*} J)\mathbf{V} = \mathbf{V}^H(\nabla_{\widehat{\mathbf{A}}^*} J)^H\mathbf{U}. \tag{4.47}$$

It then follows from (4.21), (4.20), and (4.42) that

$$\nabla_{\mathbf{B}^*} J = \mathbf{0} \impliedby \nabla_{\widehat{\mathbf{A}}^*} J - \widehat{\mathbf{A}}(\nabla_{\widehat{\mathbf{A}}^*} J)^H\widehat{\mathbf{A}} = \mathbf{0}. \tag{4.48}$$

□

Since $\nabla_{\mathbf{B}^*} J \in \mathcal{T}_{\widehat{\mathbf{A}}}(M, N)$, the gradient of the modified optimization problem will adapt to the curvature of the Stiefel manifold. Moreover, (4.15) and (4.22)

imply that $\mathbf{B}$ is a stationary point of (4.19) if and only if $\widehat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}$ is a stationary point of (4.18). This implies that the proposed approach is equivalent to optimizing directly on the Stiefel manifold. Nevertheless, the proposed method is unconstrained in the Euclidean space, and therefore one can employ existing algorithms such as the classical gradient descent, Nesterov accelerated gradient [123], or the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [124]. Furthermore, the proposed re-parameterization approach can be used in other semi-unitary constrained problems such as orthogonal sparse principle component analysis [125], ICA, and unitary recurrent neural network [126].

In practice, $\|\nabla_{\mathbf{B}^*} J\|_F$ is inversely proportional to the singular values of $\mathbf{B}$ due to (4.20) and (4.21). Therefore $\|\nabla_{\mathbf{B}^*} J\|_F$ can be arbitrarily small if $\|\mathbf{B}\|_F$ is large. This implies that the convergence rate is reduced. This issue can be mitigated by adding a penalty term that is dependent on $\|\mathbf{B}\|_F$ or by restarting the optimization procedure with the initial value of $(\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}$ when $\|\mathbf{B}\|_F$ is larger than a pre-defined threshold. Another important concern is that $\mathbf{B}$ may not maintain its row rank during the optimization process. Lemma 4.2 below highlights that this is not the case if $\mathbf{B}$ is initialized as a full row-rank matrix and the search direction of the optimization procedure belongs to the tangent space of $\mathcal{T}_{\widehat{\mathbf{A}}}(M, N)$, where $\widehat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}$.

Lemma 4.2. *Given a full row-rank matrix $\mathbf{B} \in \mathbb{C}^{M\times N}$, the matrix $\mathbf{B} + \mathbf{T}$ is full row rank for any $\mathbf{T} \in \mathcal{T}_{\widehat{\mathbf{A}}}(M, N)$, where $\widehat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}$.*

Proof. This is an extension of [120, Lemma 13]. First, one can factor any $\mathbf{T} \in \mathcal{T}_{\widehat{\mathbf{A}}}(M, N)$ as $\mathbf{T} = \mathbf{Y}\widehat{\mathbf{A}} + \mathbf{O}\widehat{\mathbf{A}}_\perp$ [120], where $\mathbf{Y} \in \mathbb{C}^{M\times M}$ is skew-Hermitian (i.e., $\mathbf{Y} + \mathbf{Y}^H = \mathbf{0}$), $\mathbf{O} \in \mathbb{C}^{M\times(N-M)}$ is an arbitrarily defined matrix, and $\widehat{\mathbf{A}}_\perp \in \mathbb{C}^{(N-M)\times N}$ is an arbitrarily chosen matrix such that $\widehat{\mathbf{A}}_\perp\widehat{\mathbf{A}}^H = \mathbf{0}$ and $\widehat{\mathbf{A}}_\perp\widehat{\mathbf{A}}_\perp^H = \mathbf{I}_{N-M}$. Suppose that $\mathbf{B} + \mathbf{T}$ is not full row rank, then there exists a non-zero vector $\mathbf{w} \in \mathbb{C}^{M\times 1}$ such that $\mathbf{w}^H(\mathbf{B} + \mathbf{T}) = \mathbf{0}$. Therefore,

$\mathbf{w}^H(\mathbf{B}+\mathbf{T})\widehat{\mathbf{A}}^H = \mathbf{w}^H(\mathbf{B}\mathbf{B}^H)^{1/2} + \mathbf{w}^H\mathbf{Y} = \mathbf{0}$. This implies that $\mathbf{w}^H\mathbf{Y}\mathbf{w} = \mathbf{w}^H\mathbf{Y}^H\mathbf{w} = -\mathbf{w}^H(\mathbf{B}\mathbf{B}^H)^{1/2}\mathbf{w}$. At the same time, $\mathbf{w}^H\mathbf{Y}\mathbf{w} = -\mathbf{w}^H\mathbf{Y}^H\mathbf{w}$ since $\mathbf{Y}$ is skew-Hermitian. Consequently, it is required that $\mathbf{w}^H(\mathbf{B}\mathbf{B}^H)^{1/2}\mathbf{w} = \mathbf{0}$. This is not feasible because $(\mathbf{B}\mathbf{B}^H)^{1/2}$ is positive definite, given that $\mathbf{B}$ is full row rank. □

4.2.5 Implementation of the proposed DSF algorithm

For the proposed cost function in (4.13), the gradient of $J^{(\mathrm{DSF})}(\widehat{\mathbf{A}};\ r)$ w.r.t to each $\widehat{\mathbf{a}}_n^*$ is given by

$$\nabla_{\widehat{\mathbf{a}}_n^*} J^{(\mathrm{DSF})}(\widehat{\mathbf{A}};\ r) = \frac{\mathbf{g}}{\|\widehat{\mathbf{a}}_n\|_2} - \frac{\widehat{\mathbf{a}}_n}{\|\widehat{\mathbf{a}}_n\|_2^3}\Re\{\widehat{\mathbf{a}}_n^H\mathbf{g}\}, \tag{4.49}$$

where

$$\mathbf{g} = 2E\Big\{\frac{\mathbf{x}}{\|\mathbf{x}\|_2}\Big(\frac{\mathbf{x}^H\widehat{\mathbf{a}}_n}{\|\mathbf{x}\|_2\,\|\widehat{\mathbf{a}}_n\|_2}\Big)\frac{\partial}{\partial u_n}g^{(\mathrm{DSF})}(\mathbf{u};\ r)\Big\}. \tag{4.50}$$

Assuming T data samples in total where $T \gg M$, the time complexity to evaluate $\nabla_{\widehat{\mathbf{A}}^*}J$ using (4.49) and (4.50) is $O(TNM)$. The time complexity to compute $\nabla_{\mathbf{B}^*}J$ from $\nabla_{\widehat{\mathbf{A}}^*}J$, which requires $O(NM^2)$, is negligible. Therefore, the total time complexity of each gradient update is $O(TNM)$. For comparison, the cost per iteration of the mixture model of line orientations [14, 16, 17] is $O(TNM^2)$ due to the computation of N weighted covariance matrices. As opposed to [127] where the re-parameterized cost is minimized using the Nesterov accelerated gradient, the proposed algorithm employs the L-BFGS algorithm since L-BFGS exhibits higher convergence rate [124] [123].

The proposed DSF method is implemented using the minFunc toolbox [128]. Since minFunc expects a vectorized real-valued gradient, $\Re\left\{\nabla_{\mathbf{B}^*}J^{(\mathrm{DSF})}\right\}$ and $\Im\left\{\nabla_{\mathbf{B}^*}J^{(\mathrm{DSF})}\right\}$ are vectorized and concatenated into a real-valued composite vector. No restarting procedure or penalty of $\|\mathbf{B}\|_F$ has been adopted in the proposed DSF method. The stopping conditions are the default values of min-

Func, i.e., the maximum number of iterations is 500, the largest element-wise change of the flatten composite gradient is less than 10^{-5}, or the functional change is less than 10^{-9}. The hyper-parameter of the proposed method, in general, depends on the shape of the mixing matrix and should be determined from the cross-validation process. Since KHL is computationally efficient and is a special case of the proposed method when $r = -\infty$, KHL is used to initialize the mixing matrix.

4.3 Simulation results

4.3.1 Simulation results for synthetic data

Simulations were conducted to evaluate the performance of the proposed DSF method. The proposed DSF method is compared with KHL [15], GMM [17], and SF [44]. For fair comparison, these methods share the same pre-whitening process, initial mixing matrix, and stopping conditions. For SF, the original source code was modified so that it can work with complex-valued data[2]. For GMM, it employs the Dirichlet prior,

$$p(\alpha_1, \ldots, \alpha_N) = \frac{\Gamma(N\phi)}{\Gamma(\phi)^N} \prod_{n=1}^{N} \alpha_n^{\phi-1}, \tag{4.51}$$

on the mixture weights $\alpha_1, \ldots, \alpha_N$ where N is the number of sources, Γ is the gamma function, and ϕ is a hyper-parameter. A small value of ϕ will reflect that some sources are more active than the others, whereas a large value of ϕ implies that all sources have similar activity, i.e., $\alpha_1, \ldots, \alpha_N$ are approximately equal. In [17], it is recommended that the value of ϕ should be large when the number of sources is known so that the mixture weights will be almost equal. Empirically, the variation in performance of GMM is modest when $\phi \geq 50$ for a

[2]The original implementation of sparse filtering can be found at https://github.com/jngiam/sparseFiltering.

wide range of dimensions M. Nevertheless, $\phi = 107$ is used since it yielded the best performance within the range of $\phi \in \{1, 2, ..., 150\}$ that was tested for the speech separation task described in Section 4.3.2. For DSF, we note that the hyper-parameter r is sensitive to source sparseness and the shape of the mixing matrix. Since the derivation of an optimal r is not within the scope of this work, for a selected value of input dimension M, cross validation was performed to select an optimal value of r.

The average mixing error ratio (MER) was employed as performance criterion. A higher value of MER implies higher performance. All simulations were performed on a Windows 7 64-bit computer with i7-4770 processor and 8GB RAM running a MATLAB 2018a with parallel toolbox.

The effects of sparseness, sample size, and the number of sources are studied in two scenarios where $M = 2$ and $M = 16$. The choice of $M = 2$ is derived from that of a typical smart device having two microphones, and that applying blind source separation on such devices have a lot of potential applications [129, 130]. The value of r used for DSF in this section are -0.25 and -16 respectively for M=2 and 16. In all simulations, the average performance of 100 trials are reported. For each trial, new sources are drawn from

$$f_S(s) = c \exp\left(-|s|^p\right), \ p > 0, \tag{4.52}$$

where c is a normalizing constant and p is the shape parameter. A new mixing matrix is generated in each trial such that the smallest angle between its columns is at least 10 degrees. The real and imaginary part of the mixing matrix follow standard normal distribution. The sample size and the shape parameter are set as default to $K = 1000$ and $p = 0.4$. This value of $p = 0.4$ is chosen as it matches well with realistic speech signals in the time-frequency domain [131]. For the case where $M = 2$, Figs. 4.2(a) and 4.2(d) depict the average MER and average

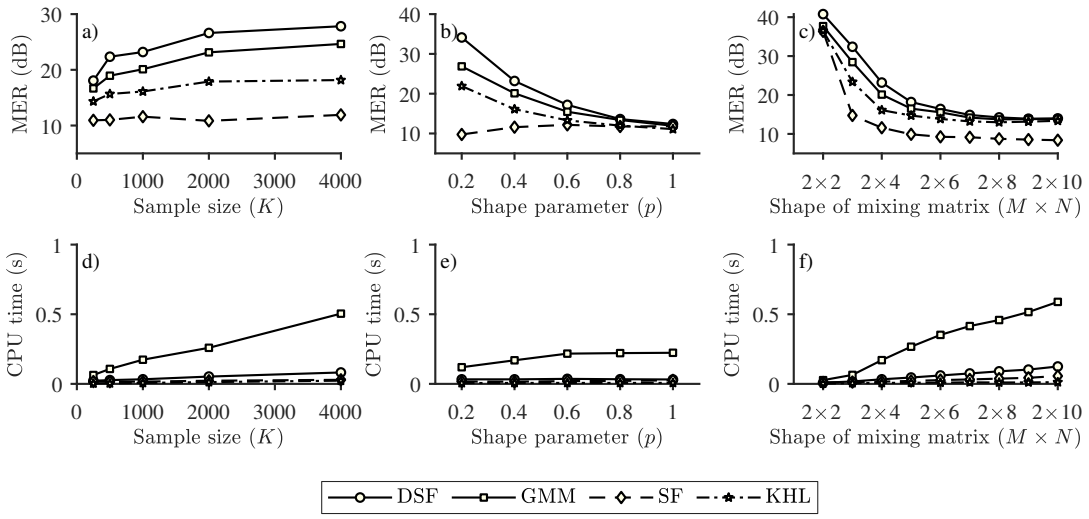


Fig. 4.2: Variation of the average MER with: a) sample size for the estimation of a 2×4 mixing matrix, b) sparseness for the estimation of a 2×4 mixing matrix, c) number of sources for $M = 2$. Variation of the average CPU time as a function of: d) sample size for the estimation of a 2×4 mixing matrix, e) sparseness for the estimation of a 2×4 mixing matrix, f) number of sources for $M = 2$.

CPU time when the sample size is increased from 250 to 4000. Figures 4.2(b) and 4.2(e) show the average MER and average CPU time for $0.2 \leq p \leq 1$, where the sparseness of the sources decreases with increasing p. Figures 4.2(c) and 4.2(f) illustrate the average MER and average CPU time when the dimension of the observed data is fixed and the number of sources is increased. Similar results for the case of $M = 16$ are reported in Fig. 4.3.

As observed from Figs. 4.2 and 4.3 ((a), (b) and (c)), SF fails completely in the under-determined mixing matrix estimation task because its sparse penalty is not suitable for this task. On the other hand, the proposed DSF method, GMM, and KHL are able to recover the mixing matrix when the sources possess sufficient degree of disjointness and the number of samples is sufficient. We note that, in general, the average MER will decrease with reducing degree of source disjointness, i.e., when either the source sparsity decreases or the number of sources increases. We also note that DSF is able to achieve higher MER than other baseline algorithms. For low degree of disjointness, while the performance

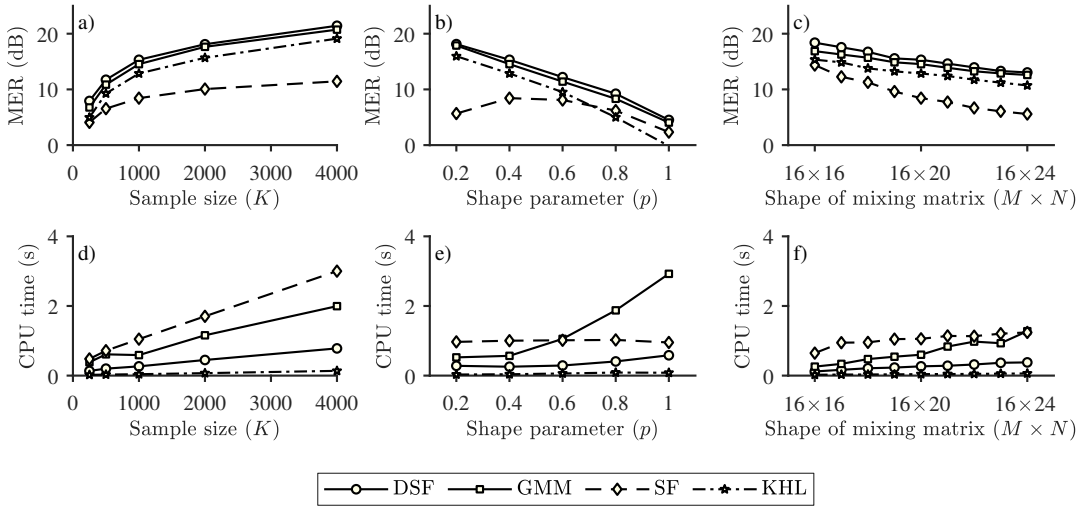


Fig. 4.3: Variation of the average MER with: a) sample size for the estimation of a 16×20 mixing matrix, b) sparseness for the estimation of a 16×20 mixing matrix, c) number of sources for $M = 16$. Variation of the average CPU time as a function of: d) sample size for the estimation of a 16×20 mixing matrix, e) sparseness for the estimation of a 16×20 mixing matrix, f) number of sources for $M = 16$.

of the proposed method and GMM are comparable, the former requires lower computational cost.

4.3.2 Simulation results for blind separation of under-determined convolutive speech mixtures

The performance of the proposed DSF algorithm is validated for the separation of under-determined convolutive speech mixtures. Similar to the time-frequency masking approach described in [17, 70], the convolutive mixtures were first converted from the time domain to instantaneous mixtures at each frequency bin in the STFT domain. This is achieved using a sufficiently long STFT window. For multi-channel speech mixtures sampled at 16 kHz, Hamming windows of length 2048 samples and 75% overlap were used. The mixing matrix at each frequency bin was then estimated and a mask for that bin is derived, which is subsequently

used to extract the sources. The separation mask

$$M_\beta(n) = \frac{\exp(\beta \cos^2 \theta_H(\mathbf{x}, \hat{\mathbf{a}}_n))}{\sum_l \exp(\beta \cos^2 \theta_H(\mathbf{x}, \hat{\mathbf{a}}_l))} \tag{4.53}$$

is computed from each pre-whitened input sample and estimated mixing matrix, where $\beta = 12.5$ specifies the softness of the mask. The above mask is a smooth version of the binary mask based on the softmax function. In fact, if β approaches ∞, the mask in (4.53) will be equivalent to the binary mask. Since natural speech signals exhibit spectral smoothness [132] and unlike the binary mask, the use of smooth mask will reduce any abrupt change in the spectral components of the separated speech signals. The permutation problem is then solved using a simplified version of [72] to align the mask across frequencies. The STFT of the sources are then extracted by applying mask on the mixtures and converting them back to the time domain. We note that, in addition to time-frequency masking, L_p-norm inverse filtering [37] can also be used for source extraction. However, the scaling ambiguity of the estimated mixing matrix is still an issue for L_p-norm inverse filtering. Such scaling ambiguity is more significant in the proposed DSF method, KHL, and GMM as they only estimate column directions of the mixing matrix.

For the evaluation of speech separation, SDR, SIR, and ESTOI are employed as performance metrics. Higher SDR, SIR, and ESTOI indicate better separation. All the algorithms are tested on 2-mixtures-and-4-sources subset of task T3 of dataset D1 in Signal Separation Evaluation Campaign (SiSEC) [115, 133, 134]. This dataset consists of live speech samples recorded in a real room environment. Table 4.1a shows the speech separation performance of the proposed DSF algorithm by varying r from -1 to -0.5. While the ESTOI scores remain the same for all r, it can be seen that $r = -0.9$ yields the highest performance in SDR and SIR. Hence, for the remaining simulations, we use $r = -0.9$.

Since the separation performance is affected by permutation alignment and masking procedure, for the evaluation of the mixing matrix estimation algorithm, baseline algorithms using KHL [13, 15, 41] and GMM [17] were created, where all the stages of the BSS algorithms are the same except for the mixing matrix estimation stage. These results are shown in Table 4.1b. For GMM, the performance for $\phi \in \{1, 2, ..., 150\}$ was evaluated and the best performance was achieved for $\phi = 107$. For reference, the performance of GMM for $\phi = 107$, 50 and 1.0 are shown in the table. Results in Table 4.1b show that DSF achieves higher average SDR, SIR, and ESTOI than KHL and GMM in six out of the eight scenarios considered. In addition, considering the fact that DSF requires lower computational complexity, the proposed DSF algorithm is a better alternative to KHL and GMM for real-world applications.

Mic. spacing	5 cm												1 m												Average
RT_{60}	130 ms						250 ms						130 ms						250 ms						performance over
Source type	females			males			females			males			females			males			females			males			all the conditions
Perf. metric	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI
DSF($r=-1.0$)	4.15	7.28	0.52	4.54	8.38	0.49	3.44	5.60	0.45	3.67	6.04	0.45	5.77	10.32	0.54	4.89	9.19	0.50	4.42	7.46	0.45	4.11	7.38	0.45	4.37	7.71	0.48
DSF($r=-0.9$)	4.27	7.44	0.50	4.46	8.24	0.48	3.50	5.61	0.46	3.74	6.21	0.45	6.08	10.87	0.55	4.90	9.19	0.50	4.40	7.37	0.45	4.19	7.55	0.45	4.44	7.81	0.48
DSF($r=-0.8$)	4.10	6.95	0.52	4.59	8.37	0.49	3.24	5.11	0.46	3.78	6.42	0.46	6.06	10.81	0.55	4.98	9.33	0.50	4.24	7.11	0.44	4.08	7.32	0.44	4.39	7.68	0.48
DSF($r=-0.7$)	3.75	6.12	0.52	4.45	8.08	0.49	3.54	5.71	0.47	3.86	6.56	0.46	6.07	10.76	0.55	5.03	9.44	0.50	4.18	7.03	0.44	4.31	7.80	0.45	4.40	7.69	0.48
DSF($r=-0.6$)	3.73	6.19	0.52	4.65	8.43	0.49	3.40	5.74	0.47	3.63	5.99	0.45	5.87	10.31	0.54	5.00	9.39	0.50	3.52	5.84	0.43	4.22	7.48	0.45	4.25	7.42	0.48
DSF($r=-0.5$)	3.69	6.15	0.52	4.49	8.07	0.49	3.45	5.58	0.47	3.65	6.01	0.44	6.01	10.53	0.55	4.95	9.19	0.50	4.47	7.41	0.45	3.54	5.97	0.42	4.28	7.36	0.48

(a) Speech separation performance of DSF at different value of r on 2mic_4src subset of SiSEC dev1 dataset.

Mic. spacing	5 cm												1 m												Average
RT_{60}	130 ms						250 ms						130 ms						250 ms						performance over
Source type	females			males			females			males			females			males			females			males			all the conditions
Perf. metric	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI
DSF($r=-0.9$)	4.27	7.44	0.50	4.46	8.24	0.48	3.50	5.61	0.46	3.74	6.21	0.45	6.08	10.87	0.55	4.90	9.19	0.50	4.40	7.37	0.45	4.19	7.55	0.45	4.44	7.81	0.48
GMM($\phi=107$)	3.31	5.49	0.49	4.14	7.41	0.48	3.53	5.93	0.46	3.18	4.77	0.43	5.95	10.60	0.52	4.71	8.58	0.49	4.41	7.44	0.46	4.18	7.47	0.45	4.18	7.21	0.47
GMM($\phi=50$)	2.93	4.69	0.48	4.07	7.26	0.47	2.85	4.37	0.45	2.63	3.60	0.41	5.88	10.39	0.52	4.39	8.08	0.49	4.40	7.42	0.45	4.12	7.35	0.45	3.91	6.65	0.46
GMM($\phi=1.0$)	3.19	4.78	0.45	2.42	3.21	0.37	3.22	4.97	0.42	1.95	1.97	0.37	5.65	10.00	0.51	4.08	7.26	0.46	5.01	8.70	0.47	3.56	6.28	0.41	3.64	5.90	0.43
KHL	3.67	6.27	0.47	3.43	5.94	0.43	2.45	3.38	0.43	2.35	3.05	0.38	5.61	10.25	0.52	2.53	3.66	0.39	2.94	4.73	0.33	3.31	5.81	0.41	3.29	5.39	0.42

(b) Speech separation performance of DSF compared with KHL and GMM on 2mic_4src subset of SiSEC dev1 dataset.

Mic. spacing	5 cm												1 m												Average
RT_{60}	130 ms						250 ms						130 ms						250 ms						performance over
Source type	females			males			females			males			females			males			females			males			all the conditions
Perf. metric	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI	SDR	SIR	ESTOI
DSF($r=-0.9$)	4.27	7.44	0.50	4.46	8.24	0.48	3.50	5.61	0.46	3.74	6.21	0.45	6.08	10.87	0.55	4.90	9.19	0.50	4.40	7.37	0.45	4.19	7.55	0.45	4.44	7.81	0.48
[70]	4.38	6.48	0.51	4.47	6.38	0.49	3.84	4.85	0.48	3.81	5.05	0.46	4.88	7.18	0.52	4.70	7.11	0.49	3.81	5.26	0.46	3.82	5.44	0.44	4.22	5.97	0.48
[135]	4.42	10.46	0.56	3.64	8.34	0.48	3.94	7.75	0.48	2.88	5.37	0.41	3.98	6.79	0.51	3.79	7.14	0.48	4.80	8.15	0.50	4.64	8.32	0.45	4.01	7.79	0.48
ACGMM [92]	3.95	5.82	0.46	3.50	4.96	0.43	2.91	3.43	0.42	2.57	2.31	0.38	4.79	7.31	0.46	3.64	5.30	0.43	3.61	4.94	0.40	2.83	2.99	0.38	3.47	4.63	0.42
Input	-4.76	-4.68	0.29	-4.81	-4.60	0.27	-4.83	-4.71	0.29	-4.79	-4.64	0.27	-4.84	-4.59	0.28	-4.90	-4.59	0.27	-4.85	-4.61	0.28	-4.88	-4.60	0.26	-4.83	-4.63	0.28

(c) Speech separation performance of the proposed method versus other heterogeneous algorithms on 2mic_4src subset of SiSEC dev1 dataset.

Table 4.1: Speech separation performance of DSF on 2mic_4src subset of SiSEC dev1 dataset.

Table 4.1c shows the comparison between the proposed DSF method with other heterogeneous BSS algorithms namely the angular central Gaussian mixture model (ACGMM) [92], interleaved ICA [135], and Bayes risk minimization [70]. As opposed to KHL and GMM, ACGMM can model rotationally asymmetric cluster, and therefore it can potentially fit the observed data better than KHL and GMM. We note that ACGMM requires the same computational complexity as that of GMM. In our implementation of [92], we employed KHL to cluster the observed data into disjoint groups before utilizing the intra-covariance matrices of these groups to initialize ACGMM. As can be seen from the results, the proposed DSF method achieves the highest average SDR of 4.44 dB among the algorithms. This method also improves the average SDR and SIR by 9.27 dB and 12.44 dB over the input, while the intelligibility score (quantified by ESTOI) is improved by 0.20. In addition, the proposed DSF algorithm outperforms that described in [70] by 0.22 dB in SDR and 1.84 dB in SIR. Compared to [135], the proposed method achieves an average of 0.43 dB higher in SDR while having the same ESTOI and modestly higher SIR.

4.4 Summary

In this chapter, a new under-determined mixing matrix estimation method is proposed by combining the benefits of directional clustering approach and sparse filtering. The contrast function of the proposed method is based on the power mean of magnitude-squared cosine distance which has the ability to better model the directional data than existing approaches. A simple but efficient substitution is also proposed to optimize the proposed cost function as well as other semi-unitary constrained optimization problems. Simulation results on both synthetic data and recorded speech mixtures highlight the improvement in performance with reduced computational complexity for the proposed directional

sparse filtering method.

Chapter 5

Stability Condition for Directional Methods

5.1 Introduction

Directional methods differ from ICA in three aspects: the mixed data is L_2-normalized, the multi-variate contrast function may be not a summation of univariate contrast functions, and the sources are assumed to be approximately disjoint, i.e., highly sparse. With sufficient degree of source sparseness, most of the observed data will lie along directions defined by columns of the mixing matrix. Despite being the core assumption of directional SCA methods, the degree of source sparseness has not been quantified for either determined scenario nor under-determined scenario.

In this chapter, it will be shown that for the L_2-normalized data to be directional, the assumption of the source sparseness can be relaxed to super-Gaussianity in the determined case. We then extend the stability analysis of complex-valued ICA and find a stability condition for methods based on the magnitude-squared cosine similarity. Similar to ICA, analysis and discussions presented in this chapter highlight that the directional methods can be inter-

preted as quasi-maximum likelihood estimation under certain conditions. While such connection has previously been drawn for real-valued clustering in [45], the recovery condition given in [45] is specific for Laplacian-distributed sources. It also depends on a conjecture which is confirmed only for a two-dimensional real vector space. The recovery condition is not only hard to verify in practice [46] but also does not take L_2-normalization into account. The derived stability condition does not suffer from such limitations.

5.2 The L_2-normalization of i.i.d. generalized Gaussian mixtures

Similar to [6, 9, 37, 70, 131, 136], the probability density function (p.d.f.) of each source is assumed to be

$$f_S(s_n) = \frac{p}{2\pi\sigma^2\Gamma(2/p)} \exp\left(-\left|s_n/\sigma\right|^p\right), \ p > 0, \tag{5.1}$$

where $\Gamma(\cdot)$ is the gamma function, $\sigma^2 = \Gamma(2/p)/\Gamma(4/p)$ is the scaling constant, and p is the shape parameter of the distribution [136, Eq. (6)]. The p.d.f. in (5.1) is the circular complex generalized Gaussian distribution (CGGD) with zero-mean and unit variance. This p.d.f. resembles a subset of circular complex sub-Gaussian distributions for $p > 2$, a subset of circular complex super-Gaussian distributions for $0 < p < 2$, circular complex Gaussian distribution for $p = 2$, and circular complex Laplacian distribution for $p = 1$. Furthermore, a source is regarded as sparse if $0 < p \leq 1$.

Let $\bar{\mathbf{x}} = \mathbf{x}/\left\|\mathbf{x}\right\|_2 = [\bar{x}_1, \ldots, \bar{x}_N]^T$ be the L_2-normalized (pre-whitened) input data. The mapping from $\mathbf{x}$ to the pair $\bar{\mathbf{x}}$ and $\left\|\mathbf{x}\right\|_2$ is a one-to-one mapping

having the Jacobian [137, pp. 221-222]

$$|\det \mathbb{J}| = \|\mathbf{x}\|_2^{2N-1} \prod_{n=1}^{N-1} |\bar{x}_n| \,. \tag{5.2}$$

Assuming that the sources are independently and identically distributed (i.i.d.) according to (5.1), the joint density of $\mathbf{s}$ is given by

$$f_S(\mathbf{s}) = \prod_n f_S(s_n) = \frac{p^N \exp\left(-\|\mathbf{s}/\sigma\|_p^p\right)}{[2\pi\sigma^2\Gamma(2/p)]^N}, \; p > 0. \tag{5.3}$$

Since $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A} = \mathbf{I}$ due to whitening, we have $f_X(\mathbf{x}) = \frac{1}{|\det \mathbf{A}|^2} f_S(\mathbf{A}^H\mathbf{x}) = f_S(\mathbf{A}^H\mathbf{x})$. Using $f_X(\mathbf{x})$, (5.2), the change of variables concept, and exploiting the homogeneity property of L_p-norm, the joint p.d.f. of $\bar{\mathbf{x}}$ and $\|\mathbf{x}\|_2$ is given by

$$\begin{aligned} f_{\bar{X},\|X\|_2}(\bar{\mathbf{x}}, \|\mathbf{x}\|_2) =& \frac{p^N}{[2\pi\sigma^2\Gamma(2/p)]^N} \exp\left(-\|\mathbf{x}\|_2^p \left\|\mathbf{A}^H\bar{\mathbf{x}}/\sigma\right\|_p^p\right) \\ & \times \|\mathbf{x}\|_2^{2N-1} \prod_{n=1}^{N-1} |\bar{x}_n| \,. \end{aligned} \tag{5.4}$$

Integrating $f_{\bar{X},\|X\|_2}(\bar{\mathbf{x}}, \|\mathbf{x}\|_2)$ over the domain of $\|\mathbf{x}\|_2$ and using $\int_0^\infty t^b e^{-at} dt = \frac{\Gamma(b+1)}{a^{b+1}}$, the marginal distribution

$$f_{\bar{X}}(\bar{\mathbf{x}}) = \int_0^\infty f(\bar{\mathbf{x}}, \|\mathbf{x}\|_2) d\,\|\mathbf{x}\|_2 = \frac{p^{N-1}\Gamma(2N/p)}{[2\pi\sigma^4\Gamma(2/p)]^N} \left\|\mathbf{A}^H\bar{\mathbf{x}}\right\|_p^{-2N} \prod_{n=1}^{N-1} |\bar{x}_n| \tag{5.5}$$

is obtained.

Since $\|\bar{\mathbf{x}}\|_2 = 1$ and $\mathbf{A}\mathbf{A}^H = \mathbf{I}$, we have

$$\left\|\mathbf{A}^H\bar{\mathbf{x}}\right\|_2^2 = \sum_{n=1}^{N} \left|\mathbf{a}_n^H\bar{\mathbf{x}}\right|^2 = 1, \; \forall\bar{\mathbf{x}}. \tag{5.6}$$

Consequently, the distribution of L_2-normalized mixtures of Gaussian sources is

$$f_{\bar{X}}(\bar{\mathbf{x}}; p = 2) = \frac{\Gamma(N)}{2(\pi)^N} \prod_{n=1}^{N-1} |\bar{x}_n| \,. \tag{5.7}$$

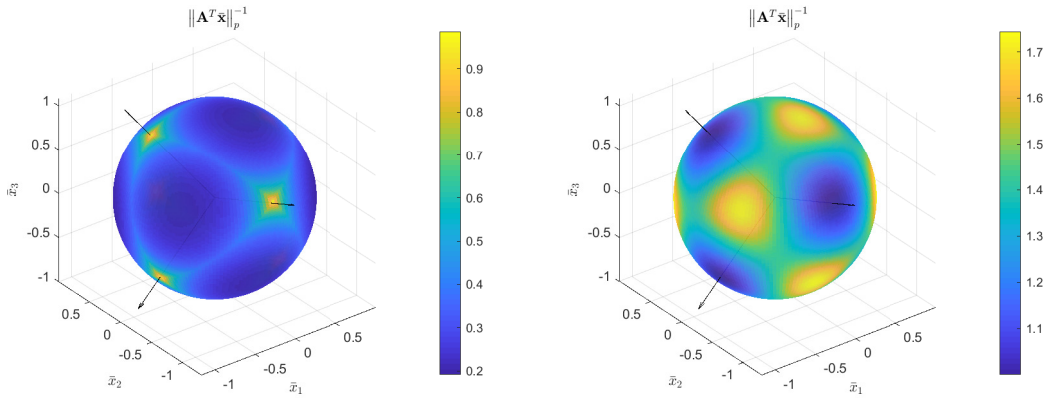


Fig. 5.1: Directional term of the marginal distribution of L_2-normalized mixtures. The black arrows denote the columns of the mixing matrix. Left: super-Gaussian CGGD mixtures ($p = 1$). Right: sub-Gaussian CGGD mixtures ($p = 3$).

This is the uniform distribution on the unit sphere of $\mathbb{C}^N$ defined by $S_2^N = \left\{\mathbf{z} \in \mathbb{C}^N : \|\mathbf{z}\|_2 = 1\right\}$ [137–139]. Similar to ICA, one cannot estimate the mixing matrix in this case since $f_{\bar{X}}(\bar{\mathbf{x}}; p = 2)$ is independent of $\mathbf{A}$. On the other hand, when the sources are non-Gaussian, the distribution on the sphere includes a "directional" term $\left\|\mathbf{A}^H\bar{\mathbf{x}}\right\|_p^{-2N}$ making it possible to estimate the mixing matrix. Interestingly, the directional term attains its maximum (or conversely minimum) when $\bar{\mathbf{x}}$ has the same direction with a column of the mixing matrix for $0 < p < 2$ (or conversely for $p > 2$). This implies that the L_2-normalized mixed data of circular complex generalized Gaussian sources will concentrate most frequently (or conversely least frequently) around the directions of columns of the mixing matrix for super-Gaussian sources (or conversely for sub-Gaussian). To verify this observation, we note that $0 \leq \left|\mathbf{a}_n^H\bar{\mathbf{x}}\right|^2 \leq 1$ due to (5.6). Therefore, for two arbitrary real numbers a and b such that $0 < a < 2 < b$, we have $\left|\mathbf{a}_n^H\bar{\mathbf{x}}\right|^b \leq \left|\mathbf{a}_n^H\bar{\mathbf{x}}\right|^2 \leq \left|\mathbf{a}_n^H\bar{\mathbf{x}}\right|^a$. Consequently,

$$\left\|\mathbf{A}^H\bar{\mathbf{x}}\right\|_b^{-2N} \geq 1 \geq \left\|\mathbf{A}^H\bar{\mathbf{x}}\right\|_a^{-2N}. \tag{5.8}$$

The equalities of (5.8) occur when there is only one index l for which $\left|\mathbf{a}_l^H\bar{\mathbf{x}}\right|^2 = 1$

and $\left|\mathbf{a}_n^H \bar{\mathbf{x}}\right|^2 = 0$ when $n \neq l$. Note that all columns of $\mathbf{A}$ have unit norm for the determined case, and therefore $\left|\mathbf{a}_l^H \bar{\mathbf{x}}\right|^2$ is the magnitude-squared cosine-similarity between $\mathbf{a}_l$ and $\bar{\mathbf{x}}$. Figure 5.1 illustrates the directional term of (5.5) in 3-dimensional real vector space where $\mathbf{A}$ is given by $\begin{bmatrix} -0.830 & -0.008 & -0.557 \\ -0.198 & -0.930 & 0.309 \\ -0.520 & 0.368 & 0.770 \end{bmatrix}$ for super-Gaussian case and sub-Gaussian case, respectively. It can be seen that the directional term attains its maximum (or conversely minimum) when $\bar{\mathbf{x}}$ has the same direction with a column of the mixing matrix for $0 < p < 2$ (or conversely for $p > 2$).

5.3 The L_2-normalization of L_p radially symmetric mixtures

Similar directional effect to the above sources can also be observed when the joint distribution of the sources is L_p radially symmetric. A source vector $\mathbf{s}$ is L_p radially symmetric if its density $f(\mathbf{s})$ can be rewritten in the form $f(\mathbf{s}) = h(\|\mathbf{s}\|_p)$ [140], i.e., the joint distribution depends only on the L_p norm of the source vector. The L_p radially symmetric sources are of interest because they are often not independent.

One notable property of L_p radially symmetric vector is that, for any scalar variable ξ which is uniform in $[0, 1]$, the vector $\mathbf{y} = \xi^{1/(2N)}\mathbf{s}/\|\mathbf{s}\|_p$ is uniformly distributed in the unit L_p-ball defined by $\mathcal{B}_p^N = \left\{\mathbf{z} \in \mathbb{C}^N : \|\mathbf{z}\|_p \leq 1\right\}$. Clearly, the projection of $\mathbf{y}$ and $\mathbf{s}$ onto the unit sphere $\mathcal{S}_2^N = \left\{\mathbf{z} \in \mathbb{C}^N : \|\mathbf{z}\|_2 = 1\right\}$ are the same.

We now consider the case where $0 < p < 2$. According to the rejection sampling [141] and the inclusion of $\mathcal{B}_p^N$ inside $\mathcal{B}_2^N$ when $p < 2$, we can generate $\mathcal{B}_p^N$ uniformly distributed samples from a $\mathcal{B}_2^N$ uniformly distributed samples by

removing samples lying inside the volume difference between $\mathcal{B}_2^N$ and $\mathcal{B}_p^N$. Intuitively, this implies that the density of $\mathbf{y}/\left\|\mathbf{y}\right\|_2$ (and hence $\mathbf{s}/\left\|\mathbf{s}\right\|_2$) at a given direction is inversely proportional to the gap between two balls $\mathcal{B}_2^N$ and $\mathcal{B}_p^N$ in the same direction. Consequently, the density of $\mathbf{s}/\left\|\mathbf{s}\right\|_2$ peaks at directions specified by the standard basis since $\mathcal{B}_2^N$ and $\mathcal{B}_p^N$ have zero gap in these directions. Since the unit ball $\mathcal{B}_2^N$ is invariant to rotation whereas $\mathcal{B}_p^N$ is only invariant to the permutation of basis vectors, the L_2-normalization of $\mathbf{As}$ peaks at directions specified by the columns of $\mathbf{A}$ given that $\mathbf{A}$ is unitary. A similar argument can be constructed for $p > 2$ and L_2-nested radially symmetric vector studied in [139]. It is worth noting that for an L_p radially symmetric or L_p-nested radially symmetric $\mathbf{s}$, the individual sources are not necessarily independent unless their joint p.d.f. is in the form of (5.3) [142]. Particularly, for the L_p-nested radially symmetric distribution, the sources can have different shape parameter. However, in the general case where each source having a different distribution, a geometrical interpretation of L_2-normalization is challenged to find.

5.4 Stability condition

As can be seen, the L_2-normalization removes the ability to distinguish super-Gaussian source and sub-Gaussian source. Therefore, all the sources are assumed to be either super-Gaussian or sub-Gaussian. For the source distribution given in (5.3), one can estimate the unitary mixing matrix from normalized non-Gaussian mixtures via maximum likelihood estimation (MLE). From (5.5), this can be achieved by minimizing the negative log-likelihood

$$\min_{\widehat{\mathbf{A}}} \ \mathcal{E}\Big\{ \ln \sum_{n=1}^{N} \left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}\right|^p \Big\}, \ \text{s.t.} \ \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}. \tag{5.9}$$

Although one may know whether $p < 2$ or $p > 2$, the exact value of p may not be available in blind source separation, and therefore one cannot directly

solve the above optimization problem. Instead, the negative log-likelihood can be replaced by a surrogate function and solving for the quasi-MLE problem.

The stability condition for surrogates of the negative log-likelihood is now derived. Due to the unitary constraint, matrix $\widehat{\mathbf{A}}$ will have unit-norm columns. Therefore, with reference to the definition of the magnitude-squared cosine-similarity in (4.3), it can be seen that $u_n = \left|\widehat{\mathbf{a}}_n^H \mathbf{x}\right|^2 / \left(\|\widehat{\mathbf{a}}_n\|_2^2 \, \|\mathbf{x}\|_2^2\right) = \left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}\right|^2$. Consequently, the contrast function corresponding to the maximum likelihood estimator given in (5.9) is $g^{(\mathrm{MLE})}(\mathbf{u}) = \ln \sum_{n=1}^{N} u_n^{p/2}$. This function is strictly Schur-concave when $0 < p < 2$ and is strictly Schur-convex when $p > 2$, according to composition rule [56, pp. 138-139]. In addition, due to (5.6), the unitary constraint becomes the standard simplex $\Delta_{N-1} = \left\{(u_1, \ldots, u_N)^T \middle| u_n \in [0,1], \sum_{n=1}^{N} u_n = 1\right\}$. In essence, the maximum likelihood estimator determines a solution that minimizes the expectation of a strictly Schur-concave or strictly Schur-convex function over a simplex. It is interesting to note that the minima of a Schur-concave function or a Schur-convex function over a simplex are independent of the choice of the function given that Schur-concavity or Schur-convexity is preserved [47, 56]. Nevertheless, due to the expectation operator, it is unclear whether one can replace $\ln \sum_{n=1}^{N} u_n^{p/2}$ by another strictly Schur-convex function when the sources are super-Gaussian or by another strictly Schur-convex function when the sources are sub-Gaussian.

To verify which functions can be used, the stability analysis of complex-valued ICA [7, 10] is extended to account for the L_2-normalization of input and the usage of non-separable contrast functions. The idea of stability analysis is that the (local) minimizers of a cost function $J(\widehat{\mathbf{A}})$ include the mixing matrix $\mathbf{A}$ if $J(\widehat{\mathbf{A}})$ always increases when a small perturbation is made around $\mathbf{A}$. The following result describes the stability condition for estimating the mixing matrix from L_2-normalized data with a (possibly non-separable) multivariate contrast function.

Lemma 5.1. *Let* $g(\mathbf{u})$*:* $\mathbb{R}^N \to \mathbb{R}$*,* $g^{(l)}(\mathbf{u}) = \frac{\partial g(\mathbf{u})}{\partial u_l}$*, and* $g^{(ln)}(\mathbf{u}) = \frac{\partial^2 g(\mathbf{u})}{\partial u_l \partial u_n}$ *be a multivariate contrast function, its first-order partial derivative, and its second-order partial derivative, respectively. Assume that* $\mathbf{x} = \mathbf{A}\mathbf{s}$*, where* $\mathbf{A}$ *is unitary and* $\mathbf{s}$ *is circular, i.e., the joint p.d.f. of the sources does not depend on phases* $\angle s_1, \ldots, \angle s_N$ *of the sources and the joint p.d.f. satisfies* $f_S(\mathbf{s}) = f_S(\mathbf{s}_{\text{abs}})$ *with* $\mathbf{s}_{\text{abs}} = [|s_1|, \ldots, |s_N|]^T$*. For* $\bar{\mathbf{x}} = \mathbf{x}/\left\|\mathbf{x}\right\|_2$*,* $\bar{\mathbf{s}} = \mathbf{s}/\left\|\mathbf{s}\right\|_2 = [\bar{s}_1, \ldots, \bar{s}_N]^T$*, and* $\mathbf{u}_0 = [|\bar{s}_1|^2, \ldots, |\bar{s}_N|^2]^T$*, the local minimizers of the following optimization problem*

$$\min_{\widehat{\mathbf{A}}} \; J(\widehat{\mathbf{A}}), \text{ s.t. } \widehat{\mathbf{A}}\widehat{\mathbf{A}}^H = \mathbf{I}, \tag{5.10}$$

where $J(\widehat{\mathbf{A}}) = \mathcal{E}\left\{g\left(\left|\hat{\mathbf{a}}_1^H\bar{\mathbf{x}}\right|^2, \ldots, \left|\hat{\mathbf{a}}_N^H\bar{\mathbf{x}}\right|^2\right)\right\}$*, include a set of unitary directions specified by columns of the mixing matrix (i.e.,* $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N$*) if, for all* $l \neq n$*, the sources and the contrast function satisfy*

$$\begin{aligned}&\mathcal{E}\Big\{-\left(|\bar{s}_l|^2 - |\bar{s}_n|^2\right)\left(g^{(l)}(\mathbf{u}_0) - g^{(n)}(\mathbf{u}_0)\right)\\ &+|\bar{s}_l|^2|\bar{s}_n|^2\left(g^{(ll)}(\mathbf{u}_0) + g^{(nn)}(\mathbf{u}_0) - 2g^{(ln)}(\mathbf{u}_0)\right)\Big\} > 0.\end{aligned} \tag{5.11}$$

Proof. Let $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$ where $\mathbf{z}_n = \mathbf{A}^H\hat{\mathbf{a}}_n$ for $n = 1, \ldots, N$. For $\bar{\mathbf{x}} = \mathbf{x}/\left\|\mathbf{x}\right\|_2$, $\bar{\mathbf{s}} = \mathbf{s}/\left\|\mathbf{s}\right\|_2$, we have $\bar{\mathbf{x}} = \mathbf{A}\bar{\mathbf{s}}$ since $\mathbf{A}$ is unitary. Therefore, $\mathbf{u}(\widehat{\mathbf{A}}; \mathbf{x}) = \left[\left|\hat{\mathbf{a}}_1^H\bar{\mathbf{x}}\right|^2, \ldots, \left|\hat{\mathbf{a}}_N^H\bar{\mathbf{x}}\right|\right] = \left[\left|\hat{\mathbf{a}}_1^H\mathbf{A}\bar{\mathbf{s}}\right|^2, \ldots, \left|\hat{\mathbf{a}}_N^H\mathbf{A}\bar{\mathbf{s}}\right|^2\right] = \mathbf{u}(\mathbf{Z}; \bar{\mathbf{s}}) = [u_1, \ldots, u_N]^T$ where $u_n = \left|\mathbf{z}_n^H\bar{\mathbf{s}}\right|^2 = \mathbf{z}_n^H\bar{\mathbf{s}}\bar{\mathbf{s}}^H\mathbf{z}_n$ for $n = 1, \ldots, N$. The cost function in (5.10) becomes

$$J(\mathbf{Z}) = \mathcal{E}\Big\{g\left(\mathbf{u}\right)\Big\}. \tag{5.12}$$

Let $\mathbf{Z}_0 = [\delta_1\mathbf{e}_1, \ldots, \delta_N\mathbf{e}_N]$, where $\delta_n \in \{\delta \in \mathbb{C} | \delta\delta^* = 1\}$ and $\mathbf{e}_n$ is the nth standard unit vector for $n = 1, \ldots, N$. Note that, $\mathbf{u}$ evaluated at $\mathbf{Z}_0$ is $\mathbf{u}_0 = [|\bar{s}_1|^2, \ldots, |\bar{s}_N|^2]^T$. By definition of $\mathbf{Z}$, $\mathbf{A}$ is a local minimizer of $J(\widehat{\mathbf{A}})$ if $\mathbf{Z}_0$ is a minimizer of $J(\mathbf{Z})$. Denoting $\Delta\mathbf{Z} = [\Delta\mathbf{z}_1, \ldots, \Delta\mathbf{z}_N]$ as a small perturbation around $\mathbf{Z}_0$, $\mathbf{Z}_0$ is a minimizer of $J(\mathbf{Z})$ if $\Delta J(\mathbf{Z}_0) = J(\mathbf{Z}_0 + \Delta\mathbf{Z}) - J(\mathbf{Z}_0) > 0, \forall\Delta\mathbf{Z}$.

The Taylor expansion of $J(\mathbf{Z})$ at $\mathbf{Z}_0$ is given by

$$\begin{aligned}\Delta J(\mathbf{Z}_0) = &\sum_{n=1}^{N}\left(\frac{\partial J(\mathbf{Z}_0)}{\partial \mathbf{z}_n^T}\Delta\mathbf{z}_n + \frac{\partial J(\mathbf{Z}_0)}{\partial \mathbf{z}_n^H}\Delta\mathbf{z}_n^*\right)\\ &+ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\Delta\mathbf{z}_n^T\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n\partial \mathbf{z}_m^T}\Delta\mathbf{z}_m + \Delta\mathbf{z}_n^H\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n^*\partial \mathbf{z}_m^H}\Delta\mathbf{z}_m^*\right)\\ &+ \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\left(\Delta\mathbf{z}_n^H\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n^*\partial \mathbf{z}_m^T}\Delta\mathbf{z}_m + \Delta\mathbf{z}_n^T\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n\partial \mathbf{z}_m^H}\Delta\mathbf{z}_m^*\right)\\ &+ o(\|\Delta\mathbf{Z}\|_F^2). \end{aligned} \tag{5.13}$$

Using Wirtinger calculus and noting that $J(\mathbf{Z})$ is real,

$$\frac{\partial J(\mathbf{Z})}{\partial \mathbf{z}_n} = \mathcal{E}\left\{\bar{\mathbf{s}}^*\mathbf{z}_n^H\bar{\mathbf{s}}g^{(n)}(\mathbf{u})\right\}, \tag{5.14}$$

$$\frac{\partial J(\mathbf{Z})}{\partial \mathbf{z}_n^*} = \left(\frac{\partial J(\mathbf{Z})}{\partial \mathbf{z}_n}\right)^*. \tag{5.15}$$

Similarly, the second-order partial derivatives are specified by

$$\frac{\partial^2 J(\mathbf{Z})}{\partial \mathbf{z}_n\partial \mathbf{z}_m^T} = \mathcal{E}\left\{\bar{\mathbf{s}}^*\bar{\mathbf{s}}^H\mathbf{z}_n^H\bar{\mathbf{s}}\mathbf{z}_m^H\bar{\mathbf{s}}g^{(nm)}(\mathbf{u})\right\}, \tag{5.16}$$

$$\frac{\partial^2 J(\mathbf{Z})}{\partial \mathbf{z}_n^*\partial \mathbf{z}_m^T} = \mathcal{E}\left\{\bar{\mathbf{s}}\bar{\mathbf{s}}^H\bar{\mathbf{s}}^H\mathbf{z}_n\mathbf{z}_m^H\bar{\mathbf{s}}g^{(nm)}(\mathbf{u}) + \frac{\partial \mathbf{z}_m^H}{\partial \mathbf{z}_n^*}\bar{\mathbf{s}}\bar{\mathbf{s}}^H g^{(m)}(\mathbf{u})\right\}, \tag{5.17}$$

$$\frac{\partial^2 J(\mathbf{Z})}{\partial \mathbf{z}_n^*\partial \mathbf{z}_m^H} = \left(\frac{\partial^2 J(\mathbf{Z})}{\partial \mathbf{z}_n\partial \mathbf{z}_m^T}\right)^*, \tag{5.18}$$

$$\frac{\partial^2 J(\mathbf{Z})}{\partial \mathbf{z}_n\partial \mathbf{z}_m^H} = \left(\frac{\partial^2 J(\mathbf{Z})}{\partial \mathbf{z}_n^*\partial \mathbf{z}_m^T}\right)^*. \tag{5.19}$$

Here, one has (5.18) and (5.19) because the Hessian matrix of complex-augmented vector is Hermitian symmetric under mild conditions. Since L_2-normalization does not affect the phase information, the phase of each component of $\bar{\mathbf{s}}$ is $\angle s_1, \ldots, \angle s_N$. Due to the circularity assumption, $\angle s_1, \ldots, \angle s_N$ are mutually independent and uniformly distributed in $[0, 2\pi]$ as well as being independent to $|s_1|, \ldots, |s_N|$. Consequently, any scalar term in (5.14)-(5.19) including $e^{\jmath\angle s_n}$ for

any n, is zero. Therefore, $\partial J(\mathbf{Z})/\partial \mathbf{z}_n$ evaluated at $\mathbf{Z}_0$ is

$$\frac{\partial J(\mathbf{Z}_0)}{\partial \mathbf{z}_n} = \delta_n^* \mathbf{e}_n \mathcal{E}\Big\{ |\bar{s}_n|^2 \, g^{(n)}(\mathbf{u}_0) \Big\}. \tag{5.20}$$

Similarly, for the second-order derivatives in (5.16) and (5.17), with $n \neq m$, we have

$$\begin{aligned}
\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n \partial \mathbf{z}_m^T} &= \mathcal{E}\Big\{ \bar{\mathbf{s}}^* \bar{\mathbf{s}}^H \delta_n^* \delta_m^* \bar{s}_n \bar{s}_m g^{(nm)}(\mathbf{u}_0) \Big\} \\
&= \delta_n^* \delta_m^* (\mathbf{e}_n \mathbf{e}_m^T + \mathbf{e}_m \mathbf{e}_n^T) \mathcal{E}\Big\{ |\bar{s}_n|^2 \, |\bar{s}_m|^2 \, g^{(nm)}(\mathbf{u}_0) \Big\}, \qquad (5.21) \\
\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n^* \partial \mathbf{z}_m^T} &= \mathcal{E}\Big\{ \bar{\mathbf{s}} \bar{\mathbf{s}}^H \delta_n \delta_m^* \bar{s}_n^* \bar{s}_m g^{(nm)}(\mathbf{u}_0) \Big\} \\
&= \delta_n \delta_m^* \mathbf{e}_n \mathbf{e}_m^T \mathcal{E}\Big\{ |\bar{s}_n|^2 \, |\bar{s}_m|^2 \, g^{(nm)}(\mathbf{u}_0) \Big\}. \qquad (5.22)
\end{aligned}$$

In the evaluation of (5.16) given in (5.21), we note that there are only two element of $\bar{\mathbf{s}}^* \bar{\mathbf{s}}^H \bar{s}_n \bar{s}_m$, that do not include the phases, result in non-zero expectation. Similarly, in the evaluation of (5.17) given in (5.22), it is noted that the second term of (5.17) becomes zeros for $m \neq n$ and that there is only one element of $\bar{\mathbf{s}} \bar{\mathbf{s}}^H \bar{s}_n^* \bar{s}_m$, that does not include the phase, results in non-zero expectation. On the other hand, when $n \equiv m$, the following derivatives are obtained

$$\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n \partial \mathbf{z}_n^T} = (\delta_n^*)^2 \, \mathbf{e}_n \mathbf{e}_n^T \mathcal{E}\Big\{ |\bar{s}_n|^4 \, g^{(nn)}(\mathbf{u}_0) \Big\}, \tag{5.23}$$

$$\frac{\partial^2 J(\mathbf{Z}_0)}{\partial \mathbf{z}_n^* \partial \mathbf{z}_n^T} = \sum_{l=1}^{N} \mathbf{e}_l \mathbf{e}_l^T \mathcal{E}\Big\{ |\bar{s}_n|^2 \, |\bar{s}_l|^2 \, g^{(nn)}(\mathbf{u}_0) + |\bar{s}_l|^2 \, g^{(n)}(\mathbf{u}_0) \Big\}. \tag{5.24}$$

From (5.20)-(5.24), with reference to (5.15), (5.18), and (5.19),

$$\begin{aligned}\Delta J(\mathbf{Z}_0) = &\sum_{n=1}^{N} 2\Re\left\{\delta_{\mathrm{n}}^{*}\Delta \mathrm{z}_{\mathrm{nn}}\right\}\mathcal{E}\left\{\left|\bar{s}_n\right|^2 g^{(n)}(\mathbf{u}_0)\right\} \\ &+ \sum_{n=1}^{N}\sum_{m\neq n} \Re\left\{\delta_n^{*}\delta_m^{*}\left(\Delta z_{nn}\Delta z_{mm} + \Delta z_{mn}\Delta z_{nm}\right)\right\}\mathcal{E}\left\{\left|\bar{s}_n\right|^2\left|\bar{s}_m\right|^2 g^{(nm)}(\mathbf{u}_0)\right\} \\ &+ \sum_{n=1}^{N}\sum_{m\neq n} \Re\left\{\delta_n\delta_m^{*}\Delta z_{nn}^{*}\Delta z_{mm}\right\}\mathcal{E}\left\{\left|\bar{s}_n\right|^2\left|\bar{s}_m\right|^2 g^{(nm)}(\mathbf{u}_0)\right\} \\ &+ \sum_{n=1}^{N} \Re\left\{\left(\delta_n^{*}\Delta z_{nn}\right)^2\right\}\mathcal{E}\left\{\left|\bar{s}_n\right|^4 g^{(nn)}(\mathbf{u}_0)\right\} \\ &+ \sum_{n=1}^{N}\sum_{l=1}^{N}\left|\Delta z_{ln}\right|^2\mathcal{E}\left\{\left|\bar{s}_n\right|^2\left|\bar{s}_l\right|^2 g^{(nn)}(\mathbf{u}_0)\right\} \\ &+ \sum_{n=1}^{N}\sum_{l=1}^{N}\left|\Delta z_{ln}\right|^2\mathcal{E}\left\{\left|\bar{s}_l\right|^2 g^{(n)}(\mathbf{u}_0)\right\} \\ &+ o\left(\left\|\Delta\mathbf{Z}\right\|_F^2\right). \end{aligned} \tag{5.25}$$

Since $(\mathbf{Z}_0 + \Delta\mathbf{Z})^H(\mathbf{Z}_0 + \Delta\mathbf{Z}) = \mathbf{I}$ due to the unitary constraint, we have

$$\Delta\mathbf{Z}^H\mathbf{Z}_0 + \mathbf{Z}_0^H\Delta\mathbf{Z} = -\Delta\mathbf{Z}^H\Delta\mathbf{Z}. \tag{5.26}$$

Evaluating both sides of (5.26) at the mth row and nth column, we obtain

$$\delta_n\Delta z_{nm}^{*} + \delta_m^{*}\Delta z_{mn} = -\sum_k \Delta z_{km}^{*}\Delta z_{kn}. \tag{5.27}$$

Setting $m = n$ in (5.27),

$$2\Re\left\{\delta_n^{*}\Delta z_{nn}\right\} = \delta_n\Delta z_{nn}^{*} + \delta_n^{*}\Delta z_{nn} = -\left\|\Delta\mathbf{z}_n\right\|_2^2. \tag{5.28}$$

This implies that any term in (5.25) including $\Delta z_{nn}\Delta z_{mm}$ or $\Delta z_{nn}^{*}\Delta z_{mm}$ or Δz_{nn}^2 or $\left|\Delta z_{nn}\right|^2$ is negligible. Furthermore, solving for Δz_{mn} and Δz_{mn}^{*} from (5.27), multiplying them together, and neglecting third-order and forth-order

cross terms, we obtain

$$|\Delta z_{nm}|^2 \approx |\Delta z_{mn}|^2 . \tag{5.29}$$

Similarly, taking the absolute square of both sides of (5.27), using (5.29) and neglecting the forth-order term,

$$\mathrm{Re}\left\{\delta_n^* \delta_m^* \Delta z_{mn} \Delta z_{nm}\right\} \approx -\left|\Delta z_{mn}\right|^2 . \tag{5.30}$$

One can finally simplify (5.25) as

$$\begin{aligned} \Delta J(\mathbf{Z}_0) = \sum_{n=1}^{N} \sum_{l \neq n} |\Delta z_{ln}|^2 \, \mathcal{E}\Big\{ \Big(|\bar{s}_l|^2 - |\bar{s}_n|^2 \Big) g^{(n)}(\mathbf{u}_0) \\ + |\bar{s}_n|^2 \, |\bar{s}_l|^2 \left(g^{(nn)}(\mathbf{u}_0) - g^{(nl)}(\mathbf{u}_0) \right) \Big\} \\ + o\left(\|\Delta \mathbf{Z}\|_F^2 \right) . \end{aligned} \tag{5.31}$$

Assuming that the contrast function is continuous, one has $g^{(nl)}(\mathbf{u}_0) = g^{(ln)}(\mathbf{u}_0)$. By grouping $|\Delta z_{ln}|^2$ with $|\Delta z_{nl}|^2$,

$$\begin{aligned} \Delta J(\mathbf{Z}_0) = \sum_{n=1}^{N-1} \sum_{l=n+1}^{N} |\Delta z_{ln}|^2 \\ \times \mathcal{E}\Big\{ -\Big(|\bar{s}_l|^2 - |\bar{s}_n|^2 \Big) \Big(g^{(l)}(\mathbf{u}_0) - g^{(n)}(\mathbf{u}_0) \Big) \\ + |\bar{s}_l|^2 \, |\bar{s}_n|^2 \left(g^{(ll)}(\mathbf{u}_0) + g^{(nn)}(\mathbf{u}_0) - 2 g^{(ln)}(\mathbf{u}_0) \right) \Big\} \\ + o\left(\|\Delta \mathbf{Z}\|_F^2 \right) . \end{aligned} \tag{5.32}$$

Consequently, $\mathbf{Z}_0$ is a local minimum if the condition in (5.11) holds. □

When the input is not normalized and the contrast function is additively separable, condition (5.11) reduces to the standard stability condition of ICA for the circular complex data given in Theorem 3.1. A similar stability condition of non-separable contrast functions for non-normalized input can be found as well.

It is important to note that Lemma 5.1 requires the source-vectors to be

circular and the mixing matrix to be unitary. When the sources are independent with zero mean and unit variance, whitening can be used to convert an arbitrary invertible mixing matrix into a unitary mixing matrix. Interestingly, since the independent assumption is not explicitly required, for unitary mixing matrices, Lemma 5.1 is applicable to two classes of generally dependent sources namely L_p radially symmetric sources [140] and L_p-nested radially symmetric sources [139]. It is worth noting that Lemma 5.1 also does not assume the sources to be i.i.d. and therefore can be applied for such cases. On the other hand, Lemma 5.1 is only derived for circular sources because the use of general multivariate contrast functions and L_2 normalization make it challenged to simplify the stability condition for non-circular sources. Extending Lemma 5.1 for non-circular sources is worth investigating in future work.

From the stability condition above, one cannot use the arithmetic mean $g(\mathbf{u}) = \sum_n u_n$ as the contrast function. On the other hand, while the analytical verification of (5.11) is difficult for other generic functions, numerical verification can be achieved via Monte Carlo simulation. From numerical simulations, it has been observed that the majority of strictly Schur-concave (or conversely strictly Schur-convex) functions are suitable for recovering the mixing matrix from mixtures of super-Gaussian (or conversely sub-Gaussian) sources. Evidently, when p is sufficiently small so that the sources are highly super-Gaussian, the second term in the stability condition which includes $|\bar{s}_l|^2 |\bar{s}_n|^2$ is negligible. Due to the requirement of Schur-concavity given in (2.12), the first term is automatically satisfied if $g(\mathbf{u})$ is strictly Schur-concave.

5.5 Stability of several directional algorithms

The stability condition of several SCA methods based on directional information is now verified in this section.

5.5.1 The stability of K-hyperlines

K-hyperlines is an extension of K-means for directional data. With reference to (3.18), K-hyperlines has the following reduced cost

$$J^{(\mathrm{KHL})}(\widehat{\mathbf{A}}) = \mathcal{E}\left\{\min_{n=1,\ldots,N}\left(1 - \frac{\left|\widehat{\mathbf{a}}_n^H \bar{\mathbf{x}}\right|^2}{\|\widehat{\mathbf{a}}_n\|_2^2}\right)\right\}. \tag{5.33}$$

By constraining $\widehat{\mathbf{A}}$ to be unitary, all columns of $\widehat{\mathbf{A}}$ have unit norm. This corresponds to the non-smooth contrast function

$$g^{(\mathrm{KHL})}(\mathbf{u}_0) = \min_{n=1,\ldots,N}(1 - |\bar{s}_n|^2). \tag{5.34}$$

The above contrast function is piece-wise linear; it is not differentiable in regions where the two largest coordinates of $\mathbf{u}_0$ are equal. If one simply ignores these regions, the first-order part of the stability condition is satisfied because this function can be shown to be Schur-concave by (2.12) and using sub-gradient; the second-order part is vanished because the contrast function is piece-wise linear leading to second-order derivatives being zeros. Consequently, K-hyperlines can recover the mixing matrix from L_2-normalized mixtures of circular complex super-Gaussian sources for $M = N \geq 2$.

5.5.2 The stability of sparse filtering

Sparse filtering (SF) is a feature learning algorithm which minimizes the L_1/L_2 norm ratio as its sparsity penalty, i.e.,

$$\min_{\widehat{\mathbf{A}}} \mathcal{E}\left\{ \left\|\widehat{\mathbf{A}}^H \mathbf{x}\right\|_1 / \left\|\widehat{\mathbf{A}}^H \mathbf{x}\right\|_2 \right\}. \tag{5.35}$$

For learning on pre-whitened data, due to norm persevering property of a unitary matrix, the objective function of sparse filtering is equivalent to $J^{(\mathrm{SF})}(\widehat{\mathbf{A}}) =$

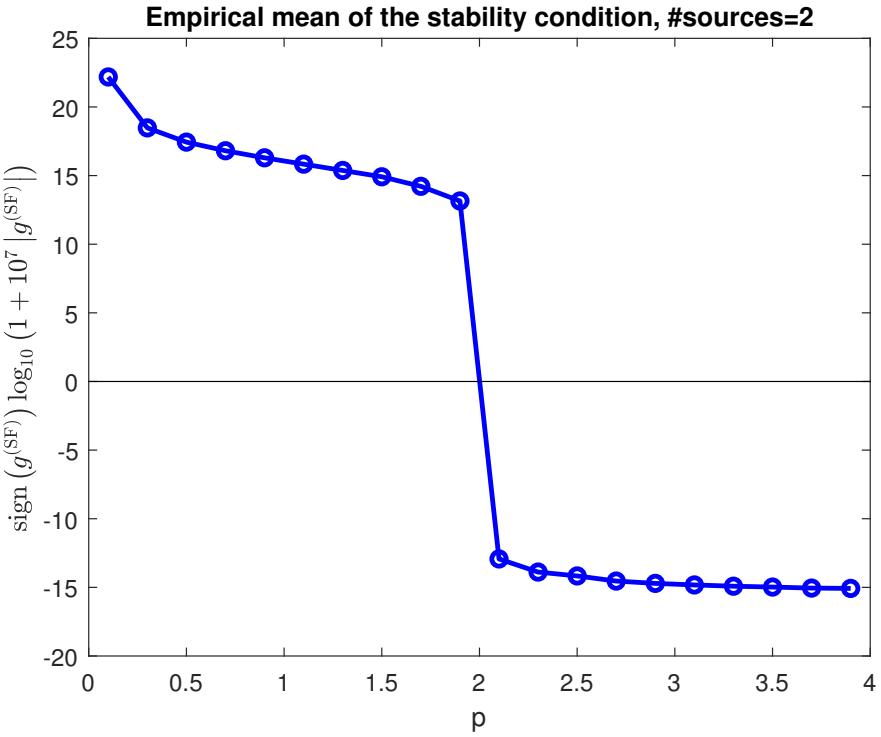


Fig. 5.2: Stability condition of sparse filtering.

$\mathcal{E}\left\{ \sum_{n=1}^{N} \left| \hat{\mathbf{a}}_n^H \bar{\mathbf{x}} \right| \right\}$. The corresponding contrast function is

$$g^{(\mathrm{SF})}(\mathbf{u}_0) = \sum_{n=1}^{N} (|\bar{s}_n|^2)^{1/2}. \tag{5.36}$$

The above contrast function is strictly Schur-concave as it is a summation of strictly concave functions. Noting the separable structure of the contrast function, it is sufficient to verify the stability condition in two-dimensional complex vector space. Consequently, the stability condition for sparse filtering can be simplified to

$$\mathcal{E}\left\{ \frac{|\bar{s}_1|^2}{|\bar{s}_2|} + \frac{|\bar{s}_2|^2}{|\bar{s}_1|} - 2|\bar{s}_1| - 2|\bar{s}_2| \right\} > 0. \tag{5.37}$$

This condition can be confirmed for super-Gaussian CGGD sources (i.e., $0 < p < 2$) by Monte Carlo simulation. Figure 5.2 depicts the average value of the above stability condition from 50 trials where the shape parameter is varied from $p = 0.1$ to $p = 4$ ($p = 2$ corresponds to Gaussian source). In each trial and

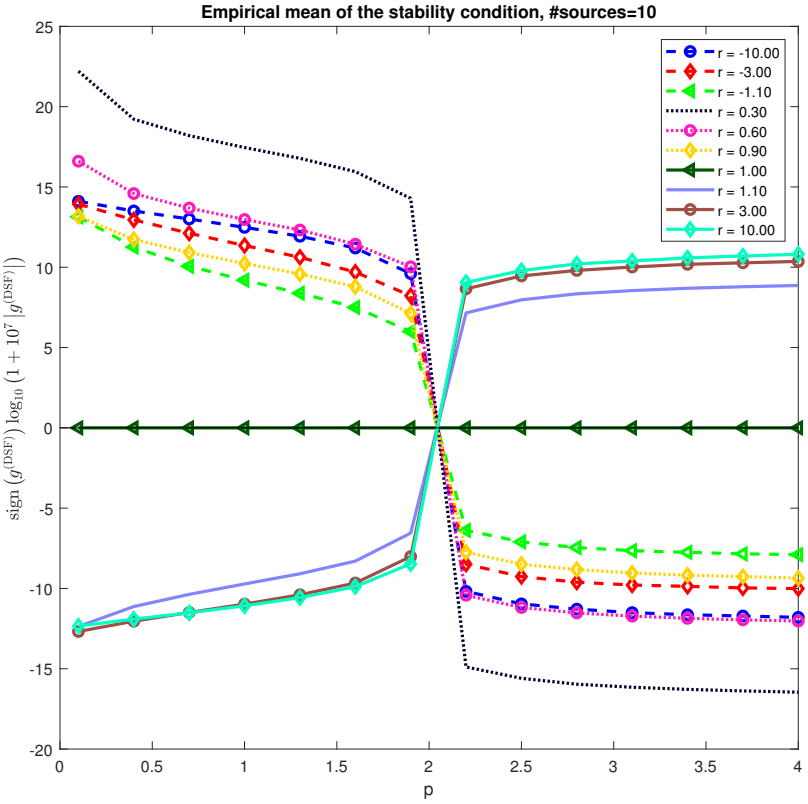


Fig. 5.3: The stability condition for the proposed contrast function.

each value of p, the stability value is computed from 5000 i.i.d. circular CGGD samples. For clarity, the stability value is scaled with the pseudo-log function, $f(x) = \text{sign(x)}\,\text{lg}(1 + 10^7\,|x|)$. It can be seen the stability condition is positive when $p < 2$. This implies that sparse filtering satisfies the stability condition for circular complex super-Gaussian sources as well.

5.5.3 The stability of the directional sparse filtering

The proposed contrast function of directional sparse filtering

$$g^{(\text{DSF})}(\mathbf{u}_0) = \left[\frac{1}{N}\sum_{n=1}^{N}\left(1 - |\bar{s}_n|^2\right)^r\right]^{1/r} \tag{5.38}$$

shown in Section 4.2.3 is non-separable and therefore verifying its stability condition is far more complicated than the above. The stability condition is given by

$$\mathcal{E}\left\{I_1 + I_2(I_3 + I_4 I_5)\right\} > 0, \tag{5.39}$$

where

$$\begin{aligned}
I_1 &= \left(|\bar{s}_l|^2 - |\bar{s}_n|^2\right)\left[\left(1-|\bar{s}_n|^2\right)^{r-1} - \left(1-|\bar{s}_n|^2\right)^{r-1}\right] \\
&\quad \times \left[\sum_{i=1}^{N}\left(1-|\bar{s}_i|^2\right)^r\right]^{(1/r)-1}, && (5.40)\\
I_2 &= \frac{(r-1)}{\left(1-|\bar{s}_l|^2\right)^2\left(1-|\bar{s}_n|^2\right)^2}|\bar{s}_l|^2|\bar{s}_n|^2\left[\sum_{i=1}^{N}\left(1-|\bar{s}_i|^2\right)^r\right]^{(1/r)-2}, && (5.41)\\
I_3 &= \left(1-|\bar{s}_l|^2\right)^r\left(1-|\bar{s}_n|^2\right)^r\left(|\bar{s}_l|^2+|\bar{s}_n|^2-2\right)^2, && (5.42)\\
I_4 &= \sum_{i\neq n\wedge i\neq l}\left(1-|\bar{s}_i|^2\right)^r, && (5.43)\\
I_5 &= \left(1-|\bar{s}_l|^2\right)^r\left(1-|\bar{s}_n|^2\right)^2 + \left(1-|\bar{s}_n|^2\right)^r\left(1-|\bar{s}_l|^2\right)^2. && (5.44)
\end{aligned}$$

Even though one cannot verify the above condition analytically, numerical experiment suggests that the stability condition of DSF holds. More specifically, with $r > 1$, it can be used to deduce the mixing matrix from sub-Gaussian mixtures. Likewise, for $r < 1$ and $r \neq 0$, it is suitable to estimate the mixing matrix from mixtures of super-Gaussian. Figure 5.3 shows the average stability value of 50 trials. The simulation setup is similar to the sparse filtering case as detailed in Section 5.5.3. One can see that the contrast function in (5.38) for $r \neq 1$ divides the probability space with regard to Gaussian distribution into two half-spaces. For super-Gaussian sources ($0 < p < 2$), the stability value is positive when $r < 1$. Similarly, for sub-Gaussian sources ($p > 2$), the stability value is positive when $r > 1$. This implies that the proposed DSF can recover the mixing matrix from mixtures of super-Gaussian sources if $r < 1$ and from mixtures of sub-Gaussian sources if $r > 1$.

5.6 Summary

In this chapter, the stability condition is derived for learning the mixing matrix from L_2-normalized mixtures using non-separable contrast function. It is shown that several SCA methods based on directional statistics can recover the mixing matrix from mixtures of super-Gaussian sources. Since not all super-Gaussian distributed sources are approximately disjoint, the stability condition implies that, for the determined case, these directional methods can work with wider class of sources than they are designed to. The implication of this result is that both these SCA methods and ICA can be interpreted as algorithms that perform maximum quasi-likelihood estimation. Furthermore, in terms of the connection between these SCA methods and ICA, the minimizers of their cost functions include the mixing matrix when the sources are i.i.d. circular complex super-Gaussian. In other words, given a set of independent super-Gaussian sources and an invertible mixing matrix, one can employ SCA methods which satisfy the proposed stability condition to recover the mixing matrix from the L_2-normalized observed mixtures.

Chapter 6

A Method Based on L-BFGS to Solve Complex-valued ICA by Entropy Bound Minimization

6.1 Introduction

This chapter addresses the problem of solving constrained objective functions of complex-valued ICA by entropy bound minimization (CEBM). In the conventional CEBM, the demixing filters are updated independently using projected gradient descending [143]. As projected gradient descending does not exploit the curvature information, this may lead to a sub-optimal performance in the flat regions of the cost function. As discussed in Section 3.2.2.1, the fixed-point iteration can mitigate this issue to a certain extent. However, the updates of the demixing filters are also computed independently in this approach. By stacking all the demixing filters into a single column vector, it can be seen that the joint Hessian of all demixing filters is restricted to be block-diagonal in fixed-point iteration. This may also affect the estimation performance when the independent model does not hold exactly. To levitate this issue, in this chapter, the

constrained CEBM is formulated as unconstrained optimization problems with respect to auxiliary variables. These problems are then solved using the vanilla L-BFGS algorithm [124]. As opposed to the conventional CEBM, the proposed method does not apply any restriction on the joint Hessian. This will reflect the true Hessian of the ICA objective function more accurately. As a consequence, the proposed method can achieve higher estimation performance than the conventional CEBM algorithm.

For ease of presentation, the mixing process in (3.2) is re-written as

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{6.1}$$

where $\mathbf{x} = [x_1, \ldots, x_N]^T$ is the vector of the observed data, $\mathbf{A} \in \mathbb{C}^{N \times N}$ is the unknown invertible mixing matrix, N is the number of sources, and $\mathbf{s} = [s_1, \ldots, s_N]^T$ is the vector of the unknown independent sources. Note that the time-frame index and frequency-bin index have been removed for brevity.

For an arbitrarily chosen complex-valued random variable $z = \Re\{z\} + \jmath\Im\{z\}$, the complex probability distribution is defined by $f_Z(z) \triangleq f_Z(\Re\{z\}, \Im\{z\})$. The entropy of z is then given by

$$H(z) \triangleq -\mathcal{E}\{\log f_Z(\Re\{z\}, \Im\{z\})\}, \tag{6.2}$$

where $\Re\{z\}$ and $\Im\{z\}$ are, respectively, the real part and imaginary part of z.

6.2 Complex-valued ICA by entropy bound minimization

Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ be the demixing matrix and let $\mathbf{y} = [y_1, \ldots, y_N]^T = \mathbf{W}^H\mathbf{x}$ be the demixed signals. To estimate $\mathbf{s}$, one can minimize the mutual information

of demixed signals given by [9, 22]

$$\mathcal{I}(\mathbf{W}) = E\left\{ \log \frac{f_Y(\mathbf{y})}{\prod_n f_{Y_n}(y_n)} \right\} \tag{6.3}$$

$$= \sum_n H(y_n) - H(\mathbf{y}) \tag{6.4}$$

$$= \sum_{n=1}^{N} H(\mathbf{w}_n^H \mathbf{x}) - H(\mathbf{x}) - 2\log|\det(\mathbf{W})|\,, \tag{6.5}$$

where $f_Y(\mathbf{y})$ is the joint distribution of all demixed signals, $f_{Y_n}(y_n)$ is the distribution of the ith demixed signal, $H(\mathbf{y})$ is the entropy of $\mathbf{y}$, $H(y_n)$ is the entropy of y_n, and $H(\mathbf{x})$ is the entropy of $\mathbf{x}$. Here, $H(\mathbf{x})$ can be ignored as it is independent of $\mathbf{W}$. Since $\mathcal{I}(\mathbf{W})$ is unbounded below, the second term can be arbitrarily small and hence certain restriction on $\mathbf{W}$ should be applied. When the data are whitened, one can constrain $\mathbf{W}$ to be unitary (i.e., $\mathbf{W}\mathbf{W}^H = \mathbf{W}^H\mathbf{W} = \mathbf{I}$). In this case, the second term of (6.5) can be ignored. A more relaxed restriction is to constrain each column of $\mathbf{W}$ to have unit norm so that $|\det(\mathbf{W})|$ is bounded above by 1. The tightest upper bound and corresponding nonlinear function(s) will subsequently be chosen to compute the value and the gradient of each $H(y_n)$.

In the so-called complex-valued ICA by entropy bound minimization (CEBM), several upper bounds of $H(y_n)$ corresponding to some predetermined measuring functions are estimated based on maximum entropy principle (MEP). The tightest upper bound and corresponding nonlinear function(s) will subsequently be chosen to compute the value and the gradient of each $H(y_i)$.

For simplicity, consider the entropy estimator for a real variable v that has zero-mean and unit-variance. Given a measuring function $g(v)$, suppose that $\mathcal{E}\{g(v)\}$ evaluated on the observed data is μ_g, a density function $q(v)$ will be

constructed according to MEP

$$\begin{aligned} \max_{q(v)} \ & \int_{-\infty}^{\infty} q(v) \log q(v) dv \\ \text{s.t.} \ & \int_{-\infty}^{\infty} q(v) dv = 1, \ \int_{-\infty}^{\infty} v q(v) dv = 0, \\ & \int_{-\infty}^{\infty} v^2 q(v) dv = 1, \ \int_{-\infty}^{\infty} g(v) q(v) dv = \mu_g. \end{aligned} \tag{6.6}$$

In the above equation, the constraint $\int_{-\infty}^{\infty} q(v)dv = 1$ is applied to make sure that $q(v)$ is a proper density function. The last three constraints are due to $\mathcal{E}\{v\} = 0$, $\mathcal{E}\{v^2\} = 1$ and $\mathcal{E}\{g(v)\} = \mu_g$. The general form for the solution of (6.6) is given by

$$q(v) = A \exp(-av^2 - bv - cg(v)), \tag{6.7}$$

where A, a, b, and c are coefficients that can be found using numerical iteration [144] to satisfy the last three constraints in (6.6). After the maximum entropy density is found, the entropy bound corresponding to $g(v)$, if exists, can be computed as follows

$$H^{(\text{bound})}(v) = -\log A + a + c\mu_g. \tag{6.8}$$

The entropy bound can further be improved by utilizing several measuring functions to calculate the smallest bound among several upper-bounds.

With the method to estimate entropy bound for real-valued variables, the entropy bound for any complex-valued variable can also be derived by representing a complex-valued variable with two real-valued variables [22].

6.3 CEBM based on L-BFGS

We propose a second-order method in Euclidean space based on L-BFGS without any structural restriction on the joint the Hessian of vec($\mathbf{W}$), a vector created

by stacking all columns of $\mathbf{W}$. For $\mathcal{I}(\mathbf{W})$ given in (6.5), suppose that the total entropy bound $\mathcal{H}(\mathbf{W}) = \sum_n H(\mathbf{w}_n^H \mathbf{x})$ and its gradient $\frac{\partial \mathcal{H}}{\partial \mathbf{W}^*}$ are estimated by numerical method in [22], the gradient direction of $\mathcal{I}(\mathbf{W})$ is then given by

$$\frac{\partial \mathcal{I}(\mathbf{W})}{\partial \mathbf{W}^*} = \frac{\partial \mathcal{H}(\mathbf{W})}{\partial \mathbf{W}^*} - \mathbf{W}^{-H}. \tag{6.9}$$

We consider two constrained optimization problems

$$\min_{\mathbf{W}} \ \mathcal{I}(\mathbf{W}), \ \text{s.t.} \ \mathbf{W}\mathbf{W}^H = \mathbf{I}_M, \tag{6.10}$$

$$\min_{\mathbf{W}} \ \mathcal{I}(\mathbf{W}), \ \text{s.t.} \ \|\mathbf{w}_1\| = \ldots = \|\mathbf{w}_N\| = 1. \tag{6.11}$$

For an arbitrary full rank matrix $\widetilde{\mathbf{W}} \in \mathbb{C}^{N\times N}$, let $\widehat{\mathbf{W}} = (\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H)^{-0.5}\widetilde{\mathbf{W}}$ be the nearest unitary matrix of $\widetilde{\mathbf{W}}$ and $\overline{\mathbf{W}} = [\bar{\mathbf{w}}_1, \ldots, \bar{\mathbf{w}}_N]$ be the column normalized matrix of $\widetilde{\mathbf{W}}$, where $\bar{\mathbf{w}}_i = \tilde{\mathbf{w}}_i / \|\tilde{\mathbf{w}}_i\|$. It is important to note that $\widehat{\mathbf{W}}$ and $\overline{\mathbf{W}}$ are functions of $\widetilde{\mathbf{W}}$. By definition, $\widehat{\mathbf{W}}$ and $\overline{\mathbf{W}}$ always satisfy their respective constraints in (6.10) and (6.11). With these re-parameterizations, (6.10) and (6.11) can, respectively, be solved via the following unconstrained optimization problems

$$\min_{\widetilde{\mathbf{W}}\in\mathbb{C}^{N\times N}} \ \mathcal{I}(\widehat{\mathbf{W}}) + \frac{\lambda_1}{2}\|\widetilde{\mathbf{W}}\|_F^2, \tag{6.12}$$

$$\min_{\widetilde{\mathbf{W}}\in\mathbb{C}^{N\times N}} \ \mathcal{I}(\overline{\mathbf{W}}) + \frac{\lambda_2}{2}\|\widetilde{\mathbf{W}}\|_F^2, \tag{6.13}$$

where λ_1 and λ_2 are two hyper-parameters and $\|\widetilde{\mathbf{W}}\|_F^2$ is the Frobenius norm of $\widetilde{\mathbf{W}}$. It is noted that $\mathcal{I}(\widehat{\mathbf{W}})$ and $\mathcal{I}(\overline{\mathbf{W}})$ are independent of $\|\widetilde{\mathbf{W}}\|_F^2$ as long as $\widetilde{\mathbf{W}}$ is full rank because of the normalization in $\widehat{\mathbf{W}}$ and $\overline{\mathbf{W}}$. Therefore, given the small values of λ_1 and λ_2, the second term of (6.12) and (6.13) will penalize $\widetilde{\mathbf{W}}$ which has large Frobenius norm. Nevertheless, for the exact minimization of $\mathcal{I}(\widehat{\mathbf{W}})$ and $\mathcal{I}(\overline{\mathbf{W}})$, one should use $\lambda_1 = \lambda_2 = 0$. Using Lemma 4.1, the gradient of the

nested cost in (6.12) is given by

$$\mathbf{K} = -\Big(\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\Big(\frac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widehat{\mathbf{W}}^*}\Big)\mathbf{V}\Big) \oslash (\mathbf{1}_N\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_N^T), \tag{6.14}$$

$$\frac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}^*} = \mathbf{U}(\mathbf{K}^H + \mathbf{K})\boldsymbol{\Sigma}\mathbf{V}^H + \mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\Big(\frac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widehat{\mathbf{W}}^*}\Big), \tag{6.15}$$

where $\oslash$ is the element-wise matrix division, $\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$ denotes the singular value decomposition of $\widetilde{\mathbf{W}}$, $\boldsymbol{\sigma}$ is the diagonal of $\boldsymbol{\Sigma}$, and $\mathbf{1}_N \in \mathbb{C}^N$ is the all-ones column vector. Similarly, using Wirtinger calculus, the gradient w.r.t. each column of the nested cost given in (6.13) is

$$\begin{aligned}
\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \tilde{\mathbf{w}}_n^*} =& \frac{\partial \bar{\mathbf{w}}_n^*}{\partial \tilde{\mathbf{w}}_n^*}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_n^*} + \frac{\partial \bar{\mathbf{w}}_n}{\partial \tilde{\mathbf{w}}_n^*}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_n} \\
=& \Big(\frac{\mathbf{I}}{\|\tilde{\mathbf{w}}_n\|} - \frac{\tilde{\mathbf{w}}_n\tilde{\mathbf{w}}_n^H}{2\|\tilde{\mathbf{w}}_n\|^3}\Big)\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_n^*} - \frac{\tilde{\mathbf{w}}_n\tilde{\mathbf{w}}_n^T}{2\|\tilde{\mathbf{w}}_n\|^3}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_n} \\
=& \frac{1}{\|\tilde{\mathbf{w}}_n\|}\Big(\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_n^*} - \bar{\mathbf{w}}_n\Re\Big\{\bar{\mathbf{w}}_n^H\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_n^*}\Big\}\Big).
\end{aligned} \tag{6.16}$$

In the last step, since $\mathcal{I}(\overline{\mathbf{W}})$ is real-valued, we have used $(\frac{\partial \mathcal{I}}{\bar{\mathbf{w}}_n})^* = \frac{\partial \mathcal{I}}{\partial \bar{\mathbf{w}}_n^*}$. The gradients, $\frac{\partial \mathcal{I}}{\partial \bar{\mathbf{w}}_n^*}$ and $\frac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widehat{\mathbf{W}}^*}$, are evaluated using (6.9).

The proposed complex-valued ICA optimization problems are novel because they are unconstrained in the Euclidean space while being able to maintain the constraints of the arguments of the objective functions. Since the cost functions in (6.12) and (6.13) are unconstrained, one can vectorize these gradients and solve the respective cost functions in real-composite space or complex-augmented space. In this work, optimization in the real-composite space is chosen. In line with this, the complex-valued gradients are converted into real-composite space using (2.3). Here, the re-parameterized objectives can be solved using any real-valued gradient-based solver. However, to address the performance reduction of ICA when the sources approach Gaussianity, (6.12) and (6.13) are solved using L-BFGS [124]. An equivalent implementation in complex-augmented space is also

1. Find initial matrix $\mathbf{W}_0$ using [10]
2. Solving (6.12) using L-BFGS, given initial matrix $\widetilde{\mathbf{W}} \equiv \mathbf{W}_0$
3. $\widetilde{\mathbf{W}}_0 \leftarrow (\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H)^{-0.5}\widetilde{\mathbf{W}}$
4. Solving (6.13) using L-BFGS, given initial matrix $\widetilde{\mathbf{W}} \equiv \widetilde{\mathbf{W}}_0$
5. $\mathbf{W} \leftarrow [\frac{\widetilde{\mathbf{w}}_1}{\|\widetilde{\mathbf{w}}_1\|}, \ldots, \frac{\widetilde{\mathbf{w}}_N}{\|\widetilde{\mathbf{w}}_N\|}]$

Table 6.1: Complex-valued ICA by entropy bound minimization based on L-BFGS.

possible using the complex-valued L-BFGS software accompanied by [51]. We note that the difference in performance between real-composite implementation and complex-augmented implementation is negligible.

The proposed method is summarized in Table 6.1. This algorithm comprises three main steps. It is initialized using [10] with the contrast function of $g(y) = y^{1.25}$. After that, an ICA estimation is performed using (6.12), then the estimate is refined using (6.13). The refinement step are necessary because the demixing matrix is not exactly unitary due to the imperfection of whitening [143]. Nevertheless, in our pilot study, we found that the refinement step is only required for highly sparse sources.

The main advantage of our proposed method is that the approximate Hessian of vec($\widetilde{\mathbf{W}}$) based on L-BFGS is not limited to be block diagonal. The second advantage is that strong Wolfe line search of L-BFGS guarantees a sufficient decrease of the objective at each step, whereas there is no such guarantee in the fixed-point iteration. Last but not least, the method can be applied for other complex-valued ICA objectives or any real ICA objectives with ease.

6.4 Simulations

We compare the proposed complex-valued entropy bound minimization based on LBFGS (CEBM+), with CEBM [22] and complex-valued ICA by maximization of non-Gaussianity (CMN) [10]. Methods such as complex-valued FastICA [7]

and complex-valued non-circular FastICA [21] are not included as they normally perform worse than CEBM and CMN. The proposed CEBM+ is implemented in MATLAB using minFunc toolbox [128] where the maximum number of iterations is 100 (40 iterations for (6.12) and 60 iterations for (6.13)). We remark that the maximum iterations of CEBM+ is equal to that of CEBM. The algorithm is stopped if the directional gradient is smaller than 10^{-5}. We set hyperparameters λ_1 and λ_2 to 0.001.

We employ two performance indexes – the normalized Amari index (NAI) and the percentage of failed trials, both of which are commonly used in ICA literature. An algorithm is deemed to have failed if the matrix $\mathbf{P} = [p_{ij}] = \mathbf{W}^H\mathbf{A}$ has two rows where the largest magnitude values are at the same column. Intuitively, the magnitude p_{ij} is the cosine-similarity of $\mathbf{w}_i$ and $\mathbf{a}_j$ (given that $\mathbf{W}$ and $\mathbf{A}$ are unitary due to prewhitening). If there are two columns $\mathbf{w}_m$ and $\mathbf{w}_n$ that are both associated to the same column of $\mathbf{A}$, the algorithm should be considered as being failed. Lower NAI and lower failure percentage indicate better performance.

For each simulation, we repeat 100 trials, where for each trial, new $\mathbf{A}$ and $\mathbf{s}$ are generated. The mixed data are subsequently generated using (6.1). The real and imaginary part of each element of $\mathbf{A}$ are drawn from standard normal distribution. Each source s_i is drawn from zero-mean complex generalized Gaussian distribution [136] where two parameter p and ρ control the shape of the distribution and the circularity of the source, respectively. The circular coefficient ρ is defined by the correlation between $\Re\{s_i\}$ and $\Im\{s_i\}$ when variances of $\Re\{s_i\}$ and $\Im\{s_i\}$ are assumed to be 1. A source is complex Gaussian distributed if $p = 2$ and a source is circular if $\rho = 0$.

Figure 6.1 depicts the performance with respect to sample size. Here, we created sixteen sources ($N = 16$) with the shape parameters being evenly spaced between $0.5 \leq p \leq 3.5$. All sources are non-circular with ρ selected in $(0, 1)$

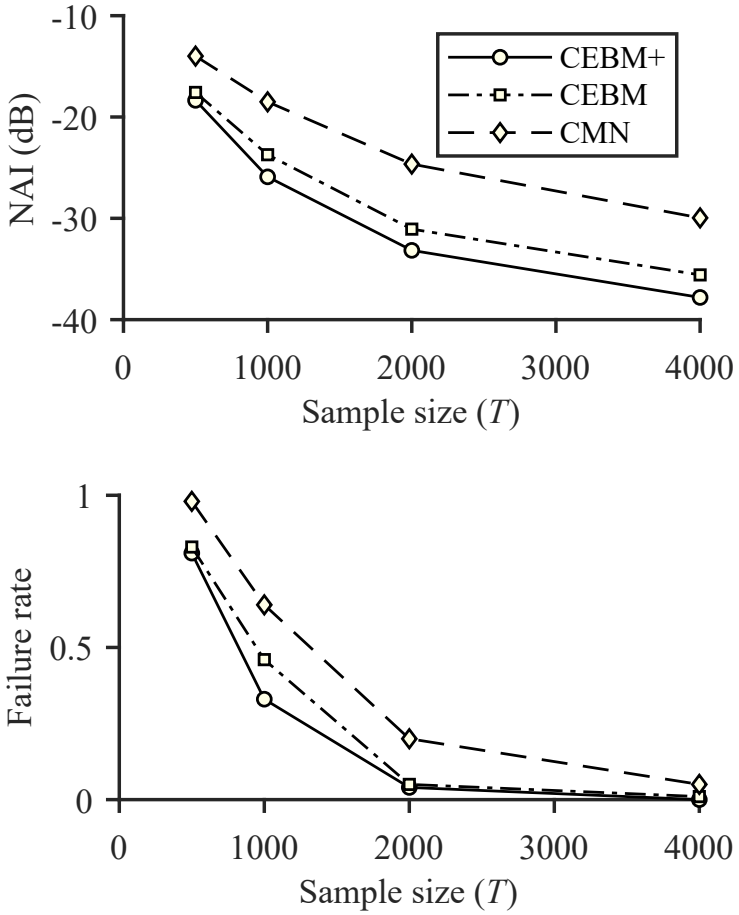


Fig. 6.1: Average NAI and failure rate as functions of the sample size.

randomly. This simulation shows that our proposed algorithm achieves lower sample complexity than CMN and CEBM.

Figure 6.2 illustrates the performance as a function of the number of sources. Similar to previous simulation, we created sources with shape parameter that is uniformly spaced in the range $[0.5, 3.5]$. The sample size is fixed at 1000 samples. In this simulation, our proposed algorithm yields better estimates than CMN and CEBM. We can also observe that the proposed method is more reliable than the baselines when the size of the mixing matrix increases. Note that, both our proposed method and CEBM share the same objective function. This highlights the benefits of joint optimization and the effectiveness of the proposed approach.

Figure 6.3 shows the effect of shape parameter on the performance of the tested complex-valued ICA algorithms. In each trial, we generated sixteen i.i.d.

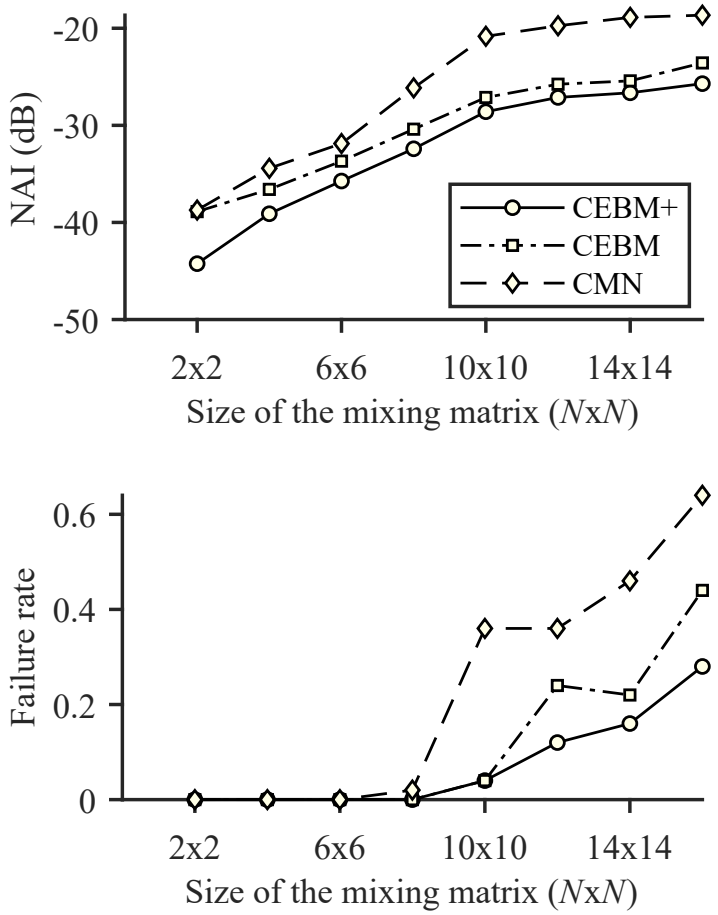


Fig. 6.2: Average NAI and failure rate as functions of the number of sources.

complex generalized Gaussian sources. The sample size is 1000 and all the sources are non-circular with ρ randomly selected in $(0, 1)$. The shape parameter p is varied from 0.4 to 3.4. As shown by this result, the estimation performance reduces when the Gaussianity of the sources increases. In all tested cases, our algorithm generally performs better than the baseline algorithms for sources with low non-Gaussianity as seen from the failure rate.

Overall, the proposed method performs equally or higher than CMN and CEBM in most of the simulations. In particular, the proposed algorithm has lower sample complexity and can yield better estimate than CMN and CEBM when the source distributions are close to Gaussian. Nevertheless, the running time of CEBM+ is twice as much as CEBM on average due to the costly L-BFGS update.

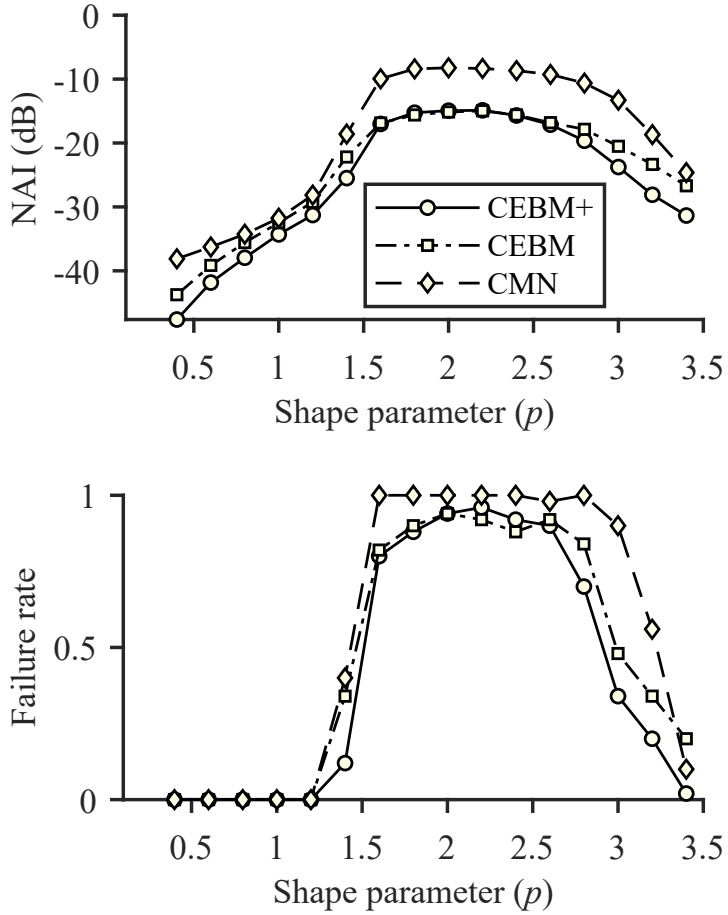


Fig. 6.3: Average NAI and failure rate as functions of the shape parameter ($p = 2$ is Gaussian).

6.5 Summary

In this chapter, an algorithm that employs L-BFGS to improve the estimation performance of complex-valued ICA by entropy bound minimization is proposed. Since the objective function is constrained by unitary condition or unit column-norm condition, it is not feasible to use the L-BFGS algorithm directly. This issue is addressed by formulating complex-valued ICA as an unconstrained optimization problem. This allows the joint Hessian of all demixing filter to be approximated jointly using L-BFGS updating rule. As a result, the proposed ICA algorithm outperforms that of its counterparts in many cases.

Chapter 7

Simultaneous Estimation of the Temporal Whitening Filters and the Mixing Matrix in Complex-valued ICA by Entropy Rate Bound Minimization

7.1 Introduction

Complex-valued ICA by entropy rate bound minimization (CERBM) [12] is a generalization of CEBM where the dependency among source samples is also taken into account. As opposed to CEBM and CEBM+, CERBM can estimate the mixing matrix from mixtures of colored Gaussian signals. In the conventional CERBM, a set of temporal whitening filters will be estimated to reduce the temporal dependency in each demixed signal. After that, the mixing matrix is then computed. In this chapter, CERBM+ is proposed. The motivation of the proposed approach lies with the fact that finding the temporal whitening filters

and the demixing matrix separately is computationally expensive. In view of this, CERBM+ seeks to estimate both the temporal whitening filters and the mixing matrix at the same time efficiently. It will be shown in simulations that the proposed approach is not only reducing the computational cost, it also improves the estimation performance.

For ease of presentation, we re-write the mixing process in (3.2) as follows

$$\mathbf{x}(\tau) = \mathbf{A}\mathbf{s}(\tau), \tag{7.1}$$

where $\tau = 1, \ldots, T$ is the time-frame index, T is the total number of samples, $\mathbf{x}(\tau) = [x_1(\tau), \ldots, x_N(\tau)]^T$ is the vector of observed data, $\mathbf{A} \in \mathbb{C}^{N \times N}$ is the unknown invertible mixing matrix, N is the number of sources, and $\mathbf{s}(\tau) = [s_1(\tau), \ldots, s_N(\tau)]^T$ is the vector of unknown independent source signals. It should be noted that the frequency-bin index have been removed for brevity.

Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ be the demixing matrix and let

$$\mathbf{y}(\tau) = [y_1(\tau), \ldots, y_N(\tau)]^T = \mathbf{W}^H \mathbf{x}(\tau) \tag{7.2}$$

be the estimated sources. For a sequence of complex-valued random variable $y_n(\tau)$, the entropy rate of y_n is defined by

$$H_r(y_n) = \lim_{T \to \infty} \frac{1}{T} H(y_n(1), \ldots, y_n(T)), \tag{7.3}$$

where $H(y_n(1), \ldots, y_n(T))$ is the joint entropy for T samples. In CERBM, one minimizes the mutual information rate

$$\mathcal{I}_r(\mathbf{W}) = \sum_{n=1}^{N} H_r(y_n) - 2 \log |\det(\mathbf{W})| - C, \tag{7.4}$$

where C is a constant that does not depend on $\mathbf{W}$ and $H_r(y_n)$ is the entropy

rate of y_n. If the samples are i.i.d., the mutual information rate $\mathcal{I}_r(\mathbf{W})$ and the entropy rate $H_r(y_n)$ reduce to the mutual information $\mathcal{I}(\mathbf{W})$ and the entropy $H(y_n)$ discussed in Chapter 6.

To simplify the estimation of the mutual information rate, CERBM assumes that there exists a causal widely linear time-invariant filter that eliminates the temporal dependency among $y_n(\tau)$,

$$v_n(\tau) = \sum_{k=0}^{K} p_{nk}^* y_n(\tau - k) + \sum_{k=0}^{K} q_{nk} y_n^*(\tau - k) = \mathbf{b}_n^H \mathbf{y}_{n\mathcal{C}}^{(\mathrm{K}+1)}(\tau), \tag{7.5}$$

where $v_n(\tau)$ is a white process, $\mathbf{b}_n = [p_{n0}, \ldots, p_{nK}, q_{n0}^*, \ldots, q_{nK}^*]^T \in \mathbb{C}^{2K+2\times 1}$ is the temporal whitening filter, p_{nk} and q_{nk} are the filtering coefficients, K is the filter order, and $\mathbf{y}_{n\mathcal{C}}^{(\mathrm{K}+1)}(\tau) = [y_n(\tau), \ldots, y_n(\tau - K), y_n^*(\tau), \ldots, y_n^*(\tau - K)]^T \in \mathbb{C}^{2K+2\times 1}$. Note that, the use of the time-invariant temporal whitening filter as in (7.5) may not yield a desirable improvement if the sources are non-stationary.

Note that, (7.5) can be rewritten in real domain as follows

$$\begin{bmatrix} \Re\{v_n(\tau)\} \\ \Im\{v_n(\tau)\} \end{bmatrix} = \sum_{k=0}^{K} \mathbf{L}(k) \begin{bmatrix} \Re\{y_n(\tau - k)\} \\ \Im\{y_n(\tau - k)\} \end{bmatrix}, \tag{7.6}$$

where

$$\mathbf{L}(k) = \begin{bmatrix} \Re\{p_{nk}\} + \Re\{q_{nk}\} & \Im\{p_{nk}\} + \Im\{q_{nk}\} \\ -\Im\{p_{nk}\} + \Im\{q_{nk}\} & \Re\{p_{nk}\} - \Re\{q_{nk}\} \end{bmatrix}. \tag{7.7}$$

Let $\mathbb{B}$ be the Jacobian matrix of the mapping from $\Re\{y_n(1)\}, \Im\{y_n(1)\}, \ldots, \Re\{y_n(\tau)\}, \Im\{y_n(\tau)\}$ to $\Re\{v_n(1)\}, \Im\{v_n(1)\}, \ldots, \Re\{v_n(\tau)\}, \Im\{v_n(\tau)\}$ specified in (7.6). From (6.2) and the change of variables concept, we have

$$H(\Re\{v_n(1)\}, \Im\{v_n(1)\}, \ldots, \Re\{v_n(\tau)\}, \Im\{v_n(\tau)) = H(\Re\{y_n(1)\}, \Im\{y_n(1)\}, \ldots, \Re\{y_n(\tau)\}, \Im\{y_n(\tau)) + \ln|\det \mathbb{B}|. \tag{7.8}$$

By definition, $\mathbb{B}$ is a block matrix where the submatrix denoting the relationship between $v_n(i)$ and $y_n(j)$ is given by

$$\lfloor \mathbb{B} \rfloor_{ij} = \begin{bmatrix} \partial\Re\{v_n(i)\}/\partial\Re\{y_n(j)\} & \partial\Re\{v_n(i)\}/\partial\Im\{y_n(j)\} \\ \partial\Im\{v_n(i)\}/\partial\Re\{y_n(j)\} & \partial\Im\{v_n(i)\}/\partial\Im\{y_n(j)\} \end{bmatrix}. \tag{7.9}$$

Since (7.6) is causal, $\lfloor \mathbb{B} \rfloor_{ij}$ is a zero 2×2 matrix for any $i < j$. Therefore, the Jacobian matrix $\mathbb{B}$ is a block lower triangular matrix with the diagonal $\lfloor \mathbb{B} \rfloor_{ii} = \mathbf{L}(0)$ for $i = 1, \ldots, \tau$. As a result, one has

$$|\det \mathbb{B}| = |\det \mathbf{L}(0)|^{\tau} = \left| |p_{n0}|^2 - |q_{n0}|^2 \right|^{\tau}. \tag{7.10}$$

Consequently, from (7.3), (7.8), and (7.10), we obtain

$$H_r(v_n) = H_r(y_n) + \ln \left| |p_{n0}|^2 - |q_{n0}|^2 \right|. \tag{7.11}$$

Hence, to persevere the entropy rate between the input and output of the filtering operation in (7.5), the following condition must be satisfied [12, 145]

$$\mathbf{b}_n^H \mathbf{D}_b \mathbf{b}_n = 1, \tag{7.12}$$

where

$$\mathbf{D}_b = \mathrm{diag}(1, \underbrace{0, \ldots, 0}_{K}, -1, \underbrace{0, \ldots, 0}_{K}). \tag{7.13}$$

The optimal filter $\mathbf{b}_n$ can then be found by solving

$$\min_{\mathbf{b}_n} \quad H(v_n), \quad \text{s.t.} \quad \mathbf{b}_n^H \mathbf{D}_b \mathbf{b}_n = 1. \tag{7.14}$$

In the case where v_n is a white Gaussian process, (7.14) has closed-form solution given by the dominant eigen-vector of $\mathcal{E}\{\mathbf{y}_{n\mathcal{C}}^{(\mathrm{K}+1)}(\mathbf{y}_{n\mathcal{C}}^{(\mathrm{K}+1)})^H\}^{-1}\mathbf{D}_b$. Once the demixed signals are temporally whitened, due to whiteness of $v_n(\tau)$, the mutual

entropy rate becomes

$$\mathcal{I}_r(\mathbf{W}) \propto \frac{1}{T}\sum_{n=1}^{N}\sum_{\tau=1}^{T} H(v_n(\tau)) - 2\log|\det(\mathbf{W})|\,. \tag{7.15}$$

In the above equation, $H(v_n(\tau))$ and its gradient can be estimated using the maximum entropy principle as described in Chapter 6. Table 7.1 summarizes the main steps of the CERBM algorithm.

1. Find initial matrix $\mathbf{W}$ using the conventional CEBM
2. Initialize $\mathbf{b}_n$ using close-form solution, $\forall n$.
3. Repeat until convergence
4. For $n = 1$ to N do
5. Solve (7.14) for optimal $\mathbf{b}_n$.
6. Update $\mathbf{w}_n \leftarrow \mathbf{w}_n - \eta\frac{\partial \mathcal{I}_r(\mathbf{W})}{\partial \mathbf{w}_n^*}$.
7. Normalize $\mathbf{w}_n \leftarrow \mathbf{w}_n / \|\mathbf{w}_n\|_2$.

Table 7.1: Conventional complex-valued ICA by entropy rate bound minimization.

To this end, it can be seen that CERBM involves the estimation of the temporal whitening filters followed by the estimation of the conventional entropy bound for $H(v_n(\tau))$. For each iteration, one must minimize $H(v_n(\tau))$ with respected to $\mathbf{b}_n$ for all n, this problem is almost as computationally heavy as a whole ICA algorithm. Furthermore, since $y_n(\tau) = \mathbf{w}_n^H\mathbf{x}(\tau)$, one has

$$v_n(\tau) = \mathbf{b}_n^H\mathbf{X}_{\mathcal{C}}^{(\mathrm{K}+1)}(\tau)\mathbf{w}_n^*, \tag{7.16}$$

where

$$\mathbf{X}_{\mathcal{C}}^{(K+1)}(\tau) = [\mathbf{x}(\tau),\ldots,\mathbf{x}(\tau-K),\mathbf{x}^*(\tau),\ldots,\mathbf{x}^*(\tau-K)]^T \in \mathbb{C}^{2K+2\times N}. \tag{7.17}$$

Clearly, the entropy $H(v_n(\tau))$ depends jointly on both $\mathbf{b}_n$ and $\mathbf{w}_n$ for all n. Therefore, the convergence of CERBM can be improved by jointly optimizing

the temporal whitening filters and the mixing matrix.

7.2 The proposed CERBM+ algorithm

We consider the joint optimization problem for CERBM under the unit column-norm constraint of the demixing matrix and the entropy preservation constraint of the temporal whitening filters,

$$\begin{aligned}
&\min_{\mathbf{W},\mathbf{B}} \mathcal{I}_r(\mathbf{W},\mathbf{B}) \triangleq \frac{1}{T}\sum_{n=1}^{N}\sum_{\tau=1}^{T} H(\mathbf{b}_n^H \mathbf{X}_{\mathcal{C}}^{(\mathrm{K}+1)}(\tau)\mathbf{w}_n^*) - 2\log|\det(\mathbf{W})|, \qquad (7.18)\\
&\text{s.t. } \mathbf{w}_1^H\mathbf{w}_1 = \ldots = \mathbf{w}_N^H\mathbf{w}_N = 1,\\
&\quad \mathbf{b}_1^H\mathbf{D}_b\mathbf{b}_1 = \ldots = \mathbf{b}_N^H\mathbf{D}_b\mathbf{b}_N = 1,
\end{aligned}$$

where $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_N] \in \mathbb{C}^{2K+2\times N}$ is the temporal demixing matrix. Two auxiliary matrices $\widetilde{\mathbf{W}} \in \mathbb{C}^{N\times N}$ and $\widetilde{\mathbf{B}} \in \mathbb{C}^{2K+2\times N}$ are now introduced. Let $\overline{\mathbf{W}} = [\bar{\mathbf{w}}_1, \ldots, \bar{\mathbf{w}}_N]$, where $\bar{\mathbf{w}}_n = \tilde{\mathbf{w}}_n / \|\tilde{\mathbf{w}}_n\|$, be the column normalized matrix of $\widetilde{\mathbf{W}}$, and let $\overline{\mathbf{B}} = [\bar{\mathbf{b}}_1, \ldots, \bar{\mathbf{b}}_N]$, where $\bar{\mathbf{b}}_n = \tilde{\mathbf{b}}_n / \sqrt{\tilde{\mathbf{b}}_n^H \mathbf{D}_b \tilde{\mathbf{b}}_n}$, be the column normalized matrix of $\widetilde{\mathbf{B}}$. It can be seen that $\overline{\mathbf{W}}$ and $\overline{\mathbf{B}}$ are matrix functions of $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{B}}$, respectively. Similarly to CEBM+, we now have the following unconstrained cost function for CERBM

$$\min_{\widetilde{\mathbf{W}},\widetilde{\mathbf{B}}} \mathcal{I}_r(\overline{\mathbf{W}}(\widetilde{\mathbf{W}}), \overline{\mathbf{B}}(\widetilde{\mathbf{B}})), \qquad (7.19)$$

where the columns of $\overline{\mathbf{W}}$ and $\overline{\mathbf{B}}$ are always satisfied their corresponding constraints by definition.

Note that, the gradients of $\mathcal{I}_r$ w.r.t. $\bar{\mathbf{w}}_1^*, \ldots, \bar{\mathbf{w}}_N^*$ and $\bar{\mathbf{b}}_1^*, \ldots, \bar{\mathbf{b}}_N^*$ are readily available in [12]. The goal is to find the gradients w.r.t to the auxiliary variables.

By referencing to (6.16) and noting that $\bar{\mathbf{w}}_n = \widetilde{\mathbf{w}}_n / \|\widetilde{\mathbf{w}}_n\|$, we have

$$\frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \widetilde{\mathbf{w}}_n^*} = \frac{1}{\|\widetilde{\mathbf{w}}_n\|} \left(\frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{w}}_n^*} - \bar{\mathbf{w}}_n \Re\Big\{ \bar{\mathbf{w}}_n^H \frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{w}}_n^*} \Big\} \right). \tag{7.20}$$

Similarly, since $\bar{\mathbf{b}}_n = \widetilde{\mathbf{b}}_n / \sqrt{\widetilde{\mathbf{b}}_n^H \mathbf{D}_b \widetilde{\mathbf{b}}_n}$ and that $(\frac{\partial \mathcal{I}_r}{\bar{\mathbf{b}}_n})^* = \frac{\partial \mathcal{I}_r}{\partial \bar{\mathbf{b}}_n^*}$ due to $\mathcal{I}_r$ being real-valued, we obtain

$$\begin{aligned}
\frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \widetilde{\mathbf{b}}_n^*} =& \frac{\partial \bar{\mathbf{b}}_n^*}{\partial \widetilde{\mathbf{b}}_n^*} \frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{b}}_n^*} + \frac{\partial \bar{\mathbf{b}}_n}{\partial \widetilde{\mathbf{b}}_n^*} \frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{b}}_n} \\
=& \Big(\frac{\mathbf{I}}{\sqrt{\widetilde{\mathbf{b}}_n^H \mathbf{D}_b \widetilde{\mathbf{b}}_n}} - \frac{\mathbf{D}_b \widetilde{\mathbf{b}}_n \widetilde{\mathbf{b}}_n^H}{2(\widetilde{\mathbf{b}}_n^H \mathbf{D}_b \widetilde{\mathbf{b}}_n)^{3/2}} \Big) \frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{b}}_n^*} \\
& - \frac{\mathbf{D}_b \widetilde{\mathbf{b}}_n \widetilde{\mathbf{b}}_n^T}{2(\widetilde{\mathbf{b}}_n^H \mathbf{D}_b \widetilde{\mathbf{b}}_n)^{3/2}} \frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{b}}_n} \\
=& \frac{1}{\sqrt{\widetilde{\mathbf{b}}_n^H \mathbf{D}_b \widetilde{\mathbf{b}}_n}} \left(\frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{b}}_n^*} - \mathbf{D}_b \bar{\mathbf{b}}_n \Re\Big\{ \bar{\mathbf{b}}_n^H \frac{\partial \mathcal{I}_r(\overline{\mathbf{W}}, \overline{\mathbf{B}})}{\partial \bar{\mathbf{b}}_n^*} \Big\} \right).
\end{aligned} \tag{7.21}$$

Consider the real-valued parameterization of $\mathcal{I}_r$ given by

$$\boldsymbol{\Theta}_{\mathcal{R}} = [\Re\{\mathrm{vec}(\widetilde{\mathbf{W}})\}^T, \Im\{\mathrm{vec}(\widetilde{\mathbf{W}})\}^T, \Re\{\mathrm{vec}(\widetilde{\mathbf{B}})\}^T, \Im\{\mathrm{vec}(\widetilde{\mathbf{W}})\}^T]. \tag{7.22}$$

It can be seen that the mapping from $\widetilde{\mathbf{W}}$, $\widetilde{\mathbf{B}}$ to $\boldsymbol{\Theta}_{\mathcal{R}}$ is a one-to-one linear map. With reference to (2.3), the gradient of the entropy rate w.r.t. $\boldsymbol{\Theta}_{\mathcal{R}}$ is given by

$$\frac{\partial \mathcal{I}_r}{\partial \boldsymbol{\Theta}_{\mathcal{R}}} = 2 \left[\Re\Big\{ \mathrm{vec}\Big(\frac{\partial \mathcal{I}_r}{\partial \widetilde{\mathbf{W}}^*} \Big) \Big\}^T, \Im\Big\{ \mathrm{vec}\Big(\frac{\partial \mathcal{I}_r}{\partial \widetilde{\mathbf{W}}^*} \Big) \Big\}^T, \Re\Big\{ \mathrm{vec}\Big(\frac{\partial \mathcal{I}_r}{\partial \widetilde{\mathbf{B}}^*} \Big) \Big\}^T, \Im\Big\{ \mathrm{vec}\Big(\frac{\partial \mathcal{I}_r}{\partial \widetilde{\mathbf{B}}^*} \Big) \Big\}^T \right]^T, \tag{7.23}$$

where $\partial \mathcal{I}_r / \partial \widetilde{\mathbf{W}}^*$ and $\partial \mathcal{I}_r / \partial \widetilde{\mathbf{B}}^*$ can be obtained by evaluating (7.20) and (7.21) for all n, respectively. To this end, one can use any existing real-valued optimization method to optimize for $\boldsymbol{\Theta}_{\mathcal{R}}$. Similarly to CEBM+, the L-BFGS algorithm in the minFunc toolbox [128] is used to compute the optimal $\boldsymbol{\Theta}_{\mathcal{R}}$. After that, the optimal $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{B}}$ can be obtained by a simple manipulation. Table (7.2) shows

1. Find initial matrix $\widetilde{\mathbf{W}}$ using the conventional CEBM
2. Initialize $\widetilde{\mathbf{B}}$ where $\widetilde{\mathbf{b}}_n = [1, \underbrace{0, \ldots, 0}_{2K+1}]$, $\forall n$.
3. Solving (7.19) using L-BFGS, given initial matrix $\widetilde{\mathbf{W}}$ and $\widetilde{\mathbf{B}}$
4. $\mathbf{W} \leftarrow [\frac{\widetilde{\mathbf{w}}_1}{\|\widetilde{\mathbf{w}}_1\|}, \ldots, \frac{\widetilde{\mathbf{w}}_N}{\|\widetilde{\mathbf{w}}_N\|}]$

Table 7.2: Complex-valued ICA by entropy rate bound minimization based on L-BFGS.

the proposed CERBM algorithm based on L-BFGS (CERBM+). Note that, the details of L-BFGS [51, 124] is not discussed here since it is a well-known algorithm.

7.3 Simulations

7.3.1 Simulation results for synthetic data

The proposed CERBM+ algorithm is evaluated in comparison to the conventional CERBM, CEBM, and the CEBM+ algorithm proposed in previous chapter. Other well-known complex-valued algorithms such as cFastICA [7], noncircular cFastICA [21], joint approximate diagonalization of eigen-matrices (JADE) [146], CMN [10] are excluded since they have subpar performance compared to CEBM and CERBM. We set the maximum iteration of CERBM+ to 100 which is equal to that of CERBM. The order of the whitening filters is fixed at $K = 4$ for both CERBM+ and CERBM.

The normalized Amari index (NAI) is employed as our performance metrics. Since the mixing matrix is assumed to be invertible, low NAI does not only indicate good estimation of the mixing matrix, it also implies good source separation. All simulations were performed on a Windows 10 64-bit computer with i7-4770 processor and 8GB RAM running a MATLAB 2018b with parallel toolbox.

For each simulation, the average performance of 100 trials is reported. New

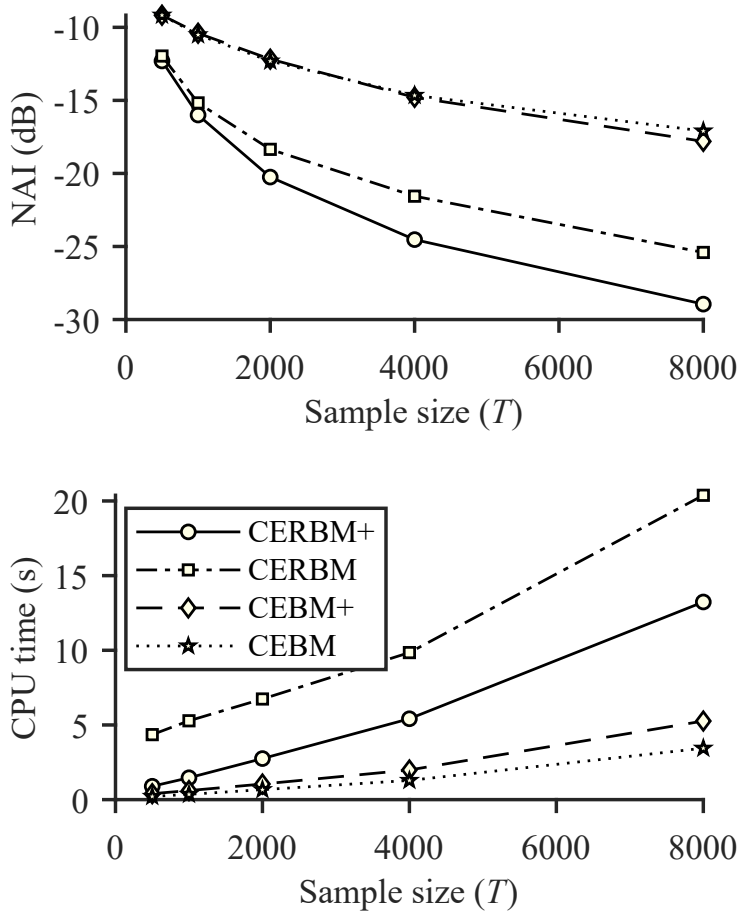


Fig. 7.1: Average NAI and CPU time as functions of the number of sources.

source signals and a new mixing matrix were created in each trial. Each element of the mixing matrix had its real part and imaginary part drawn from standard normal distribution. The sources were drawn from complex generalized Gaussian distribution [136],

$$f_S(s) \propto \exp(\rho s^2 + \rho s^{*2} - 2ss^*)^{p/2}, \tag{7.24}$$

where $p > 0$ is the shape parameter and $\rho \in [0, 1]$ is the circularity coefficient. For the shape parameter, $p = 2$ implies Gaussian distribution. For the circularity coefficient, $\rho = 0$ implies circular distribution, i.e., the real part and the imaginary part of s are independent. Additive white Gaussian noise was subsequently added to each source signal according to a preset signal noise ratio (SNR). Finally, the source signals were filtered via the following first-order

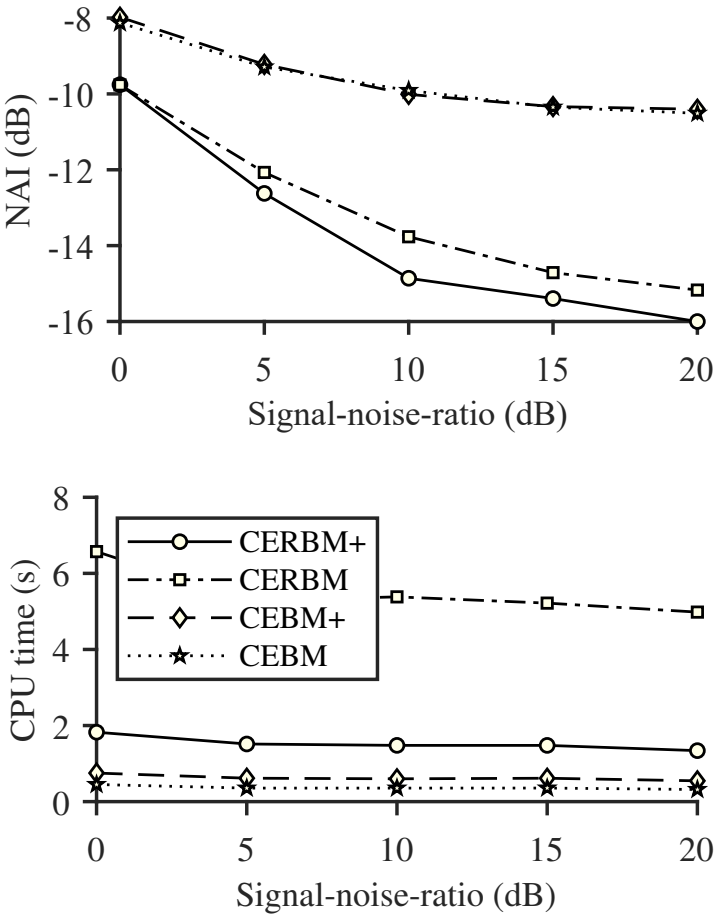


Fig. 7.2: Average NAI and CPU time as functions of the number of sources.

auto-regressive model

$$s^{(\text{color})}(\tau) = 0.9s^{(\text{color})}(\tau - 1) + s(\tau) \tag{7.25}$$

to simulate sample dependency.

Figure 7.1 shows the NAI and CPU time as functions of the sample size T for $N = 10$. Here, there are ten sources in each simulation with ρ being randomly selected from 0 to 1. The shape parameter p was varied uniformly from 1 to 3 so that there were five super-Gaussian sources and five sub-Gaussian sources. The SNR for each source was fixed at 20 dB. It can be seen that the performance of CEBM and CEBM+ are poor since they do not exploit the temporal dependency of the source signals. On the other hand, both CERBM and CERBM+ perform

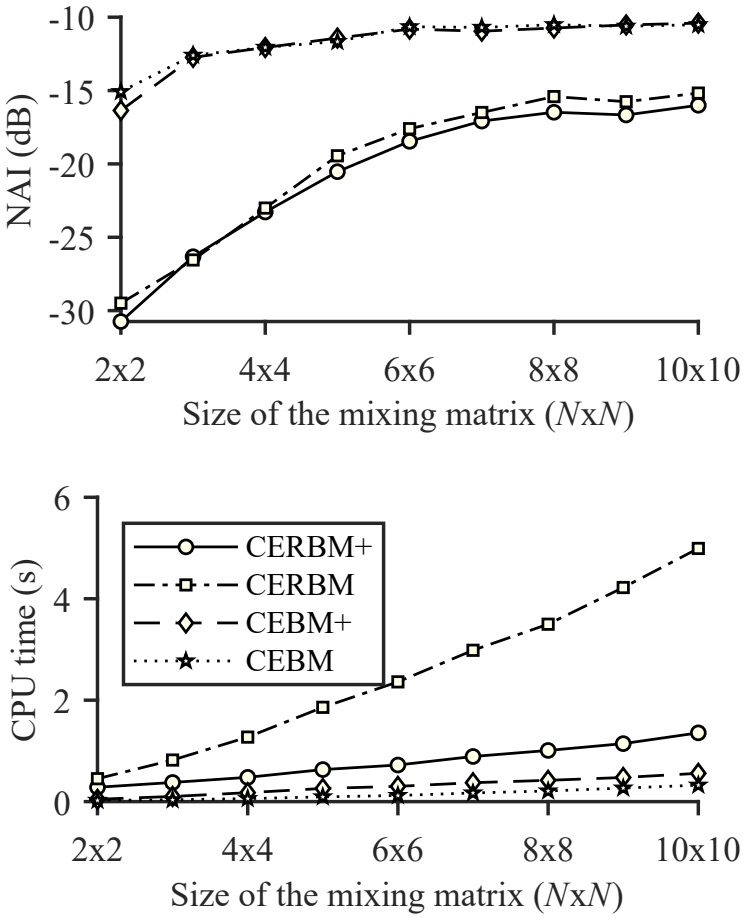


Fig. 7.3: Average NAI and CPU time as functions of the number of sources.

sufficiently well. Compared to CERBM, the proposed CERBM+ is not only faster but it can also achieve similar performance to CERBM using a much lower number of input samples.

The performance at $T = 1000$ in terms of NAI and computational time with respect to SNR is shown in Fig. 7.2. The parameters p, ρ, and N are chosen similarly to the previous simulation. The results in Fig. 7.2 indicate the superior performance of CERBM+ in comparison to other baseline algorithms.

Figure 7.3 depicts the performance with respect to the number of sources N at $T = 1000$. In this simulation, the shape parameters of the sources were uniformly space between $p = 1$ and $p = 3$. The SNR for each source was also fixed at 20 dB. The circular coefficient was randomly selected in $[0, 1]$ for each draw of a source. In this simulation, the performance of CERBM+ is modestly

higher than CERBM at $T = 1000$. However, one can observe that the proposed CERBM+ converges at a higher rate compared to CERBM.

From the above simulations, it can be seen that the proposed CERBM+ consistently outperforms CERBM in most tested cases while being less computational intensive. This implies that CERBM+ is a better alternative of CERBM for blind source separation by entropy rate bound minimization.

7.3.2 Simulation results for speech separation

The proposed CERBM+ in this chapter and the proposed CEBM+ in the previous chapter are utilized for blind speech separation. All processing steps follow the block diagrams in Figs. 3.2 and 3.3. First, the input mixtures were converted into the time-frequency domain using STFT. The demixing matrix at each frequency-bin was then estimated by CERBM+ or CEBM+ algorithm. Subsequently, the separated sources at each frequency-bin were computed using (7.2). The scaling ambiguity was then resolved using back-projection [67,78,108]. After that, similar to the method in Chapter 4, the permutation problem was resolved by utilizing [72] to align the soft-binary masks defined in (4.53) with the mixing matrix $\widehat{\mathbf{A}} = \mathbf{W}^{-H}$ and $\beta = 12.5$. The soft-binary masks were preferred over the amplitude envelopes as they yielded better permutation alignment results. Finally, the separated sources were transformed to the time domain using inverse STFT. Two baselines, whereby CERBM and CEBM were employed for mixing matrix estimation, were also created. Except for the mixing matrix estimation step, initial values of the mixing matrices and all processing steps of these two baselines were identical to CERBM+ and CEBM+. This is to ensure that the difference in separation performance is only attributed to the difference in mixing matrix estimation step.

For evaluation, a dataset of two-channel recordings was collected in three real outdoor, office, and lecture theater environments. All environments were

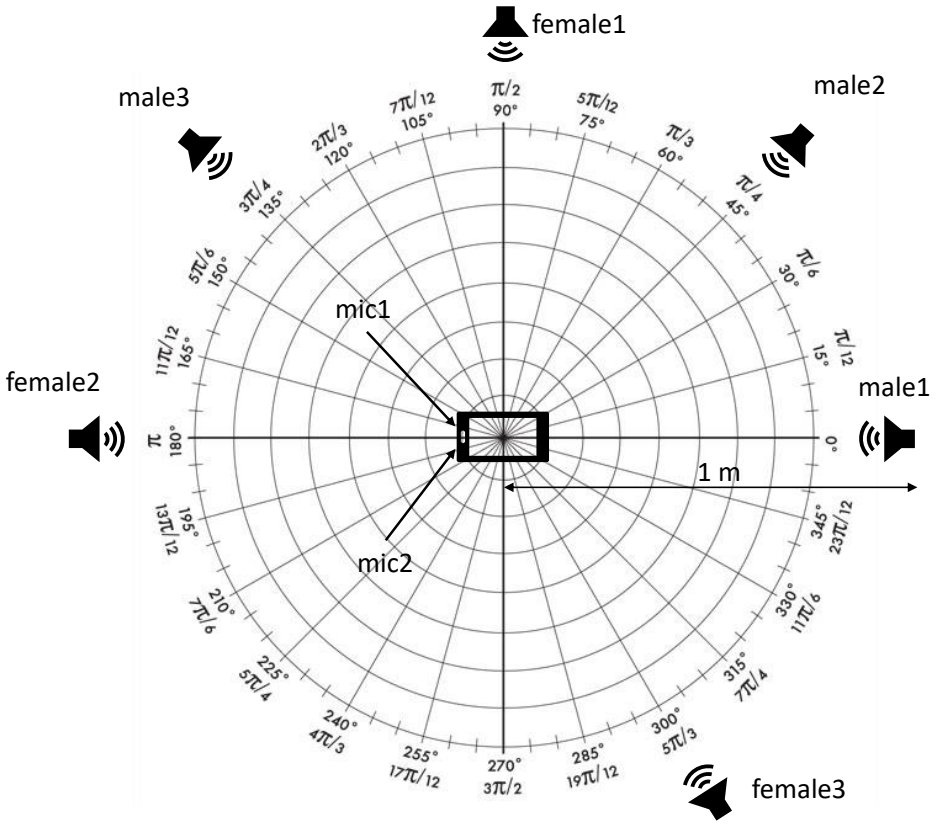


Fig. 7.4: Locations of speech sources and microphones.

within the campus of Nanyang Technological University. Outdoor recording was conducted within the university campus, office recording was carried out at a laboratory environment, and recording in lecture theater took place within the campus. Six speech sources of 10 seconds were prepared beforehand using speech signals from TIMIT dataset [147]. Three sources are females and three sources are males. For each environment, each source was played through a loudspeaker and was recorded using a Huawei Mate 10 Pro phone at sampling rate of 16000 Hz. This phone has two microphones and the spacing between two microphones is approximately 2.4 cm. The locations of the sources are shown in Fig. 7.4 where the distance between a source and the center of the mobile phone was 1 m. Finally, for each environment, fifteen two-channel mixtures of two convolutive sources were synthetically created by choosing a combination of two sources then adding their source images together.

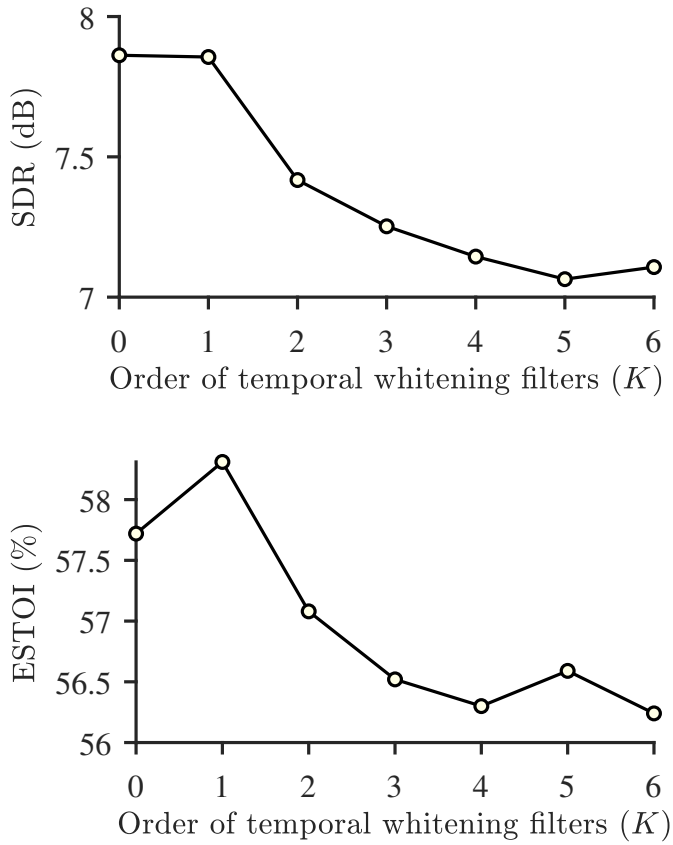


Fig. 7.5: Average SDR and ESTOI over all environments with respect to the order of the whitening filters K for blind separation of two speech sources from two convolutive mixtures.

SDR [148] and ESTOI [117] are employed as the performance metrics. The performance of two unsupervised BSS methods namely independence low-rank matrix analysis (ILRMA) [105] and OverIVA [149] are also included. For all evaluated algorithms, periodic Hamming window of 4096 samples and 75% overlap were used for both forward and inverse STFT.

Figure 7.5 illustrates the performance of CERBM+ with respect to the order of the whitening filters. Here, $K = 0$ corresponds to the performance of CEBM+. It can be seen that $K = 1$ yields the highest ESTOI while having almost the same SDR to that of $K = 0$. The peak performance at $K = 1$ can be attributed to the overlap between two consecutive STFT analysis windows. Higher values of K yield considerably lower performance because speech signals

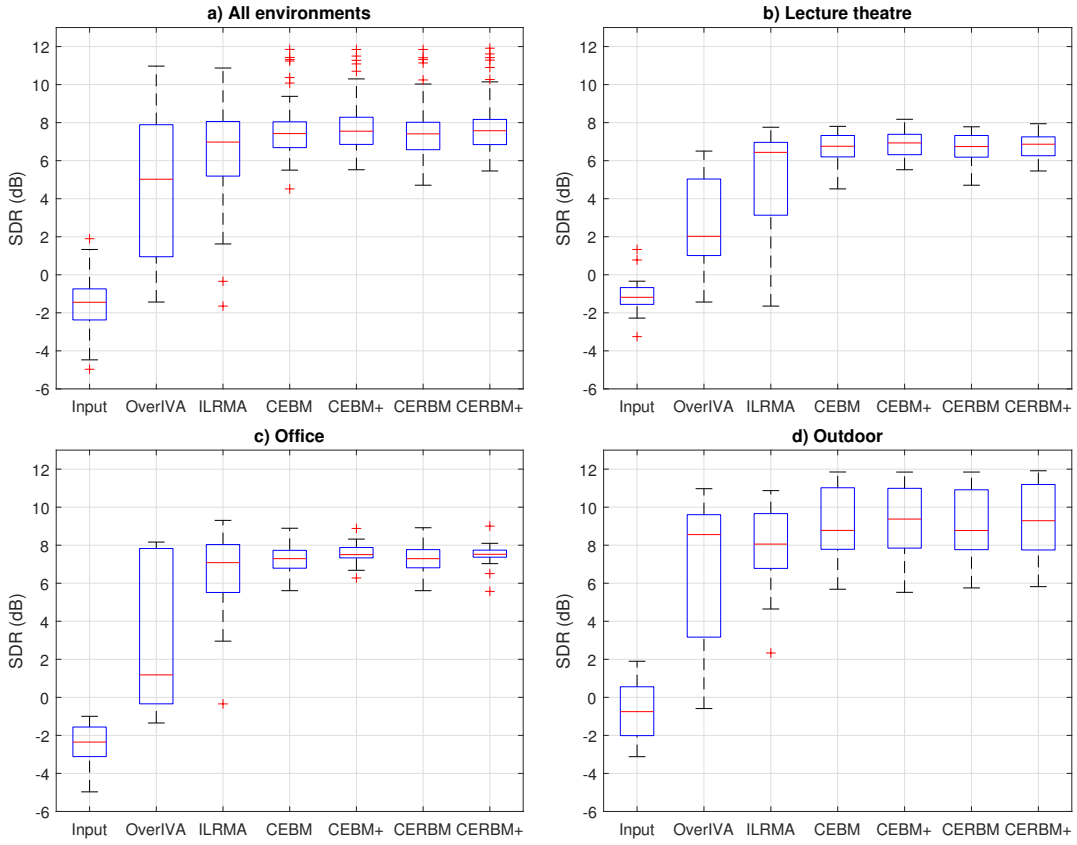


Fig. 7.6: Separation performance in SDR for blind separation of two speech sources from two convolutive mixtures in: a) All environments. b) Lecture theater. c) Office. d) Outdoor.

are usually not stationary.

Figures 7.6 and 7.7 show the performance in terms of SDR and ESTOI of the proposed methods in comparison with the baselines. Here, $K = 1$ was used for CERBM and CERBM+. Overall, it can be seen that the median separation performance is highest for outdoor environment, followed by office and lecture theater in that order. This is because there is more reverberation in office and lecture theater than outdoor environment. Furthermore, CERM, CERM+, CERBM, CERBM+ perform similarly and they are better than ILRMA and OverIVA. It can also be seen that the performance of CERM+ and CERBM+ are modestly higher in most cases than the baseline algorithms CERM and CERBM, respectively. Note that, the similarity in performance of CERM,

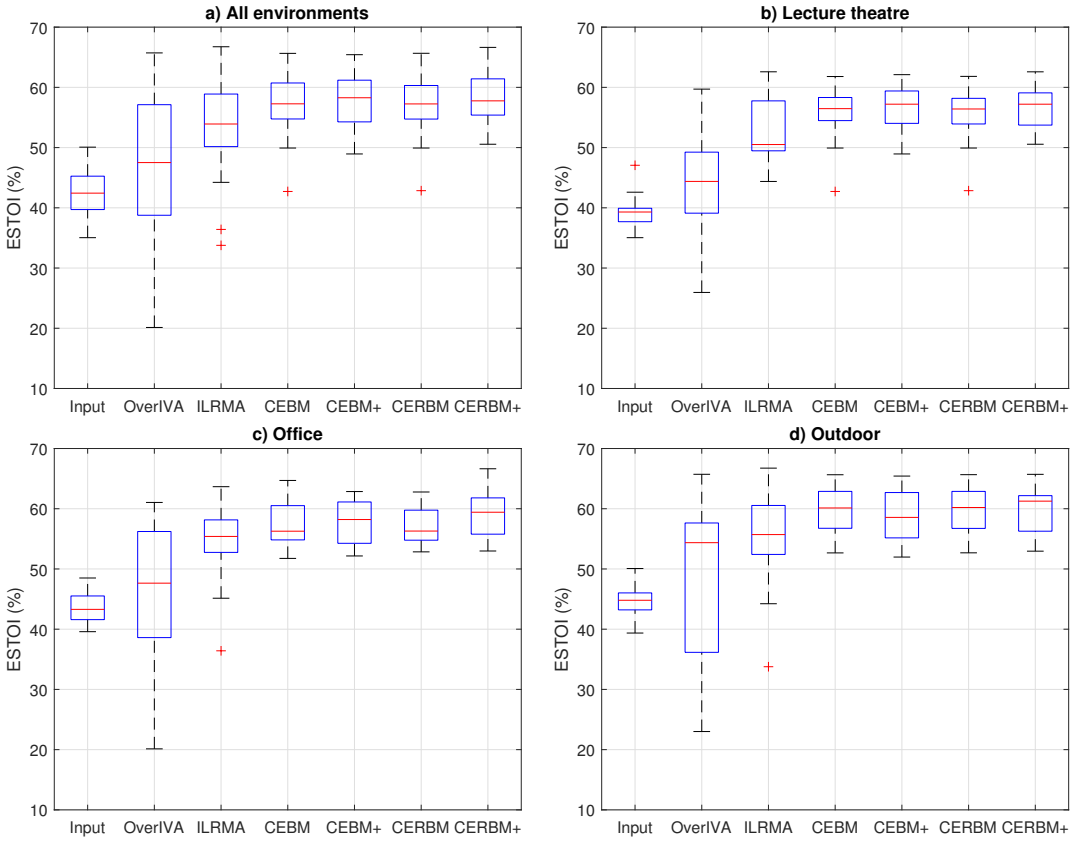


Fig. 7.7: Separation performance in ESTOI for blind separation of two speech sources from two convolutive mixtures in: a) All environments. b) Lecture theater. c) Office. d) Outdoor.

CERM+, CERBM, CERBM+ is consistent with the previous simulation results in Figs. 6.1, 6.3, and 7.1 where the sources are sparse and the number of samples is small.

7.4 Summary

In this chapter, for ICA by entropy rate bound minimization, the temporal whitening filters are directly optimized in parallel with the mixing matrix using auxiliary variables. Using the isomorphism between real-composite space and complex-augmented space, the entropy rate – a real-valued function of complex-valued arguments – is parameterized by a real-valued function that consists of

real-valued arguments. The resultant optimization problem is then solved by L-BFGS. Simulations on synthetic data show that such simple approach can lead to significantly better estimate of the mixing matrix while saving a substantial amount of computational time. Compared to the baseline algorithms, the proposed method shows a modest improvement in separation of real-recorded convolutive speech mixtures.

Chapter 8

Conclusion and Recommendations

8.1 Conclusion

In this thesis, several frequency-domain blind source algorithms have been reviewed. A frequency-domain BSS algorithm typically consists of bin-wise complex-valued mixing matrix estimation and bin-wise source separation followed by a permutation alignment step. Among these steps, complex-valued mixing matrix estimation is the most critical part since it strongly dictates the performance of every subsequent processing procedure. For mixing matrix estimation, two most common methods, ICA and SCA, are discussed.

The ICA methods search for a linear transformation such that the demixed signals are independent. Due to the central limit theorem, one can perform ICA by maximization of non-Gaussian since the mixed signal tends to be Gaussian distributed. Similarly, ICA can also be done by minimizing the mutual information of the source defined by the Kullback–Leibler divergence between the joint distribution and the factorized distribution. Nevertheless, ICA filters are often estimated separately using an approximation of the Hessian matrix. This approximation will be invalidated when the independence assumption of the sources is not exact.

The SCA methods assume that the sources are highly sparse to the point that most of the observed data concentrates about directions specified by the mixing matrix. By clustering based on direction, the mixing matrix can then be recovered. Directional clustering methods suffer from two main drawbacks: they are not completely suitable for sparse signals because each cluster is modeled by Gaussian functions, and that they are computationally expensive since the covariance matrix of each cluster must be computed for each iteration.

Chapter 4 focuses on the mixing matrix estimation for highly sparse source signals. The directional sparse filtering algorithm proposed in Chapter 4 employs a novel contrast function where clustering objective and sparse filtering objective are simultaneously achieved. The criteria for a generic contrast function to be successful in directional sparse filtering are also outlined. Simulation results show that directional sparse filtering can achieve better estimation than the baseline directional clustering methods.

For the SCA methods based on directional statistics, their core assumption is the concept of approximate sparseness. It is however unclear when the sources are sufficiently sparse so that the mixing matrix can be recovered. Furthermore, these SCA methods often employ multivariate non-separable contrast functions. This makes their analysis difficult. Chapter 5 extends the stability analysis of complex-valued ICA to include the case where the contrast function is multivariate and the observed data is normalized to unit norm. This stability condition can be verified for several directional clustering methods as well as the directional sparse filtering algorithm proposed in Chapter 4. The condition implies that one can use directional clustering to estimate the ICA filters when the sources are super-Gaussian distributed. This result shows the intertwine between ICA and SCA approaches.

Chapter 6 addresses the limitation of second-order ICA methods. By using auxiliary variables, the constrained ICA problems are converted into uncon-

strained ones. Therefore, these problems can be solved with off-the-shelf quasi-Newton method such as L-BFGS. As opposed to conventional methods to solve ICA objective function, the proposed approach does not restrict the Hessian to have any specific form. Simulations indicate that the proposed ICA method can greatly improve the estimation performance of the baseline ICA algorithm, particularly when the source distribution tends to Gaussian distribution.

Chapter 7 further develops the idea presented in Chapter 6 to jointly estimate the mixing matrix and the temporal whitening filters. It is seen from simulation results that this approach can greatly enhance the performance of ICA by entropy rate bound minimization.

8.2 Recommendations for further research

There are a few suggestions for future research:

1. **Sparse Vector Analysis for under-determined BSS:** In determined BSS, independent vector analysis (IVA) has been proposed to avoid the permutation problem. However, IVA methods rely on probabilistic models which are intractable in under-determined scenario. Similar to IVA, one can potentially solve the source separation problem and permutation problem jointly by exploiting the correlation of the source activity among frequency bins as well as the sparseness of the source across time dimension. As opposed to IVA, sparse vector analysis does not require the mixing matrix at each frequency to be invertible and therefore it may have more far-reaching applications.

2. **Robust BSS for asynchronous recording:** In practice, one may have multiple-device recordings of a same conversation. Due to the difference in starting offset and sampling-rate mismatches, BSS methods such as those discussed in this thesis are not applicable. The common approach to separate asyn-

chronous recordings is to perform offset and sampling rate compensation followed by a conventional BSS algorithm. If the BSS algorithm is robust to sampling rate mismatch among different devices, the costly compensation of sampling rate can be eliminated. One simple way to achieve this goal is to determine a distance metric that is invariant to sampling mismatch among input devices instead of the cosine similarity that has been used in directional clustering and directional sparse filtering. Another idea is to estimate both the sampling rate mismatch and the source simultaneously.

3. **Jointly learn spatial filters and spectral filters using spatially-aware neural network:** Blind source separation methods, whether relying on independence or sparsity, essentially estimate the spatial filters modeling the propagation channels from the sources to the sensors. Spectral information of the sources is often overlooked in unsupervised BSS methods such as ICA and SCA. In recent years, there have been multiple works to combine the spatial approaches with deep learning [150–152]. However, the current combination approaches are limited to the use of unsupervised BSS algorithm to guide the deep-learning-based separation networks, e.g., by generating the noisy target for supervised algorithms. Thus, the spatial information is not fully exploited. Directly incorporating the sparsity assumption or independence assumption into the neural network to help the network jointly learn both spatial information and spectral information can be worth investigating.

Author's Publications

[P1] **A. H. T. Nguyen**, V. G. Reju, and A. W. H. Khong, *"Directional Sparse Filtering for Blind Estimation of Under-determined Complex-valued Mixing Matrices,"* IEEE Transactions on Signal Processing, vol. 68, pp. 1990-2003, Mar. 2020.

[P2] Z. Tan, **A. H. T. Nguyen**, and A. W. H. Khong, *"An Efficient Dilated Convolutional Neural Network for UAV Noise Reduction at Low Input SNR,"* in Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 2019, pp. 1885-1892.

[P3] **A. H. T. Nguyen**, V. G. Reju, and A. W. H. Khong, *"A Method Based on L-BFGS to Solve Constrained Complex-valued ICA,"* in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2019, pp. 4370–4374.

[P4] V. G. Reju, R. S. Rashobh, **A. H. T. Nguyen**, and A. W. H. Khong, *"An Efficient Multi-Source DOA Estimation Algorithm for Underdetermined System,"* in Proceedings of International Workshop on Acoustic Signal Enhancement, 2018, pp. 86–90.

[P5] **A. H. T. Nguyen**, V. G. Reju, A. W. H. Khong, and I. Y. Soon, *"Learning complex-valued latent filters with absolute cosine similarity,"* in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2017, pp. 2412–2416.

Bibliography

[1] P. A. Rodriguez, M. Anderson, V. D. Calhoun, and T. Adali, "General nonunitary constrained ica and its application to complex-valued fmri data," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 922–929, Mar. 2015.

[2] T. Adali, M. Anderson, and G. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 18–33, May 2014.

[3] J. Eriksson, A.-M. Seppola, and V. Koivunen, "Complex ica for circular and non-circular sources," in *methods*, vol. 7, 2005, pp. 1–4.

[4] N. Mitianoudis and M. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 11, no. 5, pp. 489–497, 2003.

[5] T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind separation for convolutive mixtures of speech based on multistage ica using subarray processing," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 1, 2004, pp. I–225.

[6] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks,*, vol. 10, no. 3, pp. 626–634, May 1999.

[7] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 01, pp. 1–8, Feb. 2000.

[8] T. Adali, T. Kim, and V. Calhoun, "Independent component analysis by complex nonlinearities," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 5, 2004, pp. V–525–8.

[9] T. Adali, H. Li, M. Novey, and J.-F. Cardoso, "Complex ica using nonlinear functions," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4536–4544, Sep. 2008.

[10] M. Novey and T. Adali, "Complex ica by negentropy maximization," *IEEE Trans. Neural Networks,*, vol. 19, no. 4, pp. 596–609, Apr. 2008.

[11] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adali, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, Sep. 2014.

[12] G.-S. Fu, R. Phlypo, M. Anderson, and T. Adali, "Complex independent component analysis using three types of diversity: Non-gaussianity, non-whiteness, and noncircularity," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 794–805, Feb. 2015.

[13] P. D. O'Grady and B. A. Pearlmutter, "Hard-lost: Modified k-means for oriented lines," in *Proc. Irish Signals Syst. Conf.*, 2004, pp. 247–252.

[14] ——, "Soft-lost: Em on a mixture of oriented lines," in *Proc. Int. Conf. Indep. Compon. Anal. Signal Sep.* Springer, 2004, pp. 430–436.

[15] Z. He, A. Cichocki, Y. Li, S. Xie, and S. Sanei, "K-hyperline clustering learning for sparse component analysis," *Signal Processing*, vol. 89, no. 6, pp. 1011–1022, Jun. 2009.

[16] D. H. T. Vu and R. Haeb-Umbach, "Blind speech separation employing directional statistics in an expectation maximization framework," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2010, pp. 241–244.

[17] H. Sawada, S. Araki, and S. Makino, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 19, no. 3, pp. 516–527, Jan. 2011.

[18] A. Asaei, M. J. Taghizadeh, S. Haghighatshoar, B. Raj, H. Bourlard, and V. Cevher, "Binary sparse coding of convolutive mixtures for sound localization and separation via spatialization," *IEEE Trans. Signal Process.*, vol. 64, no. 3, pp. 567–579, Feb. 2016.

[19] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Underdetermined blind separation for speech in real environments with sparseness and ica," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2004, pp. iii–881.

[20] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors," *Signal Processing*, vol. 87, no. 8, pp. 1833–1847, Aug. 2007.

[21] M. Novey and T. Adali, "On extending the complex fastica algorithm to noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2148–2154, May 2008.

[22] X.-L. Li and T. Adali, "Complex independent component analysis by entropy bound minimization," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 57, no. 7, pp. 1417–1430, Jul. 2010.

[23] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411–430, Jun. 2000.

[24] H. Chen, D. Yao, and L. Zeng, "A bfgs-ica algorithm and application in localization of brain activities," *Neurocomputing*, vol. 64, pp. 513–519, 2005.

[25] P. Ablin, J. Cardoso, and A. Gramfort, "Faster independent component analysis by preconditioning with hessian approximations," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4040–4049, Aug. 2018.

[26] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster ica under orthogonal constraint," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 4464–4468.

[27] L. Albera, A. Ferréol, P. Comon, P. Chevalier, F. Anne, P. Comon, P. Chevalier, L. Albera, F. Anne, P. Comon, P. Chevalier, B. Identification, L. Albera, and A. Ferr, "Blind Identification of Overcomplete MixturEs of sources (BIOME)," *Linear Algebra Appl.*, vol. 391, pp. 3–30, 2004.

[28] A. Ferréol, L. Albera, and P. Chevalier, "Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM)," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1640–1653, May 2005.

[29] L. De Lathauwer, J. Castaing, and J. F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 55, no. 6 II, pp. 2965–2973, Jun. 2007.

[30] L. De Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1096–1105, Mar. 2008.

[31] A. Karfoul, L. Albera, and G. Birot, "Blind underdetermined mixture identification by joint canonical decomposition of HO cumulants," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 638–649, Feb. 2010.

[32] X. Luciani, A. L. F. de Almeida, and P. Comon, "Blind identification of underdetermined mixtures based on the characteristic function: The complex case," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 540–553, Feb. 2011.

[33] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, Nov. 2001.

[34] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," in *Proc. SPIE 44th Annu. Meet.*, vol. 3813, 1999, pp. 188–195.

[35] S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 1, 2002, pp. 529–532.

[36] F. Georgiev, F. Theis, and A. Cichocki, "Blind source separation and sparse component analysis of overcomplete mixtures," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 5, 2004, pp. V–493–6 vol. 5.

[37] E. Vincent, "Complex nonconvex lp norm minimization for underdetermined source separation," in *Proc. Int. Conf. Indep. Compon. Anal. Signal Sep.*, 2007, pp. 430–437.

[38] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Blind sparse source separation with spatially smoothed time-frequency masking," in *Proc. Int. Work. Acoust. Echo Noise Control*, 2006.

[39] V. G. Reju, S. N. Koh, and I. Y. Soon, "Underdetermined convolutive blind source separation via time-frequency masking," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 18, no. 1, pp. 101–116, Jan. 2010.

[40] B. Liu, V. G. Reju, and A. W. H. Khong, "A linear source recovery method for underdetermined mixtures of uncorrelated ar-model signals without sparseness," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 4947–4958, Oct. 2014.

[41] L. Drude, C. Boeddeker, and R. Haeb-Umbach, "Blind speech separation based on complex spherical k-mode clustering," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2016, pp. 141–145.

[42] C. F. J. Wu, "On the convergence properties of the em algorithm," *Ann. Stat.*, pp. 95–103, Mar. 1983.

[43] F. M. Zennaro and K. Chen, "Towards understanding sparse filtering: A theoretical perspective," *Neural Networks*, vol. 98, pp. 154–177, Feb. 2018.

[44] J. Ngiam, Z. Chen, S. A. Bhaskar, P. W. Koh, and A. Y. Ng, "Sparse filtering," in *Proc. Adv. neural Inf. Process. Syst.*, 2011, pp. 1125–1133.

[45] A. Vinnikov and S. Shalev-Shwartz, "K-means recovers ica filters when independent components are sparse," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 712–720.

[46] C. Pehlevan and D. B. Chklovskii, "A hebbian/anti-hebbian network derived from online non-negative matrix factorization can cluster and discover sparse features," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, 2014, pp. 769–775.

[47] K. Kreutz-delgado and B. D. Rao, "A general approach to sparse basis selection : Majorization , concavity , and affine scaling," San Diego, Tech. Rep., 1997.

[48] K. Kreutz-Delgado, B. D. Rao, K. Engan, T.-W. Lee, and T. Sejnowski, "Convex/schur-convex (csc) log-priors and sparse coding," in *Proc. Jt. Symp. Neural Comput.*, 1999, pp. 65–71.

[49] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T.-W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, no. 2, pp. 349–396, 2003.

[50] E. Kreyszig, *Advanced engineering mathematics*, 10th ed. John Wiley & Sons, 2010.

[51] L. Sorber, M. V. Barel, and L. D. Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM J. Optim.*, vol. 22, no. 3, pp. 879–898, Jul. 2012.

[52] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[53] T. Adali and P. J. Schreier, "Optimization and estimation of complex-valued signals: Theory and applications in filtering and blind source separation," *IEEE Signal Process. Mag.*, vol. 5, no. 31, pp. 112–128, Sep. 2014.

[54] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[55] A. Hjørungnes, *Complex-valued matrix derivatives: with applications in signal processing and communications*. Cambridge University Press, 2011.

[56] A. W. Marshall, I. Olkin, and B. Arnold, *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

[57] A. W. Marshall and I. Olkin, "Majorization in multivariate distributions," *Ann. Stat.*, pp. 1189–1200, 1974.

[58] P. D. O'Grady, B. A. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," *Int. J. Imaging Syst. Technol.*, vol. 15, no. 1, pp. 18–33, Jan. 2005.

[59] S. Choi, A. Cichocki, H.-M. Park, and S.-Y. Lee, "Blind source separation and independent component analysis: A review," *Neural Inf. Process. Rev.*, vol. 6, no. 1, pp. 1–57, Jan. 2005.

[60] Q. V. Le, A. Karpenko, J. Ngiam, and A. Y. Ng, "Ica with reconstruction cost for efficient overcomplete feature learning," in *Adv. Neural Inf. Process. Syst.*, 2011, pp. 1017–1025.

[61] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.* IEEE, 2013, pp. 8595–8598.

[62] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1, pp. 21–34, Nov. 1998.

[63] M. Joho and H. Mathis, "Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture," in *Indep. Compon. Anal. Blind Signal Sep.*, 2000, pp. 81–86.

[64] L. Zhang, A. Cichocki, and S.-i. Amari, "Multichannel blind deconvolution of nonminimum-phase systems using filter decomposition," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1430–1442, Apr. 2004.

[65] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-

order statistics," *IEEE Trans. speech audio Process.*, vol. 13, no. 1, pp. 120–134, Dec. 2004.

[66] S. C. Douglas, M. Gupta, H. Sawada, and S. Makino, "Spatio–temporal fastica algorithms for the blind separation of convolutive mixtures," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 15, no. 5, pp. 1511–1520, 2007.

[67] S. Ikeda and N. Murata, "An approach to blind source separation of speech signals," in *Proc. ICANN*, 1998, pp. 761–766.

[68] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Processing*, vol. 85, no. 7, pp. 1389–1403, Jul. 2005.

[69] J. Cho, J. Choi, and C. D. Yoo, "Underdetermined convolutive blind source separation using a novel mixing matrix estimation and mmse-based source estimation," in *Proc. IEEE Int. Work. Mach. Learn. Signal Process.*, 2011, pp. 1–6.

[70] J. Cho and C. D. Yoo, "Underdetermined convolutive bss: Bayes risk minimization based on a mixture of super-gaussian posterior approximation," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 23, no. 5, pp. 828–839, May 2015.

[71] F. Nesta, P. Svaizer, and M. Omologo, "Convolutive bss of short mixtures by ica recursively regularized across frequencies," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 19, no. 3, pp. 624–639, Mar. 2011.

[72] L. Wang, "Multi-band multi-centroid clustering based permutation alignment for frequency-domain blind speech separation," *Digit. Signal Process.*, vol. 31, pp. 79–92, Aug. 2014.

[73] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Work. Appl. Signal Process. to Audio Acoust.* IEEE, 2011, pp. 189–192.

[74] H. Shen and M. Kleinsteuber, "A matrix joint diagonalization approach for complex independent vector analysis," in *Latent Var. Anal. Signal Sep.* Springer Berlin Heidelberg, 2012, vol. 7191, ch. 9, pp. 66–73.

[75] N. Ito, S. Araki, and T. Nakatani, "Permutation-free convolutive blind source separation via full-band clustering based on frequency-independent source presence priors," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2013, pp. 3238–3242.

[76] N. Ito, S. Araki, T. Yoshioka, and T. Nakatani, "Relaxed disjointness based clustering for joint blind source separation and dereverberation," in *Proc. Int. Work. Acoust. Signal Enhanc.*, 2014, pp. 268–272.

[77] K. Matsuoka, "Minimal distortion principle for blind source separation," in *Proc. 41st SICE Annu. Conf.*, vol. 4, 2002, pp. 2138–2143.

[78] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1, pp. 1–24, 2001.

[79] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," *Multichannel Speech Process. Handb.*, pp. 1065–1084, 2007.

[80] L. Hualiang and T. Adali, "A class of complex ica algorithms based on the kurtosis cost function," *IEEE Trans. Neural Networks,*, vol. 19, no. 3, pp. 408–420, 2008.

[81] Q. Su, Y. Wei, K. Xu, C. Deng, H. Li, and Y. Shen, "Using Derivatives of Second Generating Function for Underdetermined Blind Identification," *Circuits, Syst. Signal Process.*, vol. 39, no. 9, pp. 4578–4595, 2020.

[82] T. Kanamori and A. Takeda, "Non-convex optimization on stiefel manifold and applications to machine learning," in *Proc. Int. Conf. Neural Inf. Process.* Springer, 2012, pp. 109–116.

[83] M. Zibulevsky, P. Kisilev, Y. Y. Zeevi, and B. A. Pearlmutter, "Blind source separation via multinode sparse representation," *Adv. Neural Inf. Process. Syst.*, pp. 1049–1056, 2002.

[84] P. Georgiev and A. Cichocki, "Sparse component analysis by improved basis pursuit method," *Neural Networks,*, pp. 28–35, 2002.

[85] R. Balan and J. Rosca, "Statistical properties of stft ratios for two channel systems and applications to blind source separation," *Proc. Int. Work. Indep. Compon. Anal.*, vol. 1, no. X2, p. X2, 2000.

[86] S. Rickard, R. Balan, and J. Rosca, "Real-time time-frequency based blind source separation," in *Proc. Int. Work. Indep. Compon. Anal.*, 2001, pp. 651–656.

[87] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, 2004.

[88] P. Georgiev, F. Theis, A. Cichocki, and H. Bakardjian, "Sparse component analysis: a new tool for data mining," in *Data Min. Biomed.* Springer, 2007, pp. 91–116.

[89] Y. Li, A. Cichocki, and S.-I. Amari, "Sparse component analysis for blind source separation with less sensors than sources." Citeseer, 2003.

[90] V. G. Reju, S. N. Koh, and I. Y. Soon, "An algorithm for mixing matrix estimation in instantaneous blind source separation," *Signal Processing*, vol. 89, no. 9, pp. 1762–1773, Sep. 2009.

[91] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "Optimality and stability of the k-hyperline clustering algorithm," *Pattern Recognit. Lett.*, vol. 32, no. 9, pp. 1299–1304, 2011.

[92] N. Ito, S. Araki, and T. Nakatani, "Complex angular central gaussian mixture model for directional statistics in mask-based microphone array signal processing," in *Proc. Eur. Signal Process. Conf.*, pp. 1153–1157.

[93] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011.

[94] G. M. Del Corso, "Estimating an eigenvector by the power method with a random start," *SIAM J. Matrix Anal. Appl.*, vol. 18, no. 4, pp. 913–937, 1997.

[95] L. Drude and R. Haeb-Umbach, "Integration of Neural Networks and Probabilistic Spatial Models for Acoustic Blind Source Separation," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 4, pp. 815–826, Aug. 2019.

[96] J. Yang, Y. Guo, Z. Yang, and S. Xie, "Under-Determined Convolutive Blind Source Separation Combining Density-Based Clustering and Sparse Reconstruction in Time-Frequency Domain," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 66, no. 8, pp. 3015–3027, Aug. 2019.

[97] A. Ozerov, Fe, X, and C. Votte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 18, no. 3, pp. 550–563, 2010.

[98] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 21, no. 5, pp. 971–982, May 2013.

[99] S. Arberet, A. Ozerov, N. Q. K. Duong, E. Vincent, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Nonnegative matrix factorization and spatial covariance model for under-determined reverberant audio source separation," in *Inf. Sci. Signal Process. their Appl.* IEEE, 2010, pp. 1–4.

[100] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 18, no. 7, pp. 1830–1840, 2010.

[101] ——, "Spatial location priors for gaussian model based reverberant audio source separation," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, pp. 1–11, 2013.

[102] Y. Xie, K. Xie, J. Yang, Z. Wu, and S. Xie, "Underdetermined Reverberant Audio-Source Separation Through Improved Expectation–Maximization Algorithm," *Circuits, Syst. Signal Process.*, vol. 38, no. 6, pp. 2877–2889, Jan. 2019.

[103] Y. Xie, K. Xie, J. Yang, and S. Xie, "Underdetermined blind source separation combining tensor decomposition and nonnegative matrix factorization," *Symmetry (Basel).*, vol. 10, no. 10, p. 521, Oct. 2018.

[104] Y. Xie, K. Xie, and S. Xie, "Underdetermined convolutive blind separation of sources integrating tensor factorization and expectation maximization," *Digit. Signal Process.*, vol. 87, pp. 145–154, Feb. 2019.

[105] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined Blind Source Separation Unifying Independent Vector Analysis and Nonnegative Matrix Factorization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, Sep. 2016.

[106] K. Sekiguchi, Y. Bando, K. Yoshii, and T. Kawahara, "Bayesian Multichannel Speech Enhancement with a Deep Speech Prior," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2018, pp. 1233–1239.

[107] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, "Fast Multichannel Source Separation Based on Jointly Diagonalizable Spatial Covariance Matrices," in *Proc. Eur. Signal Process. Conf.*, 2019, pp. 1–5.

[108] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. ICA*, 1999.

[109] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, no. 4, pp. 231–251, 1995.

[110] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[111] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[112] K. Xie, Z. He, and A. Cichocki, "Convergence Analysis of the FOCUSS Algorithm," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 26, no. 3, pp. 601–613, Mar. 2015.

[113] I. Jafari, S. Haque, R. Togneri, and S. Nordholm, "Evaluations on under-determined blind source separation in adverse environments using time-frequency masking," *EURASIP J. Adv. Signal Process.*, vol. 2013, no. 1, p. 162, 2013.

[114] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inf. Process. Syst.*, pp. 757–763, 1996.

[115] E. Vincent, S. Araki, and P. Bofill, "The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation," in *Proc. Int. Conf. Indep. Compon. Anal. Signal Sep.*, 2009, pp. 734–741.

[116] E. Vincent, S. Araki, F. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, B. V. Gowreesunker, D. Lutter, and N. Q. K. Duong, "The signal separation evaluation campaign (2007-2010): Achievements and remaining challenges," *Signal Process*, vol. 92, pp. 1928–1936, 2012.

[117] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 11, pp. 2009–2022, Aug. 2016.

[118] C. Wang and X. S. Wang, "Indexing very high-dimensional sparse and quasi-sparse vectors for similarity searches," *VLDB Journal—The Int. J. Very Large Data Bases*, vol. 9, no. 4, pp. 344–361, 2001.

[119] P. S. Bullen, *Handbook of means and their inequalities.* Springer Science & Business Media, 2013, vol. 560.

[120] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.

[121] T. E. Abrudan, J. Eriksson, and V. Koivunen, "Steepest descent algorithms for optimization under unitary matrix constraint," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1134–1147, Mar. 2008.

[122] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[123] Y. E. Nesterov, "A method for solving the convex programming problem with convergence rate o (1/kˆ2)," *Dokl. akad. Nauk Sssr*, vol. 269, pp. 543–547, 1983.

[124] D. C. Liu and J. Nocedal, "On the limited memory bfgs method for large scale optimization," *Math. Program.*, vol. 45, no. 1-3, pp. 503–528, Aug. 1989.

[125] K. Benidis, Y. Sun, P. Babu, and D. P. Palomar, "Orthogonal sparse eigenvectors: A procrustes problem," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2016, pp. 4683–4686.

[126] M. Arjovsky, A. Shah, and Y. Bengio, "Unitary evolution recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 1120–1128.

[127] A. H. T. Nguyen, V. G. Reju, A. W. H. Khong, and I. Y. Soon, "Learning complex-valued latent filters with absolute cosine similarity," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2017, pp. 2412–2416.

[128] M. Schmidt, "Minfunc: Unconstrained differentiable multivariate optimization in matlab," 2005. [Online]. Available: http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html

[129] D. N. Parikh, M. Z. Ikram, and D. V. Anderson, "Implementation of blind source separation and a post-processing algorithm for noise suppression in

cell-phone applications," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2010, pp. 1634–1637.

[130] N. Ono, "Blind source separation on iphone in real environment," in *Proc. Eur. Signal Process. Conf.*, 2013, pp. 1–5.

[131] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multi-microphone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.

[132] A. P. Klapuri, "Multipitch estimation and sound separation by the spectral smoothness principle," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 5, 2001, pp. 3381–3384.

[133] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 signal separation evaluation campaign (sisec2011): Audio source separation," in *Proc. Int. Conf. Latent Var. Anal. Signal Sep.* Springer, 2012, pp. 414–422.

[134] N. Ono, Z. Koldovsky, S. Miyabe, and N. Ito, "The 2013 signal separation evaluation campaign," in *Proc. IEEE Int. Work. Mach. Learn. Signal Process.*, 2013, pp. 1–6.

[135] F. Nesta and M. Omologo, "Convolutive underdetermined source separation through weighted interleaved ica and spatio-temporal source correlation," in *Proc. Int. Conf. Latent Var. Anal. Signal Sep.* Springer, 2012, pp. 222–230.

[136] M. Novey, T. Adali, and A. Roy, "A complex generalized gaussian distribution - characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.

[137] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized algorithms for analysis and control of uncertain systems: with applications.* Springer Science & Business Media, 2012.

[138] G. Calafiore, F. Dabbene, and R. Tempo, "Radial and uniform distributions in vector and matrix spaces for probabilistic robustness," in *Top. Control its Appl.* Springer, 1999, pp. 17–31.

[139] F. H. Sinz, "Lp-nested symmetric distributions," *J. Mach. Learn. Res.*, vol. 11, pp. 3409–3451, Dec. 2010.

[140] D. Song and A. Gupta, "Lp-norm uniform distribution," *Proc. Am. Math. Soc.*, vol. 125, no. 2, pp. 595–601, 1997.

[141] L. Devroye, "Sample-based non-uniform random variate generation," in *Proc. ACM Conf. Winter Simul.*, 1986, pp. 260–265.

[142] F. Sinz, S. Gerwinn, and M. Bethge, "Characterization of the p-generalized normal distribution," *J. Multivar. Anal.*, vol. 100, no. 5, pp. 817–820, May 2009.

[143] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, 2007.

[144] X.-L. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, 2010.

[145] ——, "Blind separation of noncircular correlated sources using gaussian entropy rate," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2969–2975, Feb. 2011.

[146] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE Proc. Radar Signal Process.*, vol. 140, no. 6, Dec. 1993, pp. 362–370.

[147] J. S. Garofolo, "Timit acoustic phonetic continuous speech corpus," *Linguistic Data Consortium*, 1993.

[148] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[149] R. Scheibler and N. Ono, "Independent Vector Analysis with More Microphones Than Sources," in *Proc. IEEE Work. Appl. Signal Process. to Audio Acoust.*, 2019, pp. 185–189.

[150] Z. Q. Wang and D. L. Wang, "Combining spectral and spatial features for deep learning based blind speaker separation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 2, pp. 457–468, 2019.

[151] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, May 2018.

[152] Z. Chen, T. Yoshioka, X. Xiao, L. Li, M. L. Seltzer, and Y. Gong, "Efficient integration of fixed beamformers and speech separation networks for multi-channel far-field speech separation," in *Proc. EEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 5384–5388.