AN ALGORITHMIC FRAMEWORK FOR INSTANTANEOUS AND CONVOLUTIVE BOUNDED COMPONENT ANALYSIS

by

Hüseyin Atahan İnan

A Thesis Submitted to the

Graduate School of Engineering

in Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Electrical & Electronics Engineering

Koç University

June, 2014

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Hüseyin Atahan İnan

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

Assoc. Prof. Alper Tunga Erdoğan

Prof. Alper Demir

Assoc. Prof. Emre Alper Yıldırım

Date: ____________________

To my family...

Acknowledgments

First of all, I would like to express my deepest appreciation and thanks to my supervisor Assoc. Prof. Alper Tunga Erdoğan for all of his guidance and encouragement throughout this process. I can honestly say that due to my supervisor, I have always enjoyed my studies and been very motivated since the first day I have started my M.Sc. degree. I cannot express how lucky I feel I am to have Assoc. Prof. Alper Tunga Erdoğan as a supervisor and it has been priceless to work with him. I would also like to thank my committee members, Prof. Alper Demir, and Assoc. Prof. Emre Alper Yıldırım for their encouragement, insightful comments, and serving as my committee members.

I would like to thank to Koç University, TUBITAK and TURKCELL for providing me with financial support which made this study possible. This thesis is supported in part by TUBITAK 112E057 Project.

I would also like to thank all of my friends who supported me and incented me to strive towards my goal. Finally, I express my gratitude to my family for their lifetime support and encouragement.

ABSTRACT

Bounded Component Analysis (BCA) is a recent concept proposed as an alternative method for Blind Source Separation problem. BCA enables the separation of dependent as well as independent sources from their mixtures under the practical assumption on source boundedness. Therefore, Bounded Component Analysis (BCA) is a framework that can be considered as a more general framework than Independent Component Analysis (ICA) under the boundedness constraint on sources. In this thesis, we provide a stationary point analysis for recently introduced instantaneous BCA algorithms. We then extend the instantaneous BCA method providing the ability to generate a variety of BCA algorithms. We illustrate the advantages of proposed BCA examples regarding the correlated source separation capability over the state of the art ICA based approaches.

Furthermore, we extend the instantaneous BCA approach to the convolutive BCA problem. We first introduce a family of convolutive BCA criteria and corresponding algorithms based on the stationarity assumption on sources. We prove that the global optima of the proposed criteria, under generic BCA assumptions, are equivalent to a set of perfect separators. The algorithms introduced in this approach are capable of separating not only the independent sources but also the sources that are dependent/correlated in both component (space) and sample (time) dimensions. Therefore, under the condition that the sources are bounded, they can be considered as "Extended Convolutive ICA" algorithms with additional dependent/correlated source separation capability. We illustrate the space-time correlated source separation capability through a Copula distribution based example. Furthermore, they have potential to provide improvement in separation performance especially for short data records. A frequency-selective MIMO Equalization example demonstrates the clear performance advantage of the proposed BCA approach over the state of the art ICA based approaches in setups involving convolutive mixtures of digital communication sources.

Contrary to this stochastic convolutive framework, we propose novel deterministic convolutive BCA frameworks for the blind source extraction and blind source separation problems which allow the sources to be potentially non-stationary. The global maximizers of the proposed deterministic BCA optimization settings are proved to be perfect separators. We illustrate that the iterative algorithms corresponding to these frameworks are capable of extracting/separating convolutive mixtures of non-stationary as well as stationary independent and/or dependent sources.

ÖZETÇE

Sınırlı Bileşenler Analizi (BCA), kaynak ayrıştırma problemi için yeni tasarlanmış bir metot olup kaynakların sınırlı olduğu varsayımından faydalanarak bağımlı ve bağımsız kaynakları birbirinden ayırmaya olanak sağlamaktadır. Bu yüzden, Sınırlı Bileşenler Analizi (BCA) kaynakların sınırlı olması varsayımı altında Bağımsız Bileşenler Analizinden (ICA) daha genel bir yöntemdir. Bu tezde, yeni bir anlık BCA yönteminde öne sürülen algoritmaların yakınsama analizini yapıyoruz. Daha sonra, bu anlık BCA yöntemini çeşitli BCA algoritmaları üretilebilecek şekilde geliştiriyoruz. Geliştirilen yeni metotla oluşturulan algoritma örneklerinin literatürde bulunan bazı ICA yöntemlerine göre bağımlı kaynak ayrıştırma performanslarındaki avantajlarını gösteriyoruz.

Bu çalışmalara ilave olarak, anlık BCA metotunu geliştirerek evrişimsel BCA yöntemi üretiyoruz. Öncelikle, kaynakların durağan olduğunu kabul ederek evrişimsel BCA kriterleri ve karşılık gelen algoritmaları tanımlıyoruz. Tanımlanan kriterlerin, soysal BCA varsayımları altında, evrensel maksimumlarının mükemmel ayrıştırıcılardan oluşan bir kümeye denk geldiğini ispatlıyoruz. Bu yöntemde tanımlanan algoritmaların sadece bağımsız değil, bileşenlerinde ve zamanda bağımlı kaynaklarında ayrıştırmasını yapabildiğini gösteriyoruz. Bu yüzden, kaynakların sınırlı olduğu varsayımı altında, BCA algoritmalarını bağımlı ve ilintili kaynakları ayırabilme özelliğine sahip genişletilmiş evrişimsel ICA algoritmaları olarak düşünebiliriz. Copula dağılımıyla üretilen kaynaklar örneğiyle, BCA algoritmalarının bileşenlerinde ve zamanda ilintili kaynakları ayrıştırabildiğini gösteriyoruz. Ayrıca, veri sayısının az olduğu durumlarda, ayrıştırma performansında daha iyi sonuçlar verdiklerini ortaya çıkarıyoruz. Çok girişli çok çıkışlı frekans seçimli denkleştirme örneği, dijital iletişim kaynaklarının evrişimsel karışımlarında, önerilen BCA yönteminin literatürde geçen ICA yöntemlerinden üstünlüğünü gösteriyor.

Bir önceki yöntemin tersine, evrişimsel karışımların kaynak ayrıştırma ve özütleme prob-

lemleri için gerekirci evrişimsel BCA metotları tasarlıyoruz. Böylelikle kaynakların durağan olmasını varsaymıyoruz. Tanımladığımız gerekirci kriterlerin evrensel maksimumlarının mükemmel ayrıştırıcılar kümesine denk geldiğini ispatlıyoruz. Bunun yanı sıra, ortaya çıkan algoritmaların evrişimsel durağan yada durağan olmayan bağımlı veya bağımsız kaynakları ayrıştırabilme ve özütleyebilme kapasitelerine sahip olduklarını örneklerle gösteriyoruz.

Contents

1 Introduction **1**

1.1 Contributions 4

1.1.1 Convergence Analysis for Instantaneous BCA Algorithms 5

1.1.2 Extension of Instantaneous BCA Approach 5

1.1.3 Convolutive BCA Algorithms for Stationary Independent and/or Dependent Source Separation 5

1.1.4 A Convolutive BCA Analysis Framework for Potentially Non-Stationary Independent and/or Dependent Sources 7

1.2 Outline 7

2 Notation and BCA Setups **9**

2.1 Notation 9

2.2 Instantaneous BCA Setup 10

2.3 Convolutive BCA Setup 11

3 Review of Instantaneous BCA Approach in [1] **16**

4 Convergence Analysis for Instantaneous BCA Algorithms **19**

4.1 Iterative BCA Algorithms 19

4.2 Convergence Analysis Corresponding to Objective Function $J_1(\boldsymbol{W})$ . . . 21

4.3 Convergence Analysis Corresponding to Objective Function $J_{2,1}(\boldsymbol{W})$ . . . 27

4.4 Convergence Analysis Corresponding to Objective Function $J_{2,2}(\boldsymbol{W})$ . . . 29

4.5 Convergence Analysis Corresponding to Objective Function $J_{2,\infty}(\boldsymbol{W})$ . . 30

4.6 Extension to Complex Signals 31

4.7 Convergence Analysis Corresponding to Objective Function $Jc_1(\psi(\boldsymbol{W}))$. 33

4.8 Convergence Analysis Corresponding to Objective Function $Jc_{2,1}(\psi(\boldsymbol{W}))$ 35

4.9 Convergence Analysis Corresponding to Objective Function $Jc_{2,2}(\psi(\boldsymbol{W}))$ 36
4.10 Convergence Analysis Corresponding to Objective Function $Jc_{2,\infty}(\psi(\boldsymbol{W}))$ 37
4.11 Conclusion . . . 38

5 Extension of Instantaneous BCA Approach . . . 39
5.1 Extended BCA Optimization Framework . . . 40
5.2 Adaptive Implementations . . . 44
5.3 Numerical Examples and Conclusion . . . 45

6 Convolutive BCA Algorithms for Stationary Independent and/or Dependent Source Separation . . . 48
6.1 A Family of Convolutive BCA Algorithms . . . 48
6.1.1 A Convolutive BCA Optimization Framework . . . 48
6.1.2 The Global Optimality of the Perfect Separators . . . 50
6.1.3 Extension of Convolutive BCA Optimization Framework . . . 53
6.2 Iterative BCA Algorithms . . . 56
6.3 Extension to Complex Signals . . . 59
6.3.1 Complex Extension of the Convolutive BCA Optimization Framework . . . 60
6.3.2 Complex Extension of the Alternative Objective Functions . . . 64
6.4 Numerical Examples . . . 65
6.4.1 Separation of Space-Time Correlated Sources . . . 65
6.4.2 MIMO Blind Equalization . . . 68
6.5 Conclusion . . . 70

7 A Convolutive BCA Analysis Framework for Potentially Non-Stationary Independent and/or Dependent Sources . . . 72
7.1 Blind Source Extraction . . . 72
7.1.1 Criterion . . . 73
7.1.2 Algorithm . . . 74
7.1.3 Extension to Complex Signals . . . 75
7.2 Blind Source Separation . . . 76

7.2.1 Criteria . . . 77
7.2.2 Algorithms . . . 79
7.2.3 Extension to Complex Signals . . . 81
7.3 Examples . . . 83
7.3.1 Blind Source Extraction . . . 84
7.3.2 Blind Source Separation . . . 86
7.4 Conclusion . . . 89
7.5 Appendix . . . 90
7.5.1 Proof of Theorem 1 . . . 90
7.5.2 Proof of Theorem 2 . . . 92
7.5.3 Proof of Theorem 3 . . . 93
7.5.4 Analysis of the Family of Objective Functions ($J_{s2,r}$) . . . 96
7.5.5 Proof of Theorem 4 . . . 99

8 Conclusion and Future Work . . . 101

Bibliography . . . 102

List of Figures

2.1 Instantaneous Blind Source Separation Setup. 10

2.2 Convolutive Blind Source Separation Setup. 12

3.1 Bounding Hyper-rectangle and Principal Hyper-ellipsoid 17

5.1 Result of the proposed BCA algorithms' performances for the mixtures of dependent sources for various correlation parameters when the mixture length is 100000. 46

5.2 Result of the proposed BCA algorithms' performances for the mixtures of exponentially distributed dependent sources for various correlation parameters when the mixture length is 10000. 47

6.1 Dependent convolutive mixtures separation performance results for SNR = 45dB. 67

6.2 Dependent convolutive mixtures separation performance results for SNR = 20dB. 68

6.3 Dependent convolutive mixtures separation performance results for SNR = 5dB. 69

6.4 Signal to Interference Ratio as a function of Sample Length 70

7.1 Result of the proposed blind source extraction algorithm performance for the convolutive mixtures of dependent sources (ρ_s and ρ_t is set as 0.5) for various sample lengths under SNR = 45dB. 85

7.2 Result of the proposed blind source extraction algorithm performance for the convolutive mixtures of dependent sources (ρ_s and ρ_t is set as 0.5) for various sample lengths under SNR = 20dB. 86

7.3 Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 45dB. 87

7.4 Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 20dB. 88

7.5 Result of the proposed blind source separation algorithm performance for the convolutive mixtures of digital communication sources for various sample lengths. 89

7.6 Two dimensional example for the global maxima of (4.63). 92

List of Tables

2.1 Notation used in the thesis. 9

Chapter 1

Introduction

Blind Source Separation (BSS) is one of the basic problems in signal processing and machine learning with a diverse set of applications [2]. BSS aims to extract individual components (or sources) from their mixture samples where there is no, or very limited, prior information about their nature or the mixing process. We can state some prominent application examples of BSS as:

- **Cocktail Party Problem :** A number of people are talking simultaneously in a room resulting in a mixture of speeches.
- **Brain Signal Processing :** Measuring of electromagnetic signals from different brain regions where the muscle artefacts mix with the brain signal of interest.
- **Digital Communications :** Transmission of the digital communication signals.

The blindness property is the key to the flexibility of this approach which leads to its widespread use. However, the blindness feature also makes BSS a challenging problem to solve. The hardship caused by the lack of training data and relational statistical information is generally overcome by exploiting some "side information"/assumptions about the model.

The most common assumption is the mutual statistical independence of sources. The approach based on this assumption is referred to as the Independent Component Analysis (ICA) and it is the most popular and successful BSS approach [2–4]. Its success resides in the simple and generic nature of the independence assumption and its applicability ensures that ICA has a diverse range of BSS application domains. There have been several other assumptions mostly fortifying the independence assumption such as time structure (e.g., [5, 6]), sparsity (e.g., [7]) and special constant modulus or finite alphabet structure of communications signals (e.g. [8–10]).

In typical practical BSS applications source values take their values from a compact set. This property has been exploited especially in some recent ICA algorithms. The potential for utilizing the boundedness property as an additional assumption in the ICA framework was first put forward in [11]. In this work, Pham reformulated the mutual information cost function in terms of order statistics. In the bounded case, this formulation leads to the effective minimization of the separator output ranges. In the similar direction, Cruces and Duran showed that an optimization framework based on Renyi's Entropy leads to support length minimization to extract sources from their mixtures [12]. Vrins et. al, utilized range minimization approach to obtain alternative ICA algorithms [13–15]. Parallel to these contributions, Erdogan extended the infinity norm minimization based blind equalization approach in [16, 17] to obtain source separation algorithms, again within ICA framework, based on infinity norm minimization [18], [19]. These algorithms assumed peak symmetry for the bounded sources, which is later abandoned in [20].

Following these contributions related to exploitation of boundedness of signals within the ICA framework, recently, Cruces showed that boundedness can be utilized to replace mutual statistical independence assumption with a weaker assumption [21]. This fact led to a new framework, named Bounded Component Analysis (BCA), which enables separation of independent and dependent (even correlated) sources.

For bounded sources, BCA provides a more general framework than ICA, since the joint density factorability requirement of the mutual independence assumption is replaced by the weaker domain separability assumption. The domain separability assumption refers to the condition that the convexified effective support of the joint pdf to be written as the cartesian product of their marginal pdf counterparts. Therefore, it is a necessary condition for the independence. However, for the independence assumption to hold there is a more stringent requirement about the factorizability of the joint pdf in terms of product of marginals. BCA framework removes this requirement, therefore provides a more flexible framework for bounded sources including ICA as a special case.

Within the newly introduced BCA framework, Cruces introduced a source extraction algorithm in [21]. A deflationary approach for BCA was recently proposed in [22]. In [23], total range minimization is posed as a BCA approach for uncorrelated sources and the characterization of the stationary points of the corresponding symmetric orthogonalization algorithm are provided. More recently, Erdogan proposed a new BCA approach which enables separation of both independent and dependent, including correlated, bounded sources from their instantaneous mixtures [1]. In this approach, two geometric objects, namely principal hyperellipsoid and bounding hyperrectangle, concerning separator outputs are introduced. The separation problem is posed as maximization of the relative sizes of these objects, in which the size of hyperellipsoid is chosen as its volume. When the size of bounding hyperrectangle is chosen as its volume, a generalized form of Pham's objective in [11], which was derived by manipulating the mutual information objective in ICA framework, is obtained. When the size of the bounding hyperrectangle is chosen as a norm of its main diagonal, this leads to a set of BCA algorithms whose global optima correspond to a fixed relative scalings of sources at the separator outputs.

In this thesis, we first provide a stationary point analysis for the instantaneous BCA algorithms introduced in [1]. We prove that all stationary points of the instantaneous BCA algorithms rather than perfect separators are saddle points. We then extend the instantaneous BCA approach by considering generalized functions of ranges of separator outputs which can be used to generate a variety of instantaneous BCA algorithms. Through simulations, we illustrate the advantages of proposed BCA examples regarding the correlated source separation capability over the state of the art ICA based approaches.

We furthermore extend the instantaneous BCA method to a convolutive BCA framework where we assume the stationary of sources and utilize process based correlation information among sources in the optimization setting. We show that with this approach, it is possible to generate algorithms that are capable of separating not only independent sources but also dependent (even correlated) sources from their convolutive mixtures when the sources are assumed to be stationary. We note that this is the first convolutive BCA approach in the literature. Contrary to this stochastic convolutive framework, we additionally propose novel deterministic convolutive BCA frameworks for the blind source extraction and blind source separation problems where the sources are allowed to be non-stationary. We show that the proposed scheme can generate algorithms that can extract/separate convolutive mixtures of non-stationary as well as stationary independent and/or dependent sources. We point out that even when the sources are independent, the samples may not reflect this behaviour especially for short data records. Hence, we show the potential for the significant performance improvement offered by the proposed BCA approach over the state of the art ICA based approaches, especially for short data records.

1.1 Contributions

The contributions of this thesis are as follows :

1.1.1 Convergence Analysis for Instantaneous BCA Algorithms

- A stationary point analysis for the BCA algorithms introduced in [1] is presented. It is shown that all stationary points of the instantaneous BCA algorithms besides perfect separators are saddle points.

- This work is to be submitted to IEEE Transactions on Signal Processing.

1.1.2 Extension of Instantaneous BCA Approach

- The instantaneous BCA approach introduced in [1] is extended by considering generalized functions of ranges of output samples (corresponding to the side lengths of bounding hyper-rectangle) which also covers the size of bounding hyper-rectangle.

- It is shown that the extended approach can be used to generate a variety of instantaneous BCA algorithms.

- The advantages of proposed BCA examples regarding the correlated source separation capability over the state of the art ICA based approaches are illustrated through simulations.

- This work is submitted to Asilomar Conference on Signals, Systems, and Computers.

1.1.3 Convolutive BCA Algorithms for Stationary Independent and/or Dependent Source Separation

- A convolutive BCA approach is proposed which can be used to generate an algorithm that is capable of separating not only real independent sources but also dependent (even correlated) sources from their convolutive mixtures when the sources are stationary. Therefore, for bounded sources, a more general framework than ICA is proposed for the convolutive source separation problem which replaces the strong mutual independence assumption with weaker and more generic domain separability assumption leading to additional capability to separate sources which are potentially dependent/correlated in both space and time directions.

- A convolutive BCA objective is offered whose global optima are proven to correspond to perfect separators.

- This part is presented in the 38th International Conference on Acoustics, Speech and Signal Processing.

- The optimization setting of the approach is then extended which generates a family of convolutive BCA algorithms. Furthermore, the complex sources case is included in the approach.

- Through a digital communications example, the potential for the significant performance improvement offered by the proposed BCA approach over the state of the art ICA based approaches, especially for short data records is illustrated.

- The framework prescribes an approach based on the update of time-domain filter parameters which is free of permutation alignment problem suffered by the algorithms using frequency domain updates. Furthermore, unlike many time-domain convolutive ICA approaches, the proposed approach does not require prewhitening operation, which is problematic in convolutive settings.

- The journal version is accepted for publication in IEEE Transactions on Neural Networks and Learning Systems.

1.1.4 A Convolutive BCA Analysis Framework for Potentially Non-Stationary Independent and/or Dependent Sources

- Novel deterministic convolutive BCA approaches are proposed for the blind source extraction and blind source separation problems where the objectives are directly defined in terms of mixture samples rather than some stochastic measures or their sample based estimates allowing the sources to be potentially non-stationary. The introduced framework does not exploit any non-stationarity feature, therefore it is applicable to both stationary/non-stationary sources.

- It is proved that the global optima of proposed BCA objectives correspond to perfect extractors/separators.

- The capability of proposed algorithms regarding the extracting/separating convolutive mixtures of dependent (even correlated) sources are illustrated. The performances of proposed algorithms with the state of the art convolutive ICA approaches are further compared and through a digital communications example, the potential for the significant performance improvement offered by the proposed BCA approach is shown, especially for short data records.

- This work is submitted to IEEE Transactions on Signal Processing.

1.2 Outline

Chapter 2 begins with the notation and the BCA setups that are considered throughout the thesis. In chapter 3, we provide an essential summary of the instantaneous BCA approach introduced in [1]. In chapter 4, we first recall the instantaneous BCA algorithms and then provide the convergence analysis results of the algorithms considered for both real and complex signals. Chapter 5 extends the instantaneous BCA approach using generalized functions of ranges of separator outputs including the size of hyperrectangle.

Chapter 6 presents the convolutive BCA approach that assumes the stationarity of sources and produces a family of convolutive BCA algorithms. The numerical examples illustrating the separation capability of proposed algorithms for the convolutive mixtures of stationary independent and/or dependent sources are provided based on Copula distribution. Especially for short data records, the performance improvement offered by the proposed BCA approach over the state of the art ICA based approaches is shown through a digital communications example. Chapter 7 provides a deterministic approach for the blind source extraction and blind source separation problems where the sources are allowed to be potentially non-stationary. In this approach, there is no stationarity assumption. The sources can be both stationary and non-stationary. It is shown that the convolutive BCA algorithms generated by this scheme are capable of separation of stationary as well as non-stationary independent and/or dependent sources.

Finally, Chapter 8 is for concluding comments and discussing future works.

Chapter 2

Notation and BCA Setups

In this chapter, we will present the notation and BCA setups that we consider in our derivations for the instantaneous and convolutive blind source separation problems.

2.1 Notation

Let $\mathbf{A} \in \mathbb{C}^{p \times q}$ and $\mathbf{a} \in \mathbb{C}^{p \times 1}$ be arbitrary. The notation used in the thesis is summarized in Table 2.1.

Notation	**Meaning**
$\mathbf{A}_{m,:}$ ($\mathbf{A}_{:,m}$)	m^{th} row (column) of $\mathbf{A}$
$\mathbb{R}\{\mathbf{A}\}$ ($\mathbb{I}\{\mathbf{A}\}$)	The real (imaginary) part of $\mathbf{A}$
$\lVert\mathbf{a}\rVert_r$	Usual r-norm given by $(\sum_{m=1}^{p} \lvert a_m\rvert^r)^{1/r}$.
$diag(\mathbf{a})$	Diagonal matrix whose diagonal entries starting in the upper left corner are a_1 , . . . , a_p.
$\prod(\mathbf{a})$	$a_1 a_2 \ldots a_p$, i.e. the product of the elements of $\mathbf{a}$.
$\mathcal{S}_{\mathbf{a}}$	The convex support for random vector $\mathbf{a}$
$\mathbf{e}_m$	Standard basis vector pointing in the m direction.
I	Identity matrix
$\otimes$	Kronecker product

Table 2.1: Notation used in the thesis.

Indexing: m is used for (source, output) vector components, k is the sample index and i is the algorithm iteration index.

2.2 Instantaneous BCA Setup

The instantaneous BSS setup assumed throughout the thesis is summarized in Figure 2.1:

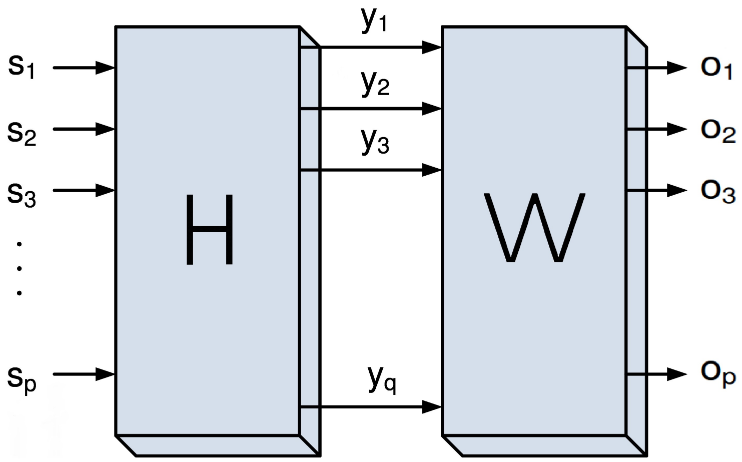

Figure 2.1: Instantaneous Blind Source Separation Setup.

- We consider a deterministic setup consisting of p real sources which are represented by the vector $\boldsymbol{s} = [\ s_1 \quad s_2 \quad \dots \quad s_p\ ]^T$. We assume that the sources are bounded such that $s_m(k) \in [\alpha_m, \beta_m]$ where $\alpha_m, \beta_m \in \Re, \beta_m > \alpha_m$ for $m = 1, ..., p$ and $k \in \mathbb{Z}$. We point out that we do not assume the sources are independent, or uncorrelated. In fact, the sources are allowed to be potentially correlated.

- The sources are mixed by a linear and instantaneous system. The mixing matrix is assumed to be full rank and represented as $\boldsymbol{H} \in \Re^{q \times p}$. We further assume that $q \geq p$, therefore, we consider the (over)determined BSS problem. The mixtures are represented with $\boldsymbol{y} = [\ y_1 \quad y_2 \quad \dots \quad y_q\ ]^T$ where the sources and the mixtures are related by $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{s}$.

- $\boldsymbol{W} \in \Re^{p \times q}$ is the separator matrix of the system which produces the outputs as $\boldsymbol{o} = \boldsymbol{W}\boldsymbol{y}$.

- The overall system function is defined as $\boldsymbol{G} = \boldsymbol{WH} \in \Re^{p \times p}$ where the relation between the sources and the outputs can be written as $\boldsymbol{o} = \boldsymbol{Gs}$.

$Y = \{\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(N)\}$ is the finite set consisting of observations of mixture samples. The main goal in BSS problems is to adapt the separator system based on these observations. We denote the corresponding set of unobservable source samples by $S = \{\boldsymbol{s}(1), \boldsymbol{s}(2), \ldots, \boldsymbol{s}(N)\}$. The following assumption is introduced regarding the set S:

Assumption: S contains the vertices of its (non-degenerate) bounding hyper-rectangle (**A1**).

The separator system and the corresponding overall system produce the set of output samples as

$$\begin{aligned} O &= \{\boldsymbol{Wy}(1), \boldsymbol{Wy}(2), \ldots, \boldsymbol{Wy}(N)\} \\ &= \{\boldsymbol{Gs}(1), \boldsymbol{Gs}(2), \ldots, \boldsymbol{Gs}(N)\}. \end{aligned}$$

The optimization settings are proposed based on the set O.

2.3 Convolutive BCA Setup

The convolutive BSS setup assumed throughout the thesis is summarized in Figure 2.2:

- The setup consists of p real sources. The sources are represented by the vector $\boldsymbol{s} = [\ s_1 \quad s_2 \quad \ldots \quad s_p\ ]^T$. We assume that ranges of the sources are bounded, i.e., $s_m(k) \in [\alpha_m, \beta_m]$ where $\alpha_m, \beta_m \in \mathbb{R}, \beta_m > \alpha_m$ for $m = 1, ..., p$ and $k \in \mathbb{Z}$. We define $\gamma_m = \mathcal{R}(s_m) = \beta_m - \alpha_m$ as the range of s_m where $\mathcal{R}(\cdot)$ is the range

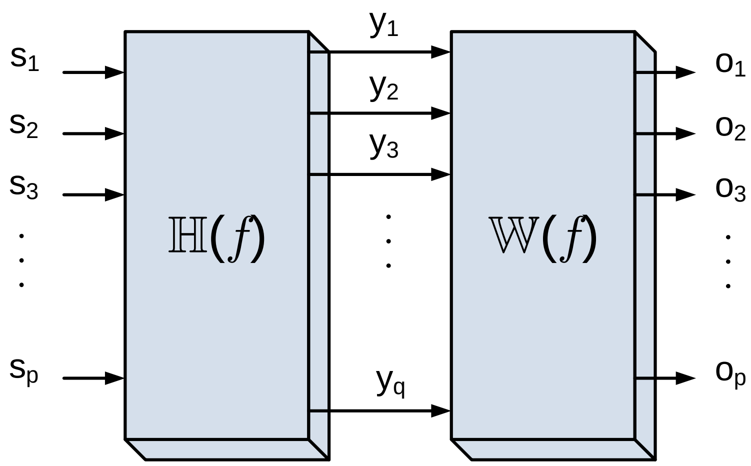

Figure 2.2: Convolutive Blind Source Separation Setup.

operator providing the support length for the pdf of its argument. We decompose the sources as

$$\boldsymbol{s}(k) \stackrel{\triangle}{=} \Upsilon \underline{\boldsymbol{s}}(k), \quad k \in \mathbb{Z},$$

where $\Upsilon = diag(\gamma_1, \gamma_2, \ldots, \gamma_p)$ is the range matrix of $\boldsymbol{s}$ and $\{\underline{\boldsymbol{s}}(k) \in \mathbb{R}^p\,; k \in \mathbb{Z}\}$ is the normalized source process whose components have unit ranges.

- The source signals are mixed by a MIMO system with a q x p transfer matrix

$$\mathbb{H}(f) = \sum_{l=0}^{L-1} \boldsymbol{H}(l) e^{-j2\pi f l},$$

where $\{\boldsymbol{H}(l);\ l \in \{0, \ldots, L-1\}\}$ are the impulse response coefficients, $\{\mathbb{H}(f);\ f \in [-\frac{1}{2}, \frac{1}{2})\}$ represents the Discrete Time Fourier Transform (DTFT) of the impulse response. We assume that $\mathbb{H}(f)$ is an equalizable transfer function of order $L-1$ [25]. The output of the mixing MIMO system is denoted by $\{\boldsymbol{y}(k) \in \mathbb{R}^q\,; k \in \mathbb{Z}\}$.

We have

$$\boldsymbol{Y}(f) = \mathbb{H}(f)\mathbf{S}(f),$$

where $\boldsymbol{Y}(f)$ represents the Discrete Time Fourier Transform (DTFT) of $\{\boldsymbol{y}(k) \in \mathbb{R}^q\,; k \in \mathbb{Z}\}$, $\mathbf{S}(f)$ is the DTFT of $\{\boldsymbol{s}(k) \in \mathbb{R}^p\,; k \in \mathbb{Z}\}$. Equivalently, in the time domain

$$\boldsymbol{y}(k) = \sum_{l=0}^{L-1} \boldsymbol{H}(l)\boldsymbol{s}(k-l), \quad k \in \mathbb{Z}.$$

Defining $\tilde{\boldsymbol{H}} = [\ \boldsymbol{H}(0) \quad \boldsymbol{H}(1) \quad \ldots \quad \boldsymbol{H}(L-1)\ ]$ as the mixing coefficient matrix and $\tilde{\boldsymbol{s}}_L(k) = [\ \boldsymbol{s}^T(k) \quad \boldsymbol{s}^T(k-1) \quad \ldots \quad \boldsymbol{s}^T(k-L+1)\ ]^T$, we can also write

$$\boldsymbol{y}(k) = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{s}}_L(k), \quad k \in \mathbb{Z}.$$

- Separator of the system

$$\mathbb{W}(f) = \sum_{l=0}^{M-1} \boldsymbol{W}(l)e^{-j2\pi fl}$$

is a p x q FIR transfer matrix of order $M-1$ where $\{\boldsymbol{W}(l);\ l \in \{0, \ldots, M-1\}\}$ are the impulse response coefficients and $\{\mathbb{W}(f);\ f \in [-\frac{1}{2}, \frac{1}{2})\}$ represents the DTFT of the coefficients. The separator output sequence is denoted by $\{\boldsymbol{o}(k) \in \mathbb{R}^p\,; k \in \mathbb{Z}\}$ whose DTFT can be written as

$$\boldsymbol{O}(f) = \mathbb{W}(f)\boldsymbol{Y}(f).$$

Equivalently, in the time domain

$$\boldsymbol{o}(k) = \sum_{l=0}^{M-1} \boldsymbol{W}(l)\boldsymbol{y}(k-l), \quad k \in \mathbb{Z}.$$

Defining $\tilde{\boldsymbol{W}} = [\ \boldsymbol{W}(0) \quad \boldsymbol{W}(1) \quad \ldots \quad \boldsymbol{W}(M-1)\ ]$ as the separator coefficient matrix and $\tilde{\boldsymbol{y}}_M(k) = [\ \boldsymbol{y}^T(k) \quad \boldsymbol{y}^T(k-1) \quad \ldots \quad \boldsymbol{y}^T(k-M+1)\ ]^T$, we can also write

$$\boldsymbol{o}(k) = \tilde{\boldsymbol{W}}^T \tilde{\boldsymbol{y}}_M(k), \quad k \in \mathbb{Z}.$$

.

- The overall system function is defined as

$$\mathbb{G}(f) = \mathbb{W}(f)\mathbb{H}(f) = \sum_{l=0}^{P-1} \boldsymbol{G}(l)e^{-j2\pi f l},$$

where $\{\boldsymbol{G}(l);\, l \in \{0, \ldots, P-1\}\}$ are the impulse response coefficients, $\{\mathbb{G}(f);\, f \in [-\frac{1}{2}, \frac{1}{2})\}$ represents the DTFT of the coefficients and $P-1$ is the order of overall system. Therefore, in the time domain, the sources $\{\boldsymbol{s}(k) \in \mathbb{R}^p\,;k \in \mathbb{Z}\}$ and the separator outputs $\{\boldsymbol{o}(k) \in \mathbb{R}^p\,;k \in \mathbb{Z}\}$ are related by

$$\boldsymbol{o}(k) = \sum_{l=0}^{P-1} \boldsymbol{G}(l)\boldsymbol{s}(k-l), \quad k \in \mathbb{Z}.$$

We similarly define $\tilde{\boldsymbol{G}} = [\ \boldsymbol{G}(0) \quad \boldsymbol{G}(1) \quad \ldots \quad \boldsymbol{G}(P-1)\ ]$ and $\tilde{\boldsymbol{s}}_P(k) = [\ \boldsymbol{s}(k) \quad \boldsymbol{s}(k-1) \quad \ldots \quad \boldsymbol{s}(k-P+1)\ ]^T$, we have $\boldsymbol{o}(k) = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{s}}_P(k)$, for $k \in \mathbb{Z}$. We obtain the range matrix of $\tilde{\boldsymbol{s}}$ as $\tilde{\Upsilon} = I \otimes \Upsilon$.

Similarly, $Y = \{\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots, \boldsymbol{y}(N)\}$ is the finite set consisting of observations of mixture samples . Since the mixing channel $\tilde{\boldsymbol{H}}$ is convolutive having order of $L-1$, the corresponding set of unobservable source samples could be denoted by $S = \{\boldsymbol{s}(-L+2), \ldots, \boldsymbol{s}(0), \boldsymbol{s}(1), \boldsymbol{s}(2), \ldots, \boldsymbol{s}(N)\}$ such that

$$Y = \{\tilde{\boldsymbol{H}}\tilde{\boldsymbol{s}}_L(1), \tilde{\boldsymbol{H}}\tilde{\boldsymbol{s}}_L(2), \ldots, \tilde{\boldsymbol{H}}\tilde{\boldsymbol{s}}_L(N)\}.$$

We point out that the source samples in the set S could be generated from stationary distributions.

For a given convolutive separator channel $\tilde{\boldsymbol{W}}$ having order of $M-1$ with a corresponding convolutive overall channel $\tilde{\boldsymbol{G}}$ having order of $P-1$, the convolutive nature of channel generates $N-M+1$ outputs and we illustrate the generated set of separator outputs $O = \{\boldsymbol{o}(1), \boldsymbol{o}(2), \ldots, \boldsymbol{o}(N-M+1)\}$ as

$$\begin{aligned} O &= \{\tilde{\boldsymbol{W}}\tilde{\boldsymbol{y}}_M(M), \tilde{\boldsymbol{W}}\tilde{\boldsymbol{y}}_M(M+1), \ldots, \tilde{\boldsymbol{W}}\tilde{\boldsymbol{y}}_M(N)\} \\ &= \{\tilde{\boldsymbol{G}}\tilde{\boldsymbol{s}}_P(M), \tilde{\boldsymbol{G}}\tilde{\boldsymbol{s}}_P(M+1), \ldots, \tilde{\boldsymbol{G}}\tilde{\boldsymbol{s}}_P(N)\}. \end{aligned}$$

Chapter 3

Review of Instantaneous BCA Approach in [1]

In this chapter, an essential summary of the deterministic instantaneous BCA approach introduced in [1] is provided. We recall that we will extend this approach to the convolutive BCA problem. We first start with the definitions of two geometric objects used by this approach:

- **Principal Hyperellipsoid** is the hyperellipse whose principal semi-axis directions are determined by the eigenvectors of the covariance matrix and whose principal semi-axis lengths are equal to principal standard deviations, i.e., the square roots of the eigenvalues of the covariance matrix.
- **Bounding Hyperrectangle** corresponds to the box defined by the Cartesian product of the support sets of the individual components. This can be also defined as the minimum volume box containing all samples and aligning with the coordinate axes.

An example, for a case of 3-sources to enable 3D picture, is provided in Figure 3.1. In this figure, a realization of separator output samples and the corresponding bounding hyperrectangle and principal hyper-ellipsoid are shown.

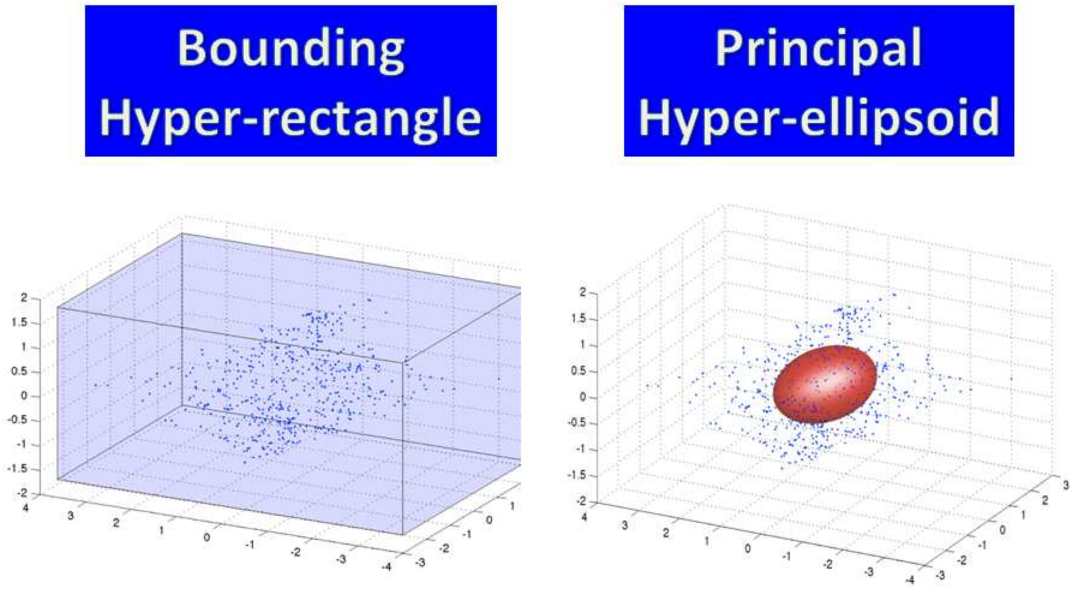

Figure 3.1: Bounding Hyper-rectangle and Principal Hyper-ellipsoid

The separation problem is posed as maximization of the relative sizes of these objects. First optimization setting of this approach is given as:

$$maximize \quad J_1(\boldsymbol{W}) = \frac{\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{O}})}}{\prod \hat{\mathcal{R}}(O)}. \tag{3.1}$$

In this optimization setting, $\hat{\boldsymbol{R}}_{\boldsymbol{O}}$ represents the sample covariance matrix of the output samples in the set O and $\hat{\mathcal{R}}(O)$ contains the range values of the components of the output samples in the set O. Note that the numerator of (3.1) is the (scaled) volume of the principal hyperellipse, whereas the denominator is the volume of the bounding hyperrectangle for the output vectors.

The principal hyperellipsoid in the output domain is the image (with respect to the instantaneous overall mapping defined as $\boldsymbol{G}$) of the principal hyper-ellipsoid in the source domain. We note that the instantaneous overall mapping causes $|det(\boldsymbol{G})|$ scaling in the volume of principal hyperellipsoid. However, the image of the bounding hyperrectangle in the source domain is a hyperparallelopiped which is a subset of the bounding hyperrectangle in the output domain. Hence, the volume scaling for the source and separator output bounding boxes is more than or equal to $|det(\boldsymbol{G})|$. As an important observation, for the latter, the scaling would be $|det(\boldsymbol{G})|$ if and only if $\boldsymbol{G}$ is a perfect separator matrix. We observe this by simple geometrical reasoning that the bounding box in the source domain is mapped to another hyper-rectangle (aligning with coordinate axes) if and only if $\boldsymbol{G}$ can be written as $\boldsymbol{G} = \boldsymbol{D}\boldsymbol{P}$, where $\boldsymbol{D}$ is a diagonal matrix with non-zero diagonal entries, and $\boldsymbol{P}$ is a permutation matrix. The set of such $\boldsymbol{G}$ matrices is referred as Perfect Separators. Therefore, the optimization setting given as an example provides Perfect Separators.

When the size of the bounding hyper-rectangle is chosen as a norm of its main diagonal, a family of alternative optimization settings is proposed as:

$$maximize \quad J_{2,r}(\boldsymbol{W}) = C_p \frac{\sqrt{\det(\hat{\boldsymbol{R}}_{\boldsymbol{O}})}}{||\hat{\mathcal{R}}(\boldsymbol{o})||_r^p}. \tag{3.2}$$

In this case, it is shown that the global optima set is in the form $\boldsymbol{G} = d\boldsymbol{P}(\boldsymbol{U} - \boldsymbol{L})^{-1}diag(\sigma)$ where d is a non-zero value, $\sigma \in \{-1, 1\}^p$. Therefore, all the members of the global optima set share the same relative source scalings, unlike the set of global optima for J_1 which has arbitrary relative scalings.

Chapter 4

Convergence Analysis for Instantaneous BCA Algorithms

In this chapter, we provide a stationary point analysis for the instantaneous BCA algorithms introduced in [1].

4.1 Iterative BCA Algorithms

We first provide the iterative algorithms corresponding to the objective functions introduced in [1] as follows:

- Objective function $J_1(\boldsymbol{W})$:

$$\boldsymbol{W}^{(t+1)} = \boldsymbol{W}^{(t)} + \mu^{(t)}\left(\left(\boldsymbol{W}^{(t)}\hat{\mathbf{R}}(Y)\boldsymbol{W}^{(t)^T}\right)^{-1}\boldsymbol{W}^{(t)}\hat{\mathbf{R}}(Y) - \sum_{m=1}^{p}\frac{1}{\hat{\mathcal{R}}_m(O^{(t)})}\mathbf{e}_m\boldsymbol{b}_m^{(t)^T}\right), \tag{4.1}$$

with

$$\boldsymbol{b}_m^{(t)} = \sum_{k_{m,+}\in\mathcal{K}_{m,+}(O^{(t)})}\lambda_{m,+}^{(t)}(k_{m,+})\boldsymbol{y}(k_{m,+}) - \sum_{k_{m,-}\in\mathcal{K}_{m,-}(O^{(t)})}\lambda_{m,-}^{(t)}(k_{m,-})\boldsymbol{y}(k_{m,-}).$$

where

- $\boldsymbol{W}^{(t)}$ is the separator matrix at the t^{th} iteration,

– $\mu^{(t)}$ is the step size at the t^{th} iteration,

– $\mathcal{K}_{m,+}(O^{(t)})$ is the set of indexes where the m^{th} separator output reaches its maximum value at the t^{th} iteration,

– $\mathcal{K}_{m,-}(O^{(t)})$ is the set of indexes where the m^{th} separator output reaches its minimum value at the t^{th} iteration,

– $\{\lambda_{m,+}^{(t)}(k_{m,+}) : k_{m,+} \in \mathcal{K}_{m,+}(O^{(t)})\}$ is the convex combination coefficients, used for combining the input vectors causing the maximum output, at the t^{th} iteration, which satisfy

$$\lambda_{m,+}^{(t)}(k_{m,+}) \geq 0, \quad k_{m,+} \in \mathcal{K}_{m,+}(O^{(t)}), \qquad \sum_{k_{m,+} \in \mathcal{K}_{m,+}(O^{(t)})} \lambda_{m,+}^{(t)}(k_{m,+}) = 1,$$

for $m = 1, 2, \ldots, p$.

– $\{\lambda_{m,-}^{(t)}(k_{m,-}) : k_{m,-} \in \mathcal{K}_{m,-}(O^{(t)})\}$ is the convex combination coefficients, used for combining the input vectors causing the minimum output, at the t^{th} iteration, which satisfy

$$\lambda_{m,-}^{(t)}(k_{m,-}) \geq 0, \quad k_{m,-} \in \mathcal{K}_{m,-}(O^{(t)}), \qquad \sum_{k_{m,-} \in \mathcal{K}_{m,-}(O^{(t)})} \lambda_{m,-}^{(t)}(k_{m,-}) = 1,$$

for $m = 1, 2, \ldots, p$.

- Objective function $J_{2,r}(\boldsymbol{W})$:

– For $r = 1, 2$, the update equation is

$$\boldsymbol{W}^{(t+1)} = \boldsymbol{W}^{(t)} + \mu^{(t)} \Bigg(\left(\boldsymbol{W}^{(t)} \hat{\mathbf{R}}(Y) \boldsymbol{W}^{(t)^T} \right)^{-1} \boldsymbol{W}^{(t)} \hat{\mathbf{R}}(Y) - \sum_{m=1}^{p} \frac{p \hat{\mathcal{R}}_m(O^{(t)})^{r-1}}{||\hat{\mathcal{R}}(O^{(t)})||_r^r} \mathbf{e}_m \boldsymbol{b}_m^{(t)^T} \Bigg). \tag{4.2}$$

– For $r=\infty$, the update equation is

$$\boldsymbol{W}^{(t+1)}=\boldsymbol{W}^{(t)}+\mu^{(t)}\Bigg(\left(\boldsymbol{W}^{(t)}\hat{\mathbf{R}}(Y)\boldsymbol{W}^{(t)^T}\right)^{-1}\boldsymbol{W}^{(t)}\hat{\mathbf{R}}(Y) \\ -\sum_{m\in\mathcal{M}(O^{(t)})}\frac{p\beta_m^{(t)}}{||\hat{\mathcal{R}}(O^{(t)})||_\infty}\mathbf{e}_m\boldsymbol{b}_m^{(t)^T}\Bigg), \tag{4.3}$$

where $\mathcal{M}(O^{(t)})$ is the set of indexes for which the peak range values is achieved, i.e.,

$$\mathcal{M}(O^{(t)})=\{m:\hat{\mathcal{R}}_m(O^{(t)})=||\hat{\mathcal{R}}(O^{(t)})||_\infty\},$$

and $\beta_m^{(t)}=\dfrac{1}{|\mathcal{M}(O^{(t)})|}$.

4.2 Convergence Analysis Corresponding to Objective Function $J_1(\boldsymbol{W})$

In order to identify stationary points, we first rewrite the iterative algorithm (4.1) in terms of $\boldsymbol{G}$ as

$$\begin{aligned}\boldsymbol{G}^{(t+1)}&=\boldsymbol{G}^{(t)}+\mu^{(t)}\Bigg(\left(\boldsymbol{G}^{(t)}\hat{\mathbf{R}}(S)\boldsymbol{G}^{(t)^T}\right)^{-1}\boldsymbol{G}^{(t)}\hat{\mathbf{R}}(S)\boldsymbol{H}^T\boldsymbol{H}-\sum_{m=1}^{p}\frac{1}{\hat{\mathcal{R}}_m(O^{(t)})}\mathbf{e}_m\boldsymbol{b}_m^{(t)^T}\boldsymbol{H}\Bigg),\\ &=\boldsymbol{G}^{(t)}+\mu^{(t)}\Bigg(\left(\boldsymbol{G}^{(t)}\right)^{-T}-\sum_{m=1}^{p}\frac{1}{\hat{\mathcal{R}}_m(O^{(t)})}\mathbf{e}_m\Bigg[\sum_{k_{m,+}\in\mathcal{K}_{m,+}(O_{\mathbf{G}^{(t)}})}\lambda_{m,+}^{(t)}(k_{m,+})\boldsymbol{s}^T(k_{m,+})\\ &-\sum_{k_{m,-}\in\mathcal{K}_{m,-}(O_{\mathbf{G}^{(t)}})}\lambda_{m,-}^{(t)}(k_{m,-})\boldsymbol{s}^T(k_{m,-})\Bigg]\Bigg)\boldsymbol{H}^T\boldsymbol{H}.\end{aligned} \tag{4.4}$$

Note that under the assumption (**A1**), the inputs in the expression (4.4) can be written as

$$\boldsymbol{s}(k_{m,+})^T = \text{sign}\left\{\mathcal{P}\left\{\boldsymbol{G}^{(t)}_{m,:}\right\}\right\}\boldsymbol{U} - \text{sign}\left\{\mathcal{N}\left\{\boldsymbol{G}^{(t)}_{m,:}\right\}\right\}\boldsymbol{L} + \mathbf{a}^{(t)}_m(k_{m,+})^T, \tag{4.5}$$

$$\boldsymbol{s}(k_{m,-})^T = -\text{sign}\left\{\mathcal{N}\left\{\boldsymbol{G}^{(t)}_{m,:}\right\}\right\}\boldsymbol{U} + \text{sign}\left\{\mathcal{P}\left\{\boldsymbol{G}^{(t)}_{m,:}\right\}\right\}\boldsymbol{L} + \boldsymbol{c}^{(t)}_m(k_{m,-})^T, \tag{4.6}$$

where $\boldsymbol{U} = \text{diag}\,(\,\max(s_1),\ \max(s_2),\ \ldots,\ \max(s_p)\,)$ and $\boldsymbol{L} = \text{diag}(\,\min(s_1),\ \min(s_2),\ \ldots,\ \min(s_p)\,)$ are the diagonal matrices containing maximum (minimum) values for the components of the source samples in the set S and $\mathbf{a}^{(t)}_m(k_{m,+})^T$ and $\boldsymbol{c}^{(t)}_m(k_{m,-})^T$ are additive source terms which have zero values for the components corresponding to nonzero components of $\boldsymbol{G}^{(t)}_{m,:}$ and arbitrary values (from the support) for the other components. Therefore, the input dependent update part in the right side of (4.4) can be written as

$$\begin{aligned}\text{sign}\left\{\boldsymbol{G}^{(t)}_{m,:}\right\}(\boldsymbol{U}-\boldsymbol{L}) + &\sum_{k_{m,+}\in\mathcal{K}_{m,+}(O_{\mathbf{G}^{(t)}})} \lambda^{(t)}_{m,+}(k_{m,+})\mathbf{a}^{(t)}_m(k_{m,+})^T \\ - &\sum_{k_{m,-}\in\mathcal{K}_{m,-}(O_{\mathbf{G}^{(t)}})} \lambda^{(t)}_{m,-}(k_{m,-})\boldsymbol{c}^{(t)}_m(k_{m,-})^T \approx \text{sign}\left\{\boldsymbol{G}^{(t)}_{m,:}\right\}(\boldsymbol{U}-\boldsymbol{L}),\end{aligned} \tag{4.7}$$

where we assume that the contributions of the $\mathbf{a}$ and $\boldsymbol{c}$ dependent terms average out.

We identify $\boldsymbol{G}_*$ as a stationary point if and only if it is mapped to itself after an iteration of the algorithm. We note that since $\boldsymbol{H}$ is assumed to be a full rank matrix $\boldsymbol{H}^T\boldsymbol{H}$ is invertible hence from (4.4) and (4.7), this is equivalent to the condition that

$$\text{diag}(\hat{\mathcal{R}}_1(O),\ldots,\hat{\mathcal{R}}_p(O))^{-1}\text{sign}\left\{\boldsymbol{G}_*\right\}(\boldsymbol{U}-\boldsymbol{L})\boldsymbol{G}_*^T = \mathbf{I}. \tag{4.8}$$

Defining $\boldsymbol{Q} = \boldsymbol{G}_*(\boldsymbol{U}-\boldsymbol{L})$ with noting that $\text{sign}\left\{\boldsymbol{G}_*\right\} = \text{sign}\left\{\boldsymbol{Q}\right\}$ and $\hat{\mathcal{R}}_m(O) = ||\boldsymbol{Q}_{m,:}||_1$

for $m = 1, \ldots, p$ yields

$$\text{sign}\left\{\boldsymbol{Q}\right\}\boldsymbol{Q}^T = \text{diag}\left(||\boldsymbol{Q}_{1,:}||_1, \ldots, ||\boldsymbol{Q}_{p,:}||_1\right).$$

We define $\tilde{\mathbf{Q}} = \text{diag}\left(||\boldsymbol{Q}_{1,:}||_1, \ldots, ||\boldsymbol{Q}_{p,:}||_1\right)^{-1}\boldsymbol{Q}$ and obtain

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.9}$$

We provide some examples for the set of stationary points:

- *Perfect Separators:* If $\tilde{\mathbf{Q}} = \boldsymbol{P}\text{diag}(\sigma)$ where $\sigma \in \{-1, 1\}^p$ and $\boldsymbol{P}$ is a permutation matrix, then $\text{sign}\left\{\tilde{\mathbf{Q}}\right\} = \tilde{\mathbf{Q}}$ hence $\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}$. This yields $\boldsymbol{G}_* = \boldsymbol{D}\boldsymbol{P}$ hence the corresponding $\boldsymbol{G}_*$ is a perfect separator matrix.

- *Orthogonal matrices where each row has entries with constant magnitude:* Suppose that $\tilde{\mathbf{Q}}$ is an orthogonal matrix whose i'th row has α_i non-zero values and the magnitude of corresponding non-zero entries is $1/\alpha_i$. Therefore, we can write $\tilde{\mathbf{Q}} = \text{diag}(1/\alpha_1, 1/\alpha_2, \ldots, 1/\alpha_p)\text{sign}\left\{\tilde{\mathbf{Q}}\right\}$ This implies that $\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_p)\ \text{diag}(1/\alpha_1, 1/\alpha_2, \ldots, 1/\alpha_p) = \mathbf{I}$.

- *Matrices whose entries are powers of 0.5:* Defining $\mathbf{T} = \begin{bmatrix} (0.5)^{n-1} & (0.5)^{n-1} & (0.5)^{n-2} & \ldots & (0.5)^2 & 0.5 \\ (0.5)^{n-1} & (0.5)^{n-1} & (0.5)^{n-2} & & (0.5)^2 & -0.5 \\ (0.5)^{n-2} & (0.5)^{n-2} & (0.5)^{n-3} & & -0.5 & 0 \\ \vdots & \vdots & \vdots & & 0 & \vdots \\ (0.5)^2 & (0.5)^2 & -0.5 & & \vdots & 0 \\ 0.5 & -0.5 & 0 & \ldots & 0 & 0 \end{bmatrix}$ a set of stationary points can be defined of the form $\tilde{\mathbf{Q}} = \text{diag}(\sigma)\boldsymbol{P}_1\mathbf{T}\boldsymbol{P}_2$ where $\boldsymbol{P}_1$ and $\boldsymbol{P}_1$ are permutation matrices.

We note that we do not provide the set of all stationary points, however, we will show that if a stationary point of the algorithm (4.1)/(4.4) is not a perfect separator, then it is a saddle point.

Lemma : If a stationary point does not belong to the set of perfect separators, then its rows and columns can be permuted such that upper-left 2 by 2 sub-matrix of its sign matrix becomes

$$\mathbf{\Lambda}_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \text{ or } \mathbf{\Lambda}_{-1} = \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Proof : Let $\tilde{\mathbf{Q}}$ be a stationary point which is not a perfect separator. There exists a row of $\tilde{\mathbf{Q}}$ which has more than one non-zero entries (wlog, assume $\tilde{\mathbf{Q}}_{1,:}$). From (4.9), we have

(p.1) $\left\{\text{sign}\left\{\tilde{\mathbf{Q}}_{j,:}\right\}, \ j = 1, 2, \ldots, p\right\}$ are linearly independent.

(p.2) $\text{sign}\left\{\tilde{\mathbf{Q}}_{j,:}\right\}\tilde{\mathbf{Q}}_{1,:}^T = 0$ for $j = 2, \ldots, p$.

Note that **(p.1)** implies that at least one $\tilde{\mathbf{Q}}_{j,:}$ has a non-zero entry overlapping with one of the non-zero entries of $\tilde{\mathbf{Q}}_{1,:}$. Otherwise, non-overlap condition restricts span of $\left\{\text{sign}\left\{\tilde{\mathbf{Q}}_{j,:}\right\}, \ j = 2, \ldots, p\right\}$ to at most $p-2$ dimensional space which conflicts with linear independence.

Furthermore, **(p.2)** implies that number of overlapping entries should be greater than one with an alternating sign.

Therefore, rows and columns of a stationary point which is not a perfect separator can be permuted such that upper-left 2 by 2 sub-matrix of its sign matrix becomes $\mathbf{\Lambda}_1$ or $\mathbf{\Lambda}_{-1}$.

The following theorem shows that the stationary points other than perfect separators are saddle points:

Theorem 1: If a stationary point of the algorithm (4.1)/(4.4) does not belong to the set of perfect separators, then it is a saddle point.

Proof : We note that $\boldsymbol{G}$ is a perfect separator matrix implies that $\tilde{\mathbf{Q}}$ is a perfect separator matrix and vice versa. Therefore, it is equivalent to show that all $\tilde{\mathbf{Q}}$ matrices satisfying (4.9) are saddle points if they are not perfect separators.

We note that the cost function in terms of $\tilde{\mathbf{Q}}$ is equivalent to

$$J_1(\tilde{\mathbf{Q}}) = |det(\tilde{\mathbf{Q}})|.$$

From Lemma, $\tilde{\mathbf{Q}}$ can be permuted such that upper-left 2 by 2 sub-matrix of its sign matrix becomes $\mathbf{\Lambda}_1$ or $\mathbf{\Lambda}_{-1}$. We define the permuted matrix as $\breve{\mathbf{Q}}$ and wlog we assume $\text{sign}\left\{\breve{\mathbf{Q}}_{1:2,1:2}\right\} = \mathbf{\Lambda}_1$. We also observe that $J(\tilde{\mathbf{Q}}) = J(\breve{\mathbf{Q}})$.

We partition

$$\breve{\mathbf{Q}} = \begin{bmatrix} \breve{\mathbf{Q}}^{(a)} & \breve{\mathbf{Q}}^{(b)} \\ \breve{\mathbf{Q}}^{(c)} & \breve{\mathbf{Q}}^{(d)} \end{bmatrix}, \quad \begin{matrix} \breve{\mathbf{Q}}^{(a)} = \breve{\mathbf{Q}}_{1:2,1:2}, & \breve{\mathbf{Q}}^{(b)} = \breve{\mathbf{Q}}_{1:2,3:p} \\ \breve{\mathbf{Q}}^{(c)} = \breve{\mathbf{Q}}_{3:p,1:2}, & \breve{\mathbf{Q}}^{(d)} = \breve{\mathbf{Q}}_{3:p,3:p} \end{matrix}$$

Note that, $\text{sign}\left\{\breve{\mathbf{Q}}^{(a)}\right\} = \mathbf{\Lambda}_1$. We will prove that $\breve{\mathbf{Q}}^{(d)}$ is non-singular. From (4.9), we have

$$\begin{bmatrix} \breve{\mathbf{Q}}^{(a)} & \breve{\mathbf{Q}}^{(b)} \\ \breve{\mathbf{Q}}^{(c)} & \breve{\mathbf{Q}}^{(d)} \end{bmatrix} \begin{bmatrix} \text{sign}\left\{\breve{\mathbf{Q}}^{(a)^T}\right\} & \text{sign}\left\{\breve{\mathbf{Q}}^{(c)^T}\right\} \\ \text{sign}\left\{\breve{\mathbf{Q}}^{(b)^T}\right\} & \text{sign}\left\{\breve{\mathbf{Q}}^{(d)^T}\right\} \end{bmatrix} = \mathbf{I},$$

which yields

$$\breve{\mathbf{Q}}^{(c)}\text{sign}\left\{\breve{\mathbf{Q}}^{(a)^T}\right\} + \breve{\mathbf{Q}}^{(d)}\text{sign}\left\{\breve{\mathbf{Q}}^{(b)^T}\right\} = \mathbf{0}.$$

If $\breve{\mathbf{Q}}^{(d)}$ is singular, then there exists a non-zero vector $\mathbf{x} \in \Re^{p-2}$ such that $\mathbf{x}^T\breve{\mathbf{Q}}^{(d)} = \mathbf{0}$. Therefore,

$$\mathbf{x}^T\breve{\mathbf{Q}}^{(c)}\text{sign}\left\{\breve{\mathbf{Q}}^{(a)^T}\right\} = \mathbf{0},$$

which yields $\mathbf{x}^T\breve{\mathbf{Q}}^{(c)} = \mathbf{0}$ since $\text{sign}\left\{\breve{\mathbf{Q}}^{(a)}\right\} = \mathbf{\Lambda}_1$ is non-singular. Defining the non-zero vector $\hat{\mathbf{x}} \in \Re^p$ such that $\hat{\mathbf{x}} = [0\ 0\ \mathbf{x}^T]^T$, we have

$$\begin{bmatrix} \breve{\mathbf{Q}}^{(a)^T} & \breve{\mathbf{Q}}^{(c)^T} \\ \breve{\mathbf{Q}}^{(b)^T} & \breve{\mathbf{Q}}^{(d)^T} \end{bmatrix} \hat{\mathbf{x}} = \mathbf{0}.$$

This yields contradiction since $\breve{\mathbf{Q}}^T$ is non-singular. Therefore, $\breve{\mathbf{Q}}^{(d)}$ is non-singular.

We now prove that $\breve{\mathbf{Q}}$ is a saddle point. Using Schur's Complement, we have

$$J_1(\breve{\mathbf{Q}}) = \left|\det\left(\breve{\mathbf{Q}}^{(d)}\right)\det\left(\breve{\mathbf{Q}}^{(a)} - \breve{\mathbf{Q}}^{(b)}\left(\breve{\mathbf{Q}}^{(d)}\right)^{-1}\breve{\mathbf{Q}}^{(c)}\right)\right|.$$

Defining $\mathbf{\Delta} = \breve{\mathbf{Q}}^{(a)} - \breve{\mathbf{Q}}^{(b)}\left(\breve{\mathbf{Q}}^{(d)}\right)^{-1}\breve{\mathbf{Q}}^{(c)}$, we note that

$$\mathbf{\Delta}^{-1} = \left(\breve{\mathbf{Q}}^{-1}\right)_{1:2,1:2} = \text{sign}\left\{\tilde{\mathbf{Q}}^T\right\}_{1:2,1:2} = \mathbf{\Lambda}_1.$$

Hence we obtain $\mathbf{\Delta} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$.

- If we perturb $\breve{\mathbf{Q}}^{(a)}$ matrix with $\begin{bmatrix} \epsilon & -\epsilon \\ -\epsilon & -\epsilon \end{bmatrix}$, we then note that $\det(\breve{\mathbf{Q}}^{(d)})$ does not change and $|\det(\mathbf{\Delta})|$ becomes $|\det(\mathbf{\Delta})| + 2\epsilon^2$. Hence, we have $J_1(\tilde{\mathbf{Q}}_\epsilon) > J_1(\tilde{\mathbf{Q}})$.

- If we perturb $\breve{\mathbf{Q}}^{(a)}$ matrix with $\begin{bmatrix} \epsilon & -\epsilon \\ \epsilon & \epsilon \end{bmatrix}$, we then note that $\det(\breve{\mathbf{Q}}^{(d)})$ does not change and $|\det(\mathbf{\Delta})|$ becomes $|\det(\mathbf{\Delta})| - 2\epsilon^2$. Hence, we have $J_1(\tilde{\mathbf{Q}}_\epsilon) < J_1(\tilde{\mathbf{Q}})$.

Therefore, if a stationary point of the algorithm (4.1)-(4.4) does not belong in the set of perfect separators, then it is a saddle point.

4.3 Convergence Analysis Corresponding to Objective Function $J_{2,1}(\boldsymbol{W})$

Following similar steps as in the previous section, we have

$$\boldsymbol{G}^{(t+1)} = \boldsymbol{G}^{(t)} + \mu^{(t)} \Bigg(\left(\boldsymbol{G}^{(t)}\right)^{-T} - \sum_{m=1}^{p} \frac{p}{||\hat{\mathcal{R}}(O^{(t)})||_1} \mathbf{e}_m \\ \Bigg[\sum_{k_{m,+} \in \mathcal{K}_{m,+}(O^{(t)})} \lambda_{m,+}^{(t)}(k_{m,+}) \boldsymbol{s}^T(k_{m,+}) - \sum_{k_{m,-} \in \mathcal{K}_{m,-}(O^{(t)})} \lambda_{m,-}^{(t)}(k_{m,-}) \boldsymbol{s}^T(k_{m,-}) \Bigg] \Bigg) \boldsymbol{H}^T \boldsymbol{H}. \tag{4.10}$$

The stationary points in this case satisfies

$$\frac{p}{||\hat{\mathcal{R}}(O)||_1} \text{sign}\{\boldsymbol{G}_*\} (\boldsymbol{U} - \boldsymbol{L}) \boldsymbol{G}_*^T = \mathbf{I}. \tag{4.11}$$

Similarly, we define $\boldsymbol{Q} = \boldsymbol{G}_*(\boldsymbol{U} - \boldsymbol{L})$ with noting that $\text{sign}\{\boldsymbol{G}_*\} = \text{sign}\{\boldsymbol{Q}\}$ and $||\hat{\mathcal{R}}(O)||_1 = \sum_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_1$ yields

$$\text{sign}\{\boldsymbol{Q}\} \boldsymbol{Q}^T = \left(\frac{\sum_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_1}{p} \right) \mathbf{I}.$$

This implies that

$$||\boldsymbol{Q}_{1,:}||_1 = ||\boldsymbol{Q}_{2,:}||_1 = \ldots = ||\boldsymbol{Q}_{p,:}||_1.$$

We define $\tilde{\mathbf{Q}} = \frac{1}{||\boldsymbol{Q}_{1,:}||_1}\boldsymbol{Q}$ and obtain

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.12}$$

We note that we reach the same condition as (4.9) for the objective function $J_{2,1}(\boldsymbol{W})$, therefore, the examples of the $\tilde{\mathbf{Q}}$ matrices also applies here. However, the difference is that for the objective function $J_1(\boldsymbol{W})$ corresponding $\boldsymbol{Q}$ matrices can be obtained by arbitrary scaling of the rows of $\tilde{\mathbf{Q}}$ whereas here we should multiply the rows of $\tilde{\mathbf{Q}}$ matrices with the same parameter.

Similar to the objective function $J_1(\boldsymbol{W})$, we will show that if a stationary point of the algorithm (4.2)/(4.10) is not a perfect separator, then it is a saddle point.

Theorem 2: If a stationary point of the algorithm (4.2)-(4.10) does not belong to the set of perfect separators, then it is a saddle point.

Proof : In this case, we note that the cost function in terms of $\tilde{\mathbf{Q}}$ is equivalent to

$$J(\tilde{\mathbf{Q}}) = \frac{|det(\tilde{\mathbf{Q}})|}{p^p}.$$

Therefore, the proof of the **Theorem 1** also applies here.

4.4 Convergence Analysis Corresponding to Objective Function $J_{2,2}(\boldsymbol{W})$

For the objective function $J_{2,2}(\boldsymbol{W})$, we have

$$\boldsymbol{G}^{(t+1)} = \boldsymbol{G}^{(t)} + \mu^{(t)}\Bigg(\left(\boldsymbol{G}^{(t)}\right)^{-T} - \sum_{m=1}^{p} \frac{p\hat{\mathcal{R}}_m(O^{(t)})}{||\hat{\mathcal{R}}(O^{(t)})||_2^2}\mathbf{e}_m$$
$$\Bigg[\sum_{k_{m,+}\in\mathcal{K}_{m,+}(O^{(t)})} \lambda_{m,+}^{(t)}(k_{m,+})\boldsymbol{s}^T(k_{m,+}) - \sum_{k_{m,-}\in\mathcal{K}_{m,-}(O^{(t)})} \lambda_{m,-}^{(t)}(k_{m,-})\boldsymbol{s}^T(k_{m,-})\Bigg]\Bigg)\boldsymbol{H}^T\boldsymbol{H}. \tag{4.13}$$

The stationary points in this case satisfies

$$\frac{p}{||\hat{\mathcal{R}}(O)||_2^2}\mathrm{diag}(\hat{\mathcal{R}}_1(O),\ldots,\hat{\mathcal{R}}_p(O))\mathrm{sign}\left\{\boldsymbol{G}_*\right\}(\boldsymbol{U}-\boldsymbol{L})\boldsymbol{G}_*^T = \mathbf{I}.$$

Similarly, we define $\boldsymbol{Q} = \boldsymbol{G}_*(\boldsymbol{U}-\boldsymbol{L})$ and obtain

$$\mathrm{diag}\left(||\boldsymbol{Q}_{1,:}||_1,\ldots,||\boldsymbol{Q}_{p,:}||_1\right)\mathrm{sign}\left\{\boldsymbol{Q}\right\}\boldsymbol{Q}^T = \left(\frac{\sum_{m=1}^{p}||\boldsymbol{Q}_{m,:}||_1^2}{p}\right)\mathbf{I}.$$

This implies that

$$||\boldsymbol{Q}_{1,:}||_1 = ||\boldsymbol{Q}_{2,:}||_1 = \ldots = ||\boldsymbol{Q}_{p,:}||_1.$$

Similarly, defining $\tilde{\mathbf{Q}} = \frac{1}{||\boldsymbol{Q}_{1,:}||_1}\boldsymbol{Q}$ yields

$$\mathrm{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.14}$$

We note that this condition is equivalent to the condition for the objective function $J_{2,1}(\boldsymbol{W})$.

4.5 Convergence Analysis Corresponding to Objective Function $J_{2,\infty}(\boldsymbol{W})$

For the objective function $J_{2,\infty}(\boldsymbol{W})$, we have

$$\boldsymbol{G}^{(t+1)} = \boldsymbol{G}^{(t)} + \mu^{(t)}\Bigg(\left(\boldsymbol{G}^{(t)}\right)^{-T} - \sum_{m\in\mathcal{M}(O^{(t)})} \frac{p\beta_m^{(t)}}{||\hat{\mathcal{R}}(O^{(t)})||_\infty}\mathbf{e}_m \\ \Bigg[\sum_{k_{m,+}\in\mathcal{K}_{m,+}(O^{(t)})} \lambda_{m,+}^{(t)}(k_{m,+})\boldsymbol{s}^T(k_{m,+}) - \sum_{k_{m,-}\in\mathcal{K}_{m,-}(O^{(t)})} \lambda_{m,-}^{(t)}(k_{m,-})\boldsymbol{s}^T(k_{m,-})\Bigg]\Bigg)\boldsymbol{H}^T\boldsymbol{H}. \tag{4.15}$$

The stationary points satisfies

$$\sum_{m\in\mathcal{M}(O)} \frac{p\beta_m}{||\hat{\mathcal{R}}(O)||_\infty}\mathbf{e}_m \text{sign}\left\{(\boldsymbol{G}_*)_{m,:}\right\}(\boldsymbol{U}-\boldsymbol{L})\boldsymbol{G}_*^T = \mathbf{I}.$$

Defining $\boldsymbol{Q} = \boldsymbol{G}_*(\boldsymbol{U}-\boldsymbol{L})$ yields

$$\sum_{m\in\mathcal{M}(O)} \frac{p\beta_m}{||\hat{\mathcal{R}}(O)||_\infty}\mathbf{e}_m \text{sign}\left\{\boldsymbol{Q}_{m,:}\right\}\boldsymbol{Q}^T = \mathbf{I}. \tag{4.16}$$

We note that in order to satisfy (4.29), we must have $\mathcal{M}(O) = \{1, 2, \ldots, p\}$ which implies that the ranges of outputs are equal, i.e.,

$$||\boldsymbol{Q}_{1,:}||_1 = ||\boldsymbol{Q}_{2,:}||_1 = \ldots = ||\boldsymbol{Q}_{p,:}||_1.$$

Hence, $\beta_m = 1/p$ for $m = 1, 2, \ldots, p$. We similarly define $\tilde{\mathbf{Q}} = \frac{1}{||\boldsymbol{Q}_{1,:}||_1}\boldsymbol{Q}$ and obtain from (4.29) that

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.17}$$

We note that this condition is equivalent to the condition for the objective function $J_{2,1}(\boldsymbol{W})$.

4.6 Extension to Complex Signals

In the complex case, the source vectors and output vectors belong to $\mathbb{C}^p$ and the mixture vectors belong to $\mathbb{C}^q$. The mixing and separator matrices are complex matrices, i.e., $\boldsymbol{H} = \mathbb{C}^{q\times p}$ and $\boldsymbol{W} = \mathbb{C}^{p\times q}$. For a given complex vector $\mathbf{x} \in \mathbb{C}^p$, we define the corresponding isomorphic real vector $\grave{\mathbf{x}} \in \mathbb{R}^{2p}$ as $\grave{\mathbf{x}} = \begin{bmatrix} \mathbb{R}(\mathbf{x}^T) & \mathbb{I}(\mathbf{x}^T) \end{bmatrix}^T$. We also define the operator $\psi : \mathbb{C}^{p\times q} \to \mathbb{R}^{2p\times 2q}$ as

$$\psi(\mathbf{X}) = \begin{bmatrix} \mathbb{R}(\mathbf{X}) & -\mathbb{I}(\mathbf{X}) \\ \mathbb{I}(\mathbf{X}) & \mathbb{R}(\mathbf{X}) \end{bmatrix}.$$

We note that since $\boldsymbol{o} = \boldsymbol{W}\boldsymbol{y}$, we have $\grave{\mathbf{o}} = \psi(\boldsymbol{W})\grave{\mathbf{y}}$.

Using these definitions, J_1 objective function in (3.1) has been modified for the complex case as

$$Jc_1(\psi(\boldsymbol{W})) = C_p \frac{\sqrt{\det(\hat{\boldsymbol{R}}_{\grave{\mathbf{o}}})}}{\prod \hat{\mathcal{R}}(\grave{O})}, \tag{4.18}$$

where in this case $\grave{O} = \{\grave{\mathbf{o}}(1), \grave{\mathbf{o}}(2), \ldots, \grave{\mathbf{o}}(N)\}$.

For the modified objective function, it is proved that the global optima for (4.18) (in terms of $\boldsymbol{G}$) is given by

$$\begin{aligned} O_c = \{ & \boldsymbol{G} = \boldsymbol{DP} : \boldsymbol{P} \in \mathbb{R}^{p \times p} \text{ is a permutation matrix}, \\ & \boldsymbol{D} \in \mathbb{I}^{p \times p} \text{ is a full rank diagonal matrix with}, \\ & D_{ii} = \alpha_i e^{\frac{j\pi k_i}{2}}, \alpha_i \in \mathbb{R}, k_i \in \mathbb{Z}, i = 1, \ldots, p\}, \end{aligned}$$

which corresponds to a subset of complex perfect separators with discrete phase ambiguity.

Similar to the real case, complex approach is extended to the Jc_2 family by defining

$$Jc_{2,r}(\psi(\boldsymbol{W})) = C_p \frac{\sqrt{\det(\hat{\boldsymbol{R}}_{\grave{\mathbf{o}}})}}{||\hat{\mathcal{R}}(\grave{O})||_r^{2p}}. \tag{4.19}$$

The corresponding iterative updates for $\psi(\boldsymbol{W})$ can similarly be written as

- Objective function $Jc_1(\psi(\boldsymbol{W}))$:

  $$\begin{aligned} \psi(\boldsymbol{W}^{(t+1)}) = \psi(\boldsymbol{W}^{(t)}) + \mu^{(t)} \Bigg(& \left(\psi(\boldsymbol{W}^{(t)}) \hat{\mathbf{R}}(\grave{Y}) \psi(\boldsymbol{W}^{(t)})^T \right)^{-1} \psi(\boldsymbol{W}^{(t)}) \hat{\mathbf{R}}(\grave{Y}) \\ & - \sum_{m=1}^{2p} \frac{1}{\hat{\mathcal{R}}_m(\grave{O}^{(t)})} \mathbf{e}_m \boldsymbol{b}_m^{(t)T} \Bigg), \end{aligned} \tag{4.20}$$

  where

  $$\boldsymbol{b}_m^{(t)} = \sum_{k_{m,+} \in \mathcal{K}_{m,+}(Z^{(t)})} \lambda_{m,+}^{(t)}(k_{m,+}) \grave{\mathbf{y}}(k_{m,+}) - \sum_{k_{m,-} \in \mathcal{K}_{m,-}(Z^{(t)})} \lambda_{m,-}^{(t)}(k_{m,-}) \grave{\mathbf{y}}(k_{m,-}),$$

  for $m = 1, 2, \ldots, 2p$.

- Objective function $Jc_{2,r}(\psi(\boldsymbol{W}))$:

For $r = 1, 2$, the update equation is

$$\psi(\boldsymbol{W}^{(t+1)}) = \psi(\boldsymbol{W}^{(t)}) + \mu^{(t)} \Bigg(\left(\psi(\boldsymbol{W}^{(t)}) \hat{\mathbf{R}}(\grave{Y}) \psi(\boldsymbol{W}^{(t)})^T \right)^{-1} \psi(\boldsymbol{W}^{(t)}) \hat{\mathbf{R}}(\grave{Y}) \\ - \sum_{m=1}^{2p} \frac{2p \hat{\mathcal{R}}_m(\grave{O}^{(t)})^{r-1}}{||\hat{\mathcal{R}}(\grave{O}^{(t)})||_r^r} \mathbf{e}_m \boldsymbol{b}_m^{(t)T} \Bigg). \tag{4.21}$$

For $r = \infty$, the update equation is

$$\psi(\boldsymbol{W}^{(t+1)}) = \psi(\boldsymbol{W}^{(t)}) + \mu^{(t)} \Bigg(\left(\psi(\boldsymbol{W}^{(t)}) \hat{\mathbf{R}}(\grave{Y}) \psi(\boldsymbol{W}^{(t)})^T \right)^{-1} \psi(\boldsymbol{W}^{(t)}) \hat{\mathbf{R}}(\grave{Y}) \\ - \sum_{m \in \mathcal{M}(\grave{O}^{(t)})} \frac{2p \beta_m^{(t)}}{||\hat{\mathcal{R}}(\grave{O}^{(t)})||_\infty} \mathbf{e}_m \boldsymbol{b}_m^{(t)T} \Bigg). \tag{4.22}$$

4.7 Convergence Analysis Corresponding to Objective Function $Jc_1(\psi(\boldsymbol{W}))$

We rewrite the iterative algorithm (4.20) in terms of $\psi(\boldsymbol{G})$ as

$$\psi(\boldsymbol{G}^{(t+1)}) = \psi(\boldsymbol{G}^{(t)}) + \mu^{(t)} \Bigg(\psi(\boldsymbol{G}^{(t)})^{-T} - \sum_{m=1}^{2p} \frac{1}{\hat{\mathcal{R}}_m(\grave{O}^{(t)})} \mathbf{e}_m \\ \Bigg[\sum_{k_{m,+} \in \mathcal{K}_{m,+}(\grave{O}^{(t)})} \lambda_{m,+}^{(t)}(k_{m,+}) \grave{\mathbf{s}}^T(k_{m,+}) - \sum_{k_{m,-} \in \mathcal{K}_{m,-}(\grave{O}^{(t)})} \lambda_{m,-}^{(t)}(k_{m,-}) \grave{\mathbf{s}}^T(k_{m,-}) \Bigg] \Bigg) \psi(\boldsymbol{H})^T \psi(\boldsymbol{H}). \tag{4.23}$$

We can similarly write the input dependent update part in the right side of (4.23) as

$$\text{sign}\left\{ \psi(\boldsymbol{G}^{(t)})_{m,:} \right\} (\boldsymbol{U}_T - \boldsymbol{L}_T), \tag{4.24}$$

where $\boldsymbol{U}_T = \begin{bmatrix} \boldsymbol{U}_R & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_I \end{bmatrix}$ and $\boldsymbol{L}_T = \begin{bmatrix} \boldsymbol{L}_R & \mathbf{0} \\ \mathbf{0} & \boldsymbol{L}_I \end{bmatrix}$ such that

$$\boldsymbol{U}_R = \text{diag}\left(\max(\mathbb{R}(s_1)), \max(\mathbb{R}(s_2)), \ldots, \max(\mathbb{R}(s_p))\right),$$
$$\boldsymbol{L}_R = \text{diag}\left(\min(\mathbb{R}(s_1)), \min(\mathbb{R}(s_2)), \ldots, \min(\mathbb{R}(s_p))\right),$$
$$\boldsymbol{U}_I = \text{diag}\left(\max(\mathbb{I}(s_1)), \max(\mathbb{I}(s_2)), \ldots, \max(\mathbb{I}(s_p))\right),$$
$$\boldsymbol{L}_I = \text{diag}\left(\min(\mathbb{I}(s_1)), \min(\mathbb{I}(s_2)), \ldots, \min(\mathbb{I}(s_p))\right).$$

Here we assume that $\max(\mathbb{R}(s_i)) > 0$, $\min(\mathbb{R}(s_i)) < 0$, $\max(\mathbb{I}(s_i)) > 0$ and $\min(\mathbb{I}(s_i)) < 0$ for $i = 1, 2, \ldots, p$. Therefore, the stationary points satisfies

$$\text{diag}(\hat{\mathcal{R}}_1(\grave{O}), \ldots, \hat{\mathcal{R}}_{2p}(\grave{O}))^{-1}\text{sign}\left\{\psi(\boldsymbol{G})_*\right\}(\boldsymbol{U}_T - \boldsymbol{L}_T)\psi(\boldsymbol{G})_*^T = \mathbf{I}. \tag{4.25}$$

Defining $\boldsymbol{Q} = \psi(\boldsymbol{G})_*(\boldsymbol{U}_T - \boldsymbol{L}_T)$ with noting that $\text{sign}\left\{\psi(\boldsymbol{G})_*\right\} = \text{sign}\left\{\boldsymbol{Q}\right\}$ and $\hat{\mathcal{R}}_m(\grave{O}) = ||\boldsymbol{Q}_{m,:}||_1$ for $m = 1, \ldots, 2p$ yields

$$\text{sign}\left\{\boldsymbol{Q}\right\}\boldsymbol{Q}^T = \text{diag}\left(||\boldsymbol{Q}_{1,:}||_1, \ldots, ||\boldsymbol{Q}_{2p,:}||_1\right).$$

We define $\tilde{\mathbf{Q}} = \text{diag}\left(||\boldsymbol{Q}_{1,:}||_1, \ldots, ||\boldsymbol{Q}_{2p,:}||_1\right)^{-1}\boldsymbol{Q}$ and obtain

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.26}$$

We first provide the set of perfect separators in this case and then prove that if a stationary point of the algorithm (4.20)/(4.23) is not a perfect separator, then it is a saddle point.

- *Perfect Separators:* We note that in this case, due to the structure of $\psi(\boldsymbol{G})$, positions of non-zero values of $\tilde{\mathbf{Q}}_{:,1:p}$ suffices to have the positions of non-zero values of $\tilde{\mathbf{Q}}$. If $\tilde{\mathbf{Q}}_{1:p,:} = \boldsymbol{P}\text{diag}(\sigma)$, then sign $\left\{\tilde{\mathbf{Q}}\right\} = \tilde{\mathbf{Q}}$ hence sign $\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}$. This yields $\boldsymbol{G}_* = \boldsymbol{D}\boldsymbol{P}$ where $D_{ii} = \alpha_i e^{\frac{j\pi k_i}{2}}, \alpha_i \in \mathbb{R}, k_i \in \mathbb{Z}, i = 1, \ldots, p$. We note that the entries of $\boldsymbol{G}$ matrices' can only be real or purely imaginary due to the structure of $\psi(\boldsymbol{G})$.

Theorem 3: If a stationary point of the algorithm (4.20)/(4.23) does not belong to the set of perfect separators, then it is a saddle point.

Proof : We note that $\boldsymbol{G}$ is a perfect separator matrix implies that $\tilde{\mathbf{Q}}$ is a perfect separator matrix and vice versa. Therefore, it is equivalent to show that all $\tilde{\mathbf{Q}}$ matrices satisfying (4.26) are saddle points if they are not perfect separators.

From (4.18), the cost function in terms of $\tilde{\mathbf{Q}}$ is equivalent to

$$Jc_1(\tilde{\mathbf{Q}}) = |det(\tilde{\mathbf{Q}})|.$$

Therefore, the proof of the **Theorem 1** also applies here.

4.8 Convergence Analysis Corresponding to Objective Function $Jc_{2,1}(\psi(\boldsymbol{W}))$

Following similar steps as in the previous section, the stationary points satisfy

$$\frac{2p}{||\hat{\mathcal{R}}(\grave{O})||_1}\text{sign}\left\{\psi(\boldsymbol{G})_*\right\}(\boldsymbol{U}_T - \boldsymbol{L}_T)\psi(\boldsymbol{G})_*^T = \mathbf{I}.$$

We define $\boldsymbol{Q} = \psi(\boldsymbol{G})_*(\boldsymbol{U}_T - \boldsymbol{L}_T)$ with noting that $\text{sign}\{\psi(\boldsymbol{G})_*\} = \text{sign}\{\boldsymbol{Q}\}$ and $||\hat{\mathcal{R}}(\grave{O})||_1 = \sum_{m=1}^{2p} ||\boldsymbol{Q}_{m,:}||_1$ and obtain

$$\text{sign}\{\boldsymbol{Q}\}\,\boldsymbol{Q}^T = \left(\frac{\sum_{m=1}^{2p} ||\boldsymbol{Q}_{m,:}||_1}{2p}\right)\mathbf{I}.$$

This implies that

$$||\boldsymbol{Q}_{1,:}||_1 = ||\boldsymbol{Q}_{2,:}||_1 = \ldots = ||\boldsymbol{Q}_{2p,:}||_1.$$

We define $\tilde{\mathbf{Q}} = \frac{1}{||\boldsymbol{Q}_{1,:}||_1}\boldsymbol{Q}$ and obtain

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.27}$$

We note that this is the same condition (4.26) for the objective function $Jc_{2,1}(\psi(\boldsymbol{W}))$, therefore, the derivations also apply here. However, the difference is that for the objective function $Jc_1(\psi(\boldsymbol{W}))$, corresponding $\boldsymbol{Q}$ matrices can be obtained by arbitrary scaling of the rows of $\tilde{\mathbf{Q}}$ whereas here we should multiply the rows of $\tilde{\mathbf{Q}}$ matrices with the same parameter.

4.9 Convergence Analysis Corresponding to Objective Function $Jc_{2,2}(\psi(\boldsymbol{W}))$

The stationary points in this case satisfies

$$\frac{2p}{||\hat{\mathcal{R}}(\grave{O})||_2^2}\text{diag}(\hat{\mathcal{R}}_1(\grave{O}), \ldots, \hat{\mathcal{R}}_{2p}(\grave{O}))\text{sign}\{\psi(\boldsymbol{G})_*\}(\boldsymbol{U}_T - \boldsymbol{L}_T)\psi(\boldsymbol{G})_*^T = \mathbf{I}.$$

Similarly, we define $\boldsymbol{Q} = \psi(\boldsymbol{G})_*(\boldsymbol{U}_T - \boldsymbol{L}_T)$ and obtain

$$\text{diag}\left(||\boldsymbol{Q}_{1,:}||_1, \ldots, ||\boldsymbol{Q}_{2p,:}||_1\right) \text{sign}\left\{\boldsymbol{Q}\right\} \boldsymbol{Q}^T = \left(\frac{\sum_{m=1}^{2p} ||\boldsymbol{Q}_{m,:}||_1^2}{2p}\right) \mathbf{I}.$$

This implies that

$$||\boldsymbol{Q}_{1,:}||_1 = ||\boldsymbol{Q}_{2,:}||_1 = \ldots = ||\boldsymbol{Q}_{2p,:}||_1.$$

Similarly, defining $\tilde{\mathbf{Q}} = \frac{1}{||\boldsymbol{Q}_{1,:}||_1}\boldsymbol{Q}$ yields

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\} \tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.28}$$

We note that this condition is equivalent to the condition for the objective function $J_{2,1}(\psi(\boldsymbol{W}))$.

4.10 Convergence Analysis Corresponding to Objective Function $Jc_{2,\infty}(\psi(\boldsymbol{W}))$

The stationary points satisfies

$$\sum_{m \in \mathcal{M}(\grave{O})} \frac{2p\beta_m}{||\hat{\mathcal{R}}(\grave{O})||_\infty} \mathbf{e}_m \text{sign}\left\{(\psi(\boldsymbol{G})_*)_{m,:}\right\} (\boldsymbol{U}_T - \boldsymbol{L}_T)\psi(\boldsymbol{G})_*^T = \mathbf{I}.$$

Defining $\boldsymbol{Q} = \psi(\boldsymbol{G})_*(\boldsymbol{U}_T - \boldsymbol{L}_T)$ yields

$$\sum_{m \in \mathcal{M}(\grave{O})} \frac{2p\beta_m}{||\hat{\mathcal{R}}(\grave{O})||_\infty} \mathbf{e}_m \text{sign}\left\{\boldsymbol{Q}_{m,:}\right\} \boldsymbol{Q}^T = \mathbf{I}. \tag{4.29}$$

We note that in order to satisfy (4.29), we must have $\mathcal{M}(\grave{O}) = \{1, 2, \ldots, 2p\}$ which implies that the ranges of outputs are equal, i.e.,

$$||\boldsymbol{Q}_{1,:}||_1 = ||\boldsymbol{Q}_{2,:}||_1 = \ldots = ||\boldsymbol{Q}_{2p,:}||_1.$$

Hence, $\beta_m = 1/2p$ for $m = 1, 2, \ldots, 2p$. We similarly define $\tilde{\mathbf{Q}} = \frac{1}{||\boldsymbol{Q}_{1,:}||_1}\boldsymbol{Q}$ and obtain from (4.29) that

$$\text{sign}\left\{\tilde{\mathbf{Q}}\right\}\tilde{\mathbf{Q}}^T = \mathbf{I}. \tag{4.30}$$

We note that this condition is equivalent to the condition for the objective function $J_{2,1}(\psi(\boldsymbol{W}))$.

4.11 Conclusion

In this section, we presented the stationary points of instantaneous BCA algorithms introduced in [1]. We provide some examples for the set of stationary points, however, we prove that all the stationary points of the algorithms are saddle points except perfect separators for each algorithm.

Chapter 5

Extension of Instantaneous BCA Approach

In this chapter, we extend the instantaneous BCA approach introduced in [1] by considering generalized functions of ranges of separator outputs. We recall that the approach in [1] exploits two geometric objects defined on output samples which are principal hyper-ellipsoid and bounding hyper-rectangle. The approach is the optimization of the relative sizes of these objects where the volume and the main diagonal length is considered to determine the size of bounding hyper-rectangle. In this chapter, we consider more general functions of ranges of output samples (corresponding to the side lengths of bounding hyper-rectangle) which also covers the size of bounding hyper-rectangle. It has been proved that when the assumption (**A1**) holds, the global maxima of the introduced objective functions which are the perfect separators are reached. However, in some real world applications, this assumption may not hold or we may not know if this assumption holds or not which can cause a variation in the performances of the introduced algorithms. Therefore, in this chapter, we define a more general optimization framework and correspondingly a variety of objective functions and prove that the corresponding global maxima are in the set of perfect separators under same conditions. Hence, with this approach, we are able to generate a variety of instantaneous BCA algorithms that can be exploited to obtain better performances in different applications.

5.1 Extended BCA Optimization Framework

We provide the updated instantaneous BCA optimization framework by considering the objective functions of [1] in a general case as

$$\boxed{J(\boldsymbol{W}) = \frac{\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{O}})}}{f\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right)},} \tag{5.1}$$

where $\hat{\boldsymbol{R}}_{\boldsymbol{O}} = \frac{1}{N}\sum_{l=1}^{N}(\boldsymbol{o}(l) - \hat{\boldsymbol{\mu}}(\boldsymbol{o}))(\boldsymbol{o}(l) - \hat{\boldsymbol{\mu}}(\boldsymbol{o}))^T$, $\hat{\boldsymbol{\mu}}(\boldsymbol{o}) = \frac{1}{N}\sum_{l=1}^{N}\boldsymbol{o}(l)$ is the sample covariance matrix of $\boldsymbol{o}$, $\hat{\mathcal{R}}(o_m)$ is the range of the m'th component of the vectors in the set O and f is any function that satisfies the following:

$$f\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) \geq c_p \prod_{m=1}^{p} \hat{\mathcal{R}}(o_m), \tag{5.2}$$

such that the equality is achievable for a finite constant c_p and the ranges $\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2)$, $\ldots, \hat{\mathcal{R}}(o_p)$ with some specific requirements. In the proposed objective function, the modification is in the denominator where we consider generalized functions of ranges of outputs. Here, we recall the assumption regarding the set S:

Assumption: S contains the vertices of its (non-degenerate) bounding hyper-rectangle (**A1**).

In the following theorem, we show that the global maxima of the objective function (5.1) correspond to the perfect separators.

Theorem: Assuming the setup in Section 2.2, $\boldsymbol{H}$ is a full rank matrix and $\hat{\boldsymbol{R}}_{\boldsymbol{S}} \succ 0$, the set of global maxima for J in (5.1) is equal to a set of perfect separator matrices.

Proof: We first note that since $\boldsymbol{o} = \boldsymbol{G}\boldsymbol{s}$, we have $\hat{\boldsymbol{R}}_{\boldsymbol{O}} = \boldsymbol{G}\hat{\boldsymbol{R}}_{\boldsymbol{S}}\boldsymbol{G}^T$ where $\hat{\boldsymbol{R}}_{\boldsymbol{S}}$ is the sample covariance matrix of $\boldsymbol{s}$. Therefore, $\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{O}})} = |det(\boldsymbol{G})|\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{S}})}$.

When assumption **(A1)** holds, we can write the range of m^{th} component of $\boldsymbol{o}$ as $\hat{\mathcal{R}}(o_m) = ||\boldsymbol{G}_{m,:}\boldsymbol{\Upsilon}||_1$ where $\boldsymbol{G}_{m,:}$ is the m^{th} row of $\boldsymbol{G}$ and $\boldsymbol{\Upsilon} = diag(\hat{\mathcal{R}}(s_1), \hat{\mathcal{R}}(s_2), \ldots, \hat{\mathcal{R}}(s_p))$ is the diagonal matrix containing range values of the source samples in the set S. We can further define $\boldsymbol{A} = \boldsymbol{G}\boldsymbol{\Upsilon}$ and write the objective function (5.1) in terms of $\boldsymbol{A}$ as

$$J(\boldsymbol{W}) = \frac{|det(\boldsymbol{A}\boldsymbol{\Upsilon}^{-1})|\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{s}})}}{f\left(||\boldsymbol{A}_{1,:}||_1, ||\boldsymbol{A}_{2,:}||_1, \ldots, ||\boldsymbol{A}_{p,:}||_1\right)},$$
$$= \frac{\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{s}})}}{\prod_{m=1}^{p} \hat{\mathcal{R}}(s_m)} \frac{|det(\boldsymbol{A})|}{f\left(||\boldsymbol{A}_{1,:}||_1, ||\boldsymbol{A}_{2,:}||_1, \ldots, ||\boldsymbol{A}_{p,:}||_1\right)}$$

Using the Hadamard inequality [26] and the ordering $||\boldsymbol{q}||_1 \geq ||\boldsymbol{q}||_2$ for any $\boldsymbol{q}$ yields

$$det(\boldsymbol{A}) \leq \prod_{m=1}^{p} ||\boldsymbol{A}_{m,:}||_2 \tag{5.3}$$
$$\leq \prod_{m=1}^{p} ||\boldsymbol{A}_{m,:}||_1. \tag{5.4}$$

Since the function f satisfies (5.2), we obtain

$$\frac{|det(\boldsymbol{A})|}{f\left(||\boldsymbol{A}_{1,:}||_1, ||\boldsymbol{A}_{2,:}||_1, \ldots, ||\boldsymbol{A}_{p,:}||_1\right)} \leq \frac{\prod_{m=1}^{p} ||\boldsymbol{A}_{m,:}||_1}{c_p \prod_{m=1}^{p} ||\boldsymbol{A}_{m,:}||_1},$$

which further implies

$$J(\boldsymbol{W}) \leq \frac{1}{c_p} \frac{\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{s}})}}{\prod_{m=1}^{p} \hat{\mathcal{R}}(s_m)}. \tag{5.5}$$

To achieve the equality in (5.5), the equalities in (5.2), (5.3) and (5.4) must be achieved. The equality in (5.3) is achieved if and only if the rows of $\boldsymbol{A}$ are orthogonal to each other and the equality in (5.4) is achieved if and only if the rows of $\boldsymbol{A}$ align with the coordinate axes. Therefore, the equality in (5.3) and (5.4) holds if and only if $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}$, hence, $\boldsymbol{G} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{\Upsilon}^{-1}$ where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{D}$ is a nonsingular diagonal matrix which corresponds to the perfect separators. Hence, with the specific requirement of (5.2) to achieve the equality in (5.5), the global maxima of the objective function (5.1) correspond to the perfect separators or a subset of perfect separators.

We here give some examples for the function f:

- $f_1\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) = \prod_{m=1}^{p} \hat{\mathcal{R}}(o_m)$:
  This is a trivial example where the equality in (5.2) is achieved for $c_p = 1$. Hence, the global maxima is achieved when $\boldsymbol{G} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{\Upsilon}^{-1}$ which corresponds to perfect separators. We note that this function corresponds to the volume of the bounding hyperrectangle of outputs and is equivalent to the objective function $J_1^{(\boldsymbol{W})}(\boldsymbol{W})$ of [1].

- $f_{2,r}\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) = \left\| \begin{bmatrix} \hat{\mathcal{R}}(o_1) & \hat{\mathcal{R}}(o_2) & \ldots & \hat{\mathcal{R}}(o_p) \end{bmatrix}^T \right\|_r^p$ where $r \geq 1$:
  In this example, due to the ordering $||\boldsymbol{q}||_r \geq p^{\frac{1-r}{r}} ||\boldsymbol{q}||_1$ for any $\boldsymbol{q} \in \Re^p$ and the Arithmetic-Geometric-Mean-Inequality yields

$$
\begin{aligned}
f_{2,r}\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) &\geq p^{\frac{p(1-r)}{r}} \left\| \begin{bmatrix} \hat{\mathcal{R}}(o_1) & \hat{\mathcal{R}}(o_2) & \ldots & \hat{\mathcal{R}}(o_p) \end{bmatrix}^T \right\|_1^p, \\
&\geq p^{\frac{p}{r}} \prod_{m=1}^{p} \hat{\mathcal{R}}(o_m),
\end{aligned}
$$

  where the equality is achieved when $c_p = p^{\frac{p}{r}}$ and the ranges $\hat{\mathcal{R}}(o_m)$ for $m = 1, 2, \ldots p$ are equal to each other or equivalently, the non-zero entries in the rows of $\boldsymbol{A}$ are equal in magnitude. Hence, the global maxima is achieved when $\boldsymbol{G} = k\text{diag}(\rho)\boldsymbol{P}\boldsymbol{\Upsilon}^{-1}$ where k is a non-zero value and $\rho \in \{-1, 1\}^p$ which correspond to a subset of perfect separators. We note that these functions correspond to the length of the main diagonal of the bounding hyperrectangle of outputs and are equivalent to the objective functions $J_{2,r}^{(\boldsymbol{W})}(\boldsymbol{W})$ of [1] for $r = 1, 2, \infty$.

- $f_3\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \hat{\mathcal{R}}(o_3)\right) = \left(2\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_2) + 2\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_3) + 2\hat{\mathcal{R}}(o_2)\hat{\mathcal{R}}(o_3)\right)^{\frac{3}{2}}$:

This example illustrates the surface area of the bounding hyperrectange of outputs for $p = 3$. Using the Arithmetic-Geometric-Mean-Inequality (AGMI) yields

$$f_3\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \hat{\mathcal{R}}(o_3)\right) \geq 6^{3/2} \prod_{m=1}^{3} \hat{\mathcal{R}}(o_m),$$

where the equality is achieved in the same condition with the norm example for $c_p = 6^{3/2}$. We can generalize this by choosing $f_3\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) = \left(\sum_{t=1}^{p} \hat{\mathcal{R}}(o_1)^{m_{t,1}} \hat{\mathcal{R}}(o_2)^{m_{t,2}} \ldots \hat{\mathcal{R}}(o_p)^{m_{t,p}}\right)^{p/x}$ where $\sum_{t=1}^{p} m_{t,j} = x$ for $j = 1, 2, \ldots p$ and $x \in \Re^+$. According to the equality requirement of AGMI, the global maxima correspond to a subset of perfect separators.

- $f_4\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) = log\left(e^{\hat{\mathcal{R}}(o_1)} + e^{\hat{\mathcal{R}}(o_2)} + \ldots + e^{\hat{\mathcal{R}}(o_p)}\right)^{2p}$:
In this case, using the AGMI yields

$$\begin{aligned} f_4\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) &\geq log\left(p\left(e^{\hat{\mathcal{R}}(o_1)+\hat{\mathcal{R}}(o_2)+\ldots+\hat{\mathcal{R}}(o_p)}\right)^{1/p}\right)^{2p}, \\ &= \left(log(p) + \frac{1}{p}\left(\hat{\mathcal{R}}(o_1) + \hat{\mathcal{R}}(o_2) + \ldots + \hat{\mathcal{R}}(o_p)\right)\right)^{2p}, \\ &\geq \left(log(p) + \left(\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_2)\ldots\hat{\mathcal{R}}(o_p)\right)^{1/p}\right)^{2p}, \\ &\geq 2^{2p} log(p)^p \prod_{m=1}^{p} \hat{\mathcal{R}}(o_m), \end{aligned}$$

where the equality is achieved when $c_p = 2^{2p} log(p)^p$ and the ranges $\hat{\mathcal{R}}(o_m)$ for $m = 1, 2, \ldots p$ are equal to each other and $log(p) = \left(\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_2)\ldots\hat{\mathcal{R}}(o_p)\right)^{1/p}$, yielding $\hat{\mathcal{R}}(o_m) = log(p)$ for $m = 1, 2, \ldots p$. Hence, the global maxima is achieved when $\boldsymbol{G} = log(p)\boldsymbol{P\Upsilon}^{-1}$ which correspond to a subset of perfect separators.

5.2 Adaptive Implementations

In this section, we provide the adaptive algorithms corresponding to the examples of objective functions presented in the previous section. We note that rather than maximizing J , we maximize its logarithm since with the logarithm operation, we utilize the conversion of ratio expression to the difference expression since it simplifies the update components in the iterative algorithm. Therefore, the new objective function is modified as

$$\bar{J}(\boldsymbol{W}) = log\left(J(\boldsymbol{W})\right) = \frac{1}{2} log\left(det\left(\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^T\right)\right) - log\left(f\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right)\right).$$

The derivative of the first part of $\bar{J}(\boldsymbol{W})$ with respect to $\boldsymbol{W}$ is

$$\frac{1}{2}\frac{\partial log\left(det\left(\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^T\right)\right)}{\partial \boldsymbol{W}} = \left(\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^T\right)^{-1}\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}.$$

We note that since f_1 and $f_{2,r}$ functions for $r = 1, 2, \infty$ are covered in [1], we only provide the adaptive algorithms for f_3 and f_4 functions.

- Iterative algorithm for f_3:

  The subgradient based adaptive algorithm maximizing $\bar{J}(\boldsymbol{W})$ using the function f_3 can be written as

$$\boxed{\begin{aligned}\boldsymbol{W}^{(i+1)} = \boldsymbol{W}^{(i)} + \mu^{(i)}\Bigg(&\left(\boldsymbol{W}^{(i)}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^{(i)T}\right)^{-1}\boldsymbol{W}^{(i)}\hat{\boldsymbol{R}}_{\boldsymbol{y}} - \\ &\frac{3}{2}\sum_{m=1}^{3} g_m \mathbf{e}_m \left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^T\Bigg)\end{aligned}}$$

  where $g_m = \dfrac{\hat{\mathcal{R}}(o_1^{(i)}) + \hat{\mathcal{R}}(o_2^{(i)}) + \hat{\mathcal{R}}(o_3^{(i)}) - \hat{\mathcal{R}}(o_m^{(i)})}{\hat{\mathcal{R}}(o_1^{(i)})\hat{\mathcal{R}}(o_2^{(i)}) + \hat{\mathcal{R}}(o_1^{(i)})\hat{\mathcal{R}}(o_3^{(i)}) + \hat{\mathcal{R}}(o_1^{(i)})\hat{\mathcal{R}}(o_3^{(i)})}$, $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l_m^{max(i)}$ ($l_m^{min(i)}$) is the sample index for which the maximum (minimum) value for the m^{th} separator output is achieved at the i^{th} iteration.

- Iterative algorithm for f_4:

  The subgradient based adaptive algorithm maximizing $\bar{J}(\boldsymbol{W})$ using the function f_4 can be written as

$$\boxed{\begin{aligned}\boldsymbol{W}^{(i+1)} = \boldsymbol{W}^{(i)} + \mu^{(i)} \Bigg(\left(\boldsymbol{W}^{(i)} \hat{\boldsymbol{R}}_{\boldsymbol{y}} \boldsymbol{W}^{(i)^T} \right)^{-1} \boldsymbol{W}^{(i)} \hat{\boldsymbol{R}}_{\boldsymbol{y}} - 2p \\ \sum_{m=1}^{p} \frac{e^{\hat{\mathcal{R}}(o_m^{(i)})}}{log\,(h)\, h} \mathbf{e}_m \left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)}) \right)^T \Bigg)\end{aligned}}$$

  where $h = e^{\hat{\mathcal{R}}(o_1^{(i)})} + e^{\hat{\mathcal{R}}(o_2^{(i)})} + \ldots + e^{\hat{\mathcal{R}}(o_p^{(i)})}$.

5.3 Numerical Examples and Conclusion

In this section, we provide the following scenario to illustrate the separation capability of the algorithms corresponding to the examples given in the Section 5.1 for the dependent-correlated sources: We generate the sources through the zero-mean adjusted Copula-t distribution, a perfect tool for generating vectors with controlled correlation, with 4 degrees of freedom whose correlation matrix parameter is given by a Toeplitz matrix $\mathbf{R}_s$ whose first row is $\begin{bmatrix} 1 & \rho_s & \ldots & \rho_s^{p-1} \end{bmatrix}$, where the correlation parameter is varied in the range 0 to 1. Here, we consider a scenario with 3 sources and 5 mixtures and the coefficients of the 5×3 mixing matrix are randomly generated, based on i.i.d. Gaussian distribution.

Figure 5.1 shows the output total Signal energy to total Interference energy (over all outputs) Ratio (SIR) obtained for the BCA algorithm examples (corresponding to $f_1, f_{2,1}, f_3, f_4$) for various correlation parameters $\rho_s \in [0, 0.9]$ for the mixture length of $N = 100000$. The same procedure is repeated for FastICA [3], [27] and JADE [28], [29] algorithms, as representative ICA approaches.

We consider the performance criteria as the output total Signal energy to total Interference energy (over all outputs) Ratio (SIR) which is defined as

$$\text{SIR} = \frac{\text{Total Signal Power}}{\text{Total Residual Power}} = \frac{\text{Trace}(\boldsymbol{G}_{\text{sig}}\hat{\boldsymbol{R}}_{\boldsymbol{s}}\boldsymbol{G}_{\text{sig}}^T)}{\text{Trace}(\boldsymbol{G}_{\text{res}}\hat{\boldsymbol{R}}_{\boldsymbol{s}}\boldsymbol{G}_{\text{res}}^T)},$$

where $\boldsymbol{G}_{\text{sig}}$ is defined as the matrix obtained from $\boldsymbol{G}$ by keeping the maximum entries of each row and making the other entries 0 and $\boldsymbol{G}_{\text{res}}$ is defined as the matrix obtained from $\boldsymbol{G}$ by making the maximum entries of each row 0 and keeping the other entries so that we will have the total signal power and total residual power.

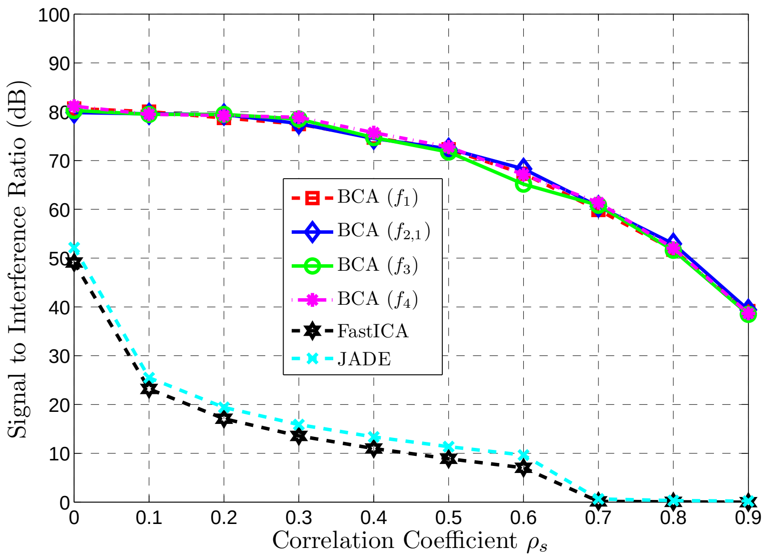

Figure 5.1: Result of the proposed BCA algorithms' performances for the mixtures of dependent sources for various correlation parameters when the mixture length is 100000.

In the second example, we generate the sources from exponentially distributed random variables by the inverse CDF method used on the first setup. Figure 5.2 illustrates the separation performances when the mixture length is $N = 10000$.

We observe from the figures that the BCA algorithms maintain high separation perfor-

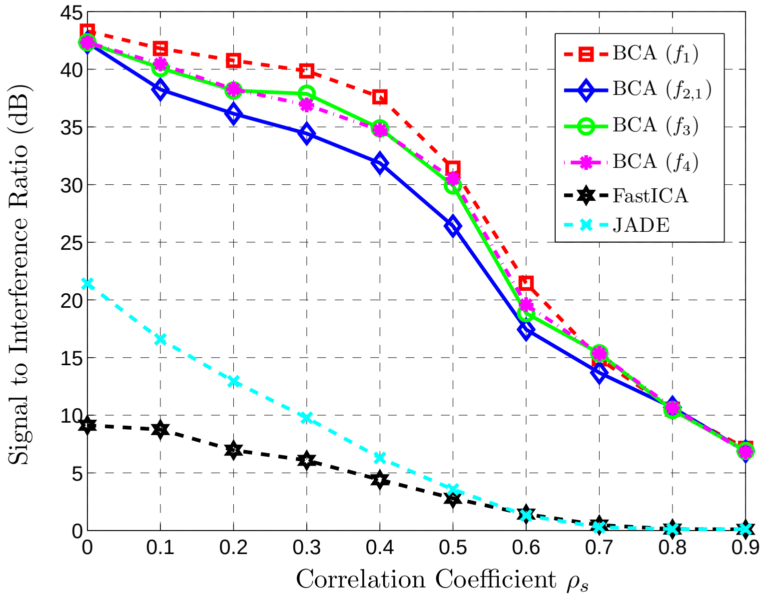

Figure 5.2: Result of the proposed BCA algorithms' performances for the mixtures of exponentially distributed dependent sources for various correlation parameters when the mixture length is 10000.

mance for a wide range of correlation parameters especially for the longer sample size case (i.e., $N = 100000$). However, both FastICA and JADE algorithms' performances degrades substantially along with increasing correlation since the independence assumption does not hold. We also point out that for $\rho_s = 0$, the performances of the BCA algorithms are better than FastICA and JADE even though the independence assumption holds. This is due to the fact that the sample sizes are sufficient for the assumption (**A1**) to hold, whereas they may not be sufficient to reflect the stochastic independence of the sources. We also note that in the second example proposed BCA algorithms have different performances, therefore, the variety of BCA algorithms which can be produced from this analysis might be useful in different scenarios. We finally note that exponential distribution decreases the likeliness of (**A1**) to hold, however, proposed BCA algorithms still have good separation performances and with the longer data records the BCA algorithms become more successful in separating correlated sources.

Chapter 6

Convolutive BCA Algorithms for Stationary Independent and/or Dependent Source Separation

In this chapter, we extend the instantaneous or memoryless BCA approach introduced in [24] for the convolutive BCA problem. We extend the objective functions proposed in [24] to cover the more general case where the observations are space-time mixtures of the original sources. In particular, we show that the algorithms corresponding to these extensions are capable of separating not only independent sources but also sources which are potentially dependent and even correlated in both space and time dimensions. This is a remarkable feature of the proposed approach which is due to a proper exploitation of the boundedness property of sources.

6.1 A Family of Convolutive BCA Algorithms

6.1.1 A Convolutive BCA Optimization Framework

In this section, we extend the instantaneous BCA approach introduced in [24] to the convolutive BSS separation problem. We modify the first (volume ratio) objective function of [24] as

$$J_1(\tilde{\boldsymbol{W}}) = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\boldsymbol{O}}(f)))df - log\left(\prod_{m=1}^{p}\mathcal{R}(o_m)\right), \tag{6.1}$$

where $\mathbf{P}_{\boldsymbol{O}}(f)$ is the PSD of the separator output sequence. In the proposed objective function, we only modify the log volume of the principal hyper-ellipse, i.e., the first term. The definition of the volume of the principal hyper-ellipse is extended from sample based correlation information to process based correlation information, capturing inter-sample correlations. We note that the objective in (6.1) is the asymptotic extension of the J_1 objective in [24]. This extension is obtained by concatenating source vectors in the source process and invoking the wide sense stationarity property of the sources along with the linear-time-invariance property of mixing and separator systems such that the determinant of the covariance in [24] converges to the integral term in (6.1). We also note that this integral term or its modified forms (due to the diagonality of the PSD for independent sources) appear in some convolutive ICA approaches such as [30], [31]. However, unlike ICA, the sources are not assumed to be independent, or uncorrelated, therefore implying that $\mathbf{P}_{\mathbf{s}}(f)$ is allowed to be non-diagonal. We assume sources satisfy "the domain separability" assumption (**C1**) which is stated as follows:

- (**C1**) The (convex hulls of the) domain of the extended source vector ($\mathcal{S}_{\tilde{\boldsymbol{s}}}$) can be written as the cartesian product of (the convex hulls of the) the individual components of the extended source vector ($\mathcal{S}_{\tilde{s}_m}$ for $m = 1, 2, \ldots, Pp$), i.e., $\mathcal{S}_{\tilde{\boldsymbol{s}}} = \mathcal{S}_{\tilde{s}_1} \times \mathcal{S}_{\tilde{s}_2} \times \ldots \times \mathcal{S}_{\tilde{s}_{Pp}}$.

Note that $\tilde{\boldsymbol{s}}$ corresponds to the collection of all random variables from all p sources in a time window of length P. Therefore, the domain separability assumption (**C1**) in effect states that the convex support of the joint pdf corresponding to all of these source random variables in a given P-length time window, is separable, i.e., it can be written as the cartesian product of convex supports of the marginals of these random variables. This assumption essentially states that the range of values for each of these random variables is not determined by the values other random variables, whereas the probability distribution defined over this range may be dependent. Therefore, the assumption (**C1**) is a quite flexible constraint, allowing arbitrary joint densities (corresponding to dependent or independent random variables) over this separable domain. The samples can, in fact, be correlated, i.e., $\mathbf{P_s}(f)$ can change with frequency. We point out that (**C1**) is a much weaker assumption than the assumption of independence of sources in both time and space dimensions. In fact, the domain separability, which is a requirement about the support set of the joint distribution, is a necessary condition for the mutual independence. However, the mutual independence assumption further dictates that the joint distribution is equal to the product of the marginals, which is rather a strong additional requirement on top of the domain separability.

6.1.2 The Global Optimality of the Perfect Separators

In this section, we show that the global optima of the objective function (6.1) corresponds to perfect separators. The following theorem shows that the proposed objective is useful for achieving separation of convolutive mixtures whose setup is outlined in Section 2.3.

Theorem: Assuming the setup in Section 2.3, $\mathbb{H}(f)$ is equalizable by an FIR separator matrix of order $M-1$ and $\mathbf{P}_{\boldsymbol{\mathcal{S}}}(f) \succ 0$ for all $f \in [-\frac{1}{2}, \frac{1}{2})$, the set of global maxima for J_1 in (6.1) is equal to a set of perfect separator matrices.

Proof: Using the fact that

$$\mathbf{P}_{\boldsymbol{O}}(f) = \mathbb{G}(f)\Upsilon\mathbf{P}_{\underline{\boldsymbol{\mathcal{S}}}}(f)\Upsilon^T\mathbb{G}(f)^H,$$

we obtain

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\boldsymbol{O}}(f)))df = \int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(|det(\mathbb{G}(f)\Upsilon)|^2 det(\mathbf{P}_{\underline{\boldsymbol{S}}}(f))\right) df$$
$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(|det(\mathbb{G}(f)\Upsilon)|^2\right) df + \int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{S}}}(f))\right) df. \quad (6.2)$$

Using the Hadamard inequality [26] yields

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(|det(\mathbb{G}(f)\Upsilon)|^2\right) df \leq \int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(\prod_{m=1}^{p} ||\left(\mathbb{G}(f)\Upsilon\right)_{m,:}||_2^2\right) df$$
$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{m=1}^{p} log\left(||\left(\mathbb{G}(f)\Upsilon\right)_{m,:}||_2^2\right) df = \sum_{m=1}^{p} \int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(||\left(\mathbb{G}(f)\Upsilon\right)_{m,:}||_2^2\right) df, \quad (6.3)$$

where $\left(\mathbb{G}(f)\Upsilon\right)_{m,:}$ is the m^{th} row of $\mathbb{G}(f)\Upsilon$. From Jensen's inequality [32], for $m = 1, ..., p$, we have

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(||\left(\mathbb{G}(f)\Upsilon\right)_{m,:}||_2^2\right) df \leq log\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} ||\left(\mathbb{G}(f)\Upsilon\right)_{m,:}||_2^2 df\right). \quad (6.4)$$

The use of Parseval's theorem yields

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} ||\left(\mathbb{G}(f)\Upsilon\right)_{m,:}||_2^2 df = ||(\tilde{\boldsymbol{G}}\tilde{\Upsilon})_{m,:}||_2^2. \quad (6.5)$$

Thus, from (6.3-6.5), we obtain

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(|det(\mathbb{G}(f)\Upsilon)|^2\right) df \leq \sum_{m=1}^{p} log\left(||(\tilde{\boldsymbol{G}}\tilde{\Upsilon})_{m,:}||_2^2\right), \quad (6.6)$$

which further implies,

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\boldsymbol{O}}(f)))df \leq \sum_{m=1}^{p} log\left(||(\tilde{\boldsymbol{G}}\tilde{\Upsilon})_{m,:}||_2^2\right) + \int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{S}}}(f))\right) df. \quad (6.7)$$

As a result,

$$J_1(\tilde{\boldsymbol{W}}) \leq \frac{1}{2}\sum_{m=1}^{p} log\left(||(\tilde{\boldsymbol{G}}\tilde{\Upsilon})_{m,:}||_2^2\right) - log\left(\prod_{m=1}^{p}\mathcal{R}(o_m)\right) + \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{s}}}(f))\right) df. \tag{6.8}$$

Under the BCA's domain separability assumption (**C1**) stated in Section 6.1.1, we can write the range of m^{th} component of $\boldsymbol{o}$ as $\mathcal{R}(o_m) = ||\tilde{\boldsymbol{G}}_{m,:}\tilde{\Upsilon}||_1$. We can further define $\boldsymbol{Q} \triangleq \tilde{\boldsymbol{G}}\tilde{\Upsilon}$, the range vector for the separator outputs can be rewritten as

$$\mathcal{R}(\boldsymbol{o}) = \left[||\boldsymbol{Q}_{1,:}||_1 \quad ||\boldsymbol{Q}_{2,:}||_1 \quad \dots \quad ||\boldsymbol{Q}_{p,:}||_1\right]. \tag{6.9}$$

If we rewrite the inequality (6.8) in terms of $\boldsymbol{Q}$ we obtain

$$J_1(\tilde{\boldsymbol{W}}) \leq \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_2\right) - \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_1\right) + \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{s}}}(f))\right) df. \tag{6.10}$$

Note that,

$$\sum_{m=1}^{p} log\left(||(\boldsymbol{Q}_{m,:})||_2\right) \leq \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_1\right), \tag{6.11}$$

due to the ordering $||\boldsymbol{q}||_1 \geq ||\boldsymbol{q}||_2$ for any $\boldsymbol{q}$. Therefore,

$$J_1(\tilde{\boldsymbol{W}}) \leq \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{s}}}(f))\right) df. \tag{6.12}$$

We note that the equalities in (6.3) and (6.11) must be achieved in order to achieve the equality in (6.12). The equality in (6.11) is achieved if and only if each row of $\boldsymbol{Q}$ has only one non-zero element which results in each row of $\tilde{\boldsymbol{G}}$ has only one non-zero element and the inequality in (6.3) is achieved if and only if the rows of $\mathbb{G}(f)$ are perpendicular to each other. Since $\mathbb{G}(f) = \sum_{l=0}^{P-1} \boldsymbol{G}(l)e^{-j2\pi fl}$ and $\tilde{\boldsymbol{G}} = \begin{bmatrix} \boldsymbol{G}(0) & \boldsymbol{G}(1) & \dots & \boldsymbol{G}(P-1) \end{bmatrix}$, the combination of these two requirements yield that the only one non-zero elements in the rows of $\tilde{\boldsymbol{G}}$ must not be positioned in the same indexes with respect to mod p, otherwise, the rows of $\mathbb{G}(f)$ would not be perpendicular to each other.

As a result, the inequality in (6.12) is achieved if and only if $\tilde{\boldsymbol{G}}$ corresponds to perfect separator transfer matrix in the form

$$\mathbb{G}(f) = diag(\alpha_1 e^{-j2\pi f d_1}, \alpha_2 e^{-j2\pi f d_2}, \ldots, \alpha_p e^{-j2\pi f d_p})\boldsymbol{P} \tag{6.13}$$

where α_k's are non-zero real scalings, d_k's are non-negative integer delays, and $\boldsymbol{P}$ is a Permutation matrix. The FIR equalizability of the mixing system implies the existence of such parameters.

Here, we point out that virtual delayed source problem does not exist in the proposed framework: In case, if one of the separator outputs is the delayed version of another output $\mathbf{P}_{\boldsymbol{O}}(f)$ becomes rank deficient, and therefore, its determinant becomes zero. Therefore, the maximizing solution for the proposed objective will avoid such cases as they will actually minimize the PSD dependent term in the objective (6.1). This fact is also reflected by the proof of the Theorem above.

6.1.3 Extension of Convolutive BCA Optimization Framework

The framework introduced in section 6.1.1 is extended by proposing different alternatives for the second term of the objective function (6.1) (measure of the bounding hyperrectangle for the output vectors). Example of such alternatives can be achieved by choosing the length of the main diagonal of the bounding hyperrectangle. As a result, we obtain a family of alternative objective functions in the form

$$J_{2,r}(\tilde{\boldsymbol{W}}) = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\boldsymbol{O}}(f)))df - log\left(||\mathcal{R}(\boldsymbol{o})||_r^p\right), \tag{6.14}$$

where $r \geq 1$. By modifying (6.10), we can obtain the corresponding objective expression in terms of scaled overall mapping $\boldsymbol{Q}$ as

$$J_{2,r}(\tilde{\boldsymbol{W}}) \leq \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_2\right) - log\left(\left\|\left[||\boldsymbol{Q}_{1,:}||_1 \quad ||\boldsymbol{Q}_{2,:}||_1 \quad \ldots \quad ||\boldsymbol{Q}_{p,:}||_1\right]^T\right\|_r^p\right) + \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{S}}}(f))\right)df. \tag{6.15}$$

The results of analyzing this objective function, for some special r values:

- $r = 1$ *Case:* In this case, we can write

$$\left\|\left[\begin{array}{cccc} ||\boldsymbol{Q}_{1,:}||_1 & ||\boldsymbol{Q}_{2,:}||_1 & \ldots & ||\boldsymbol{Q}_{p,:}||_1 \end{array}\right]^T\right\|_1^p \geq p^p \prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_1,$$

  where the inequation comes from Arithmetic-Geometric-Mean-Inequality, and the equality is achieved if and only if all the rows of $\boldsymbol{Q}$ have the same 1-norm. In consequence, we can write

$$J_{2,1}(\tilde{\boldsymbol{W}}) \leq \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_2\right) - \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_1\right) - log(p^p) + \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{S}}}(f))\right)df. \tag{6.16}$$

  Similarly, from (6.11), we have

$$J_{2,1}(\tilde{\boldsymbol{W}}) \leq \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{S}}}(f))\right)df - log(p^p). \tag{6.17}$$

As a result, $\boldsymbol{Q}$ is a global maximum of $J_{2,1}(\tilde{\boldsymbol{W}})$ if and only if it is a perfect separator matrix of the form

$$\mathbf{Q} = k\mathbf{P}\text{diag}(\rho),$$

where k is a non-zero value, $\rho \in \{-1, 1\}^p$ and $\mathbf{P}$ is a permutation matrix. This implies $\tilde{\boldsymbol{G}}$ is a global maximum of $J_{2,1}(\tilde{\boldsymbol{W}})$ if and only if it can be written in the form

$$\tilde{\boldsymbol{G}} = k\mathbf{P}\tilde{\Upsilon}^{-1}\text{diag}(\rho).$$

- $r = 2$ *Case:* In this case, using the basic norm inequality, for any $\mathbf{a} \in \mathbb{R}^p$, we have

$$||\mathbf{a}||_2 \geq \frac{1}{\sqrt{p}}||\mathbf{a}||_1,$$

where the equality is achieved if and only if all the components of $\mathbf{a}$ are equal in magnitude. As a result, we can write

$$\begin{aligned} J_{2,2}(\tilde{\boldsymbol{W}}) \leq & \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_2\right) - \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_1\right) - log(p^{p/2}) \\ & + \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{s}}}(f))\right) df. \end{aligned} \tag{6.18}$$

Similarly, $J_{2,2}$ has the same set of global maxima as $J_{2,1}$.

- $r = \infty$ *Case:* Using the basic norm inequality, for any $\mathbf{a} \in \mathbb{R}^p$,

$$||\mathbf{a}||_\infty \geq \frac{1}{p}||\mathbf{a}||_1,$$

and Arithmetic-Geometric-Mean-Inequality yields

$$||\mathcal{R}(\boldsymbol{o})||_\infty^p \geq \frac{1}{p^p}||\mathcal{R}(\boldsymbol{o})||_1^p \geq \prod_{m=1}^{p} \mathcal{R}(z_m).$$

where the equality is achieved if and only if all the components of $\mathcal{R}(\boldsymbol{o})$ are equal in magnitude. Based on this inequality, we obtain

$$J_{2,\infty}(\tilde{\boldsymbol{W}}) \leq \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_2\right) - \sum_{m=1}^{p} log\left(||\boldsymbol{Q}_{m,:}||_1\right) + \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log\left(det(\mathbf{P}_{\underline{\boldsymbol{s}}}(f))\right) df. \tag{6.19}$$

Therefore, $J_{2,\infty}$ also has the same set of global optima as $J_{2,1}$ and $J_{2,2}$.

Hence, to attain the global maximum of $J_{2,1}, J_{2,2}$ and $J_{2,\infty}$, there is also a condition that all the rows of $\boldsymbol{Q}$ have the same 1-norm. This implies $\tilde{\boldsymbol{G}}$ is a global maximum of $J_{2,1}, J_{2,2}$ and $J_{2,\infty}$ if and only if it can be written in the form

$$\tilde{\boldsymbol{G}} = k\mathbf{P}\tilde{\Upsilon}^{-1}\text{diag}(\rho).$$

where k is a non-zero value, $\rho \in \{-1,1\}^p$ and $\mathbf{P}$ is a permutation matrix which corresponds to a subset of perfect separators defined by (6.13).

6.2 Iterative BCA Algorithms

In this section, we provide the adaptive algorithm corresponding to the optimization settings presented in the previous section.

- *Objective Function* $J_1(\tilde{\boldsymbol{W}})$:

In the adaptive implementation, we assume a set of finite observations of mixtures $\{\boldsymbol{y}(0), \boldsymbol{y}(1), ..., \boldsymbol{y}(N-1)\}$ and modify the objective as

$$\bar{J}_1(\tilde{\boldsymbol{W}}) = \frac{1}{2\eta}\sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\boldsymbol{O}}(l))) - log\left(\prod_{m=1}^{p}\hat{\mathcal{R}}(o_m)\right), \tag{6.20}$$

where $\nu = N + M - 1$, $\eta = 2\nu + 1$ is the DFT size and we use the PSD estimate for the separator outputs given by

$$\hat{\mathbf{P}}_{\boldsymbol{O}}(l) = \sum_{k=-\nu}^{\nu}\hat{\boldsymbol{R}}_{\boldsymbol{O}}(k)e^{-j2\pi lk/\eta},$$

for $l \in \{-\nu, ..., \nu\}$, where N is the number of samples and $\hat{\boldsymbol{R}}_{\boldsymbol{O}}$ is the output sample autocovariance function, defined as

$$\hat{\boldsymbol{R}}_{\boldsymbol{O}}(k) = \frac{1}{\nu + 1 - |k|}\sum_{q=max(0,-k)}^{min(\nu,\nu-k)}\boldsymbol{o}(q)\boldsymbol{o}^T(q+k),$$

for $k = -\nu, ..., \nu$. We point out that we use $\hat{\mathcal{R}}(\boldsymbol{o})$ for the range vector of the sample outputs for which we have

$$\hat{\mathcal{R}}(o_m) = \max_{k\in\{1,2,...,N\}} o_m(k) - \min_{k\in\{1,2,...,N\}} o_m(k),$$

for $m = 1, 2, ..., p$.

Note that the derivative of the first part of $\bar{J}_1(\tilde{\boldsymbol{W}})$ with respect to $\boldsymbol{W}(n)$ is

$$\frac{1}{2\eta}\frac{\partial\sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\boldsymbol{O}}(l)))}{\partial\mathbf{W}(n)} = \frac{1}{\eta}\sum_{l=-\nu}^{\nu}\mathbb{R}\left\{\hat{\mathbf{P}}_{\boldsymbol{O}}(l)^{-1}\mathbb{W}(l)\hat{\mathbf{P}}_{\boldsymbol{y}}(l)e^{j2\pi nl/\eta}\right\}, \tag{6.21}$$

where

$$\hat{\mathbb{W}}(l) = \sum_{k=-\nu}^{\nu} \boldsymbol{W}(k) e^{-j2\pi lk/\eta},$$

and

$$\hat{\mathbf{P}}_{\boldsymbol{y}}(l) = \sum_{k=-\nu}^{\nu} \hat{\boldsymbol{R}}_{\boldsymbol{y}}(k) e^{-j2\pi lk/\eta}.$$

Following the similar steps as in [24] for the derivative of $log\left(\prod_{m=1}^{p} \hat{\mathcal{R}}(o_m)\right)$, the subgradient based iterative algorithm for maximizing objective function (6.20) is provided as

$$\begin{aligned}\boldsymbol{W}^{(i+1)}(n) = \boldsymbol{W}^{(i)}(n) + \mu^{(i)} \Bigg(\frac{1}{\eta} \sum_{l=-\nu}^{\nu} \mathbb{R} \left\{ \hat{\mathbf{P}}_{\boldsymbol{o}}(l)^{-1} \hat{\mathbb{W}}^{(i)}(l) \hat{\mathbf{P}}_{\boldsymbol{y}}(l) e^{j2\pi nl/\eta} \right\} \\ - \sum_{m=1}^{p} \frac{1}{\mathbf{e}_m^T \hat{\mathcal{R}}(\boldsymbol{o}_{\mathbf{W}^{(i)}})} \mathbf{e}_m \left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)}) \right)^T \Bigg), \quad (6.22)\end{aligned}$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l_m^{max(i)}$ ($l_m^{min(i)}$) is the sample index for which the maximum (minimum) value for the m^{th} separator output is achieved at the i^{th} iteration.

- *Objective Function* $J_{2,r}(\tilde{\boldsymbol{W}})$:

In the adaptive implementation, we modify the family of alternative objective functions as

$$\bar{J}_{2,r}(\tilde{\boldsymbol{W}}) = \frac{1}{2\eta} \sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\boldsymbol{o}}(l))) - log\left(||\hat{\mathcal{R}}(\boldsymbol{o})||_r^p \right) \quad (6.23)$$

For $r = 1, 2$, we can write the update equation as

$$\boldsymbol{W}^{(i+1)}(n) = \boldsymbol{W}^{(i)}(n) + \mu^{(i)}\Bigg(\frac{1}{\eta}\sum_{l=-\nu}^{\nu}\mathbb{R}\left\{\hat{\mathbf{P}}\boldsymbol{o}(l)^{-1}\hat{\mathbb{W}}^{(i)}(l)\hat{\mathbf{P}}\boldsymbol{y}(l)e^{j2\pi nl/\eta}\right\} - \\ \sum_{m=1}^{p}\frac{p\hat{\mathcal{R}}_m(\boldsymbol{o}_{\mathbf{W}^{(i)}})^{r-1}}{||\hat{\mathcal{R}}(\boldsymbol{o}_{\mathbf{W}^{(i)}})||_r^r}\mathbf{e}_m\left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^T\Bigg). \quad (6.24)$$

For $r = \infty$, the update equation has the form

$$\boldsymbol{W}^{(i+1)}(n) = \boldsymbol{W}^{(i)}(n) + \mu^{(i)}\Bigg(\frac{1}{\eta}\sum_{l=-\nu}^{\nu}\mathbb{R}\left\{\hat{\mathbf{P}}\boldsymbol{o}(l)^{-1}\hat{\mathbb{W}}^{(i)}(l)\hat{\mathbf{P}}\boldsymbol{y}(l)e^{j2\pi nl/\eta}\right\} - \\ \sum_{m\in\mathcal{M}(\boldsymbol{o}_{\mathbf{W}^{(i)}})}\frac{p\beta_m^{(i)}}{||\hat{\mathcal{R}}(\boldsymbol{o}_{\mathbf{W}^{(i)}})||_\infty}\mathbf{e}_m\left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^T \quad (6.25)$$

where $\mathcal{M}(\boldsymbol{o}_{\mathbf{W}^{(i)}})$ is the set of indexes for which the peak range value is achieved, i.e.,

$$\mathcal{M}(\boldsymbol{o}_{\mathbf{W}^{(i)}}) = \{m : \hat{\mathcal{R}}_m(\boldsymbol{o}_{\mathbf{W}^{(i)}}) = \|\hat{\mathcal{R}}(\boldsymbol{o}_{\mathbf{W}^{(i)}})\|_\infty\}, \quad (6.26)$$

and $\beta_m^{(i)}$s are the convex combination coefficients.

6.3 Extension to Complex Signals

In the complex domain, we consider p complex sources with finite support, i.e., real and imaginary part of the sources have finite support. We define the operator $\Phi : \mathbb{C}^p \rightarrow \mathbb{R}^{2p}$,

$$\Phi(\mathbf{a}) = \begin{bmatrix} \mathbb{R}\{\mathbf{a}^T\} & \mathbb{I}\{\mathbf{a}^T\} \end{bmatrix}^T \quad (6.27)$$

as an isomorphism between p dimensional complex domain and $2p$ dimensional real domain. For any complex vector $\mathbf{a}$, we introduce the corresponding isomorphic real vector as $\grave{\mathbf{a}}$, i.e., $\grave{\mathbf{a}} = \Phi\{\mathbf{a}\}$. We also define the operator $\Gamma : \mathbb{C}^{p\times q} \to \mathbb{R}^{2p\times 2q}$ as

$$\Gamma(\mathbf{A}) = \begin{bmatrix} \mathbb{R}\{\mathbf{A}\} & -\mathbb{I}\{\mathbf{A}\} \\ \mathbb{I}\{\mathbf{A}\} & \mathbb{R}\{\mathbf{A}\} \end{bmatrix}. \tag{6.28}$$

In the complex domain, both mixing and separator coefficient matrices are complex matrices, i.e., $\tilde{\boldsymbol{H}} \in \mathbb{C}^{q\times pL}$ and $\tilde{\boldsymbol{W}} \in \mathbb{C}^{p\times qM}$. The set of source vectors S and the set of separator outputs O are subsets of $\mathbb{C}^p$ and the set of mixtures Y is a subset of $\mathbb{C}^q$. We also note that since

$$\boldsymbol{y}(k) = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{s}}(k),$$

we have

$$\grave{\boldsymbol{y}}(k) = \Gamma(\tilde{\boldsymbol{G}})\grave{\tilde{s}}(k).$$

where $\Gamma(\tilde{\boldsymbol{G}}) = \begin{bmatrix} \Gamma(\boldsymbol{G}(0)) & \Gamma(\boldsymbol{G}(1)) & \ldots & \Gamma(\boldsymbol{G}(P-1)) \end{bmatrix}$ and $\grave{\tilde{s}}(k) = [\ \grave{\boldsymbol{s}}(k)\ \ \grave{\boldsymbol{s}}(k-1)\ \ldots\ \grave{\boldsymbol{s}}(k-P+1)\ ]^T$.

6.3.1 Complex Extension of the Convolutive BCA Optimization Framework

We can extend the framework introduced in the Section 6.1 to the complex signals by following similar steps with the real vector $\grave{\boldsymbol{o}}$. We define the subset of $\mathbb{R}^{2p}$ vectors which are isomorphic to the elements of O as

$$\grave{O} = \{\grave{\boldsymbol{o}} : \boldsymbol{o} \in O\}.$$

Similarly, we define

$$\begin{aligned} \grave{S} = & \quad \{\grave{\mathbf{s}} : \mathbf{s} \in S\}, \\ \grave{Y} = & \quad \{\grave{\mathbf{y}} : \mathbf{y} \in Y\}. \end{aligned}$$

By these definitions, we modify the objective function J_1 in (6.1) for the complex case as

$$Jc_1(\tilde{\boldsymbol{W}}) = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\grave{\boldsymbol{o}}}(f)))df - log\left(\prod_{m=1}^{2p} \mathcal{R}(\grave{o}_m)\right). \tag{6.29}$$

Note that the mapping between $\Phi(\underline{\boldsymbol{s}})$ and $\Phi(\boldsymbol{o})$ is given by $\Gamma(\boldsymbol{Q})$, thus the theorem proved in Section 6.1.2 implies that the set of global maxima for the objective function (6.29) have the properties that the corresponding $\Gamma(\tilde{\boldsymbol{G}})$ satisfies (6.13) and the structure imposed by (6.28). Therefore, the set of global optima for (6.29) (in terms of $\tilde{\boldsymbol{G}}$) is given by

$$\begin{aligned} GO_c \quad = \quad & \{\tilde{\boldsymbol{G}} = \mathbf{PD} : \mathbf{P} \in \mathbb{R}^{p \times pP} \text{ is a permutation matrix,} \\ & \mathbf{D} \in \mathbb{C}^{pP \times pP} \text{ is a full rank diagonal matrix with} \\ & D_{ii} = \alpha_i e^{\frac{j\pi k_i}{2}}, \alpha_i \in \mathbb{R}, k_i \in Z, i = 1, \dots pM\}, \end{aligned}$$

which corresponds to a subset of complex perfect separators with discrete phase ambiguity.

In the adaptive implementation, we modify the objective as

$$\bar{J}c_1(\tilde{\boldsymbol{W}}) = \frac{1}{2\eta}\sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\grave{\boldsymbol{o}}}(l))) - log\left(\prod_{m=1}^{2p}\hat{\mathcal{R}}(\grave{o}_m)\right). \tag{6.30}$$

Note that the derivative of the first part of $\bar{J}c_1(\tilde{\boldsymbol{W}})$ with respect to $\boldsymbol{W}(n)$ is

$$\frac{1}{2\eta}\frac{\partial\sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\grave{\boldsymbol{o}}}(l)))}{\partial\mathbf{W}(n)} = \Lambda_{11} + \Lambda_{22} + j(\Lambda_{21} - \Lambda_{12}), \tag{6.31}$$

where we define

$$\frac{1}{\eta}\sum_{l=-\nu}^{\nu}\mathbb{R}\left\{\hat{\mathbf{P}}_{\grave{\boldsymbol{o}}}(l)^{-1}\Gamma(\hat{\mathbb{W}})(l)\hat{\mathbf{P}}_{\grave{\mathbf{y}}}(l)e^{j2\pi nl/\eta}\right\} = \begin{bmatrix}\Lambda_{11} & \Lambda_{21}\\ \Lambda_{21} & \Lambda_{22}\end{bmatrix} \qquad \Lambda_{11}, \Lambda_{12}, \Lambda_{21}, \Lambda_{22} \in \mathbb{R}^{p\times q}$$

where

$$\Gamma(\hat{\mathbb{W}})(l) = \sum_{k=-\nu}^{\nu}\Gamma(\boldsymbol{W}(k))e^{-j2\pi lk/\eta}.$$

The corresponding iterative update equation for $\mathbf{W}(n)$ can be written as

$$\begin{aligned}\mathbf{W}^{(i+1)}(n) = \mathbf{W}^{(i)}(n) + \mu^{(i)}\Bigg(&\Lambda_{11}^{(i)} + \Lambda_{22}^{(i)} + j(\Lambda_{21}^{(i)} - \Lambda_{12}^{(i)})\\ &- \sum_{m=1}^{2p}\frac{1}{\mathbf{e}_m^T\hat{\mathcal{R}}(\grave{\mathbf{o}}_{\mathbf{W}^{(i)}})}\mathbf{v}_m\left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^H\Bigg),\end{aligned} \tag{6.32}$$

where we define

$$\mathbf{v}_m = \begin{cases}\mathbf{e}_m & m \leq p,\\ j\mathbf{e}_{m-p} & m > p.\end{cases} \tag{6.33}$$

A variation on the approach considered for the complex case can be obtained by observing that in the Jc_1 objective function

$$log(det(\mathbf{P}_{\grave{\boldsymbol{O}}}(f))) = log\left(|det(\Gamma(\mathbb{G})(f)\Gamma(\Upsilon))|^2 det(\mathbf{P}_{\underline{\grave{\boldsymbol{s}}}}(f))\right), \tag{6.34}$$

where $\Gamma(\mathbb{G})(f) = \sum_{l=0}^{P-1} \Gamma(\boldsymbol{G}(l))e^{-j2\pi fl}$. We also note that

$$|det(\Gamma(\mathbb{G})(f)\Gamma(\Upsilon))| = |det(\mathbb{G}(f)\Upsilon)|^2. \tag{6.35}$$

Therefore, if we define an alternative objective function

$$Jc_{1a}(\tilde{\boldsymbol{W}}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\boldsymbol{O}}(f)))df - log\left(\prod_{m=1}^{2p} \mathcal{R}(\grave{o}_m)\right), \tag{6.36}$$

it would have the same set of global optima. In the adaptive implementation, the modified objective will be as

$$\bar{J}c_{1a}(\tilde{\boldsymbol{W}}) = \frac{1}{\eta} \sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\boldsymbol{O}}(l))) - log\left(\prod_{m=1}^{2p} \hat{\mathcal{R}}(\grave{o}_m)\right). \tag{6.37}$$

The convenience of Jc_{1a} is in terms of the simplified update expression for (6.31), which results in the derivative of the first part of $\bar{J}c_{1a}(\tilde{\boldsymbol{W}})$ with respect to $\boldsymbol{W}(n)$ is

$$\frac{1}{\eta}\frac{\partial \sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\boldsymbol{O}}(l)))}{\partial \mathbf{W}(n)} = \frac{1}{\eta} \sum_{l=-\nu}^{\nu} \hat{\mathbf{P}}_{\boldsymbol{O}}(l)^{-1} \hat{\mathbb{W}}(l) \hat{\mathbf{P}}_{\mathbf{y}}(l) e^{j2\pi nl/\eta}. \tag{6.38}$$

Therefore, the corresponding iterative update equation for $\mathbf{W}(n)$ in (6.32) is updated accordingly.

6.3.2 Complex Extension of the Alternative Objective Functions

Similar to the complex extension of J_1 provided in the previous subsection, we can extend the J_2 family by defining

$$Jc_{2,r}(\tilde{\boldsymbol{W}}) = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\grave{\boldsymbol{o}}}(f)))df - log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right), \tag{6.39}$$

or alternatively,

$$Jca_{2,r}(\tilde{\boldsymbol{W}}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} log(det(\mathbf{P}_{\boldsymbol{O}}(f)))df - log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right), \tag{6.40}$$

and in the adaptive implementation by modifying these objective functions as

$$\bar{J}c_{2,r}(\tilde{\boldsymbol{W}}) = \frac{1}{2\eta}\sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\grave{\boldsymbol{o}}}(l))) - log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right), \tag{6.41}$$

or alternatively,

$$\bar{J}ca_{2,r}(\tilde{\boldsymbol{W}}) = \frac{1}{\eta}\sum_{l=-\nu}^{\nu} log(det(\hat{\mathbf{P}}_{\boldsymbol{O}}(l))) - log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right). \tag{6.42}$$

The update equation is similar to (6.32) where the derivative of first part of (6.41) or (6.42) is given by either (6.31) or (6.38) depending on the choice and the derivative of the second part depends on the selection of r, e.g.,

- $r = 1, 2$ *Case*: In this case

$$\frac{\partial log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right)}{\partial \mathbf{W}(n)} = \sum_{m=1}^{2p} \frac{p\hat{\mathcal{R}}_m(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}})^{r-1}}{||\hat{\mathcal{R}}(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}})||_r^r}\mathbf{v}_m\left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^H, \tag{6.43}$$

  where $\mathbf{v}_m$ is as defined in (7.15).

- $r = \infty$ *Case*: In this case

$$\frac{\partial log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right)}{\partial \mathbf{W}(n)} = \sum_{m \in \mathcal{M}(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}})} \frac{p\beta_m^{(i)}}{||\hat{\mathcal{R}}(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}})||_\infty} \mathbf{v}_m \left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^H \quad (6.44)$$

where $\mathbf{v}_m$ is as defined in (7.15),

$$\mathcal{M}(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}}) = \{m : \hat{\mathcal{R}}_m(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}}) = \|\hat{\mathcal{R}}(\grave{\boldsymbol{o}}_{\mathbf{W}^{(i)}})\|_\infty\}, \quad (6.45)$$

and $\beta_m^{(i)}$s are the convex combination coefficients.

6.4 Numerical Examples

In this section, we illustrate the separation capability of the proposed algorithms for the convolutive mixtures of both independent and dependent sources.

6.4.1 Separation of Space-Time Correlated Sources

We first consider the following scenario to illustrate the performance of the proposed algorithms regarding the separability of convolutive mixtures of space-time correlated sources: In order to generate space-time correlated sources, we first generate a samples of a τp size vector, $\mathbf{d}$, with Copula-t distribution in [33], a perfect tool for generating vectors with controlled correlation, with 4 degrees of freedom whose correlation matrix parameter is given by $\mathbf{R} = \mathbf{R}_t \otimes \mathbf{R}_s$ where $\mathbf{R}_t$ ($\mathbf{R}_s$) is a Toeplitz matrix whose first row is $\begin{bmatrix} 1 & \rho_t & \dots & \rho_t^{\tau-1} \end{bmatrix}$ ($\begin{bmatrix} 1 & \rho_s & \dots & \rho_s^{p-1} \end{bmatrix}$) (Note that the Copula-t distribution is obtained from the corresponding t-distribution through the mapping of each component using the corresponding marginal cumulative distribution functions, which leads to uniform marginals [33]). Each sample of $\mathbf{d}$ is partitioned to produce source vectors, $\mathbf{d}(k) = \begin{bmatrix} \mathbf{s}(k\tau) & \mathbf{s}(k\tau+1) & \dots & \mathbf{s}((k+1)\tau-1) \end{bmatrix}$. Therefore, we obtain the source vectors as samples of a wide-sense cyclostationary[1] process whose correlation structure in time direction and space directions are governed by the parameters ρ_t and ρ_s, respectively.

In the simulations, we considered a scenario with 3 sources and 5 mixtures, an i.i.d. Gaussian convolutive mixing system with order 3 and a separator of order 10. At each run, we generate 50000 source vectors where τ is set as 5. The results are computed and averaged over 500 realizations.

Figure 6.1 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithms ($\bar{J}_1, \bar{J}_{2,1}, \bar{J}_{2,\infty}$) for various space and time correlation parameters under 45dB SNR. SINR performance of Minimum Mean Square Error (MMSE) filter of the same order, which uses full information about mixing system and source/noise statistics, is also shown to evaluate the relative success of the proposed approach. A comparison has also been made with a gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35] where we take $k_{max} = 50$ and $l_{max} = 20$. We have obtained these methods from [2], [36].

We consider the performance criteria as the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) which is defined as

$$\text{SINR} = \frac{\text{Total Signal Power}}{\text{Total Residual Power + Total Noise Power}} = \frac{\text{Trace}(\boldsymbol{G}_{\text{sig}}\hat{\boldsymbol{R}}_{\boldsymbol{S}}\boldsymbol{G}_{\text{sig}}^T)}{\text{Trace}(\boldsymbol{G}_{\text{res}}\hat{\boldsymbol{R}}_{\boldsymbol{S}}\boldsymbol{G}_{\text{res}}^T) + \sigma^2||\boldsymbol{W}||_F^2},$$

where $\boldsymbol{G}_{\text{sig}}$ is defined as the matrix obtained from $\boldsymbol{G}$ by keeping the maximum entries of each row and making the other entries 0 and $\boldsymbol{G}_{\text{res}}$ is defined as the matrix obtained from $\boldsymbol{G}$ by making the maximum entries of each row 0 and keeping the other entries so

[1]This actually violates the stationarity assumption on sources when $\rho_t \neq 0$. However, we still use this as a convenient method to generate space-time correlated sources

that we will have the total signal power and total residual power. The variance of the noise is σ^2 and $||\boldsymbol{W}||_F$ is the Frobenius norm of matrix $\boldsymbol{W}$.

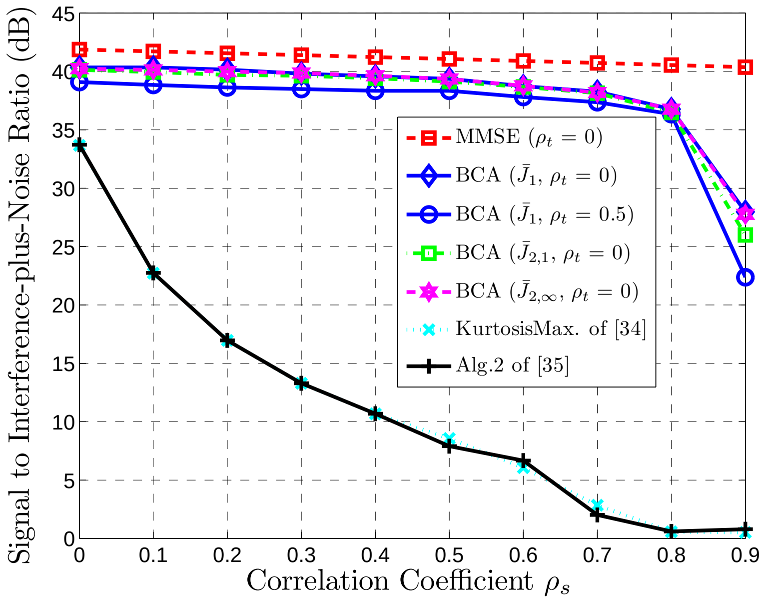

Figure 6.1: Dependent convolutive mixtures separation performance results for SNR = 45dB.

For the same setup, Figure 6.2 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithms ($\bar{J}_1, \bar{J}_{2,1}, \bar{J}_{2,\infty}$), gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.), **Alg.2** of [35], and MMSE for various space correlation parameters under 20dB SNR.

In Figure 6.3, we provide the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithms ($\bar{J}_1, \bar{J}_{2,1}, \bar{J}_{2,2}, \bar{J}_{2,\infty}$), gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.), **Alg.2** of [35], and MMSE for various space correlation parameters under 5dB SNR.

These results demonstrate that the performance of proposed algorithms closely follows

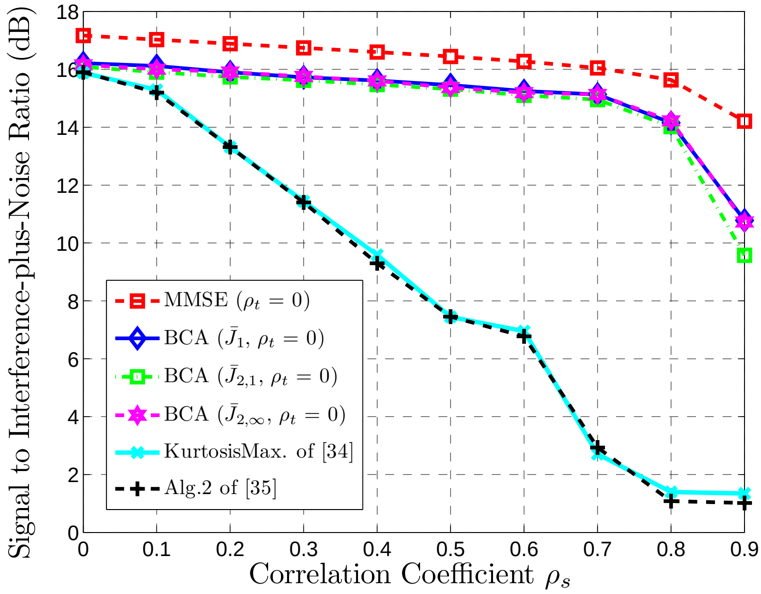

Figure 6.2: Dependent convolutive mixtures separation performance results for SNR = 20dB.

its MMSE counterpart for a wide range of correlation values. Therefore, we obtain a convolutive extension of the BCA approach introduced in [24], which is capable of separating convolutive mixtures of space-time correlated sources.

We also point out that the proposed BCA approaches maintains high separation performance for various space and time correlation parameters. On the other hand, the performance of gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35] degrades substantially with increasing correlation, since in the correlated case, independence assumption simply fails.

6.4.2 MIMO Blind Equalization

We next consider the following scenario to illustrate the performance of the proposed method for the convolutive mixtures of digital communication sources. We consider 3 complex QAM sources such that 2 sources are 16-QAM signals and 1 source is 4-QAM

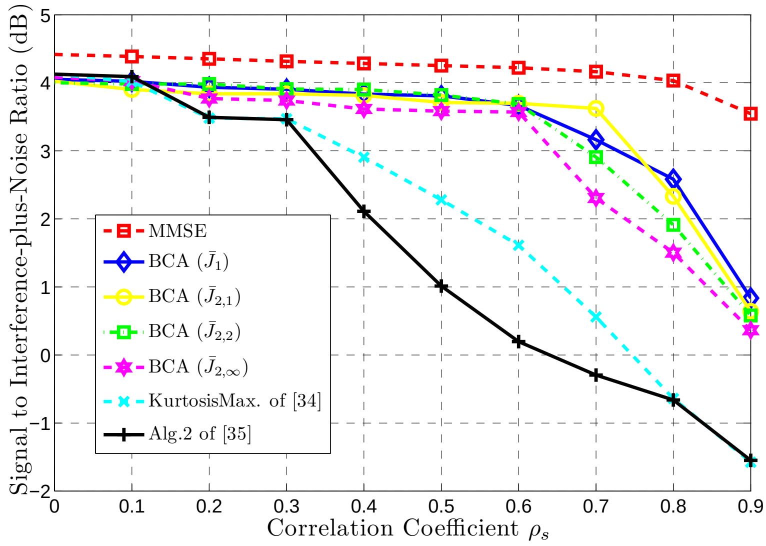

Figure 6.3: Dependent convolutive mixtures separation performance results for SNR = 5dB.

signal. We take 5 mixtures, an i.i.d. Gaussian convolutive mixing system with order 3 and a separator of order 10. The results are computed and averaged over 500 realizations. We use the objective functions $\bar{J}c_1$, $\bar{J}c_{2,1}$ and $\bar{J}c_{2,2}$ as the BCA algorithms introduced in Section 6.3 for this simulation. The resulting Signal to Interference Ratio is plotted with respect to the sample lengths in Figure 6.4. We have also compared our algorithms with a gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35].

According to Figure 6.4, the proposed BCA approaches achieve better performance than ICA based approaches. As it is mentioned earlier, the proposed method does not assume/exploit statistical independence. The only impact of short data length is on accurate representation of source box boundaries. The simulation results suggest that the shorter data records may not be sufficient to reflect the stochastic independence of the sources, and therefore, the compared algorithms require more data samples to achieve the same SIR level as the proposed approach.

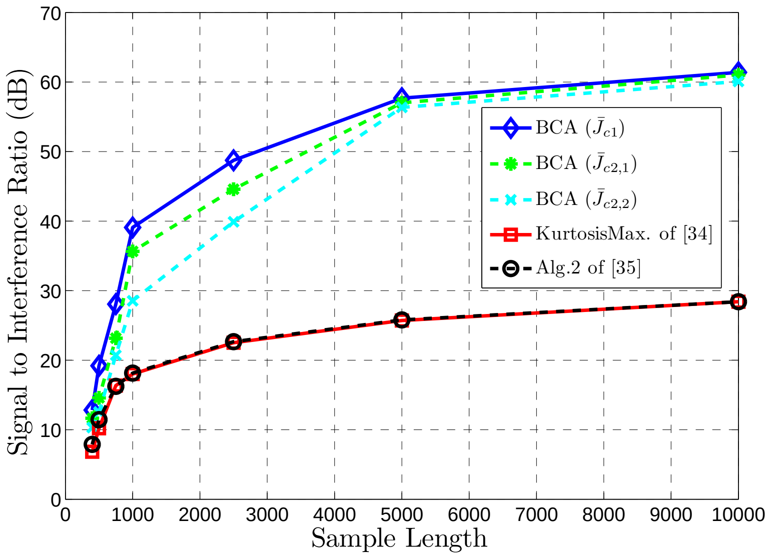

Figure 6.4: Signal to Interference Ratio as a function of Sample Length

6.5 Conclusion

In this chapter, we introduce an algorithmic framework for the convolutive Bounded Component Analysis problem. The utility of the proposed algorithms are mainly twofold:

- The proposed algorithms are capable of separating not only independent sources but also dependent, even correlated sources. The dependence/correlation is allowed to be in both source (or space) and in sample (or time) directions. The proposed framework's capability in terms of separating space-time correlated sources (as well as independent sources) from their convolutive mixtures favors it as a widely applicable approach under the practical constraint on the boundedness of sources. In fact the proposed approach can be considered as a more general convolutive approach than ICA with additional dependent source separation capability, under the condition that the sources are bounded and satisfy domain separability assumption.

- Even though the source samples may be drawn from a stochastic setting where they are mutually independent, especially for short data records, the estimation of sources based on domain separability is expected to be more robust than the estimation based on the independence feature. As illustrated in the previous section, this feature results in superior separation performances, relative to some state of the art Convolutive ICA methods, in convolutive MIMO equalization problem, which is more pronounced especially for shorter packet sizes.

We note that the dimension of the extended vector of sources ($\tilde{\boldsymbol{s}}$) increases with the order of the overall system and with the number of sources. This implies that the proposed convolutive BCA approaches' performances will depend on the sample length as the order of the convolutive system and/or the number of sources increase. We finally note that for the applications where the sources have tailed distributions, the performances of the proposed convolutive BCA algorithms are likely to suffer, which can be considered as one drawback of the algorithms.

Chapter 7

A Convolutive BCA Analysis Framework for Potentially Non-Stationary Independent and/or Dependent Sources

In this chapter, we extend the instantaneous or memoryless BCA approach introduced in [1] for the convolutive BCA problem. We propose deterministic frameworks for the blind source extraction and blind source separation problems which allows the sources to be potentially non-stationary. We point out that the sources could be stationary or non-stationary and we do not exploit non-stationary property of sources. However, the proposed scheme works for both stationary and non-stationary sources. We show that the algorithms corresponding to these frameworks are capable of extracting/separating convolutive mixtures of not only independent sources but also dependent (even correlated) sources where the correlation can be in both space and time dimensions.

7.1 Blind Source Extraction

In this section, we first introduce the objective function for the blind source extraction of real signals. We then prove that the global maxima of the introduced objective function correspond to perfect extractors. We provide the iterative algorithm corresponding to the objective function. We conclude with the complex sources extension of the proposed approach.

In this case, the mixtures are passed through an extractor system and produce the single output as $o(k) = \tilde{\boldsymbol{w}}^T \tilde{\boldsymbol{y}}_M(k)$ where $\tilde{\boldsymbol{w}} = [\ \boldsymbol{w}^T(0)\ \ \boldsymbol{w}^T(1)\ \ \ldots\ \ \boldsymbol{w}^T(M-1)\ ]^T$ is the extractor coefficient vector. Therefore, the sources $\{\boldsymbol{s}(k) \in \mathbb{R}^p; k \in \mathbb{Z}\}$ and the single extractor output $\{o(k) \in \mathbb{R}; k \in \mathbb{Z}\}$ are related by $o(k) = \tilde{\boldsymbol{g}}^T \tilde{\boldsymbol{s}}_P(k)$ where $\tilde{\boldsymbol{g}} = [\ \boldsymbol{g}^T(0)\ \ \ldots\ \ \boldsymbol{g}^T(P-1)\ ]^T$ is defined as the overall system coefficient vector. The generated set of extractor output is illustrated as $o = \{o(1), o(2), \ldots, o(N-M+1)\}$.

7.1.1 Criterion

We introduce the objective function for the blind source extraction method as

$$J_e(\tilde{\boldsymbol{w}}) = \frac{\sqrt{\frac{1}{N_1}\sum_{l=1}^{N_1}\left(o(l) - \hat{\mu}_o\right)^2}}{\hat{\mathcal{R}}(o)}, \tag{7.1}$$

where $\hat{\mu}_o = \frac{1}{N_1}\sum_{l=1}^{N_1} o(l)$, $N_1 = N - M + 1$ and $\hat{\mathcal{R}}(o)$ is the range of the single output in set o. We note that this objective function is deduced from the instantaneous BCA objectives introduced in [1].

We define

$$\begin{aligned}\boldsymbol{\mu}_{\underline{\tilde{\boldsymbol{s}}}_P} &= \frac{1}{N_1}\sum_{l=M}^{N} \underline{\tilde{\boldsymbol{s}}}_P(l),\\ \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} &= \frac{1}{N_1}\sum_{l=M}^{N}(\underline{\tilde{\boldsymbol{s}}}_P(l) - \boldsymbol{\mu}_{\underline{\tilde{\boldsymbol{s}}}_P})(\underline{\tilde{\boldsymbol{s}}}_P(l) - \boldsymbol{\mu}_{\underline{\tilde{\boldsymbol{s}}}_P})^T,\end{aligned}$$

as the sample covariance matrix of $\underline{\tilde{\boldsymbol{s}}}_P$. If sources are stationary, then $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P}$ is a block Toeplitz matrix. However, sources are allowed to be non-stationary, therefore, $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P}$ may not be a block Toeplitz matrix. Note that this approach does not exploit any structure on $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P}$ (i.e., the sources can be stationary or non-stationary). Under the condition

$\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \succ 0$, the following theorem shows that maximizing the proposed objective function (7.1) achieves the blind source extraction of convolutive mixtures whose setup is outlined in Section 2.3.

Theorem 1: Assuming the setup in Section 2.3, $\tilde{\boldsymbol{H}}$ is equalizable by an FIR extractor matrix of order $M-1$ and under the validity of (**C1**), the set of global maxima for J_e in (7.1) is equal to the set of perfect extractors.

Proof: The proof is provided in Appendix 7.5.1.

7.1.2 Algorithm

In this section, we provide the iterative algorithm corresponding to the optimization setting presented in the previous section.

Rather than maximizing J_e, we maximize its logarithm since with the logarithm operation, we utilize the conversion of ratio expression to the difference expression since it simplifies the update components in the iterative algorithm. Therefore, the new objective function is modified as

$$\bar{J}_e(\tilde{\boldsymbol{w}}) = \log\left(J_e(\tilde{\boldsymbol{w}})\right) = \frac{1}{2}\log\left(\tilde{\boldsymbol{w}}^T \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M} \tilde{\boldsymbol{w}}\right) - \log\left(\hat{\mathcal{R}}(o)\right), \tag{7.2}$$

where $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M}$ is the sample covariance matrix of $\tilde{\boldsymbol{y}}_M$.

Note that the derivative of the first part of $\bar{J}_e(\tilde{\boldsymbol{w}})$ with respect to $\tilde{\boldsymbol{w}}$ is

$$\frac{\partial \log\left(\tilde{\boldsymbol{w}}^T \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M} \tilde{\boldsymbol{w}}\right)}{\partial \tilde{\boldsymbol{w}}} = \frac{2\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M} \tilde{\boldsymbol{w}}}{\tilde{\boldsymbol{w}}^T \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M} \tilde{\boldsymbol{w}}}.$$

Following the similar steps as in [1] for the derivative of $\log\left(\hat{\mathcal{R}}(o)\right)$, the subgradient based iterative algorithm for maximizing objective function (7.2) is provided as

$$\tilde{\boldsymbol{w}}^{(i+1)} = \tilde{\boldsymbol{w}}^{(i)} + \mu^{(i)} \left(\frac{\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M} \tilde{\boldsymbol{w}}}{\tilde{\boldsymbol{w}}^T \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_M} \tilde{\boldsymbol{w}}} - \frac{1}{\hat{\mathcal{R}}(o^{(i)})} \left(\tilde{\boldsymbol{y}}_M(l^{max(i)}) - \tilde{\boldsymbol{y}}_M(l^{min(i)}) \right) \right), \tag{7.3}$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l^{max(i)}$ ($l^{min(i)}$) is the sample index for which the maximum (minimum) value for the extractor output is achieved at the i^{th} iteration.

7.1.3 Extension to Complex Signals

In the complex domain, both mixing and extractor coefficient matrices are complex matrices, i.e., $\tilde{\boldsymbol{H}} \in \mathbb{C}^{q \times pL}$ and $\tilde{\boldsymbol{w}} \in \mathbb{C}^{qM \times 1}$. The set of source vectors S is a subset of $\mathbb{C}^p$, the set of single extractor output o is a subset of $\mathbb{C}$ and the set of mixtures Y is a subset of $\mathbb{C}^q$.

In this section, we extend the approach introduced in the Section 7.1.1 to the complex signals. We modify the objective function as

$$J_{ce}(\tilde{\boldsymbol{w}}) = \frac{\sqrt{\frac{1}{N_1} \sum_{l=1}^{N_1} \left(\mathbb{R}\{o(l)\} - \mathbb{R}\{\hat{\mu}_o\} \right)^2}}{\hat{\mathcal{R}}\left(\mathbb{R}\{o\}\right)}, \tag{7.4}$$

where $\mathbb{R}\{\hat{\mu}_o\} = \frac{1}{N_1} \sum_{l=1}^{N_1} \mathbb{R}\{o(l)\}$ and $\hat{\mathcal{R}}\left(\mathbb{R}\{o\}\right)$ is the range of real parts of single output o. We define $\hat{\boldsymbol{R}}_{\underline{\grave{\boldsymbol{s}}}_P}$ as the sample covariance matrix of $\underline{\grave{\boldsymbol{s}}}_P$ where $\underline{\grave{\boldsymbol{s}}}_P(k) = [\ \mathbb{R}\{\underline{\boldsymbol{s}}^T(k)\}$ $\mathbb{I}\{\underline{\boldsymbol{s}}^T(k)\} \quad \ldots \quad \mathbb{R}\{\underline{\boldsymbol{s}}^T(k-P+1)\} \quad \mathbb{I}\{\underline{\boldsymbol{s}}^T(k-P+1)\}\ ]^T$. Under the condition $\hat{\boldsymbol{R}}_{\underline{\grave{\boldsymbol{s}}}_P} \succ 0$, the following theorem shows that maximizing the modified objective function (7.4) achieves the blind source extraction of convolutive mixtures of complex signals.

Theorem 2: Assuming the setup in Section 2.3, $\tilde{\boldsymbol{H}}$ is equalizable by an FIR extractor matrix of order $M-1$ and under the validity of (**C1**), the set of global maxima for J_{ce}

in (7.4) is equal to a subset of perfect extractors.

Proof: The proof is provided in Appendix 7.5.2.

In the iterative algorithm, we maximize the logarithm of J_{ce}, therefore, the objective function is modified as

$$\bar{J}_{ce}(\grave{\boldsymbol{w}}) = \log\left(J_{ce}(\grave{\boldsymbol{w}})\right) = \frac{1}{2}\log\left(\grave{\boldsymbol{w}}^T\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_M}\grave{\boldsymbol{w}}\right) - \log\left(\hat{\mathcal{R}}(\mathbb{R}\{o\})\right), \tag{7.5}$$

where

$$\grave{\boldsymbol{w}} = [\ \mathbb{R}\{\boldsymbol{w}^T(0)\} \ \ -\mathbb{I}\{\boldsymbol{w}^T(0)\} \ \ \ldots \ \ -\mathbb{I}\{\boldsymbol{w}^T(M-1)\}\ ]^T,$$

and $\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_M}$ is the sample covariance matrix of $\grave{\boldsymbol{y}}_M$. Following similar steps, the iterative algorithm for maximizing objective function (7.5) is provided as

$$\grave{\boldsymbol{w}}^{(i+1)} = \grave{\boldsymbol{w}}^{(i)} + \mu^{(i)}\left(\frac{\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_M}\grave{\boldsymbol{w}}}{\grave{\boldsymbol{w}}^T\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_M}\grave{\boldsymbol{w}}} - \frac{1}{\hat{\mathcal{R}}(\mathbb{R}\{o\}^{(i)})}\left(\grave{\boldsymbol{y}}_M(l^{max(i)}) - \grave{\boldsymbol{y}}_M(l^{min(i)})\right)\right), \tag{7.6}$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l^{max(i)}$ ($l^{min(i)}$) is the sample index for which the maximum (minimum) value of the real part of the extractor output is achieved at the i^{th} iteration. Finally, we can obtain $\tilde{\boldsymbol{w}}$ from $\grave{\boldsymbol{w}}$ using a simple transition $\tilde{\boldsymbol{w}}_{mq+1:(m+1)q} = \grave{\boldsymbol{w}}_{2mq+1:2(m+1)q-q} - j\grave{\boldsymbol{w}}_{2(m+1)q-q+1:2mq}$ for $m = 0, 1, \ldots, M-1$.

7.2 Blind Source Separation

In this section, we first introduce an objective function for the blind source separation of real signals. We then prove that the global maxima of the introduced objective function correspond to the perfect separators. We next provide a family of alternative objective functions. After producing the iterative algorithms corresponding to the introduced objective functions, we conclude with the complex extension of the proposed approaches.

7.2.1 Criteria

In order to define the first objective function, we use a similar geometric setting introduced in [1]. Defining the set $O_K = \{\tilde{\boldsymbol{o}}_K(K), \tilde{\boldsymbol{o}}_K(K+1), \ldots, \tilde{\boldsymbol{o}}_K(N-M+1)\}$, we introduce the following objects corresponding to the sets of output samples O_K and O:

- $\mathcal{P}(O_K)$: This is the hyper-ellipsoid whose center is given by the sample mean of the set O_K, its principal semiaxes directions are determined by the eigenvectors of the sample covariance matrix $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K}$ corresponding to O_K and its principal semiaxes lengths are equal to the principal standard deviations, i.e., the square roots of the eigenvalues of $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K}$.
- $\mathcal{B}(O)$: This is the bounding hyper-rectangle which is defined as minimum volume box covering all the samples in O and aligning with the coordinate axes.

The first objective function that we introduce for blind source separation is

$$J_{s1}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K})}\right)^{1/K}}{\prod_{m=1}^{p} \hat{\mathcal{R}}(o_m)}, \tag{7.7}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\tilde{\boldsymbol{o}}_K} &= \frac{1}{N_2} \sum_{l=K}^{N_1} \tilde{\boldsymbol{o}}_K(l), \\ \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K} &= \frac{1}{N_2} \sum_{l=K}^{N_1} (\tilde{\boldsymbol{o}}_K(l) - \boldsymbol{\mu}_{\tilde{\boldsymbol{o}}_K})(\tilde{\boldsymbol{o}}_K(l) - \boldsymbol{\mu}_{\tilde{\boldsymbol{o}}_K})^T, \end{aligned}$$

$N_2 = N_1 - K + 1$ such that $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K}$ is the sample covariance matrix of $\tilde{\boldsymbol{o}}_K$. $\hat{\mathcal{R}}(o_m)$ is the range of the m'th component of the output vectors in the set O and we choose $K \geq P$ where P is the order of the overall system.

We note that, as defined in [1],

- $\sqrt{\det(\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K})}$ refers to the scaled volume of principal hyper-ellipse for the extended output vector $\tilde{\boldsymbol{o}}_K$.
- $\prod_{m=1}^{p} \hat{\mathcal{R}}(o_m)$ is the volume of the bounding hyper-rectangle for the output vector $\boldsymbol{o}$.

Under the condition $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{s}}_{K+P-1}} \succ 0$, the following theorem shows that maximizing the objective function (7.7) achieves the blind source separation of convolutive mixtures whose setup is outlined in Section 2.3.

Theorem 3: Assuming the setup in Section 2.3, $\tilde{\boldsymbol{H}}$ is equalizable by an FIR separator matrix of order $M-1$ and under the validity of (**C1**), the set of global maxima for J_{s1} in (7.7) is equal to the set of perfect separator matrices.

Proof: The proof is provided in Appendix 7.5.3.

We can propose different alternatives for the denominator of the objective function (7.7) (measure of the size of the bounding hyperrectangle for the output vectors). We can choose the length of the main diagonal of the bounding hyperrectangle as a measure of the size instead of its volume. As a result, we obtain a family of alternative objective functions in the form

$$J_{s2,r}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K})}\right)^{1/K}}{||\hat{\mathcal{R}}(\mathbf{o})||_r^p}, \tag{7.8}$$

where $r \geq 1$. We provide the results of analysing this family of objective functions, for some special r values (i.e., $r = 1, 2, \infty$) in Appendix 7.5.4.

7.2.2 Algorithms

In this section, we provide the iterative algorithms corresponding to the optimization settings presented in the previous section.

- *Objective Function* $J_{s1}(\tilde{\boldsymbol{W}})$:

  Similar to the approach in blind source extraction, rather than maximizing $J_{s1}(\tilde{\boldsymbol{W}})$, we maximize its logarithm. Therefore, the new objective function is modified as

$$\begin{aligned}\bar{J}_{s1}(\tilde{\boldsymbol{W}}) &= \log\left(J_{s1}(\tilde{\boldsymbol{W}})\right)\\ &= \frac{1}{2K}\log\left(\det\left(\Gamma_K(\tilde{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_{K+M-1}}\Gamma_K(\tilde{\boldsymbol{W}})^T\right)\right) - \log\left(\prod_{m=1}^{p}\hat{\mathcal{R}}(o_m)\right),\end{aligned} \tag{7.9}$$

  where $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_{K+M-1}}$ is the sample covariance matrix of $\tilde{\boldsymbol{y}}_{K+M-1}$. Note that the derivative of the first part of $\bar{J}_{s1}(\tilde{\boldsymbol{W}})$ with respect to $\tilde{\boldsymbol{W}}$ is

$$\frac{\partial \log\left(\det\left(\Gamma_K(\tilde{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_{K+M-1}}\Gamma_K(\tilde{\boldsymbol{W}})^T\right)\right)}{\partial \tilde{\boldsymbol{W}}} = 2\sum_{l=0}^{K-1}\boldsymbol{A}_{lp+1:(l+1)p,lq+1:(l+M)q}$$

  where $\boldsymbol{A} = \left(\Gamma_K(\tilde{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_{K+M-1}}\Gamma_K(\tilde{\boldsymbol{W}})^T\right)^{-1}\Gamma_K(\tilde{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{y}}_{K+M-1}}$. Following the similar steps as in [1] for the derivative of $\log\left(\prod_{m=1}^{p}\hat{\mathcal{R}}(o_m)\right)$, the subgradient based iterative algorithm for maximizing objective function (7.9) is provided as

$$\tilde{\boldsymbol{W}}^{(i+1)} = \tilde{\boldsymbol{W}}^{(i)} + \mu^{(i)} \Bigg(\frac{1}{K} \sum_{l=0}^{K-1} \boldsymbol{A}_{lp+1:(l+1)p,lq+1:(l+M)q} - \sum_{m=1}^{p} \frac{1}{\mathbf{e}_m^T \hat{\mathcal{R}}(\boldsymbol{o}^{(i)})} \mathbf{e}_m \left(\tilde{\boldsymbol{y}}_M(l_m^{max(i)}) - \tilde{\boldsymbol{y}}_M(l_m^{min(i)}) \right)^T \Bigg), \tag{7.10}$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l_m^{max(i)}$ ($l_m^{min(i)}$) is the sample index for which the maximum (minimum) value for the m^{th} separator output is achieved at the i^{th} iteration.

- *Objective Function* $J_{s2,r}(\tilde{\boldsymbol{W}})$:

We note that for the family of objective functions (7.8), the update equation is similar to (7.10) where the change is in the derivative of logarithm of the denominator depending on the selection of r. For $r = 1, 2$, we can write the update equation as

$$\tilde{\boldsymbol{W}}^{(i+1)} = \tilde{\boldsymbol{W}}^{(i)} + \mu^{(i)} \Bigg(\frac{1}{K} \sum_{l=0}^{K-1} \boldsymbol{A}_{lp+1:(l+1)p,lq+1:(l+M)q} - \sum_{m=1}^{p} \frac{p\hat{\mathcal{R}}_m(\boldsymbol{o}^{(i)})^{r-1}}{||\hat{\mathcal{R}}(\boldsymbol{o}^{(i)})||_r^r} \mathbf{e}_m \left(\tilde{\boldsymbol{y}}_M(l_m^{max(i)}) - \tilde{\boldsymbol{y}}_M(l_m^{min(i)}) \right)^T \Bigg).$$

For $r = \infty$, the update equation has the form

$$\tilde{\boldsymbol{W}}^{(i+1)} = \tilde{\boldsymbol{W}}^{(i)} + \mu^{(i)} \Bigg(\frac{1}{K} \sum_{l=0}^{K-1} \boldsymbol{A}_{lp+1:(l+1)p,lq+1:(l+M)q} - \sum_{m \in \mathcal{M}(\boldsymbol{O}^i)} \frac{p\beta_m^{(i)}}{||\hat{\mathcal{R}}(\boldsymbol{o}^{(i)})||_\infty} \mathbf{e}_m \left(\tilde{\boldsymbol{y}}_M(l_m^{max(i)}) - \tilde{\boldsymbol{y}}_M(l_m^{min(i)}) \right)^T \Bigg)$$

where $\mathcal{M}(\boldsymbol{o}^{(i)})$ is the set of indexes for which the peak range value is achieved, i.e.,

$$\mathcal{M}(\boldsymbol{o}^{(i)}) = \{m : \hat{\mathcal{R}}_m(\boldsymbol{o}^{(i)}) = \|\hat{\mathcal{R}}(\boldsymbol{o}^{(i)}\|_\infty\}, \tag{7.11}$$

and $\beta_m^{(i)}$s are the convex combination coefficients.

7.2.3 Extension to Complex Signals

In the complex domain, both mixing and separator coefficient matrices are complex matrices, i.e., $\tilde{\boldsymbol{H}} \in \mathbb{C}^{q\times pL}$ and $\tilde{\boldsymbol{W}} \in \mathbb{C}^{p\times qM}$. The set of source vectors S and the set of separator outputs O are a subset of $\mathbb{C}^p$, the set of mixtures Y is a subset of $\mathbb{C}^q$.

In this section, we extend the approach introduced in the Section 7.2.1 to the complex signals. We modify the first objective function for the blind source separation of complex signals as

$$J_{cs1}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\hat{\boldsymbol{R}}_{\grave{\boldsymbol{o}}_K})}\right)^{1/K}}{\prod_{m=1}^{2p} \hat{\mathcal{R}}(\grave{o}_m)}, \tag{7.12}$$

where $\hat{\boldsymbol{R}}_{\grave{\boldsymbol{o}}_K}$ is the sample covariance matrix of $\grave{\boldsymbol{o}}_K$ where $\grave{\boldsymbol{o}}_K(k) = [\ \mathbb{R}\{\boldsymbol{o}^T(k)\}\ \ \mathbb{I}\{\boldsymbol{o}^T(k)\}\ \ \dots\ \ \mathbb{R}\{\boldsymbol{o}^T(k-K+1)\}\ \ \mathbb{I}\{\boldsymbol{o}^T(k-K+1)\}\ ]^T$ and $\prod_{m=1}^{2p} \hat{\mathcal{R}}(\grave{o}_m)$ is the product of ranges of real and imaginary parts of all separator outputs.

Under the condition $\hat{\boldsymbol{R}}_{\grave{\boldsymbol{s}}_{K+P-1}} \succ 0$, the following theorem shows that maximizing the modified objective function (7.12) achieves the blind source separation of convolutive mixtures of complex signals.

Theorem 4: Assuming the setup in Section 2.3, $\tilde{\boldsymbol{H}}$ is equalizable by an FIR separator matrix of order $M-1$ and under the validity of (**C1**), the set of global maxima for J_{cs1} in (7.12) is equal to a subset of perfect separator matrices.

Proof: The proof is provided in Appendix 7.5.5.

In the iterative algorithm, we maximize the logarithm of J_{cs1}, therefore, the first objective function is modified as

$$\bar{J}_{cs1}(\tilde{\boldsymbol{W}}) = \log\left(J_{cs1}(\tilde{\boldsymbol{W}})\right)$$
$$= \frac{1}{2K}\log\left(\det\left(\Gamma_{2K}(\grave{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_{K+M-1}}\Gamma_{2K}(\grave{\boldsymbol{W}})^T\right)\right) - \log\left(\prod_{m=1}^{2p}\hat{\mathcal{R}}(\grave{o}_m)\right), \quad (7.13)$$

where $\grave{\boldsymbol{W}} = \begin{bmatrix} \mathbb{R}\{\boldsymbol{W}_0\} & -\mathbb{I}\{\boldsymbol{W}_0\} & \dots & \mathbb{R}\{\boldsymbol{W}_{M-1}\} & -\mathbb{I}\{\boldsymbol{W}_{M-1}\} \\ \mathbb{I}\{\boldsymbol{W}_0\} & \mathbb{R}\{\boldsymbol{W}_0\} & \dots & \mathbb{I}\{\boldsymbol{W}_{M-1}\} & \mathbb{R}\{\boldsymbol{W}_{M-1}\} \end{bmatrix}$ and $\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_{K+M-1}}$ is the sample covariance matrix of $\grave{\boldsymbol{y}}_{K+M-1}$.

The corresponding iterative update equation of $\boldsymbol{W}(n)$ for $n = 0, 1, \dots, M-1$ can be written as

$$\boldsymbol{W}^{(i+1)}(n) = \boldsymbol{W}^{(i)}(n) + \mu^{(i)}\Bigg(\boldsymbol{C}_{1:p,2nq+1:(2n+1)q} + \boldsymbol{C}_{p+1:2p,(2n+1)q+1:2(n+1)q} +$$
$$j\left(\boldsymbol{C}_{p+1:2p,2nq+1:(2n+1)q} - \boldsymbol{C}_{1:p,(2n+1)q+1:2(n+1)q}\right) -$$
$$\sum_{m=1}^{2p}\frac{1}{\mathbf{e}_m^T\hat{\mathcal{R}}(\grave{\boldsymbol{o}}^{(i)})}\mathbf{v}_m\left(\tilde{\boldsymbol{y}}_M(l_m^{max(i)}) - \tilde{\boldsymbol{y}}_M(l_m^{min(i)})\right)^H\Bigg), \quad (7.14)$$

where $\boldsymbol{C} = \frac{1}{K}\sum_{l=0}^{K-1}\boldsymbol{F}_{2lp+1:2(l+1)p,2lq+1:2(l+M)q}$
and $\boldsymbol{F} = \left(\Gamma_{2K}(\grave{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_{K+M-1}}\Gamma_{2K}(\grave{\boldsymbol{W}})^T\right)^{-1}\Gamma_{2K}(\grave{\boldsymbol{W}})\hat{\boldsymbol{R}}_{\grave{\boldsymbol{y}}_{K+M-1}}$ and

$$\mathbf{v}_m = \begin{cases} \mathbf{e}_m & m \le p, \\ i\mathbf{e}_{m-p} & m > p. \end{cases} \quad (7.15)$$

Similar to the complex extension of J_{s1}, we can extend the J_{s2} family by modifying

$$J_{cs2,r}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\hat{\boldsymbol{R}}_{\grave{\boldsymbol{o}}_K})}\right)^{1/K}}{||\hat{\mathcal{R}}(\grave{\boldsymbol{o}})||_r^{2p}}. \tag{7.16}$$

The update equation is similar to (7.14) where the change is in the derivative of logarithm of the denominator depending on the selection of r, e.g.,

- $r = 1, 2$ *Case*: In this case

  $$\frac{\partial \log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right)}{\partial \tilde{\boldsymbol{W}}} = \sum_{m=1}^{2p} \frac{p\hat{\mathcal{R}}_m(\grave{\boldsymbol{o}}^{(i)})^{r-1}}{||\hat{\mathcal{R}}(\grave{\mathbf{o}}^{(i)})||_r^r} \mathbf{v}_m \left(\tilde{\boldsymbol{y}}_M(l_m^{max(i)}) - \tilde{\boldsymbol{y}}_M(l_m^{min(i)})\right)^H$$

  where $\mathbf{v}_m$ is as defined in (7.15).

- $r = \infty$ *Case*: In this case

  $$\frac{\partial \log\left(||\mathcal{R}(\grave{\boldsymbol{o}})||_r^{2p}\right)}{\partial \tilde{\boldsymbol{W}}} = \sum_{m \in \mathcal{M}(\grave{\mathbf{o}}^{(i)})} \frac{p\beta_m^{(i)}}{||\hat{\mathcal{R}}(\grave{\mathbf{o}}^{(i)})||_\infty} \mathbf{v}_m \left(\tilde{\boldsymbol{y}}_M(l_m^{max(i)}) - \tilde{\boldsymbol{y}}_M(l_m^{min(i)})\right)^H$$

  where $\mathbf{v}_m$ is as defined in (7.15),

  $$\mathcal{M}(\grave{\mathbf{o}}^{(i)}) = \{m : \hat{\mathcal{R}}_m(\grave{\mathbf{o}}^{(i)}) = \|\hat{\mathcal{R}}(\grave{\mathbf{o}}^{(i)})\|_\infty\},$$

  and $\beta_m^{(i)}$s are the convex combination coefficients.

7.3 Examples

In this section, we illustrate the extraction/separation capability of the proposed algorithms for the convolutive mixtures of both independent and dependent sources.

7.3.1 Blind Source Extraction

We first consider the following scenario to illustrate the performance of the proposed blind source extraction algorithm regarding the convolutive mixtures of space-time correlated sources: In order to generate space-time correlated sources, we first generate a samples of a τp size vector, $\mathbf{d}$, with Copula-t distribution, a perfect tool for generating vectors with controlled correlation, with 4 degrees of freedom whose correlation matrix parameter is given by $\mathbf{R} = \mathbf{R}_t \otimes \mathbf{R}_s$ where $\mathbf{R}_t$ ($\mathbf{R}_s$) is a Toeplitz matrix whose first row is $\begin{bmatrix} 1 & \rho_t & \ldots & \rho_t^{\tau-1} \end{bmatrix}$ ($\begin{bmatrix} 1 & \rho_s & \ldots & \rho_s^{p-1} \end{bmatrix}$). Each sample of $\mathbf{d}$ is partitioned to produce source vectors, $\mathbf{d}(k) = \begin{bmatrix} \mathbf{s}(k\tau) & \mathbf{s}(k\tau+1) & \ldots & \mathbf{s}((k+1)\tau-1) \end{bmatrix}$. Therefore, we obtain the source vectors as samples of a wide-sense cyclostationary process whose correlation structure in time direction and space directions are governed by the parameters ρ_t and ρ_s, respectively.

In the simulations, we consider a scenario with 7 sources and 20 mixtures, an i.i.d. Gaussian convolutive mixing system with order 7 and a extractor of order 8. We set $\rho_s = 0.5$, $\rho_t = 0.5$ and $\tau = 5$. We note that the sources are non-stationary in this case (we will cover stationary sources in the digital communication sources scenario).

Figure 7.1 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithm ($\bar{J}_e$) for various sample lengths under 45dB SNR. SINR performance of Minimum Mean Square Error (MMSE) filter of the same order, which uses full information about mixing system and source/noise statistics, is also shown to evaluate the relative success of the proposed approach. A comparison has also been made with a gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35] where we take $k_{max} = 50$ and $l_{max} = 20$. We have obtained these methods from [2], [36]. As we did not encounter any convolutive BSS algorithm with correlated source separation capability, we compared our algorithm with some well known convolutive ICA approaches.

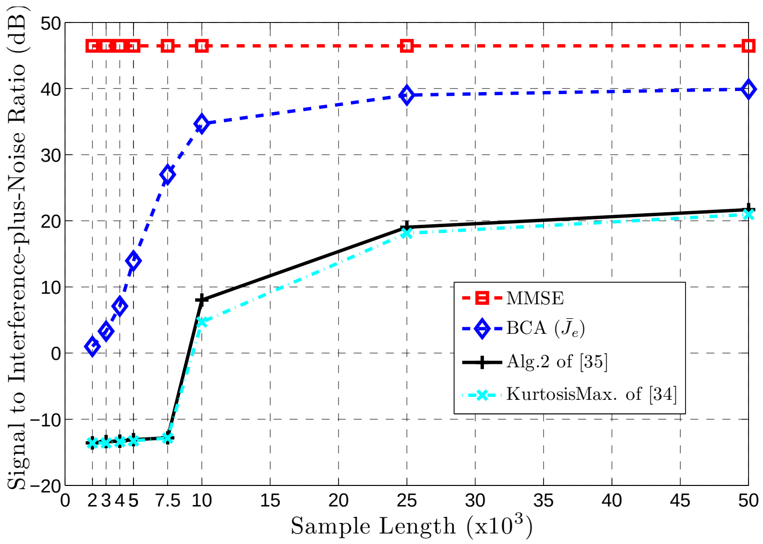

Figure 7.1: Result of the proposed blind source extraction algorithm performance for the convolutive mixtures of dependent sources (ρ_s and ρ_t is set as 0.5) for various sample lengths under SNR = 45dB.

For the same setup, Figure 7.2 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithm ($\bar{J}_e$), gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.), **Alg.2** of [35], and MMSE for various sample lengths under 20dB SNR.

These results demonstrate that the performance of the proposed blind source extraction algorithm is approaching fast to its MMSE counterpart as the sample length increases. On the other hand, the performance of gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35] is far from the performance of MMSE filter even when the sample length is increased (Figure 7.1) or they require more sample lengths to reach the same SINR performance (Figure 7.2) since in the correlated case, independence assumption simply fails. Therefore, we observe that the proposed BCA approach is capable of blind source extraction of convolutive mixtures of space-time correlated sources.

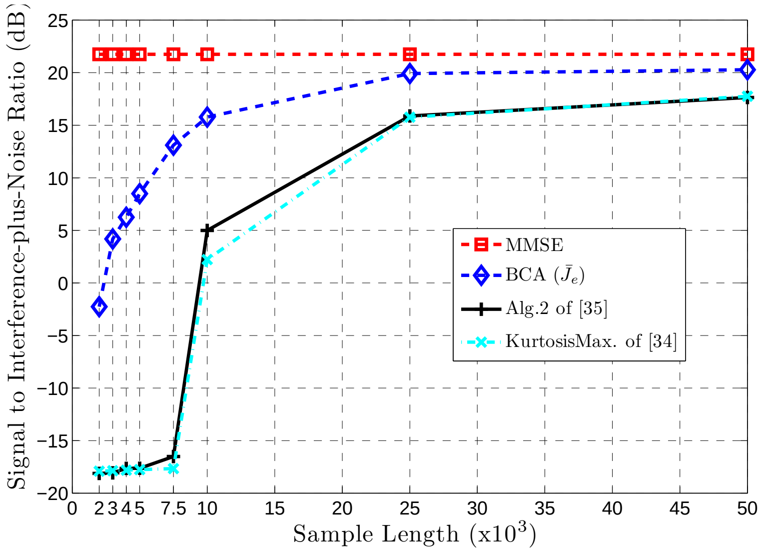

Figure 7.2: Result of the proposed blind source extraction algorithm performance for the convolutive mixtures of dependent sources (ρ_s and ρ_t is set as 0.5) for various sample lengths under SNR = 20dB.

7.3.2 Blind Source Separation

We first consider a similar scenario as in the blind source extraction examples to illustrate the performance of the proposed blind source separation algorithms regarding the separability of convolutive mixtures of space-time correlated sources.

Here, we consider a scenario with 5 sources and 15 mixtures, an i.i.d. Gaussian convolutive mixing system with order 5 and a separator of order 6 where the sample size is 50000.

Figure 7.3 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for proposed BCA algorithms ($\bar{J}_{s1}, \bar{J}_{s2,1}, \bar{J}_{s2,2}\bar{J}_{s2,\infty}$) for various space correlation parameters under 45dB SNR. The performances of MMSE, gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35] are also plotted for comparison. We note that the algorithm $\bar{J}_{s1}$ yields better performance than the other algorithms.

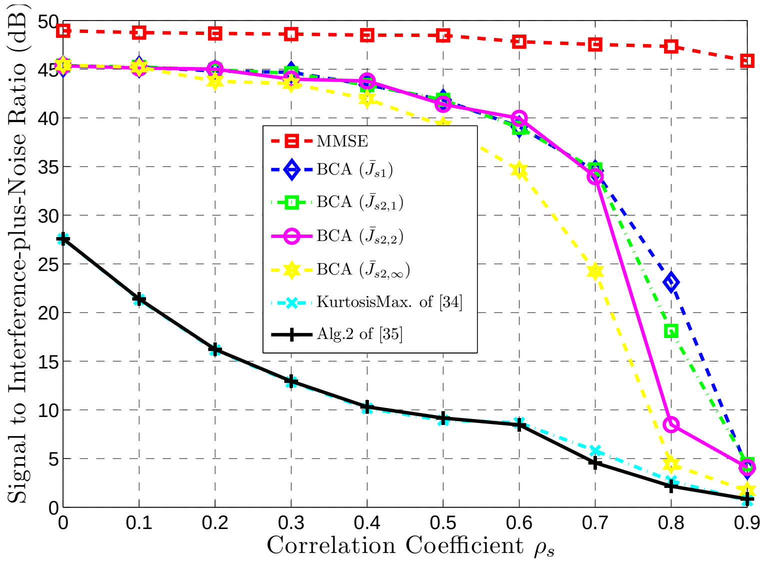

Figure 7.3: Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 45dB.

For the same setup, Figure 7.4 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the BCA algorithm ($\bar{J}_{s1}$), gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.), **Alg.2** of [35], and MMSE for various space correlation parameters under 20dB SNR.

These results demonstrate that the performance of proposed blind source separation algorithms closely follow its MMSE counterpart for a wide range of correlation values. Therefore, we obtain a convolutive extension of the BCA approach introduced in [1], which is capable of separating convolutive mixtures of space-time correlated sources.

Also note that the proposed blind source separation algorithms maintain high separation performance for various space parameters. However, the performance of gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35] degrades substantially with increasing correlation since the independence assumption does not hold. We point out that when $\rho_s = 0$ the sources are independent, yet BCA algorithms still outperforms other ICA algorithms. This result can be attributed to the finite

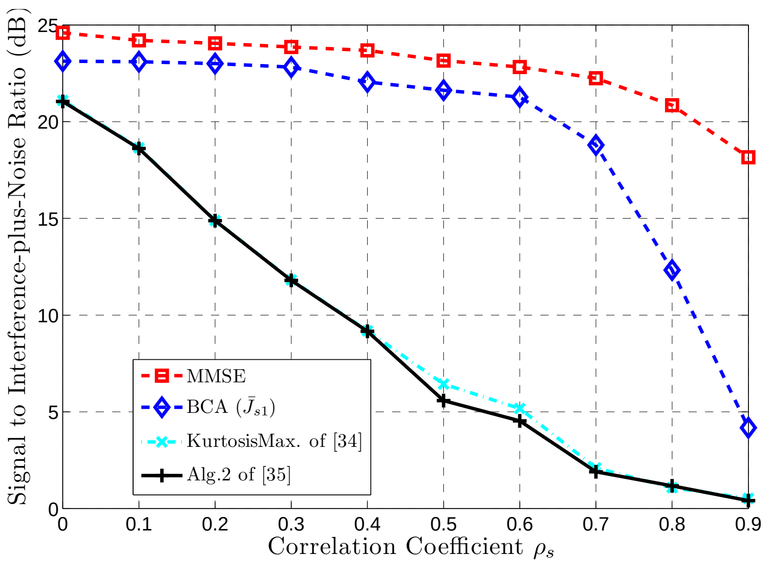

Figure 7.4: Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 20dB.

sample effects. In other words, although the sources are stochastically independent, finite samples may not reflect this behaviour and the sources may even have non-zero sample correlation. BCA algorithms being robust to such correlations can offer better performance. Effect of the sample size will be investigated in the next scenario.

We next consider the following scenario to illustrate the performance of the proposed blind source separation algorithm for the convolutive mixtures of digital communication sources. We consider 5 complex 4-QAM sources where we take 15 mixtures, an i.i.d. Gaussian convolutive mixing system with order 5 and a separator of order 6. The sources are stationary in this case. We use the objective function $\bar{J}_{cs1}$ as the BCA algorithm for this simulation. The resulting Signal to Interference Ratio is plotted with respect to the sample lengths in Figure 7.5. We have also compared our algorithm with a gradient maximization of the criterion (kurtosis) of [34] (KurtosisMax.) and **Alg.2** of [35].

As it can be observed from Figure 7.5, the proposed BCA approach achieves better performance than ICA based approaches. We again note that, the proposed method

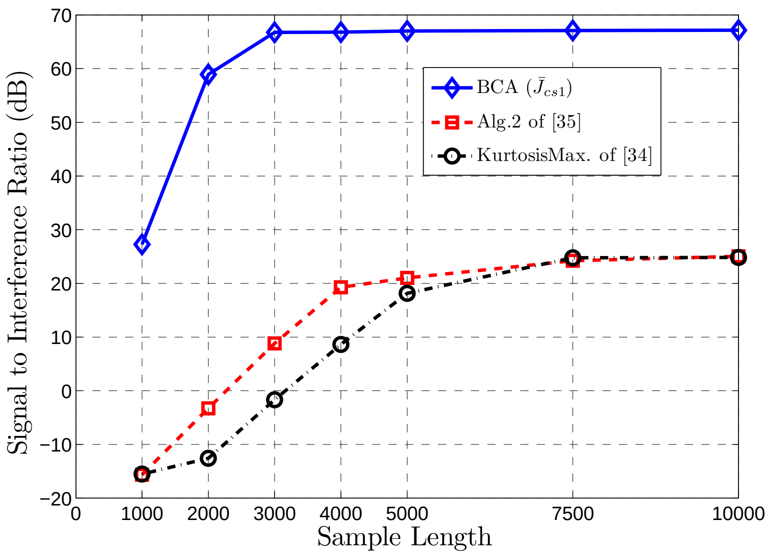

Figure 7.5: Result of the proposed blind source separation algorithm performance for the convolutive mixtures of digital communication sources for various sample lengths.

does not assume/exploit statistical independence. The only impact of short data length is on accurate representation of source box boundaries. The simulation results suggest that the shorter data records may not be sufficient to reflect the stochastic independence of the sources, and therefore, the compared algorithms require more data samples to achieve the same SIR level as the proposed approach.

7.4 Conclusion

In this section, we introduced deterministic and geometric frameworks for the convolutive BCA problem. We proposed blind source extraction and blind source separation algorithms based on certain deterministic measures obtained from the geometric objects of samples which can be used for the extraction/separation of both independent and dependent (even correlated) sources. The numerical examples illustrate that the proposed frameworks are capable of extracting/separating space-time correlated sources from their convolutive mixtures. Moreover, even when the sources are independent, having short sample lengths may not reflect the independence behaviour. Hence, the proposed approaches expectedly provide better performances than the ICA based approaches regarding separation of independent sources especially for short sample records.

7.5 Appendix

7.5.1 Proof of Theorem 1

We first note that, following similar steps as in [1], when the assumption (**C1**) holds, we can write the range of o as $\hat{\mathcal{R}}(o) = ||\tilde{\boldsymbol{g}}^T \tilde{\Lambda}||_1$ where $\tilde{\Lambda} = I \otimes \Lambda$ is the range matrix of $\tilde{\boldsymbol{s}}_P$.

Since $o(l) = \tilde{\boldsymbol{g}}^T \tilde{\boldsymbol{s}}_P(l + M - 1)$ for $l = 1, 2, \ldots, N_1$, we have

$$\frac{1}{N_1}\sum_{l=1}^{N_1}(o(l) - \hat{\mu}_o)^2 = \frac{1}{N_1}\sum_{l=M}^{N}\tilde{\boldsymbol{g}}^T(\tilde{\boldsymbol{s}}_P(l) - \boldsymbol{\mu}_{\tilde{\boldsymbol{s}}_P})(\tilde{\boldsymbol{s}}_P(l) - \boldsymbol{\mu}_{\tilde{\boldsymbol{s}}_P})^T\tilde{\boldsymbol{g}}$$
$$= \frac{1}{N_1}\sum_{l=M}^{N}\tilde{\boldsymbol{g}}^T\tilde{\Lambda}(\underline{\tilde{\boldsymbol{s}}}_P(l) - \boldsymbol{\mu}_{\underline{\tilde{\boldsymbol{s}}}_P})(\underline{\tilde{\boldsymbol{s}}}_P(l) - \boldsymbol{\mu}_{\underline{\tilde{\boldsymbol{s}}}_P})^T\tilde{\Lambda}^T\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{g}}^T\tilde{\Lambda}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P}\tilde{\Lambda}^T\tilde{\boldsymbol{g}}, \tag{7.17}$$

where $\boldsymbol{\mu}_{\tilde{\boldsymbol{s}}_P} = \frac{1}{N_1}\sum_{l=M}^{N}\tilde{\boldsymbol{s}}_P(l)$, $\boldsymbol{\mu}_{\underline{\tilde{\boldsymbol{s}}}_P} = \frac{1}{N_1}\sum_{l=M}^{N}\underline{\tilde{\boldsymbol{s}}}_P(l)$ and $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P}$ is defined as the sample covariance matrix of $\underline{\tilde{\boldsymbol{s}}}_P$.

We can further define $\boldsymbol{q}^T = \tilde{\boldsymbol{g}}^T\tilde{\Lambda}$ and rewrite the equality (7.1) in terms of $\boldsymbol{q}$ as

$$J_e(\boldsymbol{q}) = \frac{\sqrt{\boldsymbol{q}^T\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P}\boldsymbol{q}}}{||\boldsymbol{q}||_1}. \tag{7.18}$$

Note that, maximizing $J_e(\boldsymbol{q})$ is equivalent to the corresponding optimization setting

$$\begin{aligned} \text{maximize} \quad & \sqrt{\boldsymbol{q}^T \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \boldsymbol{q}} \\ \text{s.t.} \quad & ||\boldsymbol{q}||_1 \leq \gamma \end{aligned}$$

where γ is a constant. Also note that, assuming $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \succ 0$, $\sqrt{\boldsymbol{q}^T \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \boldsymbol{q}}$ is a convex function and the region of $||\boldsymbol{q}||_1 \leq \gamma$ corresponds to a convex polytope. From the definition of a convex polytope (Vertex Representation [37]), this is the convex hull of the vertices of polytope. Therefore, the maximum of $\sqrt{\boldsymbol{q}^T \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \boldsymbol{q}}$ will be attained at one of the vertices (whichever has the maximum value) and therefore, the maximum will be attained when $\boldsymbol{q}$ has only one non-zero component. To see that, we can take any vector $\boldsymbol{q}_i$ inside the convex polytope (i.e., satisfying $||\boldsymbol{q}_i||_1 \leq \gamma$). From the definition of vertex representation [37], $\boldsymbol{q}_i = \alpha_1 \boldsymbol{q}_{v_1} + \alpha_2 \boldsymbol{q}_{v_2} + \ldots + \alpha_{pP} \boldsymbol{q}_{v_{pP}}$ where $\boldsymbol{q}_{v_1}, \boldsymbol{q}_{v_2}, \ldots, \boldsymbol{q}_{v_{pP}}$ are vertices of polytope and $\sum_{l=1}^{pP} \alpha_l = 1$. Defining $f(\boldsymbol{q}) = \sqrt{\boldsymbol{q}^T \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \boldsymbol{q}}$ and using Jensen's inequality, we have

$$f(\boldsymbol{q}) \leq \alpha_1 f(\boldsymbol{q}_{v_1}) + \alpha_2 f(\boldsymbol{q}_{v_2}) + \ldots + \alpha_{pP} f(\boldsymbol{q}_{v_{pP}}) \leq \max\{f(\boldsymbol{q}_{v_1}), f(\boldsymbol{q}_{v_2}), \ldots, f(\boldsymbol{q}_{v_{pP}})\}.$$

Therefore, the maximum is attained at the vertex which has the maximum value and this yields that $\boldsymbol{q}$ has only one non-zero component.

To observe that from a geometric point of view, assuming $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \succ 0$, for any constant γ, the vectors $\boldsymbol{q}$ satisfying $\sqrt{\boldsymbol{q}^T \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \boldsymbol{q}} = \gamma$ constitutes an hyper-ellipsoid. Note that, for any constant value of $||\boldsymbol{q}||_1$, the maxima of $\sqrt{\boldsymbol{q}^T \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_P} \boldsymbol{q}}$ will be attained at one of the corner points (i.e., where $\boldsymbol{q}$ has only one non-zero component). A two dimensional example is illustrated in Figure 7.6.

Since $\tilde{\boldsymbol{g}}^T = \boldsymbol{q}^T \tilde{\Lambda}^{-1}$, $\tilde{\boldsymbol{g}}$ will also have only one non-zero component, therefore, the global

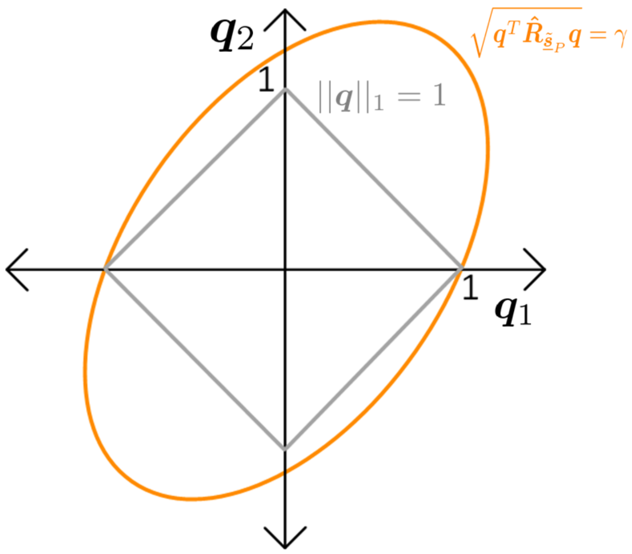

Figure 7.6: Two dimensional example for the global maxima of (4.63).

maxima of (7.1) correspond to perfect extractors.

7.5.2 Proof of Theorem 2

We begin with noting that

$$\mathbb{R}\{o(k)\} = \sum_{l=0}^{P-1} \mathbb{R}\{\boldsymbol{g}^T(l)\}\mathbb{R}\{\boldsymbol{s}(k+M-1-l)\} - \mathbb{I}\{\boldsymbol{g}^T(l)\}\mathbb{I}\{\boldsymbol{s}(k+M-1-l)\}.$$

Defining $\grave{\boldsymbol{g}} = \left[\, \mathbb{R}\{\boldsymbol{g}^T(0)\} \;\; -\mathbb{I}\{\boldsymbol{g}^T(0)\} \;\; \ldots \;\; -\mathbb{I}\{\boldsymbol{g}^T(P-1)\} \,\right]^T$ and $\grave{\boldsymbol{s}}_P(k) = [\, \mathbb{R}\{\boldsymbol{s}^T(k)\} \;\; \mathbb{I}\{\boldsymbol{s}^T(k)\} \;\; \ldots \;\; \mathbb{I}\{\boldsymbol{s}^T(k-P+1)\} \,]^T$, we obtain

$$\mathbb{R}\{o(k)\} = \grave{\boldsymbol{g}}^T \grave{\boldsymbol{s}}_P(k+M-1),$$

for $k = 1, 2, \ldots, N_1$. Following similar steps, we can write the range of $\mathbb{R}\{o\}$ as $\hat{\mathcal{R}}\left(\mathbb{R}\{o\}\right) = ||\grave{\boldsymbol{g}}^T \grave{\Lambda}||_1$ where $\grave{\Lambda} = I \otimes \Lambda$ is the range matrix of $\grave{\boldsymbol{s}}_P$. Similar to (7.17), we have

$$\frac{1}{N_1}\sum_{l=1}^{N_1}(\mathbb{R}\{o(l)\}-\mathbb{R}\{\hat{\mu}_o\})^2 = \grave{\boldsymbol{g}}^T\grave{\Lambda}\hat{\boldsymbol{R}}_{\underline{\grave{\boldsymbol{s}}}_P}\grave{\Lambda}^T\grave{\boldsymbol{g}},$$

where $\hat{\boldsymbol{R}}_{\underline{\grave{\boldsymbol{s}}}_P}$ is defined as the sample covariance matrix of $\underline{\grave{\boldsymbol{s}}}_P$. Defining $\grave{\boldsymbol{q}}^T = \grave{\boldsymbol{g}}^T\grave{\Lambda}$ and rewriting the equality (7.4) in terms of $\grave{\boldsymbol{q}}$ yields

$$J_{ce}(\grave{\boldsymbol{q}}) = \frac{\sqrt{\grave{\boldsymbol{q}}^T\hat{\boldsymbol{R}}_{\underline{\grave{\boldsymbol{s}}}_P}\grave{\boldsymbol{q}}}}{||\grave{\boldsymbol{q}}||_1}. \tag{7.19}$$

Following similar analogy, as a result, the maximum of (7.4) is attained when $\grave{\boldsymbol{g}}$ has only one non-zero component which also implies that $\tilde{\boldsymbol{g}}$ has only one non-zero component. Note that the non-zero component of $\tilde{\boldsymbol{g}}$ will be real or purely imaginary. Therefore, the global maxima of (7.4) correspond to a subset of perfect extractors for complex signals.

7.5.3 Proof of Theorem 3

We define the operator Γ_K such that $\Gamma_K(\tilde{\boldsymbol{G}})$ is a block Toeplitz matrix of dimension $Kp \times (K+P-1)p$ whose first block row is $[\boldsymbol{G}(0) \quad \boldsymbol{G}(1) \quad \dots \quad \boldsymbol{G}(P-1) \quad \mathbf{0} \quad \dots \quad \mathbf{0}]$ and first block column is $\left[\boldsymbol{G}^T(0) \quad \mathbf{0} \quad \dots \quad \mathbf{0}\right]^T$ where the zero matrices ($\mathbf{0}$) have the size $p \times p$ same as the matrices $\boldsymbol{G}(l)$ for $l = 0, \dots, P-1$. This yields,

$$\tilde{\boldsymbol{o}}_K(l) = \Gamma_K(\tilde{\boldsymbol{G}})\tilde{\boldsymbol{s}}_{K+P-1}(l+M-1),$$

for $l = K, K+1, \dots, N_1$. Defining $A = K+P-1$, we have

$$\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K} = \Gamma_K(\tilde{\boldsymbol{G}})\breve{\Lambda}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\breve{\Lambda}^T\Gamma_K(\tilde{\boldsymbol{G}})^T,$$

where $\breve{\Lambda} = I \otimes \Lambda$ is the range matrix of $\tilde{\boldsymbol{s}}_A$ and $\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}$ is the sample covariance matrix of $\underline{\tilde{\boldsymbol{s}}}_A$. Defining $\boldsymbol{Q} = \Gamma_K(\tilde{\boldsymbol{G}})\breve{\Lambda}$ yields $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{o}}_K} = \boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T$.

Following similar steps as in [1], under the assumption (**C1**), we can write the range of m^{th} component of $\boldsymbol{o}$ as $\hat{\mathcal{R}}(o_m) = ||\tilde{\boldsymbol{G}}_{m,:}\tilde{\Lambda}||_1$. Note that, $||\tilde{\boldsymbol{G}}_{m,:}\tilde{\Lambda}||_1 = ||\boldsymbol{Q}_{m,:}||_1$ for $m = 1, 2, \ldots, p$. Therefore, the range vector for the separator outputs can be rewritten as

$$\hat{\mathcal{R}}(\boldsymbol{o}) = \left[||\boldsymbol{Q}_{1,:}||_1 \quad ||\boldsymbol{Q}_{2,:}||_1 \quad \ldots \quad ||\boldsymbol{Q}_{p,:}||_1\right].$$

Rewriting the equality (7.7) in terms of $\boldsymbol{Q}$, we obtain

$$J_{s1}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T)}\right)^{1/K}}{\prod_{m=1}^{p}||\boldsymbol{Q}_{m,:}||_1}. \tag{7.20}$$

For any $\tilde{\boldsymbol{G}}$ whose rows are not linearly independent, we have $\det\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right) = 0$, therefore, corresponding $\tilde{\boldsymbol{G}}$ can not be global maxima of (7.7). Hence for any $\tilde{\boldsymbol{G}}$ whose rows are linearly independent, assuming $\hat{\boldsymbol{R}}_{\tilde{\boldsymbol{s}}_{K+P-1}} = \hat{\boldsymbol{R}}_{\tilde{\boldsymbol{s}}_A} \succ 0$, to complete $\boldsymbol{Q}$ into a full rank square matrix we introduce a $(P-1)p \times Ap$ matrix $\boldsymbol{M} = \boldsymbol{D}\boldsymbol{P}$ where $\boldsymbol{D} = \mathrm{diag}(a_1, a_2, \ldots, a_{(P-1)p})$ is a full rank diagonal matrix and $\boldsymbol{P}$ is a permutation matrix such that $\det\left(\boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^T\right) = 1$ where we define $\boldsymbol{B} = \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A} - \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right)^{-1}\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}$. This yields,

$$\begin{aligned}
&\det\left(\begin{bmatrix}\boldsymbol{Q}\\ \boldsymbol{M}\end{bmatrix}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\begin{bmatrix}\boldsymbol{Q}^T & \boldsymbol{M}^T\end{bmatrix}\right) = \det\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right)\\
&\det\left(\boldsymbol{M}\left(\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A} - \hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right)^{-1}\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\right)\boldsymbol{M}^T\right)\\
&= \det\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right)\det\left(\boldsymbol{M}\boldsymbol{B}\boldsymbol{M}^T\right) = \det\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right).
\end{aligned}$$

We note that $\begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{M} \end{bmatrix} \hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A} \begin{bmatrix} \boldsymbol{Q}^T & \boldsymbol{M}^T \end{bmatrix} \succ 0$ and $\boldsymbol{MBM}^T$ is the Schur complement of $\boldsymbol{Q}\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A}\boldsymbol{Q}^T$, therefore, $\boldsymbol{MBM}^T \succ 0$. We also note that $\det\left(\boldsymbol{MBM}^T\right) = a_1^2 a_2^2 \dots a_{(P-1)p}^2$ $\det\left([\boldsymbol{B}]_{per}\right)$ where $[\boldsymbol{B}]_{per}$ has the chosen rows and columns of $\boldsymbol{B}$ depending on the positions of $a_1, a_2, \dots, a_{(P-1)p}$. Hence by choosing appropriate values for $a_1, a_2, \dots, a_{(P-1)p}$ we can obviously introduce a matrix $\boldsymbol{M}$ such that

$$\det\left(\begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{M} \end{bmatrix} \hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A} \begin{bmatrix} \boldsymbol{Q}^T & \boldsymbol{M}^T \end{bmatrix}\right) = \det\left(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A}\boldsymbol{Q}^T\right).$$

Using Hadamard's Inequality [26] yields

$$\det\left(\begin{bmatrix} \boldsymbol{Q} \\ \boldsymbol{M} \end{bmatrix} \hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A} \begin{bmatrix} \boldsymbol{Q}^T & \boldsymbol{M}^T \end{bmatrix}\right) \leq \prod_{m=1}^{Kp} ||\boldsymbol{Q}_{m,:}||_2^2 \prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2^2 \quad \det(\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A}). \quad (7.21)$$

Note that $\prod_{m=1}^{Kp} ||\boldsymbol{Q}_{m,:}||_2^2 = \left(\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_2^2\right)^K$ since $\boldsymbol{Q}$ is block Toeplitz matrix. Hence,

$$\left(\sqrt{\det(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A}\boldsymbol{Q}^T)}\right)^{1/K} \leq \left(\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_2\right)\left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2\right)^{1/K} \det(\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A})^{1/2K}.$$

Therefore, we have

$$\begin{aligned} J_{s1}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A}\boldsymbol{Q}^T)}\right)^{1/K}}{\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_1} &\leq \frac{\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_2}{\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_1}\left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2\right)^{1/K} \det(\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A})^{1/2K} \\ &\leq \left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2\right)^{1/K} \det(\hat{\boldsymbol{R}}_{\tilde{\underline{\boldsymbol{s}}}_A})^{1/2K}, \end{aligned} \quad (7.22)$$

due to the ordering $||\boldsymbol{q}||_1 \geq ||\boldsymbol{q}||_2$ for any $\boldsymbol{q}$.

To achieve the equality in (7.22), the equalities $||\boldsymbol{Q}_{m,:}||_1 = ||\boldsymbol{Q}_{m,:}||_2$ for $m = 1, 2, \dots p$

and the equality in (7.21) should be achieved. The equalities $||\boldsymbol{Q}_{m,:}||_1 = ||\boldsymbol{Q}_{m,:}||_2$ for $m = 1, 2, \ldots p$ are achieved if and only if the first p rows of $\boldsymbol{Q}$ has only one non-zero element. Since $\Gamma_K(\tilde{\boldsymbol{G}}) = \boldsymbol{Q}\breve{\Lambda}^{-1}$, this implies that each row of $\tilde{\boldsymbol{G}}$ has only one non-zero element. The inequality in (7.21) is achieved if and only if the rows of $\boldsymbol{Q}$ are perpendicular to each other and to the rows of $\boldsymbol{M}$ which yields that the rows of $\Gamma_K(\tilde{\boldsymbol{G}})$ are perpendicular to each other and to the rows of $\boldsymbol{M}$. Note that since $K \geq P$, the structure of $\Gamma_K(\tilde{\boldsymbol{G}})$ guarantees that there is a block column which contains $\boldsymbol{G}(0), \boldsymbol{G}(1), \ldots, \boldsymbol{G}(P-1)$, therefore, the non-zero entries of $\tilde{\boldsymbol{G}}$ would not be in the same position with respect to mod p, since otherwise $J_{s1}(\tilde{\boldsymbol{W}})$ would be simply 0.

As a result, the maximum is achieved if and only if $\tilde{\boldsymbol{G}}$ corresponds to perfect separator transfer matrix in the form $\mathbb{G}(z) = \text{diag}(\alpha_1 z^{-d_1}, \alpha_2 z^{-d_2}, \ldots, \alpha_p z^{-d_p})\boldsymbol{P}$ where $\mathbb{G}(z)$ is the Z-transform of the overall system function $\{\boldsymbol{G}(l);\, l \in \{0, \ldots, P-1\}\}$, α_k's are non-zero real scalings, and d_k's are non-negative integer delays.

Here, we point out that the blind source extraction problem is a special case of the blind source separation problem. Therefore, this proof can simply be also applied to the blind source extraction method. However, we treat the blind source extraction problem as a separate case to provide alternative geometric intuition.

7.5.4 Analysis of the Family of Objective Functions ($J_{s2,r}$)

Before analysing this family of objective functions for some special r values, similar to the proof of Theorem 3, we can rewrite (7.8) in terms of $\boldsymbol{Q}$ and obtain

$$J_{s2,r}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\boldsymbol{Q}\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{s}}}_A}\boldsymbol{Q}^T)}\right)^{1/K}}{\left\| \begin{bmatrix} ||\boldsymbol{Q}_{1,:}||_1 & ||\boldsymbol{Q}_{2,:}||_1 & \ldots & ||\boldsymbol{Q}_{p,:}||_1 \end{bmatrix}^T \right\|_r^p}.$$

Following similar steps, by modifying (7.22), we can obtain the corresponding inequality

$$J_{s2,r}(\tilde{\boldsymbol{W}}) \leq \frac{\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_2}{\left\| \left[||\boldsymbol{Q}_{1,:}||_1 \quad ||\boldsymbol{Q}_{2,:}||_1 \quad \ldots \quad ||\boldsymbol{Q}_{p,:}||_1 \right]^T \right\|_r^p} \left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2 \right)^{1/K} \det(\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{S}}}_A})^{1/2K}.$$

The results of analysing this family of objective functions, for some special r values:

- $r = 1$ *Case:* In this case, we have

$$\left\| \left[\begin{array}{cccc} \|\boldsymbol{Q}_{1,:}\|_1 & \|\boldsymbol{Q}_{2,:}\|_1 & \ldots & \|\boldsymbol{Q}_{p,:}\|_1 \end{array} \right]^T \right\|_1^p = \left(\sum_{m=1}^{p} \|\boldsymbol{Q}_{m,:}\|_1 \right)^p \geq p^p \prod_{m=1}^{p} \|\boldsymbol{Q}_{m,:}\|_1,$$

where the inequality comes from Arithmetic-Geometric-Mean-Inequality, and the equality is achieved if and only if all the rows $\boldsymbol{Q}$ have the same 1-norm. Hence, we have

$$\begin{aligned} J_{s2,1}(\tilde{\boldsymbol{W}}) &\leq \frac{\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_2}{p^p \prod_{m=1}^{p} \|\boldsymbol{Q}_{m,:}\|_1} \left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2 \right)^{\frac{1}{K}} \det(\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{S}}}_A})^{\frac{1}{2K}} \\ &\leq \frac{1}{p^p} \left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2 \right)^{\frac{1}{K}} \det(\hat{\boldsymbol{R}}_{\underline{\tilde{\boldsymbol{S}}}_A})^{\frac{1}{2K}}. \end{aligned}$$

As a result, $\boldsymbol{Q}$ is a global maximum of $J_{s2,1}(\tilde{\boldsymbol{W}})$ if and only if it is a perfect separator matrix of the form

$$\mathbf{Q} = k\mathbf{P}\text{diag}(\rho),$$

where k is a non-zero value, $\rho \in \{-1, 1\}^p$ and $\mathbf{P}$ is a permutation matrix. This implies $\tilde{\boldsymbol{G}}$ is a global maximum of $J_{s2,1}(\tilde{\boldsymbol{W}})$ if and only if the corresponding form is satisfied

$$\Gamma_K(\tilde{\boldsymbol{G}}) = k\mathbf{P}\breve{\Lambda}^{-1}\text{diag}(\rho).$$

Therefore, the global maxima of the objective function $J_{s2,1}$ corresponds to a subset of perfect separators.

- $r = 2$ *Case:* In this case, using the basic norm inequality and Arithmetic-Geometric-Mean-Inequality, for any $\mathbf{x} \in \mathbb{R}^p$, we have

$$(||\mathbf{x}||_2)^p \geq \left(\frac{1}{\sqrt{p}}||\mathbf{x}||_1\right)^p \geq p^{p/2}\prod_{m=1}^{p}|x_m|$$

  where the equality is achieved if and only if all the components of $\mathbf{x}$ are equal in magnitude. As a result, this yields

$$\begin{aligned} J_{s2,2}(\tilde{\boldsymbol{W}}) &\leq \frac{\prod_{m=1}^{p}||\boldsymbol{Q}_{m,:}||_2}{p^{p/2}\prod_{m=1}^{p}||\boldsymbol{Q}_{m,:}||_1}\left(\prod_{n=1}^{(P-1)p}||\boldsymbol{M}_{n,:}||_2\right)^{\frac{1}{K}}\det(\hat{\boldsymbol{R}}_{\underline{\tilde{\mathbf{S}}}_A})^{\frac{1}{2K}} \\ &\leq \frac{1}{p^{p/2}}\left(\prod_{n=1}^{(P-1)p}||\boldsymbol{M}_{n,:}||_2\right)^{\frac{1}{K}}\det(\hat{\boldsymbol{R}}_{\underline{\tilde{\mathbf{S}}}_A})^{\frac{1}{2K}}. \end{aligned}$$

  Similarly, $J_{s2,2}$ has the same set of global maxima as $J_{s2,1}$.

- $r = \infty$ *Case:* Following similar steps, using the basic norm inequality and Arithmetic-Geometric-Mean-Inequality, for any $\mathbf{x} \in \mathbb{R}^p$, we have

$$(||\mathbf{x}||_\infty)^p \geq \left(\frac{1}{p}||\mathbf{x}||_1\right)^p \geq \prod_{m=1}^{p}|x_m|,$$

where the equality is achieved if and only if all the components of $\mathbf{x}$ are equal in magnitude. Based on this inequality, we obtain

$$J_{s2,\infty}(\tilde{\boldsymbol{W}}) \leq \frac{\prod_{m=1}^{p} ||\boldsymbol{Q}_{m,:}||_2}{\prod_{m=1}^{p} \|\boldsymbol{Q}_{m,:}\|_1} \left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2 \right)^{\frac{1}{K}} \det(\hat{\boldsymbol{R}}_{\underline{\tilde{\mathbf{s}}}_A})^{\frac{1}{K}}$$
$$\leq \left(\prod_{n=1}^{(P-1)p} ||\boldsymbol{M}_{n,:}||_2 \right)^{\frac{1}{2K}} \det(\hat{\boldsymbol{R}}_{\underline{\tilde{\mathbf{s}}}_A})^{\frac{1}{2K}}.$$

Therefore, $J_{s2,\infty}$ also has same set of global optima as $J_{s2,1}$ and $J_{s2,2}$.

7.5.5 Proof of Theorem 4

We begin with observing that

$$\mathbb{R}\{\boldsymbol{o}(k)\} = \sum_{l=0}^{P-1} \mathbb{R}\{\boldsymbol{G}^T(l)\}\mathbb{R}\{\boldsymbol{s}(k+M-1-l)\} - \mathbb{I}\{\boldsymbol{G}^T(l)\}\mathbb{I}\{\boldsymbol{s}(k+M-1-l)\},$$
$$\mathbb{I}\{\boldsymbol{o}(k)\} = \sum_{l=0}^{P-1} \mathbb{I}\{\boldsymbol{G}^T(l)\}\mathbb{R}\{\boldsymbol{s}(k+M-1-l)\} + \mathbb{R}\{\boldsymbol{G}^T(l)\}\mathbb{I}\{\boldsymbol{s}(k+M-1-l)\}.$$

Defining $\grave{\boldsymbol{G}} = \begin{bmatrix} \mathbb{R}\{\boldsymbol{G}_0\} & -\mathbb{I}\{\boldsymbol{G}_0\} & \ldots & \mathbb{R}\{\boldsymbol{G}_{P-1}\} & -\mathbb{I}\{\boldsymbol{G}_{P-1}\} \\ \mathbb{I}\{\boldsymbol{G}_0\} & \mathbb{R}\{\boldsymbol{G}_0\} & \ldots & \mathbb{I}\{\boldsymbol{G}_{P-1}\} & \mathbb{R}\{\boldsymbol{G}_{P-1}\} \end{bmatrix}$ and $\grave{\mathbf{s}}_{K+P-1}(k) = [\ \mathbb{R}\{\boldsymbol{s}^T(k)\} \quad \mathbb{I}\{\boldsymbol{s}^T(k)\} \quad \ldots \quad \mathbb{R}\{\boldsymbol{s}^T(k-K-P+2)\} \quad \mathbb{I}\{\boldsymbol{s}^T(k-K-P+2)\}\ ]^T$ yields $\grave{\mathbf{o}}_K(k) = \Gamma_{2K}(\grave{\boldsymbol{G}})\grave{\mathbf{s}}_{K+P-1}(k)$. Thus,

$$\hat{\boldsymbol{R}}_{\grave{\mathbf{o}}_K} = \Gamma_{2K}(\grave{\boldsymbol{G}})\grave{\Lambda}\hat{\boldsymbol{R}}_{\underline{\grave{\mathbf{s}}}_{K+P-1}}\grave{\Lambda}^T\Gamma_{2K}(\grave{\boldsymbol{G}})^T,$$

where $\grave{\Lambda} = I \otimes \Lambda$ is the range matrix of $\grave{\mathbf{s}}_{K+P-1}$ and $\hat{\boldsymbol{R}}_{\underline{\grave{\mathbf{s}}}_{K+P-1}}$ is defined as the sample covariance matrix of $\underline{\grave{\mathbf{s}}}_{K+P-1}$.

Defining $\grave{\boldsymbol{Q}} = \Gamma_{2K}(\grave{\boldsymbol{G}})\grave{\Lambda}$ and following similar steps, we can write $\prod_{m=1}^{2p} \hat{\mathcal{R}}(\grave{o}_m) =$

$\prod_{m=1}^{2p} ||\grave{\boldsymbol{Q}}_{m,:}||_1$. Rewriting (7.12) in terms of $\grave{\boldsymbol{Q}}$ yields

$$J_{cs1}(\tilde{\boldsymbol{W}}) = \frac{\left(\sqrt{\det(\grave{\boldsymbol{Q}}\hat{\boldsymbol{R}}_{\underline{\grave{\boldsymbol{S}}}_A}\grave{\boldsymbol{Q}}^T)}\right)^{1/K}}{\prod_{m=1}^{2p} ||\grave{\boldsymbol{Q}}_{m,:}||_1}.$$

Note that we have the similar expression as (7.20). Hence, the proof of Theorem 3 also applies here. Note that the structure of $\Gamma_{2K}(\grave{\boldsymbol{G}})$ implies that the non-zero entries of $\tilde{\boldsymbol{G}}$ can only be real or purely imaginary. Therefore, the set of global maxima for the objective function (7.12) corresponds to a subset of complex perfect separators.

Chapter 8

Conclusion and Future Work

This dissertation has presented the convergence analysis of recently introduced instantaneous BCA algorithms. Moreover, the instantaneous BCA approach has been extended by providing a general optimization framework which can be used to produce numerous instantaneous BCA algorithms. Additionally, a convolutive BCA framework has been introduced which can produce a family of convolutive BCA algorithms that are able to separate stationary independent and/or dependent sources. We point out that this is the first convolutive BCA method in the literature. Besides, a deterministic BCA analysis framework has been proposed which does not assume any stationarity of sources. With this approach, it is possible to separate convolutive mixture of non-stationary as well as stationary independent and/or dependent sources.

Further research might explore the convergence behaviour of the instantaneous BCA algorithms. Another possible area of future research would be to investigate a BCA method that can incorporate the pdf information of sources to provide a better separation performance for tailed distributions.

Bibliography

[1] A. Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *Signal Processing, IEEE Transactions on*, vol. 61, pp. 5730–5743, Nov 2013. ix, 3, 5, 7, 16, 19, 38, 39, 40, 42, 44, 72, 73, 74, 77, 78, 79, 87, 90, 94

[2] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications.* Academic Press, 2010. 1, 66, 84

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis.* John Wiley and Sons Inc., 2001. 1, 45

[4] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, pp. 287–314, April 1994. 1

[5] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *Signal Processing, IEEE Transactions on*, vol. 45, pp. 434 –444, Feb 1997. 2

[6] M. Matsuoka, M. Ohaya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, pp. 411–419, March 1995. 2

[7] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, pp. 992–996, July 2005. 2

[8] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1136–1155, May 1996. 2

[9] P. Jallon, A. Chevreuil, and P. Loubaton, "Separation of digital communication mixtures with the CMA: Case of unknown symbol rates," *Signal Processing*, vol. 90, pp. 2633–2647, September 2010. 2

[10] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. I. algorithms," *EEE Transactions on Signal Processing*, vol. 44, pp. 1184–1197, May 1996. 2

[11] D.-T. Pham, "Blind separation of instantenaous mixtrures of sources based on order statistics," *IEEE Trans. on Signal Processing*, vol. 48, pp. 363–375, February 2000. 2, 3

[12] S. Cruces and I. Duran, "The minimum support criterion for blind source extraction: a limiting case of the strenthened young's inequality," *Lecture Notes in Computer Science*, pp. 57–64, September 2004. 2

[13] F. Vrins, M. Verleysen, and C. Jutten, "SWM: A class of convex contrasts for source separation," *Proceedings of IEEE ICASSP 2005*, vol. V, pp. 161–164, March 2005. 2

[14] F. Vrins, J. A. Lee, and M. Verleysen, "A minimum-range approach to blind extraction of bounded sources," *IEEE Trans. on Neural Networks*, vol. 18, pp. 809–822, May 2007. 2

[15] F. Vrins and D. Pham, "Minimum range approach to blind partial simultaneous separation of bounded sources: Contrast and discriminacy properties," *Neurocomputing*, vol. 70, pp. 1207–1214, March 2007. 2

[16] S. Vembu, S. Verdu, R. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1952–1960, August 1994. 2

[17] A. T. Erdogan and C. Kizilkale, "Fast and low complexity blind equalization via subgradient projections," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2513–2524, July 2005. 2

[18] A. T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. on Signal Processing*, vol. 54, pp. 438–449, February 2006. 2

[19] A. T. Erdogan, "Globally convergent deflationary instantaneous blind source separation algorithm for digital communication signals," *IEEE Trans. on Signal Processing*, vol. 55, pp. 2182–2192, May 2007. 2

[20] A. T. Erdogan, "Adaptive algorithm for the blind separation of sources with finite support," *Proceedings of EUSIPCO 2008, Lausanne*, August 2008. 2

[21] S. Cruces, "Bounded component analysis of linear mixtures: A criterion for minimum convex perimeter," *IEEE Trans. on Signal Process.*, vol. 58, pp. 2141–2154, April 2010. 2, 3

[22] P. Aguilera, S. Cruces, I. Duran, A. Sarmiento, and D. Mandic, "Blind separation of dependent sources with a bounded component analysis deflationary algorithm," *IEEE Signal Processing Letters*, vol. 20, pp. 709–712, July 2013. 3

[23] A. T. Erdogan, "On the convergence of symmetrically orthogonalized bounded component analysis algorithms for uncorrelated source separation," *IEEE Transactions on Signal Processing*, pp. 6058–6063, November 2012. 3

[24] A. T. Erdogan, "A family of bounded component analysis algorithms," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1881–1884, March 2012. 48, 49, 58, 68

[25] Y. Inouye and R.-W. Liu, "A system-theoretic foundation for blind equalization of an FIR MIMO channel system," *IEEE Trans. on CAS-I: Fund. Th. and Apps.*, vol. 49, pp. 425–436, April 2002. 12

[26] D. Garling, *Inequalities: a journey into linear analysis.* Cambridge University Press, 2007. 41, 51, 95

[27] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999. 45

[28] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *Radar and Signal Processing, IEEE Proceedings F*, vol. 140, no. 6, pp. 362–370, 1993. 45

[29] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, pp. 157–192, Jan. 1999. 45

[30] T. Mei, J. Xi, F. Yin, A. Mertins, and J. Chicharo, "Blind source separation based on time-domain optimization of a frequency-domain independence criterion," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, pp. 2075–2085, Nov 2006. 49

[31] T. Mei, A. Mertins, F. Yin, J. Xi, and J. F. Chicharo, "Blind source separation for convolutive mixtures based on the joint diagonalization of power spectral density matrices," *Signal Processing*, vol. 88, no. 8, pp. 1990 – 2007, 2008. 49

[32] W. Rudin, *Real and Complex Analysis.* McGraw-Hill Book Co., 1987. 51

[33] S. Demarta and A. J. McNeil, "The t copula and related copulas," *International Statistical Review*, vol. 73, pp. 111–129, Jan. 2005. 65

[34] C. Simon, P. Loubaton, and C. Jutten, “Separation of a class of convolutive mixtures: a contrast function approach,” *Signal Processing*, vol. 81, no. 4, pp. 883 – 887, 2001. 66, 67, 68, 69, 84, 85, 86, 87, 88

[35] M. Castella and E. Moreau, “New kurtosis optimization schemes for MISO equalization,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 3, pp. 1319–1330, 2012. 66, 67, 68, 69, 84, 85, 86, 87, 88

[36] P. Jallon, M. Castella, and R. Dubroca, “Matlab toolbox for separation of convolutive mixtures @ONLINE,” 2011. 66, 84

[37] B. Grunbaum, V. Klee, A. Perles, and G. Shephard, *Convex Polytopes*. John Wiley & Sons, Incorporated, 1967. 91