Adaptive Independent Component Analysis: Theoretical Formulations and Application to CDMA Communication System with Electronics Implementation

By

Zaid Albataineh

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Electrical Engineering – Doctor of Philosophy

2014

ABSTRACT

Adaptive Algorithms for Independent Component Analysis: Formulations and Application to CDMA Communication systems with Electronic Implementation

By

Zaid Albataineh

Blind Source Separation (BSS) is a vital unsupervised stochastic area that seeks to estimate the underlying source signals from their mixtures with minimal assumptions about the source signals and/or the mixing environment. BSS has been an active area of research and in recent years has been applied to numerous domains including biomedical engineering, image processing, wireless communications, speech enhancement, remote sensing, etc. Most recently, Independent Component Analysis (ICA) has become a vital analytical approach in BSS. In spite of active research in BSS, however, many foundational issues still remain in regards to convergence speed, performance quality and robustness in realistic or adverse environments. Furthermore, some of the developed BSS methods are computationally expensive, sensitive to additive and background noise, and not suitable for a real-time or real world implementation.

In this thesis, we first formulate new effective ICA-based measures and their corresponding robust adaptive algorithms for the BSS in dynamic "convolutive mixture" environments. We demonstrate their superior performance to present competing algorithms. Then we tailor their application within wireless (CDMA) communication systems and Acoustic Separation Systems. We finally explore a system realization of one of the developed algorithms among ASIC or FPGA platforms in terms of real time speed, effectiveness, cost, and economics of scale.

Firstly, we propose a new class of divergence measures for Independent Component Analysis (ICA) for estimating sources from mixtures. The Convex Cauchy-Schwarz Divergence (CCS-DIV) is formed by integrating convex functions into the Cauchy-Schwarz inequality. The new measure is symmetric and convex with respect to the joint probability, where the degree of convexity can be tuned by a (convexity) parameter. A non-parametric (ICA) algorithm generated from the proposed divergence is developed exploiting convexity parameters and employing the Parzen window-based distribution estimates. The new contrast function results in effective parametric and nonparametric ICA-based computational algorithms. Moreover, two pairwise iterative schemes are proposed to tackle the high dimensionality of sources.

Secondly, a new blind detection algorithm, based on fourth order cumulant matrices, is presented and applied to the multi-user symbol estimation problem in Direct Sequence Code Division Multiple Access (DS-CDMA) systems. In addition, we propose three new blind receiver schemes, which are based on the state space structures. These so-called blind state-space receivers (BSSR) do not require knowledge of the propagation parameters or spreading code sequences of the users but relies on the statistical independence assumption among the source signals.

Lastly, system realization of one of the developed algorithms has been explored among ASIC or FPGA platforms in terms of cost, effectiveness, and economics of scale. Based on our findings of current stat-of-the-art electronics, programmable FPGA designs are deemed to be the most effective technology to be used for ICA hardware implementation at this time. I would like to dedicate this dissertation to my loving Father, Mohammad Taisser, who sadly passed away in 2001. I wish he had a chance to be here today. A great and special feeling of gratitude to my loving Mother, Bahjh, her constant love and caring are the reasons of who and what I am. My gratitude and my love are beyond words to her and to all my sisters and brothers; namely Jedah, Ziad, Jawdat, Jehan, Fareed and Jomana. Lastly, I would like to dedicate this dissertation to my country Jordan.

ACKNOWLEDGMENTS

First of all, I wish to express my great gratitude to my advisor, Dr. Fathi M. Salem, for his guidance and support throughout all stages of this research. His advice and encouragement have provided me with the skills needed to develop and refine this research. Special Thanks for Dr. Hayder Radha, Dr. Yang Wang and Dr. Shantanu Chakrabartty for agreeing to serve on my committee.

I would like to acknowledge and thank Michigan State University (MSU) for the educational opportunity that allowed me to conduct my study and research. I also wish to thank all my friends and colleagues in the ECE departments at MSU. A huge thanks to all my friends who have assisted and supported me in so many ways during my study, especially Reemon Haddad, Yazan Smadi, Taghleb, Saad, Zuhair, Khalil, Abu Rawhi and Mustafa who have been friends and colleagues. I would also like to thank Yarmouk University for supporting me financially during my study at MSU.

Lastly, I would like to express my great gratitude to my mother, brothers and sisters for all their support and encouragement.

PREFACE

In *Chapter 1*, we provide the background needed to understand Blind Source Separation Problems. We provide the definitions and theoretical background necessary to understand Blind Techniques. We also address our motivations, problem definitions and some concepts that are essential to grasp the topics covered by the different chapters of this dissertation; namely, Blind Techniques in Wireless Communication, State Space Approach and Adaptive Framework for Blind Source Separation.

In *Chapter 2*, we perform a thorough review of the BSS\ICA algorithms, and then we give an overview of the ICA algorithms and emphasize the approaches that influenced our work. We also study some of the methods that have been developed to solve the ICA problems in the case of instantaneous and convolutive mixtures.

In *Chapter 3*, we propose a new Convex Cauchy–Schwarz Divergence (CCS-DIV) measure for Blind Source Separation (BSS) and unsupervised learning of acoustic and speech signals. The CCS-DIV measure is developed by integrating convex functions into the Cauchy–Schwarz inequality. By including a convexity quality parameter, the measure has a broad control range of its convexity. With this measure, a new CCS–ICA algorithm is structured and a non-parametric form is developed incorporating the Parzen window-based distribution. Moreover, the CCS–ICA algorithm has a controlled speed towards timed convergence. Also, A two pairwise iterative schemes are proposed to tackle the high dimensional problem in the blind source separation BSS.

In *Chapter 4*, we present a frequency-domain method based on robust independent component analysis (RICA) to address the multichannel blind source separation (BSS)

vi

problem of the convolutive speech mixtures in highly reverberant environments. We apply an algorithm to separate the source signals in adverse conditions, i.e.: high reverberation conditions when the short observation signals are available. Furthermore, we study the impact of several parameters on the performance of separation, e.g. overlapping ratio and window type in frequency-domain method. We also compare different techniques to solve the permutation ambiguity.

In *Chapter 5*, a new blind detection algorithm, based on fourth order cumulant matrices, is presented and applied to the multi-user symbol estimation problem in Direct Sequence Code Division Multiple Access (DS-CDMA) systems. The blind detection is to estimate multiple symbol sequences in the downlink of a DS-CDMA communication system using only the received wireless data and without any knowledge of the user spreading codes. The proposed algorithm takes advantage of higher cumulant matrix properties to reduce the computational load and enhance performance.

In *Chapter 6*, we develop three update laws in order to enhance the performance of the blind detector based on the state space structures. Bit error rate (BER) simulations of these methods are shown for different number of users, signal to noise ratio (SNR) and different number of symbols per user in comparison with the Blind Multiuser Detectors (BMUD), Linear Minimum mean squared error (LMMSE) and other conventional detectors.

In *Chapter 7*, we introduce a constrained blind multiuser detection in order to improve its performance with imposing the regularization parameter to cope the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation.

vii

In *Chapter 8*, we investigate the ICA algorithms in terms of hardware implementation. Although software implementation is important to investigate the capabilities of ICA algorithms and to simulate significant aspects of applications, Hardware implementation provides real time solutions and an optimal parallelism method in terms of fast convergence. Furthermore, software implementation may suffer from insufficient memory problems because the large data sets of the ICAs' applications and its high dimensionality. Thus, hardware implementations are a promising approach to implement the ICA algorithms and they are executed by Integrated Circuit (ICs). Owning the high speed processing and the parallel architecture features make the hardware implementation outperforming the software implementation in terms of sufficient memory and fast convergence. Finally in *Chapter 9* we conclude the dissertation and highlight directions for future work.

TABLE OF CONTENTS

LIST OF TAB	LES	ciii
LIST OF FIGU	JRES	xiv
LIST OF ALG	ORITHMSxv	viii
1 Chapter 1		1
Introduction		1
1.1	Blind Techniques in Wireless Communication	5
1.1.1	Direct-Sequence Code Division Multiple Access	7
1.1.2	Why Blind	9
1.1.3	CMDA SIGNAL MODELS	10
	1.1.3.1 DS-CMDA Receiver Signal Model	10
1.0	1.1.3.2 WCDMA Receiver Signal Model	13
1.2	The State Space Framework	14
1.3	Adaptive Framework for Blind Source Separation	15
1.3.1	Invertible Mixtures	16
	1.3.1.1 Instantaneous mixtures	16
1.0.0	1.3.1.2 Convolutive Mixtures	18
1.3.2	Underdetermined mixtures	20
1.4	Dissertation Contributions	21
2 Chapter 2		26
Literature Revie	W	26
2.1	Introduction	26
2.2	Principle Component Analysis (PCA)	30
2.3	Independent Component Analysis (ICA)	32
2.3.1	The Instantaneous ICA Framework	33
	2.3.1.1 Preprocessing	34
	2.3.1.2 Nonlinear function choice (Activation function):	35
	2.3.1.3 The Learning Update Rules	36
	2.3.1.3.1 Batch Learning (Offline learning):	36
	2.3.1.3.2 Stochastic gradient (Online learning):	36
	2.3.1.4 ICA based on Maximization of non-Gaussianity	36
	2.3.1.4.1 Kurtosis Measure	37
	2.3.1.4.2 Gradient algorithm using kurtosis	38
	2.3.1.5 Fixed point algorithm	39
	2.3.1.6 Negentropy Measure	40
	2.3.1.7 ICA based on Maximum Likelihood Estimation	43
	2.3.1.8 ICA based on Entropy Maximization	46
	2.3.1.9 Bell-Sejnowski method	46
	2.3.1.10 ICA based on Tensorial Methods	48

		2.3.1.11 PARAllel FACtor (PARAFAC) algorithms	49
		2.3.1.12 Joint Approximation Diagonalisation (JAD)	50
		2.3.2 The Convolutive ICA Mixtures	50
		2.3.2.1 Time-Domain Methods	54
		2.3.2.2 TRINICON Blind Source Separation	55
		2.3.2.3 Frequency-Domain Methods:	57
		2.3.2.3.1 Lee's approach	57
		2.3.2.3.2 Smardagdis approach	59
		2.3.2.3.3 Independent Vector Analysis (IVA)	60
		2.3.2.3.4 Parra's approach:	
		2 3 2 3 5 Recursive Convolutive ICA	62
	24	Ambiguities in ICA algorithms	62
	<u> </u>	2.4.1 Scale ambiguity	63
		2.4.2 Permutation ambiguity	65
		2.4.3 Circularity of Fast Fourier Transform (FFT)	66
	2.5	Performance Metrics of ICA methods	60
	2.0	2.5.1 Instantaneous case	67
		2.5.1 Performance Matrix G	67
		2.5.1.1 Ferrormance Wattik C	67
		2.5.1.2 Sivie measure	67
		2.5.2 Convolutive cuse	68
		2.5.2.1 Performation measure:	68
		2.5.2.2 Nutual information measure	68
3 Conv	Chaj ex C	pter 3 Cauchy–Schwarz Independent Component Analysis for Blind Source Sep	70 aration
•••••			70
	3.1	Introduction	
	3.2	A Brief Description of Previous Divergence Measures	74
		3.2.1 Previous Divergence Measures	75
		3.2.2 The proposed Divergence Measure	78
		3.2.3 Link to other Divergences:	80
		3.2.4 Geometrical Interpretation of the Proposed Divergence for	$\alpha = 1$
		and $\alpha = -1$.	81
		3.2.5 Evaluation of Divergence Measures	82
	3.3	Convex Cauchy–Schwarz Divergence Independent Component A	nalysis
	(CC	S-ICA)	87
	3.4	Scenario of two or three source signals	93
	3.5	Computational Complexity	96
	3.6	Simulation Results	96
		3.6.1 Sensitivity of CCS-DIV measure	97
		3.6.2 The performance and the convergence speed of the pro-	oposed
		CCS-ICA algorithms versus the existing ICA-based algorithms	100
		3.6.3 Experiments on Speech and Music Signals	109
	3.7	Conclusion	114

A Ro	bust	ICA-Based Algorithm for Blind Separation of Convolutive Mixtures	116
	4.1	Introduction	116
	4.2	Convolutive Mixtures	120
		4.2.1 Problem Definition	120
	4.1	Recursively Regularized ICA	123
	4.2	The presented method Based on RobustICA framework	125
		4.4.1 Step1: Preprocessing (Data Whitening)	125
		4.4.2 Step 2: Determining the rotation matrix (unitary matrix) U (w)	127
	4.5	Scaling and Permutation Ambiguities	131
		4.5.1 Estimation the diagonal matrix Dw	132
		4.5.2 Estimation the permutation matrix $\Gamma(\mathbf{w})$	133
	4.6	Experiments Results	138
		4.6.1 Section 1	139
		4.6.2 Section 2	144
	4.7	Conclusion	148
5	Chap	pter 5	149
Robu	ıst Bl	ind Multiuser Detection Algorithm Using Fourth Order Cumulant Matrices	149
	5.1	Introduction	149
	5.2	DS-CMDA Signal Model	152
	5.3	Robust and Fast Independent Component Analysis (ICA)	154
		5.3.1 The Preprocessing	155
		5.3.2 The FastICA Algorithm	156
		5.3.3 The Robust ICA algorithm	158
	5.4	The Proposed Detection Algorithm Based On Cumulant Matrices	159
		5.4.1 Step1: Preprocessing (Data Whitening)	159
		5.4.2 Step 2: Determining the rotation matrix (unitary matrix) U	161
	5.5	Simulation Results	163
		5.5.1 Performance	163
		5.5.2 Measure of Computation	168
	5.6	Conclusion	172
6	Chap	pter 6	173
Adap	otive	Blind Multiuser Detection DS-CDMA Based on State Space Approach	173
	6.1	Introduction	174
	6.2	DS-CMDA Signal Model	179
	6.3	Conventional Blind Linear Multiuser Detectors	182
		6.3.1 Single user detection (SUD) Detector	182
		6.3.2 Rake Detector	182
		6.3.3 LMMSE Detector	183
	6.4	The Proposed Detection Schemes Based On state space framework	184
		6.4.1 Step1: Preprocessing (Data Whitening)	185
		6.4.2 Step 2a: Determining the rotation matrix (unitary matrix) U based	l on
		the feedforward structure.	186
		6.4.3 Step 2b: Determining the rotation matrix (unitary matrix) U based	l on
		the feedback structure I.	188

		6.4.4 Step 2c: Determining the rotation matrix (unitary matrix) U	J based on
		the feedback structure II.	
		6.4.5 The proposed adaptive detectors	193
	6.5	Simulation Results	
	6.6	Conclusion	205
7	Cha	nter 7	206
, Cor	nstrair	ed Blind Multiuser Detection for DS-CDMA System	206
001	7 1	Introduction	206
	7.2	DS-CMDA Signal Model	208
	73	Conventional Blind Multiuser Detection	209
	74	The Proposed Detection Scheme	212
	75	Simulation Results	215
	7.6	Conclusion	
8	Cha	pter 8	
Har	dware	Implementation	220
	8.1	Introduction	221
	8.2	Comparative Study of Existing Solutions to implement ICA	Algorithms
			221
		8.1.1 Analog CMOS Integration Circuit	
		8.1.2 Mixed Signal Techniques (Analog and Digital Circuit)	223
		8.1.3 ASIC Solutions	
		8.1.4 FPGA Solutions	
	8.3	Multiplier Design	
	8.4	Simulation Results	
	8.5	Conclusion	233
0	Cha	nter 9	236
	Clia	pici 9	
COI		Conclusion	
	9.1	Euture Worlz	
	9.2		
AP	PENE	DIX	
BIF	BLIO	GRAPHY	

LIST OF TABLES

Table 3.1: The performance of the ICA algorithm based on the proposed divergence.. 226

Table 3.2: The computational load, in seconds, of the ICA algorithm based on the proposed divergence and other widely used ICA algorithms, each entry averages over the corresponding number of trials. Observation mixtures consists of two source signals that follow the same
Table 3.3: The corresponding variance of the performance. 104
Table 3.4: Kurtosis Values of the different probability density functions that used in the ICA experiments 107
Table 3.5: The performance of the ICA algorithm based on the proposed divergence in terms of Amari error (multiplied by 100). Each entry averages over the corresponding number of trials.107
Table 3.6: The computational load, in seconds, of the ICA algorithm based on the proposed divergence and other widely used ICA algorithms, each entry averages over the corresponding number of trials. 108
Table 3.7: The performance of the ICA algorithm based on the proposed divergence and other widely used ICA algorithms in terms of Amari error (multiplied by 100). Each entry averages over the corresponding number of trials
Table 8.1: Comparison of Analog, Mixed signal and ASIC Solutions
Table 8.2: Comparison of FPGA Solutions 227
Table 8.3: Comparison Results Among Various ICA Implementations

LIST OF FIGURES

Figure 1.1: Illustration of the BSS Problem [2]
Figure 1.2: Block diagram of blind source separation
Figure 1.3: Wireless Communication Scenario [2]
Figure 1.4: Conceptual State-Space model which illustrates the general linear
Figure 2.1: Classifications for BSS problem
Figure 2.2: Block diagram of the Convolutive Mixtures
Figure 2.3: Lee's Block diagram
Figure 2.4: Smardagdis's Block diagram
Figure 2.5: Illustration of permutation ambiguity in frequency domain
Figure 3.1: Illustration of Geometrical Interpretation of the proposed Divergence
Figure 3.2: Different divergence measures versus the joint probability Px1, x2(A, A) 85
Figure 3.3: CCS-DIV and α -DIV versus the joint probability Px1, x2(A, A)
Figure 3.4: CCS-DIV with various alphas versus the joint probability Px1, x2(A, A) 86
Figure 3.5: The surfaces and Contours of CCS-DIV vs CS-DIV
Figure 3.6: Comparison of (a) CCS-DIV with $\alpha = 1$, (b) CCS-DIV with $\alpha = -1$, (c) KL-DIV, (d) E-DIV, (e) CS-DIV (f) C-DIV with $\alpha = -1$ of demixed signals as a function of the demixing parameters $\theta 1$ and $\theta 2$
Figure 3.7: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in parametric BSS task
Figure 3.8: Comparison of learning curves of C-ICA, E-ICA, KL-ICA, and 106
Figure 3.9: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in parametric BSS task random initial value 111
Figure 3.10: Comparison of learning curves of C-ICA, E-ICA, KL-ICA, and CCS-ICA with α =1, and α =-1 in a two-source BSS task with random initial value

Figure 3.11: Comparison of SIRs (dB) of demixed two speeches and	12
Figure 3.12: Comparison of learning curves of C-ICA, E-ICA,	12
Figure 3.13: Comparison of SIRs (dB) of demixed two speeches and	13
Figure 3.14: The original signals and de-mixed signals by using	13
Figure 3.15: Comparison of SIRs (dB) of demixed two speeches and music signals using different ICA algorithms in instantaneous BSS task with additive Gaussi noise.	by an 14
Figure 4.1: configuration of the two experimental setups that were conducted by 13	38
Figure 4.2: Results obtained in Test1 experiments. The SIR performance of	40
Figure 4.3: Results obtained in Test1 experiments. The SDR performance	41
Figure 4.4: Corresponding CPU time for each method	42
Figure 4.5: Figure 4.5: Results obtained in Test2 experiments. The SIR performance 14	43
Figure 4.6: Results obtained in Test2 experiments. The SIR performance of the 14	43
Figure 4.7: Results obtained in the Test1 experiments [130]. Best performance	45
Figure 4.8: Results obtained in the Test1 experiments [130].	46
Figure 4.9: Results obtained in the Test2 experiments [130].	46
Figure 4.10: Results obtained in the Test2 experiments [130] 14	47
Figure 4.11: Impact of FFT length, 2-by-2 case, Results obtained in the	47
Figure 5.1: Average BER as a function of SNR for 30 users and 100 runs	65
 Figure 5.2: Average BER as a function of SNR for different sample lengths T with a users and 100 runs. Black triangle right lines: M = 104. Black circle lines: T = 900 Black hexagram lines: T = 8000. Black square lines: T = 7000. Red triangle up line M = 6000. Blue circle lines: T=5000. Blue hexagram lines: T=4000. Blue squa lines: T=3000. Blue triangle right lines: T=2000. 	30)0. es: ure 66
Figure 5.3: Average BER as a function of SNR for different Users K with signal bloc composed of T = 3500 samples and 100 runs. Black triangle right lines: K = 1 Black circle lines: K = 20. Red triangle up lines: K = 30. Blue square lines: K=4 Blue hexagram lines: K=50.	ks 0. 10. 67

Figure 5.4: Average extraction quality as a function of computational cost for different mixture sizes K with signal blocks composed of T = 3500 samples and 1000 mixture realizations. Solid lines: K = 5. Dashed lines: K = 10. Dotted lines: K = 20. 169

Figure 5.5: Average extraction quality as a function of computational	cost for different
mixture sizes K with signal blocks composed of $T = 3500$ samples	and 1000 mixture
realizations. Solid lines: K = 30. Dashed lines: K = 40	

Figure 5.6: Average extraction quality as a function of computational cost for different mixture sizes K with different signal blocks T samples and 1000 mixture realizations. Solid lines: K = 20. Dashed lines: K = 30. Dotted lines: K = 40 171
Figure 6.1: Signal generation model for a typical QPSK DS-CDMA system 179
Figure 6.2: Feedback Demixing Structure I
Figure 6.3: Feedback Demixing Structure II
Figure 6.4: Average BER as a function of SNR for DS-CDMA downlink. Using Gold codes G=63. (a) Using 30 users (b) Using 50 users
Figure 6.5: Corresponding CPU time for each method
Figure 6.6: Average BER as a function of SNR for DS-CDMA downlink. Using OVSF codes G=64. (a) Using 30 users (b) Using 50 users
Figure 6.7: Average BER as a function of SNR for WS-CDMA downlink 199
Figure 6.8: Average BER as a function of SNR for WS-CDMA downlink. Using OVSF codes G=64. (a) Using 30 users (b) Using 50 users
Figure 6.9: Average BER as a function of SNR for DS-CDMA downlink 201
Figure 6.10: Average BER as a function of SNR for WCDMA downlink 202
Figure 6.11: Average BER as a function of SNR for various number of users K 203
Figure 6.12: Average BER as a function of SNR for various sample sets M 204
Figure 7.1: Average BER as a function of SNR for 15 users
Figure 7.2: Average BER as a function of SNR for 15 users with L=2N, L=3N 216
Figure 7.3: Average BER as a function of SNR for 15 users For various
Figure 7.4: Average BER as a function of SNR for 15 users with L=1000 218
Figure 8.1: The proposed Mixer (Multiplier) schematic

Figure 8.2: Voltage Conversion Gain versus IF	232
Figure 8.3: Voltage Conversion Gain versus IF	233
Figure 8.4: Voltage Conversion Gain versus IF	234
Figure 8.5: Voltage Conversion Gain versus IF	234
Figure 8.6: The proposed Mixer (Multiplier) Layout.	235

LIST OF ALGORITHMS

Algorithm 3.1: ICA Based on the gradient descent	
Algorithm 3.2: ICA Based on pairwise gradient decent scheme	
Algorithm 3.3: ICA Based on pairwise Jacobi scheme	
Algorithm 6.1: RAKE based FastICA method	192
Algorithm 6.2: RAKE based RICA method	

1 Chapter 1

Introduction

Blind techniques have been used in studies since the 1980's, when the first adaptive equalizers were designed for digital communication [1]. The problem with using this technique was to estimate an unknown linear single input signal output (SISO) stationary channel, without any knowledge about the input signal. The word "blind" implies that all signal processing techniques recover both the unknown mixing systems and unknown sources based only on the observations [1], [2].

The Blind Source Separation (BSS) problem was found in the framework of neural modeling around 1982 by Bernard Ans, Jeanny Herault and Christian Jutten [1]. It then gained considerable attention in more diverse research areas after Comon published his pioneering paper in a signal processing journal on Independent Component Analysis (ICA) in 1994 [7]. In 1995, Bell and Senjnowski were boosting the ICA topic by developing the infomax algorithm [18]. Meanwhile, the well-known JADE algorithm was proposed by Cardoso and Soul, in 1993 [16]. Although various BSS algorithms have been developed with numerous contrast functions for optimization over the last decade, BSS is still considered one of the most important research topics in signal processing. It has generated a lot of interest in the last decade [1], [3].



Figure 1.1: Illustration of the BSS Problem [2]

BSS is considered to be an unsupervised stochastic method which separates the underlying source signals from their mixtures, without any knowledge about the source signals or the mixing process. Recently, Independent Component Analysis (ICA) has become a vital algorithm in BSS, Figure 1.1. ICA has been a very important topic in many research areas [1-3], [12-27], i.e.: biomedical engineering, image processing, wireless communication systems, speech enhancement, remote sensing, etc. ICA is related to Principle Component Analysis (PCA) and Factor Analysis (FA) in multivariate analysis. It specifically corresponds to second order methods in which the components or

factors are a Gaussian distribution. However, the ICA is a statistical technique that includes higher order statistics (HOS) where; the goal is to represent a set of random variables as a linear transformation of statistically independent components.

The ICA aims to recover both the unknown source signals and mixing system or one of them from only observed system outputs. This important research topic includes several concepts of signal processing, information theory, statistics and probability, neural networks, etc. The ICA has several applications in many fields such as image processing, wireless communications, biomedical applications and audio source separation. In addition, the ICA problem appears in many multi-sensor systems [1]. ICA methods are essentially based on parameter estimation, which requires a model of separating system, objective criteria and optimization methods.





The BSS can be expressed as in Figure 1.2, where $S = \{s_1, s_2, ..., s_n\}$ represents the n source signals, and $X = \{x_1, x_2, ..., x_m\}$ represents the m observed signals, A or H represents the mixing systems which correspond to whatever is static or dynamic. Two mixture models have been considered. Firstly, *linear static mixtu*res, which assume that the mixing system A is memoryless, are called linear instantaneous mixtures. Secondly, the *linear convolutive mixtures*, which assume that the mixing matrix H is a memory system, vary with time. In order to solve the BSS problems, we need assumptions to apply; otherwise this problem is considered to be ill-posted. The common assumption in BSS\ICA field is the mutual statistical independence among the original sources. Although sometimes this assumption is problematic to establish, it is realistic and fully justified in several applications. Other suitable assumptions can be used successfully for solving BSS problems depending on the applications.

In this dissertation, we investigate and study the two models of mixing systems: instantaneous mixtures model in (1.1) and convolutive mixtures model in (1.2), respectively.

$$x(t) = As(t) + v(t)$$
 $t = 1, 2, ..., T$ (1.1)

Where

- $A = [a_{ij}]_{mxn}$ is the memoryless mixing system.
- $s(t) = [s_{jt}]_{nxT}$ is the *n* original source signals to the system where the T is the sampled length.
- $x(t) = [x_{it}]_{mxT}$ is the *m* observed signals of the system
- $v(t) = [n_{it}]_{nxT}$ is the noise and usually assume to be an additive white Gaussian noise (AWGN)

$$x(t) = \sum_{d}^{L} H_{d}s(t-d) + v(t) \quad \forall \quad t = 1, 2, ..., T$$
 (1.2)

Where

- $H_d(t) = [h_{ijd}(t)]_{mxn}$ is the mixing matrix at *dth* delay which represents d = 1, 2, ..., L
- h_{iit}(t) is the impulse response of time instant t.

We study these two models intuitively in next chapter. Also, In terms of applications, we address the problem of BSS in a cocktail party problem [3], [38] and [120]. In additional, we are going to address its application in wireless communication systems specifically using blind multiuser detections in CDMA system [89], [75], [59]. Furthermore, a system realization of one of the developed algorithms will be explored among ASIC or FPGA platforms factoring in cost, effectiveness, and economics of scale.

1.1 Blind Techniques in Wireless Communication

In telecommunication systems [74], [75], [89], [91], the most essential challenge has been to set up the system based on simultaneous multiuser access in order to get higher efficient wireless systems. However, several state-of-the-art approaches have been proposed in literature [74], [85], [94] to overcome this challenge such as the trained-based systems. These techniques periodically enforce the user to send a known training sequence for the receiver in order to make the receiver able to estimate the parameter of the propagation channel. They are caused by the multiple reflections of the radio waves on the obstacles encountered, e.g. buildings, cars, and trees etc. Furthermore, according to [60], [74], it has been reported that 20% of the bandwidth is devoted to the training sequence in a GSM system and up to 40% in a UMTS system. In spite of the good performance of the aforementioned technique, the cost tends to be significantly large in terms of bandwidth. The efficiency of most communication systems is based on the

bandwidth and the transmitted power. However, the blind multi user techniques are a promising area in wireless communication systems because of its potential to ensure the high communication rate and spectral efficiently, thereby reducing or disposing of the training sets.



Figure 1.3: Wireless Communication Scenario [2]

Blind techniques are considered to be an attractive field of work because the following reasons: 1) reduce the training sequence; 2) help the trained-based systems back up in fast-time varying channels and at severe multipath fading, respectively. Also, Blind techniques help to recover the signals in other situations such as eavesdropping, where using the training sequence is not possible [59], [74]. For these reasons, we are motivated to do additional research in this area in order to design a new multi user detection that performs well in a multipath propagation environment. And, It has to be more robust to the outlets "versus type of noise" in terms of performance.

1.1.1 Direct-Sequence Code Division Multiple Access

Code division multiple access (CDMA) is a multiplexing technique or a channel access method that allows several users to access to the same multi-point transmission medium "RF channel" asynchronously and simultaneously, to transmit over it and to share its capacity. CDMA is used by various radio communication technologies such as CDMAONE "the mobile phone standard", CDMA2000 "the third generation of CDMAONE, WCDMA "the third generation standard used by GSM carriers", and Evolved High-Speed Packet Access (HSPA+) [74]. Although, LTE (4G) is in operation in many cellular companies inside and outside the U.S., the networks are still not fully built out, and LTE coverage is still not universal. Thus, the most of the older 2G and 3G systems are still in charge or at least working in parallel with the 4G as in U.S. companies, AT&T and T-Mobile use GSM/WCDMA/HSPA while Verizon, Sprint, and MetroPCS use cdma2000/EV-DO. Moreover, The LTE wireless interface is incompatible with 2G and 3G networks, so that it must be operated on a separate wireless spectrum. Nevertheless, 3G is intended to be replaced by 4G technologies sooner or later, but it is going to take a long time before the LTE coverage is developed to be fully operated and universal, especially in some countries worldwide, such as India, Pakistan, Iraq ... etc. [141, [142].

One of the most interesting concepts in data communication is the idea of allowing multiusers to share and send information simultaneously over a single communication channel. This means, several users share a band of frequencies called "Bandwidth". Despite the fact that the CDMA is suitable for satisfying the demand for higher data rates and an inherent capability of CDMA is resisting interference and providing a secure channel, Multiuser detection in CDMA systems usually suffers from the multi-access interference (MAI) and inter-symbol interference (ISI) due to the nonideal cross-correlations between users' spreading sequences. The concept of spread spectrum communication was proposed for military purposes and used in anti-jamming techniques. Later, the spread spectrum techniques were employed to civilian purposes.

In highly loaded systems [74], [88], conventional detectors are considered an unsuitable choice, since most of them suffer from the external interference sources such as adjacent channel interference or jamming, and treat the interference as an additional background noise. These drawbacks have motivated development of numerous interference rejection techniques [89], [91].

In CDMA literature [74], [76], most of the conventional Multiuser detection methods assume the low statistical correlation between desired users and interfering ones, which motivated them to use the SOS properties of the received data. Thus, one relatively new idea is to extend the work on higher order statistics (HOS) in order to make the methods robust and secure against incomplete cross-correlation and a near-far problem, which are considered to be additional drawback factors in conventional detection methods. BSS based on HOS [1], [2] [8-21] are able to recover the signals from the mixtures without any knowledge about the waveform structure "modulation" and mixing coefficients. Lastly, the adaptive LMMSE detector has been proposed to overcome the complex matrix inversion operation, but this detector still needs the spreading codes of all users. Therefore, the Adaptive MMSE detector might not be realistic in the downlink receiver. Therefore, one of the main emphasizes in this dissertation is to develop and implement the BSS/ICA algorithm to assist Multiuser detection methods in order to

mitigate different types of interference sources in CDMA system, especially in DS/W CDMA downlink systems.

1.1.2 Why Blind

Blind Multiuser detectors are promising because they only require the received signal with neither prior knowledge of any training signals nor the user spreading codes in order to equalize the channels and estimate the multiple symbol sequences associated with all users in CDMA systems. On the other side, several state-of-the-art approaches have been proposed in literature to overcome this challenge such as the trained-based systems [74], [85- 86]. These techniques periodically enforce the user to send a known training sequence for the receiver in order to make the receiver able to estimate the parameter of the propagation channel. In spite of the good performance of the aforementioned technique, the cost tends to be significantly large in terms of bandwidth. The efficiency of most communication systems is based on the bandwidth and the transmitted power. Therefore, the blind techniques usually perform well and more robust in terms of the estimation the symbol under the ill-condition environment i.e. under severe multipath fading channels. One can also incorporate prior information such as spatial knowledge or a set of short training sequences if available, in order to construct the semi-blind detectors [75], [78], [89], [91], [96].

The reasons to apply the blind multiuser detections are as follows [2]:

- 1) Training examples for interference are often not available [74], [78].
- 2) In rapid time varying channels, training may not be efficient [79].

- Capacity of the system can be increased by eliminating or reducing training sets [60], [74].
- Multi-path fading during the training period may lead to poor source or channel estimations. [87-89]
- 5) Training in distributed systems requires synchronization and/or sending a training set each time a new link is to be set up. This may not be feasible in a multi-user scenario [90 -93].

1.1.3 CMDA SIGNAL MODELS

In this section, we briefly present the signal model for a CDMA implementation using one layer of spreading codes only. Next, we briefly describe the DS-CDMA signal and WCDMA signal models in a typical synchronous CDMA system employed for indoor ATM and certain ad hoc wireless networks [75], [81], and [89].

1.1.3.1 DS-CMDA Receiver Signal Model

In a DS-CDMA system, several users share the medium simultaneously by using their own signatures. The simplest received signal model r(t) before filtering in a symbol interval is given by

$$r(t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{l=0}^{L-1} \alpha_{lm} b_{k,m} s_k (t - mT_b - d_l T_c) + n(t)$$
(1.3)

where

- ✓ l, k, and m represent the path, user and symbol indices, respectively.
- \checkmark α_{lm} is the path gain since in downlink model; the path gain does not differ among users because all users'

signals are sent together and the path gain α_{lm} and propagation delay factor d_l depend only on the number of paths.

✓ $b_{k,m}$ is symbol.

- ✓ $s_k(.)$ is spreading code (chip sequence)
- ✓ d_l is the propagation delay factor, d_l ∈ {0, 1, ..., $\frac{C-1}{2}$ }
- ✓ C is the number of chips per symbol.
- ✓ T_b , T_c , t are time , symbol and chip duration, respectively.
- ✓ n(t) is an additive white Gaussian noise (AWGN) channel.

In this dissertation, the system is assumed to be time-invariant which means that the channel parameters are much slower than the frequency of the transmitted symbol data. However, let us assume that G is the number of code sequence, K is the number of users, and L is the number of channels; thus, the vector form of the equation (1) will become:

$$\mathbf{r} = \mathbf{H}\mathbf{S}\mathbf{b} + \mathbf{n} \tag{1.4}$$

Where **r** is the received vector signal, **H** is an $(G + L - 1) \times G$ matrix which represents the multipath propagation coefficients, **S** is an $G \times K$ block diagonal matrix, **b** is an K - d vector which represent the data symbols and **n** is the (G + L - 1) - d channel noise vector with covariance matrix **Q**. This model received signals (2) is suitable for deriving the linear symbols detectors such as the MF, the RAKE, the LMMSE and the blind Detectors based on FastICA and Robust ICA algorithms [89], [88] – [75].

An alternative signal model is proposed in [89], [81] as a linear convolutive model given by:

$$r_n = H_0 b_n + H_1 b_{n-1} + n_n = H b_n + n_n$$
(1.5)

where

- \checkmark r_n is the received signal vector;
- ✓ $\mathbf{b}_{\mathbf{n}} = [\mathbf{b}_1(\mathbf{n}), ..., \mathbf{b}_K(\mathbf{n})]^T$ is a current bits of all users;
- ✓ H₀ = [h₁,..., h_k] is signature matrix of the current bits of all users including MAI;

$$h_{k} = \begin{bmatrix} 0 \\ h_{k}(0) \\ \vdots \\ \vdots \\ h_{k}(N - d_{k} - 1) \end{bmatrix}$$
(1.6)

✓ $H_1 = [\overline{h_1}, ..., \overline{h_k}]$ is the signature matrix of the previous bits of all users including ISI; where

$$\overline{h}_{k} = \begin{bmatrix} h_{k}(N - d_{k}) \\ \vdots \\ \vdots \\ h_{k}(N + M - 1) \\ 0 \end{bmatrix}$$
(1.7)

- ✓ $H = [H_0; H_1]$ is the signature matrix of all users;
- ✓ $\boldsymbol{b}_n = [b_1(n), ..., b_K(n)]^T$ are currents bits of all user;

✓ $\boldsymbol{b}_{n-1} = [b_1(n-1), ..., b_K(n-1)]^T$ are previous bits of all

users;

$$\checkmark \ \overline{\boldsymbol{b}}_{\boldsymbol{n}} = [\boldsymbol{b}_{\boldsymbol{n}}^T; \ \boldsymbol{b}_{\boldsymbol{n-1}}^T]^T$$
 are bits of all users;

✓ $n_n = [n(nN), ..., n(nN + N - 1)]^T$ is independent white Gaussian Noise vector.

In uplink (asynchronous) CDMA systems, one can assume that H_0 and H_1 are mutually independent. Therefore, H is a full column matrix and its rank is 2K as shown in [88], [89]. In downlink (synchronous) CDMA communication [75], [81], H is a matrix, has full-rank and less restrictions as seen in [89]. In additional, our proposed algorithms are working well in asynchronous CDMA system [83], [81], the main focus in this dissertation is the synchronous CDMA communication system.

1.1.3.2 WCDMA Receiver Signal Model

In a WCDMA system, the presence of scrambling codes makes the system different from the DS-CDMA system. The main reason behind the MAI in the WCDMA system is the intra-cell multiple user signals sharing the same multipath channels. However, the simplest received signal model r(t) before filtering in a symbol interval is given by:

$$\mathbf{r}(t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{l=0}^{L} \alpha_{lm} b_{k,m} c_k (t - d_l T_c) s_k (t - m T_b - d_l T_c) + n(t)$$
(1.8)

Where $c_k(t) \in \{\pm 1 \pm j\}$ are the complex cell-specific scrambling sequences, the rest of the variables are defined in a similar manner to the model (1.8). The received signal at UE/MS is passed through a chip-matched filter and sampled at chip rate. The received vector r in this case can be expressed as

$$\boldsymbol{r} = \boldsymbol{H}\boldsymbol{C}\boldsymbol{S}\boldsymbol{b} + \boldsymbol{n} \tag{1.9}$$

Where **C** is the G x G complex diagonal scrambling matrix with $\mathbf{CC}^{H} = \mathbf{I}_{GxG}$ and the rest of the variables are defined as similar in (1.9). The form of **C** is given by:

$$\boldsymbol{C} = \boldsymbol{diag}[\boldsymbol{c_1} \ \boldsymbol{c_2} \dots \ \boldsymbol{c_3}] \tag{1.10}$$

Where

$$c_i \in \{\pm 1 \pm j\} \quad \forall \ 1 \le i \le G$$

1.2 The State Space Framework

In this dissertation, we investigate the state space framework in order to design a blind multiuser detection based on the state space approach. The Linear State Space approach is an extension of the static "instantaneous" ICA model which easily extends further to a flexible nonlinear model as in [2], [95].

Despite the fact that it is one of the most powerful tools in terms of Blind Source Separation, there are several reasons that have made us interested in this specific model [2]. [95], [135]:

- Flexible and Universal Linear Model: the mixing and filtering processes in the state space approach may have different mathematical and/or physical models, i.e. multichannel deconvolution models (MCD), Finite Impulse Response (FIR), Moving Average (MA), Autoregressive (AR) and Autoregressive Moving Average (ARMA), etc [2].
- Many canonical realizations: it provides us with many canonical realizations of the same dynamic system using the equivalent transformations.
- The Linear State Space approach is an extension of the static "instantaneous" ICA model and it is easy to extend it further to a flexible nonlinear model.
- 4. The state space models provide us two subsystems with different update approaches, 1) a linear and memory less output layer 2) a non-linear or linear recurrent networks.

5. Recoverability: the inverse of the state space representation depends on the invert-ability of the mixing matrix between the input and the output.



Figure 1.4: Conceptual State-Space model which illustrates the general linear state-space mixing and self-adaptive demixing model for Dynamic ICA [2].

1.3 Adaptive Framework for Blind Source Separation

Generally, the BSS consists of recovering the unobserved sources, denoted in vector notation as $s(t) = (s_1(t), s_2(t), ..., s_n(t))^T \in \mathbb{R}^n$, with assuming zero mean and stationary with observed mixtures, $x(t) = (x_1(t), x_2(t), ..., x_m(t))^T \in \mathbb{R}^m$, which can be written:

$$\mathbf{x}(\mathbf{t}) = \Phi(\mathbf{s}(\mathbf{t})) \tag{1.11}$$

Where Φ is an unknown mapping from \mathbb{R}^n in \mathbb{R}^m , and t represents the sample index, which can stand for instance of time; so, we can divide the BSS problems based on the type of Φ mapping into two groups, one called the invertible mixtures and second called Underdetermined mixtures "non-invertible" [1].

1.3.1 Invertible Mixtures

If the mapping Φ is invertible, which means satisfying the condition $m \ge n$, where *m* and *n* are the numbers of sensors and the number of sources, respectively. In this case, identification of Φ or of its inverse, directly leads to source separation, i.e. provides estimated sources y(t) such that:

$$y(t) = Wx(t) = W * H * s(t) = P * D * s(t)$$
 (1.12)

Where

P – is a generalized permutation matrix.

D – is a scaling matrix which is a diagonal matrix.

This equation shows typical indeterminacies of BSS problem. According to the nature mixing, BSS problem can be more or less complicated. For example, simple instantaneous mixtures when Φ is restricted to a simple mixture A, sources are estimated up to a permutation and a scale. Invertible mixtures can be divided into following two categories [1]:

1.3.1.1 Instantaneous mixtures

One modest approximation is to assume that the mixing system A in Figure 1.2 is instantaneous, assuming that A is a mixing matrix, n(t) is an additive noise, s(t) is the

source signals and x(t) is the observed signals, then the instantaneous model can be expressed as follows:

$$x(t) = As(t) + n(t)$$
 $t = 1, 2, ..., T$ (1.13)

$$\begin{bmatrix} x_{1}(t) \\ x_{2}(t) \\ \vdots \\ \vdots \\ x_{M}(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{M1} & a_{M2} & \dots & \vdots \\ a_{M1} & a_{M2} & \dots & a_{MN} \end{bmatrix} \begin{bmatrix} s_{1}(t) \\ s_{2}(t) \\ \vdots \\ \vdots \\ s_{N}(t) \end{bmatrix} + \begin{bmatrix} n_{1}(t) \\ n_{2}(t) \\ \vdots \\ \vdots \\ n_{M}(t) \end{bmatrix}$$
(1.14)

For notational and mathematical convenience, we will assume that we have the number of observed signals equal to the number of source signals, i.e.: m = n. We also assume that the mixing matrix A is a full rank matrix. In addition, we assume that the noise in the mixtures is an additive Wight Gaussian noise.

The BSS problem aims to retrieve the original sources from the given observations. In the instantaneous model, we only need to estimate un-mixing matrix W that equals the inverse of the mixing matrix $W = A^{-1}$ to recover the original signals s(t) almost directly. Estimate the unmixing matrix $W \approx A^{-1}$ from a given set of observations x(t) that can retrieve the original signals via a linear transform:

$$y(t) = Wx(t) \approx s(t) \quad t = 1, 2, ..., T$$
 (1.15)

In order to measure the performance of the separation algorithm, we use the performance matrix G,

$$G = WA \approx I \tag{1.16}$$

where I is an identity matrix. Ideally, the performance matrix G would be closed to an identity matrix for an efficient separation algorithm. Despite the fact that the separated sources y(t) may not be the same order and scale as the original sources, the matrix G should be an identity up to scale and a permutation.

1.3.1.2 Convolutive Mixtures

A convolutive mixture can be considered as a natural extension of the instantaneous BSS problem. Assume an m-dimensional vector of received discrete time signals $x(k) = [x_1(k), x_2(k), ..., x_m(k)]^T$ at time k is assumed to be produced from an n-dimensional vector of source signals $s(k) = [s_1(k), s_2(k), ..., s_m(k)]^T$, where $m \ge n$, by using a stable mixture model [2]:

$$x(k) = \sum_{p=-\infty}^{\infty} H_p s(k-p) = H_p * s(k),$$

with $\sum_{-\infty}^{\infty} ||H_p|| \le \infty$ (1.17)

Where * represents the linear convolution operator and H_p is an (m x n) matrix of mixing coefficients at time-lag p.

Assume that elements h_{jip} denote the coefficients of the Finite Impulse Response (FIR) filter H_p , and L is the maximum unknown channel length. Then, the noise-free convolutive model is written as follows:

$$\mathbf{x}(\mathbf{k}) = \sum_{p=0}^{L-1} \mathbf{H}_{p} \mathbf{s}(\mathbf{k} - \mathbf{p})$$
(1.18)

Thus, one can find an approximate inverse channel matrix W_p in order to recover the source signals $s(k) = [s_1(k), s_2(k), ..., s_m(k)]^T$ such that

$$y(k) = W_p * x(k) = \sum_{p=-0}^{Q-1} W_p x(k-p) = \hat{s}(k)$$
(1.19)

where Q is the length of the inverse of the channel impulse response. However, there are two approaches to solve this problem and recover the source signals. In time
domain approaches, they have several general drawbacks such as Q should be selected at least equal to the unknown true channel L. Therefore, for long mixing filter, which means long transfer functions, the computation will be expensive [2], [133]. Also, using the IIR filter instead of long FIR filter to overcome this problem really suffers from the instability and need to invert the non-minimum phase filters [2], [38], [133]. Moreover, timing approaches are sensitive to channel order mismatch – see [2], [133] for a recent survey. However, time domain methods are suitable and very efficient for small mixing filters such as in communication channel [133]. For all these limitations, we focus our study on frequency approaches to solve the cocktail party problem [3], [38]. Also, the main advantage of a frequency domain BSS approach is the ability to apply the set of any instantaneous ICA algorithms to solve the convolutive BSS problem. On other hand, the main challenge of BSS in the frequency domain is to deal with the permutation and scaling ambiguities - see [1], [3], [38] for a recent survey.

However, one can re-map the aforementioned BSS models into frequency domain by applying the Discrete Fourier Transform (DFT) on the observed signals x(k) in order to transform it to the instantaneous mixtures problem as follows

$$\mathbf{x}(\mathbf{k}) = \mathbf{H} * \mathbf{s}(\mathbf{t}) \ \mathbf{x}(\mathbf{q}, \mathbf{w}) \approx \mathbf{H}(\mathbf{w})\mathbf{s}(\mathbf{q}, \mathbf{w}) \tag{1.20}$$

where w is a frequency index, q is a frame index, $s(q, w) = [s_1(q, w), ..., s_m(q, w)]^T$ and $x(q, w) = [x_1(q, w), ..., x_n(q, w)]^T$.

In the previous equation, it is considered to be valid only for periodic signals s(t) [3]. However, it is approximately valid if the time-convolution is circular. Therefore, to ensure that the time convolution is circular, it requires making the Fourier Transform length significantly larger than the maximum length of the mixing channels L [3]. In [38],

they imposed the spectral smoothing approach in order to mitigate the circularity effect in frequency domain BSS methods. We will study these effects intuitively in next chapter.

1.3.2 Underdetermined mixtures

If the number of observations (sensors) m is less than the number of sources n, the mixing process is referred to be underdetermined (not invertible) [1], [52], [53].

The separation processes can be attained successfully in the frequency-domain up to scaling and permutation ambiguities under the assumption that the mixing matrix H(f)is full column rank at each frequency bin. However, when the number of source signals is more than the number of sensors, the assumption on the mixing matrix H(f) becomes not valid. So, in this case the problem is more difficult since the mixing matrix H(f)becomes ill-conditioned matrix which means the mixing matrix H(f) is not left pseudoinvertible. However, a lot of work has been done in order to perform a good separation process in the case of the instantaouse mixture [1]. However, there are not so many works that has been done on the underdetermined case in the convolutive mixture [1], [38]. In the literature, the well-known algorithm of such method is the DUET algorithm which is proposed by Rickard et al [2], [3], [38] and [117]. The DUET algorithm assumes a specific delayed model that only works for audio signals with small delay, e.g. hearing aid etc. The DUET algorithm performs the separation processes using the two sensors in order to compute two parameters amplitude differences and phase differences between the source signals. Several papers were published to develop and enhance the performance of the DUET algorithm in [3], but their performance in real reverberant environment is still limited. One of the promising approaches in this field is to convert some of the underdetermine cases of the instantounous mixture into the frequency domain in order to tackle the underdetermined problem in the convoluvtive mixture as presented in the literature [1], [3], [38].

1.4 Dissertation Contributions

The contributions of this thesis are summarized as follows:

- We perform a thorough review of the BSS\ICA algorithms, and then we give an overview of the ICA algorithms and emphasize the approaches that influenced our work. We also study some of the methods that have been developed to solve the ICA problems in the case of instantaneous and convolutive mixtures.
- Independent Component Analysis (ICA) is a crucial tool in Blind Source Separation (BSS). In this thesis, we present a new Convex Cauchy–Schwarz Divergence (CCS-DIV) measure for Blind Source Separation (BSS) and unsupervised learning of acoustic and speech signals. This CCS-DIV measure is developed by integrating convex functions into the Cauchy–Schwarz inequality. By including a convexity quality parameter, the measure has a broad control advantage of its convexity. With this measure, a new CCS–ICA algorithm is structured and a non-parametric form is developed incorporating the Parzen window-based distribution. Furthermore, the CCS–ICA algorithm has a controlled speed towards timed convergence. Several case-study scenarios were carried out on instantaneous and noisy mixtures of speech signals. Finally, the superiority of the proposed CCS–ICA algorithm is demonstrated in metric performance comparison with FAST ICA, Robust

ICA, convex ICA (C-ICA), and other existing algorithms based on mutual information and Jenson's inequality.

- Two pairwise iterative schemes are proposed to tackle the high dimensionality problem. Two pairwise schemes non-parametric independent component analysis ICA algorithms based on a new high-performance Convex Cauchy–Schwarz Divergence (CCS-DIV). These two schemes enable fast and efficient de-mixing of sources in real-world applications where the dimensionality of the sources is high. Finally, the performance superiority of the proposed schemes is demonstrated in metric-comparison with FastICA, RobustICA, convex ICA (C-ICA), and other leading existing algorithms.
- We propose a frequency-domain method based on robust independent component analysis (RICA) to address the multichannel Blind Source Separation (BSS) problem of the convolutive speech mixtures in highly reverberant environments. We impose regularization processes to tackle the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation in frequency domain methods. We apply an algorithm to separate the source signals in adverse conditions, i.e. high reverberation conditions when short observation signals are available. Furthermore, we study the impact of several parameters on the performance of separation, e.g. overlapping ratio and window type in the frequency domain method. We also compare different techniques to solve the permutation ambiguity. Through simulations and real-world experiments, we verify the superiority of the presented algorithm among other BSS algorithms, i.e.

recursive regularized ICA (RR-ICA), independent vector analysis (IVA) and others.

Code Division Multiple Access (CDMA) is a channel access method used by various radio technologies and it is based on spread-spectrum technology. In general, CDMA is used as an access method in many mobile standards such as CDMA2000, and WCDMA. We address the problem of blind multiuser equalization in the wideband CDMA system, in the noisy multipath propagation environment. Herein, we propose three new blind receiver schemes, which are based on the state space structures. This so-called blind state-space receivers (BSSR) does not require knowledge of the propagation parameters or spreading code sequences of the users but relies on the statistical independence assumption between the source signals. Also, we develop and derive three update-laws in order to enhance the performance of the blind detector. Additionally, we upgrade three semi-blind adaptive detectors based on the corporation between the RAKE receiver and the stochastic gradient algorithms which are used in several blind adaptive signal processing algorithms, namely FastICA, RobustICA, and principle component analysis PCA. Bit error rate (BER) simulations of these methods are shown for different number of users, signal to noise ratio (SNR) and different number of symbols per user in comparison with the Blind Multiuser Detectors (BMUD), Linear Minimum mean squared error (LMMSE) and other conventional detectors. The results show that the proposed algorithm outperforms the other detectors in estimating the symbol signals from the

mixed CDMA received signals. Moreover, the new blind detectors mitigate the multi access interference (MAI) in CDMA.

- A new blind detection algorithm, based on fourth order cumulant matrices, is . presented and applied to the multi-user symbol estimation problem in Direct Sequence Code Division Multiple Access (DS-CDMA) systems. The blind detection is to estimate multiple symbol sequences in the downlink of a DS-CDMA communication system using only the received wireless data and without any knowledge of the user spreading codes. The proposed algorithm takes advantage of higher cumulant matrix properties to reduce the computational load and enhance performance. Bit error rate (BER) simulations of this algorithm are shown for different number of users, signal to noise ratios (SNR) and different number of symbols per user in comparison with the FAST ICA and Robust ICA algorithms. The results show that the proposed algorithm outperforms both ICA-based detectors in estimating the symbol signals from the received mixed signals. Moreover, the proposed blind detector is computationally fast and exhibits high convergence speed in extracting user symbols.
- In direct sequence code division multiple access DS-CDMA communication system, the blind multiuser detection is presented for enhance the computational complexity and mitigate the multiple access interference (MAI) in the detector. The ill-condition of the covariance matrix of the received signals degrades the performance of the linear minimum mean-squared error LMMSE detector. Especially, when the Signal to noise ratio is high and small

data set is available for covariance matrix estimation. In this thesis, we introduce a constrained blind multiuser detection in order to improve its performance with imposing the regularization parameter to cope the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation. Through simulation results, we show that the proposed method improves the performance of the blind multiuser detection and outperforms the conventional multiuser detections.

• Lastly, we investigate the ICA algorithms in terms of hardware implementation. Although software implementation is important to investigate the capabilities of ICA algorithms and to simulate significant aspects of applications, Hardware implementation provides real time solutions and an optimal parallelism method in terms of fast convergence.

2 Chapter 2

Literature Review

In this chapter, we perform a thorough review of the BSS\ICA algorithms, and then we will give an overview of the ICA algorithms and will emphasize the approaches that influenced our work. We will study some of the methods that have been developed to solve the ICA problems in the case of instantaneous and convolutive mixtures, respectively. Finally, for a more thorough review on ICA problems, applications and methods, we recommend referencing one of these valuable books: "Handbook of Blind Source Separation Independent Component Analysis and Applications" [1], "Independent Component Analysis in frequency domain" [3] and "Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications" [2].

2.1 Introduction

Independent component analysis (ICA) considers a vital algorithm in Blind source separation (BSS). ICA algorithms based on the information theatric approach are attractive and have been considered a hot topic in signal processing for the last two decades due to its potential in areas such as biomedical, wireless communication system, audio separation and identification, etc. The goal of ICA is to recover the original source signals from the mixtures without any further knowledge about the mixing coefficients and the original sources. However, ICA is a statistical technique that includes higher order statistics (HOS), where the goal is to represent a set of random variables as a linear transformation of statistically independent components [1-3].

The idea of BSS was first introduced by J. Herault, C. Jutten and B. Ans) who worked in neurophysiologies in the early 1982s [1]. They proposed a blind method to separate the natural impulses coming from different parts of human body. [5], [40],[41]. Meanwhile, telecommunications related applications of ICA have been proposed even earlier in MIMO systems [62],[101], [117], [118]. Also, ICA has been investigated and implemented in several applications such as audio and biomedical signal processing and feature extraction. An extensive review of the history of ICA and its applications is given in [53], [56], [57]. BSS methods have several interesting applications. In finance, they use the BSS algorithms to find the independent factors in data [74]. In Image processing, they use the BSS algorithm to help estimate the best independent basis for compression or denoising [14-16]. They are also used in biomedical signals, like in an EEG [11] or an ECG, for analysis purposes [1-2]. In audio, they use the BSS algorithms to identify the sounds or separate the audio signals as in the cocktail party problem [1-3], [51]. However, one of the most interesting applications of the BSS is in wireless communication systems.



Figure 2.1: Classifications for BSS problem.

There are more than four communities that have worked on the BSS, especially Independent component analysis (ICA), refer to [1] and [2]. However, one of the best examples used to illustrate the BSS problem is the cocktail party problem. While attending a cocktail party and using ones' ears to listen and separate a specific sound source among all other sounds present in the room, for example: people talking, music, etc., in order to emulate this human behavior, researchers carry out the BSS idea from the way our brain tackles this problem. Much research has been done over the last two decades in BSS and ICA areas.

Recently, researchers proposed several BSS methods based on the frequency-time analysis [3], [50] or using regularization algorithms [20], [21]. They conducted the matrix factorization based on frequency-time analysis in order to get inproved performance and speed up the convergence as in [4], [36], [47] and [49]. They also investigated the BSS singles channel problem extensively in [14], [19] and [42]. Additionally, they employed the independent vector analysis for joint BSS over multiple datasets [2] and studied the sparse analysis in order to estimate the demixing matrix W due to a quadratic programing technique and to deal with nonnegative BSS problems. However, sparse component analysis was proposed in [50]. Despite all of these methods, they are not considered a case of non-stationary conditions, the uncertainty of parameters in general ICA model and the effects of noise signals.

On other hand, one can express the mixing system in various models based on the non-stationary mixing coefficients and source signals in real-world recording. This complicated circumstance may explain two scenarios. One can assume that the sources are moving or the sources have disappeared. However, several works have been done on both scenarios individually. Usually, for the scenario where the sources are moving, the source distributions and the number of sources is assumed to be fixed. Thus, an adaptive BSS algorithm has been proposed to compensate for the variations of a mixing matrix. In [118], they applied the Markov process to hold the variation of the sources signals. Also, the status of source signals was detected using the 3-D tracker, where if the sources were

moving, they added a beam-forming algorithm to BSS methods in [119]. Several researchers have characterized the time-varying of the source distributions by automatic relevance determination (ARD) techniques [120], [121], [122]. The switching ICA algorithm [123] was proposed in order to detect the absence sources or the ones present. S-ICA algorithm used a hidden Markov model in order to represent the status of the source signals and it assumed that the generative model was fixed. In [123], they studied the replacement of source signals. An online vibrational Bayesian (VB) learning was used in [120]; they proposed a new online ICA algorithm based on VB learning to separate the dependent source signals. In [125], they proposed the ICA algorithm based on the piecewise non-Gaussian stationary model in order to separate the non-Gaussian signals which have the varying distribution, refer to Figure 2.1 to see the root of BSS problems.

2.2 Principle Component Analysis (PCA)

Principle Component Analysis (PCA) is one of the most well-known algorithms in multivariate analysis and data mining. It was established by Pearson [2]. He proposed a general framework of the PCA algorithm in biological context. Recently, there have been many efficient and powerful adaptive algorithms PCA which have been proposed and developed for PCA [1].

Generally speaking, the PCA aims to derive a smaller set of variables with less redundancy while retaining as much of the information from the original variables as possible. In other words, PCA is a mathematical tool that uses orthogonal transformation in order to convert a set of observations, which might be correlated variables, into a set of values of linearity uncorrelated variables referred to as the principle component. The most important objectives of PCA are dimensionality reduction, determination of linear combinations of variables, feature selection (choosing the most useful variables), visualization of multidimensional data, identification of underlying variables, and identification of groups of outliers.

Assume we have a random vector x with T elements and there are m observations of this vector. In order to transform a set of observations x into a set of values of linearity uncorrelated variables u, we can apply the PCA to the observed data x as follows:

The first step will be to remove the dc component of observed data x as follows:

$$\mathbf{x} = \mathbf{x} - \mathbf{E}[\mathbf{x}] \tag{2.1}$$

The operator E[.] is the expectation value of x. In this dissertation, we use the expectation operator for the theoretical analysis. For the practical simulation of the algorithms, we will find that the expectation depends on the type of learning algorithms. For batch learning algorithms (offline) we will use the sample mean, whereas, for the stochastic learning algorithms we will drop the expectation and use the actual expression inside the expectation. Then, one can convert the PCA to the eigenvalue problem of the covariance matrix of x, which is essentially equivalent to the well-known transformation (Karhunen-Loeve transform) which is used in signal processing, as follows:

$$R_{xx} = E[x(t)x^{T}(t)] = V\Lambda V^{T} \in \mathbb{R}^{mxm} \quad \forall t = 1, 2, ..., T$$
(2.2)

Where x^{T} is x transpose, $\Lambda = \text{diag}\{\lambda_{1}, \lambda_{2}, ..., \lambda_{m}\}$ is a diagonal matrix that contains m eigenvalues and $V = [V_{1}, V_{2}, ..., V_{m}] \in \mathbb{R}^{m \times m}$ is corresponding orthogonal or unitary matrix that consists of the unit length eigenvalues called principle eigenvectors. The Karhunen-Loeve-transform [2] sets up a linear transformation of an observed vector x as follows

$$\mathbf{y}_{\mathbf{p}} = \mathbf{V}_{\mathbf{S}}^{\mathrm{T}}\mathbf{x} \tag{2.3}$$

Where $x = [x_1(t), x_2(t), ..., x_m(t)]^T$ is the zero-mean observed vector (input vector), $y_p = [y_1(t), y_2(t), ..., y_n(t)]^T$ is the output vector referred as the vector of principle components (PCs), and $V_S = [v_1, v_2, ..., v_m]^T \in \mathbb{R}^{mxn}$ is the set of signal subspace eigenvectors, with the orthonormal vectors $v_i = [v_{i1}, v_{i2}, ..., v_{im}]^T$, (i.e. $(v_i^T v_j = \delta_{ij}) \forall j \leq i, (\delta_{ij})$ is the Kronecker delta. The vectors $(v_i \forall i = 1, 2, ..., n)$ are the eigenvectors of the covariance matrix R_{XX} , while the variances of the PCs y_i are the corresponding principal eigenvalues. Therefore, we can re-formulate the equations as

$$R_{xx}v_i = \lambda_i v_i \quad \forall i = 1, 2, \dots, n$$
(2.4)

Where v_i are the eigenvectors, λ_i are the corresponding eigenvalues and $R_{xx} = E[xx^T]$ is the covariance matrix of zero-mean signal x and E is the expectation operator. Also we can re-write the equation (2.4) in a matrix form $V^T R_{xx} V = \Lambda$ where Λ is the diagonal matrix of eigenvalues of the covariance matrix R_{xx} . However, to compute the eigenvalues and corresponding eigenvectors of the covariance matrix R_{xx} , one might use the Single Value Decomposing method [2], which is referred as prewhitening or decorrelation of input data, to transform the observations to a set of orthogonal (decorrelated) signals. However, The PCA algorithms carry out principle components where they are uncorrelated, i.e. they are orthogonal. However, PCA didn't recover the sources from the observed data mixtures.

2.3 Independent Component Analysis (ICA)

The concept of Independent Component Analysis (ICA) is a vital algorithm for the BSS [1]. P. Comon [7] was the first to describe the fundamentals of this technique and defined its name in 1994. ICA has been succeeding as an attractive algorithm since it has been applied in many diverse fields as a method that can retrieve the original sources from the linearly mixed independent components.

2.3.1 The Instantaneous ICA Framework

An instantaneous ICA mixtures model with *n* source signals is defined as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{v} \tag{2.5}$$

Where A is an mxn mixing matrix, x represents a matrix with m observed mixed signal vectors as in (2.6), **s** is a matrix with n source signals as in (2.7) and **v** is an Additive Gaussian Noise.

$$\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^{\mathrm{T}}$$
(2.6)

$$\mathbf{s} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n]^{\mathrm{T}}$$

In general, ICA framework carries out with the following assumptions:

1. The source signals s are assumed to be statistically independent, which means that:

$$p(\mathbf{s}) = p(s_1, s_2, ..., s_n) = p(s_1)p(s_2) ... p(s_n)$$

No more one source signal has Gaussian distribution. Since, the mixing matrix A is not identifiable for more than one Gaussian Independent Components. [12], [14].

For simplicity, we will assume that A is square, i.e. m = n, for the rest of analysis. The main idea of ICA is to recover the source signals from the observed signals without any knowledge about the source signal or the mixing matrix. In order to achieve this purpose, the ICA algorithm computes the weighting matrix W that is equal to the inverse of matrix A. However, the estimated source signals y is given as below

$$\mathbf{y} = \mathbf{W}^{\mathrm{T}} \mathbf{x} \tag{2.8}$$

However, it is a linear transformation so we can estimate one of the independent components due to $w^T x$, where w is a column vector of the demixing matrix W in (2.8). Generally speaking, ICA methods usually divide into two steps, preprocessing (pre-whitening) and rotation. Pre-whitening or preprocessing is actually half of the ICA process. Pre-whitening is based on second order statistics (SOS) and the rotation process needed to separate the mixtures, which is based on ICA methods. In the next section, we are going to analyze a basic approach using the pre-whitening method.

2.3.1.1 Preprocessing

Some of adaptive ICA algorithms require pre-whitening, also called sphering or normalized spatial decorrelation. The preprocessing consists of two steps. The first step aims to center the mixed signals \mathbf{x} by removing its mean of mixed signals as in (2.1). After mixed signals have centered, we express the centering matrix $\dot{\mathbf{x}}$ as follows

The Eigen-Vector Decomposition SVD [28] can be used to decompose the covariance matrix of $\dot{\mathbf{x}}$ and its corresponding operation is expressed as in (2.10)

$$C_{\mathbf{x}} = \mathbf{E}[\dot{\mathbf{x}}\dot{\mathbf{x}}^{\mathrm{T}}] = \mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}}$$
(2.10)

Where E represents the eigenvectors which are orthogonal matrix of mixed signals, and D expresses in (2.11) represents the diagonal Eigenvalues of matrix C_x .

$$D = diag(d_1, d_2, \dots, d_n)$$
(2.11)

Thus, the whitening process of $\dot{\mathbf{x}}$ is expressed by (2.12)

$$Z = D^{-\frac{1}{2}} E^{T} \dot{\mathbf{x}} = V \dot{\mathbf{x}}$$
(2.12)

Where V equals $D^{-\frac{1}{2}}E^{T}$ and represents the whitening matrix or Mahalnobis transform of $\dot{\mathbf{x}}$. Equation (2.12) shows that the centered matrix $\dot{\mathbf{x}}$ is a linear transformed to a matrix Z and the covariance matrix of Z equals to identity matrix [2]. In other words, the means of the matrix Z is uncorrelated and have a unit variance. Figure 2.2 illustrates three basic transformations of observed data x: pre-whitening, PCA and ICA, respectively.

In next sections, we are going to discuss some of the basic approaches in order to conduct the ICA performance of instantaneous mixtures, and we assume that the additive weight Gaussian noise terms v in (2.5) are negligible or reduced to be at negligible levels due to the preprocessing stated in the previous section.

2.3.1.2 Nonlinear function choice (Activation function):

Nonlinear function or Activation function is the source signal model $\vartheta(\mathbf{y})$. However, it is very important to select a suitable nonlinear function depending on our source signals. There has been much research regarding this topic. The most suitable ones for super Gaussian sources and Sub Gaussian sources are proposed by Hyvarinen in [14], [15], as follows:

For super Gaussian sources which are the source signals having a positive kurtosis sign, e.g. a Laplacian signal.

$$\vartheta^{\text{SUP}}(\mathbf{y}) = -2\tanh(\mathbf{y}) \tag{2.13}$$

For sub Gaussian sources which are the source signals having a negative kurtosis sign, e.g. a uniform signal.

$$\vartheta^{SUB}(\mathbf{y}) = \tanh(\mathbf{y}) - \mathbf{y} \tag{2.14}$$

2.3.1.3 The Learning Update Rules

The update rules divide into two categories based on learning procedure, online learning and offline learning [2].

2.3.1.3.1 Batch Learning (Offline learning):

Batch learning is the kind of algorithms that have an update rule requiring the whole training sample in every step of iteration. Usually, in batch learning, the update rule relies on the expectation of observed data x. In practice, the expectation of observed data x approximated by the mean of observed data x.

2.3.1.3.2 Stochastic gradient (Online learning):

Online learning is the kind of algorithms that have an update rule that doesn't require the whole training sample in every step of iteration. Mathematically speaking, these categories of update rules don't rely on the expectation of observed data x. In other words, one might get online learning from the batch learning by dropping the expectation operator from the offline update rule.

2.3.1.4 ICA based on Maximization of non-Gaussianity

In this section, we will study the ICA algorithm based on the non- Gaussianity criteria. The non-Gaussianity approach is based on the Center Limit Centre (CLT) which states that for independent sources their sum will become closer to Gaussian distribution

than each individual source. The CLT shows that for whitened data (in section 2.3.1), finding an independent source is equal to finding the direction of w which gives a component of maximum non-gaussianty [2], [12-16].

For sake of illustration, one assumes that the observed data is $\mathbf{x} = A\mathbf{s}$ and the weight vector is w. However, in order to find one of the independent components, x that is $\mathbf{y} = \mathbf{w}^{T}\mathbf{x}$, the vector \mathbf{w}^{T} should be in the row of the inverse matrix A^{-1} .

$$\mathbf{y} = \mathbf{w}^{\mathrm{T}} \mathbf{x} = \mathbf{v}^{\mathrm{T}} \mathbf{s} \tag{2.15}$$

Therefore, this implies that, by maximizing the non-gaussianty of \mathbf{y} in terms of \mathbf{w} , then, we will get one of the independent components present in x. In addition, this is the same as we have in the whitened data before applying ICA methods. There are different criteria for measuring non-Gassianity. Next, we will study some of these criteria.

2.3.1.4.1 Kurtosis Measure

Kurtosis is a dimensionless measure and refers to a fourth order cumulant of a random variable. Mathematically, one can express the normalized kurtosis of zero mean random variable in terms of 2nd – 4th order moments as follows:

$$kurt(y) = \frac{E[y^4]}{(E[y^2])^2} - 3$$
 (2.16)

The important feature of kurtosis is that kurtosis kurt(y) is equal to zero for Gaussian random variables. So, kurtosis is a tool to measure the relative sharpness and flatness of distributions. However, kurtosis with positive sign is termed to the super Gaussian data and kurtosis with negative sign is termed to the sub Gaussian data.

For sake of optimization, the kurtosis kurt(y) expression in (2.16) can be expressed as follows:

$$kurt(y) = E[y^4] - 3(E[y^2])^2$$
(2.17)

This expression is easier to optimize, by omitting the denominator since the term $(E[y^2])^2$ is always positive. In this approach, usually the data has to be whitening in order to ensure that the source signals are uncorrelated and have unit variance, i.e. the source signals are orthonormal. Then apply the ICA methods to find the direction of $y_i = w^T x$, where kurt(y) is maximized, i.e. the direction of the most non-Gaussian component. After that the orthogonal projection $y_i = w^T x$ will give us the separated component.

2.3.1.4.2 Gradient algorithm using kurtosis

After pre-whitened, the observation signals can be expressed as z = Vx so that the estimated signals y becomes as follows:

$$\mathbf{y} = \mathbf{w}^{\mathrm{T}} \mathbf{z} = \mathbf{w}^{\mathrm{T}} \mathbf{V} \mathbf{x} \tag{2.18}$$

In general, we are beginning from the initial random vector of w, and then looking for a direction of w at which the value of kurtosis of estimated signals $y = w^{T}z$ is increasing. In fact, we consider the Maximizing of the absolute value of kurtosis and it is suitable for both super Gaussian and sub Gaussian signals. However, to perform the gradient decent method of kurtosis under the constraint that $||w||^{2} = 1$ as follows:

$$\frac{\partial |\operatorname{kurt}(w^{\mathrm{T}}z)|}{\partial w} = 4\operatorname{sgn}(\operatorname{kurt}(w^{\mathrm{T}}z))\{\mathbb{E}[z(w^{\mathrm{T}}z)^{3}] - 3w||w||^{2}\}$$
(2.19)

In terms of direction, we can simplify the gradient vector by omitting the scalar term and the second term. Then, one can update the expression as follows:

$$w = w + \gamma \Delta w$$

$$\Delta w = \text{sgn}(\text{kurt}(w^{T}z))\{\text{E}[z(w^{T}z)^{3}]\}$$

$$w = \frac{w}{||w||}$$

(2.20)

2.3.1.5 Fixed point algorithm

In [14-15], Hyvarinen proposed the ICA algorithm that takes advantage of a Newten-type method (Lagrange Multipliers) for maximizing the Kurtosis in order to increase the speed and Robustness in previous ICA algorithm. The derivation of a fixed point algorithm is discussed in depth in [14]. He forms the new update law of an ICA algorithm as follows:

$$w^{+} = E[z(w^{T}z)^{3}] - 3w$$
(2.21)

The basic scheme for one independent component estimated is as follows:

- 1. Prewithening Data, i.e. z = Vx
- 2. An initial value for the Wight vector w that has ||w|| = 1
- 3. Find the updated weighed vector

$$w^+ = E[z(w^T z)^3] - 3w$$

4. Normalize and update the weight vector

$$w = \frac{w^+}{\|w^+\|}$$
 Where $\|w\|$ is the norm of w .

5. Go back to step 3 until the convergence.

In order to estimate many components, one can apply the previous scheme Ntimes to get all sources (components) that exist in the observed data. Notably, we should always keep each new estimated component orthogonal to the previous one in order to prevent estimating the same component each time as follows:

- 1. Prewithening Data, i.e. z = Vx
- 2. An initial value for the Wight vector w that has ||w|| = 1
- 3. Find the updated weighed vector

$$w^+ = E[z(w^T z)^3] - 3w$$

4. Find a projection matrix *B* then Set

$$w^+ = w^+ - BB^T w^+$$

5. Normalize and update the weight vector

$$w = \frac{w^+}{\|w^+\|}$$

Where ||w|| is the norm of w.

6. Go back to step 3 until the convergence.

Practically, the projection matrix B contains all vectors w that calculate to find the previous components. However, the transformation in step 4 enables the algorithm to converge to a different component from the previous ones that were discovered.

2.3.1.6 Negentropy Measure

Let us define that J as a negentropy of random vector y where it represents a normalized version of entropy of a random vector s. In general, negentropy is an information theoretic tool that is used to measure the distance of random variables from the Gaussian distribution at the same covariance. Mathematically, we can express J(y)as follows:

$$J(y) = H(y^{GAUSS}) - H(y)$$
(2.22)

However, negentropy J is an appropriate measure of nonGassianity, statistically [2]. Instead of estimating the negentropy, one can use an approximation of negentroy that is proposed in [11] as follows:

$$J(y) \approx \frac{1}{12} E^2 [y^3] + \frac{1}{48} [kurt(y)]^2$$
 (2.23)

By using the higher order cumulants and taking advantage of a non-quadratic function G to simplify the approximation of negentropy, one can rewrite the approximation of negentroy J(y) as follows:

$$J(y) = (E[G(y)] - E[G(v)])^2$$
(2.24)

where v represents a Gaussian variable with zero mean and unit variance. Based on this approximation of negentroy, FastICA algorithm is structured and analyzed in depth by Hyvarinen in [12].

The gradient algorithm can be estimated in order to produce a fixed point algorithm as follows (after the pre-whitening):

$$\Delta w = \mu E[zg(w^{T}z)]$$

$$w = w/||w|| \qquad (2.25)$$
Where $\mu = E[G(w^{T}z)] - E[G(v)]$, and $g(y) = \frac{\partial G(y)}{\partial y}$.

One of the most common choice non-quadratic functions of g(y), amongst others, is as follows:

$$g(y) = \tanh(\alpha y), \forall \ 1 \le \alpha \le 2$$
(2.26)

Practically, the maximum of the approximation of the negentropy of $w^{T}z$ are occurred at certain optima of $E[G(w^{T}z)]$ under the constraint $||w||^{2} = 1$. In order to find a certain optima, one can solve the gradient of the Lagrangian to zero (Kuhn-Tucker conditions) [2].

$$F(z, w) = E[zg(w^{T}z)] + \lambda w = 0$$
(2.27)

Newton's method is used to solve this equation, we have

$$\frac{\partial F}{\partial w} = E[zz^{T}g'(w^{T}z)] + \lambda I \approx E[zz^{T}]E[g'(w^{T}z)] + \lambda I = \{E[g'(w^{T}z)] + \lambda\}I(2.28)$$

According to the Newton's method, the update rule becomes as follows:

$$w^{+} = w - \left[\frac{\partial F}{\partial w}\right]^{-1} F \qquad (2.29)$$

Finally, the update rule for the FastICA is

$$w^{+} = E\{z[g(w^{T}z)]^{T}\} - E\{g'(w^{T}Z)\}w$$
(2.30)

In general, the FastICA algorithm can be stated up as follows:

The basic FastICA scheme for one independent component estimated is as follows [12-15]

- 1. An initial value for the Wight vector w
- 2. Find the updated weighed vector

$$w^+ = E\{Z[g(w^TZ)]^T\} - E\{g'(w^TZ)\}w$$

Where the g is a non-quadratic function such as $g(y) = y^3$, $y = w^T X$ and g' is the derivative of the non-quadratic function g.

3. Normalize and update the weight vector

$$w = \frac{w^+}{\|w^+\|}$$

Where ||w|| is the norm of w.

4. Go back to step 2 until the convergence.

In order to estimate many components, one can apply the previous scheme Ntimes to get all sources (components) that exist in the observed data. Notably, we should always keep each new estimated component orthogonal to the previous one in order to prevent estimating the same component each time as follows:

- 1. Pre-whitening Data, i.e. z = Vx
- 2. An initial value for the Wight vector w that has ||w|| = 1

3. Find the updated weighed vector

$$w^+ = E\{Z[g(w^TZ)]^T\} - E\{g'(w^TZ)\}w$$

4. Set

$$w^+ = w^+ - BB^T w^+$$

5. Normalize and update the weight vector

$$w = \frac{w^+}{\|w^+\|}$$

Where ||w|| is the norm of w.

6. Go back to step 3 until the convergence.

There is a method that one can use to estimate all independent components simultaneously instead of estimating each independent component individually. This can occur by using different learning rules for all estimated signals and apply the symmetric decorrelation to ensure the convergence in ICA method. The symmetric decorrelation is as follows:

$$W = W(W^{T}W)^{\frac{-1}{2}}$$
(2.31)

Where $W = [w_1, w_2, ..., w_N]$ is the matrix of the vectors w_i .

2.3.1.7 ICA based on Maximum Likelihood Estimation

In this part, we employ Maximum Likelihood (ML) as a contrast function in ICA algorithm. ICA based on ML estimation carried out in [28], [29]. Assume that the unmixing matrix denotes $W^T \approx A^{-1}$, thus, we can recall the instantaneous mixture model in (2.5) as follows:

$$\mathbf{x} = \mathbf{A}\mathbf{s}$$

Then the estimated signals are

$$\mathbf{y} = \mathbf{W}^T \mathbf{x}$$

Due to a basic property of linear transformed random vectors

$$p_x(x) = |\det(A^{-1})|p_s(s)$$
 (2.32)

By the assumption of the statistical independence between the estimated source signals y and $p_s(s) \approx p_y(y)$, we can show that the probability density of observed data x is as follows:

$$p_{x}(x) = |\det(W)|p_{y}(y) = |\det(W)|\prod_{i=1}^{n} p_{i}(y_{i})$$
(2.33)

Where $y_i = w_i^T x$ where w_i is a column of elements in W, therefore, we can express $p_x(x)$ as follows:

$$p_{x}(x) = |\det(W)| \prod_{i=1}^{n} p_{i}(w_{i}^{T}x)$$
(2.34)

By constructing the likelihood function of W, as a product of the densities at each observed signal, we get L(W) is:

$$L(W) = \prod_{i=1}^{n} p_i(w_i^{T}x) |det(W)|$$
(2.35)

Then, by optimizing the expectation of log-likelihood function L(W) as follows:

$$E[\log\{L(W)\}] = E[\log\{\prod_{i=1}^{n} p_i(w_i^T x) | det(W)|\}] = E[\sum_{i=1}^{n} \log\{p_i(w_i^T x)\}] + \log(|det(W)|)$$
(2.36)

Then, the ML contrast function of W express as follows:

$$G(W) = E[\sum_{i=1}^{n} \log\{p_i(w_i^T x)\}] + \log(|det(W)|)$$
(2.37)

Now, we are going to use a gradient decent approach in order to maximize ML, G(W), contrast function with respect to W. However, one can show that:

$$\frac{\partial G(W)}{\partial W} = (W^{\mathrm{T}})^{-1} + E[\vartheta(Wx)x^{\mathrm{T}}]$$
(2.38)

Where $\vartheta(Wx) = \vartheta(y) = [\vartheta_1(y_1), \vartheta_2(y_2), ..., \vartheta_n(y_n)]^T$ and it is a nonlinear function or (Activation function) that performs the source signal model. We can derive the nonlinear function according to following equation:

$$\vartheta_{i}(y_{i}) = \frac{\partial}{\partial y_{i}} \log\{p(y_{i})\} = \frac{1}{p(y_{i})} \frac{\partial p(y_{i})}{\partial y_{i}}$$
(2.39)

The update rule for ML estimation using the gradient decent method is expressed

$$W = W + \gamma \Delta W \tag{2.40}$$

Where

as:

$$\Delta W \alpha \frac{\partial G(W)}{\partial W} = \{ (W^{T})^{-1} + E[\vartheta(Wx)x^{T}] \}$$

And, γ is the learning rate or the step size

In [12], the same result has been achieved when minimized the Kullback-Leibler (KL) divergence between the joint and the product of the marginal distributions of estimated signals y_i . The vital point in Amari's paper [2], [71] showed that the parameter space in this optimization problem is a Riemannian metric Structure instead of Euclidian Structure. However, the steepest decent should be given by the natural gradient instead of using the gradient decent method. In that sense, the update rule of the natural gradient is given as follows (by multiplying the right hand of previous update rule by W^TW):

$$W = W - \gamma \frac{\partial G(W)}{\partial W} W^{T} W$$
(2.41)

Then,

$$\Delta W \alpha \frac{\partial G(W)}{\partial W} W^{T} W = \{I + E[\vartheta(y)y^{T}]\} W$$

Notably, Natural gradient method is based on the attempt to implement the Newton decent method by the approximation of hessian inverse $(\nabla^2 G)^{-1} \approx W^T W$.

2.3.1.8 ICA based on Entropy Maximization

The differential entropy of a random vector s with density p(s) can be expressed as follows [2], [36]:

$$H(\mathbf{s}) = -\int p(\mathbf{s}) \log\{p(\mathbf{s})\} \, d\mathbf{s}$$
(2.42)

Let's define J a negentropy of random vector s where it represents a normalized version of entropy of a random vector s. In general, negentropy is information theoretic tool that uses to measure the distance of random variables from the Gaussian distribution at the same covariance. Mathematically, we can express J(s) as follows:

$$J(\mathbf{s}) = H(\mathbf{s}^{GAUSS}) - H(\mathbf{s})$$
(2.43)

In addition, mutual information can be a good method to measure the statistical dependence between random variables. P. Comon shows that the mutual information is a good metric of statistical dependence [7]. In that sense, if the random variables $\mathbf{s} = \{s_1, s_2, ..., s_n\}$ are statistically independent, then the mutual information I(\mathbf{s}) is equal to zero. We can define the mutual information I(\mathbf{s}) as follows:

$$I(\mathbf{s}) = \sum_{i=1}^{n} H(s_i) - H(\mathbf{s})$$
(2.44)

2.3.1.9 Bell-Sejnowski method

ICA algorithm based on minimized the mutual information metric is proposed by Bell and Sejnowski in [18], [2]. They use the mutual information as a way to measure the independent between random variables. One can assume that the un-mixing matrix is W and the estimated source signals are $\mathbf{y} = \mathbf{W}^T \mathbf{x}$. However, let's express the mutual information as follows:

$$I(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{W}^T \mathbf{x})$$
(2.45)

By using the fact that the differential entropy is in general not invariant under arbitrary invertible maps, we can express the mutual information as follows:

$$I(\mathbf{y}) = \sum_{i=1}^{N} H(y_i) - H(\mathbf{x}) - \log |\det(W)|$$
(2.46)

One can state the optimization problem as follows: we can minimize the mutual information $I(\mathbf{y})$ with respect to the un-mixing matrix W to estimate the un-mixing matrix W that makes the estimated signals \mathbf{y} more statistically independent. Then, one can re-write the expression of differential entropy as follows:

$$H(y_i) = -E[\log\{p(y_i)\}]$$
(2.47)

And, the mutual information expression becomes as follows:

$$I(\mathbf{y}) = -\sum_{i=1}^{N} E[\log\{p(y_i)\}] - H(\mathbf{x}) - \log |\det(W)|$$
(2.48)

Despite the fact that the estimated source signals y are uncorrelated, because of the ICA assumption that the source signals s are statistically independent, one can simplify the mutual information cost function to become almost identical with ML cost function G(W). One can show that the determent of un-mixing matrix det (W) is constant as follows:

Since we have uncorrelated estimated signals y, then we can state that:

$$E[yy^{T}] = I \Rightarrow WE[xx^{T}]W^{T} = I \Rightarrow then, det(W) det(E[xx^{T}]) det(W^{T}) = 1$$

This implies that the det(W) must be constant. However, $H(\mathbf{x})$ is not a function of W. So, it can be omitted from the MI contrast function $I(\mathbf{y})$ as follows:

$$I(\mathbf{y}) = -\sum_{i=1}^{N} E[\log\{p(y_i)\}] - \log |\det(W)|$$
(2.49)

$$I(\mathbf{y}) = -\left[E\left[\sum_{i=1}^{N} \log\{p_i(\mathbf{w}_i^T \mathbf{x})\}\right] + \log\left(|\det(\mathbf{W})|\right) \right]$$
(2.50)

Then, recall the ML contrast function of W that is expressed as follows:

$$G(W) = E[\sum_{i=1}^{N} \log\{p_i(w_i^T \mathbf{x})\}] + \log(|det(W)|)$$
(2.51)

Apart from the minus sign, both contrast functions look very similar. By minimizing the MI cost function with respect to W, we will end up with the same update rule of the ML estimation as follows:

$$\Delta W \alpha \{ (W^{T})^{-1} + E[\vartheta(W\mathbf{x})\mathbf{x}^{T}] \}$$
(2.52)

In conclusion, ICA algorithm based on deferent metrics of statistical independent (MI, ML and KL criterions) ends up with the same update algorithms.

2.3.1.10 ICA based on Tensorial Methods

A tensor is a multi-linear operator that is derived from the Taylor series of the log-characteristic function f(w) = E[exp(jwx)], where x is a zero mean random variable. One can express the Taylor series of the log-characteristic function log {f(w)} as follows:

$$\log f(w) = \kappa_1(jw) + \kappa_2 \frac{(jw)^2}{2!} + \dots + \kappa_r \frac{(jw)^r}{r!} + \dots \quad (2.53)$$

The coefficients $\kappa_i \forall i = 1, 2, ...$ are called Cumulants. In multivariate situations, one can call them cross cumulants, which are similar to cross conveniences. In BSS problem, Kurtosis can be expressed as a fourth-order cross-cumulant as follows:

$$Kurt(y) = cum(y_i, y_j, y_k, y_l)$$
(2.54)

Where $y_i = \sum_i w_i x_i$, then

$$\operatorname{Kurt}(\sum_{i} w_{i} x_{i}) = \operatorname{cum}\left(\sum_{i} w_{i} x_{i}, \sum_{j} w_{j} x_{j}, \sum_{k} w_{k} x_{k}, \sum_{l} w_{l} x_{l}\right)$$
$$= \sum_{ijkl} w_{i}^{4} w_{j}^{4} w_{k}^{4} w_{l}^{4} \operatorname{cum}(x_{i}, x_{j}, x_{k}, x_{l}) \quad (2.55)$$

In general, the tensor is defined as the fourth order cumulants, and it is similar to the covariance matrix for second order moments. The cumulant structure is symmetry, however the eigenvalue decomposion is always valid as it's shown in [1]. Let's assume that we have an eigenvector matrix Vand the corresponding eigenvalues λ , and then one can decompose the tensor Fas follows:

$$\mathbf{F} = \lambda \mathbf{V} \tag{2.56}$$

Likewise, the pre-whitened data $Z = VAs = W^Ts$, where the matrix W is the unmixing matrix, however, matrices W, W^T will be orthogonal. One can express the eigenmatrix $V = w_m w_m^T$, where the vector w_m is the m-th row of matrix W, of the following tensor F with the corresponding eigenvalues which represents the kurtosis of the independent components [1], [16], [27] as follows:

$$F = F_{ij} = \sum_{kl} V_{kl} cum(z_i, z_j, z_k, z_l) = \sum_{kl} w_{mk} w_{ml} cum(z_i, z_j, z_k, z_l) = \cdots$$
$$= w_{mi} w_{mj} kurt(s_m)$$
(2.57)

In other words, one can estimate the un-mixing matrix W for the independent sources at given eigenmatrices of the tensor. This case is valid if we have the distinct eigenvalues, otherwise, the problem will be difficult to solve.

2.3.1.11 PARAllel FACtor (PARAFAC) algorithms

Several BSS algorithms based on the Parallel factor the (PARAFAC) model have been proposed i.e. [11], [59], [60], [86]. PARAFAC is a multi-linear tool for tensor decomposing in sum of rank-1 tensors.

2.3.1.12 Joint Approximation Diagonalisation (JAD)

In order to overcome the problem in the tensor, Cardoso [16] is proposed JADE algorithm by diagonals of the tensor of matrix F. Since the tensor F is a linear combination of terms of $w_i w_i^T$, one can express the tensor F of any matrix as the eigenvalue decomposition form, i.e. the matrix $Q = WFW^T$. This allows estimating the unmixing matrix W by minimizing the off-diagonal terms or maximizing the diagonal terms of Q. However, the cost function of the JADE algorithm was proposed by Cardoso as follows:

$$\max_{W} J_{iade}(W) = \max_{W} \sum_{i} \|diag(WF_{i}W^{T})\|^{2}$$
(2.58)

Where F_i represents the tensor of the different matrix V_i , where V_i might be the eigenmatrices of the tensor F_i . JADE algorithm is not as effective in terms of convergence and computational especially for the high dimension [2], [16] [17].

2.3.2 The Convolutive ICA Mixtures

In the previous sections, we have investigated several methods based on the instantaneous case in the ICA framework. All previous methods perform well in terms of quality of separating source signal from the linear mixed sources. However, in practicality, if we apply these methods on real life applications, i.e. a multipath channel in communication, room environment for a sound separation, we will fail to recover source signals. The major reason is because the instantaneous model doesn't hold the varying in the mixing matrix. Figure 2.2 illustrates that a single channel convolution and a deconvolution process. A multi-channel deconvolution problem can be considered as a natural extension of the instantaneous BSS problem. With this problem, assume an m-

dimensional vector of received discrete time signals $x(k) = [x_1(k), x_2(k), ..., x_m(k)]^T$ at time k is assumed to be produced from an n-dimensional vector of source signals $s(k) = [s_1(k), s_2(k), ..., s_m(k)]^T$, where $m \ge n$, by using a stable mixture model [2].

$$\mathbf{x}(\mathbf{k}) = \sum_{p=-\infty}^{\infty} \mathbf{H}_{p} \mathbf{s}(\mathbf{k} - \mathbf{p}) = \mathbf{H}_{\mathbf{k}} * \mathbf{s}(\mathbf{k}), \text{ with } \sum_{-\infty}^{\infty} \left\| \mathbf{H}_{p} \right\| \le \infty$$
 (2.59)

Where * represents the convolution operator and H_p is an $(m \times n)$ matrix of mixing coefficients at time-lag p.

One can define that

$$H(z) = \sum_{p=-\infty}^{\infty} H_p z^{-p}$$
(2.60)

where z^{-1} represents the unit time-delay operator, i.e. $z^{-p}[s_i(k)] = s_i(k-p)$.

Generally speaking, the goal of multichannel deconvolution is to recover the source signals, up to the possibly scaled and time delayed, from the received signals by using the approximate knowledge about the source signal distributions and statistics.



a) Diagram illustrating convolution and deconvolution process of the signal channel



b) Multichannel blind deconvolution problem (MBD)

Figure 2.2: Block diagram of the Convolutive Mixtures.

Typically, we assume every source signal $s_i(k)$ is an i.i.d (independent and identically distributed) sequence. One can express the convolutive mixture model as follows:

$$\begin{bmatrix} x_{1}(n) \\ x_{2}(n) \\ \dots \\ x_{M}(n) \end{bmatrix} = \begin{bmatrix} a_{11} & \dots & a_{1N} \\ a_{21} & \dots & a_{2N} \\ & \dots & & \\ a_{M1} & \dots & a_{MN} \end{bmatrix} * \begin{bmatrix} s_{1}(n) \\ s_{2}(n) \\ \dots \\ s_{N}(n) \end{bmatrix}$$
(2.61)

Applying Short Time Fourier Transform (STFT) for this model gives us two advantages as follows:

- In frequency domain, signals become more super Gaussian, which will be more suitable for any ICA learning algorithms.
- In frequency domain, one can use the approximation of linear convolution with multiplication.

$$STFT \left\{ \begin{bmatrix} x_{1}(n) \\ x_{2}(n) \\ \dots \\ x_{M}(n) \end{bmatrix} \right\} = STFT \left\{ \begin{bmatrix} a_{11} & \dots & a_{1N} \\ a_{21} & \dots & a_{2N} \\ \dots & \dots & \dots \\ a_{M1} & \dots & a_{MN} \end{bmatrix} * \begin{bmatrix} s_{1}(n) \\ s_{2}(n) \\ \dots \\ s_{N}(n) \end{bmatrix} \right\}$$
(2.62)

Let's assume M = N, the STFT of convolutive model becomes as follows:

$$\begin{bmatrix} x_{1}(f,t) \\ x_{2}(f,t) \\ \dots \\ x_{N}(f,t) \end{bmatrix} = \begin{bmatrix} A_{11}(f) & \dots & A_{1N}(f) \\ A_{21}(f) & \dots & A_{2N}(f) \\ \dots & \dots & \dots \\ A_{N1}(f) & \dots & A_{NN}(f) \end{bmatrix} \begin{bmatrix} s_{1}(f,t) \\ s_{2}(f,t) \\ \dots \\ s_{N}(f,t) \end{bmatrix}$$
(2.63)

$$\Rightarrow \mathbf{x}(\mathbf{f}, \mathbf{t}) = \mathbf{A}_{\mathbf{f}} \mathbf{s}(\mathbf{f}, \mathbf{t}), \ \forall \mathbf{f} = 1, \dots, \mathbf{F}$$
(2.64)

Where F is the number of FFT points, and also note that we use the STFT instead of FT to preserve the stationary property of the signals and to divide the signal into shorter overlapping frames. In other words, we transform the convolutive mixture problem into L instantaneous problems by assuming the statistical independence among the frequency bins. However, one can simply transform the convolution problem into multiplication by using the windowing method, i.e. the window larger than the filter length such as $F \gg T$. But in fact, this case is not easy to implement since the data will be in complex number form, which affects the stability factor of the algorithm [3], [39], [71]. In additional, the scale and permutation will have an effect in this approach as it will be explained later.

2.3.2.1 Time-Domain Methods

One can estimate source signals by estimating the un-mixing coefficients in time domain. The convoltutive mixtures model can be expressed as follows:

$$x_{i}(n) = \sum_{j=1}^{N} \sum_{d=1}^{T} a_{ijd} s_{j}(n-d) \quad \forall i = 1, 2, ..., N$$
(2.65)

In order to estimate the source signals from the mixtures in this model, one can estimate the un-mixing coefficients filter w_{ijd} in FIR filter architecture (feedback architecture) as follows:

$$y_i(n) = \sum_{j=1}^N \sum_{d=1}^T w_{ijd} s_j(n-d) \quad \forall i = 1, 2, ..., N$$
 (2.66)

Delay-compensation problem considers a major issue in time-domain models. Several researchers carry out some methods to solve these problems in time domain. One of these methods is the use of feedback architecture which is proposed by Torkkola [2], [39], [133] in order to remove temporal dependencies and stabilize the cross-weights. Some of his research utilized the IIR structure. Lee [36], [38] presented the following IIR structure to separate the source signals from the mixtures as follows:

$$y_i(n) = x_i(n) - \sum_{j=1}^N \sum_{d=1}^T w_{jd} y_j(n-d) \quad \forall i = 1, 2, ..., N$$
 (2.67)

Or

$$y(n) = x(n) - W_0 y(n) - \sum_{k=1}^{L} W_k y(n-k)$$
(2.68)
So, in order to estimate the un-mixing matrix W, Lee maximizes the joint entropy H(g(y)), where g(.) is the sigmoid function which is used in the Bell-Sejnowski's method. He presents a new update rule for this model in time domain as follows:

$$\Delta W_0 = -(I + W_0)(I + E[\phi(y)y^T])$$
(2.69)

$$\Delta W_{k} = -(I + W_{k})E[\varphi(y)y^{T}(n - k)]$$
(2.70)

Where $\varphi(y) = -\frac{\partial \log p(y)}{\partial y}$.

Several drawbacks are noticed in time-domain methods for recovering the source signals. For long mixing filter, which means long transfer functions, the computation will be too expensive [2], [133], [71]. Also, using the IIR filter instead of long FIR filter to overcome this problem it suffers from instability and needs to invert the non-minimum phase filters [39]. However, time domain methods are suitable and very efficient for small mixing filters such as in communication channel [2]. In addition, Torkkola proposes a feedback structure to overcome the spectral whitening problem in feed forward structure in [2], [38], and [133]. Next, we will explain some methods in frequency domain to solve the convolutive mixtures problem.

2.3.2.2 TRINICON Blind Source Separation

The TRINICON algorithm is based on the time domain approach and proposed by Buchner et al. [131], [1]. The main drawback in this algorithm is its sensitivity to outlets, not robust, especially in a real world recording problems. See [133], [3] and [38].

In their work, they use the multivariate models as a cost function in order to consider the whole temporal structure of the original sources. Actually, they just simplified the optimal formula of BSS in time domain approach by windowing the observed signals in terms of blocks. Let us assume L_B is the length of each block. One can express the separated model of each block as follows

$$y(b, win) = x(b, win)W(b)$$
(2.71)

where

- b denotes the block index
- win $\in \{0, ..., L_B 1\}$ is the time-shift index within the block
- x(b, win) = [x₁(b, win), ..., x_M(b, win)] is the M observed signals which segmented in blocks of length L_B.

 $y(b, win) = [y_1(b, win), ..., y_N(b, win)]$ is the estimated N source signals W(b) is the separation matrix for a given block and it is given by:

$$W(b) = \begin{bmatrix} W_{11}(b) & \dots & W_{1M}(b) \\ W_{21}(b) & \dots & W_{2M}(b) \\ \\ W_{N1}(b) & \dots & W_{NM}(b) \end{bmatrix}$$
(2.72)

So, assume L is the length of the FIR filters; however, the m-th mixture is modeled as

$$x_{m}(b, win) = [x_{m}(bL + win), ..., x_{M}((b-2)L + 1 + win)]$$
 (2.73)

Therefore, the output signals are modeled as:

$$y(b, win) = [y_n(bL + win), ..., y_N(bL - D + 1 + win)]$$

= $\sum_{m=1}^{M} x_m(b, win) W_{mn}(b)$ (2.74)

Where D is the number of the time-lags.

The TRINICON algorithm aims to estimate each of the demixing matrixes $W_{mn}(b)$ based on three common optimization criteria:

- Minimization of the cross-correlation of the output over multiple timelags.
- Based on standard ICA algorithm "Non-Gaussianity.
- Minimization of the cross-correlation of the output at different instant time "Non-stationary".

2.3.2.3 Frequency-Domain Methods:

In this section, we will express three interesting methods of ICA convolutive mixture in frequency domain as follows:

2.3.2.3.1 Lee's approach

Lee in [3], [38] proposed a FIR un-mixing structure. He used a method that moved from time domain to the frequency domain in order to separate the sources and to avoid the convulsion in time domain. In additional, he developed an update rule of unmixing matrix W_f for each bin, which is similar to natural gradient one as follows:

$$\Delta W_{f} = (I + E[STFT\{\phi(y(n))\}y^{H}(f,t)])W_{f}$$
(2.75)

Lee's method used time-domain and frequency domain. Time domain was used to take advantage of the features of the nonlinearity function $\varphi(y)$. Whereas, he employed the frequency domain just to make the unmixing processes. The proposed framework of Lee's method can be seen in Figure 2.3.



Figure 2.3: Lee's Block diagram

The main drawback in this method is that it requires extra computational complexity. It requires moving from and to the frequency domain in order to use the nonlinearity at each update step. According to Lee's results his method didn't encounter the permutation problem.



Figure 2.4: Smardagdis's Block diagram

2.3.2.3.2 Smardagdis approach

Some researchers only employ the frequency domain for the convolutive problem. However, they perform the source modeling and un-mixing in frequency domain, in order to avoid the complexity in previous methods. Figure 2.4 shows the framework of smaragdis approach [3], [39], where the system adapted to work in frequency domain for each bin individually.

Since, the source signals tend to be more superGaussian in frequency domain; one can take advantage of minimizing the Kullback-Leibler divergence in order to estimate the source signals in frequency domain. Amari derives an update rule for a complex data as follows:

$$\Delta W_{f} = \gamma (I + E[\phi(y(f, t))y^{H}(f, t)])W_{f}$$
(2.76)

Where γ is the learning rate, and $\varphi(y) = \frac{\partial \log p_y(y)}{\partial y}$.

Smaragdis mentioned in his paper that most problems arise in convolutive mixtures come from the permutations and scale ambiguities. Also, he proposed zero-padding method before the FFT in order to smooth the spectra. Smaragdis's framework seems to be a robust and general solution for convolutive mixture problems.

2.3.2.3.3 Independent Vector Analysis (IVA)

Independent vector Analysis (IVA) is developed by Intae et al. [36], [40], it extends the ICA model to be in the multivariate model. Furthermore, they proposed the decoupling frequency in the adaptive learning rule to reduce the possibility of the permutations. Similar to time domain method, IVA updates all the variables at the same time thus it might converge into local minima. Furthermore, IVA algorithm suffers a slow convergence from the high dimensionality of its contrast function and, in terms of cost it's considered to be too expensive to be implemented in real time.

2.3.2.3.4 Parra's approach:

Parra and Spence proposed a new ICA algorithm based on the non-stationary and SOS of signals in order to solve the convolutive mixture problems. Signals are considered to be non-stationary if their statistics are varying in time. Mathematically, one can say that the signal x(n) is a non-stationary signal if $C_x(n) \neq C_x(n + d)$, where $C_x(n)$ is the covariance matrix of x, it represents as $C_x(n) = E[x(n)x^T(n)]$, and d is a constant time.

Now, assume a noisy convolutive mixture model, as follows:

$$x(n) = A * s(n) + e(n)$$
 (2.77)

And the STFT form will be

$$x(f, t) = A(f)s(f, t) + e(f, t) \quad \forall f = 1, 2, ..., F$$
(2.78)

Then, the covariance of the observed data x in frequency domain is

$$C_{x}(f,k) = E[xx^{T}] = A_{f}C_{s}(f,t)A_{f}^{H} + C_{e}(f,k)$$
 (2.79)

Where $C_s(f, t)$ is the source covariance and $C_e(f, t)$ is the noise covariance; next, one can assume that the estimated source covariance is $\widehat{C_s}(f, t)$ and the estimated noise covariance is $\widehat{C_e}(f, t)$. However, one of appropriate error measurements will be as follows:

$$\operatorname{Error}[k] = C_{x}(f,k) - A_{f}\widehat{C}_{y}(f,k)A_{f}^{H} - \widehat{C}_{e}(f,k)$$
(2.80)

Where y(f,t) = W(f)x(f,t) are the estimated sources, one can write the cost function as follows:

$$J(A_{f}, \hat{C}_{y}, \hat{C}_{e}) = \sum_{k} \|Error[k]\|_{F}^{2}$$
(2.81)

In order to estimate each of the parameters A_f , \hat{C}_y , \hat{C}_e , one can find the derivative of J respect to each, as follows $\partial J / \partial A$, $\partial J / \partial \hat{C}_y$, $\partial J / \partial \hat{C}_e$.

Using a stable FIR un-mixing filter W_f , we can re-write the above formula as follows:

$$\hat{C}_{y}(f,k) = E[yy^{T}] = W_{f}[C_{x}(f,t) - C_{e}(f,k)]W_{f}^{H}$$
 (2.82)

Then, the cost function will be as follows:

$$J(W_{f}, \hat{C}_{y}, \hat{C}_{e}) = \sum_{k} \| [C_{x}(f, t) - C_{e}(f, k)] \|_{F}^{2}$$
(2.83)

According to analysis in [62], one can estimate the un-mixing matrix W_f using the gradient of the above cost function in terms of W_f , C_y , and C_e . Parra proposed in his paper two methods to recover the source signals y(f, t). These methods were a least squares and a Maximum Likelihood estimator. Wang [3] addressed a cyclostationary convolutive mixture and proposed a new algorithm by combining the fourth and second order statistics of the data to enhance the performance.

2.3.2.3.5 Recursive Convolutive ICA

The RR-ICA is proposed by F. Nesta et al. [130]. It is based on frequency domain approach to separate the source signals from the short data sets in high reverberation. The RR-ICA is used to speed up the convergence and make it more robust to outlets. The main drawback in this algorithm is its sensitivity to outlets in the real world recording, refer to chapter 4.

2.4 Ambiguities in ICA algorithms

In general, there are some ambiguities described in all ICA methods, as follows: <u>Scale ambiguity:</u> one can't identify the energies or the variances of the independent components. Since, both of mixing matrix A and source signals s are unknown, then any scalar multiplication on A or s will be lost in the de-mixing process.



Figure 2.5: Illustration of permutation ambiguity in frequency domain.

<u>*Permutation ambiguity:*</u> one can't identify the order of the independent components. The mathematical model of the ambiguities of the ICA model can be expressed as follows:

$$x = As = (ADP)(D^{-1}P^{-1}s)$$
 (2.84)

Recall the performance matrix

$$G = WA \approx DP$$
,

Where D is any non-singular diagonal matrix which illustrates the scale ambiguity, and P is an identity matrix with permuted rows, which illustrates the permutation matrix. So, in general, ICA methods recover the source signals s from only a given observed signals x up to arbitrary scaling and permutation. However, in the instantaneous ICA case, these Ambiguities are not affective and can be ignored. But in some of convolutive ICA models, we will see that these ambiguities should be addressed especially, in some applications such as when working in frequency domain.

2.4.1 Scale ambiguity

Generally speaking, the ICA algorithms aren't able to determine the energies (variances) of the independent components. As a result, in instantaneous mixture problem, this ambiguity usually is ignored since one can normalized the source signals in order to rectify this ambiguity without any loss. However, the unmixed signals can be amplified or attenuated after the separation.

In frequency domain, ICA algorithm performs L instantaneous ICA algorithms for each frequency bin. So, scaling ambiguity has a real effect in this domain, where any arbitrary scaling change of each individual update rule will cause a spectral deformation to our observed signals. Also, if the arbitrary scales are not uniform along the frequency, this might cause changing in the signal envelope after separation. Researchers have worked to tackle this ambiguity. A method proposed to keep the un-mixing matrix normalized with unit norm was $||W_f|| = 1$, in order to remove the scale of data as in Smaragdis's paper [3], [39]. In addition, this also helps the natural gradient to fast convergence. This can be expressed as follows:

$$W_{f} = \frac{W_{f}}{\|W_{f}\|^{\frac{-1}{N}}}$$
(2.85)

Another smart idea to solve this ambiguity is by constraining the diagonal elements in the un-mixing matrix to be unity, such as $W_f^{ii} = 1$. This constraint ensures that there is no spectral deformation of the observed signals.

Scaling Ambiguity due to Minimal Distortion Principle (MDP)

For the sake of simplicity, let's assume there is no permutation $\Gamma(f) = 1$

ambiguity. Then, the estimated source signals S(f) as each frequency as follows:

$$Y(f) = W(f)X(f) \approx D(f)S(f)$$
(2.86)

Thus, the estimated signals Y(f) are scaled versions of the source signals S(f) by diagonal matrix D(f), however, after, multiplying both sides of the previous equation by $W^{-1}(f)$. It becomes as follows

$$W^{-1}(f) Y(f) \approx W^{-1}(f) D(f)S(f)$$
 (2.87)

Also, we have the

$$W(f) = D(f)H^{-1}(f)$$
(2.88)

Thus,

$$W^{-1}(f) Y(f) \approx H(f)S(f)$$
(2.89)

Under the Minimal Distortion Principle definition, the nth source is scaled with respect to the image at the nth microphone [3], [60], [129]. Therefore, the rescaled output signals are given by

$$Y^{scaled}(f) \approx diag(W^{-1}(f))Y(f)$$
(2.90)

2.4.2 Permutation ambiguity

In general, ICA algorithms suffer from the permutation problem [38], since it is unable to recover the source signals in order. Although this ambiguity usually is ignored in instantaneous mixtures especially in time domain, it has a place and a real effect in convolutive mixtures especially in frequency domain as shown in Figure 2.5. Any arbitrary permutation of the source signals along frequency axis will cause uncompleted separation among the sources. Thus, several researchers have proposed methods to impose some coupling between frequency bins to withstand the permutation along frequency.

The main cause of the permutation ambiguity is the statistically independency assumption between the frequency bins. Lee applies this assumption in time domain especially in the source model, which is the nonlinearity in time domain, thus, he never reported the permutation algorithm.

Permutation Ambiguity

Permutation Ambiguity is one of the main challenges in frequency domain for BSS. Many techniques have been proposed to cope with this ambiguity in frequency domain, but it is still an open issue [3]. Since, in this dissertation, we are choosing to develop the robust ICA algorithm in frequency domain, we pay a lot of attention to investigating this ambiguity and to developing a robust method to overcome this ambiguity.

There are three main solution groups to solve the permutation ambiguity in frequency domain as follows:

- ✓ Group based on the geometric information such as Time Direction of Arrivals (TDOA) and Direction of Arrivals (DOA) [3], [38], [72], [128].
- \checkmark Group based on the clustering-based techniques [57], [60], and [3].

In terms of performance, the first group generally performs better than the second group especially with a small data sample. But it is not optimal in a practical sense, since we don't usually have geometric information about real environment conditions. A second group performed better than first group especially when we had a large sample set of data, because they are based on the clustering-based techniques i.e. correlation, distance, etc. And, they are more robust for real world scenarios. For more details, refer to [3], [64].

2.4.3 Circularity of Fast Fourier Transform (FFT)

It has been known that the time domain signal can be transformed to frequencydomain by Fourier Transform. We are computed by the mean of the Discrete Fourier Transform over sample time blocks *F*. This approximation means that we enforce the signal to be a periodical signal with period equal to the sample frequency over the sample time block $T = \frac{f_s}{F}$. However, in [55], [50], [72] and [63] they have reported that this simplification is not a realization in sense of time domain filters. Therefore, the transfer function of this filters are unstable and having overshoots in his frequency response. For more details refer to [38], [3].

There are two solutions to mitigate the circularity effect of FFT; 1) by increasing the length of the DFT F [72], and 2) by imposing smoothed function to modify the frequency response of such a filter as in [38].

2.5 Performance Metrics of ICA methods

2.5.1 Instantaneous case

2.5.1.1 Performance Matrix G

In order to measure a performance for ICA algorithms, one can use the Performance matrix G [2] as follows:

$$G = WA \approx I$$

Ideally, our un-mixing matrix W should equal the inverse of mixing matrix A^{-1} . However, one would expect the matrix P to be closed to identity matrix. But, since ICA method separates the sources up to permutation and scale. Then, the performance matrix G should be a good indication measure of the quality of separation. Additionally, performance matrix G presents the relation between the permutation of the original sources and the estimated ones.

2.5.1.2 SNR measure:

In practice, one can use the signal to noise ratio (SNR) as a separation quality measurement [2] as follows:

$$SNR = 10\log\left[\frac{\sum_{i} s^{2}(i)}{\sum_{i} (s(i) - y(i))^{2}}\right]$$
(2.91)

In other words, it shows the comparison between the energies of an original signal and estimated signals. Notably, to use this metric, we should compare the signals with the same variance and polarity, since ICA method separates the sources up to a permutation and scale.

2.5.2 Convolutive case

2.5.2.1 Performance Index:

From a statistical view, the performance index was established in [2] by employing the performance matrix G as follows:

$$PI = \sum_{i=1}^{m} \left[\sum_{j=1}^{m} \frac{|G_{ij}|}{\max_{k}|G_{ik}|} - 1 \right] + \sum_{j=1}^{m} \left[\sum_{i=1}^{m} \frac{|G_{ij}|}{\max_{k}|G_{ik}|} - 1 \right]$$
(2.92)

Obviously, for an ideal performance matrix G, this index tends to minimum (to zero). However, the larger performance index value PI is the worst performance for the algorithm.

2.5.2.2 Mutual Information measure:

Reiss et al [56] takes advantage of the mutual information as a measure of statistical independence and creates a performance index. He develops the time series method to estimate the mutual information in [71].

2.5.2.3 Performance Evaluation

From (2), the separated sources are given by

$$s_i(t) = \sum_{j=1}^{m} W_{ij} * x_j(t)$$
 (2.93)

According to [64], one can divide the power of one of the separated sources $s_i(t)$ into two portions; the first portion belongs to the source coming from the source i, p_{ii} ,

second one belongs to the crosstalk signals $s_k(t)$, p_{ik} . Therefore, one can define the output SIR as the ratio of the power of the first portion p_{ii} to the power of the second one p_{ik} as follows:

$$SIR_{i} = \frac{p_{ii}}{p_{ik}} = 10 \log \frac{\sum_{t} s_{ii}^{2}(t)}{\sum_{t} \sum_{i \neq k} s_{ik}^{2}(t)}$$
(2.94)

In this dissertation, we will calculate the SIR for source i as follows

$$SIR_{i} = \frac{p_{ii}}{p_{ik}} = 10 \log \frac{\sum_{t} \left(\sum_{j=1}^{m} W_{ij} * x_{ji}(t) \right)^{2}}{\sum_{t} \sum_{i \neq k} \left(\sum_{j=1}^{m} W_{ij} * x_{jk}(t) \right)^{2}}$$
(2.95)

We will deal with the convolve speech signals with premeasured real-word recordings or artificially generated room impulse responses (RIRs). However, we only have access to the observed signals $x_{ji}(t)$ (microphone signals) recorded when only the *ith* source is active. We will set the input SIR as a baseline, i.e. the SIR obtained without any processing. Or, we will refer to the most interesting evaluation criteria in [126], [127] to study our algorithm performance.

3 Chapter 3

Convex Cauchy–Schwarz Independent Component Analysis for Blind Source Separation

Independent Component Analysis (ICA) is a powerful tool in Blind Source Processing (BSP). We present a new high-performance Convex Cauchy–Schwarz Divergence (CCS-DIV) measure for Blind Source Separation (BSS) and unsupervised learning of acoustic and speech signals. The CCS-DIV measure is developed by integrating convex functions into the Cauchy–Schwarz inequality. By including a convexity quality parameter, the measure has a broad control range of its convexity. With this measure, a new CCS–ICA algorithm is structured and a non-parametric form is developed incorporating the Parzen window-based distribution. Furthermore, a pairwise iterative scheme is employed to tackle the high dimensional problem in BSS. We present two schemes of pairwise non-parametric ICA algorithms based on gradient decent and the Jacobi Iterative method. Several case-study scenarios are carried out on noise-free and noisy mixtures of speech and music signals. Finally, the superiority of the proposed CCS–ICA algorithm is demonstrated in metric-comparison performance with FastICA, RobustICA, convex ICA (C-ICA), and other leading existing algorithms.

3.1 Introduction

Blind Signal Processing (BSP) is one of the most challenging and emerging areas in signal processing. BSP has gained a solid theoretical foundation and numerous potential applications. BSP remains a very important and challenging area of research and development in many domains, e.g. biomedical engineering, image processing, communication system, speech enhancement, remote sensing, etc. BSP techniques do not assume full a priori knowledge about the mixing environment, source signals, etc. and do not require any training samples. BSP includes three major areas: Blind Signal Separation (BSS), Independent Component Analysis (ICA), and Multichannel Blind Deconvolution (MBD) [1], [2].

In the following, we provide a focused and brief overview. ICA is considered a key factor of BSS and unsupervised learning algorithms [1]. ICA specializes to Principal Component Analysis (PCA) and Factor Analysis (FA) in multivariate analysis and data mining, corresponding to second order methods in which the components are in the form of a Gaussian distribution [8 - 20], [1], [2]. However, ICA is a statistical technique that includes higher order statistics (HOS), where the goal is to represent a set of random variables as a linear transformation of statistically independent components.

ICA techniques are based on the assumption of non-Gaussianity and independence of the sources. Let an $M \times T$ observation vector $X = [x_1, x_2, ..., x_M]^T$ be obtained from M statistically independent sources $S = [s_1, s_2, ..., s_M]^T$ by X = AS, where A is an $M \times M$ invertible mixing matrix. The estimated sources can be modeled by Y = WXwhere W is a demixing matrix. The ICA goal is to determine a demixing matrix W to estimate the source signals. ICA uses the non-Gaussianity of sources and an independence measure to find a demixing matrix W. A measure could be based on the mutual information, Higher Order Statistic (HOS), such as the kurtosis, and Joint Approximate Diagonalization. In other words, the demixed matrix is obtained by optimizing such a contrast function.

Furthermore, the metrics of cumulants, likelihood function, negentropy, kurtosis, and mutual information have been developed to obtain a demixing matrix in different adaptations of ICA-based algorithms [1]. Comon [7] was the first to describe the fundamentals of ICA. Recently, he proposed the Robust Independent Component Analysis (R-ICA) in [11]. He used a truncated polynomial expansion rather than the output marginal probability density functions to simplify the estimation processes. In [14 – 15], the authors have presented ICA using mutual information. They constructed a formulation by minimizing the difference between the joint entropy and the marginal entropy of different sources.

The so-called convex ICA [20] is established by incorporating a convex function into a Jenson's inequality-based divergence measure. Xu et al [21] used the approximation of Kullback–Leibler (KL) divergence based on the Cauchy–Schwartz inequality. Boscolo et al. [22] established nonparametric ICA by minimizing the mutual information contrast function and by using the Parzen window distribution.

A new contrast function based on nonparametric distribution was developed by Chien and Chen [23], [24] to construct the ICA algorithm. They used the cumulative distribution function (CDF) to obtain a uniform distribution from the observation data. Moreover, Matsuyama et al. [25] proposed the alpha divergence approach. Also, the fdivergence was proposed by Csiszar et al. [4], [6], [26]. Alternate studies have presented the nonnegative matrix factorization (NMF) to solve the BSS problem [4]. They took advantage of imposing the nonnegative constraints to minimize and measure the approximation errors. The Euclidean distance and KL divergence were used as the error functions for NMF problems in [26].

In addition, the maximum-likelihood (ML) criterion [27] is another tool for BSS algorithms [27]–[29]. It is used to estimate the demixing matrix by maximizing the likelihood of the observed data. However, the ML estimator needs to know all the source distributions. Recently, in terms of divergence measure, Fujisawa et al. [28] have proposed a very robust similarity measure to outliers and they called it the Gamma divergence. In addition, the Beta divergence was proposed in [31] and investigated by others in [4].

In this chapter, we develop an effective and improved measure of dependency among the signals, and then we construct its corresponding (parametric and nonparametric) ICA algorithms. A novel family of dependency divergence is developed which we name Convex Cauchy Schwarz Divergence (CCS-DIV) -- due to its use of the Cauchy Schwarz Inequality "divergence." We develop this new measure by conjugating a convex function into the Cauchy–Schwarz inequality-based divergence measure. This new contrast function has a wide range of effective curvature since it is controlled by a convexity parameter. The corresponding convex Cauchy–Schwarz divergence ICA (CCS–ICA) employs the Parzen window density approximation to distinguish the non-Gaussian structure of source densities. We also present two effective pairwise ICA algorithms: one is based on the gradient descent and the other is based on the Jacobi optimization. The link between CCS_DIV, ED-DIV, KL-DIV and CS-DIV is also shown. The efficacy of the corresponding ICA algorithms based on the proposed CCS-DIV is verified by means of several ICA experiments. This CCS–ICA has succeeded effectively in solving the BSS of speech and music signals with and without additive (Gaussian) noise, and it has shown a high comparative performance outperforming other existing ICA-based algorithms.

The chapter is organized as follows. Section II presents a brief description of several divergence measures. Section III proposes the new convex Cauchy–Schwarz divergence measure. Section IV presents the CCS–ICA method. The comparative simulation results and conclusions are given in Section V and Section VI, respectively.

3.2 A Brief Description of Previous Divergence Measures

Divergence or their counterparts (dis)similarity measures play an important role in the areas of neural computation, pattern recognition, learning, estimation, inference, and optimization [4]. In general, they measure a quasi-distance or directed difference between two probability distributions p and q, which can also be expressed for unconstrained arrays and patterns. Divergence measures are commonly used to find a distance or difference between two *n*-dimensional probability distributions $p = (p_1, p_2, ..., p_n)$ and $q = (q_1, q_2, ..., q_n)$. However, the divergence measure is a fundamental and key factor in measuring the dependence between observed variables and creating an ICAbased procedure.

In this dissertation, we are mostly interested in distance-type measures that are separable, thus, satisfying the condition $D(p||q) = \sum_{i=1}^{n} d(p_i, q_i) \ge 0$; where the condition equals zero if and only if p = q. But they are not necessarily symmetric in the

sense that D(p||q) = D(q||p) and do not necessarily satisfy the triangular inequality $D(p||q) \le D(p||z) + D(z||q)$, for another distribution z.

Usually, the vector p corresponds to the observed data and the vector q is the estimated or expected data that are subject to constraints imposed on the assumed models. For the BSS (ICA and NMF) problem, p corresponds to the observed data matrix X and q corresponds to the estimated matrix Y = WX. Information divergence is a measure of distance between two probability curves. In other words, the distance-type measures under consideration are not necessarily a metric on the space P of all probability distributions [4].

The metric is the distance between two pdfs if the following conditions hold: (i) $D(\mathbf{p}||\mathbf{q}) = \sum_{i=1}^{n} d(p_i, q_i) \ge 0$ if and only if $\mathbf{p} = \mathbf{q}$, (ii) $D(\mathbf{p}||\mathbf{q}) = D(\mathbf{q}||\mathbf{p})$ and (iii) $D(\mathbf{p}||\mathbf{q}) \le D(\mathbf{p}||\mathbf{z}) + D(\mathbf{z}||\mathbf{q})$. Distances which are not a metric, are referred to as divergences [4]. Next, we review the most common divergence measures with onedimensional probability curves.

3.2.1 Previous Divergence Measures

Shannon theory shows the KL divergence (KL-DIV) [1], [4], which is the relative entropy between the joint distributions of two continuous variables x_1 and x_2 ($p(x_1, x_2)$) and the product of their marginal distributions ($p(x_1)p(x_2)$). KL-DIV is given by

$$D_{KL}(x_1, x_2) = H(p(x_1)) + H(p(x_2)) - H(p(x_1, x_2))$$
(3.1)

$$D_{KL}(x_1, x_2) = \iint p(x_1, x_2) \log\left(\frac{p(x_1, x_2)}{p(x_1) \cdot p(x_2)}\right) dx_1 dx_2$$
(3.2)

where $D_{KL}(x_1, x_2) \ge 0$ with equality if and only if $x_1 = x_2$. This means that they are independent of each other. Xu [21] developed Euclidean divergence (E-DIV) and Cauchy–Schwartz divergence (CS-DIV) by joining the terms of joint distributions of two variables and their product of marginal distributions into the Euclidean distance and the Cauchy–Schwartz inequality, respectively. E-DIV and CS-DIV are given respectively by

$$D_{E}(x_{1}, x_{2}) = \iint (p(x_{1}, x_{2}) - p(x_{1}) \cdot p(x_{2}))^{2} \cdot dx_{1} dx_{2}$$
(3.3)
$$D_{CS}(x_{1}, x_{2}) = \log \frac{\iint p(x_{1}, x_{2})^{2} \cdot dx_{1} dx_{2} \cdot \iint p(x_{1})^{2} \cdot p(x_{2})^{2} \cdot dx_{1} dx_{2}}{[\iint p(x_{1}, x_{2}) p(x_{1}) p(x_{2}) \cdot dx_{1} dx_{2}]^{2}}$$
(3.4)

where $D_E(x_1, x_2) \ge 0$ and $D_{CS}(x_1, x_2) \ge 0$ and the equality holds if and only if $x_1 = x_2$. At equality, the variables are independent of each other. These divergence measures are reasonable contrast functions to be used in the ICA method as novel measures of dependence. Furthermore, the alpha divergence (α -DIV) was developed by Amari et. al. [2], [4]. It can be used as a measure of dependence. α -DIV is given by:

$$D_{\alpha}(x_{1}, x_{2}, \alpha) = \iint \left[\frac{1-\alpha}{2} p(x_{1}, x_{2}) + \frac{1+\alpha}{2} p(x_{1}) \cdot p(x_{2}) - p(x_{1}, x_{2}) \right]^{\frac{1-\alpha}{2}} (p(x_{1}) \cdot p(x_{2}) - p(x_{1}, x_{2}))^{\frac{1-\alpha}{2}} (p(x_{1}) \cdot p(x_{2}))^{\frac{1-\alpha}{2}} (p(x_{1}) \cdot p(x_{2}))^{\frac{1-\alpha}{2$$

$$p(x_2)\Big)^{\frac{\alpha+1}{2}} \bigg] dx_1 dx_2.$$
 (3.5)

Matsuyama [25] introduced the alpha ICA algorithm by using α -DIV as a contrast function of the ICA method. In the case $\alpha = -1$, the α -DIV is equivalent to KL-DIV [4], [6]. Csiszár [26] introduced an interesting divergence measure that is called an f-divergence (f-DIV) and is given by

$$D_{f}(x_{1}, x_{2}) = \iint p(x_{1}, x_{2}) f\left(\frac{p(x_{1}, x_{2})}{p(x_{1}) \cdot p(x_{2})}\right) dx_{1} dx_{2}$$
(3.6)

where f(.) denotes a convex function satisfying $f(t) \ge 0$ for $t \ge 0$, and f(1) = 0, $\hat{f}(1) = 0$. In addition, Csiszár shows that the α -DIV is a special case of f-DIV when using the following convex function

$$f(t) = \frac{4}{1-\alpha^2} \left[\frac{1-\alpha}{2} + \frac{1+\alpha}{2} t - t^{\frac{1+\alpha}{2}} \right] \text{ For } t \ge 0$$
(3.7)

Furthermore, Zhang [31] developed a general divergence function by integrating the α -DIV and f-DIV functions into the following form:

$$D_{Z}(x_{1}, x_{2}) = \frac{4}{1-\alpha^{2}} \left\{ \frac{1-\alpha}{2} \iint f(p(x_{1}, x_{2})) \cdot dx_{1} dx_{2} + \frac{1+\alpha}{2} \iint f(p(x_{1}) \cdot p(x_{2})) \cdot dx_{1} dx_{2} - \int f\left(\frac{1-\alpha}{2}p(x_{1}, x_{2}) + \frac{1+\alpha}{2}p(x_{1}) \cdot p(x_{2})\right) \cdot dx_{1} dx_{2} \right\}$$

$$(3.8)$$

Lin [32] developed a Jensen–Shannon divergence (JS-DIV) by using the Shannon entropy H[.] into the Jensen's inequality; the JS_DIV is given by

$$D_{JS}(x_1, x_2) = H(\lambda p(x_1, x_2) + (1 - \lambda)p(x_1)p(x_2)) - \lambda H(p(x_1, x_2)) - (1 - \lambda)H(p(x_1)p(x_2))$$
(3.9)

where $0 \le \lambda \le 1$ represents a weighting parameter between the joint distribution and the product of their corresponding marginal distributions. $D_{JS}(x_1, x_2) \ge 0$, and the equality holds if and only if $x_1 = x_2$. Recently, Chien [20] proposed the convex divergence (C-DIV) by using the Jensen's inequality. C-DIV is developed by combining the convex function f(.) into the Jensen's inequality. C-DIV is given by

$$D_{C}(x_{1}, x_{2}, \alpha) = \frac{4}{1-\alpha^{2}} \left\{ \iint \lambda \left[\frac{1-\alpha}{2} + \frac{1+\alpha}{2} p(x_{1}, x_{2}) - p(x_{1}, x_{2})^{\frac{1+\alpha}{2}} \right] dx_{1} dx_{2} + (1-\lambda) \iint \left[\frac{1-\alpha}{2} + \frac{1+\alpha}{2} p(x_{1}) \cdot p(x_{2}) - \left(p(x_{1}) \cdot p(x_{2}) \right)^{\frac{1+\alpha}{2}} \right] dx_{1} dx_{2} - \left[\frac{1-\alpha}{2} + \frac{1+\alpha}{2} \left(\lambda p(x_{1}, x_{2}) + (1-\lambda) p(x_{1}) p(x_{2}) \right) - \left(\lambda p(x_{1}, x_{2}) + (1-\lambda) p(x_{1}) p(x_{2}) \right)^{\frac{1+\alpha}{2}} \right] \right\}$$

$$(3.10)$$

In the case $\alpha = 1$, C-DIV is equivalent to the JS-DIV. $D_C(x_1, x_2, \alpha) \ge 0$ and the equality holds if and only if $x_1 = x_2$, which means they are independent of each other.

3.2.2 The proposed Divergence Measure

While there exist a wide range of measures, performance especially in audio and speech applications still requires improvements. The quality of an improved measure should provide geometric properties for a contrast function in anticipation of a dynamic (e.g., gradient) search in a parameter space of de-mixing matrices. The motivation here is to introduce a simple measure and incorporate controllable convexity in order to control convergence to the optimal solution.

To improve the performance of the divergence measure and speed up the convergence, this chapter presents a novel divergence method that is based on conjugating the convex function into the Cauchy–Schwartz inequality. In this context, we take advantage of the convexity parameter alpha to control the convexity in the divergence function and to speed up the convergence in the ICA and NMF algorithms. Incorporating the joint distribution $(p(x_1, x_2))$ and the marginal distributions

 $(p(x_1)p(x_2))$ into the convex function f(.) in (3.7) and conjugating them to the Cauchy– Schwartz inequality yields

$$\begin{aligned} \left| \left\langle f(p(x_1, x_2)), f(p(x_1)p(x_2)) \right\rangle \right|^2 \\ &\leq \left\langle f(p(x_1, x_2)), f(p(x_1, x_2)) \right\rangle \cdot \left\langle f(p(x_1)p(x_2)), f(p(x_1)p(x_2)) \right\rangle (3.11) \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is an inner product; Now, based on the Cauchy–Schwartz inequality a new symmetric divergence measure is proposed, namely:

$$D_{CCS}(x_1, x_2, \alpha) = \log \frac{\iint f^2(p(x_1, x_2)) \cdot dx_1 dx_2 \cdot \iint f^2(p(x_1) \cdot p(x_2)) \cdot dx_1 dx_2}{[\iint f(p(x_1, x_2)) \cdot f(p(x_1) p(x_2)) \cdot dx_1 dx_2]^2}$$
(3.12)

where $D_{CCS}(x_1, x_2, \alpha) \ge 0$ and equality holds if and only if $x_1 = x_2$. This divergence function is then used to develop the ICA and NMF algorithms. Notably, the joint distribution and product of the marginal densities in $D_{CCS}(x_1, x_2, \alpha)$ is symmetric. This symmetrical property does not hold for KL-DIV, α -DIV, and f-DIV. Additionally, the CCS-DIV is tunable by the convexity parameter α .

In contrast to C-DIV and α -DIV, the convexity parameter α range is extendable. However, Based on l'Hopital's rule, one can derive the realization of CCS-DIV for the case of $\alpha = 1$ and $\alpha = -1$ by finding the derivatives, with respect to α , of the numerator and denominator for each parts of $D_{CCS}(x_1, x_2, \alpha)$. Thus, the CCS-DIV with $\alpha = 1$ and $\alpha = -1$ are respectively given by (3.13) and (3.14).

3.2.3 Link to other Divergences:

This CCS-DIV distinguishes itself from the previous divergences in the literature by incorporating the convex function into (not merely a function of) the Cauchy Shawarz inequality-- in order to guarantee convexity in the new divergence. This chapter thus develops a framework for generating a family of dependency measure based on conjugating the convex function into the Cauchy Shawarz inequality. Such convexity is anticipated (as is evidenced by experiments) to reduce local minimum near the optimal solution and enhance searching a non-linear surface of the contrast function. Also, it provides a flexibility of scalability to high dimensional data. The motivation behind this divergence is to render the CS-DIV to be convex similar to the f-DIV. For this work, we shall focus on one convex function f(t) as in (3.7), and its corresponding CCS-DIVs in (3.13) and (3.14). It can be seen that the CCS-DIV, for the $\alpha = 1$ and $\alpha = -1$ cases, is implicitly based on Shannon entropy (KL divergence) and Renyi's quadratic entropy, respectively. Also, it is to show that the CCS-DIVs for the $\alpha = 1$ and $\alpha = -1$ cases are convex functions in contrast to the CS-DIV.

$$D_{CCS}(x_1, x_2, 1) = \frac{\left(\iint \left\{ \left(p(x_1, x_2) \cdot \log(p(x_1, x_2)) - p(x_1, x_2) + 1 \right)^2 \right\} dx_1 dx_2 \right) \cdot \left(\iint \left\{ \left(p(x_1) \cdot p(x_2) \cdot \log(p(x_1) \cdot p(x_2)) - p(x_1) \cdot p(x_2) \right) - p(x_1) \cdot p(x_2) + 1 \right) \cdot \left(p(x_1) \cdot p(x_2) \cdot \log(p(x_1) \cdot p(x_2)) - p(x_1) \cdot p(x_2) + 1 \right) \right\}}{\left[\iint \left\{ \left(p(x_1, x_2) \cdot \log(p(x_1, x_2)) - p(x_1, x_2) + 1 \right) \cdot \left(p(x_1) \cdot p(x_2) \cdot \log(p(x_1) \cdot p(x_2)) - p(x_1) \cdot p(x_2) + 1 \right) \right\} \right]}$$
(3.13)

$$D_{CCS}(x_1, x_2, -1) = \log \frac{\left(\iint \left\{ \left(\log(p(x_1, x_2)) - p(x_1, x_2) + 1\right)^2 \right\} dx_1 dx_2 \right) \cdot \left(\iint \left\{ \left(\log(p(x_1) \cdot p(x_2)) - p(x_1) \cdot p(x_2) + 1\right)^2 \right\} dx_1 dx_2 \right)}{\left[\iint \left\{ \left(\log(p(x_1, x_2)) - p(x_1, x_2) + 1\right) \cdot \left(\log(p(x_1) \cdot p(x_2)) - p(x_1) \cdot p(x_2) + 1\right) \right\} dx_1 dx_2 \right]^2}$$

$$(3.14)$$

3.2.4 Geometrical Interpretation of the Proposed Divergence for $\alpha = 1$ and $\alpha = -1$.

For simplicity, let's define the following terms:

$$V_{J} = \iint (p(x_{1}, x_{2}))^{2} dx_{1} dx_{2}$$

$$V_{M} = \iint (p(x_{1})p(x_{2}))^{2} dx_{1} dx_{2}$$

$$V_{c} = \iint p(x_{1}, x_{2})p(x_{1})p(x_{2}) dx_{1} dx_{2}$$

$$V_{JJ} = \begin{cases} \iint \left\{ \begin{pmatrix} p(x_{1}, x_{2}) \cdot \log(p(x_{1}, x_{2})) \\ -p(x_{1}, x_{2}) + 1 \end{pmatrix}^{2} \right\} dx_{1} dx_{2} \quad \alpha = 1$$

$$\iint \left\{ \begin{pmatrix} \log(p(x_{1}, x_{2})) \\ -p(x_{1}, x_{2}) + 1 \end{pmatrix}^{2} \right\} dx_{1} dx_{2} \quad \alpha = -1$$

$$V_{MM} = \begin{cases} \iint \left\{ \begin{pmatrix} p(x_{1}) \cdot p(x_{2}) \cdot \log(p(x_{1}) \cdot p(x_{2})) \\ -p(x_{1}) \cdot p(x_{2}) + 1 \end{pmatrix}^{2} \right\} dx_{1} dx_{2} \quad \alpha = -1$$

$$\iint \left\{ \begin{pmatrix} \log(p(x_{1}) \cdot p(x_{2})) \\ -p(x_{1}) \cdot p(x_{2}) + 1 \end{pmatrix}^{2} \right\} dx_{1} dx_{2} \quad \alpha = -1$$

$$V_{CC} = \begin{cases} \iint \left\{ \begin{pmatrix} p(x_{1}, x_{2}) \cdot \log(p(x_{1}, x_{2})) \\ p(x_{1}) \cdot p(x_{2}) + 1 \end{pmatrix} \\ \iint \left\{ \begin{pmatrix} \log(p(x_{1}, x_{2}) \cdot \log(p(x_{1}) \cdot p(x_{2})) \\ -p(x_{1}) \cdot p(x_{2}) + 1 \end{pmatrix} \\ \iint \left\{ \begin{pmatrix} \log(p(x_{1}, x_{2}) \cdot \log(p(x_{1}) \cdot p(x_{2})) \\ -p(x_{1}) \cdot p(x_{2}) + 1 \end{pmatrix} \\ \iint \left\{ \begin{pmatrix} \log(p(x_{1}, x_{2})) \\ -p(x_{1}, x_{2}) + 1 \end{pmatrix} \\ \iint \left\{ \begin{pmatrix} \log(p(x_{1}, x_{2})) \\ -p(x_{1}, x_{2}) + 1 \end{pmatrix} \\ (\log(p(x_{1}) \cdot p(x_{2})) \\ -p(x_{1}) \cdot p(x_{2}) + 1 \end{pmatrix} \right\} dx_{1} dx_{2} \quad \alpha = -1$$

With these terms, one can express the CCS-DIV and the CS-DIV as

$$D_{CCS} = \log(V_{JJ}) + \log(V_{MM}) - 2\log(V_{CC})$$
(3.15)

$$D_{CS} = \log(V_J) + \log(V_M) - 2\log(V_C)$$
(3.16)

In Figure 3.1, we illustrate the geometrical interpretation of the proposed divergence (CCS-DIV), which is equivalent to Cauchy Schwarz Divergence (CS-DIV). Geometrically, we can show that the angle between the Joint pdfs and Marginal pdfs in the CCS-DIV is given as following:

$$\theta_{CCS} = \operatorname{acos}\left(\frac{V_{CC}}{\sqrt{V_{JJ}V_{MM}}}\right) \equiv \theta_{CS} = \operatorname{acos}\left(\frac{V_C}{\sqrt{V_JV_M}}\right)$$
(3.17)

where *acos* denotes the cosine inverse. As a matter of fact, the convex function f renders the CS-DIV a Convex contrast function for the $\alpha = 1$ and $\alpha = -1$ cases. Moreover, it provides the proposed measure an advantage over the CS-DIV in terms of speed and accuracy.

3.2.5 Evaluation of Divergence Measures

In this section, the relations among the KL-DIV, E-DIV, CS-DIV, JS-DIV, α -DIV, C-DIV, and the proposed CCS-DIV are discussed. C-DIV, α -DIV, and the proposed CCS-DIV with $\alpha = 1, \alpha = 0$ and $\alpha = -1$ are evaluated. Without loss of generality, a



Figure 3.1: Illustration of Geometrical Interpretation of the proposed Divergence simple case is considered. Two binomial variables $\{x_1, x_2\}$ in the presence of the binary events $\{A, B\}$ are considered as in [20], and [24].

The joint probabilities are $p_{x_1,x_2}(A, A)$, $p_{x_1,x_2}(A, B)$, $p_{x_1,x_2}(B, A)$ and $p_{x_1,x_2}(B, B)$, and the marginal probabilities are $p_{x_1}(A)$, $p_{x_1}(B)$, $p_{x_2}(A)$ and $p_{x_2}(B)$. Different divergence methods are tested by fixing the marginal probabilities, e.g., $p_{x_1}(A) = 0.7$, $p_{x_1}(B) = 0.3$, $p_{x_2}(A) = 0.5$ and $p_{x_2}(B) = 0.5$, and setting the joint probabilities of $p_{x_1,x_2}(A, A)$ and $p_{x_1,x_2}(B, A)$ free in the intervals (0, 0.7) and (0, 0.3), respectively.

Figure 3.2 shows the different divergence measures versus the joint probability $p_{x_1,x_2}(A, A)$. All the divergence measures reach the same minimum at $p_{x_1,x_2}(A, A) = 0.35$, which means that the two random variables are independent. Figure 3.3 shows the CCS-DIV and α -DIV at different values of α , which controls the slope of curves, respectively. Among these measures, the steepest curve is obtained by the CCS-DIV at $\alpha = -1$. Fig. 3.4 represents the CCS-DIV with different values of α : positive values more than +1 and negative values less than -1.

Notably, CCS-DIV works with any value of α and it effectively increases the slope of the "learning" curve by decreasing α ; on the contrary, C-DIV and α -DIV work only for $|\alpha| \leq 1$. Furthermore, the flattest curve is obtained by CCS-DIV with increasing α , see Figure. 3.4. This is similar to E-DIV [6] and C-DIV [20] with $\alpha = 1$. Moreover, as we have shown in Figure 3.2 and Figure 3.4, CCS-DIV with $\alpha \geq -1$ is comparatively sensitive to the probability model and obtains the minimum divergence effectively. However, CCS-DIV with $\alpha \geq -1$ should be a good choice as a contrast function for devising the ICA algorithm.

It is also worthwhile to compare and study the difference between the proposed measure and the Cauchy-Schwarz measure. Figure 3.5 shows the different divergence measures versus the joint probabilities $p_{x_1,x_2}(A, A)$ and $p_{x_1,x_2}(B, A)$. According to Figure 3.5, all the divergence measures reach the same minimum on the line $p_{x_1,x_2}(A, A) = 1.5p_{x_1,x_2}(B, A)$, which means that the two random variables become independent. One can observe that the CS-DIV is not a convex function of the pdfs in contrast to CCS-DIV from the graphs in Figure 3.5.



Figure 3.2: Different divergence measures versus the joint probability

 $\mathbf{P}_{\mathbf{x}_1,\mathbf{x}_2}(\mathbf{A},\mathbf{A})$



Figure 3.3: CCS-DIV and α -DIV versus the joint probability $P_{x_1,x_2}(A, A)$



Figure 3.4: CCS-DIV with various alphas versus the joint probability

 $\mathbf{P}_{\mathbf{x}_1,\mathbf{x}_2}(\mathbf{A},\mathbf{A})$



Figure 3.5: The surfaces and Contours of CCS-DIV vs CS-DIV

3.3 Convex Cauchy–Schwarz Divergence Independent Component Analysis (CCS–ICA)

In this section, we develop the ICA algorithm by using the CCS-DIV as a contrast function. Let us consider a simple system that is described by the vector-matrix form

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{v} \tag{3.18}$$

where $x = [x_1, ..., x_M]^T$ is a mixture vector, $s = [s_1, ..., s_M]^T$ is a source signal vector, $v = [v_1, ..., v_M]^T$ is an additive noise vector, and H is an unknown full rank $M \times M$ mixing matrix. However, to obtain a good estimate of Y = Wx of the source signals s, the contrast function CCS-DIV should be minimized. Then, the components of Y become least dependent, that is, when this demixing matrix W becomes a rescaled permutation of H^{-1} . Following the standard ICA procedure, the original data x should be preprocessed by removing the mean $\{E[x] = 0\}$ and by a weighting matrix $\{V = \Lambda^{-1/2}E^T\}$, where the matrix E represents the eigenvectors matrix and Λ the eigenvalues matrix of the autocorrelation, namely, $\{R_{xx} = E[xx^T]\}$. However, the whitening step obtained matrix (V) so that the MxT whitened data vector (X_t) has covariance of identity matrix, $\{R_{xx} = I_K\}$, which can be obtained as $\{X_t = \Lambda^{-\frac{1}{2}}V^Tx\}$. The demixing matrix can be estimated by, e.g., the gradient descent algorithm [2], [13]:

$$W(k+1) = W(k) - \gamma \frac{\partial D_{CCS}(X,W(k))}{\partial W(k)}$$
(3.19)

where k represents the iteration index and γ is a step size or a learning rate. Therefore, the updated term in the gradient descent is composed of the differentials of the CCS-DIV with respect to each element w_{ml} of the M × M demixing matrix W. The differentials $\frac{\partial D_{CCS}(X,W(k))}{\partial w_{ml}(k)}$, $1 \le m, l \le M$ are calculated using a different probability model and CCS-DIV measures as in [2], [20] and [24]. The update procedure (14) will stop when the absolute increment of the CCS-DIV measure meets a predefined threshold value. During the iterations, we should make the normalization step $w_m = \frac{w_m}{||w_m||}$ for each row of W, where ||.|| denotes a norm. Furthermore, we can use the CCS-DIV measure in the natural gradient format to increase the efficiency of the ICAbased algorithm, i.e.

$$W(k+1) = W(k) - \gamma \frac{\partial D_{CCS}(X,W(k))}{\partial W(k)} W^{T}(k)W(k)$$
(3.20)

The natural gradient KL-ICA algorithm [28] suffers from the problem of convergence to the matrix with large scaling values, especially, if the initial demixing matrix and learning rate are not carefully selected by the user. This kind of problem is too challenging and hard to overcome specifically when a highly non-linear function is presented in the KL-ICA. However, many regularization algorithms have been proposed to stabilize the KL-ICA algorithm and improve the convergence speed as in [2], [13], and [28].

In general, dealing with the indeterminacy of the scales of the demixed signals in the natural gradient form is, at most times, hard. Here, the ICA algorithm based on the CCS-DIV measure mitigates this problem by selecting an appropriate learning rate.

In setting up the CCS–ICA algorithm based on the proposed CCS-DIV measure, $D_{CCS}(x_1, x_2, \alpha)$, usually, the vector x_1 corresponds to the observed data and the vector x_2 corresponds to the estimated or expected data. Here, the CCS–ICA algorithm is established as follows.

Assuming that the demixed signals $Y_t = WX_t$ with the mth component denoted as $y_{mt} = w_m X_t$. Then, using CCS-DIV as the contrast function with built-in convexity parameter α , we get

$$D_{CCS}(Y_t, y_{mt}, \alpha) = \log \frac{\iint f^2(p(Y_t)) . dy_1 dy_2 \cdot \iint f^2(\prod_1^M p(y_{mt})) . dy_1 dy_2}{[\iint f(p(Y_t)) \cdot f(\prod_1^M p(y_{mt})) . dy_1 dy_2]^2}$$
(3.21)

We use the Lebsegue measure [5] to approximate the integral with respect to the joint distribution of $Y_t = \{y_1, y_2, ..., y_N\}$. The contrast function thus becomes

$$D_{CCS}(Y_t, y_{mt}, \alpha) = \log \frac{\sum_{1}^{T} f^2(p(WX_t)) \cdot \sum_{1}^{T} f^2(\prod_{1}^{N} (p(w_{mt}X_t)))}{[\sum_{1}^{T} f(p(WX_t)) \cdot f(\prod_{1}^{N} (p(w_{mt}X_t)))]^2}$$
(3.22)

The adaptive CCS-ICA algorithms are carried out by using the deferential of the

proposed divergence
$$\left(\frac{\partial D_{CCS}(Y_t, y_{mt}, \alpha)}{\partial w_{ml}} \right)$$
 which is derived in Appendix A. Note

that the derivative of determinant demixing matrix (det (W)) with respect to element (w_{ml}) equals the cofactor of entry (m, l) in the calculation of the determinant of W, which means $\left(\frac{\partial \det(W)}{\partial w_{ml}} = W_{ml}\right)$. And the joint distribution of the output is determined by $p(Y_t) = \frac{p(X_t)}{|\det(W)|}$ in Appendix A.

For simplicity, we can write $D_{CCS}(Y_t, y_{mt}, \alpha)$ as a function of three variables.

$$D_{CCS}(Y_t, y_{mt}, \alpha) = \log \frac{V_1 \cdot V_2}{(V_3)^2}$$
(3.23)

Then,

$$\frac{\partial D_{CCS}(Y_{t}, y_{mt}, \alpha)}{\partial w_{ml}} = \frac{V_1' V_2 + V_1 V_2' - 2V_1 V_2 V_3'}{V_1 V_2 V_3}$$
(3.24)

where

$$V_1 = \sum_{t=1}^{T} f^2(Y_t)$$
, $V'_1 = \sum_{t=1}^{T} 2f(Y_t)f'(Y_t)Y'_t$

$$\begin{split} V_2 &= \sum_{t=1}^T f^2(y_{mt}) , \quad V_2' = \sum_{t=1}^T 2f(y_{mt})f'(y_{mt})y'_{mt} \\ & V_3 = \sum_{t=1}^T f(Y_t) f(y_{mt}) , \\ V_3' &= \sum_{t=1}^T f'(Y_t)f(y_{mt})Y_t' + \sum_{t=1}^T f(Y_t)f'(y_{mt})y'_{mt} \\ & Y_t = p(WX_t) \text{and } y_2 = \prod_{n=1}^M p(w_mX_t) \\ & Y_t' &= \frac{\partial Y_t}{\partial w_{ml}} = -\frac{p(X_t)}{|\det(W)|^2} \cdot \frac{\partial \det(W)}{\partial w_{ml}} \cdot \text{sign}(\det(W), \end{split}$$
Where $\frac{\partial \det(W)}{\partial w_{ml}} = W_{ml};$

$$y'_{mt} = \frac{\partial y_{mt}}{\partial w_{ml}} = \left[\prod_{j=m}^{M} p(w_j X_t)\right] \frac{\partial p(w_n X_t)}{\partial (w_n X_t)} \cdot x_l.$$

where x_1 denotes the lth entry of X_t .

In general, the estimation accuracy of a demixing matrix in the ICA algorithm is limited by the lack of knowledge of the accurate source probability densities. However, non-parametric density is used in [1], [13], [15], [43] by applying the Parzen window estimation since it has a distribution shape that is data-driven and is flexibly formed based on the Kernel functions with a bandwidth h. In this work, a novel non-parametric CCS–ICA algorithm is also presented by minimizing the CCS-DIV to generate the demixed signals $Y = [y_1, y_2, ..., y_M]^T$.
The demixed signals are described by the following univariate and multivariate distributions [43],

$$p(y_m) = \frac{1}{Th} \sum_{t=1}^{T} \vartheta\left(\frac{y_m - y_{mt}}{h}\right)$$
(3.25)

$$p(Y) = \frac{1}{Th^{M}} \sum_{t=1}^{T} \varphi\left(\frac{Y - Y_{t}}{h}\right)$$
(3.26)

where the univariate Gaussian Kernel is

$$\vartheta(u) = (2\pi)^{-\frac{1}{2}} e^{-\frac{u^2}{2}}$$

and the multivariate Gaussian Kernel is

$$\varphi(\mathbf{u}) = (2\pi)^{-\frac{N}{2}} e^{-\mathbf{u}^T \frac{\mathbf{u}}{2}}.$$

The Gaussian kernel, used in the non-parametric ICA, is a smooth function. We note that the performance of a learning algorithm based on the non-parametric ICA is better than the performance of a learning algorithm based on the parametric ICA. By substituting (20) and (21) with $Y_t = WX_t$ and $y_{mt} = w_m x_t$ into (17), the nonparametric CCS-DIV becomes

$$D_{CCS}(Y_{t}, y_{mt}, \alpha) = \log \frac{\sum_{t=1}^{T} f^{2}(p(WX_{t})) \cdot \sum_{t=1}^{T} f^{2}\left(\prod_{1}^{M} \frac{1}{Th} \sum_{i=1}^{T} \vartheta(\frac{w_{m}(X_{t}-X_{i})}{h})\right)}{[\sum_{t=1}^{T} f(p(Wx_{t})) \cdot f(\prod_{1}^{M} \frac{1}{Th} \sum_{i=1}^{T} \vartheta(\frac{w_{m}(X_{t}-X_{i})}{h}))]^{2}}$$
(3.27)

and its derivative is

$$\frac{\partial D_{CCS}(Y_t, y_{mt}, \alpha)}{\partial w_{ml}} = \frac{V_1' V_2 + V_1 V_2' - 2V_1 V_2 V_3'}{V_1 V_2 V_3}$$
(3.28)

where

$$V_1 = \sum_{t=1}^{T} f^2(Y_t)$$
, $V'_1 = \sum_{t=1}^{T} 2f(Y_t)f'(Y_t)Y'_t$

$$V_{2} = \sum_{t=1}^{T} f^{2}(y_{mt}) , \quad V_{2}' = \sum_{t=1}^{T} 2f(y_{mt})f'(y_{mt})y'_{mt}$$
$$V_{3} = \sum_{t=1}^{T} f(Y_{t}) f(y_{mt}) ,$$
$$V_{3}' = \sum_{t=1}^{T} f'(Y_{t})f(y_{mt})Y'_{t} + \sum_{t=1}^{T} f(Y_{t})f'(y_{mt})y'_{mt}$$
$$Y_{t} = p(WX_{t})$$

$$Y'_{t} = \frac{\partial Y_{t}}{\partial w_{ml}} = -\frac{p(X_{t})}{|\det(W)|^{2}} \cdot \frac{\partial \det(W)}{\partial w_{ml}} \cdot \operatorname{sign}(\det(W),$$

where
$$\frac{\partial \det(W)}{\partial w_{ml}} = W_{ml}$$
; and sign(·) is the sign function. Thus

$$y_{mt} = \prod_{m=1}^{M} \frac{1}{Th} \sum_{i=1}^{T} \vartheta \left(\frac{y_m - y_{mi}}{h} \right) = \prod_{n=1}^{N} \frac{1}{Th} \sum_{i=1}^{T} \vartheta \left(\frac{w_m(X_t - X_i)}{h} \right)$$

$$y'_{mt} = \frac{\partial y_{mt}}{\partial w_{ml}} = -\frac{1}{Th} \sum_{i=1}^{T} \vartheta \left(\frac{w_m(X_t - X_i)}{h} \right) \cdot \left(\frac{w_m(X_t - X_i)}{h} \right) \cdot \left(\frac{X_{tl} - X_{il}}{h} \right)$$

$$\cdot \left[\prod_{j \neq m}^{M} p(w_j X_t) \right].$$

where X_{tl} and X_{il} denote the lth entry of X.

<u>Remark: This non-parametric CCS_DIV might suffer from insufficient data and high computation in a high dimensional space, especially when estimating the joint distribution. In this case, the pairwise iterative scheme which is proposed in [20], [38] should be used to mitigate this potential drawback.</u>

Algorithm 3.1: ICA Based on the gradient descent

Input: $(M \times T)$ matrix of realizations X, Initial demixing matrix $W = I_M$, Max. number of iterations Itr, Step Size γ i.e. $\gamma = 0.3$, alpha α i.e. $\alpha = -0.99999$ Perform Pre-Whitening $\{X = V * X = \Lambda^{(-1/2)} E^{T} X\}$, For loop: for each I Iteration do For loop: for each t = 1, ..., TEvaluate the proposed contrast function and its derivative $\begin{pmatrix} \partial D_{ccs}(Y_t, y_{mt}, \alpha) \\ \partial w_{ml} \end{pmatrix}$ End For $Update \ de-mixing \ matrix W$ $W = W - \gamma \frac{\partial D_{ccs}(X, W)}{\partial W}$ Normalization of WCheck Convergence $\|\Delta D_c\| \le \epsilon \ i.e. \ \epsilon = 10^{-4}$ End For Output: Demixing Matrix W, estimated signals y

3.4 Scenario of two or three source signals

Generally Speaking, the non-parametric ICA algorithm suffers from insufficient data and high computation in a high dimensional space, especially when estimating the joint distribution. However, in several previous reports in the literature, e.g., [13], [16], the authors suggest applying the pairwise iterative schemes to tackle the high dimensional data problem for non-parametric ICA algorithm(s). However, there are no results indicating how the performance would hold up with the pairwise scheme, especially in terms of computational complexity and in terms of the accuracy of the non-parametric ICA algorithm. In this work, we present two effective pairwise ICA algorithms: one is based on the gradient descent and the other is based on the Jacobi optimization [16].

Without loss of generality, one can represent the demixing matrixW as a series of rotational matrices in terms of unknown angle(s) $\theta_{ij} \in [-\pi/4, \pi/4]$ between each two pair (i, j) of the observed signals. Specifically, define the pairwise rotation matrix

$$\boldsymbol{W}(\theta_{ij}) = \begin{bmatrix} \cos \theta_{ij} & -\sin \theta_{ij} \\ \sin \theta_{ij} & \cos \theta_{ij} \end{bmatrix}$$
(3.29)

The idea is to make each pair of the estimated (marginal) output "independent" as possible (minimize dependency). It was proved and pointed out by Comon in [6] that the mutual independence between the M whitened observed signals might be attained by maximize the independence between each pair of them. In this work, we present two algorithms to solve the high dimensional problem in the non-parametric scheme. First, we adopt the non-parametric algorithm based on the gradient descent into the pairwise iterative scheme of *Algorithm 3.2*.

Algorithm 3.2: ICA Based on pairwise gradient decent scheme Input: $(M \times T)$ matrix of realization X, Initial demixing matrix $W = I_M$, number of iterations Itr, Step Size γ i.e. $\gamma = 0.3$, alpha α i.e. $\alpha = -0.99999$ For loop: for each $i = 1 \dots M - 1$ For loop: for each $j = i + 1 \dots M$ Initial demixing matrix $W_2 = I_2$ While: while (true) Find W_2 from due to Algorithm 1 for each pairs of X; End While Initial rotational matrix $R = I_M$ *Update rotational matrix* $R([i \ j], [i \ j]) = W_2$ Update Demixing matrix W = R * WUpdate observation Matrix X = W * XEnd For End For

Output: Demixing matrix W and demixed sources in X

Second, we proposed a CCS-ICA algorithm based on Jacobi pairwise scheme in

Algorithm 3.3. This algorithm based on finding the rotation matrix in (3.29) that attains the

minima of CCS-DIV. So, in fact, we set up the range of thetas, such that $\theta_{ij} \in \left[-\frac{pi}{4}: \theta_g: \frac{pi}{4}\right]$, where θ_g is the grid search, for instance $\theta_g = \frac{pi}{64}$. Then for each pair (i, j) of the observation data, we find the demixing matrix W_2 , which attains the minimum of the CCS-DIV. Please refer to Algorithm 3 for more details.

Algorithm 3.3: ICA Based on pairwise Jacobi scheme

Input: $(M \times T)$ matrix of realization X, Initial demixing matrix $W = I_M$, number of iterations Itr, Step Size γ i.e. $\gamma = 0.3$, alpha α i.e. $\alpha = -0.99999$ **Perform Pre-Whitening** $\{X = V * X = \Lambda^{\wedge}((-1) / 2) E^{\wedge}T X\},\$ For loop: for each $i = 1 \dots M - 1$ For loop: for each $j = i + 1 \dots M$ For loop: For each $\theta_1 = -\frac{pi}{4} : \frac{pi}{64} : \frac{pi}{4}$ $W_2 = \begin{bmatrix} \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_1 & \cos \theta_1 \end{bmatrix}$ Evaluate $D_{c}(X([i j], :), W_{2} * X([i j], :), \alpha)$ For all t = 1, ..., T. End For Find $W_2 = min_{W_2} D_c(X(i:j,:), W_2 * X, \alpha)$ *Initial rotational matrix* $R = I_M$ Update rotational matrix $R([i \ j], [i \ j]) = W_2$ *Update Demixing matrix* W = R * WUpdate observation Matrix X = W * XEnd For End For Output: Demixing matrix W and estimated Sources in X

3.5 Computational Complexity

Given T realizations of M observation signals, the computational complexity of the proposed algorithms rely on T and the number of observation signals M, and approximately is given by $0\left(\frac{M(M-1)}{2}T^2\right)$. The computational complexity has been a measure of merit for ICA algorithms. With the advent of Graphics Processing Units (GPUs) (see Nvidia.com, e.g.), and more powerful computing platforms, performance accuracy holds more merit. In our comparison among the ICA algorithms, we employ several metrics including computational load time and accuracy. In this work, we employ the adaptive sampling technique that produces improved performance in terms of accuracy and computational load together. The presented technique samples the signal into small time blocks in order to evaluate the integration of the proposed divergence and reduce the computational complexity. Thus, we have introduced sampling factor T s to evaluate the proposed divergence at each T_s instance. Therefore, the computational complexity of the proposed algorithm is reduced by the square of the sample factor Ts to be less than $O\left(\frac{M(M-1)}{2}\left(\frac{T}{T_s}\right)^2\right)$. Namely, we quantize the specific area of integration of the proposed divergence into equal $\left(\frac{T}{T_s}\right)$ segments to evaluate the proposed divergence.

3.6 Simulation Results

Several experimental results are conducted to compare the performance of different ICA-based algorithms. This work provides results that have a diversity of experimental data and conditions.

3.6.1 Sensitivity of CCS-DIV measure

This experiment evaluates the proposed CCS-DIV divergence measure in relation to the sensitivity of the probability model of the discrete variables. Results indicate that the CCS-DIV with $\alpha = 1$ and $\alpha = -1$ successfully reaches the minimum point of the measure. Let us consider the case as in [20], [34], [35], where the mixed signals X = AS, to investigate the sensitivity of CCS-DIV with $\alpha = 1$ and $\alpha = -1$, respectively. Simulated experiments in [20], [35] were performed for two sources (M = 2) and with a demixing matrix W

$$W = \begin{bmatrix} \cos \theta_1 & \sin \theta_1 \\ \cos \theta_2 & \sin \theta_2 \end{bmatrix}$$
(3.24)

where W, in this case, is a parametrized matrix that establishes a polar coordinate system. The row vectors in W have unit norms and provide the counterclockwise rotation of θ_1 and θ_2 , respectively. The orthogonal rows in W holds the relationship between θ_1 and θ_2 , which is $\theta_2 = \theta_1 \pm \frac{\pi}{2}$. Notably, the amplitude should not affect the independent sources. By varying θ_1 and θ_2 , we get different demixing matrices. However, consider the simple case, i.e., mixtures of signals of two zero mean continuous variables; one variable is of a sub-Gaussian distribution and the other variable is of a super-Gaussian distribution. For the sub-Gaussian distribution, we use the uniform distribution

$$p(s_1) = \begin{cases} \frac{1}{2\tau_1} & s_1(-\tau_1,\tau_1) \\ 0 & \text{Otherwise} \end{cases}$$

$$(3.25)$$

and for the super-Gaussian distribution, we use the Laplacian distribution

$$p(s_2) = \frac{1}{2\tau_2} \exp\left[-\frac{|s_2|}{\tau_2}\right]$$
(3.26)

In this task, data samples T = 1000 are selected and randomly generated by using $\tau_1 = 3$ and $\tau_2 = 1$. Kurtosis for the two signals are -1.2, and 2.99, respectively, and they are evaluated using Kurt(s) = $\frac{E[s^4]}{E[s^2]} - 3$.

Without loss of generality, we take the mixing matrix as the 2 × 2 identity matrix, thus, $x_1 = s_1$ and $x_2 = s_2$ [6], [20], [25]. The normalized divergence measures of the demixing signals and their sensitivity to the variation of the demixing matrix is shown in Figure 3.6. As shown in Figure 3.6, the variations of the demixing matrix are represented by the polar systems $\theta 1$ and $\theta 2$. A wide variety of demixing matrices are considered by taking the interval of angles { θ_1 and θ_2 } from 0 to π . Furthermore, Fig. 3.6 evaluates the CCS-DIV along with E-DIV, KL-DIV, and C-DIV with $\alpha = 1$ and $\alpha = -1$. The minimum (i.e., zero) divergence is achieved at the same conditions { $\theta_1 = 0$ and $\theta_2 = \frac{\pi}{2}$ } as is clearly seen.

In addition, no local minima are found. Clearly, the values of CCS-DIV with $\alpha = 1$ are low and flat within the range of $\theta 2$ between 0.5 and 2.5. This performance is similar to other divergence measures as in [20], [25]. Contrarily, the values of CCS-DIV with $\alpha = -1$ enable a relatively more convex form in the same range. Thus, the CCS-DIV with $\alpha = -1$ leads through the steepest descent to the minimum point of the CCS-DIV measure. Observe that the CCS-DIV with $\alpha = 1$ has a flat curve with respect to θ_1 and θ_2 . For other α values, the CCS-DIVas a contrast function, can produce large decremental steps of the demixing matrix towards convergence to the solution. And again, one can observe that the CS-DIV is not a convex function in contrast to CCS-DIV from the graphs in Figure 3.6.



Figure 3.6: Comparison of (a) CCS-DIV with $\alpha = 1$, (b) CCS-DIV with $\alpha = -1$, (c) KL-DIV, (d) E-DIV, (e) CS-DIV (f) C-DIV with $\alpha = -1$ of demixed signals as a function of the demixing parameters θ_1 and θ_2 .

3.6.2 The performance and the convergence speed of the proposed CCS-ICA algorithms versus the existing ICA-based algorithms

In this section, Monte Carlo Simulations are carried out. It is assumed that the number of sources is equal to the number of observations "sensors". All algorithms have used the same whitening method. The experiments have been carried out using the MATLAB software on an Intel Core i5 CPU 2.4-GHz processor and 4G MB RAM. Each entry corresponds to the average of corresponding trial "independent Monte Carlo" runs in which the mixing matrix is randomly chosen.

First, we compare the performance and convergence speed of the gradient descent ICA algorithms based on the CCS-DIV, CS-DIV, E-DIV, KL-DIV, and C-DIV with $\alpha = 1$ and $\alpha = -1$. In all tasks, the standard gradient descent method is used to devise the parameterized and non-parameterized ICA algorithms based on CCS-DIV with $\gamma=0.7$ and $\gamma=0.3$ for $\alpha=1$ and $\alpha=-1$ cases, respectively, CS-DIV with $\gamma=0.3$, E-DIV with γ =0.06, KL-DIV γ =0.17 as in [14], and C-DIV with γ =0.008 and γ =0.1 for α =-1 and α =1 cases, respectively as in [13]. During the comparison, we use the bandwidth as a function of sample size, namely, $h = 1.06T^{\frac{-1}{5}}$ [13-15]. To study the parametric scenario for the ICA algorithms, we use mixed signals that consist of two signal sources with a mixing matrix $A = \begin{bmatrix} 0.5 & 0.6 \end{bmatrix}^T \begin{bmatrix} 0.3 & 0.4 \end{bmatrix}^T$, which has a determinant det(A) = 0.02. One of the signal sources has a uniform distribution (sub-Gaussian) and the other has a Laplacian distribution with kurtosis values -1.2109 and 3.0839, respectively. T = 1000 sampled data are taken using a learning rate $\gamma = 0.3$ and for 250 iterations. The gradient descent ICA algorithms based on the CCS-DIV, CS-DIV, E-DIV, KL-DIV, and C-DIV with $\alpha = 1$ and $\alpha = -1$, respectively, are implemented to recover the estimated source signals.

The initial demixed matrix W is taken as an identity matrix. Fig. 3.7 shows the demixed signals resulting from the application of the various ICA-based algorithms. Clearly, the parameterized CCS-ICA algorithm outperforms all other ICA algorithms in this scenario with signal to interference ratio (SIR) of 41.9 dB and 32 dB, respectively. Additionally, Fig. 3.8 shows the "learning curves" of the parameterized CCS–ICA algorithm with $\alpha =$ 1 and $\alpha = -1$ when compared to the other ICA algorithms, as it graphs the DIV measures versus the iterations (in epochs). As shown in Fig. 3.8, the speed convergence of the CCS-ICA algorithm is comparable to the C-ICA and KL-ICA algorithms. Furthermore, Table 3.1 and 3.2 summarize the performance of the proposed nonparametric ICA algorithms with $\alpha = -1$ against other several algorithms, i.e. CS-DIV, E-DIV, KL-DIV, C-DIV with $\alpha = -1$ and IK-DIV in terms of accuracy and computational complexity, respectively. CCS2 and CCS3 represent Algorithm 2 and Algorithm 3, respectively. We also compare it with other benchmark algorithms such as FastICA [8], RobustICA [7], JADE [11] and RapidICA [42]. For these methods, the default setting parameters are used according to their toolboxes and their publications. In this task, we have examined the aforementioned ICA algorithms to separate mixtures of two sub-Gaussians, two sup-Gaussians, and both sub and sup-Gaussian signals. We use the following distributions: For the sub-Gaussian distribution, we use the uniform distribution

$$p(s_1) = \begin{cases} \frac{1}{2\tau_1} & s_1 in(-\tau_1, \tau_1) \\ 0 & \text{Otherwise} \end{cases}$$
(3.27)

and the Rayleigh distribution, we use the following

$$p(s_2) = s_2 \exp\left[-\frac{s_2^2}{2}\right]$$
 (3.28)

For the super-Gaussian distribution, we use the Laplacian distribution

$$p(s_3) = \frac{1}{2\tau_2} \exp\left[-\frac{|s_3|}{\tau_2}\right]$$
(3.29)

and log-normal distribution, we use the following

$$p(s_4) = \exp\left[-\frac{(\log s_4)^2}{2}\right]$$
(3.30)

Also, data samples, T = 1000, are selected and randomly generated by using $\tau_1 = 3$ and $\tau_2 = 1$. Kurtoses for all aforementioned signals are -1.2, 2.99, -0.7224, and 8.4559 respectively, and they are evaluated using Kurt(s) = $E[s^4] / (E[s^2])^2 - 3$.

One can observe several patterns from *Tables 3.1, 3.2* and *3.3*. The presented algorithms based on the proposed measure show the best performance in terms of accuracy (in most cases) and stability. The proposed algorithm CCS3 exhibits the comparable behavior in terms of speed and stability with KL and ED. Clearly, the proposed divergence improves the CS-DIV in terms of stability and performance. Notably, most the presented divergences struggle to separate the Rayleigh distributions (s_2, s_2) (including the KL-DIV) except the proposed divergence and C-DIVs. Moreover, *Table 3.3* verifies our point in this letter, thanks to the convexity; the stability of the proposed algorithm outperforms the CS-DIV and makes the divergence more robust against variation of parameters.

Also, it is obvious that the non-parametric methods perform betters in terms of performance and stability than the non-Gaussian methods such as JADE, FastICA and other algorithms. Nevertheless JADE performs better than each of FastICA, RobustICA and Rapid ICA in terms of accuracy in some cases, but in terms of speed, we find that these later algorithms outperform the JADE algorithm, especially the rapid ICA and Robust ICA. However, Table 3.5 summarizes the performance of the aforementioned

algorithms in a more complex separation process. A different, randomly generated source signals (refer to Table 3.4) and mixing matrices are employed. As a result, Table 3.4 summarizes the performance of each algorithm in terms of the standard error metric (multiplied $\times 100$), see [1], [13]. All results have been averaged over a number of independent Monte Carlo runs. Table 3.4 demonstrates again that the non-parametric ICA based on the proposed divergence provides the best performance in terms of accuracy (in most cases). However, in terms of speed, RapidICA, FastICA, RobustICA and JADE perform better. So, these algorithms could be chosen to initialize for methods of higher performance in order to reduce the overall computational load.

Since, the comparison between the ICA algorithms has relied on two criteria, namely, accuracy and computational load, a tradeoff between these two criteria has always been assessed for each targeted application. We also note that with the advent of Graphics Processing Units (GPUs), computational load/speed becomes less of a factor, and the true metric becomes accuracy. **Table 3.6** summarizes the performance of CCS-ICA (see **Algorithm 3.3**) based on the different values of T_s (1,10,100,1000), and **Table 3.7** shows their corresponding computational load in seconds. Based on these results, one observes that the best performance of the CCS-ICA, **Algorithm 3.3** scheme, in terms of accuracy and speed occurs with $T_s = 100$. For brevity, Readers can get more results of non-parametric of CCS-ICA algorithm at http://www.egr.msu.edu/bsr/.

Also, to check the robustness of the proposed algorithm, we have modified the initial demixed matrix W to be random. Figure 3.9 and Figure 3.10 show the results of the SIR of the demixed signals and the learning curve of C-ICA, E-ICA, KL-ICA, and

CCS-ICA with $\alpha = 1$, and $\alpha = -1$ in a two-source BSS task with a random initial demixing matrix, respectively.

	Each entry averages over the												
		corresp	onding	number	of tria	uls. Ol	oserv	ation	mixtu	res co	onsists	of two	o source
	sig	nals tha	nt follo	w the san	ne dist	ributi	on as	deno	ted in	the c	orres	oondin	g exam
2S	Trials	FastICA	JADE	RobustICA	Rapid ICA	IK- DIV	CS- DIV	KL- ICA	ED- DIV	C- DIV+	C- DIV-	CCS DIV1+	CCS- DIV1-
	100	6.16	4.77	5.27	5.07	3.32	2.66	1.75	2.04	2.17	2.36	2.25	2.40
	100	22.34	18.51	28.29	20.26	6.78	7.39	8.13	5.12	5.38	3.83	8.92	5.80
	100	2.45	2.10	2.24	2.14	2.31	2.21	2.31	2.31	2.65	2.50	2.19	1.94
	100	3.34	3.03	3.13	3.29	1.93	2.02	1.93	1.71	2.04	1.90	1.97	1.93
	100	5.11	4.53	5.39	5.17	2.44	2.07	2.06	2.24	2.56	2.10	2.50	2.33

vergence and other widely used ICA algorithms, each entry averages over the corresponding nutrials. Observation mixtures consists of two source signals that follow the same distribution as denoted in the corresponding example.

mples	Trials	FastICA	JADE	RobustICA	Rapid	IK-	CS-	KL-	ED-	C-	C-	CCS-	CCS-	CC DU
					- ICA					DIVT	$D_1 v =$	Div2+	DIV2-	
.000	100	0.0	0.0	0.0	0.0	20.1	22.1	19.5	20.1	24.1	24,1	22.2	22.2	19
000	100	0.0	0.1	0.0	0.0	20.1	-21.3-	19.2	20,2	23.3	23.3	19,1	19.1	21
000	100	0.0	0.0	0.0	0.0	19,1	20.7	19.1	22.1	25.1	25.1	18.1	18.1	20
000	100	0.0	0.1	0.0	0.0	20.4	24.3	19	23.1	24.1	24.1	19.1	19.1	19
000	100	0.0	0.0	0.0	0.0	20.2	-20.1	20.1	22.1	21.4	21.4	18.1	18.1	19
			т	able 3.3: 1	The co	rrespo	nding	varia	nce o	f the p	erforn	nance.		
						1				1				
mples	Triale	EastICA	IADE	RobustICA	Rapid	TK.	CS.	KI I	ED.	C	C.	CCS	CCS	CCS.
mpica	111415	rastien	JADL	RODUSIICA	ICA	DIV	DIV	ICA	DIV	DIV+	DIV-	DIV2+	DIV2-	DIV3-
000	100	11.02	12.07	38.05	11.74	3.76	2.58	0.81	1.15	1.39	0.98	1.53	1.91	0.72
000	100	102.53	211.75	332.76	95.06	16.43	37.71	8.87	8.87	6.76	3.92	28.35	10.63	6.19
000	100	1,11	1.80	1.71	1.27	1.63	1.54	1.66	1.66	1.60	1.35	2.39	1.93	1.17
000	100	18.47	15.34	17.44	14.64	1.51	1.52	0.95	0.78	1.07	1.08	1.19	1.51	0.83

13.90 14.16 2.14 2.75 1.63 1.72 2.25 1.16

These results agree with those in the previous sections. We also report the time of each epoch for using different divergence measures in their implementation. Furthermore, Figure 3.11 shows the "learning curves" of the CCS-DIV measure with several convexity parameter values in a three-source BSS task. The mixed signals are a result of the mixing matrix

 $A = \begin{bmatrix} [0.3 \ 0.2 \ 0.4]^T & [0.4 \ 0.8 \ 0.7]^T & [0.5 \ 0.6 \ 0.3]^T \end{bmatrix}$

and three Laplacian distributions with $\tau_1 = 1$, $\tau_2 = 0.5$, and $\tau_3 = 1.5$, respectively. The sampled data of each source has samples T = 1000. The values of the kurtosis of the three sources are 3.22, 3.08, and 2.57, respectively. In this task, the standard gradient descent method (3.19) is used to devise the parametrized ICA algorithms based on CCS-DIV with $\gamma = 0.7$, CS-DIV with $\gamma = 0.3$, E-DIV with $\gamma = 0.06$, KL-DIV $\gamma = 0.17$ as in [25], and C-DIV with $\gamma = 0.008$ as in [20].Clearly, the CCS-ICA with $\alpha = -1$ (as well as the C-ICA with $\alpha = -1$) attains the same convergence speed, see Figure 3.12. Moreover, Figure 3.13 depicts the SIR of the demixed signals of all the algorithms. It is obvious that the CCS-ICA with $\alpha = -1$ has a better performance when compared to all others algorithms.



Figure 3.7: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in parametric BSS task.



Figure 3.8: Comparison of learning curves of C-ICA, E-ICA, KL-ICA, and CCS-ICA with α =1, and α =-1 in a two-source BSS task.

probability	ole 3.4: Kurtosis / density functio experim	Values of the ns that used i ents	e different in the ICA
Signals' Notation	Kurtosis	Signals' Notation	Kurtosis
<i>s</i> ₁	-1.2116	\$ ₁₂	-0.65419
\$2	2.9324	\$ ₁₃	-0.33421
<i>S</i> 3	-1.3995	s ₁₄	-1.6935
s ₄	136.0108	\$15	-0.86239
S5	11.6452	S ₁₆	-0.60566
<i>s</i> ₆	4.219	\$ ₁₇	-0.75488
\$7	-1.2065	\$ ₁₈	-0.65645
\$8	3.1965	\$19	-0.81022
S9	3.4302	\$ ₂₀	-0.7692
\$10	-1.3049	\$ ₂₁	-0.27737
S ₁₁	-1.6805	s ₂₂	-0.56816

Table 3.5: The performance of the ICA algorithm based on the proposed divergence in terms of Amari error (multiplied by 100). Each entry averages over the corresponding number of trials.

Dimensions	Samples	Trials	CCS3	CCS3	CCS3	CCS3
М	1		0.1T	0.01T	0.001T	1
2	1000	1024	4.6	2.9	2.1	2
	2000	1024	3.6	2.3	1.9	1.8
	4000	1024	2.8	1.9	1.6	1.4
	8000	1024	2.2	1.6	1.1	1.2
4	1000	250	5.8	3.8	2.4	2.5
	2000	250	5	2.9	2	1.8
	4000	250	3.5	2.5	1.6	1.6
	8000	250	2.7	2.2	1.3	1.3
8	1000	100	5.6	3.8	.5	3.2
	2000	100	3.7	3.1	2.2	3
	4000	100	3.1	2.6	2.2	2.8
	8000	100	3.0	2.2	1.9	1.9
16	1000	25	20.5	15.8	8.6	5.5
	2000	25	12.6	10.1	7	5.1
	4000	25	8.6	8	4.5	4.2
	8000	25	5.8	3.9	1.9	2.9
20	1000	10	27.7	15.1	13.7	8.9
	2000	10	22.8	11.3	12	7.2
	4000	10	15.6	9	7.2	5.3
	8000	10	9.8	6.3	3	2.3

Dimensions	Samples	Trials	JADE	FastICA	RapidICA	RobustICA	CS	CDIV	KLDIV	CCS2	CCS3
2	1000	512	5.6	7.3	6.1	7.2	2.5	2.2	2.3	2.1	2
	2000	512	5.1	5.9	5.5	6	1.9	1.9	1.9	1.9	1.8
	4000	512	3.1	4.1	3.5	4.3	1.6	1.6	1.6	1.6	1.4
	8000	512	2.4	2.6	2.5	2.6	1.4	1.4	1.4	1.4	1.1
4	1000	200	8	9.7	9.1	9.8	3.1	3.1	3.1	3.1	2.5
	2000	200	5.4	7.3	6.5	7.2	2.9	2.9	2.9	2.9	1.8
	4000	200	4.2	4.2	4.1	4.3	1.4	1.4	1.4	1.4	1.6
	8000	200	2.1	2.7	2.5	2.7	1.5	1.5	1.5	1.5	1.2
8	1000	75	10.5	10.3	9.6	11.2	4.6	4.6	4.6	4.6	3.2
	2000	75	8.1	8.0	7.6	8.2	3.9	3.9	3.9	3.9	3
	4000	75	5.7	4.1	4.4	3.9	2.3	2.3	2.3	2.3	2.8
	8000	75	2.7	3.1	3.0	3.2	2	2	2	2	1.9
16	1000	15	8	9.7	9.1	9.8	8.1	8.1	8.1	8.1	5.5
	2000	15	5.4	7.3	6.5	7.2	6.9	6.9	6.9	6.9	5.1
	4000	15	4.2	4.2	4.1	4.3	5.6	5.6	5.6	5.6	4.2
	8000	15	2.1	2.7	2.5	2.7	3.6	3.6	3.6	3.6	2.9
20	1000	5	22.3	21.1	20.1	26.2	14.1	11.1	14.1	11.1	8.9
	2000	5	15.7	15.6	15.2	16.2	9.3	13.3	10.3	8.3	7.2
	4000	5	7.8	7.2	7.1	7.2	7.5	7.6	6.4	6.7	5.3
	0000	~	4.5	4 1	2.0	4.0	4 7	6.2	A /	A A	2.2

Table 3.7: The performance of the ICA algorithm based on the proposeddivergence and other widely used ICA algorithms in terms of Amari error (multipliedby 100). Each entry averages over the corresponding number of trials.

Table 3.6: The computational load, in seconds, of the ICA algorithm based on the proposed divergence and other widely used ICA algorithms, each entry averages over the corresponding number of trials.

Dimansions	Samplas T	Trials	CCS3	CCS3 At	CCS3	CCS3
Dimensions	Sumples 1	111015	ccb5	0.01T	4+	4+
141			0 1T	0.011	Al	AI 1
	1000	1024	0.11	2.0	0.0011	1
2	1000	1024	0.4	2.8	29.8	28
	2000	1024	0.5	4.8	44.8	96.4
	4000	1024	0.8	8	77.9	342.9
	8000	1024	1.5	10.6	137	1073
4	1000	250	1.8	24	218.1	237.9
	2000	250	4.3	39	344.8	630.3
	4000	250	5.9	47.9	593.4	2348.6
	8000	250	10.2	83.6	1105	7737.1
8	1000	100	19.3	128.7	1053	1174
	2000	100	31.5	201.7	1743	3347
	4000	100	46.5	266.4	3109	11705
	8000	100	74.2	241.8	5534	42115
16	1000	25	170.6	909.5	6282	4376.2
	2000	25	242.3	1171	9320	17918.3
	4000	25	305.5	1403	14717	58894.6
	8000	25	329.9	2297	25658	10483.4
20	1000	10	339	1195.7	9605	11355.2
	2000	10	427.4	1724.2	14708	27504.8
	4000	10	607.6	2398.3	23634	52536.6
	8000	10	900	3754.5	42538	97312.1

3.6.3 Experiments on Speech and Music Signals

Two experiments are presented in this section to evaluate the CCS–ICA algorithm. Both experiments are carried out involving speech and music signals under different conditions. The source signals are two speech signals of different male speakers and a music signal. The first experiment is to separate three source signals from their mixtures given by X = AS where the 3 x 3 mixing matrix

$$A = [[0.8 \ 0.3 \ -0.3]^{T} \ [0.2 \ -0.8 \ 0.7]^{T} \ [0.3 \ 0.2 \ 0.3]^{T}].$$

The three speech signals are sampled from the ICA '99 conference BSS test sets at http://sound.media.mit.edu/ica-bench/ [24], [66] with an 8 kHz sampling rate. The non-parametrized CCS–ICA algorithms (as well as the other algorithms) with $\alpha = 1$ and $\alpha = -1$ are applied to this task. The resulting waveforms are acquired and the signal to interference ratio (SIR) of each estimated source is calculated. We use the following to calculate the SIR:

Given the source signals $S = \{s_1, s_2, ..., s_M\}$ and demixed signals $Y = \{y_1, y_2, ..., y_M\}$, the SIR in decibels is calculated by

SIR (dB) =
$$10 \log \frac{\sum_{t=1}^{M} ||s_t||^2}{\sum_{t=1}^{M} ||y_t - s_t||^2}$$
 (3.31)

The summary results are depicted in Figure 3.14. In addition, Figure 3.14 shows the SIRs for the other algorithms, namely, JADE, Fast ICA, Robust ICA, KL-ICA and C-ICA with $\alpha = 1$ and $\alpha = -1$. As shown in Figure 3.14, the proposed CCS–ICA algorithm achieves significant improvements in terms of SIRs. As shown in the previous figures also, the proposed algorithm has consistency and obtains the best performance among the host of algorithms

Moreover, a second experiment is conducted to examine the comparative performance in the presence of additive noise. We now consider the model x = As + vthat contains the same source signals with additive noise and with a different mixing matrix

$$A = \begin{bmatrix} 0.8 & 0.3 & -0.3 \end{bmatrix}^{T} \begin{bmatrix} 0.2 & -0.8 & 0.7 \end{bmatrix}^{T} \begin{bmatrix} 0.3 & 0.2 & 0.3 \end{bmatrix}^{T}$$

The noise v is an M x T vector with zero mean and $\sigma^2 I$ covariance matrix. In addition, it is independent from the source signals. Figure 3.15 shows the separated source signals in the noisy BSS model with SNR = 20 dB. In comparison, Fig. 3.16 presents the SNRs of all the other algorithms. Clearly, the proposed algorithm has the best performance when compared to others even though its performance decreased in the noisy BSS model. Notably, the SNRs of JADE, Fast ICA and Robust ICA were very low as they rely on the criterion of non-Gaussianity, which is unreliable in the Gaussian-noise environment. In contrast, C-ICA, KL-ICA, and the proposed algorithm, which are based on different mutual information measures, achieved reasonable results. We note that one can also conduct and use the CCS-DIV to recover the source signals from the convolutive mixtures in the frequency domain as in [3], [38].



Figure 3.9: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in parametric BSS task-- random initial value.



Figure 3.10: Comparison of learning curves of C-ICA, E-ICA, KL-ICA, and CCS-ICA with α =1, and α =-1 in a two-source BSS task with random initial value.



Figure 3.11: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in parametric BSS task.



Figure 3.12: Comparison of learning curves of C-ICA, E-ICA, KL-ICA, and CCS-ICA with α =1, and α =-1 in a three-source BSS task.



Figure 3.13: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in instantaneous BSS task.



Figure 3.14: The original signals and de-mixed signals by using CCS-ICA algorithm in instantaneous BSS task with additive Gaussian noise.



Figure 3.15: Comparison of SIRs (dB) of demixed two speeches and music signals by using different ICA algorithms in instantaneous BSS task with additive Gaussian noise.

3.7 Conclusion

A novel divergence measure is presented based on integrating convex functions into the Cauchy–Schwarz inequality. This divergence measure is used as a contrast function to develop new ICA algorithms to solve the Blind Source Separation (BSS) problem. The CCS-DIV derived algorithms can be controlled to attain the steepest descent towards the minimum value. Also, a pairwise iterative scheme is employed to address the high dimensional problem in BSS. Two schemes of pairwise non-parametric ICA algorithms are developed based on the proposed divergence. Several examples and experiments are carried out to show the improved performance of the proposed divergence. Furthermore, this chapter compares the metric performance with a host of leading ICA algorithms. We have developed also nonparametric CCS–ICA approaches to demixing where the source signals are estimated by the Parzen Window density. The conver¬gence speed of the parameterized CCS–ICA procedure is evaluated and compared to other algorithms. The proposed CCS–ICA algorithms attained the highest SIR in separation of speech and music signals relative to other leading ICA-based algorithms.

4 Chapter 4

A RobustICA-Based Algorithm for Blind Separation of Convolutive Mixtures

We propose a frequency-domain method based on robust independent component analysis (RICA) to address the multichannel Blind Source Separation (BSS) problem of the convolutive speech mixtures in highly reverberant environments. We impose regularization processes to tackle the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation in frequency domain methods. We apply an algorithm to separate the source signals in adverse conditions, i.e. high reverberation conditions when short observation signals are available. Furthermore, we study the impact of several parameters on the performance of separation, e.g. overlapping ratio and window type in the frequency domain method. We also compare different techniques to solve the permutation ambiguity. Through simulations and real-world experiments, we verify the superiority of the presented algorithm among other BSS algorithms, i.e. recursive regularized ICA (RR-ICA), independent vector analysis (IVA) and others.

4.1 Introduction

Blind Source Separation (BSS) has a solid theoretical foundation and many potential applications. In fact, BSS has remained a very important topic of research and development for a long time in many areas, such as biomedical engineering, image processing, communication systems, speech enhancement, remote sensing, etc. BSS techniques do not require any prior knowledge about a mixing matrix or source signals and do not require any training data [1], [2].

Independent Component Analysis (ICA) is a powerful tool in BSS and Multichannel Blind Deconvolution (MBD). ICA is a key factor of BSS and unsupervised learning algorithms. ICA is related to Principle Component Analysis (PCA) and Factor Analysis (FA) in multivariate analysis and data mining. This is especially the case when corresponding to second order methods in which the components or factors are in the form of a Gaussian distribution [1], [3], [6]. However, ICA is a statistical technique that includes higher order statistics (HOS), where the goal is to represent a set of random variables as a linear transformation of statistically independent components [1]. ICA methods usually assume certain properties on the sources or mixing system in order to exploit a separation criterion which imposes the same properties on their estimates.

In ICA of speech signals, several approaches have been proposed in a simple case of instantaneous linear mixtures [12-18]. However, the convolutive linear mixtures are considered more suitable in real-world applications [1-3]. Several convolutive ICA approaches have been proposed for time domain [3], [4], and frequency domain [35]-[42] methods. Also, refer to [3], [38] for more details of existing convolutive ICA methods.

In speech signals, one can exploit the inherent non-stationary attribute of natural speech signals by using the second order statistics (SOS) method [2]. Mixing environments are considered to be stationary environments and even on a short period, one can exploit the Higher order statistics e.g. Joint Approximation Diagonalization

117

(JAD) problem as in [16], [17]. According to [42], [143], online BSS algorithms can be adapted in time domain under non-stationary conditions. The time domain approach suffers from slow convergence, lack of stability and high computational complexity.

Alternatively, a block on-line frequency domain BSS algorithm is proposed in [38]. Then, one can apply the separation processes on individual blocks of the input data over time. Furthermore, one can assume that the mixing environment is stationary on short time windows. This means that the source signals don't change their location during this interval of time. This requires choosing the right time frame to grantee that the separation algorithms are accurate enough with this given observed data within this window. For more details, refer to [133], where there is a recent ICA algorithm based on the time domain framework for the short mixtures.

The recursive regularized ICA [130] algorithms proposed allow estimating a large number of demixing matrix even with a short amount of data. Despite the good performance of the aforementioned algorithm, it is considered to be under a semi blind category since it is based on prior knowledge about the acoustic source signals, i.e.: the acoustic propagation and the spectral characteristic of the source signals. In [37], [50], they studied the relationship between the number of frames of the STFT analysis and the BSS algorithms based on frequency framework. They carry out that the BSS algorithms in frequency domain are significantly affected by the number of the mixing matrices. Also in [37], [50], they proposed the method that applying the ICA adaptation to a group of frequencies in order to leave the size of the STFT large enough to achieve accurate separation processes. However, this method assumes that the acoustic propagation approximated is based on an anechoic model, i.e.: as the DRR decreases. However, there

are several drawbacks for separating the acoustic sources based on frequency domain methods [130]. First of all, when we have a high reverberation environment, this enforces us to increase the number of demixing matrices to ensure an efficient estimation for the source signals. However, this requirement is not easy to satisfy especially if we have short observation signals of the source signals. Therefore, inspired by the works of V. Zarzoso,P. Comon [11], this chapter considers several challenges for the convolutive mixtures in the frequency domain in order to carry out the RobustICA based algorithm in frequency domain. We can summarize these challenges as follows.

- Increasing the immunity of the BSS algorithm towards the outlets, e.g. signals' length, additive noise, reverberation time and source moving etc.
- Implementing should be optimized to be suitable for the real-time operation [42] in order to make the real-time DSP processor handle the computational cost without interruptions or distortions.
- *Effectively treating the scaling and permutation problems in the frequency domain.*
- Reducing the computational complexity of the ICA algorithms based on the frequency framework.
- Controlling the accuracy of the ICA algorithm especially when short mixtures are available and the demixing matrices are not constrained by any anechoic model.

The remainder of the chapter is organized as follows: Section II, a brief description of convolutive mixture and problem statement. Section III reviews the Recursively Regularized ICA. Section IV presents the RobustICA-Based method in frequency domain. In Section IV, we perform solving the ambiguities in ICA algorithm based on frequency domain. The comparative experiments results and conclusions are given in Section V and Section VI, respectively.

4.2 Convolutive Mixtures

A convolutive mixture can be considered as a natural extension of the instantaneous BSS problem. Assume an m-dimensional vector of received discrete time signals $x(k) = [x_1(k), x_2(k), ..., x_m(k)]^T$ at time k is to be produced from an n-dimensional vector of source signals $s(k) = [s_1(k), s_2(k), ..., s_m(k)]^T$, where $m \ge n$, by using a stable mixture model [2]:

$$x(k) = \sum_{p=-\infty}^{\infty} H_p s(k-p) = H_p * s(k),$$

with $\sum_{-\infty}^{\infty} ||H_p|| \le \infty$ (4.1)

Where * represents the linear convolution operator and H_p is an (m x n) matrix of mixing coefficients at time-lag p.

4.2.1 **Problem Definition**

Assume that elements h_{jip} denote the coefficients of the Finite Impulse Response (FIR) filter H_p , and L is the maximum unknown channel length. Then, the noise-free convolutive model is written as follows:

$$x(k) = \sum_{p=0}^{L-1} H_p s(k-p)$$
(4.2)

Thus, one can find an approximate inverse channel matrix W_p in order to recover the source signals $s(k) = [s_1(k), s_2(k), ..., s_m(k)]^T$ such that

$$y(k) = W_q * x(k) = \sum_{q=-0}^{Q-1} W_q x(k-q) = \hat{s}(k)$$
(4.3)

where Q is the length of the inverse of the channel impulse response. However, there are two approaches to solve this problem and recover the source signals. In time domain approaches, they have several general drawbacks such as Q should be selected at least equal to the unknown true channel L. Therefore, for a long mixing filter, which means long transfer functions, the computation will be too expensive [2], [3]. Also, using the IIR filter instead of long FIR filter to overcome this problem really suffers from the instability and will need to invert the non-minimum phase filters [2], [3], [133]. Moreover, time approaches are sensitive to channel order mismatch [3]. However, time domain methods are suitable and very efficient for small mixing filters such as in communication channel [2], [36]. With all these limitations, we focus our study on frequency approaches to solve the cocktail party problem. The main advantage of a frequency domain BSS approach is the ability to apply the set of any instantaneous ICA algorithms to solve the convolutive BSS problem. On other hand, the main challenge of BSS in the frequency domain is to deal with the permutation and scaling ambiguities, see [3], [38] for a recent survey. However, one can re-map the aforementioned BSS models into frequency domain by applying the Discrete Fourier Transform (DFT) on the observed signals x(k) in order to transform it to the instantaneous mixtures problem as follows:

$$\mathbf{x}(\mathbf{k}) = \mathbf{H} * \mathbf{s}(\mathbf{t}) \ \mathbf{x}(\mathbf{q}, \mathbf{w}) \approx \mathbf{H}(\mathbf{w})\mathbf{s}(\mathbf{q}, \mathbf{w}) \tag{4.3}$$

where w is a frequency index, q is a frame index, $s(q, w) = [s_1(q, w), ..., s_m(q, w)]^T$ and $x(q, w) = [x_1(q, w), ..., x_n(q, w)]^T$.

In the previous equation, it is considered to be valid only for periodic signals s(t). However, it is approximately valid if the time-convolution is circular. Therefore, to ensure that the time convolution is circular [1], it requires making the Fourier Transform length significantly larger than the maximum length of the mixing channels L [6]. In [38], [130], they imposed the spectral smoothing approach in order to mitigate the circularity effect in frequency domain BSS methods. In practice, to avoid the convergence into local minima during the separation processes, one can separate the observed signal at each frequency bins. Thus, the sampled observed signals $x_s(t)$ are sampled at the discrete time instant n_s using the sampling frequency f_s . And then transforming it into time-frequency domain $x_s(q, w)$ using the short time frequency transform (STFT) applied to T overlapped samples of the observed signals. However, one can express the time-frequency of the nth sensor at t frame as follows

$$x_{s}(q,w) = \sum_{n_{s}} x_{s}(n_{s}) win\left(\frac{n_{s}-q.shift}{f_{s}}\right) e^{-j2\pi w \frac{n_{s}}{f_{s}}} \quad \forall w = \frac{t}{T} f_{s}, q \in [0, ..., T-1] \quad (4.4)$$

Where win(\cdot)denotes the windowing function, here, we usually use the Hanning window since it is typically for acoustic signals. The Hanning window [6] is given by

$$win(n_s) = \frac{1}{2} \left(1 + \cos\left(\frac{2\pi n_s}{T}\right) \right)$$
(4.6)

In a real-world scenario, we use the reverberation time T_{60} to approximately define the length of the impulse response, since the impulse response functions h(t) are theoretically being infinite. The reverberation time T_{60} is the required time that reduces the energy of sounds into 60 dB where the sound signal becomes no longer active or "dies away". Therefore, the convolutive ICA model can be approximated into a series of the instantaneous ICA model as follows:

$$\mathbf{x}(\mathbf{q}, \mathbf{w}) = \mathbf{H}(\mathbf{w})\mathbf{s}(\mathbf{q}, \mathbf{w}) \tag{4.7}$$

Where w represents the frequency bin, t denotes the time domain frame, e.g. in a short time frequency transform, x(q, w) is a column vector of the observed signals in frequency domain, s(q, w) is a column vector of the original source signals and H(w) is an M x N mixing matrix in frequency domain.

For the sake of simplicity: let us assume that the number of source signals N equals to the number of the observed signals M. Thus, by applying the ICA algorithm to the x(q, w) at each frequency bins, one can recover the estimated source signals as following

$$y(q, w) = W(w)x(q, w)$$
(4.8)

Where W(w) is the demixing matrix at w frequency bin. Also, due to the well-known symmetry property of the Fourier Transform, one can just find demixing matrices (W(w)) a half of the frequency bins $w \in (0, ..., \frac{T}{2})$, and then using the symmetry property to find the others.

4.1 Recursively Regularized ICA

The recursive regularized ICA [130] algorithms are proposed to allow estimating a large number of demixing matrix even with a short amount of data. Although this algorithm performed well, it is considered to be under a semi blind category since it is based on a prior knowledge about the acoustic source signals i.e. the acoustic propagation and the spectral characteristic of the source signals.

Naturally, BSS assumes that the source signals are usually overlapping in time. In acoustic signals, one can assume that the source signals have a sparse in timefrequency domain which means that at each time-frequency point there is one dominant energy source signal. Also, acoustic source signals have usually temporal continuity in the frequency domain.

First of all, let's recall the estimated source signals y(t, w)

$$y(t,w) = W(t)x(t,w)$$
(4.9)

Thus, the update law of the mixing matrix based on the natural gradient optimization [2] as follows:

$$y(t,w) = [H(t)]^{-1}x(t,w)$$
(4.10)

$$\Delta H(t) \leftrightarrow H(t)(I - E[\varphi(y(t, w))y(t, w)^H])$$
(4.11)

Then

$$H_{new}(t) = H(t) + \mu \Delta H(t) \tag{4.12}$$

During the updating processes, we updated all coefficients of the mixing matrix $H_{new}(t)$ due to the gradient based on the Kullback Laibur divergenceg = $E[\phi(y(t,w))y(t,w)^{H}]$.

According to [72], [130], weighting the instantaneous gradient will improve the estimation technique in previous adaptation processes. Therefore, the developed gradient expectation is given by [130] as follows:

$$\Delta \widehat{H}_{i}(k) = \widehat{H}_{i}(k) \left[I - E \left[\varphi \left(y(t, w) \right) y(t, w)^{H} \right] \right]$$

$$\Delta \widehat{H}_{i}(k) = E \left[\widehat{H}_{i}(k) (I - \varphi (y(t, w)) y(t, w)^{H}) \right]$$

$$\cong \sum_{q} \mathbb{C}(q, w) \odot \widehat{H}_{i}(k) (I - \varphi (y(t, w)) y(t, w)^{H}) \quad (4.13)$$

Where \odot is the Hadamard product (i.e., element-wise), and $\mathbb{C}(q, w)$ is a weight matrix constructed as

$$\mathbb{C}(q, w) = [C_1(q, w), C_2(q, w), \dots, C_N(q, w)]$$
(4.14)

And the generic weighting column vector $C_m(q, w)$ is defined as

$$C_m(q, w) = \frac{c_m(q, w)}{N_t} [1, 1, ..., 1]^T$$
(4.15)

Where N_t is the number of time frames on which the gradient is averaged and $c_m(q, w)$ is a weight.

4.2 The presented method Based on RobustICA framework

In this section, a new strategy is proposed, based on the RobustICA method of the kurtosis framework [11] and [38]. Here, one needs to first recall the time-frequency representation of the observed vector equation (2),

$$\mathbf{x}(\mathbf{q}, \mathbf{w}) = \mathbf{H}(\mathbf{w})\mathbf{s}(\mathbf{q}, \mathbf{w}) \tag{4.16}$$

The aim of this study is to estimate the demixing matrix W(w) from the observed vector x(q, w) under the assumption that the impulse response of all mixing filters is assumed constant during the recording. The estimated source vector is given as the following at each frequency bin

$$y(q,w) = W(w)x(q,w) \tag{4.17}$$

4.4.1 Step1: Preprocessing (Data Whitening)

In the preprocessing step, the demixing matrix W(w) are detected up to a unitary matrix U(w) using the second order statistic (SOS). This step was used to reduce the noise and to eliminate redundancy in the data at each frequency bin. The KxK covariance matrix (R) of the noise free observed signals can be expressed by

$$R(q, w) = E[x(q, w)x(q, w)^{H}] \quad \forall w = 0, ..., \frac{T}{2}$$
(4.18)

By substituting x(q, w) in (21), one gets R as follows

$$R(q,w) = H(w)E[s(q,w)s(q,w)^{H}]H(w)^{T} = H(w)H(w)^{T}$$
(4.19)

By imposing the Tikhonov regularization techniques [76] to avoid the ill-posed problem, where it is well-known that the regularization is effective way to avoid the ill-conditioned matrix, the equation (4.19) becomes as follows

$$\mathbf{R}(q, w) + \mathbf{cI} = H(w)H(w)^{\mathrm{T}}$$
(4.20)

Where I is an K x K identity matrix, and $c = m.(tr(R(q,w)) + \lambda_{max})$, it is regularization parameter with m is a positive constant and λ_{max} is a maximum eigenvalue of the estimation covariance matrix R(q, w). Note that the regularization method here just adds energy constraint to boosting the covariance matrix to be a well-conditioned matrix. Therefore, the R(q, w) + cI can be decomposed as

$$R(q, w) + cI = V(w)\Lambda(w)V^{T}(w)$$
(4.21)

where V(w) is a KxK matrix satisfying

$$V(w)V^{H}(w) = V^{H}(w)V(w) = I_{K}$$
 (4.22)

And $\Lambda(w)$ is an KxK diagonal matrix. So, from (22), the KxK matrix H(w) will be

$$H(w) = V(w)\Lambda(w)^{-\frac{1}{2}}U^{H}(w)$$
(4.23)

where U(w) is a KxK full rank unitary matrix and $UU^{H} = I_{K}$.

However, the whitening step obtained matrix V(w) so that the KxT whitened data vector Z(q,w) has covariance of identity matrix, $R_{ZZ}(q,w) = I_K$, which can be obtained as follows:

$$Z(q, w) = \Lambda^{-\frac{1}{2}} V^{T} x(q, w)$$
 (4.24)

$$Z(q, w) = U^{H}(w)s(q, w)$$
(4.25)

The estimated source signals can be recovered with a linear Zero-Forcing (ZF) equalizer. Then the estimated KxT source vector
$$y(q, w) = U(w)Z(q, w)$$
(4.26)

After the preprocessing step, the estimation of the source signals y(q, w) reduces to determining the KxK unitary matrix U(w) (rotation matrix).

4.4.2 Step 2: Determining the rotation matrix (unitary matrix) U(w).

One way of finding the rotational matrix U(w) is by maximizing the normalized fourth-order marginal cumulant (Kurtosis contrast) of the whitened data Z in (4.25). To estimate U(w) in (4.26), this chapter exploits the statistical independence of equalized source vector. More precisely, the unitary matrix U(w) will be estimated by utilizing the independent property of estimated source vector at each frequency bin (y(q, w)) in the normalized fourth-order marginal cumulant of whitened data Z(q, w)as follows:

$$K(q,w) = \frac{E[|y(q,w)|^4] - 2E^2[|y(q,w)|^2] - |E[y(q,w)^2]|^2}{E^2[|y(q,w)|^2]}$$
(4.27)

Where $E[\cdot]$ represents the expectation operator. Based on the deflation approach to ICA [97], one can extract the one of the estimated source signal as follows

$$y_i(q, w) = u_i^H(w) x(q, w)$$
 (4.28)

Where $(\cdot)^{H}$ represent the conjugate-transpose operator, $u_{i}(w)$ is the *ith* column vector of the demixing matrix U(w) and $y_{i}(q, w)$ is *ith* source signal at each wth frequency bin and qth frame time. According to [2], [3], the column vector $u_{i}(w)$ of the demixing matrix U(w) can be estimated for all users due to the batch adaptation by a gradient decent method as follows

$$u_i^{l+1} = u_i^l - \mu \Delta G_i^l \tag{4.29}$$

Where *l* denotes the iteration index, u_i^l is the *ith* column vector of the demixing matrix U(w) at *lth* iteration and ΔG_i^l is the gradient of the contrast measure that updates the demixing vector u_i^l in the demixing matrix U(w). Gradient function depends on the cost function that ICA would maximize /minimizes in order to extract the source signal. Herein, this chapter refers to use the ICA techniques based on the kurtosis criterion which is given in (4.27) as follows:

$$K(q,w) = \frac{E[|y(q,w)|^4] - 2E^2[|y(q,w)|^2] - |E[y(q,w)^2]|^2}{E^2[|y(q,w)|^2]}$$
(4.30)

Owning the RobustICA's search-method of the kurtosis criterion in (4.30) in order to choose the optimal step size as follows:

$$\mu_{\text{opt}} = \arg \max_{\mu} |\mathsf{K}(y(q, w) + \mu g(q, w))|$$
(4.31)

Where *g* is the gradient of Kurtosis contrast *K*(.). One can easily choose the optimal step size μ_{opt} based on one of the algebraic methods instead of using the exact line search as in [8], [28] to avoid the intensive computation and other limitations as in [11]. Therefore, it is easy to find the global optimum step size μ_{opt} for the criteria that can be expressed as polynomial function of μ due to its roots, e.g. the criteria kurtosis [11-16], the constant modulus [13], [82] and the constant power [2].

Therefore, RobustICA performs an optimal step-size of estimating *ith* source signal, based on optimization, for *lth* iteration, wth each frequency bin, and *qth* frame as follows:

- Step 1) An initial value for the Wight vector u(w)
- Step 2) Compute the optimal step size polynomial coefficients; for Kurtosis contrast, the optimal step size polynomial is given by

$$p(u(w)) = \sum_{k=0}^{4} a_k \mu^k$$
 (4.32)

where the coefficients a_k can be obtained at each iteration by the observed signal block and the current values of w and g. Details can be found in [7], [11].

- Step 3) Extract the optimal step size polynomial root μ^k. The root can be obtained by using the Ferrari's formula as in [134].
- Step 4) Select the optimal step size polynomial root μ^k as follows

$$\mu_{\text{opt}} = \arg \max_{\mu} |K\left(y_i^{l}(q, w) + \mu g_i^{l}(q, w)\right)|$$
(4.33)

• Step 5) Find the updated weighed vector

$$u_i^{l+1} = u_i^l - \mu_{opt} g_i^l \tag{4.34}$$

where g_i^l is the ith gradient of Kurtosis contrast K(.) at lth iteration.

• Step 6) Normalize and update the weight vector

$$u_i^{l+1} = \frac{u_i^{l+1}}{\|u_i^{l+1}\|}$$
(4.35)

Where ||u|| is a norm of u.

• Step 7) Go back to step 2 until the convergence.

To prevent locking onto a previous extracted source, or when the old and new vectors w are in the same direction, the learning converges and their absolute dot-product value reaches close to 1. Thus, owning the deflation method proposed in [97], avoids different vectors from converging at the same maxima. However, each vector of $U = \{u_1, u_2, ..., u_n\}$ needs to be orthogonalized before each iteration. Based on the Gram-Schmidt orthogonalization, the deflation scheme estimates each independent component

at each iteration step. Gram-Schmidt orthogonalization of (i + 1)th component can be expressed as follows

$$u_{i+1}^{l+1} = u_{i+1}^{l} - \sum_{j=1}^{K} \left(u_{i+1}^{l} u_{j}^{l} \right) u_{j}^{l}$$
(4.36)

$$u_{i+1}^{l+1} = \frac{u_{i+1}^{l+1}}{\|u_{i+1}^{l+1}\|}$$
(4.37)

where a new weight vector u_{i+1} is obtained by subtracting the vector projected from the old weight vector.

The following steps summarize the presented algorithm procedure:

> Start

- \blacktriangleright Perform the time-frequency representation as in (4.4).
- For each frequency bin $w = 1, ..., \frac{T}{2}$
- > Pre-processing of the observed data x(t,w) and imposing the Tikhonov regularization parameter to avoid the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation.

$$\alpha = m. tr(E[x(t, w)x^{H}(t, w)])$$
(4.38)

where m is a positive constant and $tr(\cdot)$ represent the trace of estimation covariance matrix of the observation signals.

- Initialize K x K matrix W equals identify matrix I. Where K is the number of users.
- For each user k = 1, ..., K
- \blacktriangleright Initialize w_k column vector of the demixing matrix W
- ➤ While
- \succ Evaluate y(t, w) in (4.13)

- > Select the optimal step size polynomial root μ^k in (4.14)
- ▶ update weighed vector in (4.33)
- > Do the orthogonalization and normalization in (4.36) and (4.37), respectively
- \blacktriangleright Find kth users in (4.10).

$$y_i(q, w) = u_i^H(w) z(q, w)$$
 (4.39)

> Do deflation by subtracting the estimated kth source contribution to the z(q,w) as follows [97]:

$$z_{i+1}(q,w) = z_i(q,w) - h * y_i(q,w)$$
(4.40)

Where h is the symbol direction estimated via least squares, is given by

$$h = \frac{z_i(q,w).y_i^H(q,w)}{y_i(q,w).y_i^H(q,w)}$$
(4.41)

- Check the convergence point. if so, End while loop, otherwise, go back until the convergence.
- Save u_k in the U(w);
- \succ End for loop k.
- Save the demixing matrix U(w)
- \succ End w loop

4.5 Scaling and Permutation Ambiguities

Assume U(w) is the unitary matrix that computed at each bin, however, the least square estimation of the mixing matrix H(w) is given by

$$H_{LS}(w) = x(q, w)y'(q, w)(y(q, w)y'(q, w))^{+}$$
(4.42)

Where

$$y(q,w) = U(w)x(q,w) \tag{4.43}$$

However, one can express the estimated mixing matrix $H_{LS}(w)$ in term of the perfect mixing matrix H(w) as follows

$$H_{LS}(w) = H(w)D^{-1}(w)\Gamma^{-1}(w)$$
(4.44)

Where D(w) is an unknown diagonal matrix and $\Gamma(w)$ is an unknown permutation matrix. Therefore, we have to estimate D(w) and $\Gamma(w)$ matrices to solve the scaling and permutation ambiguities.

4.5.1 Estimation the diagonal matrix D(w)

Several methods to compensate the scale ambiguity have been proposed in the literature. Thus, we choose to estimate the diagonal matrix D(w) using the minimal distortion principle [3], [38], [129]. The D(w) is given in [3] as

$$D(w) = diag[AH_{LS}(w)]$$
(4.45)

$$D(w) = diag[Ax(q, w)y'(q, w)(y(q, w)y'(q, w))^{+}]$$
(4.46)

Where $A \in \mathbb{R}^{n \times m}$ is a matrix which has all its entries $\operatorname{are}^{1}/_{m}$; where m is the number of observations whereas, n is the number of sources, and $\operatorname{diag}[C]$ returns a matrix C that contains the diagonal elements of matrix C and sets the other non-diagonal elements of matrix C zeros. The interpretation of (4.46), in a sense of perfect separation, is that each estimated source averages along the sensors in the sense of all other sources have turned off. In other words, the Minimal Distortion Principle assumes that the nth source is scaled with respect to the image at the nth microphone [129]. Therefore, the rescaled source signals can be expressed as follows:

$$y^{rescaled}(q,w) \cong D(w)x(q,w)$$
 (4.47)

$$y^{rescaled}(q,w) \cong diag[Ax(q,w)y'(q,w)(y(q,w)y'(q,w))^+]x(q,w) \quad (4.48)$$

4.5.2 Estimation the permutation matrix $\Gamma(w)$

In this subsection, despite the fact that it estimates the permutation matrix $\Gamma(w)$ proposed in several current works in the literature, it is still considered a very challenging problem that needs to be addressed. Assume that we have *n* sources signals which are presented in BSS problem; then, there are *n* factorial times the possible permutations at each bin, which yields a complex combinational problem.

Mentioned previously, there are several techniques used to solve the permutation problem in the literature [3]. In this chapter, we will review and evaluate them in terms of computational complexity and performance.

One can divide these methods into two main solution groups to solve the permutation ambiguity in frequency domain as follows:

- Group based on the geometric information such as Time Direction of Arrivals (TDOA) and Direction of Arrivals (DOA) [3], [37], [38], [50].
- Group based on the clustering-based techniques [64], [67], [69], [70].
 Many of these techniques are based on the geometric information, such as

estimation of the direction of arrival (DOA) and Time difference of Arrival (TDOA) as in [50]. Other techniques depend on the coherence of the un-mixing filter coefficients. In other words, these techniques take advantage of some prior knowledge about mixing filters and restrict the mixing matrix H(f) to be continuous in frequency domain [130]. Furthermore, in [62], Parra imposes smoothness to the de-mixing filter values in the frequency domain.

Also, a restriction is made with the frequency domain update rule to be associated with the limited length filter in the time domain. Such a restriction may not be considered sophisticated especially in a case of reverberant environment since it is necessary to have a long length filter to cover all reverberations.

Although it can be avoided by choosing a large frame size, it still causes more overall complexity, especially, when the short mixtures are available. In terms of the properties of speech, there are other categories which have been proposed in literature, to estimate the permutation matrix and make the spectral alignment.

The most common is based on the inter-frequency correlation of speech envelopes [61], [65]. The inter-frequency correlation technique exploits the nature of speech production, where it's known that all spectral components of speech signal increase as the talker speaks louder. In that sense, several weighted techniques and criteria have been proposed to impose the frequency-coupling between the adjacent frequency bins, for more details [60], [61], [65]. Although these techniques perform well in the simulations, they are not sophisticated when they are applied to a real recoding room. They suffer from the propagation error or delays. For example, if an error occurs at a certain frequency bin it may increase the possibility to occur again at the following frequency bins.

Therefore, in the literature [3], [38], they avoid propagation error by estimating a frequency independent reference profile, which is called centroid, due to using a clustering based method for each separated source, and then structuring the n frequency dependent profiles such that they are all matched with a different frequency independent reference profile at each frequency bin.

The main steps of the clustering-based techniques are as follows:

134

- Define the quantities that are used in the clustering, such as signal envelopes of the source profiles and the log-power of the source profiles etc.
- 2) Choose the measure that is use to determine the matching level between the centroids and the profiles such as correlation and distance etc.
- 3) Choose the cluster technique.

In [69], the profile $\Psi_i(q, w)$ of a separated signal y_i is chosen to be the envelope of the separated source s_i where $\Psi_i(q, w) = |y_i(q, w)|$. In [70], they are chosen for the profile $\Psi_i(q, w)$ of a separated signal y_i to be a certain dominance measure. Whereas, in [67], the profile $\Psi_i(q, w)$ of a separated signal y_i is defined to be its centered log-power spectral density where the log-power profile is given as follows:

$$\Psi_{i}(q, w) = \log \left[W_{i,:}(w) R_{\chi}(q, w) W_{i,:}^{H}(w) \right]$$
(4.49)

In clustering based approaches, the length of the profiles T_f is also an important parameter in terms of accuracy; especially for short signals. Practically, we are going to set up the profiles for the overlapping frames over the whole signal. Once we get the profiles of the separated signals, then, we compute the centroids in order to perform the clustering.

The clustering based techniques is essentially based on the assumption that profiles coming from the same source at different frequency bins still have more match level than those coming from other sources.

Actually, the most common methods to associate each source profile to a centroid at each frequency bin are based on; 1) maximize correlation measures [69], [70]; 2) Minimize distance measures across the n factorial times of the possible permutations at each frequency bin [67]. However, they employ the iterative techniques to update the centroids and the permutation matrices. In other words, they update the centroids first, and then they permute the source profiles to each desired centroid and match them together using one of the two previous measures, i.e. distance [67] or correlation in [69] and [70]).

In spite of the fact that the aforementioned iterative methods perform well, they tend to be significantly more expensive in terms of cost and computational complexity since they have the n factorial times of the possible permutations at each frequency bin.

To avoid this drawback in the aforementioned iterative methods in [60], Nion proposes a more efficient modification of the clustering strategy which is updated the whole permutation matrices and centroids simultaneously. In other words, the update of the centroids and permutation matrices are not interleaved. Thus, their modification has improved these iterative methods in terms of computational complexity. Their methods can be summarized as follows:

Step 1. Determine the centroids and Compute them as following:

Consider the $n \ge T_f$ matrix $\mathcal{F}(w)$ that is structured from the *n* profiles $\Psi_i(w)$, $\forall i = 1, ..., n$. Furthermore, one can extend the $n \ge T_f$ matrix F(w) to the $Fn \ge T_f$ matrix G(w) by concatenating the matrices $\mathcal{F}(w) \forall w = 1, ..., F$. In order to enforce the Fn profile points in matrix G(w) varying smoothly with time, we have to encounter the computation of the profiles for overlapping frames. Hereafter, we just need to classify these Fn profile points into an *n* clusters due to apply the k-mean algorithm on the $Fn \ge T_f$ matrix G(w) to carry out a frequency independent $n \ge T_f$ centroid matrix $M = [m_1^T, m_2^T, ..., m_n^T]^T$. The centroid matrix is structured by summing all the points within a cluster, which have attained a minimum distance regarding to the centroid cluster. Furthermore, the k-means algorithm also gives the list of indices that attains each one of n clusters. Thus, our simulation shows that almost about F points are assigned to each cluster, which implies that the aforementioned property of the speech is valid. Therefore, we only need to exploit the frequency independent $n \times T_f$ centroid matrix M to do the computation processes and mitigate the computational complexity.

Step 2. Estimating the permutation matrices

In previous step, we reduced the computational processes to finding the $n \times n$ permutation matrix $\Gamma(w)$ subject to $G(w)\Gamma(w)$ match the frequency independent $n \times T_f$ centroid matrix M at each frequency bin. Therefore, one can choose to minimize the distance that is given in [67] as follows

$$\min_{\Gamma(w)} \|M - G(w)\Gamma(w)\|_F^2 \ \forall \ w = 1, 2, \dots F$$
(4.50)

Or choosing the correlation criteria that is given by [69], [70] as follows

$$\max_{\Gamma(w)} \sum_{i=1}^{n} \Phi(m_i, [G(w)\Gamma(w)]_{:,i}) \forall w = 1, 2, ... F$$
(4.51)

Where $\Phi\langle \cdot, \cdot \rangle$ is the correlation coefficient.

In terms of performance, the first group generally does well and better than the second group, especially at the small data sample available. But it is not optimal in practical sense, since we don't usually have geometric information about the real environment conditions. In that sense, the second group performed better than first group especially if we have a large sample set of data, because they are based on the clustering-based techniques i.e.: correlation, distance, etc. and they are more robust to the real world scenarios. For more details, refers to [3] [38].

4.6 **Experiments Results**

In this section, we examine the performance of the RobustICA-based algorithm developed in this chapter. The time-frequency representation of the observed data is computed as explained in section due to the Short-Time-Fourier-Transform. Then, for each frequency bin, we find the demixing matrix as we will present in this section. Then, we will solve the scale and permutation ambiguities based on the aforementioned techniques.

However, we divided this section into two subsections. Firstly, we illustrate the performance of the RobustICA-based algorithm with different permutation methods in the literature [3], [38]. Also, we study the effect of the type of the windows on the performance of the presented algorithm as well as the effect of overlapping parameter.



Figure 4.1: configuration of the two experimental setups that were conducted by Francesco Nesta1 in [130], a) room is characterized for Test1, b) class-room is characterized for Test2

Secondly, we provide the performance of the presented algorithm in two realworld scenarios that are generated in adverse conditions by, in [130], and compare it with others state-of-the-arts in the [130], [36], and [62] and [67], labeled as "RR-ICA", "IVA", "Parra", "Pham", respectively. In this chapter, we evaluate the performance of the presented algorithm due to the BSS_EVAL toolbox, which is proposed in [126], [127]. We use time-invariant filters of 1024 taps to represent the signal to interference ratio (SIR) and source to distortion ratio (SDR).

4.6.1 Section 1

In this subsection, we study the computational complexity and the performance of the presented algorithm based on several following criteria to solve the scale and permutation ambiguities in the frequency domain BSS problem. Let's define these criteria as following:

<u>Method1</u> is the RobustICA-based algorithm with clustering of envelope profiles with a distance measure iterative procedure [8].

<u>Method2</u> is the RobustICA-based algorithm with clustering of log-power -profiles with a correlation measure iterative procedure [21].

<u>Method3</u> is the RobustICA-based algorithm with clustering of envelope profiles with a distance measure kmeans procedure [8], [60].

<u>Method4</u> is the RobustICA-based algorithm with clustering of log-power profiles with a correlation measure kmeans procedure [21], [60].

<u>Method5</u> is the RobustICA-based algorithm with clustering of dominance-profiles with a correlation measure iterative procedure [22].

<u>Method6</u> is the RobustICA-based algorithm with clustering of dominance-profiles with a correlation measure iterative kmeans procedure [22], [60].



Figure 4.2: Results obtained in Test1 experiments. The SIR performance of the presented algorithm with various permutation solvers

In this section, we have used real world recordings, resulting from the experiments were conducted in [130] named: Test1. We would like to thank the authors who provided these recordings on their website "http://bssnesta.webatu.com/testhscma.html". The two sources were recorded at $f_s = 16 \, kHz$ with two microphones spaced by $d = 0.02 \, m$ apart to avoid the spatial aliasing. The chosen room was characterized by a moderate reverberant time of 160 ms. The room had dimensions of (3.5m x 5.1m x 2.6m) as shown in Figure 4.1. The signal duration was fixed to be 9 sec.



Figure 4.3: Results obtained in Test1 experiments. The SDR performance of the presented algorithm with various permutation solvers

In Figure 4.2 and Figure 4.3, we show the performance of RobustICA-based algorithm with various aforementioned techniques of permutation solvers in terms of the SIR and SDR, respectively. In comparison, we notice that the *dominance*-profiles provide more robustness in terms of the signal's length, although the envelope *profiles are more* sensitive to the signal's length than the *log-power profiles*. Moreover, the *dominance-profiles approach* with the iterative procedure *has the same performance as with the kmean procedure*. Also, Figure 4.4 shows the corresponding CPU time of each permutation method that need to solve the permutation ambiguity.



Figure 4.4: Corresponding CPU time for each method.

Based on these observations, we will use the *dominance-profiles approach* with the iterative procedure after the RobustICA-based algorithm in the rest of these experiments. In Figure 4.5, we illustrate the impact of the window's types on the performance of the proposed algorithm in terms of SIR and SDR respectively. And, we test the performance of the presented algorithm versus the overlapping parameter as it shown in Figure 4.6.The best performance of the presented algorithm was achieved during the certain range of the overlapping percentage. Therefore, based on these results, we use the overlapping parameters to be 0.65 with Hamming window type.



Figure 4.5: Figure 4.5: Results obtained in Test2 experiments. The SIR performance of the presented algorithm with various window types



Figure 4.6: Results obtained in Test2 experiments. The SIR performance of the presented algorithm with various overlap ratios

4.6.2 Section 2

In this section, we perform the separation of the two mixture observations what consists of two sources. We have used the two tests "Test1 and Test2" of the real world recordings, resulting from the experiments were conducted in [130], see fig.23. Test2 uses the real world recordings of adverse reverberant conditions as in fig 23. The room is a reverberant class-room with dimensions 4.75 m Length x 5.92 m Width x 4.5 m Height. The reverberation time is around 700 mile second $T_{60} = 700ms$. The performance is also averaged over ten random pairs of sources. The signal duration was fixed to be 9 sec. After we got the demixing matrix W for each frequency bins, we used the Inverse Fourier Transform to obtain the mixing matrix in the time domain.

- The independent vector analysis IVA [17] used with step size 0.1 and number of iterations is 1000.
- ▶ Parra's method [38], used with number of iterations is 1000.
- Pham's algorithm [12] and [20]. Used with FFT overlapping equals 75% and a window size equal to 5.
- ➢ RR-ICA algorithm reported in [130].

Figures 4.7 & 4.8 and 4.9 & 4.10, shows the summary analysis of the presented algorithm versus other algorithms presented in [130] for Test1 and Test2 configurations, respectively. These graphs are reported the best performance of each algorithm over the FFT size. Obviously, the RobustICA-based algorithm are outperforms the others algorithms for any signal length in terms of SIR and SDR.

Moreover, in Figure 4.11, we illustrate the impact of the FFT length on the performance of the proposed algorithm in terms of SIR. Clearly, the presented algorithm

performs well especially during reasonable FFT length in regards to other corresponding algorithms as it shown in Figure 4.11.

Based on these results, one can show that the presented algorithm is stable in terms of the high reverberation environment and variations of the observations' parameters. Furthermore, the presented algorithm performs well in terms of stability and speed convergence. Owning the optimal step size, deflation and regularization techniques make the presented algorithm more robust and perform well even though in the adverse conditions. Therefore, the presented algorithm performs well for solving the convloutive BSS problem of the real-world recordings in adverse conditions.



Figure 4.7: Results obtained in the Test1 experiments [130]. Best performance is reported in terms of SIR, by applying the given algorithms with different signal lengths







Figure 4.9: Results obtained in the Test2 experiments [130]. Best performance is reported in terms of SIR, by applying the given algorithms with different signal lengths



Figure 4.10: Results obtained in the Test2 experiments [130]. Best performance is reported in terms of SDR, by applying the given algorithms with different signal lengths



Figure 4.11: Impact of FFT length, 2-by-2 case, Results obtained in the Test2 experiments [130].

4.7 Conclusion

This chapter presented the RobustICA-based algorithm to solve the frequencydomain BSS problem for convolutive acoustic mixtures in several adverse conditions. Through the real-world experiments, we show the superiority of the presented algorithm among other popular algorithms in the literature in terms of the performance and complexity computation. Moreover, we compared several permutation solvers in terms of computation complexity and performance to provide the RobustICA-based algorithm with an efficient frequency-dependent permutation scheme. Finally, we studied the effect of several parameters on the separation performance of the presented algorithm. We also presented the effect of the type of the window on the separation performance and we also showed that the performance improves at a certain range of overlapping between the signals. Lastly, in this chapter, we showed the performance of a system that can work efficiently with around 0.5-10 seconds of input data, which is close to the real-time implementation. Accordingly, our proposed algorithm is optimized to be suitable for the real-time operation. As a result, it is suitable for a large number of applications to ensure the real-time implementation.

5 Chapter 5

Robust Blind Multiuser Detection Algorithm Using Fourth Order Cumulant Matrices

A new blind detection algorithm, based on fourth order cumulant matrices, is presented and applied to the multi-user symbol estimation problem in Direct Sequence Code Division Multiple Access (DS-CDMA) systems. The blind detection is to estimate multiple symbol sequences in the downlink of a DS-CDMA communication system using only the received wireless data and without any knowledge of the user spreading codes. The proposed algorithm takes advantage of higher cumulant matrix properties to reduce the computational load and enhance performance. Bit error rate (BER) simulations of this algorithm are shown for different number of users, signal to noise ratios (SNR) and different number of symbols per user in comparison with the FAST ICA and Robust ICA algorithms. The results show that the proposed algorithm outperforms both ICA-based detectors in estimating the symbol signals from the received mixed signals. Moreover, the proposed blind detector is computationally fast and exhibits high convergence speed in extracting user symbols.

5.1 Introduction

Communication systems performance hinges on speedy and reliable data/symbol

transfer among users. In that context, data reaches each user with as few errors as possible. Code Division Multiple Access (CDMA) family of systems continue to be the most deployed and popular multiple access scheme. This is mainly due to its soft multiple access characteristics, robustness against fading and its anti-interference capability. In schemes accepting non-orthogonal multiple-access designs, like direct sequence code division multiple-access (DS-CDMA), multiple-access interference (MAI) is the limiting feature of the scheme's capacity. To alleviate MAI, a variety of multi-user detectors/receivers have been proposed; whereas, most of them need either the data of the preferred user's dispersion sequence or a preparation (pilot) sequence. When neither is easily possible, due to computational delays or time constraints, the challenge of extracting the broadcast information generally belongs to the domain of blind source separation (BSS) [74-80].

The conventional or single-user detection (SUD) methods consider MAI as external noise. In an alternative approach, the structure of MAI is modeled as in the work on optimum multi user detection [76]. However, it has been shown that these detectors either need a complete knowledge of the MAI and training data, or involve a long decoding delay [74]. To overcome these limitations, classes of efficient blind detectors are proposed in the literature. However, the blind detection techniques in the wireless communication literature, e.g., in [75] utilize primarily the Second Order Statistics (SOS) and some Higher Order Statistics (HOS) of the received data. Independent Component Analysis (ICA) is a statistical technique that includes HOS, where the goal is to represent a set of random variables as a linear transformation of statistically independent components [78]. ICA based techniques are based on the assumption of non-Gaussianity and independence of the sources. For example, the Fast ICA algorithm is applied for the detection of symbols in DS-CDMA systems in [77], while its convergence is not ensured. RAKE-ICA is also proposed in [75]. Recently, blind multiuser detection based on Tikhonov regularization has been proposed in [76], which requires the prior knowledge of the signature sequence and timing of the desired user.

ICA based algorithms have been the most active in solving BSS problems during the last decade [1] - [2]. There are numerous targeted applications in speech, image and biomedical signal processing. Several approaches have been established for constructing the ICA algorithms [9– 20], and one such an approach is based on information theory. The Information Maximum (InfoMax) algorithm proposed in [18], e.g., is derived from HOS. It extracts the source signals from the mixed signals; however, it involves high computational complexity. To reduce this complexity, the Fast ICA algorithm, based on an approximation of Negentropy and Newton iteration, has been subsequently developed.

A comparative study of ICA-based algorithms for the DS-CDMA detection problem [75], [88] shows that the Fast ICA algorithm performs comparatively well in extracting the unknown symbols. Whilst, it also involves high computational complexity and exhibits slow convergence.

Recently, Zarzoso et al [11] have proposed an ICA-based algorithm with a new contrast function that can avoid permutation ambiguity and has quantitatively better separation quality than the so-called conventional ICA algorithm (C-ICA). Furthermore, Zarzoso and Comon [1] have introduced a search line into the iterative maximizing Kurtosis contrast function to render the Fast ICA more Robust and to increase its computational efficiency. By using an algebraic optimal step size, they have shown that

the extraction quality would be better than the Fast ICA algorithm. Therefore, the Robust ICA algorithm has efficient extraction primarily for smaller number of components (sources). Moreover, there exist some non-linear ICA-based algorithms that are used to separate dependent source signals [89] and can extract only one signal [90].

This chapter presents a Blind DS-CDMA detection technique based on fourth order cumulant matrices [21], [34]. This technique is very effective in multi-user CDMA environments where no prior information of the user's code is required to be known at the receiver. This approach is considered blind because the spreading codes of all the users, the characteristics of the environment, as well as the transmitted symbols are assumed unknown. The simulation scenarios are carried out to observe variations in the bit error-rate as a function of (i) signal to noise ratio, (ii) number of users and/or (iii) number of symbols per user. Furthermore, the performance of the proposed algorithm is quantified and a comparison is made among the three ICA-based algorithms in terms of their performance and computational complexity. The remainder of the chapter is organized as follows. In Section II, a brief description of DS-CDMA signal model and multi-path fading is presented. In Section III, Robust and Fast Independent Component Analysis (ICA) and their signal model are discussed briefly. Section IV proposes the new detector based on the fourth order cumulant matrices. The comparative simulation results and conclusions are given in Sections V and VI, respectively.

5.2 DS-CMDA Signal Model

In a typical downlink (Synchronous) CDMA system, the CDMA employed in Evolved High-Speed Packet Access (HSPA+), in 4G systems, to keep the transmitted bandwidth constant regardless of the bit-rate through solving the non-symmetric user-

152

bandwidth problem. The synchronous DS-CDMA system has been used in Satellite systems, indoor ATM, and in certain ad hoc wireless network because of its attractive features (namely, anti-interference capability and robustness against fading). DS-CDMA systems allow several users to share the medium simultaneously by using their own signatures (spreading codes). The synchronous DS-CDMA system assigns shorter spreading codes to higher-rate users and longer spreading codes to lower-rate users, while keeping the chip rate constant. A typical DS-CDMA system model is given, e.g., in [75]. The simplest received signal model r(t) is

$$r(t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{l=0}^{L} \alpha_{lm} b_{k,m} s_k (t - mT_b - d_l T_c) + n(t)$$
(5.1)

where

- l, k, m are path, user and symbol indices, respectively.
- α_{lm} is the path gain since, in a downlink model, the path gain is the same because all users' coded signals are transmitted together and the path gain α_{lm} and propagation delay factor d_l depend only on the signal paths.
- b_{k,m} is the kth user m symbol.
- $s_k(.)$ is the kth user spreading code
- d₁ is the propagation delay factor, d_{kl} ∈ {0, 1, ..., C-1/2} (assumed to be of duration of at most half the sequence C).
- t, T_b, T_c are time, symbol, and chip duration, respectively.
- n(t) is an additive white Gaussian noise (AWGN).

The received signal is assumed to be properly sampled and synchronized discrete data. However, let us assume that G is the number of code sequence, K is the number of

users, and L is the number of channels. Thus, the vector form of the equation (1) will change to be as:

$$\mathbf{r} = \mathbf{A}\mathbf{S}\mathbf{b} + \mathbf{n} \tag{5.2}$$

Where r is the received vector signal, A is an $(G + L - 1) \times G$ matrix which represents the multipath propagation coefficients, S is an G x K block diagonal matrix, b is an K – d vector which represent the data symbols, n is the (G + L - 1) - d channel noise vector with covariance matrix Q. This model received signals (5.2) is suitable for deriving the linear symbols detectors such as the MF, the RAKE, the LMMSE and the blind Detectors based on FastICA and Robust ICA algorithms.

5.3 Robust and Fast Independent Component Analysis (ICA)

A simple Blind Source Separation (BSS) model with n Source Signals and n observations is defined as follows:

Let the sources form the vector

$$S = [S_1, S_2, ..., S_n]^T$$
 (5.3)

Let the observations form the vector

$$X = [x_1, x_2, ..., x_n]^T$$
(5.4)

The (static) linear BSS model is

$$X = AS + N \tag{5.5}$$

where A is an n x n (invertible) mixing matrix, and the vector N is an Additive Gaussian Noise. The ICA algorithm assumes that no more than one source signal has Gaussian distribution.

The main idea of ICA is to recover the source signals from the observed signals without a priori knowledge of the vector source signal or the mixing matrix. To achieve this, an ICA-based algorithm iteratively computes a weighting matrix W that incrementally approximates the inverse of the mixing matrix A. The estimated source signal Y is thus given as follows:

$$Y = WX \tag{5.6}$$

where

$$W = \left[w_{jk} \right], \ 1 \le j, k \le n \tag{5.7}$$

As it is a linear transformation, we can choose to estimate one of the independent components in the form of wX, where w is a row vector of the matrix W in (5.7). The estimation can be achieved by maximizing the "non-Gaussianity" of wX.

Generally Speaking, It can be said that the Fast ICA algorithm entails two steps: one is the preprocessing step and the other is finding the rotating matrix W that maximizes the non-Gaussianity of Y = WX.

5.3.1 The Preprocessing

This involves two computations. The first aims at centering the mixed matrix signal X by removing its mean as follows:

$$\overline{\mathbf{X}} = \mathbf{X} - \mathbf{E}[\mathbf{X}] \tag{5.8}$$

where \overline{X} represents the zero-mean mixed signal of X, and E[.] is (an estimate of) the expected value operator. The second whitens the mixed signals to have a unitary variance after centering the mixed signal. To that end, we compute The Singular-Value Decomposition (SVD) [29] is applied to the covariance matrix of \overline{X} to obtain

$$C_{x} = E[\overline{X} \, \overline{X}^{T}] = EDE^{T}$$
(5.9)

where E represents the eigenvectors which are orthogonal matrices of mixed signals, and D is a diagonal of the Eigenvalues C_x , expressed as:

$$D = diag(d_1, d_2, ..., d_n)$$
(5.10)

Thus, the whitening operation of \overline{X} is expressed as

$$Z = (D^{-\frac{1}{2}}E^{T})\overline{X} = V\overline{X}$$
(5.11)

where V equals $D^{-\frac{1}{2}}E^{T}$ and represents the pre-multiplying whitening matrix of \overline{X} . Equation (11) shows that the centered matrix \overline{X} is linearly transformed to the matrix Z. Observe now that the covariance matrix of Z equals the identity matrix (I_K) [2]. In other words, the componenets of the vector Z are uncorrelated.

5.3.2 The FastICA Algorithm

The batch algorithm FastICA [25], [26] is derived by maximizing the non-Gaussianity of a measure based on the fixed point iteration. The Center-Limit-Theorem (CLT) [2] infers that (a large) sum of independent sources will become closer to a Gaussian distribution. The CLT shows that, for whitened data, finding an independent source is achieved by finding the direction vector w which gives a component its "maximum non-gaussianity." (Further discussions are in section II-A.) Moreover, if just one of the independent components (ICs) is Gaussian, the ICA algorithm can still work and estimate the ICs. The authors in [25, 26] use the Fast ICA, which is a fixed point iteration scheme, for finding the maximum of the non-gaussianity. The basic Fast ICA scheme for one independent component is outlined as follows:

- Choose an initial value for the weight vector w
- Find the updated weight vector

$$w^+ = E\{Z[g(wZ)]^T\} - E\{g'(wZ)\}w$$

(5.12)

- where g is a non-quadratic function such as $g(y) = y^3$, y = wZ and g' is the derivative of the non-quadratic function g.
- Normalize and update the weight vector, i.e.,

$$w^{+} = \frac{w^{+}}{\|w^{+}\|} \tag{5.13}$$

where ||. || is the norm operator.

• Go back to step 2 until the convergence.

When the old and new vectors w are in the same direction, the update converges and their absolute dot-product value reaches close to 1. In the same context, the deflation method proposed in [1], [25] avoids different vectors from converging to the same value. It enforces each vector of $W = \{w_1, w_2, ..., w_n\}$ to be orthogonalized based on the Gram-Schmidt orthogonalization. The deflation scheme estimates each independent component at each iteration step. The Gram-Schmidt orthogonalization of the (k + 1)th component can be expressed as follows

$$w_{k+1}^{+} = w_{k+1} - \sum_{j=1}^{k} (w_{k+1}^{T} w_{j}) w_{j}$$
(5.14)

$$\mathbf{w}_{k+1} = \frac{\mathbf{w}_{k+1}^+}{\|\mathbf{w}_{k+1}^+\|} \tag{5.15}$$

where a new weight vector w_{k+1} is obtained by subtracting the vector projected from the old weight vectors.

5.3.3 The Robust ICA algorithm

Recently, Zarzoso and Comon [1] improved the robustness of the Fast ICA algorithm by using a line search direction to choose the optimal step size $\mu_{opt} = \arg \frac{\max}{\mu} |K(y + \mu g)|$ where g is the gradient of the kurtosis contrast function K(.). At each iteration, the Robust ICA algorithm performs an optimal step-size as follows [11]:

- \checkmark Choose an initial value for the weight vector w
- ✓ Compute the optimal step size polynomial coefficients; for the kurtosis contrast function, the optimal step size polynomial is given by

$$p(u) = \sum_{k=0}^{4} a_k \mu^k \tag{5.16}$$

- ✓ where the coefficients a_k can be obtained at each iteration from the observed signal block and the current value of *w* and *g*. Details can be found in [7].
- ✓ Compute the optimal step size polynomial roots µ^k. The roots can be obtained by using the Ferrari's formula [36].
- ✓ Select the optimal step size polynomial root μ^k as follows:

$$\mu_{opt} = \arg \, \max_{\mu} |K(y + \mu g)| \tag{5.17}$$

✓ Find the updated weight vector.

$$w^+ = w + \mu_{opt}g \tag{5.18}$$

where g is the gradient of the kurtosis contrast function K(.).

✓ Normalize and update the weight vector

$$w^{+} = \frac{w^{+}}{\|w^{+}\|} \tag{5.19}$$

where ||. || is a norm operator.

5.4 The Proposed Detection Algorithm Based On Cumulant Matrices

In this section, a new blind detection strategy is proposed, based on the cumulant matrices that uses the algorithms JADE, SHIBBS and JAD [2] [16], [17], and [19]. Here, one needs to first recall the received matrix form (5.2),

$$r = ASb + n$$

The aim here is to detect the b symbol vector from the received data r under the following assumptions:

- AS1) the G x K matrix, H := RA, is of full column rank.
- AS2) the symbol signals, b, are non-Gaussian, independent and identically distributed (i.i.d)
- AS3) the Additive Noise vector is white and independent of source signals
- AS4) the power of the transmitted symbol signals, normalized to be unity.
- AS5) in this chapter, all signals are real and the number of users K is given.
 <u>The Method involves two steps:</u>

5.4.1 Step1: Preprocessing (Data Whitening)

This is a common preprocessing step. As a consequence, the symbol signals are detected up to a unitary matrix using second order statistics (SOS). This step is used to reduce the noise and to eliminate correlations among the data components. Under the Assumptions AS1, AS2, AS3 and AS4, the G x G covariance matrix (R) of the noiseless transmitted signals can be expressed by

$$R = E[rr^{T}] - \sigma^{2}I_{K}$$
(5.20)

By substituting X from (5.2) into (5.20), one gets

$$\mathbf{R} = \mathbf{H} \, \mathbf{E}[\mathbf{b}\mathbf{b}^{\mathrm{T}}] \, \mathbf{H}^{\mathrm{T}} = \mathbf{H}\mathbf{H}^{\mathrm{T}} \tag{5.21}$$

Under AS1, R can be decomposed as

$$R = V\Lambda V^{T}$$
(5.22)

where V is a G x G matrix of eigenvectors satisfying

$$VV^{T} = V^{T}V = I_{G}$$

$$(5.23)$$

and Λ is a G x G diagonal matrix containing real (eigenvalues) entries. Thus, the G x K matrix H can be expressed in its singular value decomposition (SVD) as [29]

$$\mathbf{H} = \mathbf{V} \Lambda^{-\frac{1}{2}} \mathbf{U}^{\mathrm{T}} \tag{5.24}$$

where U is a KxK full rank unitary matrix and $UU^{T} = I_{K}$.

The whitening step obtained the matrix V so that the Kx1 whitened data vector Z has covariance equals to the identity matrix, $R_{ZZ} = I_{K.}$ Specifically,

$$Z = \Lambda^{-\frac{1}{2}} V^{T} r \tag{5.25}$$

$$\mathbf{Z} = \mathbf{U}^{\mathrm{T}}\mathbf{b} + \Lambda^{-\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\mathbf{n}$$
(5.26)

Thus, the transmitted symbols can be recovered with a linear Zero-Forcing (ZF) equalizer. Consequently, the estimated Kx1 symbols

$$\hat{\mathbf{b}} = \mathbf{U}\mathbf{Z} \tag{5.27}$$

After the preprocessing step, the detection of the symbol signal \hat{b} reduces to determining the KxK unitary matrix U (a rotation matrix).

5.4.2 Step 2: Determining the rotation matrix (unitary matrix) U.

One way of finding the rotation matrix U is based on the fourth order cumulant matrices. To estimate U in (27), we exploit the statistical independence of the equalized symbols vector. More precisely, the unitary matrix U will be estimated by using the independent property of symbols in the fourth order cumulant matrices (say, Q), so that it has many of the cumulant elements equal to zero [18]. The well-known Joint Approximation Diagonalization of Eigen Matrices algorithm [18] is efficient and robust when separating a small number of sources. JADE uses both second order decorrelation (whitening process) and fourth order cumulant matrices (rotation process) to separate the source signals from the mixed signals. Hyvarinen et al [18], [19], show that JADE is not efficient when the sources are numerous, because the size of the 4th order cumulant sets increases with the 4th power of the number of sources. This requires a large memory in addition to incurring a high cost for Eigen matrix calculation. From [21, 34], however, one can get a good sense of the JADE algorithm by using only a portion of the fourth order cumulant matrices-- up to the number of users (K), and thus avoiding the Eigen matrix calculation. The Fourth order cumulant matrices of the independent zero-mean symbols \hat{b} for $1 \leq i, j, k, l \leq K$, are

$$Q_{\hat{b}} = \operatorname{cum}(\hat{b}_{i}, \hat{b}_{j}, \hat{b}_{k}, \hat{b}_{l})$$

$$Q_{\hat{b}} = E[\hat{b}_{i}\hat{b}_{j}\hat{b}_{k}\hat{b}_{l}] - E[\hat{b}_{i}\hat{b}_{j}]E[\hat{b}_{k}\hat{b}_{l}] - E[\hat{b}_{i}\hat{b}_{k}]E[\hat{b}_{j}\hat{b}_{l}]$$

$$-E[\hat{b}_{i}\hat{b}_{l}]E[\hat{b}_{j}\hat{b}_{k}] \qquad (28).$$

Because of AS3, the symbols of vector \hat{b} are assumed to have unitary variance. Thus, the 4th order cumulant matrices become:

$$Q_{\hat{b}} = \operatorname{cum}(\hat{b}_{i}, \hat{b}_{j}, \hat{b}_{k}, \hat{b}_{l}) = E[\hat{b}_{i}\hat{b}_{j}\hat{b}_{k}\hat{b}_{l}] - \delta_{ij}\delta_{kl} - \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk} \quad (29).$$

where δ_{ij} is the Kronecker delta, and equals $1 \quad \forall i = j$; otherwise zero. Because of the independent property of the symbols \hat{b} , many of the cumulant elements are zero. However, one can rewrite the fourth order cumulant set Q for symbols \hat{b} (28) and for the whitened vector Z (27) as follows, respectively:

$$Q_{\hat{b}} = \operatorname{cum}(\hat{b}_{i}, \hat{b}_{j}, \hat{b}_{k}, \hat{b}_{l})$$
$$Q_{\hat{b}} = \begin{cases} \neq 0 & \forall i = j = k = l \\ = 0 & \text{Otherwsie} \end{cases} \forall 1 \le i, j, k, l \le K$$
(5.30).

$$Q_{\hat{b}} = \operatorname{cum}(Z_{i}, Z_{j}, Z_{k}, Z_{l}) \neq 0 \quad \forall \ 1 \le i, j, k, l \le K$$
(5.31)

Owing to this difference between (31) and (32), the symbols could be detected from the received signals. Here, the objective function, which was proposed in [21], is used to determine the unitary matrix U and then to estimate the symbols vector \hat{b} (27), specifically,

$$U: \begin{cases} C(U, \hat{b}) = \sum_{i=k=1}^{K} \operatorname{cum}^{2}(\hat{b}_{i}, \hat{b}_{j}, \hat{b}_{k}, \hat{b}_{l}) \quad \forall \ 1 \le i = j, k = l \le K \\ \text{subject to min} \ (1 - \frac{\operatorname{trace}(U)}{K}) \end{cases}$$
(5.32)

Note that the iteration stop criterion of the proposed algorithm ends the iteration when the unitary matrix U tends to be almost the identity matrix I. So, we can express it mathematically by:

$$(1 - \frac{\operatorname{trace}(U)}{K}) \le \epsilon \tag{5.33}$$

where ϵ is a threshold value, e.g., $\epsilon = e^{-3}$.

The following are the bulleted steps of the proposed algorithm, based on the Fourth order cumulant matrices:
- Pre-processing
- Zero mean of received signal
- Compute the covariance Matrix R
- Compute the whiten vector Z
- Repeat
- Calculate Cumulant Matrices $Q_{\hat{b}}$ (5.32)
- Call Joint Diagonlization U in (33)
- Continue until $\min\left(1 \frac{\operatorname{trace}(U)}{K}\right) \le \epsilon$ where ϵ is the threshold.
- End

5.5 Simulation Results

5.5.1 Performance

The simulated DS-CDMA downlink data, in the presence of AWGN, is used to verify the effectiveness of the proposed algorithm and to compare it to the detectors based on the Fast ICA and Robust ICA algorithms. We used spreading codes as short gold codes with length of chips to be NC=31. The maximum number of users is K =30. We assumed all signals for all users to be sent with the same power. Monte Carlo Simulations are executed to verify the validity and the effectiveness of the proposed algorithm in comparison to the two detectors. Fig. 34 shows the simulation results of BER vs. SNR of the presented detectors. The parameters are set as: Number of symbols M=3500, Number of users K=30, Number of paths L=1, with various values of SNR from -10 dB to 5dB. Figure 5.1 demonstrates that the proposed algorithm, based on the

fourth order cumulant matrices (SJAD), improves the performance of the CDMA system, producing the lowest BER consistently in comparison. It is also observed that the performance of the FastICA-detector and RobustICA-detector is almost identical.

Figure 5.2 shows the simulation results of BER vs. SNR with various numbers of symbols (M) with 30 users (K=30). It is obvious that the performance of the proposed algorithm performs more consistently as M increases and is better than that of the RobustICA Detector. Furthermore, Figure 5.2.c shows that the RobustICA detector, in contrast to the proposed method, performs poorly as M decreases. Figure 5.3 shows the simulation results of BER vs. SNR with medium number of symbols, M= 3500, various K users. It is obvious that the proposed method performs more consistently and exhibits improvement in the performance as K decreases. The algorithm also mitigates the MAI, and has better performance than the other Detectors as shown in Figure. 5.3. Also, the FastICA detector performs well as K decreases in Figure. 5.3.b to the contrary of the RobustICA detector in Figure 5.3.c. As a result, the proposed detector performs well in estimating and recovering symbols in DS-CDMA system.



Figure 5.1: Average BER as a function of SNR for 30 users and 100 runs.



Figure 2.a Detector based sjad algorithm



Figure 2.b Detector based FAST ICA algorithm



Figure 2.c Detector based RICA algorithm

Figure 5.2: Average BER as a function of SNR for different sample lengths T with 30 users and 100 runs. Black triangle right lines: M = 104. Black circle lines: T = 9000. Black hexagram lines: T = 8000. Black square lines: T = 7000. Red triangle up lines: M = 6000. Blue circle lines: T=5000. Blue hexagram lines: T=4000. Blue square lines: T=3000. Blue triangle right lines: T=2000.



Figure 3.a Detector based sjad algorithm



Figure 3.b Detector based FAST ICA algorithm



Figure 3.c Detector based RICA algorithm

Figure 5.3: Average BER as a function of SNR for different Users K with signal blocks composed of T = 3500 samples and 100 runs. Black triangle right lines: K = 10. Black circle lines: K = 20. Red triangle up lines: K = 30. Blue square lines: K=40. Blue hexagram lines: K=50.

5.5.2 Measure of Computation

A measure of Complexity in Computation for the algorithm, in terms of iterations, is widely used. This kind of measure is not justifiable in the sense that an algorithm may require only a few iterations to converge, but each iteration may involve heavy computation. Also, such a measure does not take into account the fact that computation time depends on actual algorithmic implementation. However, for this work, the measure followed here is based on the number of real-valued floating point operations (flops) that are required to reach a solution for an algorithm. Flop count details can be found in [7], [11]. As a natural measure of extraction quality, the average signal of mean square error SMSE was employed as a contrast function of independent criteria and is defined as follows:

$$SMSE = \frac{1}{2} \sum_{k}^{K} SMSE_{k,l(k)}$$
(5.34)

where

$$\text{SMSE}_{k,l} = \text{E}[|s_k - \alpha_l \hat{s}_l|^2]$$

and with

$$\alpha_{l} = \frac{E[s_{k}\hat{s}_{l}^{*}]}{E[|\hat{s}_{l}|^{2}]}$$

However, to study the performance of the proposed detector and to compare the three detectors in terms of computational complexity, the SMSE performance index in (5.34) is used and averaged over 1000 independent realizations of the received data. The extracted symbols are computed directly from the observed received data. Fig. 39 summarizes the performance and complexity variations obtained for T= 3500, with different values of the users K. As can be seen in Figures 5.4 and 5.5, the best and faster

performance is provided by the proposed algorithm based on fourth order cumulant matrices, especially, when K is relatively small. Moreover, the FastICA algorithm performs well as the proposed algorithm when K increases. In contrast, the RobustICA detector performs poorly when K increases. In the same context, the performance of the algorithms for a varying block of sample size T is evaluated.



Figure 5.4: Average extraction quality as a function of computational cost for different mixture sizes K with signal blocks composed of T = 3500 samples and 1000 mixture realizations. Solid lines: K = 5. Dashed lines: K = 10. Dotted lines: K = 20.



Figure 5.5: Average extraction quality as a function of computational cost for different mixture sizes K with signal blocks composed of T = 3500 samples and 1000 mixture realizations. Solid lines: K = 30. Dashed lines: K = 40.

Figure 5.6 shows the average SMSE curves for different numbers of symbols K with different T samples. From this figure, one can see that the proposed detector is considerably more efficient than the other two detectors. Overall, the best performance is provided by the proposed detector which achieves the given performance level with lower cost.



Figure 5.6: Average extraction quality as a function of computational cost for different mixture sizes K with different signal blocks T samples and 1000 mixture realizations. Solid lines: K = 20. Dashed lines: K = 30. Dotted lines: K = 40.

5.6 Conclusion

In this chapter, we have investigated three adaptive algorithms for user-detection in CDMA systems, the proposed one based on fourth order cumulant matrices, the Fast ICA and the Robust ICA algorithms. The results show that the proposed algorithm exhibits better performance relative to the other two user detectors. The results also show that the proposed algorithm can mitigate Multiple Access Interference (MAI), thus improving the performance of conventional detection. Furthermore, the performance of the proposed detector displays the most consistent improvement as M (the number of symbols) increases. Also, we assess the performance of computational complexity of the three user detection algorithms employing the average signal of mean square error SMSE, as a contrast function of independent criteria. The results show that the proposed detector provides a faster and more robust performance.

6 Chapter 6

Adaptive Blind Multiuser Detection DS-CDMA Based on State Space Approach

Code Division Multiple Access (CDMA) is a channel access method used by various radio technologies and it is based on spread-spectrum technology. In general, CDMA is used as an access method in many mobile standards such as CDMA2000, and WCDMA. We address the problem of blind multiuser equalization in the wideband CDMA system, in the noisy multipath propagation environment. Herein, we propose three new blind receiver schemes, which are based on the state space structures. This socalled blind state-space receivers (BSSR) does not require knowledge of the propagation parameters or spreading code sequences of the users but relies on the statistical independence assumption between the source signals. Also, we develop and derive three update-laws in order to enhance the performance of the blind detector. Additionally, we upgrade three semi-blind adaptive detectors based on the corporation between the RAKE receiver and the stochastic gradient algorithms which are used in several blind adaptive signal processing algorithms, namely FastICA, RobustICA, and principle component analysis PCA. Bit error rate (BER) simulations of these methods are shown for different number of users, signal to noise ratio (SNR) and different number of symbols per user in comparison with the Blind Multiuser Detectors (BMUD), Linear Minimum mean squared error (LMMSE) and other conventional detectors. The results show that the proposed algorithm outperforms the other detectors in estimating the symbol signals from the mixed CDMA received signals. Moreover, the new blind detectors mitigate the multi access interference (MAI) in CDMA.

6.1 Introduction

Code Division Multiple Access (CDMA) is a channel access method used by various radio technologies and it is based on spread-spectrum technology such as in thirdgeneration (3G) cellular telephony, terrestrial and satellite communications systems, and indoor wireless networks [1-2], [9]. Although, LTE (4G) is in operation in many cellular companies inside and outside the U.S., the networks are still not fully built out, and LTE coverage is still not universal. Thus, the most of the older 2G and 3G systems are still in charge or at least working in parallel with the 4G as in U.S. companies, AT&T and T-Mobile use GSM/WCDMA/HSPA while Verizon, Sprint, and MetroPCS use cdma2000/EV-DO. Moreover, The LTE wireless interface is incompatible with 2G and 3G networks, so that it must be operated on a separate wireless spectrum. Nevertheless, 3G is intended to be replaced by 4G technologies sooner or later, but it is going to take a long time before the LTE coverage is developed to be fully operated and universal, especially in some countries worldwide, such as India, Iraq ... etc. [141-142].

As any radio communication systems, CDMA systems consider as interference limited and it suffers from deferent types of interference, namely an internal multiple access interference (MAI) due to the non-ideal cross-correlations between users' spreading sequences, narrow-band interference, inter-symbol interference (ISI) and noise at the receiver. These drawbacks, in general, affect the performance of a CDMA system. In highly loaded systems, conventional detectors are considered unsuitable choice, since most of them suffer from the external interference sources such as adjacent channel interference or jamming, and treat the interference as an additional background noise. In practical, the primary source of interference is MAI in the CDMA system. This has motivated to the development of numerous interference rejection techniques to overcome the MAI and the near-far problem in the conventional receiver. Several state-of-art approaches are proposed in literature to overcome this challenge such as the trainedbased systems. Also, the most conventional detection for the CDMA signals are based on the second order statistics among user codes.

In CDMA system, Multiuser detection has been presented in several works in order to enhance channel capacity and mitigate multiple access interference (MAI). Multiuser detection was firstly established to obtain an optimum multiuser detector for multi-Gaussian channel in [1]. In additional, several suboptimum detectors have been proposed in [6-8], because of the computational complexity in the optimal detector which make it unrealistic. In [1] and [32-36], the training sequence techniques were used to present suboptimal detectors, i.e. adaptive linear detector and zero-forcing detector. In [6], they proposed suboptimal detector based on the Liner minimum mean square error (LMMSE) method. In [8], X. Wang and H. Poor proposed the blind MMSE and the blind de-correlating detectors. In [31-36], adaptive blind detectors were proposed based on the minimum output energy incorporating with constrained optimization methods. Several subspace approaches were proposed in in the literature, i.e. [20], [20], [36]. In [10], several types of group-blind linear detectors were proposed in order to enhance the performance for uplink and downlink channels. The key idea of these detectors was to

take advantage of the cross-correlation matrix which constructed by exploiting the correlation between the successive received signals. These detectors suffer from the computational complexity in order to be implemented. Also, they require information of timing and spreading waveform of the desired user.

The aforementioned techniques periodically enforce the user to send a training sequence that has to be known for the receiver in order to make the receiver able to estimate the parameter of the propagation channel which are caused by the multiple reflections of the radio waves on the obstacles encountered, i.e. buildings, cars, and trees etc. Furthermore, according to [42], it has been reported that 20% of the bandwidth is devoted to the training sequence in GSM and up to 40% in UMTS. In spite of the good performance of the aforementioned techniques, the cost tends to be significantly large in terms of bandwidth. However, adaptive signal processing techniques seemed to be more sufficient methods to CDMA systems in the presence of the high dynamic conditions due to the mobility of the mobile terminal, the short code case and the fortuitous of the channel access. In particular, blind adaptive techniques tend to be a hot topic for last decade in order to ensure the high communication rate due to its potential to eliminate/ reduce the training sets. Moreover, blind techniques are considered an attractive field so as to work besides the trained-based systems to 1) reduce the training sequence, 2) to help the trained-based systems to back up especially in fast-time varying channels and at severe multipath fading. Also, Blind techniques help to recover the signals in some other situations such as 1) eavesdropping, where using the training sequence is not possible or not available, 2) when the receiver fails to keep on tracks the desired user. However, the underlying user symbol sequences are usually mutually independent. Therefore, this is the key assumption that makes the CDMA system a suitable environment to take advantage of the blinds techniques, such as information maximization [1] and minimum mutual information [6]. Moreover, the adaptive LMMSE detector has been proposed to overcome the complex matrix inversion operation, but this detector still needs the spreading codes of all users. Therefore, the MMSE detector might not be realistic in the downlink receiver. Also, it might be insecure in the downlink case. However, it seems more suitable to work in the uplink case. Therefore, this chapter aims to recover the source symbol sequences from the linear convolutive received mixture without any knowledge of the user spreading codes and in the absence of channel identification. Simply put, this chapter proposes blind adaptive detections, based on state space structures using natural gradient method for multipath channels in CDMA systems. Three update-laws are derived based on the state space structures, and then, three blind statespace receivers (BSSR) are developed for MAI, ISI suppression and symbol estimations. The second contribution of this chapter is three semi-blind adaptive algorithms based on the corporation between the RAKE receiver and the stochastic gradient algorithms which are used in several blind adaptive signal processing algorithms, namely FastICA, RobustICA, and principle component analysis PCA. Nevertheless, this chapter explores a higher order statistics (HOS) in order to make the methods robust and secure against incomplete cross-correlation and a near-far problem which consider other drawback factors in conventional detection methods. The simulations are carried out to study and verify the effectiveness of the proposed methods for solving the symbols estimations. Moreover, we observe variations in the bit error-rate as a function of signal to noise ratio, number of users and number of symbols per user. Finally, a comparison is attempted between the proposed algorithms and conventional ones in terms of performance and their computational complexity.

Through this chapter, lower case letters denote scalars, bold lower case letters denote vectors, and bold upper case letters denote matrices. The following symbols are used to present our work:

- \succ (·)^{*T*} refers to transpose operator;
- \triangleright (·)^{*H*} refers to Hermitian transpose operator;
- $\blacktriangleright \qquad trace(\cdot) refers to the trace operator.$
- → $j = \sqrt{-1}$ refers to the imaginary unite.
- \blacktriangleright diag(·) refers to diagonal matrix;
- Sgn(·) refers to sign operator;
- \blacktriangleright $E[\cdot]$ refers to statistical expectation.

The remainder of the chapter is organized as follows. In Section II, a brief description and derivation of synchronous CDMA signal models in multi-path fading are presented. Section III proposes the new detection scheme based on State space approach. The comparative simulations results and conclusion are given in Section IV and Section V, respectively.



Figure 6.1: Signal generation model for a typical QPSK DS-CDMA system

6.2 DS-CMDA Signal Model

In this section, we briefly present the signal model for a DS-CDMA implementation using one layer of spreading codes only. In a typical synchronous DS-CDMA system employed for indoor ATM and certain ad hoc wireless networks [74], [75]. In a DS-CDMA system, several users share the medium simultaneously by using their own signatures. The simplest received signal model r(t) before filtering in a symbol interval, as in Figure 6.1, is given by [75]

$$r(t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \sum_{l=0}^{L} \alpha_{lm} b_{k,m} s_k (t - mT_b - d_l T_c) + n(t)$$
(6.1)

where

- l, k, m are path, user and symbol indices, respectively.
- α_{lm} is the path gain since in downlink model; the path gain does not differ among users because all users' signals are sent together and the path gain α_{lm} and propagation delay factor d_l depend only on the number of paths.
- b_{k,m} is symbol.

- s_k(.) is spreading code (chip sequence)
- d_l is the propagation delay factor, $d_{kl} \in \{0, 1, ..., \frac{C-1}{2}\}$
- C is the number of chips per symbol.
- T_b, T_c, t are time, symbol and chip duration, respectively.
- n(t) is an additive white Gaussian noise (AWGN) channel.

In this chapter, the system is assumed to be time-invariant which means that the channel parameters are much slower than the frequency of transmitted symbol data. However, let us assume that G is the number of code sequence, K is the number of users, and L is the number of channels. Thus, the vector form of the equation (6.1) will change to be as:

$$\mathbf{r} = \mathbf{HSb} + \mathbf{n} \tag{6.2}$$

Where r is the received vector signal, H is an $(G + L - 1) \times G$ matrix which represents the multipath propagation coefficients, S is an G x K block diagonal matrix, b is an K – d vector which represent the data symbols, n is the (G + L - 1) - d channel noise vector with covariance matrix Q. This model received signals (6.2) is suitable for deriving the linear symbols detectors such as the MF, the RAKE, the LMMSE and the blind Detectors based on FastICA and Robust ICA algorithms [2], [11], [15]. In additional, an alternative signal model is proposed in [80], [81] as a linear convolutive model is given by:

$$r_n = H_0 b_n + H_1 b_{n-1} + n_n = H b_n + n_n$$
(6.3)

where

- r_n is the received signal vector;
- $\mathbf{b}_n = [\mathbf{b}_1(n), \dots, \mathbf{b}_K(n)]^T$ is a current bits of all users;

H₀ = [h₁, ..., h_k] is signature matrix of the current bits of all users including MAI;

$$h_{k} = \begin{bmatrix} 0 & -1 \\ h_{k}(0) & -1 \\ -1 & -1 \\$$

H₁ = [h
₁, ..., h_k] is the signature matrix of the previous bits of all users including ISI;

where

$$\bar{\mathbf{h}}_{k} = \begin{bmatrix} \mathbf{h}_{k}(\mathbf{N} - \mathbf{d}_{k}) \\ \cdot \\ \cdot \\ \mathbf{h}_{k}(\mathbf{N} + \mathbf{M} - 1) \\ \mathbf{0} \end{bmatrix}$$

- $H = [H_0; H_1]$ is the signature matrix of all users;
- $\mathbf{b}_n = [\mathbf{b}_1(n), \dots, \mathbf{b}_K(n)]^T$ are currents bits of all user;
- $b_{n-1} = [b_1(n-1), ..., b_K(n-1)]^T$ are previous bits of all users;
- $\bar{\mathbf{b}}_n = [\mathbf{b}_n^T; \mathbf{b}_{n-1}^T]^T$ are bits of all users;
- $n_n = [n(nN), ..., n(nN + N 1)]^T$ is independent white Gaussian Noise vector.

In uplink (asynchronous) CDMA systems, one can assume that H_0 and H_1 are mutually independent, therefore H is a full column matrix and its rank is 2K as it's shown in [87], [89]. Whereas for downlink (synchronous) CDMA communication, H is a matrix and its rank is full-rank with hardly restricted as in [91]. The main focus in this chapter is on the synchronous CDMA communication system although our proposed algorithms are working well in asynchronous CDMA system [92], [88].

6.3 Conventional Blind Linear Multiuser Detectors

In this section, we briefly describe the conventional linear multiuser detectors such as the Match Filter (MF), the Rake receiver and the LMMSE detector in multipath environment. In additional, we briefly describe the Blind-Group detectors in [g] which include the blind decorrelation detector and Blind linear hybrid detector. For more details on these linear detectors, see [74], [86-95].

6.3.1 Single user detection (SUD) Detector

The SUD is a standard MF detector which exploits the user's signatures to make the best estimation of the user's sequences from the data received at MS. This detector is completely ignoring the MAI due to other users sharing the resources [74]. One can express the MF Detector for ith user in DS-CDMA system as follows:

$$b_{i,MF}^D = S_i^H x \tag{6.4}$$

where $S_i = diag(\bar{s}_i, \bar{s}_i, ..., \bar{s}_i)$, $\bar{s}_i = [0 \ 0 \ ... \ s_i \ ... \ 0]$. s_i is the ith user's signature code, x is received data, and $b_{i,MF}^D$ is the estimated DS-CDMA symbol vector, see [74].

6.3.2 Rake Detector

Perhaps, the most special case of linear multiuser detection is the Rake Detector which consists of multiple chip-delayed SUD fingers in parallel. In this chapter, we implement the Rake receiver with the knowledge of both channel delays and channel coefficients. However, one can express the RAKE for DS-CDMA system mathematically as follows:

$$b_{i,RAKE}^{D} = S_{i}^{H} \overline{H}^{H} x \tag{6.5}$$

Where *H* is the estimated channel matrix, and $b_{i,RAKE}^{D}$ is the estimated ith user's symbol;

6.3.3 LMMSE Detector

Despite the fact that the conventional linear detectors based on the Least Square (LS), Zero-Force (ZF) and BLUE algorithms perform poor especially in colored noise presence, the LMMSE detector is considered as one of the best linear detector for DS-CDMA system [74], [75]. However, one can express the LMMSE as follows:

$$b_{i,LMMSE}^{D} = S_i^{H} \overline{H}^{H} (\sigma^2 \overline{H} \overline{H}^{H} + \overline{Q})^{-1} x$$
(6.6)

Where $(\sigma^2 \overline{H} \overline{H}^H + \overline{Q}) = R = E[xx^H]$ is the auto-correlation of the received data at the MS, and σ^2 is the average power of the transmitted power. There are several drawbacks in implementation the LMMSE receiver; however, the main drawback is that the computation of the auto-correlation *R* is very expensive. One can use the eigendecomposition instead of inverting the auto-correlation matrix *R* as follows [5], [74]:

$$b_{i,LMMSE}^{W} = S_i^H \bar{H}^H (V_s D_s^{-1} V_s^H) x$$
(6.7)

Where V_s is the estimated Eigen-vectors of the auto-correlation matrix R, and D_s is the corresponding eigen-values of the auto-correlation matrix R. Additionally, one can use adaptive algorithms to estimate the LMMSE user's symbols as in [135].

6.4 The Proposed Detection Schemes Based On state space framework

In this section, a new blind detection strategy is proposed, based on the state space framework [75]. We proposed the three blind multiuser detectors based on feed-forward structure, feedback structure I, and feedback structure II, respectively.

Here, one needs to first recall the received signal model (3)

$$\mathbf{r}_{n} = \mathbf{H}_{0}\mathbf{b}_{n} + \mathbf{H}_{1}\mathbf{b}_{n-1} + \mathbf{n}_{n}$$

The aim of this chapter is to detect the b symbol vector from the received data r under the following assumptions:

- AS1) the GxK matrices, H_0 , H_1 is of full column rank.
- AS2) the symbol signals, b, are white, independent and identically distributed (i.i.d)
- AS3) the Additive Noise vector is white and independent of source signals
- AS4) the power of the transmitted symbol signals, normalized to be unity.
- AS5) the maximum lag in the entire multipath channels (max(τ₁)) is smaller than the spreading gain of the CDMA G.
- AS6) the CDMA system is not over-saturated, which means the number of users (K) is less the the number of the spreading gain (G).
- AS7) the channel is assumed to be a slowly fading wide sense stationary.

Each method involves two steps: First, a preprocessing stage. Second, the rotation stage based on the state space structures. In the next subsection, we will explain the preprocessing stage (whitening processes), and then we will derive the three methods based on each state space structure in individual subsections.

6.4.1 Step1: Preprocessing (Data Whitening)

In the preprocessing step, the symbol signals are detected up to a unitary matrix using the second order statistic (SOS). This step was used to reduce the noise and to eliminate redundancy in the data. Under the Assumptions AS1, AS2, AS3 and AS4, the GxG covariance matrix (C) of the noiseless transmitted signals can be expressed by

$$C = E[rr^{H}] - \sigma^{2}I_{NG}$$
(6.8)

Without loss of generality, we consider a simpler two tap models then we will generalize them using induction techniques. Therefore, substituting r in (6.8), under our assumptions AS1-AS7, one can expressed the covariance matrix C as follows

$$C = H_0 E[bb^H] H_0^H + H_1 E[bb^H] H_1^H = H_0 H_0^H + H_1 H_1^H$$
(6.9)

Under AS1, the $H_0H_0^{H}$ and $H_1H_1^{H}$ can be decomposed, respectively as

$$C_0 = V_0 \Lambda_0 V_0^{H} \tag{6.10}$$

$$C_1 = V_1 \Lambda_1 V_1^{H} \tag{6.11}$$

where V_0 and V_1 are an KxK matrix satisfying

$$V_0 V_0^{\ H} = V_0^{\ H} V_0 = I_K \tag{6.12}$$

$$V_1 V_1^{\ H} = V_1^{\ H} V_1 = I_K \tag{6.13}$$

and Λ_0 and Λ_1 are an KxK diagonal matrix containing significant eigenvalue entries. So, from aforementioned equation (6.3), the GxG H₀ and H₁ matrices will be represented respectively as

$$H_0 = V_0 \Lambda_0^{-\frac{1}{2}} U_0^{H}$$
(6.14)

$$H_1 = V_1 \Lambda_1^{-\frac{1}{2}} U_1^{H}$$
(6.15)

where U_0 is a KxK full rank unitary matrix and $U_0U_0^T = I_K$ and U_1 is a KxK full rank unitary matrix and $U_1U_1^T = I_K$. However, the whitening step obtained matrix V_0 and V_1 so that the Kx1 whitened data vector r_n^w has a covariance of the identity matrix, $C = I_K$, which can be obtained as follows:

$$r_{n}^{w} = \Lambda_{0}^{-\frac{1}{2}} V_{0}^{H} r_{n} + \Lambda_{1}^{-\frac{1}{2}} V_{1}^{H} r_{n}$$
(6.16)

Therefore,

$$r_{n}^{w} = U_{0}^{H}b_{n} + U_{1}^{H}b_{n-1} + \left(\Lambda_{0}^{-\frac{1}{2}}V_{0}^{H} + \Lambda_{1}^{-\frac{1}{2}}V_{1}^{H}\right)n_{n}$$
(6.17)

The transmitted symbols can be recovered based on the state space structures. However, after the preprocessing step, the detection of the symbol signal \hat{b}_n reduces to determining the KxK unitary matrices U_k (rotation matrices). Next, the derivations for the three proposed algorithms, based on feedforward structure, feedback structure I and feedback structure II, respectively, are presented.

6.4.2 Step 2a: Determining the rotation matrix (unitary matrix) U based on the feedforward structure.

The output from the feedforward structure is given in [75] as follows:

$$y_n = U_0 r_n^w + \sum_{k=1}^K U_k r_{n-k}^w$$
(6.18)

Again, we consider a simpler two tap models, thus the two tapes of the FeedForword model represents as

$$y_n = U_0 r_n^w + U_1 r_{n-1}^w \tag{6.19}$$

However, one can re-write the previous convolutive filter in the following augmented form

$$\begin{bmatrix} y_n \\ r_{n-1}^w \end{bmatrix} = \begin{bmatrix} U_0 & U_1 \\ 0 & I \end{bmatrix} \begin{bmatrix} r_n^w \\ r_{n-1}^w \end{bmatrix}$$
(6.20)

Let's define that

$$\widetilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{y}_n \\ \mathbf{r}_{n-1}^w \end{bmatrix}$$

$$\widetilde{\mathbf{U}} = \begin{bmatrix} \mathbf{U}_0 & \mathbf{0} \\ \mathbf{U}_1 & \mathbf{I} \end{bmatrix}$$

$$\widetilde{\mathbf{R}} = \begin{bmatrix} \mathbf{r}_{n}^{\mathbf{W}} \\ \mathbf{r}_{n-1}^{\mathbf{W}} \end{bmatrix}$$

So, the expression in (15) becomes as

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{U}}^{\mathrm{T}} \widetilde{\mathbf{R}} \tag{6.21}$$

Based on ICA algorithm [2], [13], [14], the update law for the weight column of de-mixing matrix Ũ, we have

$$u^{+} = u - \mu E \left[\widetilde{R} \left(G \left(u^{T} \widetilde{R} \right) \right) \right]$$
(6.22)

Where u is the column vector of $\widetilde{U},\,\mu$ is the step size and G is the score function. However,

$$\mathbf{u} = \begin{bmatrix} \mathbf{U}_0 \\ \mathbf{U}_1 \end{bmatrix} \tag{6.23}$$

Then

$$\begin{bmatrix} u_0^+ \\ u_1^+ \end{bmatrix} = \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} - \mu \begin{bmatrix} r_n^w \\ r_{n-1}^w \end{bmatrix} G(y_n)$$
(6.24)

Where u_0 , u_1 are the column vector of U_0 and U_1 , respectively. Therefore, the update laws for the individual columns are

$$u_0^{+} = u_0 - \mu r_n^w G(y_n) \tag{6.25}$$

$$u_1^{+} = u_1 - \mu r_{n-1}^{w} G(y_n)$$
(6.26)

By induction, the update law for the kth lag element ukis

$$u_{k}^{+} = u_{k} - \mu r_{n-k}^{w} G(y_{n})$$
(6.27)



Figure 6.2: Feedback Demixing Structure I

6.4.3 Step 2b: Determining the rotation matrix (unitary matrix) U based on the feedback structure I.

The output of the feedback structure I is given in Figure 6.2 as follows:

$$y_{n} = U_{0}^{-1} \left(r_{n}^{W} + \sum_{k=1}^{K} U_{k} y_{n-k} \right)$$
(6.28)

Consider two tapes of the Feedback Configuration I model

$$y_n = U_0^{-1} (r_n^w + U_1 y_{n-1})$$
(6.29)

However, one can re-write the previous convolutive filter in the following augmented form

$$\begin{bmatrix} r_n^{\mathsf{W}} \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} U_0 & U_1 \\ 0 & I \end{bmatrix} \begin{bmatrix} y_n \\ y_{n-1} \end{bmatrix}$$
(6.30)

Or

$$\begin{bmatrix} y_n \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} U_0 & U_1 \\ 0 & I \end{bmatrix}^{-1} \begin{bmatrix} r_n^w \\ y_{n-1} \end{bmatrix}$$

$$\begin{bmatrix} y_{n} \\ y_{n-1} \end{bmatrix} = \frac{1}{\det(U_{0})} \begin{bmatrix} I & -U_{1} \\ 0 & U_{0} \end{bmatrix} \begin{bmatrix} r_{n}^{W} \\ y_{n-1} \end{bmatrix}$$
(6.31)

Let's define that

$$\widetilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n-1} \end{bmatrix}$$
$$\widetilde{\mathbf{W}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{U}_1 & \mathbf{U}_0 \end{bmatrix}$$
$$\widetilde{\mathbf{R}} = \begin{bmatrix} \mathbf{r}_n^w \\ \mathbf{y}_{n-1} \end{bmatrix}$$

So, the previous augmented expression becomes

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{U}}^{\mathrm{T}} \widetilde{\mathbf{R}} \tag{6.32}$$

Based on ICA algorithm [2], the update law for the weight column of de-mixing matrix \tilde{U} , we have

$$u^{+} = u - \mu E \left[\widetilde{R} \left(G(u^{T} \widetilde{R}) \right) \right]$$
(6.33)

Where u is the column vector of $\widetilde{U},\,\mu$ is the step size and G is the score function. However,

$$\mathbf{u}^{0} = \begin{bmatrix} \mathbf{I}^{0} \\ -\mathbf{U}_{1}^{0} \end{bmatrix} \tag{6.34}$$

The update law is

$$\begin{bmatrix} i^+\\ u_1^+ \end{bmatrix} = \begin{bmatrix} i\\ u_1 \end{bmatrix} - \mu \begin{bmatrix} r_n^w\\ y_{n-1} \end{bmatrix} G(r_n^w - u_1 y_{n-1})$$
(6.35)

And

$$\mathbf{u}^{1} = \begin{bmatrix} \mathbf{0}^{1} \\ \mathbf{U}_{\mathbf{0}}^{-1} \end{bmatrix} \tag{6.36}$$

Then

$$\begin{bmatrix} 0^+\\ u_0^+ \end{bmatrix} = \begin{bmatrix} 0\\ u_0 \end{bmatrix} - \mu \begin{bmatrix} r_n^w\\ y_{n-1} \end{bmatrix} G(u_0 y_{n-1})$$
(6.37)

The update laws for the individual columns are

$$u_0^+ = u_0 - \mu y_{n-1} G(u_0 y_{n-1})$$
(6.38)

And

$$u_1^{+} = u_1 - \mu y_{n-1} G(r_n^w - u_1 y_{n-1})$$
(6.39)

By induction, the update law for the kth lag element u_k is

$$u_{k}^{+} = u_{k} - \mu y_{n-k} G(r_{n}^{w} - u_{k} y_{n-k})$$
(6.40)



Figure 6.3: Feedback Demixing Structure II

6.4.4 Step 2c: Determining the rotation matrix (unitary matrix) U based on the feedback structure II.

The output of the feedback structure II is given in Figure 6.3 as follows:

$$y_n = U_0 r_n^w + \sum_{k=1}^K U_k y_{n-k}$$
(6.41)

Again, consider two tapes of the Feedback structure II model

$$y_n = U_0 r_n^w - U_1 y_{n-1} \tag{6.42}$$

However, one can re-write the previous convolutive filter in the following augmented form

$$\begin{bmatrix} y_n \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} U_0 & -U_1 \\ 0 & I \end{bmatrix} \begin{bmatrix} r_n^W \\ y_{n-1} \end{bmatrix}$$
(6.43)

Let's define that

$$\begin{split} \widetilde{\mathbf{Y}} &= \begin{bmatrix} \mathbf{y}_n \\ \mathbf{y}_{n-1} \end{bmatrix} \\ \widetilde{\mathbf{W}} &= \begin{bmatrix} \mathbf{U}_0 & \mathbf{0} \\ -\mathbf{U}_1 & \mathbf{I} \end{bmatrix} \\ \widetilde{\mathbf{R}} &= \begin{bmatrix} \mathbf{r}_n^{\mathbf{W}} \\ \mathbf{y}_{n-1} \end{bmatrix} \end{split}$$

So, the previous expression becomes

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{U}}^{\mathrm{T}} \widetilde{\mathbf{R}} \tag{6.44}$$

Based on ICA algorithm [2], the natural gradient update laws for the weight column of de-mixing matrix \tilde{U} , we have

$$u^{+} = u - \mu E \left[\widetilde{R} \left(G (u^{T} \widetilde{R}) \right) \right]$$
(6.45)

Where u is the column vector of $\widetilde{U},\,\mu$ is the step size and G is the score function. However,

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_0 \\ -\mathbf{u}_1 \end{bmatrix} \tag{6.46}$$

Then

$$\begin{bmatrix} u_0^+ \\ u_1^+ \end{bmatrix} = \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} - \mu \begin{bmatrix} r_n^w \\ y_{n-1} \end{bmatrix} G(y_n)$$
(6.47)

The update laws for the individual columns are

$$u_0^{+} = u_0 - \mu r_n^w G(y_n) \tag{6.48}$$

And

$$u_1^{+} = u_1 - \mu y_{n-1} G(y_n) \tag{6.49}$$

By induction, the update law for the kth lag element ukis

$$u_{k}^{+} = u_{k} - \mu y_{n-k} G(y_{n}) \tag{6.50}$$

Algorithm 6.1: RAKE based FastICA method

Input: $(M \ x \ T)$ matrix of realization \mathbf{r} , Initial demixing matrix $\mathbf{W} = \mathbf{I}_{\mathbf{G}}$, number of iterations Itr, Step Size γ i.e. $\gamma = 0.3$, \mathbf{H} is the estimated channel matrix, $\mathbf{g}(\mathbf{y}) = \mathbf{y}^3$

Perform Pre-Whitening $\{r = V * r = \Lambda^{((-1) / 2)} E^{T} r\},$ For loop: for each $i = 1 \dots N$ $R = WH^{H}r(:, i)$ $W = E\{[g(WR)]^{T}\} - E\{g'(WR)\}W$

W=W/norm(W)

$$\boldsymbol{b}_{i,ICA}^{D}(:,i) = \boldsymbol{S}_{i}^{H} \mathbf{W} \overline{\boldsymbol{H}}^{H} \boldsymbol{r}$$

End For Output: the estimated Symbols \boldsymbol{b}_{ICA}^{D}

6.4.5 The proposed adaptive detectors

In this section, we develop three adaptive detectors based the independent component Analysis (ICA), Robust ICA and Principle Component Analysis (PCA). Having the RAKE receiver structure in (10), one can express the adaptive weight RAKE for DS-CDMA system mathematically as follows:

$$\boldsymbol{b}_{i,RAKE}^{D} = \boldsymbol{S}_{i}^{H} \boldsymbol{W} \boldsymbol{H}^{H} \boldsymbol{r}$$
(53)

Where **H** is the estimated channel matrix, $\boldsymbol{b}_{i,RAKE}^{D}$ is the estimated *ith* user's symbol; and **W** is an *G* x *G* weighting matrix. However, we present *Algorithms 6.1, 6.2 and 6.3* to estimate the matrix **W** adaptively based on the FastICA, Robust ICA and PCA algorithms, respectively.

Algorithm 6.2: RAKE based RICA method

Input: $(M \ x \ T)$ matrix of realization r, Initial demixing matrix $W = I_G$, number of iterations Itr, Step Size γ i.e. $\gamma = 0.3$, H is the estimated channel matrix, here g is the gradient of the Kurtosis contrast K(.)Perform Pre-Whitening $\{r = V * r = \Lambda^{((-1)} / 2) E^{T} r\}$, For loop: for each $i = 1 \dots N$ $R = WH^{H}r(:, i)$ $\Delta W = (I_G - R * R^{H})W$ $\mu_{opt} = \arg \max_{\mu} |K(y + \mu g)|$ $W = W + \mu_{opt} \Delta W$ W = W/norm(W) $b_{i,RICA}^{D}(:, i) = S_i^{H}W\overline{H}^{H}r$ End For Output: the estimated Symbols b_{RICA}^{D}

6.5 Simulation Results

In this section, a series of simulations are carried out in order to verify the proposed algorithms in downlink DS-CDMA system in the presence of AWGN. We assume a constant spreading gain, which is NG = 63 for Gold Code. The received CDMA signal is taken in five multipath channels L = 5 with delays of {0,1,2,3,4} chips, respectively. Also, we use the complex attenuation coefficients to represent the multipath channels, which are $h_0 = 0.25 + j0.18$, $h_1 = 0.21 + j0.14$, $h_2 = 0.18 + j0.11$, $h_3 = 0.14 + j0.11$, and $h_4 = 0.11 + j0.07$, respectively. We use the following model function for sub Gaussian sources which the source signals are having a negative kurtosis sign.

$$G^{SUB}(\hat{b}) = \hat{b} - (tanh(Re{\hat{b}}) + jtanh(Im{\hat{b}}))$$

Monte Carlo Simulation was run to verify the validity of the algorithm simulations. Also, we use the signal to noise ratio (SNR) as a figure of merit which merely represents the ratio of the energy per bit and the power spectral density (PSD) of the noise. Moreover, all the user symbols are assumed to be transmitted with the same power.

Figure 6.4 (a) and (b) sow the simulation results of SNR vs. BER for three proposed detectors regarding to other ones for number of users K = 30 and K = 50, respectively. The other parameters were set as: Number of symbols M=1000, Number of paths L = 5, with various values of SNR -10 dB to 30dB. Rake based on RobustICA algorithm is used with these parameters: the source Kurtosis signs is considered to be maximize absolute normalized kurtosis for all sources, 1e - 3 is used to be the threshold for statistically-significant termination test, the maximum number of iterations per extracted source is 1000, prewhitening (via SVD of the observed data matrix); orthogonalization deflation type is used and the extracting vectors initialization is an identity matrix of suitable dimensions.







Figure 6.4: Average BER as a function of SNR for DS-CDMA downlink. Using Gold codes G=63. (a) Using 30 users (b) Using 50 users

In Figure 6.4, the proposed algorithms improve the performance of the CDMA system; blind multiuser detection based on the second feedback structure has given the lowest BER regards to the others, and outperforms the performance of the other detectors. We also observe that the proposed algorithms work in the high SNR ratio,

which most likely cause ill-posed problem for LMMSE receiver especially when the sample set T is fairly small. Moreover, the performance of the blind multiuser detection degrades as the number of the user increases as shown in Fig. 6.4 (b).

Also, to complete our discussion, Figure. 6.5 shows the performance of all the proposed blind multiuser equalizers in terms of computational complexity. It shows the average of the corresponding CPU time of each proposed method. However, the price to pay for the enhancement in the BER performance is represented by the computational complexity in terms of CPU times. We also study the effect of OVSF codes in Fig. 6.6. Generally, Figure. 6.6 shows that using the OVSF codes enhances the performance of the proposed methods.



Figure 6.5: Corresponding CPU time for each method.



Figure 6.6: Average BER as a function of SNR for DS-CDMA downlink. Using OVSF codes G=64. (a) Using 30 users (b) Using 50 users

In the WCDMA System, we assume that the channel coefficients are $h_0 = 0.25 + j0.18$, $h_1 = 0.21 + j0.14$, $h_2 = 0.18 + j0.11$, $h_3 = 0.14 + j0.11$, and $h_4 = 0.11 + j0.07$,

respectively, the bandwidth of channel is 1.25 mega chips per second (MCPs), all userspecific codes use two types of spreading codes, namely, Gold codes with a spreading gain G=63, and OVSF or (Walsh-Hadamard) codes with spreading gain G=64. The long scrambling code has a frame-length of 10 ms. In Figure 6.7, we demonstrate the performance of the various methods in terms of BER for the WCDMA downlink scenario.

We observe that the LMMSE is slightly better some presented detectors under the good SNR conditions. But the proposed algorithm based on the second feedback structure outperforms all detectors again at all SNR and has given the lowest BER regards to the others.






(b)

Figure 6.7: Average BER as a function of SNR for WS-CDMA downlink. Using Gold codes G=63. (a) Using 30 users (b) Using 50 users



Figure 6.8: Average BER as a function of SNR for WS-CDMA downlink. Using OVSF codes G=64. (a) Using 30 users (b) Using 50 users

It is also worthwhile to compare the presented algorithms with a large data sample set. Thus, Figure 6.8, Figure 6.9 and Figure 6.10 are the performance of the various detectors with fairly long sample M=3000 in the DS_CDMA and WCDMA systems, respectively. A plausible notice, the LMMSE detector gets better than other detectors

under good SNR conditions. But, still the proposed algorithm based on the feed backward second configuration has exceeded the LMMSE detector at all SNRs less than 22dB.



Figure 6.9: Average BER as a function of SNR for DS-CDMA downlink. For 30 users (a) Using Gold codes G=63. (b) Using OVSF codes G=64.







Figure 6.10: Average BER as a function of SNR for WCDMA downlink. For 30 users (a) Using Gold codes G=63. (b) Using OVSF codes G=64

Finally, we study the effect of the size of the sample set and the number of users on the performance of the proposed method in Figure 6.11 and Figure 6.12, respectively. In Figure 6.11, the simulation results show the BER vs. SNR with various K users at 500 symbols for each user for blind multiuser detection based on the second feedback structure detector. It shows that the proposed detector perform less improvement of the performance as K increases. Thus, Figure 6.12 shows the simulation results of BER vs. SNR with 30 users (K=30) for various data samples (M). Although the proposed algorithm seems robust for sample sets and performs well, it is obvious that the proposed algorithm also improves more consistently in the performance as M increases and mitigates the MIA. Overall, the proposed algorithm outperforms other algorithms in most cases and performs better to solve the symbol estimation problem in DS/WCDMA downlink system, especially when the size of the sample set is relatively small.



Figure 6.11: Average BER as a function of SNR for various number of users K



Figure 6.12: Average BER as a function of SNR for various sample sets M

6.6 Conclusion

This chapter carried out both simulation and theoretical demonstrations of the blind multiuser detector based on the space state structures in the CDMA system. Also, we develop the three blind multiuser detectors based on the three adaptive algorithms; namely ICA, RICA and PCA. The results appear to show that the proposed algorithm perform well in the symbol estimation problem in DS/CDMA systems and outperform the other conventional detectors and the Adaptive MMSE. Our results also show that the Multiple Access Interference (MAI) can be mitigated by the proposed algorithm, thus improving the performance of blind multiuser detection. Although the proposed method improves as the size of the sample set increases, the results show the proposed detector performs well even though the sample sets are small, unlike the LMMSE detector. Moreover, unlike the complexity of the LMMSE detector, the complexity of the proposed methods, being a constant, didn't increase exponentially. Finally, the proposed method, unlike the adaptive LMMSE detector, has no restriction about the spreading codes since they do not require the spreading codes of the interfering users. Therefore, it is a more suitable choice in the downlink case and it does work in the uplink case as well.

7 Chapter 7

Constrained Blind Multiuser Detection for DS-CDMA System

In direct sequence code division multiple accesses DS-CDMA communication system, the blind multiuser detection is presented for enhance the computational complexity and mitigate the multiple access interference (MAI) in the detector. The illcondition of the covariance matrix of the received signals degrades the performance of the linear minimum mean-squared error LMMSE detector. Especially, when the Signal to noise ratio is high and small data set is available for covariance matrix estimation. In this chapter, we introduce a constrained blind multiuser detection in order to improve its performance with imposing the regularization parameter to cope the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation. Through simulation results, we show that the proposed method improves the performance of the blind multiuser detection and outperforms the conventional multiuser detections.

7.1 Introduction

Multiuser Detection has been one of the significant topics in communication system for past decades because of its potentials to suppress the multiple access interference (MAI) efficiently in CDMA systems. Recently, significant attention has been given on multiuser detection in blind manner, which only requires a prior knowledge about of the signature sequence and timing of desired user [74-79].

In CDMA system, Multiuser detection has been presented in several works in order to enhance channel capacity and to mitigate multiple access interference (MAI). Multiuser detection was firstly established to obtain an optimum multiuser detector for multi-Gaussian channel in [76]. In additional, several suboptimum detectors have been proposed in [76] and [77], because of the computational complexity in the optimal detector which make it unrealistic. In [76], [81] and [82], the training sequence techniques were used to present suboptimal detectors, i.e. adaptive linear detector and zero-forcing detector. In [75], they proposed suboptimal detector based on the Liner minimum mean square error (LMMSE) method.

In general, the ill-posed linear equation problem has been arisen in blind multiuser detection through the ill- covariance matrix which degrades the performance of the LMMSE detectors [76], [81]. This problem affects the MUD when small observed numbers of symbol users are considered, especially, in high signal to noise ratio environment. For example, when a small data set transmits over slowly time varying channels, we use a small data blocks within the channel coherence time. Therefore, several works have suggested to use regularization techniques, i.e. [76], [80] to deal with the ill-posed problem in order to gives a stable solution of the blind multiuser detection.

This chapter develops a Blind linear DS-CDMA detection technique, based on minimum of energy output. The key idea of the proposed detector is to improve the conventional blind detector by imposing a new constraint on the cost function and add a regularization parameter to the covariance matrix to avoid the singularity especially in the presence of high SNR environment [74]. The main focus of this chapter is to study the effect of the new constraint and regularization parameter in the presence of the high additive Gaussian noise AWGN for long and small data sets. Furthermore, we study the robustness of the proposed detector due to see the effect of the mismatch between the original code sequences and estimated ones on the performance of the proposed detector, which most likely happens in multipath channels. The simulations are carried out to observe variations in the bit error-rate as a function of signal to noise ratio, number of users and number of symbols per user. Furthermore, the performance of the proposed algorithm and subspace blind multiuser detection [87] in terms of their performance. The remainder of the chapter is organized as follows. In Section II, a brief description and derivation of synchronous CDMA signal models are presented. Section Scheme. The comparative simulations results and conclusion are given in Section VI and Section VII, respectively.

7.2 DS-CMDA Signal Model

Assume we have a downlink (synchronous baseband) DS-CDMA system with K users. At the receiver, the sampled received signal during the ith symbol through the match filter with chip rate T_c is given in vector form by [75]

$$\mathbf{r}(i) = \sum_{k=1}^{K} A_k b_k[i] s_k + n(i) \quad \forall i = 0, 1, 2, \dots, L-1$$
(7.1)

where

 $r(i) = [r_1(i), r_2(i), ..., r_N(i)]^T$, where N denotes the number of processing gain.

• A_k is the received amplitude

- b_k[i] ∈ {±1} are the ith transmitted signal and are assumed to be independent
- s_k is the normalized signature waveform of the kth user.
- n[i] is the additive white Gaussian noise (AWGN) and assumed to have zero mean and covariance matrix equals $\sigma^2 I_N$, where I_N is an N identity matrix.
- L is the number of user symbol

Without loss of generality, we assume that the signature waveforms s_k are linearly independent and the noise n[i] is independent of user data. However, the covariance matrix of received signal { $C = E[rr^T]$ } is given by:

$$C = E[rr^{T}] = \sum_{k=1}^{K} A_{k}^{2} s_{k} s_{k}^{T} + \sigma^{2} I_{N}$$
(7.2)

According to [74] the decision output of the linear detector for ith transmitted symbol of the user one can be described as a weight vector $m_1 \in \mathbb{R}^N$, however, it is given by

$$\hat{\mathbf{b}}_1(\mathbf{i}) = \operatorname{sgn}\left(\mathbf{m}_1^{\mathrm{T}}\mathbf{r}(\mathbf{i})\right) = \operatorname{sgn}(\mathbf{r}^{\mathrm{T}}(\mathbf{i})\mathbf{m}_1) \tag{7.3}$$

Let's denote that R = [r(0), r(1), ..., r(L-1)] and $b_L = [b_1(0), b_1(1), ..., b_1(L-1)]$

1)], so equation (3) becomes in vector form as

$$\boldsymbol{R} \cdot \boldsymbol{m_1} = \boldsymbol{b_L} \tag{7.4}$$

7.3 Conventional Blind Multiuser Detection

Lespite the fact that the conventional linear detectors based on the Least Square (LS), Zero-Force (ZF) and BLUE algorithms perform poor especially in colored noise

presence, the LMMSE detector is considered as one of the best linear detector for DS-CDMA system [74] –[85]. However, one can express the LMMSE as follows:

$$m_{opt} = \operatorname{argmin}\{E[\|\mathbf{b}_{k} - \mathbf{m}^{H}\mathbf{r}\|^{2}]\}$$
(7.5)

So, the MOE detector is given by [1], [7] as follow:

$$m_1 = (s_1^T C^{-1} s_1)^{-1} C^{-1} s_1$$
 (7.6)
Subject to $m^H s_k = 1$

According to literature [74], the ideal bit error rate (BER) of Minimum of Energy detector (MOE) under the constraint $m^{H}s_{k} = 1$ is given by following the approximation

$$P_{e} \approx Q\left(\frac{A_{1}}{\left(\sigma\sqrt{m_{1}^{T}m_{1}}\right)}\right) = Q\left(\frac{A_{1}}{\sigma\|m_{1}\|_{2}}\right)$$
(7.7)

where

$$Q(x) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_{x}^{\infty} \exp\left(\frac{-x^{2}}{2}\right) dx$$
(7.8)

However, in practice, the covariance matrix is computed using the observation value as follows:

$$C \approx \tilde{C} = \sum_{i=1}^{L} r[i]r^{T}[i]$$
(7.9)

As it has been reported in literature, the detector suffers from the ill-conditioned problem, especially, when the SNR is high and the data sample set is small, which cause degrading the performance of the detector.

Whilst the covariance matrix \tilde{C} sometimes in practice becomes in the illconditioned matrix, as a result, the inversion of the ill-conditioned matrix will lead to illposed problem, which degrades the robustness of the detector. Several methods have been proposed to avoid this ill-posed problem in literature, the most two important methods are subspace decomposition algorithm and Regularization method, which have already been proposed in [76] and [79], respectively. Here, we impose a new constraint to speed up the convergence and improve the performance; also, we use a new regulation rule to avoid the ill-posed problem. The detector based on the subspace decomposition algorithm (SBMUD) is given by [74]-[79]

$$m_{SBMUD} = \left[s_k^H [U_s D_s^{-1} U_s^H]^{-1} s_k\right]^{-1} [U_s D_s^{-1} U_s^H]^{-1} s_k$$
(7.10)

SBMUD can solve the ill-posed problem by dividing the ill-conditioned matrix into two subspaces, which are the signal subspace and the noise subspace. The main problem in this method is the performance of the detector totally depends on the estimation covariance, which increases the chance to degrade the performance of the detector if the estimation covariance matrix presents a large deviation from the ideal covariance.

Li Hu, in [76], applies Tikhonov regularization [80] to mitigate the ill-posed algorithm. It is well-known that the regularization is effective way to avoid the ill-conditioned matrix. The detector based on Tikhonov is given by

$$m_{\text{regular}} = \left[s_k^{\text{H}} [\tilde{C} + \alpha I]^{-1} s_k \right]^{-1} [\tilde{C} + \alpha I]^{-1} s_k$$
(7.11)

Where α is the regularization parameter, note that the regularization method just adds energy constraint to boosting the covariance matrix to be well-conditioned matrix. In [76], they use two rule for α based on Tikhonov [80] which are a) $\alpha = m. tr(\tilde{C})$ where m is a positive constant and tr(·) represent the trace of estimation covariance matrix \tilde{C} and b) $\alpha = c. \lambda_{max}$ where c is a positive constant and the λ_{max} is a maximum eigenvalue of estimation covariance matrix \tilde{C} .

7.4 The Proposed Detection Scheme

In this section, a new blind detection strategy is proposed, based on the minimum of energy of the output [74]. Here, one needs to first recall the received signal model (3)

$$\mathbf{r} = \mathbf{H}\mathbf{b} + \mathbf{n} \tag{7.12}$$

Without loss of generality, we assume that the first user to be the desired user. However, one can estimate the desired user as a weight vector $w \in \mathbb{R}^{N}$. Therefore, the output is given by:

$$\mathbf{y} = \mathbf{w}^{\mathrm{H}}\mathbf{r} \tag{7.13}$$

The output power $E[y^2]$ can be expressed as follows

$$\mathbf{E}[\mathbf{y}^2] = \mathbf{w}^{\mathrm{H}} \mathbf{R} \mathbf{w} \tag{7.14}$$

where $R = E[rr^{H}]$ is the covariance matrix, and E[.] is expectation operator. In order to avoid the ill-conditioning of the covariance matrix R of the received signal, we are going to impose the regularization parameter into the energy functions under the following two constraints

$$\begin{cases} w^{H}s_{1} = 1\\ \sum_{k=2}^{K} w^{H}s_{k} = 0 \end{cases}$$
(7.15)

However, the proposed blind linear detector can be expressed as the following constrained optimization problem:

$$w_{opt} = \arg\min_{w} \{ E[\|b_1 - w^H r\|^2] + c\|w\|^2 \}$$
(7.16)

Subject to
$$\begin{cases} w^{H}s_{1} = 1\\ \sum_{k=2}^{K} w^{H}s_{k} = 0 \end{cases}$$

Therefore, one can solve the above constraint problem using the augmented Lagrangian method, so the cost (energy) function is given by

$$J = w^{H}Rw + c||w||^{2} + \gamma_{1}(w^{H}s_{1} - 1) + \gamma_{2}(\sum_{k=2}^{K} w^{H}s_{k})$$
(7.17)

where = m. $(tr(C) + \lambda_{max})$, it is regularization parameter with m is a positive constant. γ_1 and γ_2 are the Lagrangian multipliers. Therefore, the gradient of J expression is

$$g = 2Rw + 2cw - \gamma_1 s_1 + \gamma_2 \sum_{k=2}^{K} s_k$$
(7.18)

Let's define that

$$s_2 = \sum_{k=2}^{K} s_k$$

Therefore, the gradient g in (4.18) becomes

$$g = 2Rw + 2cw + \gamma_1 s_1 + \gamma_2 s_2$$
(7.19)

Then

$$w_{opt} = \frac{1}{2} [R + cI]^{-1} [\gamma_1 s_1 + \gamma_2 s_2]$$
(7.20)

Where I is an N x N identity matrix. Under the first constraint, we have

$$w^{H}s_{1} = 1$$
 (7.21)

However,

$$\left[\frac{\gamma_1}{2}[R+cI]^{-1}s_1 + \frac{\gamma_2}{2}[R+cI]^{-1}s_2\right]^H s_1 = 1$$
(7.21)

Therefore,

$$\gamma_1 = [s_1^H[R+cI]^{-1}s_1]^{-1} \left[2 - \gamma_2[s_2^H[R+cI]^{-1}s_1]\right]$$
(7.23)

Now, under second constraint, we have

$$\sum_{k=2}^{K} w^{H} s_{k} = 0$$
 (7.24)

Then

$$\sum_{k=2}^{K} \left[\frac{\gamma_1}{2} [R + cI]^{-1} s_1 + \frac{\gamma_2}{2} [R + cI]^{-1} s_2 \right]^{H} s_k = 0$$

However,

$$\frac{\gamma_1}{2} s_1^H [R + cI]^{-1} \sum_{k=2}^K s_k + \frac{\gamma_2}{2} \sum_{k=2}^K s_k^H \ [R + cI]^{-1} \sum_{k=2}^K s_k = 0$$
$$\frac{\gamma_1}{2} s_1^H [R + cI]^{-1} s_2 + \frac{\gamma_2}{2} \sum_{k=2}^K s_k^H \ [R + cI]^{-1} s_2 = 0$$

Let's define that

$$s_{11} = s_1^{H} [R + cI]^{-1} s_1$$

$$s_{21} = \sum_{k=2}^{K} s_k^{H} [R + cI]^{-1} s_1$$

$$s_{12} = s_1^{H} [R + cI]^{-1} s_2$$

$$s_{22} = \sum_{k=2}^{K} s_k^{H} [R + cI]^{-1} s_2$$

Therefore,

$$\gamma_1 = s_{11}^{-1} [2 - \gamma_2 s_{21}] \tag{7.25}$$

Then

$$s_{11}^{-1} [2 - \gamma_2 s_{21}] s_{12} + \frac{\gamma_2}{2} s_{22} = 0$$

$$[2s_{11}^{-1} s_{12} - \gamma_2 s_{11}^{-1} s_{21} s_{12}] + \frac{\gamma_2}{2} s_{22} = 0$$

$$\gamma_2 = 2 \left[s_{11}^{-1} s_{21} s_{12} - \frac{1}{2} s_{22} \right]^{-1} s_{11}^{-1} s_{12}$$
(7. 26)

7.5 Simulation Results

In this section, the simulated DS-CDMA downlink data in the presence of AWGN is carried out to verify the proposed method and compare it to the Subspace Blind Multiuser Detector (SBMUD). We used spreading codes with short gold for the length of chips to be C=31. Thus, the maximum number of users K =30 and assumed all signals for all users are sent at the same power. Monte Carlo Simulation was run to verify the validity of the algorithm simulations. Figure 7.1. shows the simulation results of SNR vs. BER for all detectors. The parameters were set as: Number of symbols L=500 and 1000 respectively, Number of users K=15, with various values of SNR 0 dB to 12dB. In Figure 7.1, we can see that the proposed algorithm improves the performance of the system, given the lowest BER regards to the SBMUD, and outperforms the performance of the some of the substance of the proposed algorithm slightly outperforms the performance of the regularized algorithm [76] at small sample sets, i.e. L=2N and 3 N, in Figure 7.2.



Figure 7.1: Average BER as a function of SNR for 15 users



Figure 7.2: Average BER as a function of SNR for 15 users with L=2N, L=3N.

Fig. 7.3 show the simulation results of BER vs. Regularization parameter with 15 users and 1000 sample sets for various SNR at $SIR_k = -20 \text{ dB}$ and $SIR_k = 0 \text{ dB}$. It is obvious that the BER performance at $SIR_k = -20 \text{ dB}$ is worse than that at $SIR_k = 0 \text{ dB}$. Furthermore, the BER performance performs well and outperforms the direct matrix inversion (DMI), which occurs at g = 0, under all SNR values. Moreover, the regularization parameter can be chosen as g = 0.025. In order to study the effect of the signature waveform mismatch on the BER performance as in [3].



Fig.3a: L = 1000 and SIR = 0 dB



Fig.3d: L = 500 and and SIR =-20 dB

Figure 7.3: Average BER as a function of SNR for 15 users For various L sample sets

Figure 7.4 represents the BER performance corresponding to channel coefficients $H_1 = [0.7, 0.2, 0.1]$ with $\langle s_1 | \tilde{s}_1 \rangle = 0.7032$, and $H_2 = [0.65, 0.15, 0.3]$ and $\langle s_1 | \tilde{s}_1 \rangle = 0.6597$, respectively; where \tilde{s}_1 represent the effective signature waveform vector. And for L=1000 and K= 15 users. Herein, it is obvious that the proposed detector has almost the same performance in the mismatch case and without mismatch. Furthermore, it is clear that the performance degrades as mismatch increases as shown in Figure 7.4. Despite the performance degrades with mismatch increases, the proposed algorithm still performs well and gives a reasonable performance close to the match one. Overall, the proposed detector performs better for solving symbol estimation problem in DS-CDMA system and avoids the ill-condition in inversion matrix.



Figure 7.4: Average BER as a function of SNR for 15 users with L=1000.

7.6 Conclusion

In this study, we have developed the constraint blind multiuser detection based on minimum of energy. Furthermore, we use the regularization method to avoid the singularity in covariance matrix and ill-posed problem. The results appear to show that the proposed method performs well in the symbol estimation problem in DS-CDMA systems and outperforms the other detectors. Our results also show that the Multiple Access Interference (MAI) can be mitigated by the proposed method, thus improving the performance of conventional Detection. Furthermore, the results show the proposed detector displays most robustness in the performance as mismatch between the original sequence and estimated one increases.

8 Chapter 8

Hardware Implementation

In this chapter, we investigate the ICA algorithms in terms of hardware implementation. Although software implementation is important to investigate the capabilities of ICA algorithms and to simulate significant aspects of applications, Hardware implementation provides real time solutions and an optimal parallelism method in terms of fast convergence. Furthermore, software implementation may suffer from insufficient memory problems because the large data sets of the ICAs' applications and its high dimensionality. Thus, hardware implementations are a promising approach to implement the ICA algorithms and they are executed by Integrated Circuit (ICs). Owning the high speed processing and the parallel architecture features make the hardware implementation outperforming the software implementation in terms of sufficient memory and fast convergence [105].

8.1 Introduction

During the last decade, several works have been presented to implement some of ICA algorithms on fully analog CMOS circuits, mixed (analog and digital) signal integrated circuit, application-specific integrated circuit (ASICs) and field programmable gate arrays (FPGAs).

In both analog and mixed CMOS integrated circuits, designers can integrate and create a fully customized design based on analog CMOS technologies or Mixed technologies. Although these aforementioned methods use the silicon in a more efficient way, the costs of these methods are significantly high, especially in terms of design expense and process. Therefore, the digital ASICs and FPGAs are considered to be an efficient ways to implement the ICA algorithms in general. Furthermore, one can consider the FPGAs based on the reconfigurable technology are the most promising technique to implement the ICA algorithm in terms of cost, since it allows the end user to modify and re-configure their designs multiple times.

One can refer to the survey in [99], where they studied and investigated the implementation of ICA algorithm based on very large scale integration (VLSI) approaches. Also, as we aforementioned, several designers implemented some of ICA algorithms based on the analog and mixed integrated technologies i.e. [100], [104] and [106].

8.2 Comparative Study of Existing Solutions to implement ICA Algorithms

In the last decade, VLSI technologies have been presented with several advantages which make it great choice to implement the ICA algorithms.

8.1.1 Analog CMOS Integration Circuit

Implementation of ICA algorithms based on an analog integrated circuit is usually the first choice in terms of low circuit delay and power consumption. Analog circuit design allows the end users to work at transistor level and necessary interconnections. Therefore, it emphasizes that the application based on analog integrated circuit has the minimum amount of transistors and the shortest interconnections to achieve the high circuit density and then low circuit delay and power consumption.

In implementation process, the implementation of ICA algorithms can be utilized by dividing the design process into several groups based on the functions. However, one can design a simple module structure such as $2 \ge 2$ input-output structures in order to extend it to any size readily. Also, this method has the ability to control the design area based on the application and it can be connected to external peripherals due to Analog-Digital Converter / Digital-Analog Converter on or off the chip.

In [101], Cohen and Andreou proposed two chips to implement the H-J ICA algorithm of speech signals. Fabrication was based on a 2-um, n-well CMOS process. Also, Gharbi and Salem in [136], proposed a chip design for the H-J algorithm using the 2 um CMOS technology. Lately, Cho and Lee in [100] presented an analog CMOS chip to implement the InfoMax ICA algorithm using a 0.6 um, p-well, AMS CMOS process. Their design consisted of a multiplier circuit, weight update circuits and quadrature function circuit.

The analog integrated circuit design has several drawbacks such as high expense of the workstation-based development system and slow turnaround time (approximately eight weeks). Thus, one can consider it an insufficient method to fast implementation of

222

most ICA designs [99]. Moreover, the analog CMOS circuits sometimes suffer from the transistor mismatch which affects the performance of the CMOS circuit. Transistor mismatch occurs due to edge effects, striation effects and random variations, for more details refer to [137].

In spite of all the aforementioned drawbacks, the analog integrated circuit is still very efficient in terms of design. Also, one can solve the transistor mismatch problem through increasing the current due to increasing the length or the width of the transistors and using the concentric structure to ensure the matched transistors by sharing the same surrounding structures.

8.1.2 Mixed Signal Techniques (Analog and Digital Circuit)

Mixed Signal Techniques provide an alternative approach to the analog integrated circuit. Actually, it combines the analog and digital circuit to take advantage of digital circuit in terms of fast switch and easy implement for certain application i.e.: Digital to Analog circuit. Although, mixed signal circuit outperforms some of ICA implantation based on analog integrated circuit, it still suffers from the expensive cost in terms of the workstation-based and the long time of the prototype turnaround period.

In [102], they simplified the H-J and infomax ICA algorithms and proposed a new chip design based on the mixed signal design. Also in [105], they used the parallel VLSI architecture to implement the feed-forward network. The chip was fabricated by 0.5 um two poly three metal CMOS technology.

According to aforementioned reasons, one can say that the analog CMOS and Mixed signal techniques provides the end user with an efficient full-custom solutions to

223

ICA algorithms. But it still requires having sufficient knowledge about the transistor level design and the physical problem. Also, it considers costing expensive in sense of time and cost of the work-station.

8.1.3 ASIC Solutions

Application Specific Integrated Circuit (ASIC) is considered to be one of the digital VLSI technologies which also include FPGAs. ASICs typically contain about ten million gates. They allow the end user to take advantage of the large number of libraries that are provided by IC vendors. Thus, one can call ASICs semi-custom solutions. Actually, although ASICs somehow increase the design risk and the cost in the sense of a nonprogrammable feature; it provides solutions for the very complex ICA algorithms with a good compact circuit design and low power consumption.

Table 8.1 compares the Implementation of ICA algorithms based on analog IC and mixed IC and ASIC solutions. Clearly, compact circuit design (ASIC) achieve the best performance in terms of low power circuit as a result of small chip design.

8.1.4 FPGA Solutions

Field programmable gate arrays (FPGAs) are products that are fabricated with specific standards and general-purposed by hardware companies. They allow the end user to implement a specific task or design on them. Also, the end users can modify their designs several times and program the interconnections in a few hours instead of waiting several weeks for the final fabrications. Therefore, FPGAs have outperformed the other VLSI technologies in terms of the turnaround period and the development expense. Typically, FPGAs contains more than 2000 gates and up to 2 million.

In the literature [103-115], many works have been proposed to implement the ICA algorithm based on the FPGA technologies [103], [104] and [106]. Also, according to [138], the current growth of the FPGA/ASIC technologies has reached far beyond the Moore's Law. Table 8.2 and Table 8.3 presents the recent FPGA solutions to implement the ICA algorithms. Although FPGAs outperforms the other VLSI technologies by having reconfigurable and reusable features, they usually suffer from the higher circuit delay which restricts their capacities. For more details refer to [99].

Table 8.1: Comparison of Analog, Mixed signal and ASIC Solutions						
Research	ICA algorithm	VLSI Category				
Cohen and	Herault -Jutten	Analog				
Gharbi and	Herault-jutten	Analog				
Cho and lee	Infomax ICA	Analog				
Celik et al.	H-J, infomax ICA	Mixed				
Du et al. [104]	Parallel ICA	ASIC				
Research Group	Fabrication Technology	Chip Size				
Cohen and Andreou [101]	2 um n-well 2M2P CMOS	N/A				
Gharbi and Salam [136]	2um CMOS	2.22 x 2.25				
Cho and lee	0.6 um p-well 3M2P CMOS	2.8 x 2.8				
Celik et al. [102]	0.5 um 3M2P CMOS	3 x 3				
Du et al. [104]	0.18 um 6 M1P CMOS	1.191 x 1.191				
Pagaarah	Input V Output	Voltage (V)				
Group	input X Output	voltage (v)				
Cohen and Andreou [101]	2 x 2	-2.5 - 2.5				
Gharbi and Salam [136]	2 x 2	N/A				
Cho and lee	4 x 4	-2.5 - 2.5				
Celik et al.	3 x 3	N/A				
Du et al. [104]	1 x 1 sequential	-1.8 - 1.8				

Table 8.2: Comparison of FPGA Solutions						
Research Group	ICA algorithm	FPGA				
Lim et al. [99]	MI and DO ICNNs	Xilinx Virtex XCV 812E				
Nordin et al. [109]	Pipelined InfoMax	N/A				
Satter and Charay. [113]	Infomax ICA	Xilinx Virtex E				
Wei and Charo. [111]	Infomax ICA	Xilinx Virtex E				
Kim et al. [103]	InfoMax ICA	Altera EP20K600E				
Du and Qi [104]	Parallel ICA	Xilinx Virtex V1000E				
Research Group	Frequency	Samples				
Lim et al. [99]	155K Hz MI, 3.62 KHz (DO)	1500				
Nordin et al. [109]	N/A	N/A				
Satter and Charay. [113]	60M Hz	2500				
Wei and Charo. [111]	12.288 M Hz	N/A				
Kim et al. [103]	N/A	N/A				
Du and Qi [104]	20.161 MHz	60000				
Research Group	Capacity (Million Gates)	Design Utilization				
Lim et al. [99]	25	0.7 % MI, 0.5% DO				
Nordin et al. [109]	N/A	N/A				
Satter and Charay. [113]	0.6	6%				
Wei and Charo. [111]	0.6	15%				
Kim et al. [103]	0.6	N/A				
Du and Qi [104]	1.0	92 %				

Table 8.3: Comparison Results Among Various ICA Implementations								
	[101]	[103]	[102]	[103]	[104]	[112]		
Application	Speech	Image	Image	Speech	EEG	EEG		
Algorithm	ICA	pICA	pICA	FastICA	InfoMax	FastICA		
Number of Channels / Weight Vectors	2	20 (WVs)	4 (WVs)	2	4	8		
Speed (M Hz)	12.288	35.92	20.161	50	68	100		
Power Dissipation (mW)	98.8	N/A	N/A	N/A	N/A	16.35		
Gates (Million)	0.0114	N/A	0.2295	N/A	0.315	0.272		
Computation Time (Sec)	60	1129.5	N/A	0.003	N/A	0.29		
Implementatio n Approach	ASIC	FPGA	ASIC	FPGA	FPGA	FPGA		

8.3 Multiplier Design

In this section, we design a down-conversion mixer (multiplier). So, the proposed mixer is shown in Figure 8.1. The structure is a modified variation of the Gilbert cell double balance mixer which has benefits of good port-to-port isolation and low even-order distortion.

The circuit consists of RF stage(M_{N1} , M_{N2}), LO Switching Circuit($M_{N3} - M_{N6}$), Current Injection Circuit($M_{P1} - M_{P3}$), Boosting Inductor(L_1 , L_2), a load resistance stage(R_1 , R_2), and an output driver stage(M_{N1} , M_{N2}). The trans-conductance stage "RF stage" amplifies the input differential RF signals. This stage is composed of the stacked NMOS-PMOS transistors. We mixed differential signals from RF and LO input ports through operating LO signals as an ideal switch function. We added Current Injection Circuit($M_{P1} - M_{P3}$) to improve the conversion gain and linearity. Output driver transistors(M_{N1} , M_{N2}) are common-source stages to match the output characteristic impedance of 50. The parasitic capacitance at the source node of LO Switching Circuit($M_{N3} - M_{N6}$) affects the mixer performance significantly.



Figure 8.1: The proposed Mixer (Multiplier) schematic.

So, we used boosting inductor(L_1 , L_2) to resonate with the parasitic capacitance in order to improve the performance and specifically to improve the linearity. Transistors $(M_{N3} - M_{N6})$ form bias circuit to provide bias current to other stages in the circuit.

We used the current injection technique to maintain the total current of the mixer of the mixer and to reduce the switching stage flicker noise since less noise is generated at the output with less current flowing through the switching stage. The drawback from current injection is the parasitic capacitance at the source of the switching stage where this capacitance becomes larger. So, the two inductors are placed between the RF input stage and the switching stage. Furthermore, the series resonated inductors provide high impedance so as to improve the conversion gain with good gain flatness and linearity. The NF is defined by amount of noise contributed by the circuit. The mixer carries out both the RF and the image signals to the same IF.

So for a noiseless mixer the output SNR is half the input SNR then NFSSB of a noiseless mixer is 3 dB. So, the NF is

$$NF = 10 \log \left[\frac{SNR_{RF}}{SNR_{IF}} \right] \quad [dB]$$

 $NF_{SSB} = 3 dB + NF_{DSB}$

To measure mixer's performance depends on power consumption, conversion gain, linearity, and noise figure. A Figure of Merit (FoM) is a quantity used to characterize the performance of a device that attempts to combine all the important parameter values that describe the performance of a circuit.

This value could be used to measure the performance of the mixer circuit so, FoM is:

$$FoM = \frac{Gain(abs). IIP3(mW)}{NF(abs). V_{dd}. Power(mw)}$$

Where IIP3 is input third-order intercept point, Vdd is the power supply; NF is Noise Figure of the circuit. And power is the power consumption.

8.4 Simulation Results

In this section, we present simulation results for the Mixer circuits. The presented mixer is designed by TSMC The circuit is $0.18 \ \mu m$ CMOS RF process and is simulated using the Cadence tool.

The proposed Mixer design described in previous section is operated around 1.9 GHz. biased at 1.2V supply voltage. As shown in Figure 8.2, with an RF power of - 30dBm and an LO power of 5dBm, IF frequency of 250MHz, the conversion gain

is 15.9 \pm 0.4dB. The results show good gain flatness within the IF band. Figure 8.3 shows the DSB-NF of our design as a function of the IF frequency. It is clear that the DSB noise figure is less than 7.2 dB. So, the SSB Noise figure would be less than 10 dB.

Figure 8.4 presents the conversion gain versus the power supply, it is clear that the mixer can work very well with low voltage supply. Furthermore, Figure 8.4 shows the gain conversion versus the power of LO, it is so obvious the stability of our design and the higher gains over the wide band LO frequency. For the IIP3 point, we use the two tones test to measure it. As shown in Figure 8.5, we get a suitable IIP3 of 10.25 dBm. And the 1-dB compression point of -0.8 dBm. Then FoM is 0.194, which outperforms results reported in the literature.



Figure 8.2: Voltage Conversion Gain versus IF



Figure 8.3: Voltage Conversion Gain versus IF

8.5 Conclusion

In this chapter, we used the TSMC 0.18um CMOS to simulate the mixer design. The mixer demonstrates high linearity and high gain performance. Furthermore, we achieved a good noise figure. On the other side, we achieved a mixer that has a 15.9 dB conversion gain and 10 dBm IIP3 linearity with 7.2 dB DSB-noise figure. In this chapter, the injection current with boosting inductor achieve a high performance for the mixer (multiplier) at low voltage supply.



Figure 8.4: Voltage Conversion Gain versus IF



Figure 8.5: Voltage Conversion Gain versus IF


Figure 8.6: The proposed Mixer (Multiplier) Layout.

9 Chapter 9

Conclusion and Future Work

In this chapter we conclude the dissertation and highlight directions for future work.

9.1 Conclusion

In this dissertation, *Chapter 1* provided the background needed for the discussion of blind source separation problem. The benefits of blind techniques were discussed along with its applications in wireless communications, and speech enhancement.

Chapter 2 performs a thorough review of the BSS\ICA algorithms, and then it gives an overview of the ICA algorithms and emphasizes the approaches that influenced our work. It also studies some of the methods that have been developed to solve the ICA problems in the case of instantaneous and convolutive mixtures.

In *Chapter 3*, a novel divergence measure class is presented based on integrating convex functions into the Cauchy–Schwarz inequality. This divergence measure is used as a contrast function to develop new ICA algorithms to solve the Blind Source Separation (BSS) problem. The CCS-DIV derived algorithms can be controlled to attain the steepest descent towards the minimum value. Also, a pairwise iterative scheme is employed to address the high dimensional problem in BSS. Two schemes of pairwise non-parametric ICA algorithms are developed based on the proposed divergence. Several

examples and experiments are carried out to show the improved performance of the proposed divergence. Furthermore, this chapter compares the metric performance with a host of leading ICA algorithms. We have developed also nonparametric CCS–ICA approaches to demixing where the source signals are estimated by the Parzen Window density. The convergence speed of the parameterized CCS–ICA procedure is evaluated and compared to other algorithms. The proposed CCS–ICA algorithms attained the highest SIR in separation of speech and music signals relative to other leading ICA-based algorithms.

In chapter 4, we presented the RobustICA-based algorithm to solve the frequency-domain BSS problem for convolutive acoustic mixtures in several adverse conditions. Through the real-world experiments, we show the superiority of the presented algorithm among other popular algorithms in the literature in terms of the performance and complexity computation. Moreover, we compared several permutation solvers in terms of computation complexity and performance to provide the RobustICA-based algorithm with an efficient frequency-dependent permutation scheme. Finally, we studied the effect of several parameters on the separation performance of the presented algorithm. We also presented the effect of the type of the window on the separation performance and we also showed that the performance improves at a certain range of overlapping between the signals. Lastly, in this chapter, we showed the performance of a system that can work efficiently with around 0.5-10 seconds of input data, which is close to the real-time implementation. Accordingly, our proposed algorithm is optimized to be suitable for the real-time operation. As a result, it is suitable for a large number of applications to ensure the real-time implementation.

Chapter 5 has investigated three adaptive algorithms for user-detection in CDMA systems, the proposed one based on fourth order cumulant matrices, the Fast ICA and the Robust ICA algorithms. The results show that the proposed algorithm exhibits better performance relative to the other two user detectors. The results also show that the proposed algorithm can mitigate Multiple Access Interference (MAI), thus improving the performance of conventional detection. Furthermore, the performance of the proposed detector displays the most consistent improvement as M (the number of symbols) increases. Also, we assess the performance of computational complexity of the three user detection algorithms employing the average signal of mean square error SMSE, as a contrast function of independent criteria. The results show that the proposed detector provides a faster and more robust performance.

Chapter 6 carried out both simulation and theoretical demonstrations of the blind multiuser detector based on the space state structures in the CDMA system. Also, we develop the three blind multiuser detectors based on the three algorithms ICA, RICA and PCA. The results appear to show that the proposed algorithms perform well in the symbol estimation problem in DS/CDMA systems and outperform the other conventional detectors and the Adaptive MMSE. Our results also show that Multiple Access Interference (MAI) can be mitigated by the proposed algorithms, thus improving the performance of blind multiuser detection. Although the proposed method improves as the size of the sample set increases, the results show the proposed detector performs well even though the sample sets are small, unlike the LMMSE detector. Moreover, unlike the complexity of the LMMSE detector, the complexity of the proposed methods, being constant, didn't increase exponentially. Finally, the proposed algorithms, unlike the

adaptive LMMSE detector, have no restriction regarding the spreading codes since they do not require the spreading codes of the interfering users. Therefore, it is a more suitable choice in the downlink case and it does work in the uplink case as well. Moreover, In *Chapter 7*, we introduce a constrained blind multiuser detection in order to improve its performance with imposing the regularization parameter to cope the ill-conditioning problem of the covariance matrix and to mitigate the performance degradation.

In *Chapter 8*, we investigate the ICA algorithms in terms of hardware implementation. Although software implementation is important to investigate the capabilities of ICA algorithms and to simulate significant aspects of applications, Hardware implementation provides real time solutions and an optimal parallelism method in terms of fast convergence. Furthermore, software implementation may suffer from insufficient memory problems because the large data sets of the ICAs' applications and its high dimensionality. Thus, hardware implementations are a promising approach to implement the ICA algorithms and they are executed by Integrated Circuit (ICs). Owning the high speed processing and the parallel architecture features make the hardware implementation outperforming the software implementation in terms of sufficient memory and fast convergence.

9.2 Future Work

This section provides directions of future work for the area of Blind Source Separation and its implementation. Specifically, we itemize the research activities in the following:

Optimization: We presented a new Convex Cauchy–Schwarz Divergence (CCS-DIV) measure for Blind Source Separation (BSS) and unsupervised learning of acoustic and

239

speech signals. The CCS-DIV measure is developed by integrating convex functions into the Cauchy–Schwarz inequality. By including a convexity quality parameter, the measure has a broad control range of its convexity. With this new measurement technique, a new CCS-ICA algorithm is structured and a non-parametric form is developed incorporating the Parzen window-based distribution. Moreover, the CCS-ICA algorithm has a controlled speed towards timed convergence. Several case-study scenarios were carried out on instantaneous and noisy mixtures of speech signals. Finally, the superiority of the proposed CCS-ICA algorithm is demonstrated in metric-comparison performance with FAST ICA, Robust ICA, convex ICA (C-ICA), and other leading existing algorithms. The gradient-type algorithms can be considered to be robust optimization techniques; but they usually suffer from several drawbacks in terms of convergence and stability. Also, the convergence of the gradient-type algorithms is relatively slow and their stability relies on the choice of the learning rate. Therefore, one can upgrade the optimization method that is faster and more robust algorithms such as decoupled and fast relative newton optimization as in [8] and [9] respectively.

Online implementation: one of the most challenging questions about any proposed algorithm is that if it can work on-line or not. Real-time implementation is very important to measure the efficiency of the proposed algorithm. Therefore, our new algorithms will be extended to online implementation. An interesting approach to implement the algorithm is to work in a mixed block-based and real time methods such as a block LMS-type structure [2], [42], [48], [71]. In this approach, some data is stored in a series of buffers in order to be processed sequentially and the results are the sequential blocks. In this case, the challenge is to find the optimal length of this local buffer in order to

perform the separation process with acceptable performance. Moreover, the real-time DSP processor can handle the computational cost without interruptions or distortions. The challenge here is to determine the length of this interval which needs to be selected based on two parameters:

- It should be short enough for the mixing environment to be considered non-stationary
- It should be long enough to perform a separation processes successfully with an excellent.

This is the same idea as non-stationary mixing case in [1], [98], [111], [114], and [120]. Therefore, finding an optimal length of the block of data "interval" might solve both problems at the same time.

Underdetermined mixtures

If the number of observations (sensors) m is less than the number of sources n, the mixing process is referred to be underdetermined (not invertible) [1], [52], [53]. The separation processes can be attained successfully in the frequency-domain up to scaling and permutation ambiguities under the assumption that the mixing matrix H(f) is full column rank at each frequency bin. However, when the number of source signals is more than the number of sensors, the assumption on the mixing matrix H(f) becomes not valid. So, in this case the problem is more difficult since the mixing matrix H(f) becomes ill-conditioned matrix which means the mixing matrix H(f) is not left pseudo-invertible. However, a lot of work has been done in order to perform a good separation process in the case of the instantaouse mixture [1]. However, there are not so many works that has been done on the underdetermined case in the convolutive mixture [1], [38]. In the literature, the well-known algorithm of such method is the DUET algorithm which is proposed by Rickard et al [2], [3], [38] and [117]. The DUET algorithm assumes a

specific delayed model that only works for audio signals with small delay, e.g. hearing aid etc. The DUET algorithm performs the separation processes using the two sensors in order to compute two parameters amplitude differences and phase differences between the source signals. Several papers were published to develop and enhance the performance of the DUET algorithm in [3], but their performance in real reverberant environment is still limited. One of the promising approaches in this field is to convert some of the underdetermine cases of the instantounous mixture into the frequency domain in order to tackle the underdetermined problem in the convoluvtive mixture as presented in the literature [1], [3], [38].

APPENDIX

Convex Cauchy–Schwarz Divergence and its Derivative

Assume the demixed signals $Y_t = WX_t$ where the mth component $y_{mt} = w_mX_t$. Now express the CCS-DIV as a contrast function with a convexity parameter α as follows:

$$D_{CCS}(Y_t, y_{mt}, \alpha) = \log \frac{\iint f^2(p(Y_t)) dy_1 dy_2 \cdot \iint f^2(\prod_1^N p(y_{mt})) dy_1 dy_2}{[\iint f(p(Y_t)) \cdot f(\prod_1^N p(y_{mt})) dy_1 dy_2]^2}$$

By using the Lebsegue measure [5] to approximate the integral with respect to the joint distribution of $Y_t = \{y_1, y_2, ..., y_N\}$, the contrast function becomes

$$D_{CCS}(Y_t, y_{mt}, \alpha) = \log \frac{\sum_{1}^{T} f^2(p(WX_t)) \cdot \sum_{1}^{T} f^2(\prod_{1}^{N} (p(w_{mt}X_t)))}{[\sum_{1}^{T} f(p(WX_t)) \cdot f(\prod_{1}^{N} (p(w_{mt}X_t)))]^2}$$

For simplicity, let us assume

$$V_{1} = \sum_{t=1}^{T} f^{2}(Y_{t}), \quad V_{1}' = \sum_{t=1}^{T} 2f(Y_{t})f'(Y_{t})Y_{t}'$$

$$V_{2} = \sum_{t=1}^{T} f^{2}(y_{mt}), \quad V_{2}' = \sum_{t=1}^{T} 2f(y_{mt})f'(y_{mt})y_{mt}'$$

$$V_{3} = \sum_{t=1}^{T} f(Y_{t})f(y_{mt}),$$

$$V_{3}' = \sum_{t=1}^{T} f'(Y_{t})f(y_{mt})Y_{t}' + \sum_{t=1}^{T} f(Y_{t})f'(y_{mt})y_{mt}'$$

and the convex function is

$$f(t) = \frac{4}{1 - \alpha^2} \left[\frac{1 - \alpha}{2} + \frac{1 + \alpha}{2} t - t^{\frac{1 + \alpha}{2}} \right]$$
$$f'(t) = \frac{2}{1 - \alpha} \left[1 - t^{\alpha - 1/2} \right]$$

then,

$$\begin{split} Y_t &= p(WX_t) \text{and } y_{mt} = \prod_{m=1}^M p(w_m X_t) \\ Y'_t &= \frac{\partial Y_t}{\partial w_{ml}} = -\frac{p(X_t)}{|\det(W)|^2} \cdot \frac{\partial \det(W)}{\partial w_{ml}} \cdot \text{sign}(\det(W), \end{split}$$

where
$$\frac{\partial \det(W)}{\partial w_{ml}} = W_{ml}$$
;
 $y'_{mt} = \frac{\partial y_{mt}}{\partial w_{ml}} = \left[\prod_{j=m}^{M} p(w_j X_t)\right] \frac{\partial p(w_n X_t)}{\partial (w_n X_t)} \cdot x_l$.

where x_l denotes the l^{th} entry of X_t .

Thus, we re-write the CCS-DIV as

$$D_{CCS}(Y_t, y_{mt}, \alpha) = \log \frac{V_1 \cdot V_2}{[V_3]^2}$$

and its derivative becomes

$$\frac{\partial D_{CCS}(Y_t, y_{mt}, \alpha)}{\partial w_{ml}} = \frac{V_3^2}{V_1 \cdot V_2} \cdot \frac{V_1' V_2 V_3 + V_1 V_2' V_3 - 2V_1 V_2 V_3 V_3'}{V_3^4}$$

$$\frac{\partial D_{CCS}(Y_t, y_{mt}, \alpha)}{\partial w_{ml}} = \frac{V_1' V_2 + V_1 V_2' - 2V_1 V_2 V_3'}{V_1 V_2 V_3}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] P. Comon, C. Jutten (eds.), "Handbook of Blind Source Separation Independent Component Analysis and Applications." (Academic Press, Oxford, 2010).

[2] A. Cichocki, S.-I. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, John Wiley & Sons, Inc., 2002.

[3] M. S. Pedersen, J. Larsen, U. Kjems, and L. C. Parra, "A survey of convolutive blind source separation methods," in Springer Handbook of Speech Processing. New York: Springer, 2007.

[4] A. Cichocki, R. Zdunek, S.-I. Amari, Nonnegative matrix and tensor factorizations: applications to exploratory multi-way analysis and Blind Source Separation, John Wiley & Sons, Inc., 2009.

[5] S. Boyd and L. Vandenberghe Convex Optimization, 2004: Cambridge Univ. Press.

[6] C. E. Shannon "A mathematical theory of communication," Bell Syst. Tech. J., vol. 27, pp. 379–423, 1948.

[7] P. COMON, ``Independent Component Analysis, a new concept," Signal Processing, Elsevier, 36(3):287--314, April 1994PDF Special issue on Higher-Order Statistics.

[8] M. Zibulevsky, "Blind source separation with relative Newton method," in Proc. ICA 2003, 2003, pp. 897–902.

[9] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation performance analysis," IEEE Trans. Signal Process. vol. 60, no. 4, pp. 1672–1683, Apr. 2012.

[10] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," IEEE Trans. Signal Process., vol. 55, no. 51, pp. 1803–1814, 2007.

[11] V. Zarzoso and P. Comon, "Robust Independent Component Analysis by Iterative Maximization of the Kurtosis Contrast with Algebraic Optimal Step Size," IEEE Transactions on Neural Networks, vol. 21, no. 2, pp. 248–261, 2010.

[12] E. Oja and Z. Yuan "The fastICA algorithm revisited: Convergence-analysis", IEEE Trans. Neural Netw., vol. 17, no. 6, pp.1370 -1381 2006

[13] A.-J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," IEEE Trans. Signal Process., vol. 44, pp. 1136–1155, 1996.

[14] Hyvarinen. A, "Fast and robust fixed-point algorithm for independent component analysis". IEEE Transactions on Neural Network, vol. 10, no. 3, pp. 626–634, May 1999.

[15] Hyvarinen. A. E.Oja, "A fast fixed-point algorithm for independent component analysis," Neural Computation, vol. 9, no. 7, pp. 1483–1492, 1997.

[16] F. Cardoso. On the performance of orthogonal source separation algorithms. In Proc. EUSIPCO, pages 776–779, 1994a.

[17] Jean-François Cardoso, "High-order contrasts for independent component analysis," Neural Computation, vol. 11, no 1, pp. 157–192, Jan. 1999.

[18] A. Bell, T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution", Neural Computation, 7:1129-1159, 1995.

[19] L, Xianhua; Cardoso, J-Francois; Randall, Robert B. " Very fast blind source separation by signal to noise ratio based stopping threshold for the SHIBBS/SJAD algorithm", Mechanical Systems and Signal Processing, Volume 24, Issue 7, p. 2096-2103. 2010.

[20] Jen-Tzung Chien, Hsin-Lung Hsieh, "Convex Divergence ICA for Blind Source Separation," Audio, Speech, and Language Processing, IEEE Transactions on, On page(s): 302–313 Volume: 20, Issue: 1, Jan. 2012

[21] D. Xu, J. C. Principe, J. Fisher III and H.-C. Wu "A novel measure for independent component analysis (ICA)," Proc. Int. Conf. Acoust., Speech, Signal Process., vol. 2, pp. 1161–1164, 1998

[22] R. Boscolo, H. Pan and V. P. Roychowdhury "Independent component analysis based on nonparametric density estimation," IEEE Trans. Neural Netw., vol. 15, no. 1, pp. 55–65, 2004.

[23] Y. Chen "Blind separation using convex function," IEEE Trans. Signal Process., vol. 53, no. 6, pp. 2027–2035, 2005.

[24] J.-T. Chien and B.-C. Chen "A new independent component analysis for speech recognition and separation," IEEE Trans. Audio, Speech Lang. Process., vol. 14, no. 4, pp. 1245–1254, 2006

[25] Y. Matsuyama, N. Katsumata, Y. Suzuki and S. Imahara "The \$\alpha\$-ICA algorithm," Proc. Int. Workshop Ind. Compon. Anal. Blind Signal Separat., pp. 297–302, 2000

[26] A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's Divergences for Non-Negative Matrix Factorization: Family of New Algorithms," 6th International Conference on Independent Component Analysis and Blind Signal Separation, Charleston SC, USA, March 5–8, 2006 Springer LNCS 3889, pp. 32–39.

[27] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals uses mixture models," in Proc. Int. Conf. Acoust. Speech Signal Process. Apr. 1997, vol. 5, pp. 3617–3620.

[28] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," IEEE Trans. Signal Process, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.

[29] B. A. Pearlmutter and L. C. Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," Adv. Neural Inf. Process. Syst., pp. 613–619, Dec. 1996.

[30] Fujisawa, H. and Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. J. Multivariate Anal. 99 2053–2081.

[31] J. Zhang "Divergence function, duality, and convex analysis," Neural Comput., vol. 16, pp. 159–195, 2004.

[32] J. Lin "Divergence measures based on the Shannon entropy," IEEE Trans. Inf. Theory, vol. 37, no. 1, pp. 145–151, 1991.

[33] S.C.Douglas, X.Sun, "Convolutive blind separation of speech mixtures using the natural gradient," Speech commun, vol. 39. pp. 65–78, (2002).

[34] Yoshioka, Takuya Nakatani, Tomohiro Miyoshi, Masato Okuno, Hiroshi G. "Blind Separation and Dereverberation of Speech Mixtures by Joint Optimization," IEEE Transactions on Audio Speech and Language Processing, Volume. 19, Issue. 1, pp. 69, 2011.

[35] A. Cichocki, R. Zdunek, S. Amari, G. Hori and K. Umeno, "Blind Signal Separation Method and System Using Modular and Hierarchical-Multilayer Processing for Blind Multidimensional Decomposition, Identification, Separation or Extraction," Patent pending, No. 2006-124167, RIKEN, Japan, March 2006.

[36] Intae Lee, Taesu Kim, and Te-Won Lee. Independent vector analysis for convolutive blind speech separation. In Blind Speech Separation. Springer, September 2007.

[37] Solvang, Hiroko Kato Nagahara, Yuichi Araki, Shoko Sawada, Hiroshi Makino, Shoji "Frequency-Domain Pearson Distribution Approach for Independent Component Analysis (FD-Pearson-ICA) in Blind Source Separation," IEEE Transactions on Audio Speech and Language Processing, Volume 17, Issue. 4, pp. 639, 2009.

[38] H. Sawada, S. Araki, and S. Makino. Frequency-domain blind source separation. In Blind Speech Separation. Springer, September 2007. [39] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in Proc. Int. Workshop Ind. Compon. Anal. Blind Signal Separation 1999, pp. 261–266.

[40] Saruwatari, H. Kawamura, T. Nishikawa, T. Lee, A. Shikano, K. "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," IEEE Transactions on Audio Speech and Language Processing, Volume 14, Issue 2, pp. 666, 2006.

[41] Low, S.Y. Nordholm, S. Togneri, R. "Convolutive Blind Signal Separation With Post-Processing," IEEE Transactions on Speech and Audio Processing, Volume 12, Issue 5, pp. 539, 2004.

[42] Takahashi, Yu. Saruwatari, H., Shikano, K., "Real-time implementation of blind spatial subtraction array for hands-free robot spoken dialogue system," Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, On page(s): 1687–1692.

[43] Yueyue Na, Jian Yu, "Kernel and spectral methods for solving the permutation problem in frequency domain BSS," Neural Networks (IJCNN), The 2012 International Joint Conference on, On page(s): 1 - 8

[44] A. Chen, "Fast kernel density independent component analysis," in Proc. 6th Int. Conf. ICA BSS, 2006, vol. 3889, pp. 24–31.

[45] F. Nest, P. Svaizer and M. Omologo "Convolutive BSS of short mixturesby ICA recursively regularized across frequencies", IEEETrans. Audio, Speech, Lang. Process., vol. 19, no. 3, pp.624 -639 2011

[46] M. Triki and D. T. M. Slock "Iterated delay and predictequalization for blind speech dereverberation", Proc. Int. Workshop Acoust. Echo, Noise Contr., 2006

[47] S. C. Douglas and M. Gupta "Scaled natural gradient algorithms for instantaneous and convolutive blind source separation", Proc. ICASSP, vol. II, pp.637 - 640 2007

[48] Y. S. Choi , H. C. Shin and W. J. Song "Robust regularization for normalized LMS algorithms", IEEE Trans. Circuits Syst., vol. 53, no. 8, pp.627-631 2006

[49] T. S. Wada and B.-H. Juang "Acoustic echo cancellation based on independent component analysis and integrated residual echo enhancement", Proc. WASPAA, pp.205 -208 2009

[50] H. Sawada , S. Araki , R. Mukai and S. Makino "Blind extraction of dominant target sources using ICA and time–frequency masking", IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 6, pp.2165 -2173 2006

[51] P. Bofill and M. Zibulevsky "Underdetermined blind source separation using sparse representations", Signal Process., vol. 81, no. 11, pp.2353 -2362 2001

[52] P. Georgiev, F. Theis and A. Cichocki "Sparse component analysis and blind source separation of underdetermined mixtures", IEEE Trans. Neural Netw., vol. 16, no. 4, pp.992 -996 2005

[53] S. Araki , S. Makino , T. Nishikawa and H. Saruwatari "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech", Proc. ICASSP, pp.2737 -2740 2001

[54] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa and H. Saruwatari "Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures", EURASIP J. Appl. Signal Process., vol. 2003, pp.1157 -1166 2003

[55] E. Robledo-Arnuncio , H. Sawada and S. Makino "Frequency domain blind source separation of a reduced amount of data using frequency normalization", Proc. ICASSP, vol. 5, pp.837-840 2006

[56] J. Reiss, N. Mitianoudis, and M. Sandler. Computation of generalized mutual information from multichannel audio data. In 110th audio engineering society intern. Conf. Amsterdam, Netherlands, 2001.

[57] N. Murata, S. Ikeda and A. Ziehe "An approach to blind source separation based on temporal structure of speech signals", Neurocomputing, vol. 41, no. 1–4, pp.1 -24 2001

[58] L. De Lathauwer and A. de Baynast "Blind deconvolution of DS-CDMA signals by means of decomposition in rank-(1,L,L) terms", IEEE Trans. Signal Process., vol. 56, no. 4, pp.1562 -1571 2008

[59] D Nion and L De Lathauwer "A block component model-based blind DS-CDMA receiver" IEEE Trans. Signal Process. 56 5567-5579, 2008

[60] D. Nion, K. N. Mokios, N. D. Sidiropoulos and A. Potamianos "Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures", IEEE Trans. Audio, Speech Lang. Process., vol. 18, no. 6, pp.1193 -1207 2010

[61] K. Rahbar and J.-P. Reilly "A frequency domain method for blind source separation of convolutive audio mixtures", IEEE Trans. Speech Audio Process., vol. 13, no. 5, pp.832 -844 2005 [online] Available: http://www.ece.mcmaster.ca/~reilly/kamran/id18.htm

[62] L. Parra and C. Spence "Convolutive blind separation of non-stationary sources", IEEE Trans. Speech Audio Process., vol. 8, no. 3, pp.320 -327 2000 [online] Available: http://ida.first.fhg.de/~harmeli/download/download_convbss.html

[63] Dinh-Tuan Pham, Zaher El-Chami, Alexandre Guerin, and Christine Serviere. Modelling the short time fourier transform ratio and application to underdetermined audio source separation. In Tulay Adali, Christian Jutten, Jo ao Marcos Travassos Romano, and Allan Kardec Barros, editors, Independent Component Analysis and Signal Separation, pages 98–105. Springer, 2009.

[64] C. Serviegrave; re and D.-T. Pham "Permutation correction in the frequency domain in blind separation of speech mixtures", EURASIP J. Appl. Signal Process., no. 1, pp.1-16 2006

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

[66] A. Westner and J. V. M. Bove "Blind separation of real world audio signals using overdetermined mixtures", Proc. ICA\'99, 1999 [online] Available: http://sound.media.mit.edu/ica-bench

[67] D.-T. Pham, C. Servière and H. Boumaraf "Blind separation of convolutive audio mixtures using nonstationarity", Proc. Int. Workshop Indep. Compon. Anal. Blind Signal Separation (ICA\'03), pp.981 -986 2003

[68] L. Xianhua, "Blind Source Separation methods and their mechanical applications", PhD thesis, University of New South Wales, august, 2006.

[69] H. Sawada , R. Mukai , S. Araki and S. Makino "A robust and precise method for solving the permutation problem of frequency-domain blind source separation", IEEE Trans. Speech Audio Process., vol. 12, no. 5, pp.530-538 2004

[70] H. Sawada , S. Araki and S. Makino "MLSP 2007 data analysis competition: Frequency-domain blind source separation for convolutive mixtures of speech/audio signals", Proc. MLSP\'07, pp.45 -50 2007

[71] N. Mitianoudis, "Audio Source Separation using Independent Component Analysis", PhD thesis, University of London, April, 2004.

[72] F. Nesta, "Techniques for robust source separation and localization in adverse environment", PhD thesis, University of Trento, April, 2010.

[73] Lai-Wan Chan, Siu-Ming Cha, "Selection of Independent Factor Model In Finance", In Pros. Int. Workshop on Independent Component Analysis ICA and Blind Source Separation, San Diageo USA, 2001.

[74] X. D. Wang and H. V. Poor, Wireless Communication Systems: Advanced Techniques for Signal Reception. Prentice Hall, 2004.

[75] K. Waheed and F. Salem, "Blind information-theoretic multiuser detection algorithms for DS-CDMA and WCDMA downlink systems," IEEE Trans. Neural Netw., vol. 16, no. 4, pp. 937–948, Jul. 2005.

[76] L. Hu, X. Zhou, L. Zhang, "Blind Multiuser Detection Based on Tikhonov Regularization", IEEE Communications Letters, vol. 15, no. 5, May 2011.

[77] G. T. Raja and O. Reddy, "Improved ICA based multi-user detection of DS-CDMA," ICETET, 2008, pp.238-241

[78] X. G. Doukopoulos and G. V. Moustakides "Adaptive power techniques for blind channel estimation in CDMA systems", IEEE Trans. Signal Process., vol. 53, no. 3, pp.1110 -1120 2005

[79] S. Cui , M. Kisialiou , Z.-Q. Luo and Z. Ding "Robust blind multiuser detection against signature waveform mismatch based on second-order cone programming", IEEE Trans. Commun., vol. 4, no. 4, pp.1285 -1291 2005

[80] Tikhonov AN, Arsenin VI. Solutions of ill-posed problems. Washington/New York; Winston, distributed by Halsted Press. 1977.

[81] G. B. G. Zhang and L. Zhang "Group-blind intersymbol multiuser detection for downlink cdma with multipath", IEEE transaction on Wireless Communications, vol. 4, no. 2, pp.434 -443 2005

[82] R. C. de Lamare, M. Haardt and R. Sampaio-Neto "Blind adaptive constrained reduced-rank parameter estimation based on constant modulus design for CDMA interference suppression", IEEE Trans. Signal Process., vol. 56, no. 6, pp.2470 -2482 2008

[83] Z. Albataineh, F. Salem, "robust blind multiuser detection DS-CDMA algorithm using simplified fourth order cumulant matrices" IEEE International Symposium on Circuit and System (ISCAS), pp. 1946-1949, 2013.

[84] Z. Albataineh, F. Salem, "New blind multiuser detection DS-CDMA algorithm based on Extension of Efficient FAST Independent Component Analysis (EF-ICA)." 4th International Conference on Intelligent Systems Modelling & Simulation (ISMS), pp. 543-548, 2013.

[85] W. N. Yuan , Y. F. Tu and P. Z. Fan "Optimal training sequences for cyclicprefix-based single-carrier multi-antenna systems with space-time block-coding", IEEE Trans. Wireless Commun., vol. 7, no. 11, pp.4047 -4050 2008

[86] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA systems," IEEE Trans. Signal Process. vol. 48, pp. 810–823, 2000.

[87] X. Wand and H. V. Poor, "Blind multiuser detection: A subspace approach", IEEE Trans. Inform. Theory, vol. 44, pp.677-690 1998

[88] T. Huovinen, "Independent Component Analysis in DS-CDMA Multiuser Detection and Interference Cancellation", PhD thesis, Tampere University of Technology, 2008.

[89] X. Wang and A. Host-Madsen, "Group-blind multiuser detection for uplink CDMA", IEEE J. Sel. Areas Commun., vol. 17, no. 11, pp.1971 -1984 1999

[90] P-Y. Qiu, Z.-T. Huang, W.-L. Jiang, C.Zhang. Improved Blind Spreading Sequence Estimation Algorithm for the Direct Sequence Spread Spectrum Signals [J]. IET Signal Processing, 2008, 2(2):139-146.

[91] X. X. Zhang and T. S. Qiu, "Blind multiuser detection based on improved Infomax and Fast ICA," 2010, 2nd International Conference on Advanced Computer Control, Shenyang, 2010, pp. 476-479.

[92] Y. Washizawa , Y. Yamashita , T. Tanaka and A. Cichocki "Blind extraction of global signal from multi-channel noisy observations", IEEE Trans. Neural Netw., vol. 21, no. 9, pp.1472 -1481 2010

[93] M. K. Tsatsanis and Z. Xu "Performance analysis of minimum variance CDMA receivers", IEEE Trans. Signal Process., vol. 46, no. 11, pp.3014 -3022 1998

[94] A. Bayati, S. Prakriya and S. Prasad, "Semi-blind space-time receiver for multiuser detection of DS/CDMA signals in multipath channels", The Institution of Engineering and Technology, vol. 153, no. 3, 2006.

[95] F. M. Salam and G. Erten, "The state space framework for blind dynamic signal extraction and recovery", Proc. IEEE Int. Symp. Circuits and Systems (ISCAS', 99), vol. 5, pp.66 -69 1999.

[96] Koivisto T, Koivunen V. Blind Despreading of Short-Code DS-CDMA Signals in Asynchronous Multi-user Systems [J]. Signal Processing, 2007, 11(87): 2560-2568.

[97] E. Ollila "The deflation-based fastICA estimator: Statistical analysis revisited", IEEE Trans. Signal Process., vol. 58, no. 3, pp.1527 -1541 2010

[98] G. Zhou, Z. Yang, S. Xie and J. M. Yang "Online blind source separation using incremental nonnegative matrix factorization with volume constraint", IEEE Trans. Neural Netw., vol. 22, no. 4, pp.550 -560 2011

[99] H. Du , H. Qi and X. Wang "Comparative study of VLSI solutions to independent component analysis", IEEE Trans. Ind. Electron., vol. 54, no. 1, pp.548 - 558 2007

[100] K. S. Cho and S. Y. Lee "Implementation of InfoMax ICA algorithm with analog CMOS circuits", Proc. Int. Workshop Independent Compon. Anal. Blind Signal Separat., pp.70-73 2001

[101] M. H. Cohen and A. G. Andreou "Analog CMOS integration and experimentation with an autoadaptive independent component analyzer", IEEE Trans. Circuits Syst. II-Anal. Digital Signal Process., vol. 42, no. 2, pp.65 -77 1995

[102] A. Celik, M. Stanacevic and G. Cauwenberghs "Mixed-signal real-time adaptive blind source separation", Proc. IEEE Int. Symp. Circuits Syst., pp.760-763 2004

[103] C. M. Kim, H. M. Park, T. Kim, Y. K. Choi and S. Y. Lee "FPGA implementation of ICA algorithm for blind signal separation and adaptive noise canceling", IEEE Trans. Neural Netw., vol. 14, no. 5, pp.1038-1046 2003

[104] H. Du and H. Qi "A reconfigurable FPGA system for parallel independent component analysis", EURASIP J. Embedded Syst., vol. 2006, no. 23025, pp.1 -12 2006

[105] H. Du, H. Qi and G. D. Peterson "Parallel ICA and its hardware implementation in hyperspectral image analysis", Proc. SPIE, vol. 5439, pp.74 -83 2004

[106] C. Charoensak and F. Sattar"A single-chip FPGA design for real-time ICA-based blind source separation algorithm", Proc. IEEE Int. Symp. Circuits Syst., vol. 6, pp.5822 -5825 2005

[107] K. K. Shyu, M. H. Lee, Y. T. Wu and P. L. Lee "Implementation of pipelined fastICA on FPGA for real-time blind source separation", IEEE Trans. Neural Netw., vol. 19, no. 6, pp.958 -970 2008

[108] M. Kim , K. Ichige and H. Arai "Design of Jacobi EVD processor based on CORDIC for DOA estimation with MUSIC algorithm", Proc. PIMRC Conf., vol. 1, pp.120-124 2002

[109] A. Nordin, C. Hsu and H. Szu "Design of FPGA ICA for hyperspectral imaging processing", Proc. SPIE, Wavelet Appl. VIII, vol. 4391, pp.444 -454 2001

[110] Yi-Hsin Shih; Tsan-Jieh Chen; Chia-Hsiang Yang; Herming Chiueh "Hardwareefficient EVD processor architecture in FastICA for epileptic seizure detection", Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific, On page(s): 1 - 4, Volume: Issue: , 3-6 Dec. 2012

[111] Wei-Yeh Shih; Kuan-Ju Huang; Chiu-Kuo Chen; Wai-Chi Fang; Cauwenberghs, G.; Tzyy-Ping Jung "An effective chip implementation of a real-time eight-channel EEG signal processor based on on-line recursive ICA algorithm", Biomedical Circuits and Systems Conference (BioCAS), 2012 IEEE, On page(s): 192 – 195

[112] Lan-Da Van, Di-You Wu, and Chien-Shiun Chen, "Energy-Efficient FastICA Implementation for Biomedical Signal Separation," IEEE Trans. Neural Networks, vol.22, no.11, pp.1809-1822, Nov. 2011.

[113] F. Sattar and C. Charayaphan "Low-cost design and implementation of an ICAbased blind source separation algorithm", Proc. 15th Annu. IEEE Int. ASIC/SOC Conf., pp.15 -19 2002. [114] Muhammad Tahir AKHTAR, Tzyy-Ping Jung, Scott Makeigy, and Gert Cauwenberghs, "Recursive Independent Component Analysis for online Blind Source Separation," in Proc. IEEE Int. Symp. on Circuits and Systems, May 20-23, 2012.

[115] Rodriguez-Andina, J.J.; Moure, M.J.; Valdes, M.D. "Features, Design Tools, and Application Domains of FPGAs", Industrial Electronics, IEEE Transactions on, On page(s): 1810 - 1823 Volume: 54, Issue: 4, Aug. 2007

[116] Lopez, O.; Alvarez, J.; Doval-Gandoy, J.; Freijedo, F.D.; Nogueiras, A.; Lago, A.; Penalver, C.M. "Comparison of the FPGA Implementation of Two Multilevel Space Vector PWM Algorithms", Industrial Electronics, IEEE Transactions on, On page(s): 1537 - 1547 Volume: 55, Issue: 4, April 2008

[117] A. Jourjine, S. Rickard, and O. Yilmaz. Blind Separation of disjoint orthogonal signals: Demixing n sources from 2 Mixtures. In proc. ICASSP'00, Pages 2985 – 2988, Istanbul Turkey, 2000.

[118] R. Everson and S. Roberts "Blind source separation for non-stationary mixing", J. VLSI Signal Process., vol. 26, no. 1–2, pp.15-23 2000

[119] S. M. Naqvi, Y. Zhang and J. A. Chambers "Multimodal blind source separation for moving sources", Proc. Int. Conf. Acoust., Speech Signal Process., pp.125-128 2009

[120] J.-T. Chien and H.-L. Hsieh "Nonstationary source separation using sequential and variational Bayesian learning", IEEE Trans. Neural Netw. Learn, Syst., vol. 24, no. 5, pp.681 -694 2013

[121] D. J. C. MacKay "Probable networks and plausible predictions—a review of practical Bayesian methods for supervised neural networks", Netw., Comput. Neural Syst., vol. 6, no. 3, pp.469 -505 1995

[122] M. E. Tipping "Sparse Bayesian learning and the relevance vector machine", J. Mach. Learn. Res., vol. 1, pp.211 -244 2001

[123] R. A. Choudrey and S. J. Roberts "Bayesian ICA with hidden Markov sources", Proc. Int. Workshop Independ. Compon. Anal. Blind Signal Separat., pp.809 -814 2003

[124] Q. Huang , J. Yang and Y. Zhou "Bayesian nonstationary source separation", Neurocomputing, vol. 71, no. 7–9, pp.1714-1729 2008

[125] Z. Koldovsky, J. Malek, P. Tichavsky, Y. Deville and S. Hosseini "Blind separation of piecewise stationary non-Gaussian sources", Signal Process., vol. 89, no. 12, pp.2570 -2584 2009

[126] E. Vincent, C. Fevotte, and R. Gribonval. Performance measurement in blind audio source separation. IEEE Trans. Audio, Speech and Language Processing, 14(4):1462–1469, 2006.

[127] E. Vincent, S. Araki, and P. Bofill. The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation. In ICA '09: Proceedings of the 8th International Conference on Independent Component Analysis and Signal Separation, pages 734–741, Berlin, Heidelberg, 2009. Springer-Verlag.

[128] F. Nesta, P. Svaizer, and M. Omologo. Robust two-channel tdoa estimation for multiple speaker localization by using recursive ICA and a state coherence transform. ICASSP, Taipei, Taiwan, 2009.

[129] K. Matsuoka and S. Nakashima. Minimal distortion principle for blind source separation. In Proceedings of International Symposium on ICA and Blind Signal Separation, San Diego, CA, USA, December 2001.

[130] F. Nesta, P. Svaizer and M. Omologo "Convolutive BSS of short mixturesby ICA recursively regularized across frequencies", IEEETrans. Audio, Speech, Lang. Process., vol. 19, no. 3, pp.624 -639 2011 http://bssnesta.webatu.com/testhscma.html

[131] Herbert Buchner, Robert Aichner, and Walter Kellermann. TRINICON: A versatile framework for multichannel blind signal processing. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, volume 3, pages 889–892, Montreal, Canada, May 17-21 2004.

[132] Robert Aichner, Herbert Buchner, Fei Yan, and Walter Kellermann. A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments. Signal Process. 86(6):1260–1277, 2006.

[133] Z. Koldovsky and P. Tichavsky. Time-domain blind audio source separation using advanced component clustering and reconstruction. In Proceedings of HSCMA, Trento, Italy, May 2008.

[134] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C. The Art of Scientific Computing, 2nd. Cambridge, UK: Cambridge University Press, 1992.

[135] K. Whaeed, "Blind Source Recovery: Theoretical Formulations and Application to CDMA Communication Systems", PhD thesis, Michigan State University, 2003.

[136] Gharbi, "Blind Source of Unknown Sources in dynamic Environments: Theoretical Formulation and Micro-Electronic Implementation", PhD thesis, Michigan State University, 1996.

[137] T. Serrano-Gotarredona and B. Linares-Barranco "CMOS transistor mismatch model valid from weak to strong inversion", Proc. Conf. Eur. Solid-State Circuits, pp.627-630 2003

[138] Silicon Moore\'s Law, 2004 :Intel Corporation [online] Available: http://www.intel.com/research/silicon/ mooreslaw.htm

[139] III Frost, O.L. An algorithm for linearly constrained adaptive array processing. Proceedings of the IEEE, 60(8):926–935, Aug. 1972.

[140] Z. Albataineh, F. Salem, "blind multiuser detection DS-CDMA algorithm using H-DE and ICA algorithms" 4th International Conference on Intelligent Systems Modelling & Simulation (ISMS), pp. 569-574, 2013.

 $[141] \ http://www.bloomberg.com/slideshow/2013-09-19/countries-with-the-most-4g-mobile-users.html \# slide1$

[142] http://gigaom.com/2013/09/20/mapping-out-the-worlds-lte-coverage-its-in-fewer-places-than-you-think/