CONTRIBUTIONS TO BLIND SYSTEM IDENTIFICATION/DECONVOLUTION: METHODS, ALGORITHMS AND PERFORMANCE EVALUATION

BY

QADRI MAYYALA

A Dissertation Presented to the

DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

In

ELECTRICAL ENGINEERING

May, 2017

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **QADRI MAYYALA** under the direction of his thesis adviser and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING**.

Dissertation Committee

azzedine zerguine

Prof. Azzedine Zerguine (Adviser)

Prof. Karim Abed-Meraim (Co-adviser)

Assoc. Prof. Mohamed A. Deriche (Member)

Assoc. Prof. Tareq Al-Naffouri (Member)

Assoc. Prof. Ali Muqaibel (Member)

Dr. Ali Al-Shaikhi
Department Chairman

Dr. Salam A. Zummo
Dean of Graduate Studies

14/6/17
Date

©Qadri Mayyala
2017

To My Beloved Parents, Siblings, Wife & Son (Islam), for their Endless Love and Prayers

ACKNOWLEDGMENTS

I am deeply grateful to Allah, the most Compassionate and the most Merciful, who blessed me with all reasons and rewards to get this work done. I pray that this work and the knowledge that I gained during my education will be used in His cause.

I would like to thank KFUPM for providing me the opportunity to study at this reputable institute. KFUPM with its highly professional faculty members nourished my research skills. Also, I acknowledge the Deanship of Scientific Research (DSR) at KFUPM for funding my research work, and the EE Department for the healthy and friendly work environment. In particular, I would like to thank Dr. Maan Kousa and Dr. Ali Muqaibel for giving me the chance to join their team and work on one of the cutting-edge educational projects. Indeed, their exceptional teamwork ability inspired me a lot. I also thank PRISME laboratory at the University of Orleans, the place where I pursued part of my research work over six months, for the nice welcome and host and for providing me with excellent research facilities.

I would like also to express my sincere gratitude to the most warm-hearted adviser Prof. Azzedine Zerguine. He helped me a lot and pushed me to break down all the difficulties and visit France to pursue my research work. I appreciate

his wonderful, lovely, and boundless help and guidance all over the way since I met him in the adaptive filtering course. He influenced me a lot in all life aspects. Thank you so much Prof. Azzedine.

I fell very blessedly to have the most respectful co-adviser Prof. Karim Abed-Meraim. I truly a knowledge the endless hours of guidance that he devoted to me. His research ability and vast mathematical knowledge that he is blessed with, driven my enthusiasm of doing good and worthy research. Without his priceless efforts and guidance, it would not have been possible to do this work. I also thank him for inviting me to the University of Orleans for two times and hosting me at his apartment during my visits there. Indeed, it was a very interesting experience which has a significant impact on my life. Thank you so much Prof. Karim.

I should confess that, besides to the knowledge that I gained from both of my advisers, they influenced me a lot with their decent morals. For me, meeting them was one of the lovely things which have happened in my life.

I would like to thank the thesis committee members, Dr. Ali Muqaibel, Dr. Mohamed A. Deriche and Dr. Tareq Al-Naffouri for the valuable discussions and suggestions which improved the final presentation of this work.

More thanks to my friends at KFUPM and PRISME laboratory for their love, support, and encouragement. It would be impossible for me to forget the time I spent with my friends and colleagues at both institutions during my Ph.D. study. Especially, I thank Syed Awais Shah for the fruitful discussion that we sometimes had, and for sharing some of his codes.

I am also thankful to my lovely parents, brothers, sisters and wife for their continuous support, prayers, and encouragement, as being always the source of every success I am blessed with. The words cannot express their splendid influence on my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENT **v**

LIST OF TABLES **xii**

LIST OF FIGURES **xiii**

LIST OF ABBREVIATIONS **xv**

NOTATIONS **xvii**

ABSTRACT (ENGLISH) **xx**

ABSTRACT (ARABIC) **xxii**

CHAPTER 1 INTRODUCTION **1**

1.1 Channel Estimation 6

1.1.1 Training-Based Channel Estimation 6

1.1.2 Blind Channel estimation 6

1.1.3 Semi-Blind Channel Estimation 8

1.2 Thesis Purpose 9

1.3 Contributions of the Dissertation 14

1.4 Organization of the Dissertation 17

1.4.1 Part I: Blind Deconvolution 18

1.4.2 Part II: Blind System Identification 19

I Blind Deconvolution 21

CHAPTER 2 PRELIMINARIES, SYSTEM MODEL AND BD PRINCIPLE 24

2.1 Chapter Introduction 24
2.2 System Model 26
2.3 Blind Deconvolution 28
2.3.1 Spatio-Temporal Framework 28
2.3.2 Batch Setup 30
2.4 Solution to the BD Problem 31
2.4.1 Pre-whitening and Rank reduction Operation 32
2.4.2 BD Procedure 35
2.5 Indeterminacies in BD 36
2.6 Review of Givens and Shear Rotations 37
2.6.1 The Givens Rotations 37
2.6.2 The Hyperbolic Rotations 39
2.7 Conclusion 40

CHAPTER 3 DECONVOLUTION METHODS 41

3.1 Introduction 42
3.2 Cost Function 44
3.3 Two-Stage Deconvolution Algorithms 46
3.3.1 Full-BSS Deconvolution 47
3.3.2 Equalization-BSS approach 51
3.4 One-Stage Deconvolution Algorithms 52
3.4.1 Deconvolution Using Hybrid Criterion 53
3.4.2 Deconvolution Using Deflation Approach 71
3.5 Simulation Results 76
3.6 Conclusion 83

CHAPTER 4 FAST MULTI-MODULUS ALGORITHMS **84**

4.1 Introduction 85

4.2 Problem Formulation: Revisited 87

4.3 Fast MMA for Instantaneous Mixture 89

4.3.1 Deflation Based Estimation of $\boldsymbol{V}$ (fastMMAd) 91

4.3.2 Full Estimation of $\boldsymbol{V}$ (fastMMAf) 92

4.4 fastMMDA for Convolutive Mixture 93

4.5 MMA with optimized Step Size 95

4.6 Computational Complexity 98

4.7 Simulation Results 99

4.8 Conclusion 104

II Blind System Identification **105**

CHAPTER 5 STRUCTURE-BASED SUBSPACE SYSTEM IDENTIFICATION **107**

5.1 Introduction 108

5.2 Problem Formulation 109

5.2.1 Multi-channel model 109

5.2.2 Subspace method revisited 111

5.3 Structure-Based SS method (SSS) 113

5.4 Discussion 116

5.5 SIMULATION RESULTS 118

5.6 CONCLUSION 122

CHAPTER 6 PERFORMANCE ANALYSIS **123**

6.1 Introduction 124

6.2 Problem Formulation 126

6.2.1 Blind data model 126

6.2.2 Semi-Blind data model 127

6.3 Channel's CRBs Derivations 128

6.4 Unified Framework for Different CRBs' Development 128
6.4.1 Deterministic Gaussian CRB 131
6.4.2 Stochastic and Non-Circularity Based CRB 132
6.4.3 Deterministic Semi-Blind CRB 134
6.4.4 Bayesian CRB 135
6.4.5 Bayesian Semi-Blind CRB 135
6.4.6 Structurally-Aided CRBs (Sparse Channel) 136
6.5 Discussion 139
6.6 Experimental Performance Analysis 140
6.7 Conclusion 145

CHAPTER 7 CONCLUSION & FUTURE WORK 146
7.1 Conclusion 146
7.2 Future Work 149

APPENDIX A 152

APPENDIX B 154

APPENDIX C 156

APPENDIX D 158

REFERENCES 160

VITAE 180

LIST OF TABLES

1.1 Literature Classification. 13

3.1 G-MMDA and HG-MMDA Algorithms 64
3.2 G-DMMDA and HG-DMMDA Algorithms 75

4.1 fastMMAd Algorithm 93
4.2 fastMMAf Algorithm 94
4.3 Numerical Cost of Different BSS Algorithms 98

5.1 Duality Table 113

6.1 Summary of CRBs 140

LIST OF FIGURES

1.1 Training-Based, Blind and Semi-Blind Principles. 10

2.1 MIMO wireless communications system. 25
2.2 MIMO wireless communications system with QAM signaling. . . . 26

3.1 Two-Stage Full-BSS Deconvolution Approach. 47
3.2 Paring and sorting mechanism. 48
3.3 Two-Stage Equalization-BSS approach. 51
3.4 Average SER of "Full-BSS" algorithms (using ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. N_s with $SNR = 30\ dB$ and 4-QAM. 77
3.5 Average SER of Equalization-BSS (using ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. N_s with $SNR = 30\ dB$ and 16-QAM. 78
3.6 Average SER of G-MMDA and HG-MMDA a vs. N_s with $SNR = 30\ dB$ and 4-QAM. 79
3.7 Average SER of G-DMMDA and HG-DMMDA a vs. N_s with $SNR = 30\ dB$ and 4-QAM. 80
3.8 Average SER of Full-BSS algorithms (ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. SNR for $N_s = 250$ and 4-QAM. 80
3.9 Average SER of Equalization followed by BSS (using ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. SNR for $N_s = 120$ and 16-QAM. 81
3.10 Average SER vs. SNR for the G-MMDA and HG-MMDA algorithms at $N_s = 300$ and 4-QAM. 81

3.11 Average SER vs. SNR at both $N_s = 200$ and $N_s = 350$ and 4-QAM. 82

4.1 Speed of Convergence ($\mathcal{J}$ vs. N_{Sweeps}), for VSS and FSS, at SNR=30 dB, 4-QAM case. 100
4.2 Speed of Convergence ($\mathcal{J}$ vs. N_{Sweeps}), for VSS and FSS, at SNR=30 dB, $N_t = 5, N_r = 7, N_s = 200$, 16-QAM case. 100
4.3 Average SER for different CMA and MMA, for different scenarios, 4-QAM case. 101
4.4 Average SER for different CMA and MMA at different constellation setups, $N_t = 3, N_r = 7, N_s = 300$. 102
4.5 Average SER for HG-MMDA, G-MMDA, and fastMMDA at 4-QAM and 16-QAM, $N_t = 3, N_r = 5, M = 2$. 103

5.1 Well-conditioned channels, $\theta = \pi/10, \delta = \pi$. 120
5.2 Ill-conditioned system, $\theta = \pi/10, \delta = \pi/10$. 120
5.3 MSE versus δ, $SNR = 15\ dB$ 121
5.4 MSE versus SNR for different window size N_w. 122

6.1 NMSE for DCRB$_{blind}$,BCRB$_{blind}$, DCRB$_{SB}$ and BCRB$_{SB}$. 141
6.2 Impact of using sparse structure on the CRB, when different channels' coefficients are overlapping and not. 142
6.3 Impact of channels' non-zero entries order overestimation in sparse scenario. 143
6.4 Impact of using specular channel structure on the CRB. 143
6.5 a. Impact of using non-circularity of the source, b. Impact of non-circularity factor $|\rho|$ on the CRB 144

LIST OF ABBREVIATIONS

ACMA	Analatical Constant Modulus Algorithm
AWGN	Additive White Gaussian Noise
BD	Blind Deconvolution
BSI	Blind System Identifications
BSS	Blind source separation
CRB	Cramer-Rao Bound
CM	Constant Modulus
CMA	Constant Modulus Algorithm
fastMMAf	fast Multi-Modulus (full-based) Algorithm
fastMMAd	fast Multi-Modulus (deflation-based) Algorithm
fastMMDA	fast Multi-Modulus Deconvolution Algorithm
FIR	Finite Impulse Response
ISI	Inter-symbol Interfere
ISI	Inter-User Interference
QAM	Quadrature Amplitude Modulation
LTE	Long Term Evolution
MIMO	Multiple-Input Multiple-output
MM	Multi-Modulus
MMA	Multi-Modulus Algorithm
MMSE	Minimum Mean Squared Error
NMSE	Normalized Minimum Mean Squared Error
PAM	Pulse Amplitude Modulation
PSK	Phase-Shift Keying
QAM	Quadrature Amplitude Modulation
SIMO	Single-Input Multiple-output
SISO	Single-Input Single-output
SNR	Signal to Noise Ratio
SS	Standard Subspace Method
SSS	Structure Based Subspace Method
SVD	Singular Value Decomposition
TTM	Truncated Transfer Matrix Method
WLAN	Wireless Local Area Networks

NOTATIONS

$\boldsymbol{z}$	Column vector
z_i	ith entry of a vector
z_{ij}	(i,j)th entry of a matrix
$\boldsymbol{Z}$	Matrix
$\boldsymbol{I}_m$	Square identity matrix of dimension m
$(.)^T$	Array transpose
$(.)^*$	Complex conjugate
$(.)^H$	Array complex conjugate transpose
$(.)^+$	Pseudo inverse
$E(.)$	Statistical expectation
δ_{nm}	Kronecker delta. ($\delta_{nm} = 1$ if $m = n$ and 0 otherwise)
$(.)_R$, $(.)_I$	$Re(.)$ and $Im(.)$
$\|.\|_2$	Euclidean norm
$\otimes$	Kronecker Product

$\odot$	Hadamard Product
vec$\{.\}$	Operator maps the matrix into vector form, i.e., $\mathbb{C}^{a\times b} \rightarrow \mathbb{C}^{ab\times 1}$
mat$_{a,b}\{.\}$	Operator maps the vector into matrix form, i.e., $\mathbb{C}^{ab\times 1} \rightarrow \mathbb{C}^{a\times b}$
$avg\{.\}$	The average value of the variable in brackets
$tr\{.\}$	The trace of the matrix in brackets
$diag\{.\}$	Diagonal Matrix where the main diagonal is the vector in brackets

THESIS ABSTRACT

NAME:	Qadri Mayyala
TITLE OF STUDY:	Contributions to Blind System Identification/Deconvolution: Methods, Algorithms and Performance Evaluation
MAJOR FIELD:	Electrical Engineering
DATE OF DEGREE:	May, 2017

Blind source separation (BSS) and blind system identifications (BSI) date back to 1970s. Due to the rapid growth of the wireless communication systems, this subject starts receiving an increase attention as it leads to an increased throughput in case of deploying it in conjunction with the training-based channel estimation approach, namely semi-blind scenario. So far, it turns out that the standing literature related to the BSS and BSI subjects is not well addressed for the emerging semi-blind system identification technique.

This work addresses some of the earlier stated issues. In particular touches upon BSI, BSS and blind deconvolution (BD) in the context of communication systems. One of the main objectives is to use the side information which is available

in such systems to obtain much efficient blind algorithms/methods and evaluate their impact on the estimation quality. This is done by using different sufficient optimization techniques, namely Givens/Shear rotations, fixed point optimization rule, and exact line search strategy. In the first part,we use the non-constant modulus nature of QAM signals to propose dedicated BSS algorithm for instantaneous mixtures. Furthermore, to cope with a real case scenario, we extend different BSS algorithm to perform BD and cover convolutive MIMO mixtures.

In the second part, the BSI problem in the context of SIMO systems is treated. Here, we propose a novel BSI method, which exploits the inherent Toeplitz structure that exists in most of communication systems. This method indeed leads to a significant enhancement over the standard subspace method (SS) in terms of performance, especially in the ill-conditioned scenario. On the other hand, by means of the Cramer-Rao lower bound (CRB), the influence of some a priori side information on the channel system identification is investigated. In particular, a priori information includes channel sparsity and some other statistical properties such as the non-circularity. These side information are classified into statistical and structural ones. Their impact on the channel system identification's quality is handled at the end of this part.

CHAPTER 1

INTRODUCTION

In the last two decades, research efforts succeeded in introducing new technologies to mankind. This has stimulated researchers appetite to advance the technology further and further to cope with the market desire for fast, accurate and robust systems. The advancement includes digital signal processing, antenna theory, and semi-conductors fabrications. Telecommunications have been also witnessing a very fast development. Yet, this development is confronted with very fundamental problems which exist due to the system's nature. Limited channel bandwidth and the introduced Inter-symbol Interfere (ISI), due to the multipath propagation behavior of the wireless communication systems, are among these fundamental limitations. Besides, these systems suffer from the Inter-User Interference (IUI) in the case of sharing the common radio resources among multiple users [1].

Many transmission, modulation, configuration and coding schemes have been successfully introduced to overcome the challenges of the limited bandwidth and provide the user with high-speed wireless communications [1, 2]. Traditionally,

training data sequences/Pilots are used to estimate the channel state information and this is a kind of wasting the communication resources.

The subject addressed in this thesis has been inspired with different new research efforts in the context of wireless communications. The main motivation is the lack, as will be shown later, of existing dedicated blind algorithms for certain kind of modulation, such as QAM signaling, which is heavily used in many standards. Also, modern wireless communications are very demanding in terms of throughput, high transmission rate, and oriented to become greener. The use of blind techniques has been considered for many purposes including waiving the training sequence [3], reducing the pilot size using semi-blind approaches [4], or mitigating the pilot contamination problem in massive MIMO systems [5]. In particular, it has been shown in [4, 5] that, thanks to the semi-blind solution, the pilot sequence size can be reduced by a factor of 80 to 90% without affecting the channel estimation quality. Besides, in certain systems, this represents close to 25% of energy reduction which is the main objective of the current research activities for 'Green Communications'. These motivations were mainly behind this research work. In this thesis, we contribute to the non-stop streaming contribution in the literature throughout many axes.

The first axis targets the convolutive communication channels which usually arises in wireless communication systems. Mainly, Multi-Modulus (MM) criterion is used along with the simple unitary and/or J-unitary rotations to achieve the Blind Deconvolution (BD). Four different BD classes have been proposed. The

first two classes are classified as two-step method, while the rest are one-step recursive methods. The first class performs the BD by applying Blind Source Separation (BSS) routine on a spatio-temporal system model, this results in the separation of each source with its possible replicas where second order statistics (SOS) rule is used to be able to perform pairing and sorting. The second class is basically based on inverting the consequence of multipath effect by using certain channel estimation and equalization algorithms. This reduces the convolutive system into an instantaneous one and hence enabling the BSS step. While the third method is based on optimizing a hybrid cost function and the last class is a deflation based method and is conducted by extracting a single source each time followed by a course of orthogonalization technique based on the SOS approach. These contributions come as a natural extension of the recently proposed BSS algorithms in [6].

The main rational behind introducing the four different classes of deconvolution algorithms is to present an extensive comparison between the potential method that can be employed to identify and separate signals of a predefined constellation structure. Moreover, these class of algorithms, which are mainly derived based on the elementary Givens and Shear rotations, in addition to their simplicity and low computational cost as compared to other state of the art methods, show to a have a very interesting convergence behavior [6, 7, 8].

In the same context, we developed another new class of BSS algorithms which are shown to be fast and less costly as compared to the above described methods.

These methods are given in two implementations: the first one is a deflationary MM-based method which is able to extract the sources one after another, while the second one is a full MM-based method which is able to recover all sources at once. These are obtained in favor of the fixed point optimization (FPO) rule. The later implementation is shown to have an improved estimation quality. Moreover, the deflationary MM-based implementation is extended to cover systems of convolution nature and perform the blind deconvolution in an efficient way. Furthermore, these algorithms are shown to belong to the gradient descent fixed step size iterative methods with a similar update recursive equation. In doing so, the exact line search technique is used to derive an algebraic optimal step size and eventually, by tolerating a slight increase in the computational cost, boosting the algorithm's speed of convergence further.

The second axis contributes to the multi-channel Blind System Identification (BSI) problem. In particular to the case of ill-conditioned system identification, whereby channel's zeros become close, and hence most of the existing BSI fails to identify the system's response. In this regard, we exploit directly the impeded Toeplitz channel structure in the signal linear model to build a quadratic form whose minimization leads to the desired channel estimation up to a scalar factor. This method can be extended to estimate any predefined linear structure, e.g., Hankel, that is usually encountered in linear systems. As will be shown later, this method can be shown as a dual approach of the standard subspace (SS) method.

In brief, the SS method achieves the channel estimation by exploiting the

subspace information (i.e., orthogonality principle between signal and noise subspaces) as well as the block Sylvester (block-Toeplitz) structure of the channel matrix. More precisely, it enforces the latter matrix structure and minimizes the subspace orthogonality error. As will be shown later, our proposed method, which is shown to be a dual approach of the SS method, enforces the subspace information while minimizing a cost function representing the deviation of channel matrix from the Sylvester structure in the mean square sense. Theoretical and experimental findings support our method for being able to maintain satisfactory performance under challenging environments such as poor channel diversity (ill-conditioned) scenario and small window size.

In the last axis, we explore alternative ideas to improve the channel estimation and quantifying their impact. Such ideas are based on the side information which is usually available in most of the standing systems. The side information can be deployed to assist in the channel estimation and it can be of different shapes: structural, such as sparsity, or statistical, such as cyclostaionarity, non-circularity and Gaussianity. The impact of this side information is measured by the Cramer-Rao lower bound (CRB). Moreover, most of the popular blind SOS methods fail when the channel length is overestimated. Therefore, we evaluated the response of both blind and semi-blind approaches against the channel length overestimation effect by means of the CRB. This part is developed to explore the advantages of utilizing the side information on the estimation problem.

1.1 Channel Estimation

Three major sorts of channel estimation are encountered in the literature which are briefly discussed in the ensuing subsections.

1.1.1 Training-Based Channel Estimation

As discussed earlier, wireless communication systems crucially suffer from ISI and IUI. Both emerge an extra processing and waste of the scarce wireless communication resources, i.e., power and bandwidth. That said, the design of the conventional communication receivers mitigate the earlier mentioned problems either by knowing the communication channel or by estimating it by requesting an access for the transmitted signals throughout transmitting pilots. The later is the case in the standing wireless communication systems. Although, in the case of invariant channels, the training signals insignificantly lessen the throughput, yet it becomes significant for the time variant scenario. For instance, in high-frequency communication systems, the pilot transmitting time can be as much as 50% of the total transmission [9].

1.1.2 Blind Channel estimation

Estimating the channel state information based on the received data only without depending on the training pilots refers to the blind channel estimation technique. At first glance, this estimation seems to be untraceable, but it does! How would it be possible to separate the source from the channel when no one of them is

known?

The existence of blind channel estimation resides in its ability to exploit the properties and/or the structure of the input and/or the channel. A well-known scenario is when the input has a known probabilistic model or statistics. In such case, the separation is possible, for example, in communication systems the input signals have a finite alphabet property or sometimes exhibit cyclostationarity. The later property was deployed in [10] to show that the SOS can be successfully used to identify non-minimum phase channels. This led to an immense development of many subspace-based algorithms. Most of SOS based techniques impose diversity conditions and deteriorate in case of channel overestimation. This concludes that SOS techniques should not be used alone in communication systems and this is perhaps the reason for not using blind techniques in this context. For more information about SOS based methods and overview see [11] and the references therein.

Besides to the SOS approach, there is a family of algorithms which is based on the High Order Statistics (HOS). These algorithms waive the diversity conditions and accordingly can be applied for single-input single-output (SISO) systems. In the contrary to SOS-based algorithms which are phase blind methods, HOS statistics, which are known as cumulants and their respective Fourier transforms known as poly-spectra, reveal both amplitude and phase information. On the other hand, cumulants are blind to all kind of Gaussian processes, whereas the correlation (SOS-based) is not. Consequently, since many real world scenarios are

basically non-Gaussian, HOS are applicable to deal with non-Gaussian processes successfully.

HOS-based algorithms are classified into two classes [12]: implicit and explicit methods. The earlier, which uses the HOS implicitly, includes Sato algorithm [13] and Godard-2 algorithm (or constant modulus algorithm (CMA)) [14]. However, the explicit methods employ directly the cumulants or the poly spectra criterion, see for example in the case of SISO systems [15] and [16], and for multi-channel systems [17].

The fundamental drawback of the cumulant-based methods is that they require a much longer data samples than the correlation-based methods. This is needed in order to reduce the variance associated with the estimation of the higher order moments from the sample averaging techniques. For a rich overview of the HOS ttechniques kindly refer to [17] and [18].

1.1.3 Semi-Blind Channel Estimation

Lastly, hybridized techniques are traditionally emerged to solve pitfalls that may arise in one or both of the hybridized techniques. Hence, one might conclude that the remedy of the drawbacks revealed in both parties; training based and blind-based channel estimation might lie in the semi-blind techniques. To understand the meaning of this new terminology, we commence with the definition of both of the hybrid methods. Training based methods depend completely on the known received samples to retrieve the channel's response, hence the unknown samples

are ignored. On the other hand, blind techniques rely fully on the received data which contain known and unknown samples and using the characteristics of the input samples, for instance, i.i.d while ignoring the possible use of some known data. Now, the target of the semi-blind methods is to rely on the received data and use a few known data (see Fig. 1.1). Hence, one gain the positive aspects of both, the blind and the training-based methods. Also, these techniques allow estimating a longer impulse response with short training pilots. Such application is of special interest for wireless communication applications, especially in mountainous and harsh areas. Moreover, in the case of SOS methods, only one pilot is needed to make any channel to be identifiable. Besides to the aforementioned advantages, the semi-blind methods are very attractive from a performance point of view as compared to both individuals techniques. With this, semi-blind methods are becoming very promising and popular to be deployed in the next generation of wireless communications. For an overview of the semi-blind estimation methods, the reader is urged to refer to [19].

1.2 Thesis Purpose

It is not surprising arguing that wireless communications standards continue to develop to raise the data throughput capability. Boosting the data rate is basically accomplished through enhancements in the physical layers of the protocols, which usually take years to show up and see the light. Nowadays, there are two hottest wireless standards namely IEEE 802.11ac which is used in wireless local area

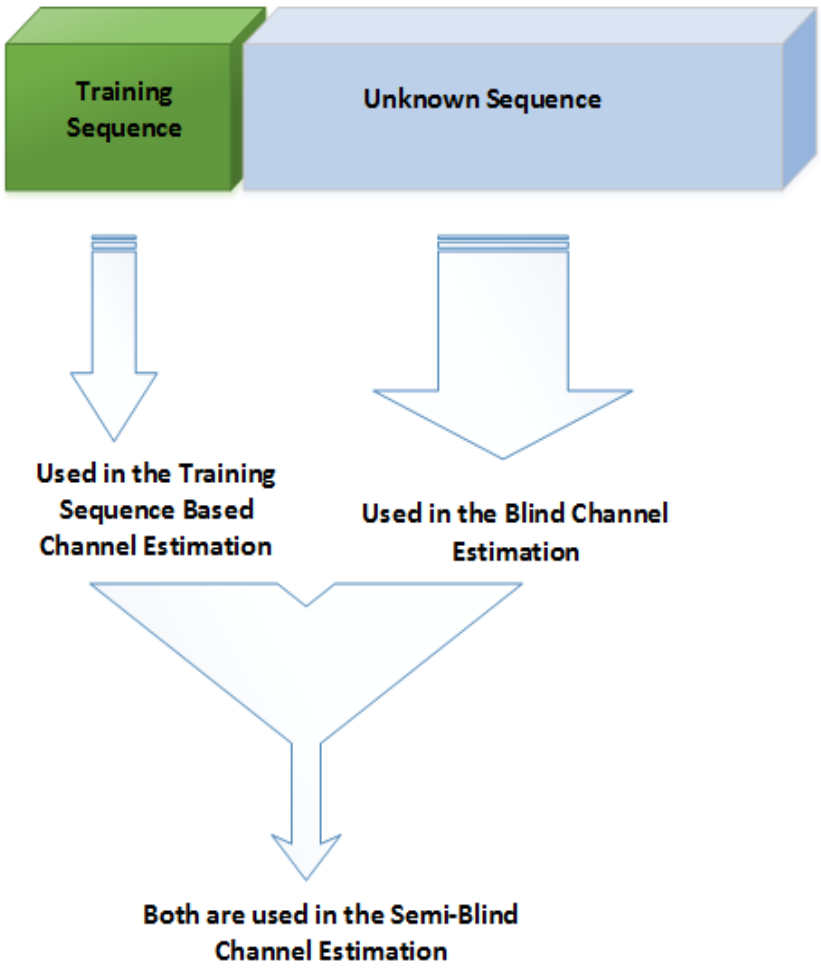


Figure 1.1: Training-Based, Blind and Semi-Blind Principles.

network (WLAN) products and marketed under the name of Wi-Fi, and 3GPP LTE-Advanced used in cellular communications. These standards featuring more spatial streams (i.e., more MIMO channels), wider bandwidths, and higher order modulations as compared to the earlier standards such as IEEE 802.11a/g/n. All of these standards use the regular channel estimation throughout injecting some known pilots. Lately, research efforts in [4] and [5] revealed another key feature, thanks to the semi-blind solution, which is able to increase the throughput further by reducing pilots sequence. It is shown that in some scenarios, a reduction up

to 90% of pilots without affecting the channel estimation quality can be achieved. This, as well, corresponds to 25% power saving which is a main trend of the current green communications.

Standing literature leaves the door open for new research in the field of blind identification and source separation for SIMO/MIMO systems. Among the earlier mentioned features which enable the new wireless communication systems, MIMO systems are receiving an increased interest. As the existing communication systems are build based on the training pilots philosophy to estimate the channel information, the literature arena is very rich in this topic. However, in the case of blind MIMO system identification and BSS for wireless communications, there is a shallow advancement in the advised methods and techniques, probably due to the dominance of the training-based approach. Also, the band-width efficient constellation-like signals, i.e., QAM and PSK are used tremendously in wireless communications, as outlined earlier, they are one of the key features for improving throughput. Since the modulation techniques used in communication system are known a priori, this property has been used recently to maintain the same property for the output signals and hence recover the source signals blindly. Depending on the source signal, different cost functions can be found in the literature to perform the source demixing [20, 21], including the CM criterion for phase/frequency modulated signals and MM criterion for QAM signals.

The CM criterion is used tremendously and lead to a number of CM algorithms (CMA) used in blind equalization [22], blind beamforming [23], and BSS [24,

7]. Moreover, there is a more general family of algorithms, known as multiuser Kurtosis (MUK) [25], which are obtained by minimizing the kurtosis as a contrast function. Further, The MM criterion [26] which is dedicated for square QAM signals lead to a plenty of MM algorithms (MMA) used in different applications, including blind equalization [26] and BSS. The BSS is performed either in an adaptive way [27, 28] or in a batch way [7, 6]. In the existing literature, it is shown that MMA outperforms both the CMA and MUK, specifically for the case of QAM signals. So far, the non-constant modulus signals' BSS is not explored satisfactorily.

On the other hand, as for convolutive MIMO mixtures, there are limited efforts which have been done for providing successful extensions for the earlier mentioned algorithms, namely MUK, CMA, and MMA. For instance, the adaptive MUK has been extended in [29] and the CMA in [30], where all were implemented in an adaptive manner. More importantly, the existing MMA, which fits the phase modulated signals, do not address the convolutive MIMO systems, which are more practical. Table 1.1 summarizes the relevant literature. In addition, among the different approaches which are followed in performing the extensions, there is a lack of the material which compares and outlines the main differences among them.

As mentioned earlier, the semi-blind technique is becoming very attractive especially for mitigating the pilot contamination problem and increasing the throughput gain. Hence, some new dedicated and efficient algorithms need to

be developed in order to open the door for such promising approaches.

BSS Alg.	MUK	CMA	MMA
Instantaneous (Adaptive)	[25, 31, 32]	[7, 33, 34]	[27, 28]
Instantaneous (Batch)	[35, 36]	[24, 25, 37, 38, 39]	[6]
Convolutive (Adaptive)	[29]	[30, 40]	open
Convolutive (Batch)	[41]	open	open

Table 1.1: Literature Classification.

Also, as can be seen in Chapter 5, multi-channel FIR systems encountered in certain applications could suffer from ill-conditioned scenario [42, 43]. This might happen in the case of high receiver diversity, or when the channel response is of small tails [44]. A shared weakness of most of SOS methods, including the standard subspace method (SS) [45, 46], the cross-relation (CR) method [47] and the two-step maximum likelihood (TSML) method [48], is that their performance is poor under ill-conditioned case. Despit of several iterative approaches have been proposed to put up with the ill-conditioned channels [43, 49], these solutions suffer from local minima problem and hence results in wrong channel estimation.

As will be shown in Chapter 6, the existing side information that can be used in blind system identification methods falls into two categories: statistical, i.e. non-circularity, Gaussianity, cyclostationarity, and structural, i.e., sparsity. The statistical side information has been quantified partially in case of MIMO systems [50, 51, 52, 53, 54, 55] and in case of SIMO systems [50, 56, 57, 58, 59, 60, 61, 62]. Whereas, the structural one has not been explored fully in this context so far.

Moreover, most of the existing blind system identification methods' performance depends on the robustness and accuracy of the channel order estimation.

The solution for this deelma can be classified into two approaches: the first approach is based on developing methods for estimating the channel order accurately [63, 64, 65, 66, 67]. The second one is based on developing robust blind method which are able to tolerate the channel order mismatch [49, 68, 67].

Interestingly, the promising semi-blind technique can also be shown as a new approach to compete this shortcoming. This is gained thanks to the lest squares structure impeded in this approach. However, so far, there is no exploration on the impact of these added pilots on the degree of channel overestimation. Hence, it is important to quantify and explore the effect of the structural side information as some did not explore statistical yet the side information on the blind system identification problem.

1.3 Contributions of the Dissertation

This dissertation touches upon different axises in the field of blind system identification, blind source separation, blind deconvolution, and performance evaluation. Meanwhile, the ultimate goal of this dissertation is to contribute opportunistically into this growing area of research by proposing new methods to fill some existing gaps in the literature (where it is applicable) and propose performance-wise and cost-wise enhanced techniques. This is to increase the opportunity of the powerful blind methods to be deployed in the wireless communications and probably in some other emerging applications.

The contributions of this dissertation are in the multi-channel systems, i.e,

SIMO and MIMO systems, in wireless communications. As will be discussed in the last chapter, other applications are possible. The contributions are summarized as follows:

- We filled the existing gap in the literature by proposing a new class of deconvolution methods to recover the constellation-like signals. This is achieved by optimizing the MM criterion using the efficient unitary/non-unitary elementary rotations. The work resulted into four different approaches which are presented in Chapter 3.

- Review and compare the state-of-the-art deconvolution approaches which can be considered to perform BD.

- In match with the earlier contributions, we proposed new methods which are more competitive and further efficient. The new methods are able to perform the BSS for instantaneous mixtures with less computational cost and improved performance quality. This is done by minimizing the MM criterion using the simple Fixed Point Optimization (FPO) rule. These algorithms are extended to perform blind deconvolution in a deflation way. Furthermore, we were able to boost the speed of convergence further by proposing an algebraic optimal step size using the exact line search strategy. This, eventually, lead to a new family of variable step size (VSS) algorithms.

- We proposed a novel structural-based subspace method. Although this new method is argued to be a dual approach of the standard subspace method, it

has an interesting gain, especially in the ill-conditioned scenario where the standard subspace method has a degraded performance in this environment.

- We explored the side information's impact, which usually exists in most of the communications systems, on the channel estimation quality. In particular, side information includes channel's sparsity and some other statistical properties such as the non-circularity. These side information are shown to be either structural or statistical. By means of the CRB, we demonstrated that these freely available side information, if they are properly exploited, would improve the channel estimation quality. Moreover, in the context of semi-blind system identification, we explored the channel overestimation consequence on the estimation quality.

Eventually, the contributions of this thesis are published/submitted in:

1. Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "Structure-Based Subspace Method for Multi-Channel Blind System Identifications",*IEEE Signal Processing Letters*, Accepted.

2. Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "Fast Multi Modulus Algorithms", submitted to *IEEE Transactions on Communications*

3. Qadri Mayyala, K. Abed-Meraim, and A. Zerguine, "A Class of Multi-Modulus Blind Deconvolution Algorithms Using Hyperbolic and Givens Rotations for MIMO Systems", submitted to *IEEE Transactions on Communications*.

4. Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "New Blind Deflation-Based Deconvolution Algorithms Using Givens and Shear Rotations", *Proc. of IEEE ICC*, Paris, France, 2017.

5. Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "On the Performance Evaluation of Blind System Identification in Presence of Side Information", *Proc. of IEEE IWCMC*, Valencia, Spain, 2017.

1.4 Organization of the Dissertation

Besides to the introduction chapter, this dissertation comprises two parts and a general conclusion chapter. The first part comprises Chapter 2, Chapter 3 and Chapter 4. In this part, we treat the general deconvolution problem blindly and we propose a class of BD algorithms which are dedicated for the non-constant modulus signals. In the second part, we elaborate on the Blind System Identification (BSI) problem and propose a subspace dual approach which exploits the convolutive nature of the communication channel. This is achieved by optimizing directly the Toeplitz structure that usually appears anywhere convolutive system exists.

Ultimately, a comprehend conclusion chapter is closing this dissertation, in addition to some perspectives for potential future work.

1.4.1 Part I: Blind Deconvolution

Blind deconvolution is a terminology referred to a composite task of performing equalization as well as BSS. It is usually encountered in communication systems where the multipath effect (results into ISI) and multi-user access (results into MUI) appear. BD includes a wide spectrum of techniques which are used to recover the sources from its convolved mixtures. Most of the existing algorithms assume the mutual independence between the transmitted sources. This assumption is pretty general. In addition to this assumption, some other assumptions characterizing the transmitted sources have recently been used, such as time coherence [69], cyclostationarity [70], sparsity [71, 72], bounded magnitude [73], kurtosis [25, 29], constant modulus [23, 7, 40] and multi-modulus [6], just to mention a few.

In this part, based on the fact that many communication signals have the constant and non-constant modulus property such as the phase modulated signals, i.e., PSK and QAM, we adopt the MM criterion.

Part one comprises the first three chapters target the BSS and BD problem in the context of wireless communication systems. In Chapter 2, a preliminary one, some of the used terminologies are covered. Also it outlines the system's model, and presents the solution approach. Chapter 3 mainly forms an extension and a generalization of [6] where BSS for instantaneous mixtures was extended to cover the convolutive mixtures. The efficient unitary and non-unitary Givens and Shear rotations are used to solving this problem in conjunction with other techniques

such as orthogonalization, sorting and pairing, and enabling the hybrid criterion. We were able to come up with a class of algorithms which summarizes and adds to the existing methodologies in the literature.

Besides, in Chapter 4, we devise even more efficient methods, called fastMMA algorithms, which are experiencing faster convergence and low computational complexity. These algorithms cover both BSS and BD problems. Eventually, this chapter ended by proposing an optimal algebraic variable step size using the exact line search technique. The impact of the VSS is demonstrated in its ability to boost the speed of the convergence of these algorithms and enable them to converge in a single iteration in certain scenarios.

1.4.2 Part II: Blind System Identification

The idea of blind system identification is built upon the use of available side information to substitute for the lack of the training phase. This side of information can be argued to be any free piece of information which is naturally impeded in the signals and/or the systems. By doing so, one can develop dedicated algorithms to perform system identification. For instance, the well-known subspace methods [11, 45] exploit the subspace structure to minimize the orthogonality between noise and signal subspaces. These kind of algorithms are attractive; since the channels estimates can be obtained in a closed form through optimizing a quadratic cost function, which is obtained here by minimizing the orthogonality property.

In this part, Chapter 5 introduces a dual subspace-based method based on the

channel's Toeplitz structure which is employed directly to formulate our quadratic cost function. This is of great interest since the Toeplitz structure is an inherent nature that exists in most of the linear systems due to their convolutive nature. The proposed method gain is demonstrated with a particular emphasis on the ill-conditioned channels.

In Chapter 6, we elaborate more on different side information and we pose the following question: What can different side information bring into the BSI problem in terms of estimation quality?

Various structural and statistical side information, such as the channel's sparsity and signals' non-circularity, have been considered. Accordingly, their effects have been assessed in terms of the CRB lower bound. These evaluations are conducted for the blind scenario, however, the semi-blind scenario is covered by evaluating their tolerance to the channel overestimation problem.

Eventually, Chapter 7 concludes the covered work in this dissertation and accordingly motivate some possible future work.

Part I

Blind Deconvolution

The first part composes three Chapters, which addresses both BD and BSS problems. These problems are investigated in the context of communication signals, in particular, the non-constant modulus signals. Mainly, four contributions are presented in this part, i.e., (i) propose, for the first time, four deconvolution algorithms based on the alphabet nature of the constellation signals (MM criterion) using the elementary Givens and Shear rotations, (ii) review and compare the state-of-the-art deconvolution approaches (i.e., the four previously mentioned solutions) that can be considered to perform BD, (iii) propose a novel set of fast and efficient BSS multi-modulus algorithms (MMA) and expand them again to cover the general BD problem, and (iv) boost the speed of convergence, for the former set of algorithms, further by proposing an optimal algebraic VSS strategy.

In the sequel, Chapter 2 formulates the BD problem, develops the solution methodology, and discusses the need for the pre-whitening as a preprocessing step on the received data before it starts beamforming or BSS. Furthermore, elaborates on some preliminaries such as the inherent permutation and scale ambiguity which exist in the BSS/BD problems. Finally, reviews the used rotations, i.e., Givens and Shear, to perform the BD in the subsequent chapter.

Chapter 3 proposes four different approaches, which resulted in eight algorithms, to perform the BD for QAM signals. These algorithms are classified into two main groups, the first one is able to perform the BD in two steps, while the second class is a one -step methods. Eventually, this chapter highlights the main differences between these approaches in terms of complexity, design, and

performance.

In the same context, Chapter 4 addresses again both BSS and BD problem and proposes much faster algorithms as compared to the class of algorithm which are recently presented in [74] and extended in Chapter 3. This is done using the FPO method. Interestingly, these algorithms are shown to be free of any user-defined parameters. Also, so far they are much cheaper than all existing counterpart. Finally, an optimal VSS is proposed and integrated with the proposed algorithms. This VSS leads to much faster algorithms.

CHAPTER 2

PRELIMINARIES, SYSTEM MODEL AND BD PRINCIPLE

2.1 Chapter Introduction

MIMO systems are becoming of great interest in differnt signal processing applications, including speech processing [75], bio-medical applications [76], multi-access communication [77], multi-sensor/radar systems [78], [79]. In wireless communications systems, for instance, usually N_t devices transmitting at the same time using a shared frequency band, such systems are called muli-access communications systems, which is described in Fig. 2.1. However, at the receiver side, the signal is received by N_r element array, and the assembled signal at each element can be viewed as a weighted sum of the N_t transmitted signals. Besides, due to the likely multi-path fading effect, this weighted sum of the transmitted waves would be a sum of a convolved versions of the transmitted waves in the mullti-path

channels. Usually, the multi-path channels are modeled with an FIR filter.

To recover the transmitted signals at the receiver side, the channel state information is crucially need to be known. In doing so, the effect of the channel is waived using the so called equalization process. Classically, training sequence (pilots), which are known a priori for both sides, are periodically transmitted to help the receiver in revealing the channel information. Although, this approach is shown to be a very successful approach, it still scarifying the communication resources, i.e., bandwidth and power. For example, in wireless distributed systems, usually the users are battery powered, and pilot transmission reduces the battery life, also it reduces the data transmission rate. As for now, both battery life and transmission rate are of great interest especially in the obligation current market.

In comparison, blind MIMO system identification approaches, which is described in the subsequent section, only requires some a priori structural and/or statistical information about the transmitted sources and/or the channel model, and eventually works with no need for the training phase.

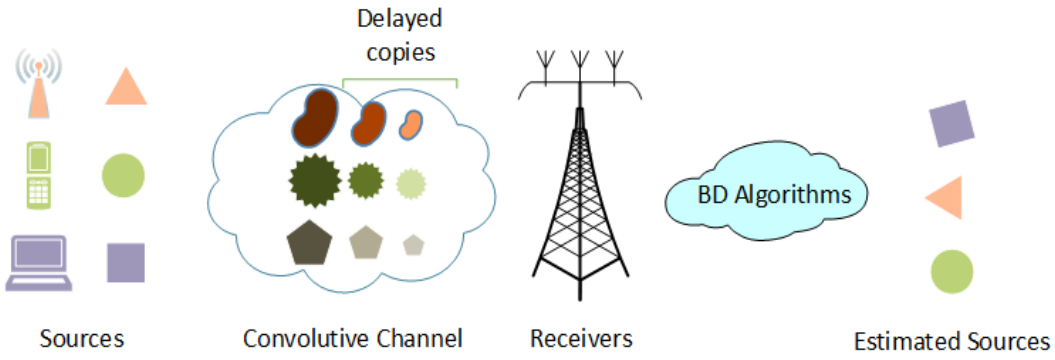


Figure 2.1: MIMO wireless communications system.

2.2 System Model

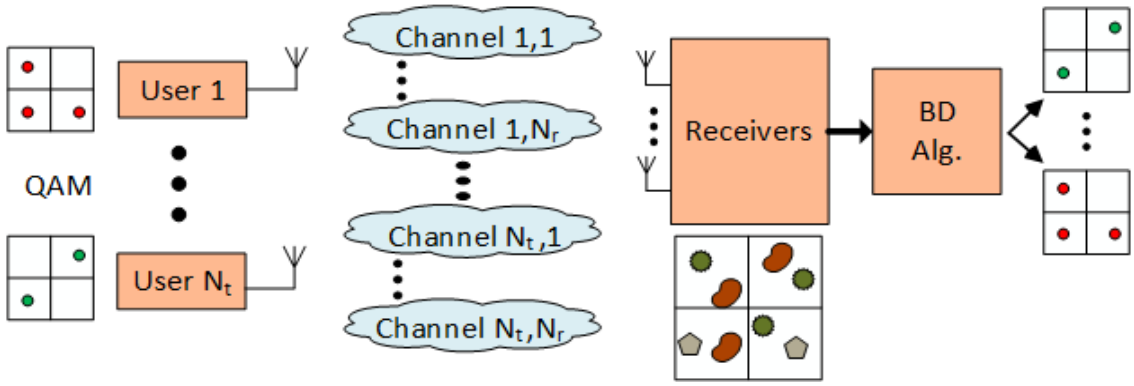


Figure 2.2: MIMO wireless communications system with QAM signaling.

Figure 2.2 reflects the considered MIMO wireless communication system: N_t transmitters emit signals independently from QAM constellation. They use the same time slot and frequency band. Each transmitted QAM signal $\boldsymbol{s}(k) = \boldsymbol{s}_R(k) + j\boldsymbol{s}_I(k)$ is drawn from T-ray square constellation, i.e., $\boldsymbol{s}_R(k), \boldsymbol{s}_I(k) \in \left\{\pm 1, \pm 3, \cdots, \pm(\sqrt{T}-1)\right\}$. The transmitted data is assumed to pass through a noisy frequency selective (convolutive) channel, i.e., ISI is presented, so that the collected $N_r > N_t$ samples at time instant k can be modeled as:

$$\boldsymbol{y}(k) = \sum_{n=0}^{M} \boldsymbol{H}(n)\boldsymbol{s}(k-n) + \boldsymbol{e}(k), \tag{2.1}$$

where

$$\boldsymbol{H}(k) = \begin{bmatrix} h_{11}(k) & \dots & h_{1N_t}(k) \\ \vdots & \ddots & \vdots \\ h_{N_r1}(k) & \cdots & h_{N_rN_t}(k) \end{bmatrix},$$

$$\boldsymbol{s}(k) = [s_1(k), \cdots, s_{N_t}(k)]^T,$$

$$\boldsymbol{y}(k) = [y_1(k), \cdots, y_{N_r}(k)]^T,$$

$$\boldsymbol{e}(k) = [e_1(k), \cdots, e_{N_r}(k)]^T.$$

Here

- $\boldsymbol{H}(n) \in \mathbb{C}^{N_r \times N_t}$ is the unknown matrix represents the channel impulse response which captures the introduced ISI. $h_{i,j}(k)$ is the channel path between the i-th transmitter and the j-th receiver at the time instant k. Without loss of generality, M is assumed to be the degree of the longest channel impulse response. Also, by defining $\boldsymbol{H}(z) \stackrel{def}{=} \sum_{n=0}^{M} \boldsymbol{H}(n) z^{-n}$ as the unknown FIR transfer function, we assume $\boldsymbol{H}(z)$ is irreducible and column reduced [80, 81].

- $\boldsymbol{s}(k) \in \mathbb{C}^{N_t \times 1}$ is the vector of unknown i.i.d and uncorrelated phase modulated (PSK or QAM) communication signals.

- $\boldsymbol{y}(k) \in \mathbb{C}^{N_r \times 1}$ is the received data vector.

- $\boldsymbol{e}(k) \in \mathbb{C}^{N_r \times 1}$ is the additive (spatially and temporally) white noise vector.

BD seeks to find the source signal matrix $\boldsymbol{s}(k)$ from the observation signal

$\boldsymbol{y}(k)$ with knowing neither the channel nor part of the source signal (pilots). Instead, some statistical and/or structural side information is utilized. Note that the convolutive model (4.1) becomes an instantaneous one at zero channel memory, i.e., $M = 0$, hence the effect of ISI diminish. This case turns the BD problem into BSS one. Next section briefs the BD problem and its solution approach.

2.3 Blind Deconvolution

This section elucidates the followed methodology to solve the BD problem.

2.3.1 Spatio-Temporal Framework

The target of BD algorithms is to revoke the convolutive channel effect. In other words, retrieve the transmitted sources $\boldsymbol{s}(k)$ using only the received data $\boldsymbol{y}(k)$ without resorting to the help of the pilot symbols. As pointed earlier, batch framework will be considered. To this end, consider the following spatio-temporal model

$$\mathbf{y}_w(k) = \boldsymbol{\mathcal{H}}\mathbf{s}_w(k) + \mathbf{e}_w(k), \tag{2.2}$$

where

$$\begin{aligned}
\mathbf{y}_w(k) &= \left[\mathbf{y}^T(k), \cdots, \mathbf{y}^T(k - N_w + 1)\right]^T, \quad (N_r N_w \times 1) \\
\mathbf{s}_w(k) &= \left[\boldsymbol{s}^T(k), \cdots, \ s^T(k - N_w - M + 1)\right]^T, \quad (N_d \times 1) \\
\mathbf{e}_w(k) &= \left[\boldsymbol{e}^T(k), \cdots, \boldsymbol{e}^T(k - N_w + 1)\right]^T, \quad (N_r N_w \times 1)
\end{aligned}$$

and $N_d = N_t(N_w + M)$ is the overall expected number of sources (including the delayed copies). It is noteworthy to mention that the temporal effect is due to introducing the window size N_w, whereas, the spatial one is because of the multiple receivers existence, i.e., N_r. $\boldsymbol{\mathcal{H}}$ is an $(N_r N_w \times N_d)$ block Toipletz convolution matrix defined as

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \boldsymbol{H}(0) & \boldsymbol{H}(1) & \cdots & \boldsymbol{H}(M) & 0 & \cdots & 0 \\ 0 & \boldsymbol{H}(0) & \boldsymbol{H}(1) & \cdots & \boldsymbol{H}(M) & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \boldsymbol{H}(0) & \boldsymbol{H}(1) & \cdots & \boldsymbol{H}(M) \end{bmatrix}, \quad (N_r N_w \times N_d)$$

As for the convolution matrix $\boldsymbol{\mathcal{H}}$ to be identified, the following condition must hold

$$N_w \geq \frac{N_t}{N_r - N_t} M \tag{2.3}$$

so as, the convolution matrix $\boldsymbol{\mathcal{H}}$ is of full column rank. The previous condition, which is justified in section 2.4.2, lead to constraints on the considered window length or the number of received antennas.

BD is achieved using deconvolution filters (DF) which are a collection of filtering vectors, $\boldsymbol{w}_j \in \mathbb{C}^{N_r N_w \times 1}, j = 1, \cdots, N_d$. To recover all the transmitted sources including the delayed ones, N_d deconvolution filters are needed to be estimated. Each DF applied on the received vector $\boldsymbol{y}_w(k)$ produces an estimate of the source signal $s_j(k)$, such that $\boldsymbol{w}_j^H \boldsymbol{y}_w(k) = z_j(k) = \hat{s}_j(k)$. These DF's are collected in an

$(N_d \times N_w N_r)$ deconvolution matrix $\boldsymbol{W} = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{N_d}]^H$. Applying the deconvolution matrix to the received signal $\boldsymbol{y}_w(k)$ produce the desired output as

$$\boldsymbol{z}(k) = \boldsymbol{W}\boldsymbol{y}_w(k) = \boldsymbol{W}\boldsymbol{\mathcal{H}}\boldsymbol{s}_w(k) + \boldsymbol{W}\boldsymbol{e}_w(k) = \boldsymbol{G}\boldsymbol{s}_w(k) + \underline{\boldsymbol{e}}_w(k) \tag{2.4}$$

where $\boldsymbol{z}(k) = \left[s_1(k), \cdots, s_{N_d(k)}\right]^T$ is the $(N_d \times 1)$ estimated source signal, $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{\mathcal{H}}$ is the $(N_d \times N_d)$ global deconvolution matrix and $\underline{\boldsymbol{e}}_w(k) \in \mathbb{C}^{N_d \times 1}$ is the filtered noise vector at the receiver output. Literally, to perform the BD, one need to estimate the deconvolution matrix $\boldsymbol{W}$.

As discussed earlier, we will be using the batch approach to perform BD. Hence, Next subsection layout the mathematical system model appropriately so that the Batch processing is enabled.

2.3.2 Batch Setup

Our implementation focus on batch (block) processing. In contrary to the common belief, batch implementations are not necessary more costly than the adaptive ones, also, they are able to exploits the information contained in the received data block more effectively [82, 36].

In batch BD algorithms, N_s samples of the received data are usually collected before processing, so that K vectors $\mathbf{y}_w(k),\ k = 1, \cdot, K$ can be formed, where $K = N_s - N_w + 1$, and concatenated in a matrix $\mathbf{Y}$ of a size $N_r N_w \times K$. In doing so, the transmitted and received source signals are related in a similar way

to relations (4.2) as

$$\mathbf{Y} = \boldsymbol{\mathcal{H}}\mathbf{S} + \mathbf{E} \tag{2.5}$$

and similar to (2.4), the estimated sources can be written in matrix form as

$$\mathbf{Z} = \mathbf{W}\mathbf{Y} \tag{2.6}$$

where

$$\begin{aligned}
\boldsymbol{Y} &= [\boldsymbol{y}_w(N_w), \cdots, \boldsymbol{y}_w(N_s)], \quad (N_r N_w \times K) \\
\boldsymbol{S} &= [\boldsymbol{s}_w(N_w), \cdots, \boldsymbol{s}_w(N_s)], \quad (N_d \times K) \\
\boldsymbol{E} &= [\boldsymbol{e}_w(N_w), \cdots, \boldsymbol{e}_w(N_s)], \quad (N_r N_w \times K) \\
\boldsymbol{Z} &= [\boldsymbol{z}(N_w), \cdots, \boldsymbol{z}(N_s)], \quad (N_d \times K)
\end{aligned}$$

2.4 Solution to the BD Problem

As outlined earlier, the main objective of BD problem is to find the deconvolution matrix $\boldsymbol{W}$. Generally, before performing the BD, the received data is subjected to a pre-processing whitening step. The whitining step is mainly initiated to shrink the searching space of the matrix $\boldsymbol{W}$ among the unitary matrices.

This section mainly explains the reasons behind deploying rank reduction and whitening operation as a pre-processing step. Then, give a nice review of the used pre-whitening method. Finally, explains the BD Procedure. Next, we will be outlining some principles associated with the blind MIMO system identification,

i.e., pre-whitening, rank-reduction, and ambiguities.

2.4.1 Pre-whitening and Rank reduction Operation

This section basically explains the pre-whitening and rank reduction procedure and the reason behind introducing them.

Rank reduction

As shown earlier, the BD problem has a spatio-tempral structure model. Based on the model, we considered that $N_r > N_t$, this means that the number of the virtual (spatio-temporal) receivers ($N_r N_w$) are greater than the number of the virtual transmitters ($N_d = N_r(N_w + M)$), i.e., $N_r N_w > N_d$, where the matrix $\boldsymbol{\mathcal{H}}$ becomes tall. Henceforth, from matrix algebra, this might lead to undesirable null-space solutions, i.e. $\boldsymbol{W}^H \boldsymbol{Y} = 0$. This problem is treated by using Rank reduction operation, which reduces the number of rows of $\boldsymbol{Y}$ form ($N_r N_w$) to (N_d), as shown hereafter.

Using underscore (-) to refer to the pre-filtered variables, and applying the (N_d) square pre-filter matrix $\boldsymbol{B}$ to (2.5) leads to

$$\underline{\mathbf{Y}} = \underline{\boldsymbol{\mathcal{H}}}^H \boldsymbol{S} + \underline{\boldsymbol{E}}, \quad \text{where} \quad \underline{\boldsymbol{\mathcal{H}}}^{\mathrm{H}} = \boldsymbol{B}\boldsymbol{\mathcal{H}},\ \underline{\boldsymbol{E}} = \boldsymbol{B}\boldsymbol{E} \tag{2.7}$$

The BD problem is now replaced by finding the square deconvolution matrix $\underline{\boldsymbol{\mathcal{H}}}$: $(N_d \times N_d)$. After Finding $\underline{\boldsymbol{\mathcal{H}}}$, the deconvolution matrix on the original (non-filtered) data will be $\boldsymbol{W} = \underline{\boldsymbol{\mathcal{H}}}\boldsymbol{B}$. In general, many methods are found for

dimension reduction, however, two conditions are opposed on finding $\boldsymbol{B}$: $\underline{\boldsymbol{\mathcal{H}}}^H = \boldsymbol{B}\boldsymbol{\mathcal{H}}$ should be of full rank, also to avoid noise enhancement, it should be well conditioned. The ratio between the largest and smallest singular values of the matrix defines its conditioning. The well-conditioned matrix has a ratio close to 1. The two mentioned conditions lead to choose the matrix $\boldsymbol{B}$ such that its columns are orthogonal and at the same time span the signal space (column span of $\boldsymbol{\mathcal{H}}$). This justifies the use the whiting technique which can accomplish this requirements, as described in the subsequent part.

Whitening

The whitening process described here is based on the work of Van Der Veen published in 2006. The whitening matrix $\boldsymbol{B} \in \mathbb{C}^{N_d \times N_d}$, which makes the matrix $\underline{\mathbf{Y}}$ white, can be obtained as follows:

1. Estimate the received data noisy covariance matrix: $\hat{\boldsymbol{R}}_{\boldsymbol{Y}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{y}_w(k)\boldsymbol{y}_w^H(k) = \frac{1}{K}\boldsymbol{Y}\boldsymbol{Y}^H$.

2. Find the corresponding eigenvalue decomposition:

$$\hat{\boldsymbol{R}}_{\boldsymbol{Y}} = \hat{\boldsymbol{U}}\hat{\boldsymbol{\Sigma}}^2\hat{\boldsymbol{U}}^H = \begin{bmatrix} \hat{\boldsymbol{U}}_s & \hat{\boldsymbol{U}}_e \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_s^2 & \\ & \hat{\boldsymbol{\Sigma}}_e^2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}_s^H \\ \hat{\boldsymbol{U}}_e^H \end{bmatrix} \tag{2.8}$$

   where $\hat{\boldsymbol{U}}$ is an $N_r N_w$ square unitary matrix and $\hat{\boldsymbol{\Sigma}}^2$ is a diagonal matrix contains the singular values of $\boldsymbol{Y}/\sqrt{K}$. The largest N_d eigenvalues are sorted in the diagonal matrix $\hat{\boldsymbol{\Sigma}}_s^2$ and their corresponding eigenvectors, which spans

the signal space, in $\hat{\boldsymbol{U}}_s$.

3. The whitening matrix is defined as

$$\boldsymbol{B} = \hat{\boldsymbol{\Sigma}}_s^{-1} \hat{\boldsymbol{U}}_s^H \tag{2.9}$$

This whitening procedure resulted into a reduced dimension data matrix: $\underline{\mathbf{Y}} \in \mathbb{C}^{N_d \times K}$, and a unitary deconvolution matrix $\underline{\boldsymbol{\mathcal{H}}}$ which whiten the received filtered data, i.e., $\hat{\boldsymbol{R}}_{\underline{\boldsymbol{Y}}} = \boldsymbol{I}_{N_d}$. To see this, let $\boldsymbol{\mathcal{H}} = \boldsymbol{U}_{\boldsymbol{\mathcal{H}}} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}} \boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^H$ be an economy-size SVD of $\boldsymbol{\mathcal{H}}$ ($\boldsymbol{U}_{\boldsymbol{\mathcal{H}}} \in \mathbb{C}^{N_r N_w \times N_d}$ is a unitary sub-matrix, $\boldsymbol{V}_{\boldsymbol{\mathcal{H}}} \in \mathbb{C}^{N_d \times N_d}$ and is unitary, and $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}} \in \mathbb{C}^{N_d \times N_d}$ contains the singular values). Assuming a large sample size, i.e., $\hat{\boldsymbol{R}}_{\boldsymbol{Y}} \approx \boldsymbol{R}_{\boldsymbol{Y}}$, then

$$\boldsymbol{R}_{\boldsymbol{Y}} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^H + \sigma_e^2 \boldsymbol{I} = \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{\mathcal{H}}} & \boldsymbol{U}_{\boldsymbol{\mathcal{H}}}^{\perp} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}}^2 + \sigma_e^2 \boldsymbol{I} & \\ & \sigma_e^2 \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{\boldsymbol{\mathcal{H}}}^H \\ \left(\boldsymbol{U}_{\boldsymbol{\mathcal{H}}}^{\perp}\right)^H \end{bmatrix} \tag{2.10}$$

By comparing (2.8) with (2.10), we find that $\boldsymbol{U}_s = \boldsymbol{U}_{\boldsymbol{\mathcal{H}}}$ and $\boldsymbol{\Sigma}_s^2 = \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}}^2 + \sigma_e^2 \boldsymbol{I}$. Hence,

$$\boldsymbol{B}\boldsymbol{\mathcal{H}} = \hat{\boldsymbol{\Sigma}}_s^{-1} \hat{\boldsymbol{U}}_s^H \boldsymbol{U}_{\boldsymbol{\mathcal{H}}} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}} \boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^H = (\hat{\boldsymbol{\Sigma}}_s^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}}) \boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^H = \boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^H \tag{2.11}$$

is true if $(\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}}^2 + \sigma_e^2 \boldsymbol{I})^{-1/2} \boldsymbol{\Sigma}_{\boldsymbol{\mathcal{H}}} = c\boldsymbol{I}$, where c is a scalar constant. This is possible at noise free case or when the columns of $\boldsymbol{\mathcal{H}}$ are orthonormal, i.e., all singular values of $\boldsymbol{\mathcal{H}}$ are equal). If this is not the case (more practical scenario), then $\boldsymbol{\mathcal{H}}$ is transformed to a matrix $\boldsymbol{V}_{\boldsymbol{\mathcal{H}}}^H$ close to unitary which is always better condi-

tioned (close to well-conditioned) than $\boldsymbol{\mathcal{H}}$ itself. Considering the above explained covariance-based whitening procedures, we find the whitening matrix $\boldsymbol{B}$. For simplicity, assume noise-free case, then the whitened received signal can be written as

$$\underline{\mathbf{Y}} = \boldsymbol{BY} = \boldsymbol{B\mathcal{H}S} = \boldsymbol{V}^H\boldsymbol{S} \tag{2.12}$$

Clearly, whitening operation reduces the problem of BD into finding the unitary matrix $\boldsymbol{V}$, we will be refering to it as filtered doconvolution matrix throughout this part. Finding the unitary matrix $\boldsymbol{V}$ is a simpler constraint compared to finding the matrix $\boldsymbol{W}$ of linearly independent rows, i.e., independent deconvolution rows $\boldsymbol{w}_j$. This is because we only need to assure the orthogonality between the rows of $\boldsymbol{V}$. Also, it is well established in adaptive filtering theory that moving into whitened domain boosts the convergence speed.

The subsequent subsection will discuss the procedures to find the filtered deconvolution matrix $\boldsymbol{V}$.

2.4.2 BD Procedure

Now, once the matrix $\boldsymbol{V}$ is obtained, the BD matrix is given as $\boldsymbol{W} = \boldsymbol{VB}$, resulting into the output

$$\boldsymbol{Z} = \boldsymbol{WY} = \boldsymbol{VBY} = \boldsymbol{V}\underline{\mathbf{Y}} = \boldsymbol{VV}^H\boldsymbol{S} = \hat{\boldsymbol{S}} \tag{2.13}$$

The derivation of the new BD methods is based on the existence of a deconvolution matrix $\boldsymbol{W}$, which is a composite of a whitening matrix $\boldsymbol{B}$ and unitary matrix $\boldsymbol{V}$, such that $\boldsymbol{WY} = \hat{\boldsymbol{S}}$. This means that the row span of $\hat{\boldsymbol{S}}$ is contained in (or equivalent to the row span of $\boldsymbol{Y}$. This is true if the channel matrix $\boldsymbol{\mathcal{H}}$ has a full column rank matrix, which justifies the suggested condition on the processing window length N_w given in(2.3). Based on the cost or contrast function, many methods are developed to estimate the channel deconvolution matrix [83, 8, 84, 26, 85, 25]. However, most of them are tackling the instantaneous mixture scenario, and non of them has considered the MM nature as a minimizing criterion in the light of convolutive mixture.

2.5 Indeterminacies in BD

After discussing one source of indeterminacy in BD problem in subsection 2.4.1, we list here another two inherent sources of indeterminacy that ever encountered in multichannel BSI and BSS theory. Knowing nothing about the source signal except some of the statistical and non-statistical properties, enables to recover the sources up to two scenarios of non-uniqueness [23]:

1. Source ordering is arbitrary: mislabeling the source ordering means that the columns of the beamformer $\boldsymbol{W}$ are mislabeled and this is referred to as the permutation ambiguity.

2. The output of each beamformer can be resolved up to an arbitrary scale factor (phase ambiguity).

That is to say

$$\boldsymbol{W}\boldsymbol{\mathcal{H}} = \boldsymbol{P}\boldsymbol{\Lambda}, \tag{2.14}$$

where $\boldsymbol{P} \in \mathbb{C}^{N_d \times N_d}$ is the permutation matrix and $\boldsymbol{\Lambda} \in \mathbb{C}^{N_d \times N_d}$ is the scaling non-singular diagonal matrix. As it is revealed by (2.14), the ambiguities are always presented in BD problem. However, such problem might be solved using some kind of Differential coding or/and injecting some short pilot data [86].

2.6 Review of Givens and Shear Rotations

The next subsections brief the unitary and non-unitary rotations which will be used as an optimization tools in the subsequent chapter.

2.6.1 The Givens Rotations

The Givens rotation is a unitary transformation. Givens matrix $\boldsymbol{\varphi}_{pq}(\theta, \alpha)$ is an identity matrix with ones in the main diagonal except for the two elements φ_{pp} and φ_{qq}. Also, all the off-diagonal elements are filled with zeros except for the two elements φ_{pq} and φ_{qp}. The exempted four elements are defined as follows:

$$\boldsymbol{\varphi}_{p,q} = \begin{bmatrix} \varphi_{pp} & \varphi_{pq} \\ \varphi_{qp} & \varphi_{qq} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & e^{i\alpha}\sin(\theta) \\ -e^{-i\alpha}\sin(\theta) & \cos(\theta) \end{bmatrix} \tag{2.15}$$

where $\{\theta, \alpha \in [-\pi/2, -\pi/2]\}$ are the Givens' angle parameters. Note that, in the real case scenario, $\alpha = 0$, and subsequently, Givens rotations has the following distinctive properties:

1. Givens rotations are unitary, i.e.,

$$\boldsymbol{\varphi}\boldsymbol{\varphi}^H = \boldsymbol{\varphi}^H\boldsymbol{\varphi} = \boldsymbol{I}.$$

2. Givens rotations preserves the vector norms. To see this, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two column vectors that are related by Givens transformation $\boldsymbol{\varphi}$, say $\boldsymbol{\varphi}\boldsymbol{x} = \boldsymbol{y}$, then

$$\|\boldsymbol{y}\|^2 \triangleq \boldsymbol{y}^H\boldsymbol{y} = \boldsymbol{x}^H\boldsymbol{\varphi}^H\boldsymbol{\varphi}\boldsymbol{x} = \boldsymbol{x}^H\boldsymbol{x} = \|\boldsymbol{x}\|^2$$

   Thus, the related vectors have a preserved Euclidean norms.

3. Givens rotations preserves inner products between vectors, i.e., if $\mathbf{b}$ and $\mathbf{c}$ are two other vectors related by the same transformation, say $\boldsymbol{\varphi}\mathbf{b} = \mathbf{c}$, then we get

$$\boldsymbol{y}.\mathbf{c} \triangleq \boldsymbol{y}^H\mathbf{c} = \boldsymbol{x}^H\boldsymbol{\varphi}^H\boldsymbol{\varphi}\mathbf{b} = \boldsymbol{x}^H\mathbf{b} = \boldsymbol{x}.\mathbf{b}$$

   This property is called angle preservation property as well as it preserves the angle between transformed vectors.

As a summary, say that vectors $(\boldsymbol{x}, \mathbf{b})$ are transformed into $(\boldsymbol{y}, \mathbf{c})$, this means that

their norms as well as the angles between them are all preserved.

2.6.2 The Hyperbolic Rotations

The Hyperbolic(Shear) rotations matrix $\boldsymbol{\phi}_{pq}(\gamma, \chi)$ is a non-unitary (or $\boldsymbol{J}$-unitary) rotations defined in a similar way to the Givens rotations. Whereby, the exempted four elements in the main and off-diagonal are defined as follows:

$$\boldsymbol{\phi}_{p,q} = \begin{bmatrix} \phi_{pp} & \phi_{pq} \\ \phi_{qp} & \phi_{qq} \end{bmatrix} = \begin{bmatrix} \cosh(\gamma) & e^{i\chi}\sinh(\gamma) \\ e^{-i\chi}\sinh(\gamma) & \cosh(\gamma) \end{bmatrix} \tag{2.16}$$

where $\chi \in [-\pi/2, \pi/2]$ and $\{\gamma \in [-\bar{\gamma}, \bar{\gamma}], \bar{\gamma} > 0\}$ are Hyperpolic rotations parameters. Also, for the real case scenario, $\chi = 0$. The Shear matrix is $\boldsymbol{J}$-unitary, i.e.,

$$\boldsymbol{\phi}\boldsymbol{\phi}^H = \boldsymbol{\phi}^H \boldsymbol{J} \boldsymbol{\phi} = \boldsymbol{J}.$$

for some signature diagonal matrices $\boldsymbol{J}$, i.e. matrix with diagonal entries equal ± 1.

For further information about Givens, Shear and some other useful unitary and non-unitary rotations, one can refer to [87, 88].

2.7 Conclusion

This chapter presents different terminologies which are regularly used throughout the dissertation, such as the pre-whitening filtering, optimization techniques (Givens and Shear rotations), and some other BD indeterminacy. Moreover, the convolutive MIMO system's model is described and the unified spatio-temporal framework is formulated. Eventually, this chapter declares that the batch processing approach will be the framework of this dissertation.

CHAPTER 3

DECONVOLUTION METHODS

This chapter targets the blind deconvolution problem for multiple-input multiple-output communication system, using small and moderate constellation's size phase modulated signals, i.e. PSK and QAM. We introduce four different blind deconvolution algorithms based on four different techniques. These algorithms come as a natural extension of the successful work done by Shah et al in 2015. The first two methods are considered as two-step based methods, where the first one is accomplished by performing a cascaded linear equalization, using one of the existing subspace-methods, followed by the BSS routine. While the second performs the BSS for the spatio-temporal system and then a pairing and sorting phase. The third method is based on the minimization of a hybrid cost function. The last one is a deflation-based method. These solutions summarize the main possible paths that can be followed to extend any of the existing instantaneous de-mixing algorithms. Experimental results are provided to show the unique characteristics of each of the four different methods.

3.1 Introduction

In the context of wireless communications, blind methods have been considered for many purposes including waiving the training sequence [3], reducing the pilot size using semi-blind approaches [4] or mitigating the pilot contamination problem in massive MIMO systems [5]. In particular, it has been shown in [4, 5] that, thanks to the semi-blind solution, the pilot sequence size can be reduced by a factor of 80 to 90% without affecting the channel estimation quality[1]. To achieve these objectives, different cost functions exist for the blind deconvolution of phase modulated signals such as PSK and QAM signals [3, 20]. Among them are the constant modulus and Multi-modulus criteria. The latter has shown to be more appealing for high order QAM signals [85], as well as the ability of the associated algorithms, namely multi-modulus algorithms (MMA), to perform joint blind equalization and carrier recovery without the need for a separate carrier-recovery system [14, 89].

Wireless communications suffer from the multipath fading effect which results into the intersymbol interference (ISI). Moreover, MIMO systems experience an additional interuser interference (IUI) problem. Hence, the blind deconvolution for MIMO systems is a twofold problem: a blind equalization to get rid of the ISI and a blind source separation for the IUI removal.

Recently, efficient BSS methods used to seperate QAM signals based on the optimization of the Multi-Modulus Critera using the elementary Givens and Shear[2]

[1]Note that, in certain systems, this represents close to 25% of energy reduction which is a main objective of the current research activities for 'Green Communications'.

[2]The terms Hyperbolic and Shear will be used interchangebly throughout this chapter.

rotations have been introduced in [6]. These methods are designed for the special case when the signal undergoes a flat fading, where the channels have an instantaneous mixing effect. However, in most practical cases, the channels are convolutive and the previous methods [6] do not apply directly.

Our first motivation is to extend the previous approach to the more general convolutive mixture case. This extension can be done in several different ways that we have developed and compared in this chapter. More precisely, we consider four different approaches. The first one is refered to as Full-BSS deconvolution, whereby we form a space-time (spatio-temporal) array from the collected data followed by a source de-mixing using BSS techniques, ended by paring and selection using a correlation-based method. The second method is a cascade of linear Second Order Statistics (SOS)-based equalization followed by a BSS demixing approach. The equalization is needed to remove the channel ISI effect, so that the problem is reduced from being of convolutive nature into an instantaneous mixture of signals, whereas the BSS is needed to get rid of IUI effect. Such a cascaded approach is referred to as a two-stage deconvolution approach, and it has been reported in several earlier works, such as [90], [40] and [91].

In the third method, we perform blind deconvolution (BD) through considering one composite cost function which penalizes the Multi-Modulus (MM) criterion and the cross-correlation between the different source signals. The optimization of the cost function is done using elementary Givens and Shear rotations resulting in two different algorithms, namely, Givens and Hyperbolic Givens MM Deconvo-

lution algorithms G-MMDA and HG-MMDA, respectively. The last approach is a deflation-based method, such that each time one single source is extracted, then it is waived from the convolutive mixture using an appropriate subspace projection. This is repeated until all sources are extracted.

This chapter has mainly two contributions:

- Propose, for the first time, deconvolution algorithms based on the alphabet nature of the constellation signals (MM criterion) using the elementary Givens and Shear rotations.

- Review and compare the state-of-the-art deconvolution approaches (i.e. the four previously mentioned solutions) that can be considered to perform blind deconvolution.

The chapter is organized as follows: Section 3.3 presents the data model, briefs the BD principle, and defines the used cost function as well as Givens and Shear rotations. Section 3.3 outlines the four different approaches and the corresponding derived algorithms. Section **??** compares the four different approaches in terms of computational cost and source signal restoration quality. Finally, Section 3.6 concludes the chapter.

3.2 Cost Function

Communication signals experience Multi-Modulus (MM) nature (e.g., QAM), and hence one proposes to estimate $\boldsymbol{V}$ in (2.12) by optimizing the MM criterion in-

troduced by [26] as

$$J_{MM}(\mathbf{V}) = \sum_{i=1}^{N_d} E\left[\left(z_{i,R}^2(k) - R_R\right)^2 + \left(z_{i,I}^2(k) - R_I\right)^2\right] \tag{3.1}$$

where $R_R = E[s_R^4(k)]/E[s_R^2(k)]$ and $R_I = E[s_I^4(k)]/E[s_I^2(k)]$ are, respectively, the real and imaginary dispersion constants[3]. $z_i(k)$ is the (i,k)-th element of $\mathbf{Z}$. The MM function enjoys several advantages over the Constant Modulus (CM) one [84], whereby, it drives to: a) faster convergence algorithms [92], b) carrier phase recovery [89], c) less undesirable minima [93], and d) ease of hardware implementations [94].

The aim of this chapter is to find a sub-unitary (close to unitary matrix) $\mathbf{V}$ as a product of elementary Shear and/or Givens rotations (see Section 2.6). The use of this approach is motivitated by Jacobi-like algoritms [83, 95]. These algorithms are used for diagonalization of symmetric matrices. The idea is to sweep a sequence of unitary rotations such that $\boldsymbol{Y} \longleftarrow \boldsymbol{\varphi}^H \boldsymbol{Y} \boldsymbol{\varphi}$, whereby the updated matrix $\boldsymbol{Y}$ is more diagonal than the earlier one. In this chapter, we adopt the same idea to find the matrix $\boldsymbol{V}$ and at the same time $\boldsymbol{Z}$ in (2.13) using a sequence of unitary/non-unitary transformations where the optimization parameters, i.e., the angels, are evaluated according to the MM criterion (3.1).

The fast convergence, ease of implementation and storage [95] are the motivations behind using these elementary rotations. Also, these matrices have a

[3]For ease of simplicity, we assume that the different dispersion constants are equal, i.e., $R_R = R_I$. Also we assumed implicitly that all emmiting sources belong to the same QAM constellation.

determinant equal to 1 and hence they guarantee the non-singularity of the build up beamforming matrix. It is customarily to mention that, by using the whitening matrix $\boldsymbol{B}$ in (2.9) with a large sample size N_s, we guarantee a good whitening step. Therefore, the claim for searching the matrix $\boldsymbol{V}$ as a unitary one becomes valid, thus using only Givens rotations is enough. Yet, for moderate and small sample sizes, we use Shear and Givens rotations together so that we enlarge the searching space by dropping the unitary condition, that said, it leads to an improved estimation accuracy.

In the subsequent section four different methods are deployed to blindly deconvolve and separate QAM source signals. The rational for introducing these methods is that, each one of them has different unique pros and cons that allow the designer to have more selection freedom as well as to conduct a comparative study of their performance.

3.3 Two-Stage Deconvolution Algorithms

By imitating the Jacobi-like algorithms [83, 95], one can write the unitary matrix $\boldsymbol{V}$ as a product of elementary Givens rotations, according to:

$$\boldsymbol{V} = \prod_{N_{Sweeps}} \prod_{1 \leq p,q \leq N_d} \boldsymbol{\varphi}_{p,q}(\theta, \alpha) \tag{3.2}$$

where N_{Sweeps} stands for the number of needed sweeps (iterations) until convergence. The Givens rotations' angles (θ and α) are obtained by optimization of

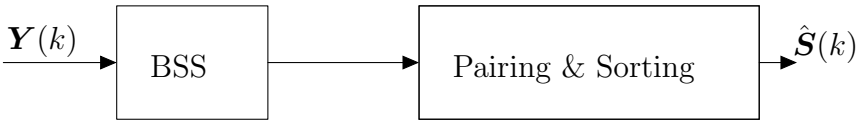


Figure 3.1: Two-Stage Full-BSS Deconvolution Approach.

the MM cost function given in (3.1).

Unfortunately, it can be shown that closed form or simple solutions for this optimization problem are not possible due to the complex non-linear expression of the cost function[4] w.r.t. parameters α and θ. To overcome this issue, the authors in [6] used an appropriate complex to real data transformation before resorting to the real valued Givens rotations (with $\alpha = 0$). In this work, we opt for using special versions of complex rotations by fixing the values of parameters α (for the Givens) and χ (for the Shear) to specific values, as detailed in the subsequent sections, which leads to an easy-go derivation.

Next subsections are dedicated to the development of the different BD approaches mentioned in section 3.1.

3.3.1 Full-BSS Deconvolution

To find matrix $\mathbf{V}$ in (2.12), many algorithms exist in the litrature which are designed for instantaneous mixtures [6], [8], [83]. Our first approach uses the ones in [6] to the built spatio-temporal system model ($\boldsymbol{Y} = \boldsymbol{\mathcal{H}}\boldsymbol{S} + \boldsymbol{E}$) given in (2.5). The resulted signals are separated up to two kinds of permutation; the first permutation is within the symbol scale (within the window size), while the

[4]All what have been discussed so far applies to both Given and Shear rotations.

(a) Signals out of BSS (b) Signals out of pairing. (c) Signals out of sorting.

Figure 3.2: Paring and sorting mechanism.

second is a window scale-type permutation. Symbol scale permutation is the consequence of the full-BSS deconvolution method, where the window scale one is the inherent BSS permutation. To get rid of the former kind of permutation, a second step is carried through the use of *pairing* and *sorting* mechanism. In this step, pairing is used to collect each source and its replicas in one window, then sorting is followed to arrange the signal within the window according to the delay index. This procedure is illustrated in Figure 3.2.

Pairing and sorting step is built around the correlation between the signals and their delayed copies, it can be summarized as follows:

1. *Pairing*: at first, randomly pick any source signal from the N_d set, then the time-lagged correlation matrix, $\boldsymbol{R}_i$ between the picked source $\boldsymbol{z}_i$ and the rest of the separated sources denoted as $\bar{\boldsymbol{Z}}_i$ (where i refers to the order of the extracted source), is defined as

$$\left[\begin{array}{c|c|c} E\left[\bar{\boldsymbol{z}}_i(k) z_i^*(k+N_\delta)\right] & \cdots & E\left[\bar{\boldsymbol{z}}_i(k) z_i^*(k-N_\delta)\right) \end{array}\right] \tag{3.3}$$

where $\bar{\boldsymbol{z}}_i(k)$ is the k-th column vector of matrix $\bar{\boldsymbol{Z}}_i$ and $N_\delta = M + N_w$ represents the number of delayed replicas for each source. Note that $\boldsymbol{z}_i$ has non-zero correlation with its $N_\delta - 1$ delayed copies and it is uncorrelated with the other source signals. Hence, ideally $\boldsymbol{R}_i$ should have only $N_\delta - 1$ non-zero rows. Therefore, the pairing is achieved as follows: one computes the norms of the row vectors of $\boldsymbol{R}_i$, sort them in a descending order and select the first $N_\delta - 1$ that correspond to $\boldsymbol{z}_i$.

2. *Sorting*: since the source signals are i.i.d, the non-zero rows of $\boldsymbol{R}_i$ have ideally only one non-zero entry corresponding to the time shift in between the two correlated sources. Then, the N_δ paired signals are arranged by sorting the correlation delays in ascending order. The next example gives a clear idea about this step.

3. Extract the paired and sorted set of signals from $\boldsymbol{Z}_i$, go back to step 1 to extract the next set as long as $i < N_t$.

Example: After completing the *pairing* step, consider sorting the following picked source and its paired ones:

$$\boldsymbol{Q}^T(k) = \begin{matrix} r_1 & r_2 & r_3 & r_4 \end{matrix} \\ \begin{bmatrix} \boldsymbol{s}^T(k-1) & \boldsymbol{s}^T(k) & \boldsymbol{s}^T(k-3) & \boldsymbol{s}^T(k-2) \end{bmatrix}$$

assume $\boldsymbol{z}_i$ is chosen to be $\boldsymbol{s}(k-2)$, i.e.,$\boldsymbol{z}_i = \boldsymbol{s}(k-2)$, then one evaluate (3.56)

for N_δ as follows:

$$[E\left[\boldsymbol{q}(k)s_i^*(k-2+3)\right], E\left[\boldsymbol{q}(k)s_i^*(k-2+2)\right], E\left[\boldsymbol{q}(k)s_i^*(k-2+1)\right], E\left[\boldsymbol{q}(k)s_i^*(k-2)\right]$$
$$, E\left[\boldsymbol{q}(k)s_i^*(k-2-1)\right], E\left[\boldsymbol{q}(k)s_i^*(k-2-2)\right], E\left[\boldsymbol{q}(k)s_i^*(k-2-3)\right]]$$

where $\boldsymbol{q}(k)$ is the k-th column vector of $\boldsymbol{Q}$. Ideally this resulted into the following:

$$\begin{array}{c} \\ r_1 \\ r_2 \\ r_3 \end{array} \begin{array}{c} \begin{array}{ccccccc} -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 0 & \sigma_s^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_s^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_s^2 & 0 & 0 \end{bmatrix} \end{array}$$

Now, by sorting the correlation delays in ascending order, i.e., $-2, -1, 0, 1$, and comparing locations of σ_s^2 with that of $\boldsymbol{Q}$, i.e., r_2, r_1, r_4, r_3, we conclude the following appropriate sorting:

$$\boldsymbol{Q}(k) = \begin{bmatrix} \boldsymbol{s}^T(k) & \boldsymbol{s}^T(k-1) & \boldsymbol{s}^T(k-2) & \boldsymbol{s}^T(k-3) \end{bmatrix}^T$$

■

The above detailed method results in separating multiple replica of the desired sources. This has the advantage to provide a controlled delay equalizer which is known to be an important issue for MIMO systems [96]. However, from the computational complexity point of view, this is costly. This method can work in the case of $N_r \leq N_t$ as opposed to our data model assumption, but with a tolerated performance degradation.

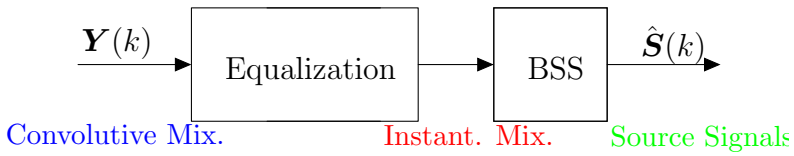


Figure 3.3: Two-Stage Equalization-BSS approach.

In the ensuing, methods with less complexity will be proposed.

3.3.2 Equalization-BSS approach

This is a two-step method, whereby the first stage is needed to remove the ISI introduced by the channel memory using one of the SOS-based blind equalization methods. Among the powerful methods to perform blind channel identification and equalization is the linear prediction method [90], the subspace (SS) method [81], the minimum noise subspace (MNS) method [97] and the structured-based subspace (SSS) method [98], to mention a few. These techniques convert the convolutive MIMO system into an instantaneous one[5]. Thus, the second stage is necessitated to perform the BSS for the equalized mixture. In other words, the decomposition of blind deconvolution problem assumes two successive stages (see Figure 3.3):

1. Equalizer construction by estimating $\boldsymbol{\mathcal{H}}$ (up to the inherent indeterminacies of MIMO blind identification). The output of this step is an instantaneous mixture of source signals. This is done based on the SOS.

2. BSS, i.e., retrieving the source signals by estimating the linear separator $\boldsymbol{W}$

[5]Note that the model (2.1) and the subsequent formulation reduce to the instantaneous model by setting $M = 0$ and $N_w = 1$.

in (2.5).

In the first stage, we used the MNS method to perform the blind channel identification due to its desirable low complexity and excellent performance [97]. Now, having $\boldsymbol{\mathcal{H}}$ estimated and the equalization performed, the instantaneous mixture will be separated using any appropriate BSS methods/algorithms. As mentioned earlier, many BSS algorithms are dedicated to seperate instananeous mixtures, and in this work, we intentionally use the recently proposed effective G-MMA and HG-MMA algorithms [6].

As it is clear from the above, in the first step we rely on the SOS method for the blind system identification. Since the SOS methods depend on the availability of channel diversity, therefore, this method sustains powerful as long as $N_r > N_t$. Yet, for the case of $N_r \leq N_t$, one can resort to higher order statistics based methods (see for example [99]).

So far, the two considered methods proceed in two stages. With this analogy, one can refer to the methods that are proposed in the ensuing section as one-stage methods.

3.4 One-Stage Deconvolution Algorithms

We introduce in this section two BD methods based on the use of a hybrid criterion optimization and on a deflation technique, respectively.

3.4.1 Deconvolution Using Hybrid Criterion

In order to separate different signals and avoid extracting multiple replica of same source signal while missing the others, one devise to perform the deconvolution in one stage using the following hybrid cost function[6]:

$$\mathcal{J}_{hyb}(\mathbf{V}) = \mathcal{J}_{MM}(\mathbf{V}) + \mathcal{J}_{\times}(\mathbf{V}) \tag{3.4}$$

where $\boldsymbol{V}$ is a $N_t \times N_d$ beamformer and

$$\begin{aligned} \mathcal{J}_{MM}(\mathbf{V}) &= \tfrac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N_t} \left(z_{i,R}^2(k) - R_R\right)^2 + \left(z_{i,I}^2(k) - R_I\right)^2 \\ \mathcal{J}_{\times}(\mathbf{V}) &= \sum_{1 \le i < j \le N_t} \sum_{\delta=-N_\delta}^{N_\delta} |r_{ij}(\delta)|^2 \end{aligned} \tag{3.5}$$

$r_{ij}(\delta)$ stands for the cross-correlation function between user i and user j, at time lag δ, defined as

$$r_{ij}(\delta) = E\left[z_i(k) z_j^*(k-\delta)\right] \tag{3.6}$$

The hybrid cost function (3.4) is of two sides: $\mathcal{J}_{MM}(\boldsymbol{V})$ penalizes the deviation of the real and imaginary parts of the equalized signals, while $\mathcal{J}_{\times}(\boldsymbol{V})$ penalizes the cross-correlation between the estimated signals to ensure unique extraction of each source signal. The parameter N_δ is an integer that should be chosen in consistent with the channel's delay spread, so that it counts for all expected replicas that might be generated for each single user due to the channels' convolutive nature.

[6]Similar cost function has been adopted in [40] with some minor differences.

Givens MM Deconvolution Algorithm (G-MMDA)

We propose here to build our beamformer $\boldsymbol{V}$ as the first N_t rows of the matrix obtained by using a sequence of the special version of the complex Givens rotations[7] that have been observed in [6, 8], whereas the two free parameters α and θ are replaced by two successive rotations with one free parameter for each rotation, i.e., θ and θ', respectively according to:

$$\begin{aligned}\mathbf{V} = \boldsymbol{J} \prod_{N_{Sweeps}} \prod_{1\leq p\leq q\leq N_t}^{a} \boldsymbol{\varphi}_{p,q}^{a}\left(\theta,0\right)\boldsymbol{\varphi}_{p,q}^{a}\left(\theta',-\tfrac{\pi}{2}\right)\mathbf{D}_p(\theta'') \\ \prod_{\substack{1\leq p\leq N_t \\ N_t<q\leq N_d}}^{b} \boldsymbol{\varphi}_{p,q}^{b}\left(\theta,0\right)\boldsymbol{\varphi}_{p,q}^{b}\left(\theta',-\tfrac{\pi}{2}\right)\end{aligned} \tag{3.7}$$

where $\boldsymbol{J} = [\boldsymbol{I}_{N_t}, \mathbf{0}]$ is a row selection matrix, $\boldsymbol{\varphi}_{p,p}^{a}$ is, by convention, the identity matrix and $\mathbf{D}_p(\theta'')$ is a diagonal matrix equal to the identity except for its p-th diagonal entry which is equal to $e^{-j\theta''}$. The novelty of our design resides in deconvolving only N_t sources out of the N_d mixed sources and their delayed versions[8]. In that sense, the first product set of rotations $\prod^{a}$ are designed to target the first N_t sources, while the second set $\prod^{b}$ limits index p within the first N_t sources and allows the q index to span the whole set of N_d mixed sources. The phase term $e^{-j\theta''}$ is called upon to compensate for the phase shift introduced by

[7]Indeed, it has been observed in [6] that working on real Givens (and Shear) rotation is equivalent to working on a special versions of Complex Givens (and Shear) rotation by fixing the value of angles α (and χ) appropriately.

[8]To the best of our knowledge, the elementary Givens and Shear products are usually covers the whole matrix space of dimension N_d. Yet our design exclusively aims to reduce the computational complexity by targeting the recovery of the transmitting sources without their delayed versions, i.e., the sweeping space is partially reduced to recover only N_t sources.

the diagonal entries of the convolution matrix $\boldsymbol{\mathcal{H}}$.

Shortly, we will be expressing $\mathcal{J}_{hyb}(\mathbf{V})$ given in (3.4) as a function of the desired angle parameters. Thus, we break down the problem of finding matrix $\boldsymbol{V}$ into estimating the angles solely, so that (3.4) is optimized. To this end, consider a single unitary rotation, given as

$$\mathbf{Z} = \boldsymbol{\varphi}_{p,q}^{a}\left(\theta, 0\right) \underline{\mathbf{Y}} \tag{3.8}$$

According to (2.15), only rows p and q change. Then, the first part of (3.4) can be re-written in terms of Givens angle (θ) (after omitting the terms that are independent of (θ) and for the sake of simplicity let $R_R = R_I = R$), as

$$\begin{aligned} \mathcal{J}_{MM}(\theta) = \tfrac{1}{K} \sum_{k=1}^{K} \Big[\big(z_{p,R}^2(k) - R\big)^2 + \big(z_{p,I}^2(k) - R\big)^2 \\ + \big(z_{q,R}^2(k) - R\big)^2 + \big(z_{q,I}^2(k) - R\big)^2 \Big] \end{aligned} \tag{3.9}$$

Hence, due to the transform in (3.8), $z_p(k)$ and $z_q(k)$ can be read as[9]

$$\begin{aligned} z_p^2(k) &= \boldsymbol{\Theta}^T \mathbf{t}_{pq}(k) + \tfrac{1}{2}\left(\underline{y}_p^2(k) + \underline{y}_q^2(k)\right) \\ z_q^2(k) &= -\boldsymbol{\Theta}^T \mathbf{t}_{pq}(k) + \tfrac{1}{2}\left(\underline{y}_p^2(k) + \underline{y}_q^2(k)\right) \end{aligned} \tag{3.10}$$

[9]The indexes R and I will be dropped intentionally, since both lead to a similar results.

where

$$\begin{aligned}\mathbf{\Theta} &= \begin{bmatrix} \cos(2\theta) & sin(2\theta) \end{bmatrix}^T, \\ \mathbf{t}_{pq}(k) &= \begin{bmatrix} \frac{1}{2}\left(\underline{y}_p^2(k) - \underline{y}_q^2(k)\right) & \underline{y}_p(k)\underline{y}_q(k) \end{bmatrix}^T\end{aligned} \tag{3.11}$$

This leads to expressing the two real terms of (3.9) into a quadratic term plus a constant term, given as

$$\begin{aligned}&\left(z_{p,R}^2(k) - R\right)^2 + \left(z_{q,R}^2(k) - R\right)^2 = \\ &\mathbf{\Theta}^T \left[2\sum_{k=1}^{K} \mathbf{t}_{pq,R}(k)\mathbf{t}_{pq,R}^T(k)\right]\mathbf{\Theta} + 2c_{pq,R}^2(k)\end{aligned} \tag{3.12}$$

where $c_{pq,R}(k) = (\underline{y}_{p,R}^2(k) + \underline{y}_{q,R}^2(k))/2 - R$. Likewise, the two imaginary terms of (3.9) are expressed by replacing the real parts of (3.12) with the imaginary ones as

$$\begin{aligned}&\left(z_{p,I}^2(k) - R\right)^2 + \left(z_{q,I}^2(k) - R\right)^2 = \\ &\mathbf{\Theta}^T \left[2\sum_{k=1}^{K} \mathbf{t}_{pq,I}(k)\mathbf{t}_{pq,I}^T(k)\right]\mathbf{\Theta} + 2c_{pq,I}^2(k)\end{aligned} \tag{3.13}$$

where $c_{pq,I}(k) = (\underline{y}_{p,I}^2(k) + \underline{y}_{q,I}^2(k))/2 - R$. By combining (3.12) and (3.13), and after dropping the constant terms that are irrelevant to the determination of the optimum value θ, we express the first part of the hybrid cost function (3.4) as a quadratic form

$$\mathcal{J}_{MM}(\theta) = \mathbf{\Theta}^T\mathbf{T}\mathbf{\Theta} \tag{3.14}$$

where

$$\mathbf{T} = 2\sum_{k=1}^{K} \mathbf{t}_{pq,R}(k)\mathbf{t}_{pq,R}^{T}(k) + \mathbf{t}_{pq,I}(k)\mathbf{t}_{pq,I}^{T}(k) \tag{3.15}$$

In an analogous way, we simplify $\mathcal{J}_{\times}$ and express it as a function of θ while we are evaluating the first set of rotations (case 'a'). To follow up with the introduced rotation in (3.8), only rows p and q are adjusted, that allows to re-write $\mathcal{J}_{\times}$ in (3.5) as

$$\mathcal{J}_{\times}(\mathbf{V}) = \sum_{\substack{j=1 \\ j\neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} |r_{pj}(\delta)|^2 + \sum_{\substack{i=1 \\ i\notin\{p,q\}}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} |r_{iq}(\delta)|^2 \tag{3.16}$$

Beginning with the first summation, two scenarios are encountered: the first one is when $j \neq q$, in this case only p is updated, according to (2.15) and using the double angle trigonometric identities, $|r_{pj}(\delta)|^2$ is expressed by a linear relation[10]

$$|r_{pj}(\delta)|^2 = \mathbf{\Theta}^T \mathbf{r}_{pq,j}(\delta) \tag{3.17}$$

where $\mathbf{r}_{pq,j}(\delta) = [\frac{1}{2}(|r_{pj}|^2 - |r_{qj}|^2), (r_{pj}r_{qj}^*)_R]^T$. The second scenario is when $j = q$, here p and q rows are modified, accordingly, $|r_{pq}(\delta)|^2$ leads to a quadratic and

[10]For the sake of simplifying the presentation, from now on all the irrelevant terms that are independent of estimation parameters (i.e., θ) will be dropped). Also, δ will be sometimes dropped to shorten the equations presentation.

linear term, given by

$$|r_{pq}(\delta)|^2 = \mathbf{\Theta}^T \left(\mathbf{r}_{pq}(\delta)\mathbf{r}_{pq}^H(\delta)\right)_R \mathbf{\Theta} + \mathbf{\Theta}^T \left(2\Delta_{pq}^*(\delta)\mathbf{r}_{pq}(\delta)\right)_R \tag{3.18}$$

with

$$\begin{aligned}
\Delta_{pq}(\delta) &= \tfrac{1}{2}\left(r_{pq} - r_{qp}\right), \\
\mathbf{r}_{pq}(\delta) &= \begin{bmatrix} \tfrac{1}{2}\left(r_{pq} + r_{qp}\right) & \tfrac{1}{2}\left(r_{qq} - r_{pp}\right) \end{bmatrix}^T
\end{aligned}$$

Likewise, the evaluation of the second summation in (3.16), $|r_{iq}(\delta)|^2$ leads to the linear relation

$$|r_{iq}(\delta)|^2 = -\mathbf{\Theta}^T \mathbf{r}_{i,pq}(\delta) \tag{3.19}$$

where $\mathbf{r}_{i,pq}(\delta) = [\frac{1}{2}(|r_{ip}|^2 - |r_{iq}|^2), (r_{ip}r_{iq}^*)_R]^T$. This allows to re-write $J_\times(\theta)$ in a quadratic-linear format as

$$\mathcal{J}_\times(\theta) = \mathbf{\Theta}^T \mathbf{R} \mathbf{\Theta} + \mathbf{\Theta}^T \mathbf{r} \tag{3.20}$$

where

$$\begin{aligned}
\mathbf{R} &= \sum_{\delta=-N_\delta}^{N_\delta} \left(\mathbf{r}_{pq}(\delta)\mathbf{r}_{pq}^H(\delta)\right)_R, \\
\mathbf{r} &= \sum_{\delta=-N_\delta}^{N_\delta} \left(2\Delta_{pq}^*(\delta)\mathbf{r}_{pq}(\delta)\right)_R \\
&\quad + \sum_{\substack{j=1 \\ j\neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{r}_{pq,j}(\delta) - \sum_{\substack{i=1 \\ i\notin\{p,q\}}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{r}_{i,pq}(\delta)
\end{aligned} \tag{3.21}$$

Finally, J_{hyb} is summarized as a sum of linear and quadratic terms

$$\begin{aligned} \mathcal{J}_{hyb}(\theta, 0) &= \mathbf{\Theta}^T \left[\mathbf{T} + \mathbf{R}\right] \mathbf{\Theta} + \mathbf{\Theta}^T \mathbf{r} \\ &= \mathbf{\Theta}^T \mathbf{Q} \mathbf{\Theta} + \mathbf{\Theta}^T \mathbf{q} \end{aligned} \tag{3.22}$$

The optimization problem (3.22) can be solved exactly using a constrained minimization problem. However, other approximate solutions are possible by linearly approximating the sine and cosine around zero point. Using the Lagrangian method, (3.22) can be set as

$$\min_{\mathbf{\Theta}} \quad \mathcal{F}(\mathbf{\Theta}) = \mathbf{\Theta}^T \mathbf{Q} \mathbf{\Theta} + \mathbf{\Theta}^T \mathbf{q} \quad \text{s.t} \quad \mathbf{\Theta}^T \mathbf{J}_G \mathbf{\Theta} = 1 \tag{3.23}$$

where $\mathbf{J}_G = diag\{[1, 1]\}$ is a constraint which is equivalent to the trigonometric identity $\sin^2(2\theta) + \cos^2(2\theta) = 1$. Considering the Lagrangian multiplier, (3.22) can be written as

$$\mathcal{L}(\mathbf{\Theta}, \beta) = \mathbf{\Theta}^T \mathbf{Q} \mathbf{\Theta} + \mathbf{\Theta}^T \mathbf{q} + \beta \left(\mathbf{\Theta}^T \mathbf{J}_G \mathbf{\Theta} - 1 \right) \tag{3.24}$$

where β is the Lagrange parameter. The solution of (3.24) is given as

$$\mathbf{\Theta} = -\frac{1}{2} (\mathbf{Q} + \beta \mathbf{J}_G)^{-1} \mathbf{q} \tag{3.25}$$

By substituting the solution in (3.25) into the constraint that is given in (3.23),

we obtain β as a solution of

$$\frac{1}{4}\mathbf{q}^T(\mathbf{Q}+\beta\mathbf{J}_G)^{-1}\mathbf{J}_G(\mathbf{Q}+\beta\mathbf{J}_G)^{-1}\mathbf{q}=1 \tag{3.26}$$

which is a 4-th order polynomial equation (see Appendix A). Interestingly, the rooting of this polynomial can be performed at insignificant cost using non-iteritive analtical procedures such as Ferrari or Cardanou's formulas [100, 101]. Among the 4 roots of (3.26), we select the one β_0 which is real-valued and corresponds to the least value of (3.24). Eventually, we solve for $\mathbf{\Theta}^0=[\theta_1^0,\theta_2^0]^T$ in (3.25) to get sine and cosine of θ as

$$\cos(\theta)=\sqrt{\frac{1+\theta_1^0}{2}} \quad , \quad \sin(\theta)=\frac{\theta_2^0}{\sqrt{2(1+\theta_1^0)}} \tag{3.27}$$

which allows us to find the entries of Givens rotation $\boldsymbol{\varphi}_{p,q}^a(\theta,0)$ assigned in (3.7).

Now, to evaluate the second Givens rotations $\boldsymbol{\varphi}_{p,q}^a\left(\theta',-\frac{\pi}{2}\right)$, a similar procedure will be followed. Then, one might express the first part of (3.4), i.e., $J_{MM}(\theta')$, as

$$\begin{aligned}&\mathcal{J}_{MM}(\theta')=\mathbf{\Theta}'^T\mathbf{T}'\mathbf{\Theta}'\\&=\mathbf{\Theta}'^T\left[2\sum_{k=1}^{K}\mathbf{t}'_{pq}(k)\mathbf{t}'^T_{pq}(k)+\mathbf{t}'_{qp}(k)\mathbf{t}'^T_{qp}(k)\right]\mathbf{\Theta}'\end{aligned} \tag{3.28}$$

The parameters of (3.28) are defined in Appendix B. Whereas, the second part of (3.4) is derived in a similar way to (3.20) and it is given as

$$\mathcal{J}_{\times}(\theta')=\mathbf{\Theta}'^T\mathbf{R}'\mathbf{\Theta}'+\mathbf{\Theta}'^T\mathbf{r}' \tag{3.29}$$

where

$$\begin{aligned}
\mathbf{R}' &= \sum_{\delta=-N_\delta}^{N_\delta} \left(\mathbf{r}'_{pq}(\delta)\mathbf{r}'^{H}_{pq}(\delta)\right)_R, \\
\mathbf{r}' &= \sum_{\delta=-N_\delta}^{N_\delta} \left(2\Delta'^{*}_{pq}(\delta)\mathbf{r}'_{pq}(\delta)\right)_R \\
&\quad + \sum_{\substack{j=1 \\ j\neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{r}'_{pq,j}(\delta) + \sum_{\substack{i=1 \\ i\notin\{p,q\}}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{r}'_{i,pq}(\delta)
\end{aligned} \tag{3.30}$$

and

$$\begin{aligned}
\Delta'_{pq}(\delta) &= \tfrac{1}{2}\left(r_{pq} + r_{qp}\right), \\
\mathbf{r}'_{pq}(\delta) &= \begin{bmatrix} \tfrac{1}{2}\left(r_{pq} - r_{qp}\right) & \tfrac{1}{2}\left(r_{pp} - r_{qq}\right) \end{bmatrix}^T, \\
\mathbf{r}'_{pq,j}(\delta) &= \begin{bmatrix} \tfrac{1}{2}(|r_{pj}|^2 - |r_{qj}|^2) & \left(\sqrt{-1}r_{pj}r^*_{qj}\right)_R \end{bmatrix}^T, \\
\mathbf{r}'_{i,qp}(\delta) &= \begin{bmatrix} \tfrac{1}{2}(|r_{iq}|^2 - |r_{ip}|^2) & \left(\sqrt{-1}r_{ip}r^*_{iq}\right)_R \end{bmatrix}^T
\end{aligned} \tag{3.31}$$

Ultimately, the hybrid cost function corresponding to the second group of Givens rotations $\boldsymbol{\varphi}^a_{p,q}\left(\theta', -\frac{\pi}{2}\right)$ can be stated in a quadratic-linear form as

$$\begin{aligned}
\mathcal{J}_{hyb}(\theta', -\tfrac{\pi}{2}) &= \boldsymbol{\Theta}'^{T}\left[\mathbf{T}' + \mathbf{R}\right]\boldsymbol{\Theta}' + \boldsymbol{\Theta}'^{T}\mathbf{r}' \\
&= \boldsymbol{\Theta}'^{T}\mathbf{Q}'\boldsymbol{\Theta}' + \boldsymbol{\Theta}'^{T}\mathbf{q}'
\end{aligned} \tag{3.32}$$

Then, matrix $\boldsymbol{\varphi}^a_{p,q}\left(\theta', -\frac{\pi}{2}\right)$ is obtained using a similar approach to (3.25)-(3.27), and according to (3.7), it should be applied successively on matrix $\underline{\mathbf{Y}}$ so that the filtering matrix $\mathbf{V}$ can be constructed accordingly.

The last phase term $e^{-j\theta''}$ can be evaluated by considering the compensation for the concerned index p, i.e., $\boldsymbol{z}_p(k) = e^{-j\theta''}\underline{y}_p(k)$. Then, the equivalent MM cost

function, expressed in terms of θ'', is obtained simply by setting $(p = q = 1)$ in (7.8).

For the remaining rotations, and since we are interested in deconvolving the first N_t sources of $\underline{\mathbf{Y}}$, only the p-th rows are involved in the cost function during the evaluation of product set b in (3.7). In that case, considering $\boldsymbol{\varphi}_{p,q}^{b}(\theta, 0)$ and ignoring the q parts in (3.12) and (3.13), (3.9) is restructured in a linear-quadratic form as (constant terms are dropped)

$$\mathcal{J}_{MM}(\theta) = \boldsymbol{\Theta}^T \mathbf{T} \boldsymbol{\Theta} \ + \boldsymbol{\Theta}^T \mathbf{t} \tag{3.33}$$

where

$$\begin{aligned} \mathbf{T} &= \sum_{k=1}^{K} \mathbf{t}_{pq,R}(k)\mathbf{t}_{pq,R}^{T}(k) + \mathbf{t}_{pq,I}(k)\mathbf{t}_{pq,I}^{T}(k) \\ \mathbf{t} &= 2\sum_{k=1}^{K} \mathbf{t}_{pq,R}(k)c_{pq,R}(k) + \mathbf{t}_{pq,I}(k)c_{pq,I}(k) \end{aligned}$$

In regards to the cross-correlation term, the second sum in (3.16) becomes irrelevant while the first one does, also q indices are out of bound (i.e., $q > N_t$). In this case $\mathcal{J}_{\times}(\theta)$ will be given only by the linear relation (3.17) and the net hybrid cost function in (3.22) is modified accordingly and is given as

$$\begin{aligned} \mathcal{J}_{hyb}(\theta, 0) &= \boldsymbol{\Theta}^T \mathbf{T} \boldsymbol{\Theta} + \boldsymbol{\Theta}^T [\mathbf{t} + \mathbf{r}] \\ &= \boldsymbol{\Theta}^T \mathbf{Q} \boldsymbol{\Theta} + \boldsymbol{\Theta}^T \mathbf{q} \end{aligned} \tag{3.34}$$

where

$$\mathbf{r} = \sum_{\substack{j=1 \\ j \neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{r}_{pq,j}(\delta)$$

The evaluation of the second angle (θ') in rotation set b is also analogous to the evaluation of the same angle in set a. Excluding the q-th part of (3.9) and making use of the real and imaginary parts of p-th index in (7.7), results into the following linear-quadratic reduced-MM equivalent cost function

$$\mathcal{J}_{MM}(\theta') = \mathbf{\Theta}'^T \mathbf{T}' \mathbf{\Theta}' + \mathbf{\Theta}'^T \mathbf{t}' \tag{3.35}$$

where

$$\begin{aligned}
\mathbf{T}' &= \sum_{k=1}^{K} \mathbf{t}'_{pq}(k) \mathbf{t}'^T_{pq}(k) + \mathbf{t}'_{qp}(k) \mathbf{t}'^T_{qp}(k) \\
\mathbf{t}' &= 2 \sum_{k=1}^{K} \mathbf{t}'_{pq}(k) c'_{pq}(k) - \mathbf{t}'_{qp}(k) c'_{qp}(k) \\
c'_{pq}(k) &= \tfrac{1}{2} \left(\underline{y}^2_{p,R}(k) + \underline{y}^2_{q,I}(k) \right) - R \\
c'_{qp}(k) &= \tfrac{1}{2} \left(\underline{y}^2_{q,R}(k) + \underline{y}^2_{p,I}(k) \right) - R
\end{aligned}$$

Considering the first summation of (3.16), the composite cost function (3.4) reduces to

$$\begin{aligned}
\mathcal{J}_{hyb}(\theta', -\tfrac{\pi}{2}) &= \mathbf{\Theta}'^T \mathbf{T}' \mathbf{\Theta}' + \mathbf{\Theta}'^T \left[\mathbf{t}' + \mathbf{r}' \right] \\
&= \mathbf{\Theta}'^T \mathbf{Q}' \mathbf{\Theta}' + \mathbf{\Theta}'^T \mathbf{q}'
\end{aligned} \tag{3.36}$$

Table 3.1: G-MMDA and HG-MMDA Algorithms

a. Initialize $\mathbf{V} = \mathbf{I}_{N_d}$
b. Pre-Whitening: $\underline{\mathbf{Y}} = \mathbf{BY}$
c. Givens Rotations:
for $n = 1 : N_{Sweeps}$ **do**
  for $p = 1 : N_t$ **do**
    1. Evaluate $\mathbf{D}_p(\theta'')$
    2. Update $\underline{\mathbf{Y}} = \mathbf{D}_p(\theta'')\underline{\mathbf{Y}}$ and $\mathbf{V} = \mathbf{D}_p(\theta'')\mathbf{V}$
    for $q = p + 1 : N_t$ **do**
      3. Evaluate $\boldsymbol{\varphi}^a_{p,q}(\theta, 0)$ using (3.22), (3.27) and (2.15)
      4. Update $\underline{\mathbf{Y}} = \boldsymbol{\varphi}^a_{p,q}(\theta, 0)\underline{\mathbf{Y}}$ and $\mathbf{V} = \boldsymbol{\varphi}^a_{p,q}(\theta, 0)\mathbf{V}$
      5. Evaluate $\boldsymbol{\varphi}^a_{p,q}\left(\theta', -\frac{\pi}{2}\right)$ using (3.32), (3.27) and (2.15)
      6. Update $\underline{\mathbf{Y}} = \boldsymbol{\varphi}^a_{p,q}\left(\theta', -\frac{\pi}{2}\right)\underline{\mathbf{Y}}$, $\mathbf{V} = \boldsymbol{\varphi}^a_{p,q}\left(\theta', -\frac{\pi}{2}\right)\mathbf{V}$
    end for
    for $q > N_t$ **do**
      7. Evaluate $\boldsymbol{\varphi}^b_{p,q}(\theta, 0)$ using (3.34), (3.27) and (2.15)
      8. Update $\underline{\mathbf{Y}} = \boldsymbol{\varphi}^b_{p,q}(\theta, 0)\underline{\mathbf{Y}}$ and $\mathbf{V} = \boldsymbol{\varphi}^b_{p,q}(\theta, 0)\mathbf{V}$
      9. Evaluate $\boldsymbol{\varphi}^a_{p,q}\left(\theta', -\frac{\pi}{2}\right)$ using (3.36), (3.27) and (2.15)
      10. Update $\underline{\mathbf{Y}} = \boldsymbol{\varphi}^b_{p,q}\left(\theta', -\frac{\pi}{2}\right)\underline{\mathbf{Y}}$, $\mathbf{V} = \boldsymbol{\varphi}^b_{p,q}\left(\theta', -\frac{\pi}{2}\right)\mathbf{V}$
    end for
  end for
end for
d. Compute the deconvolution matrix: $\mathbf{W} = \mathbf{JVB}$
e. Recover the sources: $\hat{\mathbf{S}} = \mathbf{WY}$, which is equivalent to $\boldsymbol{J}\underline{\mathbf{Y}}$

Note: To get the HG-MMDA, we repeat steps: $3 - 4$ after 4 using $\boldsymbol{\phi}^a_{p,q}(\gamma, 0)$, steps: $5 - 6$ after 6 using $\boldsymbol{\phi}^a_{p,q}(\gamma', -\pi/2)$, steps: $7 - 8$ after 8 using $\boldsymbol{\phi}^b_{p,q}(\gamma, 0)$, and steps: $9 - 10$ after 10 using $\boldsymbol{\phi}^b_{p,q}(\gamma', -\pi/2)$.

where

$$\mathbf{r}' = \sum_{\substack{j=1 \\ j \neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{r}'_{pq,j}(\delta)$$

At the end, matrix $\mathbf{V}$ is initialized[11] as $\mathbf{V} = \mathbf{I}_{N_d}$ and the complete derived G-MMDA is summarized in Table 3.1.

[11]The selection matrix $\boldsymbol{J}$ is applied only at the last stage after the algorithm's convergence.

Hyperbolic Givens MMDA (HG-MMDA)

In the case of small number of samples N_s, the pre-whitening is not effective, accordingly, the transformed convolution channel matrix $\boldsymbol{\mathcal{H}}$ would be far from being unitary. For this reason, the performance of the proposed G-MMDA algorithm lessens. To overcome this limitation, the Hyperbolic (Shear) non-unitary rotations are applied alternatively along with the Givens rotations. This results in the Hyperbolic Givens MMDA (HG-MMDA). Hence, the matrix $\boldsymbol{V}$, can be decomposed into a product of elementary unitary $\boldsymbol{\varphi}_{p,q}$ and non-unitary $\boldsymbol{\phi}_{p,q}$ rotations as follows:

$$\begin{aligned}\mathbf{V} = \mathbf{J} \prod_{N_{Sweeps}} \prod_{1\le p\le q\le N_t}^{a} \boldsymbol{\Upsilon}^{a}_{p,q}\left(\varsigma, 0\right) \boldsymbol{\Upsilon}^{a}_{p,q}\left(\varsigma', -\tfrac{\pi}{2}\right) \mathbf{D}_p(\theta'') \\ \prod_{\substack{1\le p\le N_t \\ N_t<q\le N_d}}^{b} \boldsymbol{\Upsilon}^{b}_{p,q}\left(\varsigma, 0\right) \boldsymbol{\Upsilon}^{b}_{p,q}\left(\varsigma', -\tfrac{\pi}{2}\right)\end{aligned} \tag{3.37}$$

where

$$\boldsymbol{\Upsilon}_{p,q}(\varsigma, \alpha) = \boldsymbol{\varphi}_{p,q}(\theta, \alpha)\boldsymbol{\phi}_{p,q}(\gamma, \alpha), \quad \varsigma = [\theta, \gamma], \alpha \in \{0, \frac{\pi}{2}\} \tag{3.38}$$

Again the special version of the complex Hyperbolic rotations is used, also, the derivation flow is very similar to what has been presented earlier. Hence, we brief the findings of the Hyperbolic transformation parameters to minimize the hybrid criterion (3.4).

Consider a single unitary rotation, given as $\mathbf{Z} = \boldsymbol{\phi}^{a}_{p,q}\left(\gamma, 0\right) \underline{\mathbf{Y}}$. According to

(2.16), only rows p and q change. Then, (3.9) can be re-written in terms of Shear angle (γ), as well. Hence, as derived for the Givens rotations, we can show that

$$\begin{aligned} z_p^2(k) &= \mathbf{\Phi}^T \mathbf{u}_{pq}(k) + \tfrac{1}{2}\left(\underline{y}_p^2(k) - \underline{y}_q^2(k)\right) \\ z_q^2(k) &= \mathbf{\Phi}^T \mathbf{u}_{pq}(k) + \tfrac{1}{2}\left(\underline{y}_q^2(k) - \underline{y}_p^2(k)\right) \end{aligned} \tag{3.39}$$

where

$$\begin{aligned} \mathbf{\Phi} &= \begin{bmatrix} \cosh(2\gamma) & \sinh(2\gamma) \end{bmatrix}^T, \\ \mathbf{u}_{pq}(k) &= \begin{bmatrix} \tfrac{1}{2}\left(\underline{y}_p^2(k) + \underline{y}_q^2(k)\right) & \underline{y}_p(k)\underline{y}_q(k) \end{bmatrix}^T \end{aligned} \tag{3.40}$$

Similar to the derivations in (3.12)-(3.13), using (3.39) and ignoring the terms that are independent of γ, we express (3.9) by the compact quadratic-linear formulation as

$$\mathcal{J}_{MM}(\gamma) = \mathbf{\Phi}^T \mathbf{U} \mathbf{\Phi} \ + \mathbf{\Phi}^T \mathbf{u} \tag{3.41}$$

where

$$\begin{aligned} \mathbf{U} &= \sum_{k=1}^{K} \mathbf{u}_{pq,R}(k)\mathbf{u}_{pq,R}^T(k) + \mathbf{u}_{pq,I}(k)\mathbf{u}_{pq,I}^T(k), \\ \mathbf{u} &= -2R \sum_{k=1}^{K} \mathbf{u}_{pq,R}(k) + \mathbf{u}_{pq,I}(k), \end{aligned}$$

Also, the corresponding $\mathcal{J}_{\times}(\boldsymbol{V})$ terms are evaluated by imitating the steps (3.16)-

(3.21). One can easily express $\mathcal{J}_{\times}(\boldsymbol{V})$ as

$$\mathcal{J}_{\times}(\gamma) = \mathbf{\Phi}^T \mathbf{O} \mathbf{\Phi} + \mathbf{\Phi}^T \mathbf{o} \tag{3.42}$$

where

$$\begin{aligned}
\mathbf{O} &= \sum_{\delta=-N_\delta}^{N_\delta} \left(\mathbf{o}_{pq}(\delta)\mathbf{o}_{pq}^H(\delta)\right)_R, \\
\mathbf{o} &= \sum_{\delta=-N_\delta}^{N_\delta} \left(2\Delta_{pq}^*(\delta)\mathbf{o}_{pq}(\delta)\right)_R \\
&\quad + \sum_{\substack{j=1 \\ j\neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{o}_{pq,j}(\delta) + \sum_{\substack{i=1 \\ i\notin\{p,q\}}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{o}_{i,pq}(\delta), \\
\mathbf{o}_{pq}(\delta) &= [(r_{pq}(\delta) + r_{qp}(\delta)), (r_{pp}(\delta) + r_{qq}(\delta))]^T/2, \\
\mathbf{o}_{pq,j}(\delta) &= [(|r_{pj}(\delta)|^2 + |r_{qj}(\delta)|^2)/2, \left(r_{pj}(\delta)r_{qj}^*(\delta)\right)_R]^T, \\
\mathbf{o}_{i,pq}(\delta) &= [(|r_{ip}(\delta)|^2 + |r_{iq}(\delta)|^2)/2, \left(r_{ip}(\delta)r_{iq}^*(\delta)\right)_R]^T
\end{aligned} \tag{3.43}$$

Hence, we use (3.41) and (3.42) to express $\mathcal{J}_{hyb}$ in (3.5) as

$$\begin{aligned}
\mathcal{J}_{hyb}(\gamma, 0) &= \mathbf{\Phi}^T \left[\mathbf{U} + \mathbf{O}\right] \mathbf{\Phi} + \mathbf{\Phi}^T \left[\mathbf{u} + \mathbf{o}\right] \\
&= \mathbf{\Phi}^T \mathbf{Q} \mathbf{\Phi} + \mathbf{\Phi}^T \mathbf{q}
\end{aligned} \tag{3.44}$$

The optimization problem (3.44) can also be solved using the Lagrange multiplier as followed earlier (see (3.23)-(3.25)). This is done by imposing the constraint $\cosh^2(2\gamma) - \sinh^2(2\gamma) = 1$ and, consequently, replacing $\mathbf{J}_G$ by $\mathbf{J}_H = diag\{[1, -1]\}$,.

Then, we solve for $\mathbf{\Phi}^0 = [\gamma_1^0, \gamma_2^0]^T$ to get[12]

$$\cosh(\gamma) = \sqrt{\frac{1+\gamma_1^0}{2}} \quad , \quad \sinh(\gamma) = \frac{\gamma_2^0}{\sqrt{2(1+\gamma_1^0)}} \tag{3.45}$$

which allows us to find the entries of the Shear rotations $\boldsymbol{\phi}_{p,q}^a(\gamma, 0)$ defined in (2.16).

Now, to evaluate the second Shear rotations $\boldsymbol{\phi}_{p,q}^a\left(\gamma', -\frac{\pi}{2}\right)$, a similar procedure to the one used for $\boldsymbol{\varphi}_{p,q}^a\left(\theta', -\frac{\pi}{2}\right)$ evaluation is followed, which lead to (see Appendix C)

$$\mathcal{J}_{MM}(\gamma') = \mathbf{\Phi}'^T \boldsymbol{U}' \mathbf{\Phi}' \ + \mathbf{\Phi}'^T \mathbf{u}' \tag{3.46}$$

Similar to (3.20) and (3.29), the second part of (3.4) is found as follows:

$$\mathcal{J}_{\times}(\gamma') = \mathbf{\Phi}'^T \mathbf{O}' \mathbf{\Phi}' + \mathbf{\Phi}'^T \mathbf{o}' \tag{3.47}$$

[12]Note that the first entry of $\mathbf{\Phi}$ is forced to be non-negative.

where

$$
\begin{aligned}
\mathbf{O}' &= \sum_{\delta=-N_\delta}^{N_\delta} \left(\mathbf{o}'_{pq}(\delta)\mathbf{o}'^{H}_{pq}(\delta)\right)_R, \\
\mathbf{o}' &= \sum_{\delta=-N_\delta}^{N_\delta} \left(2\Delta'^{*}_{pq}(\delta)\mathbf{o}'_{pq}(\delta)\right)_R \\
&\quad + \sum_{\substack{j=1 \\ j\neq p}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{o}'_{pq,j}(\delta) + \sum_{\substack{i=1 \\ i\notin\{p,q\}}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{o}'_{i,pq}(\delta), \\
\mathbf{o}'_{pq}(\delta) &= [(r_{pq}(\delta) - r_{qp}(\delta)), -\sqrt{-1}\,(r_{pp}(\delta) + r_{qq}(\delta))]^T/2, \\
\mathbf{o}'_{pq,j}(\delta) &= [(|r_{pj}(\delta)|^2 + |r_{qj}(\delta)|^2)/2, \left(\sqrt{-1}r_{pj}(\delta)r^*_{qj}(\delta)\right)_R]^T \\
\mathbf{o}'_{i,pq}(\delta) &= [(|r_{ip}(\delta)|^2 + |r_{iq}(\delta)|^2)/2, \left(\sqrt{-1}r_{ip}{}^*(\delta)r_{iq}(\delta)\right)_R]^T
\end{aligned}
\tag{3.48}
$$

Then, (3.47) and (7.11) are used to evaluate $\mathcal{J}_{hyb}$ in (3.4) as

$$
\begin{aligned}
J_{hyb}(\gamma', 0) &= \mathbf{\Phi}'^T \left[\mathbf{U}' + \mathbf{O}'\right] \mathbf{\Phi}' + \mathbf{\Phi}'^T \left[\mathbf{u}' + \mathbf{o}'\right] \\
&= \mathbf{\Phi}'^T \boldsymbol{Q}' \mathbf{\Phi}' + \mathbf{\Phi}'^T \mathbf{q}'
\end{aligned}
\tag{3.49}
$$

This optimization is also solved using the Lagrange multiplier approach considered before.

Having evaluated the first set of shear rotations, i.e. 'case a', we proceed to evaluate 'case b', for which the term $\boldsymbol{u}$ in (3.41) becomes:

$$
\boldsymbol{u} = 2\sum_{k=1}^{K} \mathbf{u}_{pq,R}(k)\bar{c}_{pq,R}(k) + \mathbf{u}_{pq,I}(k)\bar{c}_{pq,I}(k) \tag{3.50}
$$

where $\bar{c}_{pq}(k) = (\underline{y}^2_p(k) - \underline{y}^2_q(k))/2 - R$. Likewise, the arguments explained to

evaluate $\mathcal{J}_{\times}(\theta)$ in (3.34), are followed here to find $\mathcal{J}_{\times}(\gamma)$ as

$$\mathcal{J}_{\times}(\gamma) = \mathbf{\Phi}^T \mathbf{o} = \mathbf{\Phi}^T \sum_{\substack{j=1 \\ j \notin \{p,q\}}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{o}_{pq,j}(\delta)$$

Accordingly,

$$\begin{aligned} \mathcal{J}_{hyb}(\gamma, 0) &= \mathbf{\Phi}^T \mathbf{U} \mathbf{\Phi} + \mathbf{\Phi}^T [\mathbf{u} + \mathbf{o}] \\ &= \mathbf{\Phi}^T \mathbf{Q} \mathbf{\Phi} + \mathbf{\Phi}^T \mathbf{q} \end{aligned} \tag{3.51}$$

Eventually, the evaluation of the second angle (γ') in rotation set b is also analogous to the evaluation of the same angle in set a. We obtain:

$$\mathcal{J}_{MM}(\gamma') = \mathbf{\Phi'}^T \mathbf{U'} \mathbf{\Phi'} \ + \mathbf{\Phi'}^T \mathbf{u'} \tag{3.52}$$

where (details are given in Appendix C)

$$\begin{aligned} \mathbf{U'} &= \sum_{k=1}^{K} \mathbf{u}'_{pq}(k) \mathbf{u}'^T_{pq}(k) + \mathbf{u}'_{qp}(k) \mathbf{u}'^T_{qp}(k) \\ \mathbf{u'} &= 2 \sum_{k=1}^{K} \mathbf{u}'_{pq}(k) \bar{c}'_{pq}(k) + \mathbf{u}'_{qp}(k) \bar{c}'_{qp}(k) \\ \bar{c}'_{pq}(k) &= \tfrac{1}{2} \left(\underline{y}^2_{p,R}(k) - \underline{y}^2_{q,I}(k) \right) - R \\ \bar{c}'_{qp}(k) &= \tfrac{1}{2} \left(\underline{y}^2_{q,I}(k) - \underline{y}^2_{p,R}(k) \right) - R \end{aligned}$$

Considering the first summation of (3.16), the composite cost function (3.4) re-

duces to

$$\begin{aligned}\mathcal{J}_{hyb}(\gamma', -\tfrac{\pi}{2}) &= \mathbf{\Phi}'^T \mathbf{U}' \mathbf{\Phi}' + \mathbf{\Phi}'^T \left[\mathbf{u}' + \mathbf{o}'\right] \\ &= \mathbf{\Phi}'^T \mathbf{Q} \mathbf{\Phi}' + \mathbf{\Phi}'^T \mathbf{q}\end{aligned} \tag{3.53}$$

where

$$\mathbf{o}' = \sum_{\substack{j=1 \\ j\neq}}^{N_t} \sum_{\delta=-N_\delta}^{N_\delta} \mathbf{o}'_{pq,j}(\delta)$$

The complete HG-MMDA is summarized in Table 3.1.

3.4.2 Deconvolution Using Deflation Approach

In the subsequent sections, we use the Givens and Shear rotations also to develop new deflation based deconvolution algorithms[13]. In the algorithm designs we resort to extract a single source each time. After this, an appropriate subspace projection procedure is used to remove the effect of the extracted source and its potential delayed versions.

Givens Deflation MMDA (G-DMMDA)

Since each time we target to extract a single source, one can reduce the MM criterion (3.9), with $p = 1$, into

$$\mathcal{J}_{MM}(\mathbf{V}) = \frac{1}{K} \sum_{k=1}^{K} \left[\left(z_{1,R}^2(k) - R_R \right)^2 + \left(z_{1,I}^2(k) - R_I \right)^2 \right] \tag{3.54}$$

[13]Part of this work have been presented in [102].

Following (3.7), the complex matrix $\mathbf{V}$ is accordingly decomposed into successive multiplications of the elementary Givens rotations as[14]

$$\mathbf{V} = \mathbf{j} \prod_{N_{sweeps}} \mathbf{D}_p(\theta'') \prod_{p=1<q\leq N_d} \boldsymbol{\varphi}_{p,q}(\theta, 0)\, \boldsymbol{\varphi}_{p,q}\left(\theta', -\frac{\pi}{2}\right) \tag{3.55}$$

where $\mathbf{j} = [1, \mathbf{0}]$ is a row selection vector. According to (3.55), p is maintained to 1 in order to retrieve the first row, while q is allowed to sweep and combine the whole signal space domain, i.e., $1 < q \leq N_d$. To reduce the MM criterion (3.54) into a criterion of finding the angles parameters, we take advantage of the earlier derivation carried for G-MMDA.

Thus, considering the first two sets of rotations $\boldsymbol{\varphi}_{p,q}(\theta, 0)$ and $\boldsymbol{\varphi}_{p,q}(\theta', -\pi/2)$, then, the equivalent MM cost function, expressed as a function of parameter θ and θ', is given in (3.33) and (3.35), respectively with $p = 1$. in similar analogy we evaluate the third rotations $e^{-j\theta''}$ is obtained by setting $p = q = 1$ in (7.8).

As pointed earlier, the deflation approach is an iterative method such that each run results into extracting a single source. Each source has an expected number of replicas which depend on the window size and the channel memory, i.e., $N_\delta = N_w + M$. To waive the effect of the extracted source and its corresponding delayed copies, define the time-lagged correlation matrix $\boldsymbol{R}_i$ between the extracted source $\boldsymbol{z}_i = \mathbf{j}\boldsymbol{V}\underline{\mathbf{Y}}_{i-1}$ and the data matrix $\underline{\mathbf{Y}}_{i-1}$, at the $(i-1)$-th deflation run (i

[14]For ease of presentation, we keep using the index p, despite it is always fixed to 1.

indicates the order of the extracted source) as

$$\boldsymbol{R}_i = \left[E\left[\boldsymbol{y}_{i-1}(k)z_i^*(k+N_\delta)\right], \cdots, E\left[\boldsymbol{y}_{i-1}(k)z_i^*(k-N_\delta)\right)\right] \tag{3.56}$$

where $\boldsymbol{y}_{i-1}(k)$ is the k-th column vector of matrix $\underline{\mathbf{Y}}_{i-1}$. The singular value decomposition of $\boldsymbol{R}_i$ has the form

$$\mathbf{R}_i = \mathbf{Q}_i diag\left(\lambda_1^2, \cdots, \lambda_{N_\delta}^2\right)\mathbf{Q}_i^H + \bar{\sigma}^2\bar{\mathbf{Q}}_i\bar{\mathbf{Q}}_i{}^H \tag{3.57}$$

where λ_i's are the singular values of $\mathbf{R}_i$, also the columns of $\mathbf{Q}_i$ and $\bar{\mathbf{Q}}_\mathbf{i}$ spans the so-called extracted source subspace, i.e., $span\{\mathbf{Q}_i\} = span\{\mathbf{z}_1\}$ and the non-extracted sources subspace (orthogonal complement), respectively. Now, given that one source is extracted successfully, then the extracted source subspace eigenvalues (λ_i) are much greater than $\bar{\sigma}^2$ and very close to 1, i.e., $\bar{\sigma}^2 \ll \lambda_i \approx 1$. Eventually, after having the basis of the extracted source subspace, the new set of data which will be ready for the next source extraction, i.e., $\underline{\mathbf{Y}}_i$ is obtained by

$$\underline{\mathbf{Y}}_i = \boldsymbol{P}_i\underline{\mathbf{Y}}_{i-1} \tag{3.58}$$

where $\boldsymbol{P}_i = \boldsymbol{I} - \boldsymbol{Q}_i\boldsymbol{Q}_i^+$ is the orthogonal projector which is supposed to "deplete the effect" of the extracted source and its replicas contribution from the data matrix $\underline{\mathbf{Y}}_{i-1}$. Again, the depleted matrix $\underline{\mathbf{Y}}_i$ will go through the process again to separate a new source signal until all the source signals are recovered. The complete derived G-DMMDA is summarized in Table 3.2.

Next, we discuss an alternative way to find the projector matrix $\boldsymbol{P}_i$ without resorting to the costly SVD decomposition. As shown earlier, the matrix $\boldsymbol{R}_i$ would have ideally N_δ non zero column vectors. In practice, the latter are obtained by selecting the N_δ column vectors of $\boldsymbol{R}_i$ of maximum norm. This set of columns forms the matrix $\boldsymbol{Q}_i$

It is customarily to mention that, the proposed method regularly suffers from local minima problems which is the consequence of shortening the sweeping space, i.e., fixing $p = 1$. One of the strategic plans that is used here is to re-randomize the data matrix $\underline{\mathbf{Y}}$ once the algorithm gets stuck in a local minima. The re-randomization can be done by initializing $\boldsymbol{V}$ to be a random matrix instead of identity and as a consequence the updated matrix $\underline{\mathbf{Y}}$ becomes $\boldsymbol{V}\underline{\mathbf{Y}}$, i.e., $\underline{\mathbf{Y}} = \boldsymbol{V}\underline{\mathbf{Y}}$.

Finally, one might use the proposed algorithms, HG-DMMDA and G-DMMDA, as new BSS algorithms to seperate instantaneous mixtures. In either case, i.e. using the algorithms for instantaneous or convolutive mixtures, our methodology has a gain in terms of computational complexity as compared to [6] and [7]. This is due to the fact that our approach considers sweeping the concerned source only by fixing $p = 1$. Moreover, as discussed earlier, we devised a less complex alternative method to evaluate $\boldsymbol{Q}_i$ rather than performing the costly SVD decomposition (Table 3.2, step III).

Hyperbolic G-DMMDA (HG-DMMDA)

to substitute for the inadequacy in the whitening phase, specially in case of small processing samples, the decomposition of $\boldsymbol{V}$ is conducted based on the successive

Table 3.2: G-DMMDA and HG-DMMDA Algorithms

Pre-Whitening: $\underline{\mathbf{Y}}_0 = \mathbf{BY}$
Initialize $\mathbf{V} = \mathbf{I}_{N_d}$
Givens Rotations and Deflation:
for $i = 1 : N_t$ **do**
  for $n = 1 : N_{Sweeps}$ **do**
    1. Evaluate $\mathbf{D}_p(\theta'')$
    2. Update $\underline{\mathbf{Y}}_i = \mathbf{D}_p(\theta'')\underline{\mathbf{Y}}_{i-1}$ and $\mathbf{V} = \mathbf{D}_p(\theta'')\mathbf{V}$
    for $p = 1, q = 1 : N_d$ **do**
      1. Evaluate $\mathbf{D}_p(\theta'')$
      2. Update $\underline{\mathbf{Y}}_i = \mathbf{D}_p(\theta'')\underline{\mathbf{Y}}_{i-1}$ and $\mathbf{V} = \mathbf{D}_p(\theta'')\mathbf{V}$
      3. Evaluate $\boldsymbol{\varphi}_{p,q}(\theta, 0)$ using (3.33), (3.27) and (2.15)
      4. Update $\underline{\mathbf{Y}}_i = \boldsymbol{\varphi}_{p,q}(\theta, 0)\underline{\mathbf{Y}}_{i-1}$ and $\mathbf{V} = \boldsymbol{\varphi}_{p,q}(\theta, 0)\mathbf{V}$
      5. Evaluate $\boldsymbol{\varphi}_{p,q}\left(\theta', -\frac{\pi}{2}\right)$ using (3.35), (3.27) and (2.15)
      6. Update $\underline{\mathbf{Y}}_i = \varphi_{p,q}\left(\theta', -\frac{\pi}{2}\right)\underline{\mathbf{Y}}_{i-1}$, $\mathbf{V} = \varphi_{p,q}\left(\theta', -\frac{\pi}{2}\right)\mathbf{V}$
    end for
  end for
  1. Recover the source number i: $\boldsymbol{z}_i = \mathbf{j}\boldsymbol{V}\underline{\mathbf{Y}}_{i-1}$
  II. Evaluate correlation matrix: $\boldsymbol{R}_i$ using (3.56)
  III. Perform SVD (or use the alternative method) to evaluate $\boldsymbol{Q}_i$
  IV. Evaluate the projector matrix: $\boldsymbol{P}_i = \boldsymbol{I} - \boldsymbol{Q}_i\boldsymbol{Q}_i^{+}$
  V. Find the new subset of data: $\underline{\mathbf{Y}}_i = \boldsymbol{P}_i\underline{\mathbf{Y}}_{i-1}$
end for

Note: To get the HG-DMMDA, we repeat steps: $3-4$ after 4 using $\boldsymbol{\phi}_{p,q}(\gamma, 0)$ and steps: $5-6$ after 6 using $\boldsymbol{\phi}_{p,q}(\gamma', -\pi/2)$.

applications of the Givens and Shear rotations. Using (3.37), the decomposition becomes

$$\mathbf{V} = \mathbf{j} \prod_{N_{Sweeps}} \mathbf{D}_p(\theta'') \prod_{p=1 \le q \le N_d} \boldsymbol{\Upsilon}_{p,q}(\varsigma, 0)\, \boldsymbol{\Upsilon}_{p,q}\left(\varsigma', -\frac{\pi}{2}\right) \tag{3.59}$$

Now, considering the first two sets of rotations $\boldsymbol{\phi}_{p,q}(\theta, 0)$ and $\boldsymbol{\phi}_{p,q}(\theta', -\pi/2)$, then, the equivalent MM cost function expressed as a function of parameter γ and γ' is given in ((3.41), (3.50)) and (3.52), respectivly. The complete HG-DMMDA is

summarized in Table 3.2.

3.5 Simulation Results

This section elaborates on the performance of the proposed algorithms by presenting simulation results. Due to the non-presence of the batch MM deconvolution algorithms, we do the comparison among the different proposed algorithms, also with the CM-like algorithms when it is applicable. The comparison is delivered in terms of the average symbol error rate (SER).

The first experiment compares algorithms' performance againest the processed samples size N_s. In this regards, we consider a MIMO communication system with 3 transmitters and 5 receivers, i.e. ($N_t = 3$, $N_r = 5$). The sources transmit uncoded samples which are drawn from 4-QAM constellation. At each Monte-Carlo run, the data are passed throughout a randomly generated convolution matrix $\boldsymbol{\mathcal{H}}$ with a channel degree $M = 2$. The used window size is set to 3, i.e., $N_w = 3$, and is selected in a compliance with (2.3). The noise variance is modified according to the handled SNR. The results are averaged over 1000 trials.

Considering the on "Full BSS" method, Fig. 3.4 depicts the effect of the sample size for different MM-BSS (G-MMA, HG-MMA) and CM-BSS (G-CMA, HG-CMA, ACMA) algorithms. It is clear that the ACMA has a degraded performance as compared to the unitary/non-unitary based algorithms. This means that at small number of samples, the ACMA does't appropriately handle QAM signals. Also, the figure shows that the rest of the "Full-BSS" methods feasibly work for

data samples greater than 200, i.e.,$N_s \geq \mathcal{O}\left(10N_d\right)$.

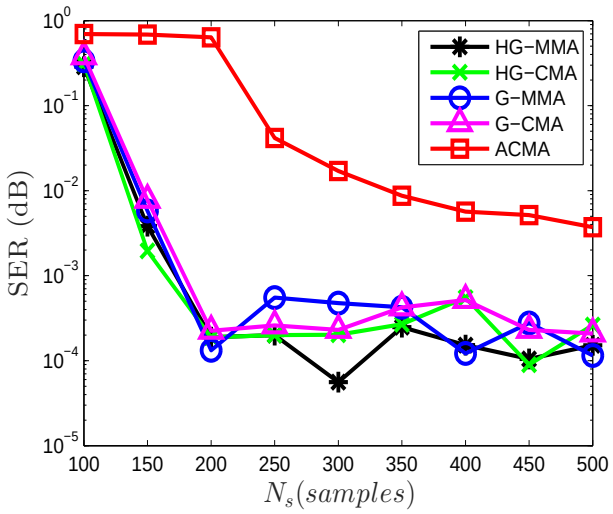


Figure 3.4: Average SER of "Full-BSS" algorithms (using ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. N_s with $SNR = 30\ dB$ and 4-QAM.

On the other hand, Fig. 3.5 demonstrate the behavior of the Equalization-BSS algorithms over different sample size. Unlike the Full-BSS methods, the algorithms' performance is monotonically iproving as the processed number of samples increases. Also, one notice that the Equalization-BSS class is more efficient than the Full-BSS class in terms of the processing data size.

The behavior of the hybrid-based deconvolution algorithms (HG-MMDA, G-MMDA) and the deflated-based algorithms (G-DMMDA and HG-DMMDA) is shown in Fig. 3.6 and Fig. 3.7, respectively. They have a comparable performance with the "Full-BSS" algorithms. Additionally, as expected, we notice that at comparably low N_s, the HG-DMMDA has a remarkable gain over the G-DMMDA. This is due to the ineffective pre-whitening operation. However, at sufficient large number of samples, all methods in both classes demonstrate a similar performance.

So far, the performance of different proposed approaches with the size of the

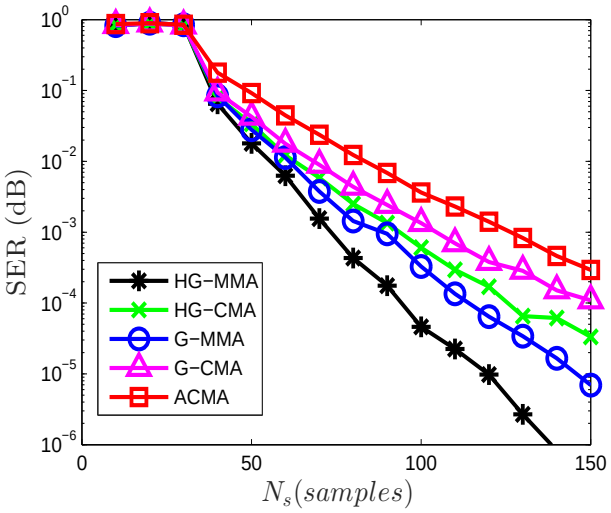


Figure 3.5: Average SER of Equalization-BSS (using ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. N_s with $SNR = 30\ dB$ and 16-QAM.

processed number of samples N_s is presented. We regarded that each class has a completely different behavior, despite that all performing the same task. The simulation is done at fixed high SNR, i.e., 30 dB. However, one might ask about their performance at different noise level. This question is incorporated in the subsequent simulation part.

Figure 3.8 compares the average SER of the proposed method "Full-BSS" for different CM and MM-based BSS algorithms, namely ACMA, G-CMA, HG-CMA, G-MMA and HG-MMA, against the SNR. As noticed previously in Fig. 3.4, the ACMA has the worst performance. Whereas, the other methods have a comparable performance, especially in the case of 4-QAM.

Figure 3.9 depicts the performance of the "Equalization-BSS" class of algorithms. It is obvious that this class behaves very well at high SNR values, however it has a poor results at low and moderate SNR values. As expected, one can notice that the HG-MMA BSS algorithm has the best performance among others

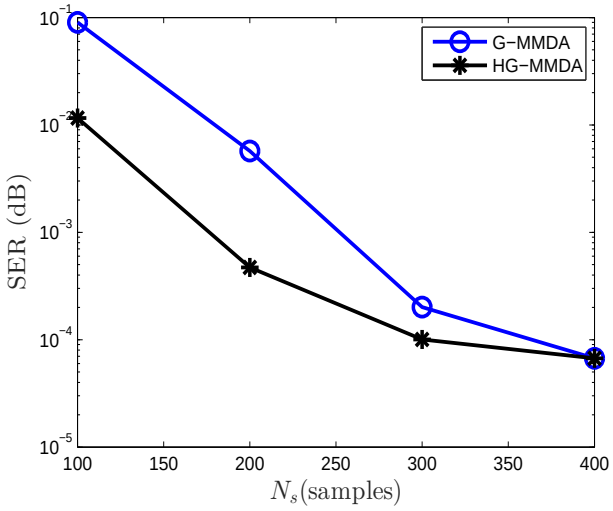


Figure 3.6: Average SER of G-MMDA and HG-MMDA a vs. N_s with $SNR = 30\ dB$ and 4-QAM.

with a tolerated increment in the computational cost.

Figure 3.10 shows the performance of the hybrid-based algorithms, namely G-MMDA and HG-MMDA. It is clear that this class of algorithms needs moderate number of samples to perform well, as already shown in Fig. 3.6, also they have an appealing results at all SNR levels. Moreover, we noticed that the HG-MMDA has an improved performance over the G-MMDA algorithm at high SNR. In the last experiment, we verify in Fig. 3.11 the deflated-based class, namely HG-DMMDA and G-DMMDA, for two different samples size, 200 and 350. It is clear that, both algorithms have an improved perormance at high N_s and it becomes more obvious at high SNR values. One may also notice that both algorithms experience a close performance in case of high N_s.

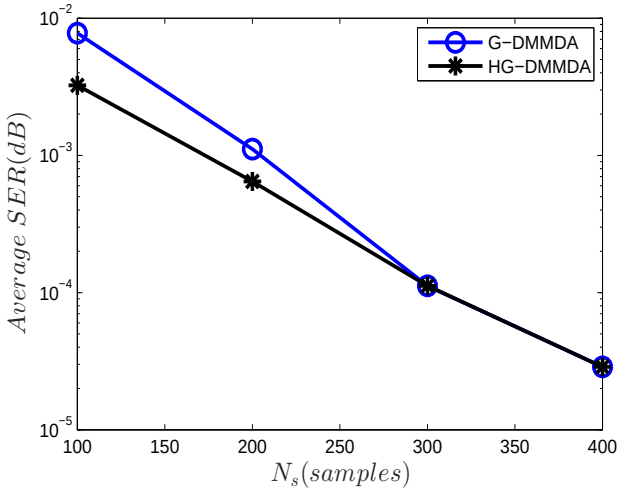


Figure 3.7: Average SER of G-DMMDA and HG-DMMDA a vs. N_s with $SNR = 30\ dB$ and 4-QAM.

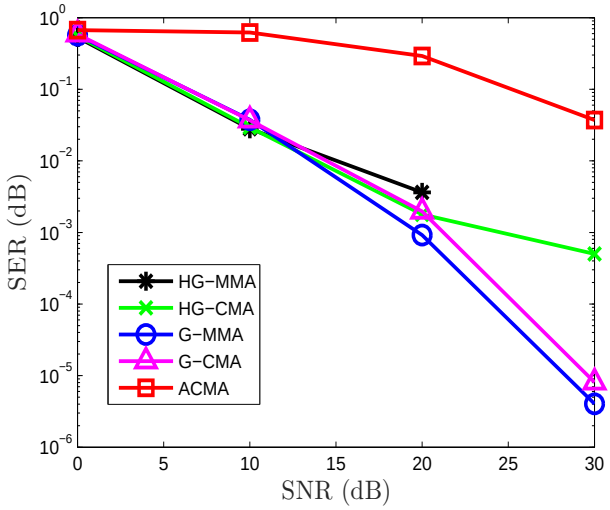


Figure 3.8: Average SER of Full-BSS algorithms (ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. SNR for $N_s = 250$ and 4-QAM.

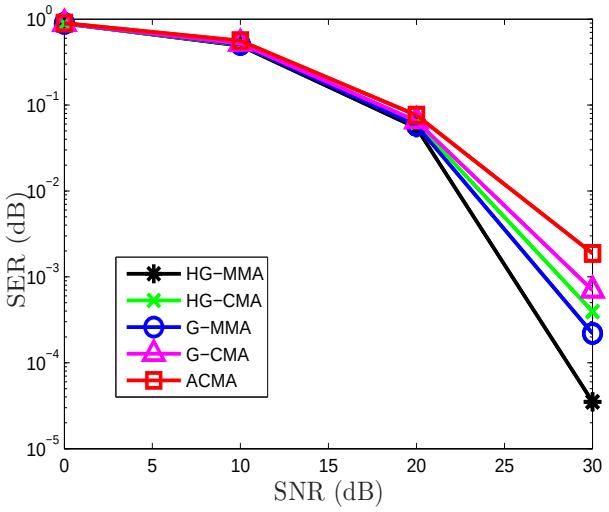


Figure 3.9: Average SER of Equalization followed by BSS (using ACMA, G-CMA, HG-CMA, G-MMA, HG-MMA) vs. SNR for $N_s = 120$ and 16-QAM.

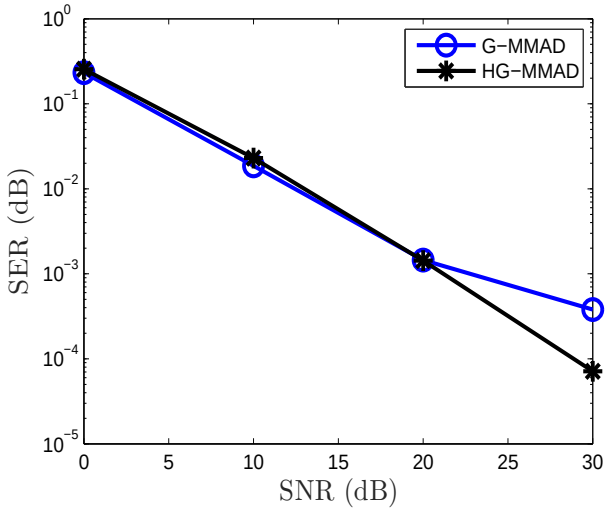


Figure 3.10: Average SER vs. SNR for the G-MMDA and HG-MMDA algorithms at $N_s = 300$ and 4-QAM.

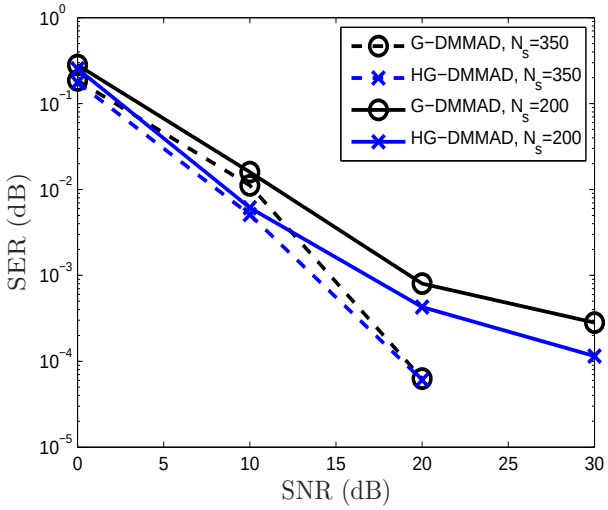


Figure 3.11: Average SER vs. SNR at both $N_s = 200$ and $N_s = 350$ and 4-QAM.

3.6 Conclusion

This work tackled one of the fundamental problems in wireless communications, more precisely, performing blind deconvolution and channel estimation for convolutive MIMO systems without the help of pilot data. In this work, for the first time, a class of MM-based algorithm combined with the SOS techniques implemented in different scenarios, to perform the blind deconvolution, are devised. Four different batch BD algorithms are presented; the first two, namely Full-BSS and Equalization-BSS, are classified as a two-step methods, while the last two, namely, Hybrid and deflation-based, are one step iterative approaches. The proposed algorithms are designed using a prewhitening operation to reduce the complexity of the optimization problem followed by applying special-versions of the unitary (Givens) and non-unitary (Shear) rotations on the whitened data. The proposed algorithms are mainly designed for the QAM signals. Finally, in addition to the computational cost comparison, simulation results are provided to highlight the favorable performance for different algorithms.

CHAPTER 4

FAST MULTI-MODULUS ALGORITHMS

A novel class of Fast Multi-Modulus (fastMMA) Blind Source Separation (BSS) and deconvolution algorithms are presented in this chapter. These are obtained through a fast fixed-point optimization rule used to minimize the Multi-Modulus (MM) criterion. In section 4.3, two BSS versions are provided to separate the sources either by finding the separation matrix at once or by separating a single source each time using a fast deflation technique. Further, the latter method is extended in section 4.4 to cover systems of convolutive nature. Interestingly, these algorithms are implicitly shown to belong to the fixed step-size gradient descent family, henceforth, an algebraic variable step-size is given in section 4.5 to make these algorithms even much faster. Apart from being computationally and performance-wise attractive, the new algorithms are free of any user-defined parameters.

4.1 Introduction

Blind Source Separation (BSS) stands for the process of retrieving N_t transmitted sources by processing N_r received signals without the need for a training phase. A prior statistical or structural information of the emitted sources is used instead.

Among the successful attempts to perform BSS for QAM signals is the Analytical Constant Modulus Algorithm (ACMA) [24], which provides an exact algebraic solution in the noise free case. It is able to solve the BSS in batch mode using few samples by solving a generalized eigenvalue problem. However, ACMA has a high computational complexity. Recently, CM and MM-based algorithms, named as Givens Constant Modulus Algorithm (G-CMA), Hyperbolic G-CMA (HG-CMA) [7], Givens Multi-Modulus Algorithm (G-MMA) and Hyperbolic G-MMA (HG-MMA) [74], have been introduced to overcome the numerical complexity of the ACMA algorithm. These algorithms use the unitary (Givens) and non-unitary (Shear) rotations to perform the BSS of the instantaneous mixtures.

In this chapter, the target is to solve both kind of mixtures, instantaneous and convolutive, in a fast and a less complex way, as compared to all existing algorithms. This is obtained through solving the MM criterion by the Fixed Point Optimization (FPO) numerical technique inspired by the famous fastICA method [35]. In the context of instantaneous mixtures, we provide a first separation algorithm which minimizes the MM criterion using FPO with a deflation approach. This algorithm is then modified in such a way all sources are extracted simultaneously (without the deflation) which improves the performance of the fastMMA.

Moreover, the previous deflation based fastMMA is extended to cover the convolutive mixtures using a simple correlation based criterion for the deflation. Finally, the proposed algorithms are shown to belong to the well-known family of gradient descent update rule, with a fixed step-size, henceforth, using the exact line search strategy, we derive an algebraic optimal step size, which increases the convergence rate of the proposed algorithms. Indeed, the variable step size (VSS) version of fastMMA is shown to converge in a very small number of iterations and sometimes in only one single iteration as illustrated in Section 5.5.

Unlike the MM stochastic Gradient Algorithm (SGA) such as MMA [27] and MUK [25], the proposed algorithms are shown to be free of user-defined parameters. Besides having fast convergence and reduced computational complexity, the proposed algorithms lead to improved estimation accuracy as illustrated in the simulation results section.

The rest of the chapter is organized as follow: Section 4.2 revise the problem formulation and explains the followed BSS approach[1]. Section 4.3 presents the two fastMMA versions which perform BSS for instantaneous mixtures. Section 4.4 extends fastMMA into fastMMDA, whereby the BSS and equalization (Deconvolution) for convolutive mixtures is enabled. The former algorithms are shown to belong to the standard step-size gradient-descent algorithms, hence in Section 4.5, we formulate a variable step procedure to make these algorithms even much faster. The performance advantages of the proposed algorithms are empowered in the simulation Section 5.5. Ultimately, Section 5.6 concludes this work.

[1]The problem formulation is already described in the preliminaries Chapter 2.

4.2 Problem Formulation: Revisited

Consider a MIMO system with N_t transmitters and N_r receiving antennas. The transmitted data is assumed to pass through a noisy frequency selective channel, so that the collected N_r samples at time instant k can be modeled as:

$$\boldsymbol{y}(k) = \sum_{n=0}^{M} \boldsymbol{H}(n)\boldsymbol{s}(k-n) + \boldsymbol{e}(k) \tag{4.1}$$

where $\boldsymbol{H}(n)$ is the channel impulse response, which captures the introduced Inter-Symbol Interference (ISI), defined as

$$\boldsymbol{H}(k) = \begin{bmatrix} h_{11}(k) & \dots & h_{1N_t}(k) \\ \vdots & \ddots & \vdots \\ h_{N_r1}(k) & \cdots & h_{N_rN_t}(k) \end{bmatrix}$$

and $\boldsymbol{H}(z) \stackrel{def}{=} \sum_{n=0}^{M} \boldsymbol{H}(n)z^{-n}$ is the unknown FIR transfer function, which is assumed to be irreducible and column reduced [80], $\boldsymbol{s}(k)$ is the unknown independent identically distributed phase modulated communication signal, and $\boldsymbol{e}(k)$ is the additive (spatially and temporally) white noise.

Considers the spatio-temporal variables:

$$\boldsymbol{y}_w(k) = \boldsymbol{\mathcal{H}}\boldsymbol{s}_w(k) + \boldsymbol{e}_w(k), \tag{4.2}$$

where

$$\boldsymbol{y}_w(k) = \left[\boldsymbol{y}^T(k), \cdots, \boldsymbol{y}^T(k - N_w + 1)\right]^T,$$

$$\boldsymbol{y}(k) = [y_1(k), \cdots, y_{N_r}(k)]^T,$$

$$\boldsymbol{s}_w(k) = \left[\boldsymbol{s}^T(k), \cdots, \boldsymbol{s}^T(k - N_w - M + 1)\right]^T,$$

$$\boldsymbol{s}(k) = [s_1(k), \cdots, s_{N_t}(k)]^T,$$

$\boldsymbol{\mathcal{H}}$ is an $(N_r N_w \times N_t(N_w + M))$ block Toeplitz convolution matrix with the first block row $[\boldsymbol{H}(0), \cdots, \boldsymbol{H}(M), \boldsymbol{0}_{Nr\times Nt}, \cdots, \boldsymbol{0}_{Nr\times Nt}]$. As for $\boldsymbol{\mathcal{H}}$ to be identified and of full column rank, the condition $N_w \geq (N_t M)/(N_r - N_t)$ must hold.

The source signals are recovered blindly using the $N_t(N_w + M) \times N_r N_w$ beamformer matrix $\boldsymbol{W}$. Hence, the receiver output is evaluated as $\boldsymbol{z}(k) = \boldsymbol{W}\boldsymbol{y}_w(k)$ which is an $N_d = N_t(N_w + M)$ dimensional estimated source signal with some delayed versions. In batch Blind Deconvolution (BD) algorithms, N_s samples of the received data are usually collected before processing, so that K vectors $\boldsymbol{y}_w$ can be formed, where $K = N_s - N_w + 1$, and concatenated in a matrix $\boldsymbol{Y} = [\boldsymbol{y}_w(N_w), \cdots, \boldsymbol{y}_w(N_s)]$. Similarly, the estimated sources can be written in matrix form as $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{Y}$.

To simplify the problem, a prewhitening operation [7], [23] is used so that it reduces the dimension of $\boldsymbol{Y}$ from $(N_r N_w \times K)$ to $(N_d \times K)$ and it transforms the channel matrix $\boldsymbol{\mathcal{H}}$ into a matrix that is close to unitary (see Section 2.4.1), according to $\underline{\boldsymbol{Y}} = \boldsymbol{B}\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{\mathcal{H}}\boldsymbol{S} = \boldsymbol{V}\boldsymbol{S}$ in the noiseless case. $\boldsymbol{B}$ is the $(N_d \times N_r N_w)$ whitening (filtering) matrix, $\boldsymbol{V} = \boldsymbol{B}\boldsymbol{\mathcal{H}}$ is a $(N_d \times N_d)$ unitary matrix[2]. This kind

[2]This is an ideal case which is satisfied with a large number of samples.

of manipulation reforms the problem of BD into finding the unitary matrix $\boldsymbol{V}$ in a way that the beamformer matrix becomes $\boldsymbol{W} = \boldsymbol{V}^H \boldsymbol{B}$, and ultimately results into the following:

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{Y} = \boldsymbol{V}^H \boldsymbol{B}\boldsymbol{Y} = \boldsymbol{V}^H \boldsymbol{V}\boldsymbol{S} = \boldsymbol{S} \tag{4.3}$$

By reducing the Blind deconvolution problem to finding the unitary matrix $\boldsymbol{V}$, we aim to use the efficient FPO technique to build $\boldsymbol{V}$ in (4.3), and hence be able to recover the source signals. Making use of the QAM signals' constellation structure, one can optimize the MM criterion [26]:

$$\mathcal{J}(\boldsymbol{V}) = \sum_{i=1}^{N_d} E\Big[\big(z_{i,R}^2(k) - R_R\big)^2 + \big(z_{i,I}^2(k) - R_I\big)^2\Big] \tag{4.4}$$

where $R_R = E[s_R^4(k)]/E[s_R^2(k)]$ and $R_I = E[s_I^4(k)]/E[s_I^2(k)]$ are, respectively, the real and imaginary dispersion constants[3]. $z_i(k)$ is the (i,k)-th element of $\boldsymbol{Z}$.

4.3 Fast MMA for Instantaneous Mixture

In this section, we focus on solving the BSS problem for memoryless systems. Hence, substituting $M = 0$ and accordingly $N_w = 1$ in (4.2) reduces the convolutive model given in (4.1) into an instantaneous one. After performing the whitening phase, we aim to find the unitary matrix $\boldsymbol{V}$ such that $\boldsymbol{Z} = \boldsymbol{V}^H \underline{\mathbf{Y}}$. Each column of $\boldsymbol{V}$, i.e., $\boldsymbol{v}_i$ extracts one source signal, hence we need to find

[3]W.l.o.g., we assume a square QAM constellation so that $R_R = R_I = R$.

$\boldsymbol{v}_i = \boldsymbol{v}_{i,R} + j\boldsymbol{v}_{i,I}$ such that the i-th source signal $\boldsymbol{z}_i = \boldsymbol{v}_i^H \underline{\mathbf{Y}}$ is obtained. To do this, define

$$\dot{\boldsymbol{v}}_i = \begin{bmatrix} \boldsymbol{v}_{i,R} \\ \\ \boldsymbol{v}_{i,I} \end{bmatrix}, \quad \boldsymbol{F}_I = \begin{bmatrix} \mathbf{0} & \boldsymbol{I} \\ \\ -\boldsymbol{I} & \mathbf{0} \end{bmatrix}, \quad \underline{\dot{\boldsymbol{Y}}} = \begin{bmatrix} \underline{\boldsymbol{Y}}_R \\ \\ \underline{\boldsymbol{Y}}_I \end{bmatrix} \tag{4.5}$$

Then, any part of the MM criterion in (4.4) becomes

$$\begin{aligned} E\left[\left(z_{i,x}^2(t) - R\right)^2\right] &= E\left[\left(\left(\dot{\boldsymbol{v}}_i^T \boldsymbol{F}_X^T \underline{\dot{\mathbf{Y}}}\right)^2 - R\right)^2\right] \\ &= E\left[\left(\dot{\boldsymbol{v}}_i^T \boldsymbol{F}_X^T \underline{\dot{\mathbf{Y}}}\right)^4\right] - 2E\left[\left(\dot{\boldsymbol{v}}_i^T \boldsymbol{F}_X^T \underline{\dot{\mathbf{Y}}}\right)^2\right] R + R^2 \end{aligned} \tag{4.6}$$

where $\boldsymbol{F}_X$ is a fixed matrix defined as

$$\boldsymbol{F}_X = \begin{cases} \boldsymbol{I}\,, & \text{when evaluating } z_{i,R} \\ \\ \boldsymbol{F}_I, & \text{when evaluating } z_{i,I} \end{cases}$$

After dropping the irrelevant terms and due to the whitening step, (4.6) becomes

$$E\left[\left(z_{i,x}^2(t) - R\right)^2\right] = E\left[\left(\dot{\boldsymbol{v}}_i^T \boldsymbol{F}_X^T \underline{\dot{\mathbf{Y}}}\right)^4\right] - 2R\left\|\dot{\boldsymbol{v}}_i\right\|_2^2 \tag{4.7}$$

To avoid the trivial solution $\boldsymbol{v}_i = 0$, the constraint $||\boldsymbol{v}_i||_2^2 = 1$ is introduced. Ultimately, the final objective function is

$$\mathcal{J}(\boldsymbol{v}_i, \lambda) = E\left[\left(\dot{\boldsymbol{v}}_i^T \underline{\dot{\mathbf{Y}}}\right)^4\right] + E\left[\left(\dot{\boldsymbol{v}}_i^T \boldsymbol{F}_I^T \underline{\dot{\mathbf{Y}}}\right)^4\right] - 4R\left\|\dot{\boldsymbol{v}}_i\right\|_2^2 + \lambda\left(1 - \left\|\dot{\boldsymbol{v}}_i\right\|_2^2\right) \tag{4.8}$$

where λ is the Lagrange parameter. To guarantee a fast convergence with an easy implementation, we propose to use the FPO numerical method. To do this, one needs to put the update equation in the format such that $\dot{\boldsymbol{v}}_i = f(\dot{\boldsymbol{v}}_i)$. Hence, the fixed point $\dot{\boldsymbol{v}}_i$ of (4.8) is obtained by zeroing the gradient vector. This results into the following iterative rule (for simplicity, we adopt here Matlab notation to avoid to rename the updated vector after each iteration):

$$\frac{\lambda}{2}\dot{\boldsymbol{v}}_i = E\left[\underline{\dot{\mathbf{Y}}} \circ \boldsymbol{e}_{N_t}\boldsymbol{a}_R^3\right] + E\left[\boldsymbol{F}_I^T\underline{\dot{\mathbf{Y}}} \circ \boldsymbol{e}_{N_t}\boldsymbol{a}_I^3\right] - 2R\dot{\boldsymbol{v}}_i \tag{4.9}$$

where $\circ$ denotes the Hadamard product, $\boldsymbol{a}_R = \boldsymbol{v}_i^T\underline{\mathbf{Y}}$, $\boldsymbol{a}_I = \boldsymbol{v}_i^T\boldsymbol{F}_I^T\underline{\mathbf{Y}}$, $\boldsymbol{e}_{N_t}$ is a column vector of size $2N_t$ filled with ones, and $\boldsymbol{a}_R^3$ refers to $(\boldsymbol{a}_R \circ \boldsymbol{a}_R \circ \boldsymbol{a}_R)$. $E[.]$ refers here to the averaging of the column vectors of its argument. After each update, λ is chosen so that $\|\dot{\boldsymbol{v}}_i\| = 1$. The significant gain of this method is that with a minimal computational cost, only a few number of iterations are needed to converge, usually 5 to 15 for small or moderate number of sources, as will be shown in the simulation section. This learning rule separates only one single source. To recover the N_t sources, two schemes are possible and are presented in the subsequent subsections.

4.3.1 Deflation Based Estimation of $\boldsymbol{V}$ (fastMMAd)

To retrieve the N_t independent components, we run (4.9) $N_t - 1$ times. To ensure that each time we extract a new different source, we install an orthogonal pro-

jection to remove the extracted source and simultaneously reduce the size of the problem, i.e., we reduce by 1 the number of sources at each run. To achieve this, let's decompose the separation vector according to $\boldsymbol{v}_i = [\tilde{\boldsymbol{v}}_i^T, \alpha]^T$ where $\alpha \geq 0$ is the last entry[4] of $\boldsymbol{v}_i$. Now, the orthogonal projection matrix $\boldsymbol{P}_i$ can be expressed as:

$$\boldsymbol{P}_i = \left[\begin{array}{c|c} \boldsymbol{I} - \frac{\tilde{\boldsymbol{v}}_i \tilde{\boldsymbol{v}}_i^H}{1+\alpha} & -\tilde{\boldsymbol{v}}_i \end{array} \right] \tag{4.10}$$

It is easy to check that $\boldsymbol{P}_i \boldsymbol{v}_i = 0$, $\boldsymbol{P}_i \boldsymbol{P}_i^H = \boldsymbol{I}$ and that $\boldsymbol{P}_i \boldsymbol{Y}_{i-1}$ reduces by one the number of rows of $\boldsymbol{Y}_{i-1}$ according to:

$$\boldsymbol{P}_i \boldsymbol{Y}_{i-1} = \tilde{\boldsymbol{Y}}_{i-1} - \tilde{\boldsymbol{v}}_i (\frac{\tilde{\boldsymbol{v}}_i^H}{1+\alpha} \tilde{\boldsymbol{Y}}_{i-1} + \tilde{\boldsymbol{y}}_{i-1})$$

where $\boldsymbol{Y}_{i-1}$ is the projected data matrix at the $(i-1)$-th run (starting with $\boldsymbol{Y}_0 = \underline{\mathbf{Y}}$) which is decomposed as: $\boldsymbol{Y}_{i-1} = [\tilde{\boldsymbol{Y}}_{i-1}^T, \tilde{\boldsymbol{y}}_{i-1}^T]^T$, where $\tilde{\boldsymbol{y}}_{i-1}$ being the last row vector of $\boldsymbol{Y}_{i-1}$. This results into the deflated fastMMA algorithm which is summarized in Table 4.1.

4.3.2 Full Estimation of $\boldsymbol{V}$ (fastMMAf)

Another possible scheme will be by iterating the up-dating of the separation matrix $\boldsymbol{V}$ such that all the N_t independent sources are separated simultaneously in each

[4] If θ is the phase argument of the last entry of $\boldsymbol{v}_i$, the latter vector is multiplied by $e^{-j\theta}$ so that its last component α becomes non-negative valued. This choice is made to guarantee that the term $1 + \alpha$ is non-zero.

Table 4.1: fastMMAd Algorithm

Pre-Whitening: $\boldsymbol{Y}_0 = \boldsymbol{B}\boldsymbol{Y}$	$\mathcal{O}(N_s N_r^2 + N_r^3)$
for $i = 1 : N_t - 1$ **do**	
1. Initialize $\boldsymbol{v}_i$ as a random vector of norm one.	
2. Build the real arrays: $\dot{\boldsymbol{Y}}_0$, $\dot{\boldsymbol{v}}_i$.	
for $n = 1 : N_{Sweeps}$ **do**	
$\boldsymbol{a}_R = \dot{\boldsymbol{v}}_i^T \dot{\boldsymbol{Y}}_{i-1}$	$2N_s(N_t - i + 1)$
$\boldsymbol{a}_I = \dot{\boldsymbol{v}}_i^T \boldsymbol{F}_I^T \dot{\boldsymbol{Y}}_{i-1}$	$2N_s(N_t - i + 1)$
$\dot{\boldsymbol{v}}_i = E\left[\dot{\boldsymbol{Y}}_{i-1} \circ \boldsymbol{e}_{N_t} \boldsymbol{a}_R^3\right] + E\left[\boldsymbol{F}_I^T \dot{\boldsymbol{Y}}_{i-1} \circ \boldsymbol{e}_{N_t} \boldsymbol{a}_I^3\right] - 2R\dot{\boldsymbol{v}}_i$	
$\dot{\boldsymbol{v}}_i = \dot{\boldsymbol{v}}_i / \left\|\dot{\boldsymbol{v}}_i\right\|^2$	$2N_s(N_t - i + 1)$
end for	
3. Extract the source number i: $\boldsymbol{z}_i = \boldsymbol{v}_i^H \boldsymbol{Y}_{i-1}$	$2N_s(N_t - i + 1)$
4. Remove its effect: $\boldsymbol{Y}_i = \boldsymbol{P}_i \boldsymbol{Y}_{i-1}$	$4N_s(N_t - i + 1)$
end for	

sweep. This implementation allows us to relax the orthogonality constraint on the matrix $\boldsymbol{V}$. In doing so, we avoid the orthogonal projection step and the restricted search of $\boldsymbol{V}$ among the unitary matrix set only[5]. Consequently, this relaxed constraint leads to a slight improvement in the estimation performance. The complete implementation is summarized in Table 4.2.

4.4 fastMMDA for Convolutive Mixture

To deal with convolutive channel systems, i.e., mitigate both ISI and inter-user interferences, we extend the attracting deflated-based fastMMAd algorithm and propose the fastMMDA algorithm (D stands for deconvolution).

Due to the channel memory and the windowing effect in the convolutive model (4.2), each single source is expected to have $N_\delta = N_w + M$ number of replicas. We

[5]If the pre-whitening step is inadequate (due for example to a short sample size) the unitary constraint might affect significantly the algorithm's performance.

Table 4.2: fastMMAf Algorithm

Pre-Whitening: $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{Y}$ $\mathcal{O}(N_s N_r^2 + N_r^3)$
1. Initialize $\boldsymbol{V}$ as a random unitary matrix.
2. Build the real arrays: $\dot{\boldsymbol{Y}}$, $\dot{\boldsymbol{V}}$.
for $n = 1 : N_{Sweeps}$ **do**
  for $i = 1 : N_t$ **do**
    $\boldsymbol{a}_{i,R} = \dot{\boldsymbol{v}}_i^T \dot{\boldsymbol{Y}}$ $(2N_s N_t)$
    $\boldsymbol{a}_{i,I} = \dot{\boldsymbol{v}}_i^T \boldsymbol{F}_I^T \dot{\boldsymbol{Y}}$ $(2N_s N_t)$
    $\dot{\boldsymbol{v}}_i = E\left[\dot{\boldsymbol{Y}} \circ \boldsymbol{e}_{N_t} \boldsymbol{a}_{i,R}^3\right] + E\left[\boldsymbol{F}_I^T \dot{\boldsymbol{Y}} \circ \boldsymbol{e}_{N_t} \boldsymbol{a}_{i,I}^3\right] - 2R\dot{\boldsymbol{v}}_i$ $(4N_s N_t)$
  end for
  Normalize the separation matrix: $\dot{\boldsymbol{V}} = \dot{\boldsymbol{V}} / ||\dot{\boldsymbol{V}}||^2$
end for
3. Recover all the transmitted sources: $\boldsymbol{Z} = \boldsymbol{V}^H \underline{\mathbf{Y}}$

---

use the orthogonalization and dimension reduction strategies to remove the effect of the source and its delayed copies after each deflation run. To this end, define the time-lagged correlation matrix between the extracted source $\boldsymbol{z}_i = \boldsymbol{v}_i^H \boldsymbol{Y}_{i-1}$ and the collected data set $\boldsymbol{Y}_{i-1}$ at the $(i-1)$-th deflation run (where i refers to the order of the extracted source) as

$$\boldsymbol{R}_i = \left[E\left[\boldsymbol{y}_{i-1}(k) z_i^*(k + N_\delta)\right], \cdots, E\left[\boldsymbol{y}_{i-1}(k) z_i^*(k - N_\delta)\right)\right]$$

where $\boldsymbol{y}_{i-1}(k)$ is the k-th column vector of matrix $\boldsymbol{Y}_{i-1}$. The correlation results ideally are either zero or the column vector of the mixing matrix of $\boldsymbol{Y}_{i-1}$ corresponding to one of the shifted version of the i-th extracted source z_i. Hence, the matrix $\boldsymbol{R}_i$ would have ideally N_δ non zero column vectors. In practice, the latter are obtained by selecting the N_δ column vectors of $\boldsymbol{R}_i$ of maximum norm. After normalization, these vectors form the columns of $\boldsymbol{Q}_i$, which is ideally unitary

thanks to the prewhitening step.

Now, to remove the extracted i-th source and its shifted versions, one uses the projection matrix $\boldsymbol{P}_i$. To do this, split $\boldsymbol{Q}_i$ into $\left[\begin{array}{c|c} \boldsymbol{Q}_{1,i}^T & \boldsymbol{Q}_{2,i}^T \end{array}\right]^T$, where the square matrix $\boldsymbol{Q}_{2,i}$ has the following polar decomposition: $\boldsymbol{Q}_{2,i} = \boldsymbol{\Omega}\boldsymbol{\Theta}$; $\boldsymbol{\Omega} \geq 0, \boldsymbol{\Theta}$ is unitary. Eventually, $\boldsymbol{P}_i$ in (4.10) will look as follows:

$$\boldsymbol{P}_i = \left[\begin{array}{c|c} \boldsymbol{I}_d - \tilde{\boldsymbol{Q}}_{1,i} \left(\boldsymbol{I} + \boldsymbol{\Omega}\right)^{-1} \tilde{\boldsymbol{Q}}_{1,i}^H & -\tilde{\boldsymbol{Q}}_{1,i} \end{array}\right]$$

where $\tilde{\boldsymbol{Q}}_{1,i} = \boldsymbol{Q}_{1,i}\boldsymbol{\Theta}^H$. Similarly to the instantaneous case, $\boldsymbol{P}_i$ achieves both orthogonal projection and dimension reduction in a fast way.

Remarks: In the convolutive case, the number of runs is N_t instead of $N_t - 1$ (as for the instantaneous case), because one needs to remove the ISI for the last step. Also, the extension of fastMMAf is not proposed here due to its high computational cost in the convolutive context. Finally, note that the full version (i.e. fastMMAf) might require several initialization trials to guarantee the algorithm's convergence.

4.5 MMA with optimized Step Size

One can read the update in equation (4.9) as[6]:

$$\dot{\boldsymbol{v}}_i = \dot{\boldsymbol{v}}_i - \frac{1}{8}\left\{4E\left[\underline{\dot{\mathbf{Y}}} \circ \boldsymbol{e}_{N_t}\boldsymbol{a}_R^3\right] + 4E\left[\boldsymbol{F}_I^T\underline{\dot{\mathbf{Y}}} \circ \boldsymbol{e}_{N_t}\boldsymbol{a}_I^3\right]\right\} \tag{4.11}$$

[6]One can fix R to 1, as it can be tolerated due to the inherent scale ambiguity in the BSS techniques. Equation (4.11) is obtained by factorizing the constant $-2R$ and including it in the normalization constant λ.

followed by a normalization step. Interestingly, similar to the work presented in [103] and [36], the former update rule is argued to be a gradient-descent based, with a fixed step size (FSS) $\mu = 1/8$, of the form

$$(\dot{\boldsymbol{v}}_i)_{new} = (\dot{\boldsymbol{v}}_i)_{old} - \mu \boldsymbol{g}_i \tag{4.12}$$

where $\boldsymbol{g}_i = \nabla_{\dot{\boldsymbol{v}}_i} \mathcal{M}(\dot{\boldsymbol{v}}_i)$ and $\mathcal{M}(\dot{\boldsymbol{v}}_i) = E[|z_{i,R}|^4] + E[|z_{i,I}|^4]$ is a sum of forth-order moments of the real and imaginary parts of the reconstructed source signals.

The FPO rule has shown to be a fast technique, in the sense that it maintains a fixed step size which has a cubic speed of convergence in the case of infinite number of samples [104]. Still, the fast convergence is not guaranteed as it might slow down and even get trapped in a local minima, especially in the short sample size case, as is shown in [105]. Recently, the exact line search optimization technique has been successfully deployed in channel identification [106], blind equalization [103] and independent component analysis [36], to provide an optimal variable step size (VSS), which has shown to speed up the convergence and improves the algorithm's robustness against local minima convergence. With this rational, faster versions of the earlier proposed algorithms can be obtained via the following step size optimization approach defined in (4.13). This is achieved by appropriately enrolling the exact line search strategy, i.e,[7]

$$\mu_{opt} = \arg \min_{\mu} \mathcal{J} \left(\frac{(\dot{\boldsymbol{v}})_{new}}{\| (\dot{\boldsymbol{v}})_{new} \|} \right) \tag{4.13}$$

[7]We dropped the index i intentionally for ease of presentation.

where $(\dot{\boldsymbol{v}})_{new}$ is given by (4.12) and $\mathcal{J}$ is the MM cost function. The search direction is typically in the negative direction of the gradient $\boldsymbol{g}_i$ in (4.12).

In addition to some other limitations [100], exact line search is known to suffer from high computational complexity, and this is maybe the reason for lack of its extensive use in practice [36]. Fortunately, some criteria, such as the constant modulus, the kurtosis, the constant power and multi-modulus, can be expressed as a polynomial or a rational function of μ. Hence, the algebraic optimal step size μ_{opt} can be easily determined by finding the roots of these functions. The fast MMA algorithms can be simply extended and become much faster by finding the optimal step size. This is done by performing the following tasks at each iteration:

1. Evaluate the coefficient $\{n_i\}_{i=0}^{4}$ of the 4-th order polynomial given by:

$$\mathcal{N}(\mu) = \sum_{i=0}^{4} n_i \mu^i \tag{4.14}$$

   The coefficients' expressions (given in Appendix D) are defined in terms of the whitened received signal's block $\underline{\dot{\mathbf{Y}}}$ and the current values of $\dot{\boldsymbol{v}}$ and $\boldsymbol{g}$. This polynomial is obtained via the derivation w.r.t. μ of the cost function in (4.13).

2. Find the four roots of the polynomial in (4.14). Interestingly, this step can be performed at an insignificant cost, compared to step 1, using for example standard algebraic procedures such as Ferrari's formula [100].

3. Select the VSS root among the real-valued candidates obtained in 2, i.e.,

the root which leads to the minimum of the MM cost function in (4.13).

4. Update (4.12) and normalize the vector $(\dot{\boldsymbol{v}})_{new}$.

It is noteworthy to mention that the above mentioned tasks are described for the deflated MMA version, i.e., fastMMAd, yet a similar procedure can be performed to find the optimal VSS for both fastMMAf and fastMMDA. Also, besides improving the convergence speed, the VSS is able to bestow the methods and enhance the robustness to the initialization and local minima problem [103, 36].

4.6 Computational Complexity

As discussed earlier, the new fast algorithms will be compared to the state of the art algorithms in the next simulation section. However, to make a fair comparison, this section provides some insight into the numerical cost of the proposed algorithms too. It is compared to the CMA and MMA-like BSS algorithms in terms of the number of flops, i.e., a real addition and a real multiplication, per sweep in Table 4.3.

Table 4.3: Numerical Cost of Different BSS Algorithms

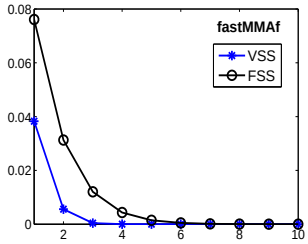
BSS Algorithm	**Complexity Order**	**VSS Extra Cost**
fastMMAf	$8N_sN_t^2 + \mathcal{O}(N_sN_t)$	$4N_sN_t^2 + \mathcal{O}(N_sN_t)$
fastMMAd	$6N_sN_t^2 + \mathcal{O}(N_sN_t)$	$2N_sN_t^2 + \mathcal{O}(N_sN_t)$
G-MMA [74]	$20N_sN_t^2 + \mathcal{O}(N_sN_t)$	-
HG-MMA [74]	$40N_sN_t^2 + \mathcal{O}(N_sN_t)$	-
ACMA [24]	$\mathcal{O}(N_sN_t^4)$	-

4.7 Simulation Results

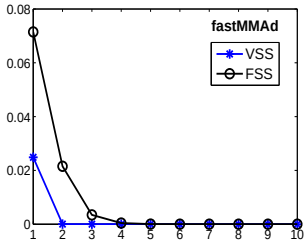
In this section, we examine the performance of the proposed algorithms. This is done by considering different simulation scenarios and comparing the achieved gain to the state of the art algorithms, such as ACMA, G-MMA, HG-MMA and other CMA like algorithms. The results are averaged over 1000 Monte-Carlo trials, whereby the channel is randomly generated at each run.

Figure 4.1 gives snap-shots for the proposed algorithm's speed of convergence by plotting the cost function, for both FSS and VSS versions, against the number of sweeps for different N_t and N_r. In this figure, parts (a) and (c) reveal that the fastMMAf algorithm, with the VSS, converges earlier than the FSS case with a reduction of $3-4$ iterations for small and intermediate number of transmitters. Meanwhile, parts (b) and (d) show that fastMMAd algorithm converges much faster than the fastMMAf for both VSS and FSS versions. Hence, it is clear that the proposed algorithms converge fast with a small number of iterations, ≤ 10 in the given examples, which varies with the system dimension and the initial value of matrix $\boldsymbol{V}$.

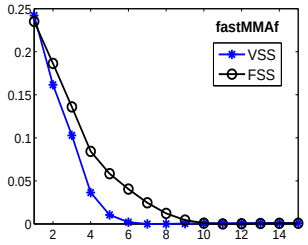
Similar setup has been conducted at low SNR values and similar conclusions are drawn. However, in Fig. 4.2, we examine the speed of convergence, once again, for 16-QAM, a non constant modulus modulation. The optimal VSS scenario, in both fastMMAf and fastMMAd, still shows to be faster with a slight increment, comparing to Fig. 4.1 (c) and (d), in the number of iterations due to the higher constellation size.



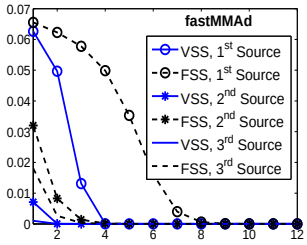
(a) $N_t = 2, N_r = 4, N_s = 50$



(b) $N_t = 2, N_r = 4, N_s = 50$

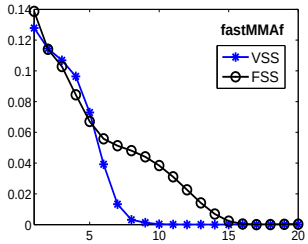


(c) $N_t = 5, N_r = 7, N_s = 150$

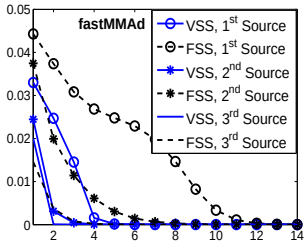


(d) $N_t = 5, N_r = 7, N_s = 150$

Figure 4.1: Speed of Convergence ($\mathcal{J}$ vs. N_{Sweeps}), for VSS and FSS, at SNR=30 dB, 4-QAM case.



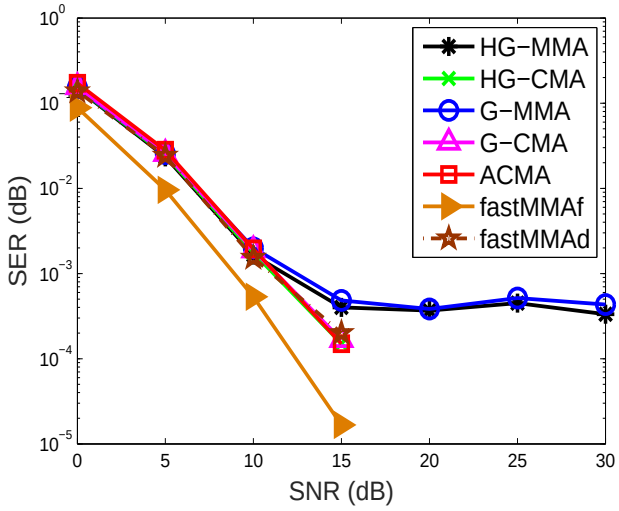
(a) fastMMAf



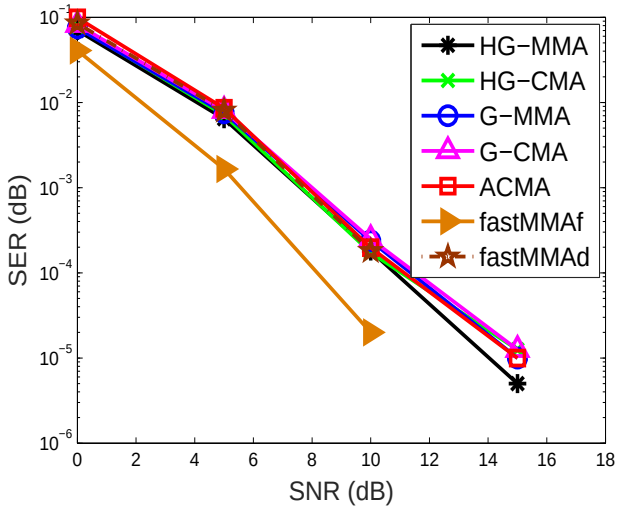
(b) fastMMAd

Figure 4.2: Speed of Convergence ($\mathcal{J}$ vs. N_{Sweeps}), for VSS and FSS, at SNR=30 dB, $N_t = 5, N_r = 7, N_s = 200$, 16-QAM case.

Figure 4.3 evaluates the SER vs. SNR for 4-QAM case at different setups. It is shown that the proposed fastMMAd has a comparable performance in comparison to other methods, yet fastMMAf has the best performance results among the rest. Also, as expected, fastMMAf has a superior performance over the fastMMAd, since the former is free to search the desired separation matrix in a wider space.



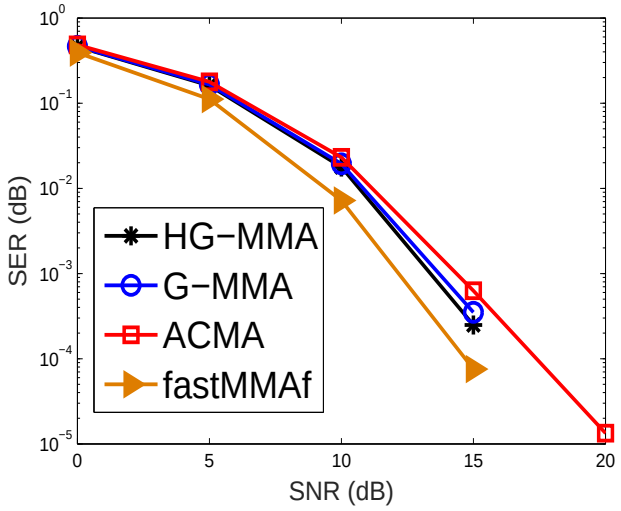
(a) $N_t = 2, N_r = 4, N_s = 30$



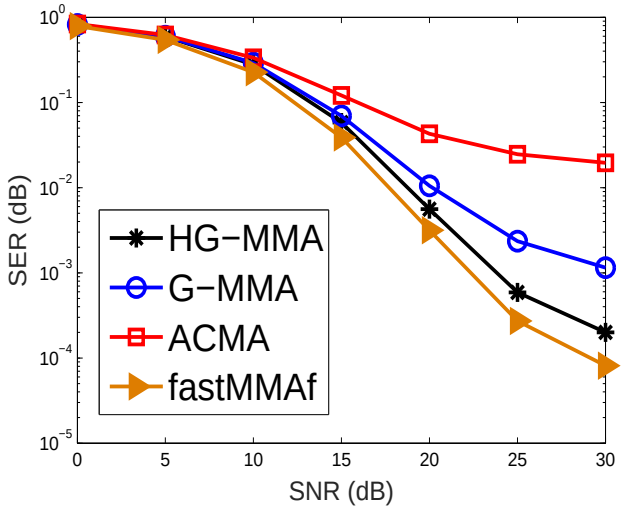
(b) $N_t = 4, N_r = 7, N_s = 100$

Figure 4.3: Average SER for different CMA and MMA, for different scenarios, 4-QAM case.

Figure 4.4 compares the SER of the considered algorithms for high order constellations. The experiment setup is $(N_t = 3, N_r = 7, N_s = 300)$. Once more, without the performance limitations imposed by the orthogonalizaiton, fastMMAf proves out-performance to other counterparts even for non-constant modulus signals.



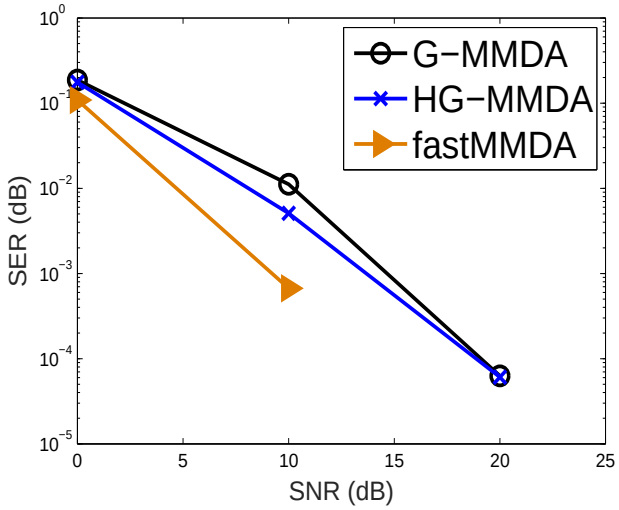
(a) 16-QAM



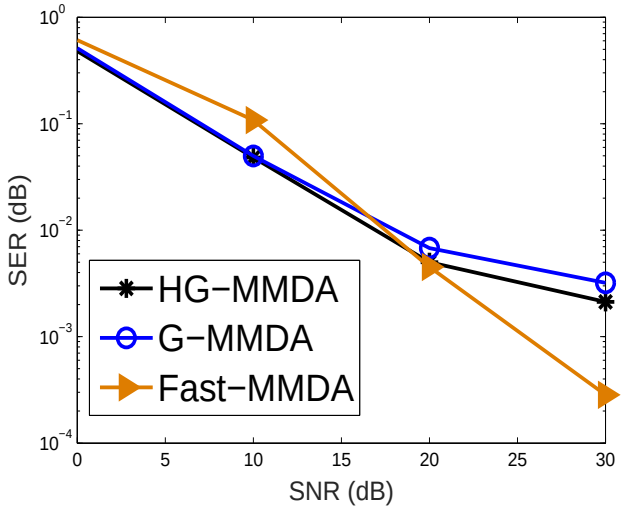
(b) 64-QAM

Figure 4.4: Average SER for different CMA and MMA at different constellation setups, $N_t = 3, N_r = 7, N_s = 300$.

The last experiment in Fig. 4.5 deals with the convolutive mixtures and compares the proposed fastMMDA with the newly proposed extensions of the G-MMA and HG-MMA, namely G-MMDA and HG-MMDA [102]. Again, our algorithm leads to the best performance at a lower computational cost.



(a) $N_s = 350$, 4-QAM



(b) $N_s = 700$, 16-QAM

Figure 4.5: Average SER for HG-MMDA, G-MMDA, and fastMMDA at 4-QAM and 16-QAM, $N_t = 3, N_r = 5, M = 2$.

4.8 Conclusion

This chapter develops a new set of fast and low cost BSS algorithms dedicated to phase modulated communication signals. Both instantaneous and convolutive mixture cases are considered. Similarly to the well-known fastICA algorithm, our methods rely on the efficient FPO technique used in conjunction with a fast deflation approach. In addition, we proposed a new version (fastMMAf) where the unitary constraint is relaxed in order to allow a wider space exploration and hence an improved separation quality. Furthermore, aiming at boosting the convergence speed further, a VSS procedure is provided using the well-known exact line search strategy. Surprisingly, our algorithms, not only have the lowest numerical cost, but also have the best performance among all standing MMA or CMA type methods.

Part II

Blind System Identification

This part composes three Chapters, which mainly addresses the BSI problem of SIMO systems in different perspectives including methods and performance evaluations. Two contributions are presented in this part,

I. Chapter 5 proposes a novel subspace method which demonstrates an appealing performance, especially in the poor spatial diversity (ill-conditioned) scenario.

II. By means of the CRB, Chapter 6 investigates the possible performance's gain by incorporating some a priori (freely accessed) side information, such as the channel's sparsity and some other statistical information such as the signals' non-circularity, on the BSI problem. Also, evaluates the impact of the semi-blind approach on the system overestimation problem.

Eventually, Chapter 7 concludes all the work covered in this dissertation and accordingly motivate some possible future work.

CHAPTER 5

STRUCTURE-BASED SUBSPACE SYSTEM IDENTIFICATION

In this chapter, a novel subspace-based method for blind identification of multichannel finite impulse response (FIR) systems is presented. Here, we exploit directly the block Toeplitz channel's structure in the signal's linear model to build a quadratic cost function whose minimization leads to the desired channel estimation up to a scalar factor. This method can be extended to estimate any predefined linear structure, e.g. Hankel, that is usually encountered in linear systems. Simulation findings are provided to highlight the appealing advantages of the new structure-based subspace (SSS) method over the standard subspace (SS) method in certain adverse identification scenarios.

5.1 Introduction

Blind system identification (BSI) is one of the fundamental signal processing problems that was initiated more than three decades ago. BSI refers to the process of retrieving the channel's impulse response based on the output sequence only. As it has so different applications, such as mobile communication, seismic exploration, image restoration, medical, and other applications, it has drawn researchers' attention and resulted in a plethora of methods. Since then, a class of subspace-based methods dedicated to BSI has been developed, among them the standard subspace method (SS) [45, 46], the cross-relation (CR) method [47], the two-step maximum likelihood (TSML) method [48], and the truncated transfer matrix (TTM) method [107], which all are second order statistics (SOS) based methods. According to the comparative studies carried out in [108] and [109], the SS method is claimed to be the most powerful one.

It is noteworthy to argue that multi-channel FIR systems encountered in certain applications could suffer from an ill-conditioned scenario [42, 43]. This might happen in the case of low receiver diversity, or when the channel response is of sparse or small tails [44]. A shared weakness of most of SOS methods is that their performance is poor in the ill-conditioned case.

In this work, a novel subspace-based method relying on the channel's Toeplitz structure, which is employed directly to formulate our cost function, is proposed. This is of great interest since the Toeplitz structure is an inherent nature that exists in most of the linear systems due to their convolutive nature. The proposed

method's gain is demonstrated with a particular emphasis on the ill-conditioned channels.

The chapter presentation focuses first on the development of the proposed structure-subspace (SSS) method. Then, the improvement obtained by the SSS method over SS, in the case of channels with closely spaced roots, is highlighted. Although, the SSS method sounds to be a promising technique, yet it has a higher computational complexity that needs to be addressed in a future work.

Notation: The invertible column vector-matrix mappings are denoted by $\text{vec}\{.\} : \mathbb{C}^{a\times b} \rightarrow \mathbb{C}^{ab\times 1}$ and $\text{mat}_{a,b}\{.\} : \mathbb{C}^{ab\times 1} \rightarrow \mathbb{C}^{a\times b}$. $(\boldsymbol{A} \otimes \boldsymbol{B})$ is the Kronecker product. $\boldsymbol{A}^T$, $\boldsymbol{A}^H$ and $\boldsymbol{A}^{\#}$ denote the transpose, the Hermitian transpose, and the pseudo-inverse, respectively.

5.2 Problem Formulation

5.2.1 Multi-channel model

In this work, a multichannel framework is considered which is obtained either by oversampling the received signal or using an array of antennas or a combination of both [11]. To further develop the multi-channel system model, consider the observed signal $y(t)$ from a linear modulation over a linear channel with additive noise given by

$$y(t) = \sum_k h(t-k)s(k) + e(t), \quad t = 0, \ldots, N-1 \tag{5.1}$$

where $h(t)$ is the FIR channel impulse response, $s(k)$ are the transmitted symbols and $e(t)$ is the additive noise. If the received signal is oversampled or recorded with N_r sensors, the signal model in (6.1) becomes N_r-variate and expressed as

$$\boldsymbol{y}(t) = \sum_{i=0}^{M-1} \boldsymbol{h}(i)s(t-i) + \boldsymbol{e}(t) \tag{5.2}$$

where $\boldsymbol{y}(t) = [y_1(t), \cdots, y_{N_r}(t)]^T$, $\boldsymbol{h}(i) = [h_1(i), \cdots, h_{N_r}(i)]^T$, $\boldsymbol{e}(t) = [e_1(t), \cdots, e_{N_r}(t)]^T$.

Define the system's transfer function $\boldsymbol{H}(z) = \sum_{k=0}^{M-1} \boldsymbol{h}(k)z^{-k}$ with $(M-1) = deg(\boldsymbol{H}(z))$ and consider the noise to be additive independent white circular noise with $E[\boldsymbol{e}(k)\boldsymbol{e}^H(i)] = \delta_{k,i}\sigma_e^2\mathbf{I}_{N_r}$ ($\delta_{k,i}$, σ_e^2, and $\mathbf{I}_{N_r}$ are the Kronecker index, the noise power and the $N_r \times N_r$ identity matrix, respectively). Assume a reception of a window of N_w samples, by stacking the data into a vector/matrix representation, we get:

$$\boldsymbol{y}_{N_w}(t) = \boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})\boldsymbol{s}_{N_w+M-1}(t) + \boldsymbol{e}_{N_w}(t) \tag{5.3}$$

where $\boldsymbol{y}_{N_w}(t) = [\boldsymbol{y}^H(t), \cdots, \boldsymbol{y}^H(t-N_w+1)]^H$, $\boldsymbol{s}_{N_w+M-1}(t) = [s(t), \cdots, s(t-N_w-M+2)]^T$, $\boldsymbol{e}_{N_w}(t)$ is stacked in a similar way to as $\boldsymbol{y}_{N_w}(t)$, and $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$ is an $mM \times (N_w+M-1)$ block Toeplitz matrix defined as

$$\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h}) = \begin{bmatrix} \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(M-1) & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{h}(0) & \cdots & \boldsymbol{h}(M-1) \end{bmatrix} \tag{5.4}$$

$\boldsymbol{h}$ is the desired parameter vector containing all channels' taps, i.e., $\boldsymbol{h} = [\boldsymbol{h}^T(0), \cdots, \boldsymbol{h}^T(M-1)]^T$. Using the observation data in (6.3), our objective is to estimate the different channels' impulse responses, i.e, recover $\boldsymbol{h}$ up to a possible scalar ambiguity. In the following subsection, we describe the subspace method, briefly.

5.2.2 Subspace method revisited

For consistency and reader's convenience, the SS method [45], which is also referred to as the noise subspace method, shall be reviewed hereafter. The SS method implicitly exploits the Toeplitz structure of the filtering matrix $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$. Let $\boldsymbol{v} = [\boldsymbol{v}_1^H, \cdots, \boldsymbol{v}_{N_w}^H]^H$, where $\boldsymbol{v}_i = [v_{(i-1)N_r+1}, \cdots, v_{iN_r}]^T$ with $i = 1, \ldots, N_w$, be in the orthogonal complement space of the range space of $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$ such that

$$\boldsymbol{v}^H \boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h}) = 0 \tag{5.5}$$

Using the block Toeplitz structure of $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$, the above linear equation can be written in terms of the channel parameter $\boldsymbol{h}$ as

$$\boldsymbol{h}^H \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_{N_w} & 0 & 0 \\ 0 & \ddots & & \ddots & 0 \\ 0 & 0 & \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_{N_w} \end{bmatrix} = \boldsymbol{h}^H \boldsymbol{\mathcal{V}} = 0 \tag{5.6}$$

The former equation can be used to estimate the channel vector $\boldsymbol{h}$ provided that (5.6) has a unique solution. Moulines et al. [45] proposed the SS method which

is based on the following theorem:

Theorem 5.1 *Assume that the components of* $\boldsymbol{H}(z)$ *have no common zeros, and* $N_w \geq M$. *Let* $\{\boldsymbol{v}_i\}_{i=1}^{d}$ *be a basis of the orthogonal complement of the column space of* $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$, *then for any* $\boldsymbol{H}'(z)$ *with* $deg(\boldsymbol{H}'(z)) = M-1$ *we have*

$$\boldsymbol{\mathcal{V}}_i^H \boldsymbol{h}' = 0, \text{for} \;\; i = 1, \cdots, d \iff \boldsymbol{H}'(z) = \alpha \boldsymbol{H}(z) \tag{5.7}$$

where α *is some scalar factor.*

One of the encountered ways to estimate the orthogonal complement of $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$, i.e., the noise subspace, is the signal-noise subspace decomposition. From the multi-channel model and noise properties, the received signal covariance matrix $\boldsymbol{R}_y = E[\boldsymbol{y}_{N_w}(t)\boldsymbol{y}_{N_w}^H(t)]$ is given as

$$\boldsymbol{R}_y = \boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})\mathbf{R}_s\boldsymbol{\mathcal{H}}_{N_w}^H(\boldsymbol{h}) + \sigma_e^2\boldsymbol{I} \tag{5.8}$$

where $\mathbf{R}_s$ is the covariance of the input signal. Consequently, the singular value decomposition of $\boldsymbol{R}_y$ has the form

$$\boldsymbol{R}_y = \boldsymbol{V}_s \text{diag}\left(\lambda_1^2, \cdots, \lambda_{\mathrm{N_w+M-1}}^2\right) \boldsymbol{V}_{\mathrm{s}}^{\mathrm{H}} + \sigma_{\mathrm{e}}^2 \boldsymbol{V}_{\mathrm{e}} \boldsymbol{V}_{\mathrm{e}}^{\mathrm{H}} \tag{5.9}$$

where λ_i^2's, $i = 1, \cdots, N_w + M - 1$, are the principal eigenvalues of the covariance matrix $\mathbf{R}_y$. Also, the columns of $\boldsymbol{V}_s$ and $\boldsymbol{V}_e$ span the so-called signal and noise subspaces (orthogonal complement), respectively. After having the basis of the

Table 5.1: Duality Table

Method	**Toeplitz Structure**	**Orthogonality**
SS	forced	minimized
SSS	minimized	forced

noise subspace, the channel identification can be performed based on the following quadratic optimization criterion:

$$\hat{\boldsymbol{h}} = \arg \min_{\boldsymbol{h}} \|\boldsymbol{V}_e^H \boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})\|^2 = \arg \min_{\boldsymbol{h}} \boldsymbol{h}^H \left[\sum_i \boldsymbol{\mathcal{V}}_i \boldsymbol{\mathcal{V}}_i^H\right] \boldsymbol{h} \tag{5.10}$$

In brief, the SS method achieves the channel estimation by exploiting the subspace information (i.e., ideally, (Range$\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$) = Range($\boldsymbol{V}_s$) $\perp$ Range($\boldsymbol{V}_e$)) as well as the block Sylvester (block-Toeplitz) structure of the channel matrix. More precisely, it enforces the latter matrix structure through the use of relations (5.5) and (5.6) and minimizes the subspace orthogonality error in (5.10). In the sequel, unlike the approach of the SS method, we propose a dual approach which enforces the subspace information (i.e., Range($\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$) = Range($\boldsymbol{V}_s$) where $\boldsymbol{V}_s$ refers to the principal subspace of the sample covariance matrix) while minimizing a cost function representing the deviation of $\boldsymbol{\mathcal{H}}_{N_w}(\boldsymbol{h})$ from the Sylvester structure as indicated in Table 5.1. This will result in the SSS method described next.

5.3 Structure-Based SS method (SSS)

In the proposed subspace method, one searches for the system matrix $\boldsymbol{\mathcal{H}}_{N_w}$ in the form $\hat{\boldsymbol{\mathcal{H}}}_{N_w} = \boldsymbol{V}_s \boldsymbol{Q}$ so that the orthogonality criterion in (5.10) is set equal to zero,

i.e. $\|\boldsymbol{V}_e^H \hat{\boldsymbol{\mathcal{H}}}_{N_w}\|^2 = 0$ while $\boldsymbol{Q}$ is chosen in such a way the resulting matrix is close to the desired block Toeplitz structure. This is done by minimizing w.r.t. $\boldsymbol{Q}$ the following structure-based cost function (informal Matlab notions are used):

$$\begin{aligned}
&\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 \\
&= \left| \sum_{j=1}^{K-1} \sum_{i=1}^{N_r(N_w-1)} \hat{w}(i,j) - \hat{w}(i+N_r, j+1) \right|^2 \\
&+ \left| \sum_{j=M+1}^{K} \hat{w}(1:N_r, j) \right|^2 + \left| \sum_{i=N_r+1}^{mM} \hat{w}(i,1) \right|^2
\end{aligned} \tag{5.11}$$

where $K = N_w + M - 1$ and $\hat{\boldsymbol{W}}$ refers to $\hat{\boldsymbol{\mathcal{H}}}_{N_w}$. The cost function in (5.11) is inspired and matched to the Toeplitz structure introduced in (5.4). It is a composite of three parts: $\mathcal{J}_1$ seeks to force the Toeplitz structure on the possibly non-zero entries, while $\mathcal{J}_2$ and $\mathcal{J}_3$ account for the zero entries in the first N_r rows and first column of $\boldsymbol{\mathcal{H}}_{N_w}(\mathbf{h})$, respectively.

Starting with $\mathcal{J}_1$, one can express it in a more compact way as follows:

$$\mathcal{J}_1 = \left\| \boldsymbol{I}_L \hat{\boldsymbol{W}} \boldsymbol{I}_R - \boldsymbol{J}_L \hat{\boldsymbol{W}} \boldsymbol{J}_R \right\|^2 \tag{5.12}$$

where:

$\boldsymbol{I}_L$ is the $(N_r N_w) \times (N_r N_w)$ left identity square matrix with setting the last N_r diagonal entries to zeros.

$\boldsymbol{I}_R$ is the $K \times K$ right identity square matrix with setting the last diagonal entry to zero.

$\boldsymbol{J}_L$ is a $(N_r N_w) \times (N_r N_w)$ square translation matrix with ones on the sub-diagonal and zeros elsewhere, i.e., $[J_L]_{i,j} = \delta_{i+N_r,j}$.

$\boldsymbol{J}_R$ is a $K \times K$ square translation matrix with ones on the super-diagonal and zeros elsewhere, i.e., $[J_R]_{i,j} = \delta_{i,j+1}$.

Now, using the Kronecker product property $\text{vec}(\boldsymbol{AGB}) = \left(\boldsymbol{B}^T \otimes \boldsymbol{A}\right) \text{vec}(\boldsymbol{G}) = \left(\boldsymbol{B}^T \otimes \boldsymbol{A}\right) \boldsymbol{g}$, one can write $\mathcal{J}_1$ as follows:

$$\begin{aligned} \mathcal{J}_1 &= \| \left(\boldsymbol{I}_R \otimes \boldsymbol{I}_L - \boldsymbol{J}_R^T \otimes \boldsymbol{J}_L\right) \text{vec}(\hat{\boldsymbol{W}}) \|^2 \\ &= \| \left(\boldsymbol{I}_R \otimes \boldsymbol{I}_L - \boldsymbol{J}_R^T \otimes \boldsymbol{J}_L\right) \left(\boldsymbol{I} \otimes \boldsymbol{V}_s\right) \boldsymbol{q} \|^2 = \| \boldsymbol{K}_1 \boldsymbol{q} \|^2 \end{aligned} \tag{5.13}$$

where $\boldsymbol{q} = \text{vec}(\boldsymbol{Q})$. In a similar way, $\mathcal{J}_2$ can be expressed as

$$\begin{aligned} \mathcal{J}_2 &= \| \hat{\boldsymbol{W}} \left(1 : N_r, M+1 : end\right) \|^2 = \| \boldsymbol{V}_{s,row} \boldsymbol{Q} \boldsymbol{I}_{row} \|^2 \\ &= \| \left(\boldsymbol{I}_{row} \otimes \boldsymbol{V}_{s,row}\right) \boldsymbol{q} \|^2 = \| \boldsymbol{K}_2 \boldsymbol{q} \|^2 \end{aligned} \tag{5.14}$$

where $\boldsymbol{V}_{s,row}$ is the sub-matrix of $\boldsymbol{V}_s$ given by its first N_r rows, and $\boldsymbol{I}_{row}$ is the $K \times K$ square identity matrix with setting the first M diagonal entries to zero. Finally, $\mathcal{J}_3$ can also be set up as

$$\begin{aligned} \mathcal{J}_3 &= \| \hat{\boldsymbol{W}} \left(N_r + 1 : N_r N_w, 1\right) \|^2 = \| \boldsymbol{V}_{s,col} \boldsymbol{Q} \boldsymbol{I}_{col} \|^2 \\ &= \| \left(\boldsymbol{I}_{col} \otimes \boldsymbol{V}_{s,col}\right) \boldsymbol{q} \|^2 = \| \boldsymbol{K}_3 \boldsymbol{q} \|^2 \end{aligned} \tag{5.15}$$

where $\boldsymbol{V}_{s,col}$ is the sub-matrix of $\boldsymbol{V}_s$ given by its last $N_r(N_w - 1)$ rows, and $\boldsymbol{I}_{col}$ is the $K \times K$ square diagonal matrix with one at the first diagonal entry and zeros elsewhere.

As a result of (5.13), (5.14) and (5.15) the optimization problem in (5.11) is reduced to the minimization of the following quadratic equation

$$\min_{\boldsymbol{q}} \quad \boldsymbol{q}^H \boldsymbol{K}^H \boldsymbol{K} \boldsymbol{q} \tag{5.16}$$

where $\boldsymbol{K} = \left[\begin{array}{c|c|c} \boldsymbol{K}_1^T & \boldsymbol{K}_2^T & \boldsymbol{K}_3^T \end{array} \right]^T$.

The optimal solution $\boldsymbol{q}$ of (5.16), under unit norm constraint of $\boldsymbol{q}$, is the least eigenvector that corresponds to the smallest eigenvalue of $\boldsymbol{K}^H \boldsymbol{K}$. The square matrix $\boldsymbol{Q}$ can be constructed by reshaping the obtained solution $\boldsymbol{q}$ from a vector into the matrix format, such that $\boldsymbol{Q} = \text{mat}_{K,K}\{\boldsymbol{q}\}$. Once matrix $\boldsymbol{Q}$ is obtained, the channel taps are estimated by averaging over the non-zero diagonal blocks of the matrix $\boldsymbol{V}_s \boldsymbol{Q}$.

5.4 Discussion

In this section, we provide some insightful comments in order to highlight the advantages and drawbacks of the proposed subspace method.

- As explained earlier the proposed approach consists of neglecting the subspace error (i.e., considering Range($\boldsymbol{V}_s$) as perfect in the sense one searches for the desired solution within that subspace) while minimizing the system's matrix (Toeplitz) structure error. The motivation behind this choice resides in the fact that the subspace error at the first order is null and hence it can be neglected in favor of more flexibility for searching the appropriate

channel matrix. Indeed, the first order expansion of the subspace error can be written as (see [110, 81] for details):

$$\delta\mathbf{\Pi} = (\boldsymbol{I} - \mathbf{\Pi})\delta\boldsymbol{R}_y(\boldsymbol{R}_y - \sigma_e^2\boldsymbol{I})^{\#} + (\boldsymbol{R}_y - \sigma_e^2\boldsymbol{I})^{\#}\delta\boldsymbol{R}_y(\boldsymbol{I} - \mathbf{\Pi}) + o(\delta\boldsymbol{R}_y)$$

where $\mathbf{\Pi} = \boldsymbol{\mathcal{H}}_{N_w}\boldsymbol{\mathcal{H}}_{N_w}^{\#}$ is the orthogonal projection matrix on the principal (signal) subspace of $\boldsymbol{R}_y$ and δ refers to the estimation error. Now, plugging the previous first order terms into the subspace error criterion [111], leads to $E(\|\delta\mathbf{\Pi}(\boldsymbol{I} - \mathbf{\Pi})\|^2) = 0$ due to the orthogonality relation $(\boldsymbol{I} - \mathbf{\Pi})(\boldsymbol{R}_y - \sigma_e^2\boldsymbol{I})^{\#} = 0$. This explains the observed gain of the SSS over SS method in certain difficult scenarios including the case of closely spaced channels roots.

- In the favorable cases where the channel matrix is well conditioned, the two subspace methods (ours and that of [45]) lead to similar performance as illustrated next in the simulation example of Fig. 5.1.

- For the SS method to apply, one needs that the noise subspace vectors generate a minimal polynomial basis of the rational subspace orthogonal to Range($\boldsymbol{H}(z)$) (see [45] for more details) and so the condition $N_w \geq M$ is considered to guarantee such requirement to hold. As the SSS does not explicitly rely on the orthogonality relation in (5.10), the latter condition might be relaxed as illustrated in the simulation example of Fig. 5.4.

- The proposed subspace method has a higher numerical cost as compared

to the SS method. However, the cost might be reduced by taking into account the Kronecker products involved in building matrix $\boldsymbol{K}$. This issue is still under investigation together with an asymptotic statistical performance analysis of SSS.

- In the case $N_w \geq M$, the solution of (5.16) is unique (up to a constant) thanks to the identifiability result of Theorem 5.1. Indeed, if $\boldsymbol{q}'$ is another solution zeroing criterion (5.11), then the FIR filter associated to matrix $\boldsymbol{\mathcal{H}}' = \boldsymbol{V}_s\boldsymbol{Q}'$ satisfies all conditions of Theorem 5.1, which leads to $\boldsymbol{V}_s\boldsymbol{Q}' = \alpha\boldsymbol{V}_s\boldsymbol{Q}$ or equivalently $\boldsymbol{Q}' = \alpha\boldsymbol{Q}$.

5.5 SIMULATION RESULTS

In this section, the devised SSS method will be benchmarked to the standard SS and the TTM methods. Three different experiments will be examined to illustrate the behavior of the SSS method in different contexts.

Two FIR channels are considered, each has a second order impulse response given by [109, 112]:

$$\boldsymbol{h}_1 = \begin{bmatrix} 1 & -2\cos(\theta) & 1 \end{bmatrix}^T,$$
$$\boldsymbol{h}_2 = \begin{bmatrix} 1 & -2\cos(\theta+\delta) & 1 \end{bmatrix}^T$$

where θ is the absolute phase value of $\mathbf{h}_1$'s zeros and δ indicates the angular distance between the zeros of the two channels on the unit circle. Small δ results

into an ill-conditioned system. In all simulations, the excitation signal is a 4-QAM, each channel receives $N = 100$ samples, and the noise is white Gaussian. Note that the SNR is defined as

$$SNR(dB) = 10\log_{10} E\frac{\|\boldsymbol{\mathcal{H}}_N \boldsymbol{s}_{N+M-1}\|^2}{mN\sigma_e^2}$$

The performance measure is the normalized mean-square-error (MSE), given as

$$MSE(dB) = 20log_{10}\left(\frac{1}{||\boldsymbol{h}||}\sqrt{\frac{1}{N_{mc}}\sum_{i=1}^{N_{mc}}||\hat{\boldsymbol{h}}_i - \boldsymbol{h}||^2}\right)$$

where $N_{mc} = 100$ refers to the number of Monte-Carlo runs and $\hat{\boldsymbol{h}}_i$ is the channel vector estimate at the i-th run.

In the first experiment given by Fig. 5.1, we show that for a well-conditioned system ($\delta = \pi$), both SS and SSS methods have a comparable performance while the TTM has a very poor performance unless the sample size is very high. This can be explained by the fact that, the SS and SSS methods share the subspace property discussed in the first item of section 5.4 while TTM does not.

In the second experiment, we consider ill-conditioned systems (i.e. poor channel diversity, $\delta = \pi/10$). In this case, as shown in Fig. 5.2, the devised SSS method beats the SS method at all SNR values while it outperforms the TTM method at high SNR. At low SNR, TTM achieves the best performance due mainly to the fact it combines two channel estimates, a costly procedure known to improve the estimation performance in noisy scenarios [113].

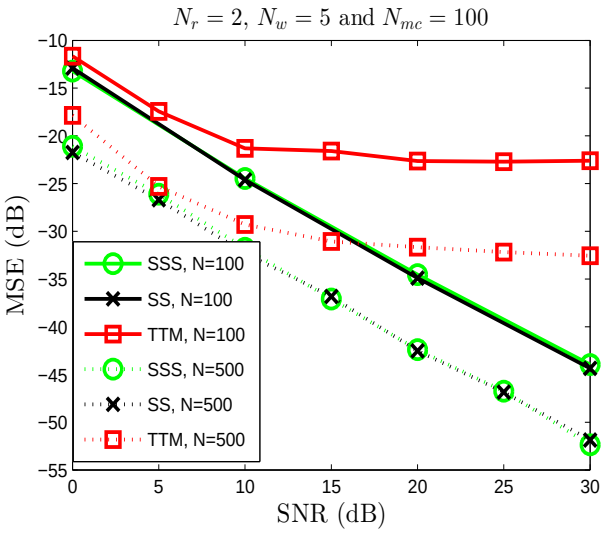


Figure 5.1: Well-conditioned channels, $\theta = \pi/10, \delta = \pi$.

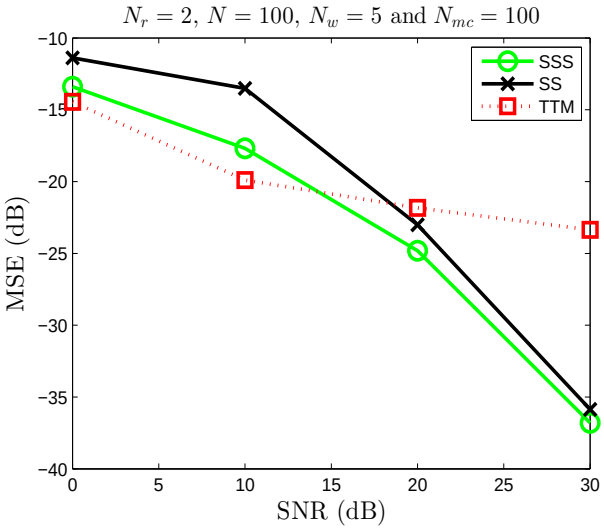


Figure 5.2: Ill-conditioned system, $\theta = \pi/10, \delta = \pi/10$.

Figure 5.3 depicts the consequence of varying δ on the MSE for SNR=15dB, once again the SSS has a clear sustainable gain at almost all the channel diversity conditions.

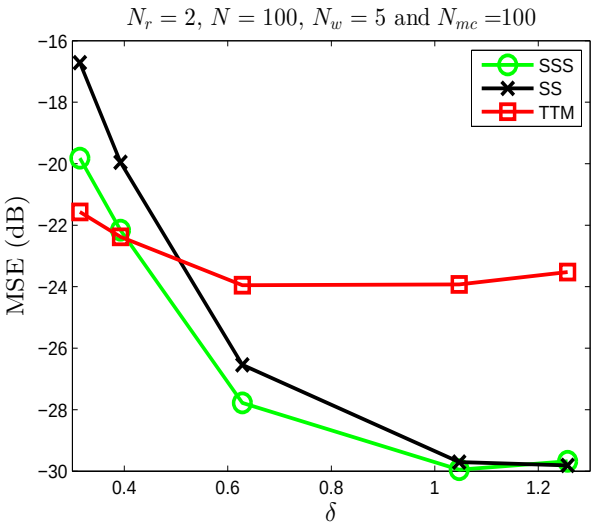


Figure 5.3: MSE versus δ, $SNR = 15\ dB$

In the last experiment, the number of channels is $N_r = 3$ and the number of taps in each channel is $M = 5$, and the transfer function of the channels are given in [81] (corresponding to a relatively well-conditioned channel case). In this experiment, we are primarily interested to look at the impact of the processing window length on the estimation performance. As can be seen from the results reported in Fig. 5.4, the performance of the SS method gets worse and degrades when the processing window length N_w becomes less than the number of the channels' taps M, while our proposed SSS is weakly affected by the window length condition, i.e. $N_w \geq M$. This allows us to reduce the dimension of the channel matrix $\boldsymbol{\mathcal{H}}_{N_w}$ with smaller window size values, especially for large dimensional systems where $N_r \gg 1$.

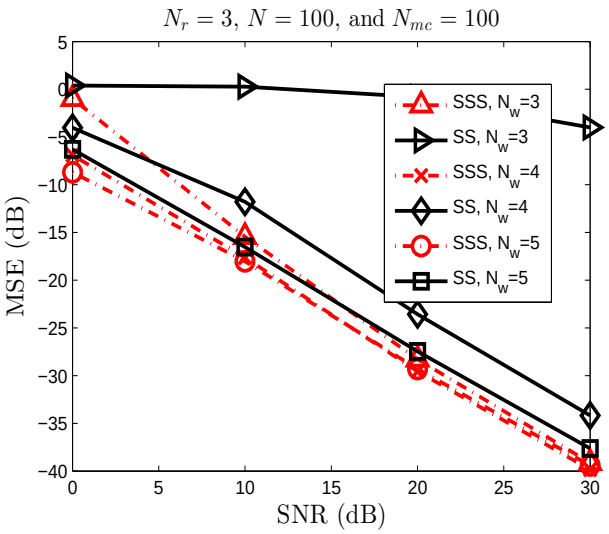


Figure 5.4: MSE versus SNR for different window size N_w.

5.6 CONCLUSION

In this chapter, we proposed a dual approach to the standard subspace method, whereby the channel matrix is forced to belong to the principal subspace of the data covariance matrix estimate while its deviation from Toeplitz structure is minimized. By doing so, we show that the channel estimation is significantly improved in the difficult context of weak channels diversity (i.e. channels with closely spaced roots). Interestingly, the principle of the proposed approach can be applied for estimation problems with other matrix structures where subspace method can be used.

CHAPTER 6

PERFORMANCE ANALYSIS

Blind channel estimation techniques have gained a considerable attention over the last two decades. The need for Blind channel estimation emerges in many engineering applications including bio-medicine, sonar, radar, radio astronomy, seismic exploration, navigation, spying and wireless communications. Recently, there is a renewed increased interest in a hybrid or assisted blind channel estimation and equalization techniques due to the emergence of new MIMO and Massive-MIMO communication systems.

In this chapter we investigate the impact of certain side information that are available in the channels and/or signals on the blind system identification through the Cramer-Rao Bound (CRB). More precisely, we considered a Single Input Multiple Output (SIMO) system, and studied, for different scenarios, the performance bounds for channel estimation in both deterministic and Bayesian cases. The latter correspond to the situations where side information is brought by either a pilot sequence (semi-blind case), channel sparsity (specular channel case) or cer-

tain signal's statistical properties such as the non-circularity. This analysis allows us to have a better understanding of the behavior of the blind channel estimation when the considered side information is taken into account.

6.1 Introduction

Recent advances in communication systems calls for new alternatives of channel estimation to solve the problem of scarce bandwidth. Traditionally, in wireless communication systems, recurrent training signals are needed as the channel varies rapidly. This process showed to result in a reduced transmission rate. Blind channel estimation was one of the solution to get rid of the pilot signal, where the channel estimation process depends completely on the received signal. Yet, unsatisfactory results of the blind channel estimation methods and algorithms emerged attention on some other approaches, one is known as semi-blind [114], whereby a superior performance is achieved by utilizing the received signal and few training pilots. Another strategy consists of reinforcing the blind estimation by exploiting structural or statistical a priori information about the signals or channels, such as the prior distribution [50], channel sparsity or signal cyclostationarity.

Recently, some efforts have been put forward to quantify the gain that could be obtained in the semi-blind scenario and develop appropriate algorithms for that. That was done in the Muli-channel context in both SIMO case [50, 56, 57, 58, 59, 60, 61, 62] and MIMO case [115, 116, 117, 54, 118, 55, 119]. The spirit of this contribution is to extend this investigation to other properties that have not

been explored carefully. Also, most of the existing blind methods such as subspace [45], least squares [48] and linear prediction [120] methods suffer from the channel length overmodeling problem. The robustness of blind estimation methods to channel overmodeling is a crucial issue that is directly related to their usefulness to real world applications. Based on that and using the most efficient lower bound accustomed by statisticians, namely the CRB, the focus of this chapter is of two folds:

- The first one is to point out and quantify the effect of channel overmodeling on the estimation quality in different scenarios, such as blind, semi-blind, Bayesian and non-Bayesian estimation [50],[58].

- The second one is to investigate the consequence of using side information, i.e. structural or statistical, on the CRB and hence on the channel estimation performance.

The rest of the this chapter is organized as follows: section II describes the blind and semi-blind SIMO[1] models. Section III reviews the corresponding CRB for the Bayesian and non-Bayesian scenarios. It also includes the non-circularity and stochastic CRB as well as the CRBs for the structure-aided models (sparse and specular channels). Comments and insightful discussions are given in section IV. The numerical simulation that illustrates the handled CRBs are presented in section V. Section VI provides the concluding remarks.

[1]We have chosen to focus on SIMO system at first before investigating the more general case of MIMO systems in a future work.

6.2 Problem Formulation

6.2.1 Blind data model

Multichannel framework can be achieved either by oversampling the received signal or from using an array of antenna or a combination of both [45]. To further develop the multi channel system model, consider the observed signal $y(t)$ from a linear modulation over a linear channel with additive noise given by

$$y(t) = \sum_{k} h(t-k)s(k) + e(t) \tag{6.1}$$

In (6.1), $h(t)$ is the FIR channel impulse response, $s(k)$ are the transmitted symbols and $e(t)$ is the additive noise. If the received signal is oversampled or recorded with N_r sensors, the signal model in (6.1) can be written as follows:

$$\boldsymbol{y}(t) = \sum_{i=1}^{M-1} \boldsymbol{h}(i)s(t-i) + \boldsymbol{e}(t) = \boldsymbol{H}\boldsymbol{s}(t) + \boldsymbol{e}(t) \tag{6.2}$$

where, $\boldsymbol{y}(t) = [y_1^H(t), \cdots, y_{N_r}^H(t)]^H$, suscript $(.)^H$ stands for conjugate transpose, $\boldsymbol{h}(i) = [h_1^H(i), \cdots, h_{N_r}^H(i)]^H$, $\boldsymbol{e}(t) = [e_1^H(t), \cdots, e_{N_r}^H(t)]^H$, $\boldsymbol{H} = [\boldsymbol{h}(M-1), \cdots, \boldsymbol{h}(0)]$ and $\boldsymbol{s}(t) = [s^H(t-M+1), \cdots, s^H(t)]^H$. Consider the noise to be additive independent white Gaussian circular process with $E[\boldsymbol{e}(k)\boldsymbol{e}^H(i)] = \delta_{k,i}\sigma_e^2\boldsymbol{I}_{N_r}$. Assume a reception of N samples, by stacking the data into a vec-

tor/matrix representation, we get:

$$\boldsymbol{y} = \boldsymbol{\mathcal{H}}\boldsymbol{s} + \boldsymbol{e} \tag{6.3}$$

where $\boldsymbol{y} = [\boldsymbol{y}^H(0), \cdots, \boldsymbol{y}^H(N-1)]^H$ and similarly applies for $\boldsymbol{e}$. $\boldsymbol{\mathcal{H}}$ is an $N_r N \times (N+M-1)$ block Toeplitz matrix with N block rows and $[\boldsymbol{H} \quad \boldsymbol{0}_{N_r \times (N-1)}]$ as its first block row, and $\boldsymbol{s} = [\boldsymbol{s}^H(1-M), \cdots, \boldsymbol{s}^H(N-1)]^H$.

6.2.2 Semi-Blind data model

In the semi-blind case, by using the convolution commutativity property, the received signal $\boldsymbol{s}$ is modeled as follows:

$$\begin{aligned} \boldsymbol{y} = \boldsymbol{\mathcal{H}}\boldsymbol{s} + \boldsymbol{e} &= \boldsymbol{\mathcal{H}}_p\boldsymbol{s}_p + \boldsymbol{\mathcal{H}}_d\boldsymbol{s}_d + \boldsymbol{e} \\ &= \boldsymbol{\mathcal{S}}\boldsymbol{h} + \boldsymbol{e} \approx (\boldsymbol{\mathcal{S}}_p + \boldsymbol{\mathcal{S}}_d)\boldsymbol{h} + \boldsymbol{e} \end{aligned} \tag{6.4}$$

where $\boldsymbol{\mathcal{H}}_p$ and $\boldsymbol{\mathcal{H}}_d$ are partitions of $\boldsymbol{\mathcal{H}}$ coresponding to pilot sequence $\boldsymbol{s}_p$ and data sequence $\boldsymbol{s}_d$, respectively, see (6.5). Also, $\boldsymbol{\mathcal{S}}$, $\boldsymbol{\mathcal{S}}_p$ and $\boldsymbol{\mathcal{S}}_d$ are block Hankel matrices filled with the elements of $\boldsymbol{s}$, $\boldsymbol{s}_p$ and $\boldsymbol{s}_d$, respectively[2]. For simplicity, all known/pilot symbols are sorted at the beginning of vector $\boldsymbol{s}$. For the data models in (6.3) and (6.4), our objective is to compare next the CRBs for channel estimation corresponding to different assumptions on the channel model or side information.

[2]The approximation in (6.4) can be used if we ignore the overlapping sequence which is common between the pilot and data sequences [58].

$$\boldsymbol{\mathcal{H}} = \left[\begin{array}{c|c|c} & & \\ \boldsymbol{\mathcal{H}}_p & | & \boldsymbol{\mathcal{H}}_d \\ & & \end{array}\right] \tag{6.5}$$

6.3 Channel's CRBs Derivations

In this section, we introduce the CRB expressions of the channel parameter vector for different scenarios representing various side information situations. At first (subsections 6.4.2,6.4.3 and 6.4.4), the side information is related to the input data, the last section corresponds to side information on the channel taps, while subsection 6.4.5 combines both. These CRBs are compared to the "no side information" case given by subsection 6.4.1.

6.4 Unified Framework for Different CRBs' Development

In this section, we shall provide a unified framework for the different CRBs' derivation. Generally, the subject of the estimator is to jointly estimate the symbols (unknown-symbols) and the channel impulse response. This is done by imposing some assumptions on both the channel and/or the symbols. It customarily to mention that in this kind of estimators the estimation of the symbols and channels is decoupled from the estimation of the noise variance, as the later has no effect

on the estimation of the earlier parameters. To this end, denote the unknown parameter vector $\boldsymbol{\Theta}$ by

$$\boldsymbol{\Theta} = [\boldsymbol{s}^H, \boldsymbol{h}^H]^H \tag{6.6}$$

then, the likelihood function is given by

$$f(\boldsymbol{Y}, \boldsymbol{\Theta}) = f(\boldsymbol{Y}/\boldsymbol{\Theta}) f(\boldsymbol{\Theta}) \tag{6.7}$$

where $f(\boldsymbol{Y}, \boldsymbol{\Theta})$ is the joint probability density function of $\boldsymbol{Y}$ and $\boldsymbol{\Theta}$, $f(\boldsymbol{\Theta})$ is the probability density function (pdf) of $\boldsymbol{\Theta}$, and $f(\boldsymbol{Y}/\boldsymbol{\Theta})$ is the conditional pdf.

Substituting (6.6) in (6.7) and noting the a priori independence of both the symbols and the channel, i.e., $f(\boldsymbol{s}, \boldsymbol{h}) = f(\boldsymbol{s}) f(\boldsymbol{h})$.

In the sequel, knowing that the CRBs, and accordingly the fisher information matrix (FIM), are developed by applying the log operator to the joint pdf. Hence, the log-likelihood function applied to the earlier developed expression is given as

$$\ln[f(\boldsymbol{Y}, \boldsymbol{\Theta})] = \ln[f(\boldsymbol{Y}/\boldsymbol{s}, \boldsymbol{h})] + \ln[f(\boldsymbol{s})] + \ln[f(\boldsymbol{h})] \tag{6.8}$$

The FIM is given by [60];

$$\begin{aligned} \boldsymbol{F}_{\boldsymbol{\Theta}\boldsymbol{\Theta}} &= E\left(\frac{\partial \ln[f(\boldsymbol{Y},\boldsymbol{s},\boldsymbol{h})]}{\partial \boldsymbol{\Theta}^*}\right)\left(\frac{\partial \ln[f(\boldsymbol{Y},\boldsymbol{s},\boldsymbol{h})]}{\partial \boldsymbol{\Theta}^*}\right)^H \\ &= -E\frac{\partial}{\partial \boldsymbol{\Theta}^*}\left(\frac{\partial \ln[f(\boldsymbol{Y},\boldsymbol{s},\boldsymbol{h})]}{\partial \boldsymbol{\Theta}^*}\right)^H \end{aligned} \tag{6.9}$$

Since we are treating complex parameters, in addition to $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}}$, we need to evaluated $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}^*}$ which is defined as

$$\begin{aligned}\boldsymbol{J}_{\boldsymbol{\Theta\Theta}^*} &= E\left(\frac{\partial \ln[f(\boldsymbol{Y},\boldsymbol{s},\boldsymbol{h})]}{\partial \boldsymbol{\Theta}^*}\right)\left(\frac{\partial \ln[f(\boldsymbol{Y},\boldsymbol{s},\boldsymbol{h})]}{\partial \boldsymbol{\Theta}}\right)^H \\ &= -E\frac{\partial}{\partial \boldsymbol{\Theta}}\left(\frac{\partial \ln[f(\boldsymbol{Y},\boldsymbol{s},\boldsymbol{h})]}{\partial \boldsymbol{\Theta}^*}\right)^H\end{aligned} \tag{6.10}$$

In case of $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}^*} \neq 0$, we resort to $\boldsymbol{\Theta}_R$ which is defined as

$$\boldsymbol{\Theta}_R = \begin{bmatrix} \mathrm{Re}\{\boldsymbol{\Theta}\} \\ \\ \mathrm{Im}\{\boldsymbol{\Theta}\} \end{bmatrix} = \boldsymbol{\mathcal{M}} \begin{bmatrix} \boldsymbol{\Theta} \\ \\ \boldsymbol{\Theta}^* \end{bmatrix}, \qquad \boldsymbol{\mathcal{M}} = \frac{1}{2}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \\ -\sqrt{-1}\boldsymbol{I} & \sqrt{-1}\boldsymbol{I} \end{bmatrix} \tag{6.11}$$

Noting that $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}} = \boldsymbol{J}^*_{\boldsymbol{\Theta}^*\boldsymbol{\Theta}^*}$ and $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}^*} = \boldsymbol{J}^*_{\boldsymbol{\Theta}^*\boldsymbol{\Theta}}$, then 6.11 implies:

$$\boldsymbol{J}_{\boldsymbol{\Theta}_R\boldsymbol{\Theta}_R} = \boldsymbol{\mathcal{M}} \begin{bmatrix} \boldsymbol{J}_{\boldsymbol{\Theta\Theta}} & \boldsymbol{J}_{\boldsymbol{\Theta\Theta}^*} \\ \\ \boldsymbol{J}^*_{\boldsymbol{\Theta\Theta}^*} & \boldsymbol{J}^*_{\boldsymbol{\Theta\Theta}} \end{bmatrix} \boldsymbol{\mathcal{M}}^H \tag{6.12}$$

Interestingly, in the case of jointly estimating the channel and symbols parameters, one can show that $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}^*} = 0$ and hence $\boldsymbol{J}_{\boldsymbol{\Theta}_R\boldsymbol{\Theta}_R}$ can be obtained totally from $\boldsymbol{J}_{\boldsymbol{\Theta\Theta}}$. After some regularity conditions and assumptions, the error covariance matrix of an unbiased estimator $\hat{\boldsymbol{h}}(\boldsymbol{Y})$ is defined as [121]

$$\boldsymbol{C}_{\hat{\boldsymbol{h}}} = E\left\{\left(\hat{\boldsymbol{h}}(\boldsymbol{Y}) - \boldsymbol{h}\right)\left(\hat{\boldsymbol{h}}(\boldsymbol{Y}) - \boldsymbol{h}\right)^H\right\} \tag{6.13}$$

satisfies the following inequality:

$$\boldsymbol{C}_{\hat{\boldsymbol{h}}} \geq \{\boldsymbol{J}_{\boldsymbol{\Theta}_R\boldsymbol{\Theta}_R}\}^{-1} \triangleq CRB \tag{6.14}$$

6.4.1 Deterministic Gaussian CRB

Considering the data model in (6.2) we assume here that the transmitted source samples and channel impulse response are all unknown deterministic variables such that the unknown parameter vector is $\boldsymbol{\Theta} = [\boldsymbol{s}^H, \boldsymbol{h}^H]^H$. Noting that, due to inherent scalar ambiguity in all blind channel identification, various constraints are imposed on the channel estimation to regularize the estimation problem [52] In what follow, we introduce the CRB that corresponds to the following constraint:

$$f(\boldsymbol{\Theta}) = h_r(i) - h_r^o = 0 \tag{6.15}$$

whereby one of the channel tap $h_r(i)$ is assumed to be known (equal to h_r^o). Hence, the constrained CRB is given by [51]

$$DCRB_{blind} = \boldsymbol{U}_r \left(\boldsymbol{U}_r^H \boldsymbol{J}_{\boldsymbol{hh}} \boldsymbol{U}_r\right)^{-1} \boldsymbol{U}_r^H \tag{6.16}$$

where

$$\boldsymbol{J}_{\boldsymbol{hh}} = \sigma_e^2 \left(\boldsymbol{\mathcal{S}}^H \boldsymbol{P}_{\boldsymbol{\mathcal{H}}}^{\perp} \boldsymbol{\mathcal{S}}\right) \tag{6.17}$$

with $\boldsymbol{P}_{\boldsymbol{\mathcal{H}}}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{\mathcal{H}}}$ and $\boldsymbol{P}_H = \boldsymbol{\mathcal{H}}(\boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}})^{-1}\boldsymbol{\mathcal{H}}^H$ is the projection matrix on $\boldsymbol{\mathcal{H}}$. $\boldsymbol{h} = vec(\boldsymbol{H})$, $vec(\boldsymbol{H})$ is an operator stacking the columns of a matrix $\boldsymbol{H}$ into a vector. $\boldsymbol{U}_r$ is the orthonormal matrix which columns span the null space of $\partial f(\boldsymbol{\Theta})/\partial\boldsymbol{\Theta}^*$ and obtained by removing the $((M-1-i)N_r+r)^{th}$ column from the identity matrix $\boldsymbol{I}_{N_rM}$.

In the light of channel order overdeterminacy, it can be shown that the blind system identification has an inherent polynomial indeterminacy problem. To this end, let $\boldsymbol{H}(z) = \sum_{i=0}^{M-1}\boldsymbol{h}(i)z^{-i} = [\boldsymbol{H}_1^H(z)\cdots\boldsymbol{H}_{N_r}^H(z)]^H$, if M is unknown and we overestimate it by $\tilde{M} = M + P$, then the system model in (6.2) will be equivalent to $\boldsymbol{y}(t) = [\boldsymbol{H}(z)\alpha(z)]\,\tilde{\boldsymbol{s}}(t) + \boldsymbol{e}(t)$, where $\alpha(z)$ is a scalar polynomial of degree P, also $\tilde{\boldsymbol{s}}(t) = [\alpha(z)^{-1}]\boldsymbol{s}(t)$. In that case, the channel parameters are not uniquely identifiable and the FIM is not invertible unless certain side information is provided as will be discussed next

6.4.2 Stochastic and Non-Circularity Based CRB

In this subsection, we consider two sorts of the statistical prior information, i.e, the circularity and non-circularity[3]. Such two piece of side information characterize the majority of the communication signals. Besides, other characteristics have been investigated, and due to the noticed neglected impact, it has not been presented here, such as cyclo-stationarity. The rest of this subsection is dedicated

[3]We have derived similar work for cyclostationary input signals but omit to present it here. Indeed, we observed that cyclostationarity has great positive impact on MIMO system identification (when sources have different cyclofrequencies) but negligible impact for SIMO systems.

to assess and quantify the statistical side information impact on the channel estimation. In the following, the received signal $\boldsymbol{s}(t)$ in (6.2) is modeled as a white, non-circular [4], Gaussian distributed random process. The noise $\boldsymbol{e}(t)$ is white, circular, zero-mean Gaussian random process of covariance $\sigma_e^2 \boldsymbol{I}_{N_r}$. Hence, the array output $\boldsymbol{y}(t)$ is a temporally white, zero-mean, non-circular, Gaussian distributed random process with covariance matrices given by:

$$\boldsymbol{R}_y = \boldsymbol{\mathcal{H}} \boldsymbol{R}_s \boldsymbol{\mathcal{H}}^H + \sigma_e^2 \boldsymbol{I}_{mM}, \quad \dot{\boldsymbol{R}}_y = \boldsymbol{\mathcal{H}} \dot{\boldsymbol{R}}_s \boldsymbol{\mathcal{H}}^T \tag{6.18}$$

where $\boldsymbol{R}_s = E[\boldsymbol{s}\boldsymbol{s}^H]$ and $\dot{\boldsymbol{R}}_s = E[\boldsymbol{s}\boldsymbol{s}^T] = \rho \sigma_s^2 \boldsymbol{I}_{(N+M-1)}$, and $\rho : (0 \leq |\rho| \leq 1)$ is a complex factor controlling the non-circularity rate. Under these assumptions, and taking advantage of scale ambiguity to assume $\sigma_s^2 = 1$, and the phase ambiguity to consider ρ as a real valued, the FIM with respect to unknowns parameter vector $\boldsymbol{\Theta} = \left[\mathrm{Re}\{\boldsymbol{h}^T\}, \mathrm{Im}\{\boldsymbol{h}^T\}, |\rho|, \sigma_e^2\right]^T$ is given by (elementwise) [53]:

$$(\boldsymbol{J})_{k,l} = \frac{1}{2} Tr \left[\frac{\partial \boldsymbol{R}_z}{\partial \theta_k} \boldsymbol{R}_z^{-1} \frac{\partial \boldsymbol{R}_z}{\partial \theta_l} \boldsymbol{R}_z^{-1} \right] \tag{6.19}$$

where $\boldsymbol{z} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{pmatrix}$ and $\boldsymbol{R}_z = \begin{pmatrix} \boldsymbol{R}_y & \dot{\boldsymbol{R}}_y \\ \dot{\boldsymbol{R}}_y^* & \boldsymbol{R}_y^* \end{pmatrix}$. In this case, no closed form expression for the CRB was found. To this end, one simply use the traditional computation way, such that evaluate the inverse of the complete FIM, then $\boldsymbol{J}_{\boldsymbol{hh}}$ can be easily evaluated by considering the top left $(2N_rM)$ block matrix of FIM

[4]Non-circularity of the source signal means that its probability distribution varies under rotation in the complex plane

inverse. As for the derivatives with respect to different parameters, they are given as below:

$$\frac{\partial \boldsymbol{R}_y}{\partial \mathrm{Re}\{h_k\}} = \frac{\partial \boldsymbol{R}_y}{\partial \mathrm{Re}\{h_r(q)\}} = \left[\boldsymbol{\mathcal{H}}\left(\frac{\partial \boldsymbol{\mathcal{H}}}{\partial \mathrm{Re}\{h_r(q)\}}\right)^H + \left(\frac{\partial \boldsymbol{\mathcal{H}}}{\partial \mathrm{Re}\{h_r(q)\}}\right)\boldsymbol{\mathcal{H}}^H\right]$$

$$\frac{\partial \dot{\boldsymbol{R}}_y}{\partial \mathrm{Re}\{h_k\}} = \frac{\partial \dot{\boldsymbol{R}}_y}{\partial \mathrm{Re}\{h_r(q)\}} = \left[\boldsymbol{\mathcal{H}}\left(\frac{\partial \boldsymbol{\mathcal{H}}}{\partial \mathrm{Re}\{h_r(q)\}}\right)^T + \left(\frac{\partial \boldsymbol{\mathcal{H}}}{\partial \mathrm{Re}\{h_r(q)\}}\right)\boldsymbol{\mathcal{H}}^T\right]$$

$$\frac{\partial \dot{\boldsymbol{R}}_y}{\partial |\rho|} = \sigma_s^2 \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^T$$

$$\frac{\partial \boldsymbol{R}_y}{\partial |\rho|} = 0$$

$$\frac{\partial \boldsymbol{R}_y}{\partial \sigma_e^2} = I_{mM}$$

$$\frac{\partial \dot{\boldsymbol{R}}_y}{\partial \sigma_e^2} = 0$$

with $r = \mathrm{mod}(k-1, N_r) + 1$ and $q = \frac{k-r}{N_r} + 1, (r = 1 : N_r, q = 1 : M)$. For the derivation w.r.t $\mathrm{Im}\{\boldsymbol{h}\}$, we just use $\frac{\partial \boldsymbol{\mathcal{H}}}{\partial \mathrm{Im}\{h_r(q)\}} = j\frac{\partial \boldsymbol{\mathcal{H}}}{\partial \mathrm{Re}\{h_r(q)\}}$ in the previous formula. Note that the stochastic circular Gaussian CRB can be obtained by setting $|\rho| = 0$ and $\boldsymbol{\Theta} = \left[\mathrm{Re}\{\boldsymbol{h}^T\}, \mathrm{Im}\{\boldsymbol{h}^T\}, \sigma_s^2, \sigma_e^2\right]^T$.

6.4.3 Deterministic Semi-Blind CRB

In this case, the channel and the data symbol sequence are both assumed to be deterministic and needs to be estimated jointly, i.e. $\boldsymbol{\Theta} = [\boldsymbol{s}_d^H, \boldsymbol{h}^H]^H$ [5], the CRB is expressed by [58]:

$$DCRB_{SB} = \boldsymbol{J}_{\boldsymbol{hh}}^{-1} = \sigma_e^2\left(\boldsymbol{\mathcal{S}}^H \boldsymbol{P}_{\boldsymbol{\mathcal{H}}_d}^{\perp} \boldsymbol{\mathcal{S}}\right)^{-1} \tag{6.20}$$

with $\boldsymbol{P}_{\boldsymbol{\mathcal{H}}_d}^{\perp} = \boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{\mathcal{H}}_d}$ and $\boldsymbol{P}_{\boldsymbol{\mathcal{H}}_d} = \boldsymbol{\mathcal{H}}_d(\boldsymbol{\mathcal{H}}_d^H\boldsymbol{\mathcal{H}}_d)^{-1}\boldsymbol{\mathcal{H}}_d^H$ is the projection matrix on $\boldsymbol{\mathcal{H}}_d$.

[5] Often, we disregard σ_e^2 as its estimation can be shown to not affect the channel CRB in the considered context.

6.4.4 Bayesian CRB

It is well noticed that wireless communication channels are typically modeled as having Rayleigh fading with a prior Gaussian distribution expressing correlations between channel coefficients. Below, we shall exploit and investigate the knowledge of channel a priori information effect on its estimation quality. Hence, both the symbol sequence and the channel are treated to be random with Gaussian distributions $f(\boldsymbol{s}) = \frac{1}{(\pi\sigma_s^2)^N}\exp(-\frac{\boldsymbol{s}^H\boldsymbol{s}}{\sigma_s^2})$ and $f(\boldsymbol{h}) = \frac{1}{(\pi)^{N_rM}|\boldsymbol{C}_h|}\exp(-\boldsymbol{h}^H\boldsymbol{C}_h^{-1}\boldsymbol{h})$, respectively. Thanks to this priori information, and considering the unknown parameter vector $\boldsymbol{\Theta} = [\boldsymbol{s}^H, \boldsymbol{h}^H]^H$, the Bayesian (stochastic) BCRB is given as [59]

$$BCRB_{blind} = \boldsymbol{J}_{\boldsymbol{hh}}^{-1} = \left(N\frac{\sigma_s^2}{\sigma_e^2}\boldsymbol{I}_{N_rM} + \boldsymbol{C}_h^{-1}\right)^{-1} \tag{6.21}$$

In this BCRB, the channel covariance matrix $\boldsymbol{C}_h$ is assumed to be known or estimated a proiori. Note that this bound is known to be 'optimistic' and even-though we use it in this chapter, ongoing investigations are currently done to verify whether the drawn conclusions (in section IV) are valid if other tighter bounds are used.

6.4.5 Bayesian Semi-Blind CRB

This lower bound [55] assumes both the channel and the unknown symbol sequence to be random with known a priori zero mean Gaussian distribution, hence, the parameter vector is $\boldsymbol{\Theta} = [\boldsymbol{s}_d^H, \boldsymbol{h}^H]^H$ and it is supposed to be jointly estimated.

The corresponding CRB is giving by:

$$BCRB_{SB} = \boldsymbol{J}_{\boldsymbol{hh}}^{-1} = \sigma_e^2\big(\boldsymbol{C}_p + \sigma_e^2\boldsymbol{C}_h^{-1}\big)^{-1} \tag{6.22}$$

where $\boldsymbol{C}_P = E\{\boldsymbol{\mathcal{S}}^H\boldsymbol{\mathcal{S}}\} = \boldsymbol{\mathcal{S}}_p^H\boldsymbol{\mathcal{S}}_p + N_d\sigma_s^2\boldsymbol{I}_{mL}$, N_d is the number of unknown data sequence, i.e., $N_d = N - N_p$. We assume here that $N_d \gg M$.

6.4.6 Structurally-Aided CRBs (Sparse Channel)

In this subsection, we pose the question of how would the channel sparsity structure improve the estimation process in terms of CRB. Afterward, we give an example where such a priori information are available. In general, the following model encounter the sparsity structure:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\mathcal{S}}\boldsymbol{h} + \boldsymbol{e}, \\ \boldsymbol{h} &= \boldsymbol{D}\boldsymbol{h}_s \end{aligned} \tag{6.23}$$

where $\boldsymbol{D}$ is a known (or eventually partially known) dictionary matrix which is a function of the considered side information, $\boldsymbol{h}_s$ is a vector with M_s active coefficients. Two cases will be studied whereby the channels' supports are overlapping and not. Considering deterministic scenario where the channel and source are both deterministic, the corresponding CRB is giving as follows:

$$DCRB_{sparse} = \boldsymbol{J}_{\boldsymbol{h_s h_s}}^{-1} = \sigma_e^2\big(\boldsymbol{U}_r^H\boldsymbol{D}^H\boldsymbol{\mathcal{S}}^H\boldsymbol{P}_{\boldsymbol{\mathcal{H}}}^{\perp}\boldsymbol{\mathcal{S}}\boldsymbol{D}\boldsymbol{U}_r\big)^{-1} \tag{6.24}$$

Next, we give an example of sparsness corresponding to specular channel model. Also, some other examples do exist such as fixed basis model in the case of time varying channels [122, 123].

In the so called specular multipath model, the channel $\boldsymbol{h}(t)$ in (6.2) can be modeled as a composite response of a propagation channel and the effective known pulse shape function[6] $g(t)$ [124], which encompasses the effect of both the transmitting and receiving filters. Hence, the channel response may be given by:

$$\boldsymbol{h}(t) = \sum_{k=1}^{d} \boldsymbol{a}_k g(t - \tau_k) \tag{6.25}$$

where d, τ_k and $\boldsymbol{a}_k = [a_{k1}^H, \cdots, a_{km}^H]^H$ are respectively the number of paths, unknown/known delays and the array vector attenuation factor associated with the k^{th} path. To match the model in (6.23), $\boldsymbol{h} = \boldsymbol{D}\boldsymbol{a}$, $\boldsymbol{D} = (\boldsymbol{G} \otimes \boldsymbol{I}_{N_r})$, $\boldsymbol{a} = [\boldsymbol{a}_1^H, \cdots, \boldsymbol{a}_d^H]^H$, $\otimes$ is the Kronecker product, and $\boldsymbol{G}$ is a matrix given by:

$$\boldsymbol{G} = \begin{pmatrix} g(0-\tau_1) & \cdots & g(0-\tau_d) \\ g(T-\tau_1) & \cdots & g(T-\tau_d) \\ \vdots & \ddots & \vdots \\ g((M-1)T-\tau_1) & \cdots & g((M-1)T-\tau_d) \end{pmatrix}$$

Next, we find the DCRB for two scenarios, when the model and delays are assumed to be known and when only the model is known.

[6]Usually, it corresponds to a raised cosine function.

Known Delays

in this case the parameter vector $\boldsymbol{\Theta} = [\boldsymbol{s}^H, \boldsymbol{a}^H]^H$, and the corresponding deterministic CRB is given by (6.24).

Unknown Delays

in this scenario, the parameter vector becomes $\boldsymbol{\Theta} = [\boldsymbol{s}^H, \boldsymbol{\tau}^H, \boldsymbol{a}^H]^H = [\boldsymbol{s}^H, \ddot{\boldsymbol{\Theta}}^H]^H$, $\boldsymbol{\tau}$ is the delays vector. Using the convolution property, $\boldsymbol{h}$ can be written as $\boldsymbol{h} = \boldsymbol{\mathcal{A}}\boldsymbol{g} = f(\ddot{\boldsymbol{\Theta}})$, with $\boldsymbol{\mathcal{A}} = (I_M \otimes \boldsymbol{A})$, $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_d]$, $\boldsymbol{g} = [\boldsymbol{g}(0), \cdots, \boldsymbol{g}((M-1)T)]^T$ and $\boldsymbol{g}(i) = [g(iT - \tau_1), \cdots, g(iT - \tau_d)]$. Hence, the corresponding CRB is given by:

$$CRB^{-1}_{spe}(\ddot{\boldsymbol{\Theta}}) = \boldsymbol{J}_{\ddot{\boldsymbol{\Theta}}\ddot{\boldsymbol{\Theta}}} = \frac{1}{\sigma_e^2}\boldsymbol{U}_r^H \left(\boldsymbol{J}_{22} - \boldsymbol{J}_{21}\boldsymbol{J}_{11}^{-1}\boldsymbol{J}_{12}\right)\boldsymbol{U}_r \tag{6.26}$$

where

$$\boldsymbol{J}_{11} = \boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{H}}, \qquad \boldsymbol{J}_{12} = \boldsymbol{J}_{21}^H = \left[\begin{array}{c|c} \boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{g'} & \boldsymbol{\mathcal{H}}^H\boldsymbol{\mathcal{S}}\boldsymbol{D} \end{array}\right],$$

$$\boldsymbol{J}_{22} = \left[\begin{array}{c|c} \boldsymbol{g'}^H\boldsymbol{\mathcal{A}}^H\boldsymbol{\mathcal{S}}^H\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{g'} & \boldsymbol{g'}^H\boldsymbol{\mathcal{A}}^H\boldsymbol{\mathcal{S}}^H\boldsymbol{\mathcal{S}}\boldsymbol{D} \\ \hline \boldsymbol{D}^H\boldsymbol{\mathcal{S}}^H\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{A}}\boldsymbol{g'} & \boldsymbol{D}^H\boldsymbol{\mathcal{S}}^H\boldsymbol{\mathcal{S}}\boldsymbol{D} \end{array}\right]$$

Now, the channel CRB is obtained as:

$$CRB_{spe}(\boldsymbol{h}) = \nabla f(\ddot{\boldsymbol{\Theta}})CRB_{spe}(\ddot{\boldsymbol{\Theta}})\nabla f(\ddot{\boldsymbol{\Theta}})^H \tag{6.27}$$

with $\boldsymbol{h} = \boldsymbol{D}(\tau)\boldsymbol{a} = f(\ddot{\boldsymbol{\Theta}})$, $\nabla f(\ddot{\boldsymbol{\Theta}})_{ij} = \partial h_i/\partial\theta_j = \left[\begin{array}{cc} \boldsymbol{G}, & \boldsymbol{g'}\boldsymbol{A} \end{array}\right]$ and $\boldsymbol{g'} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\tau}}$;

6.5 Discussion

The following remarks can be pointed out of this study:

- Using the available side information in any system would add more to the estimation performance. However, as shown in our simulation examples, the most significant gains are obtained for the semi-blind and the sparse channel (with known dictionary) cases.

- In the case of Bayesian scenario, using pilot symbols has insignificant effect on the estimation performance. However, we must admit that, since the considered bound is too optimistic, we are currently re-investigating this surprising result via other tighter bounds.

- In the case of channel order overestimation (OE) problem:

  - For the deterministic Semi-Blind case, the channel can still be estimated without ambiguity as long as the overestimation length (i.e. $\tilde{M} > M$) does not exceed the number of pilots.

  - Bayesian Blind and Semi-Blind estimators do not suffer from OE problem. Indeed, the channel power profile (given by matrix $\boldsymbol{C}_h$) takes into account the decaying of the channel taps power and hence the OE is not any more an issue in that case.

  - In case of sparse channels, the estimation is also robust to the non-zero entries order overestimation.

Clearly, this chapter considers extensions and analysis of some existing blind and semi-blind CRBs, meanwhile, it has some novel ones. That said, makes the picture noticeably broadened, therefore, to summarize we provide a clarification in Table 6.1.

Table 6.1: Summary of CRBs

CRB Type	**Novel**
$DCRB_{blind}$, $DCRB_{SB}$, $BCRB_{blind}$, $BCRB_{SB}$	No
$CRB_{Non-Circularity}$ Eq. (6.19)	Yes
$DCRB_{sparse}$	Yes
CRB_{spe}	Yes

6.6 Experimental Performance Analysis

The derived CRBs throughout this chapter will be illustrated in this section. The considered CRBs are averaged over several channel and signal (for the deterministic case) realizations. In each Monte-Carlo run, different realizations of the symbols, channel and noise will be generated randomly. With regards to the channel, we generate a Rayleigh fading channel with an exponentially decaying power delay profile (PDP) e^{-jwm}, where $m = 0 \cdots M-1$ and $w = 2$. Accordingly, $\boldsymbol{C}_h$ becomes a diagonal matrix $\boldsymbol{C}_h = \boldsymbol{I}_{N_r} \otimes diag\{e^{-wm}, m = 0 \cdots M-1\}$. As for the transmitted symbols, a random QPSK is considered but treated as deterministic. The performance of the different CRBs are evaluated in terms of the normalized mean squares error $NMSE$ against the SNR. Both are defined as: $SNR = E\frac{\|\boldsymbol{\mathcal{H}}\boldsymbol{s}\|^2}{mM\sigma_e^2}$ and $NMSE = \frac{avg\, tr(CRB)}{\|\boldsymbol{h}\|^2}$, where avg, tr and $\|\|$ stand for the

average, trace and norm, respectively.

Channel's OE impact

In the first experiment we shed light on the advantages and robustness/sensitivity of Blind/Semi-blind Bayesian against non-Bayesian to the channel overestimation problem. Also, we highlight the maximum allowable channel overestimation in the case of DCRB_{SB}. In figure 6.1 we plot the CRB for different scenarios to check their behavior in case of channel order OE. DCRB_{SB} is the most affected and it inflate when the length of the channel impulse response become greater than the number of known pilots i.e. $\tilde{M} > N_p$, whereas the BCRB_{SB} and BCRB_{blind} are not affected, yet their behavior is moderately affected at high SNR.

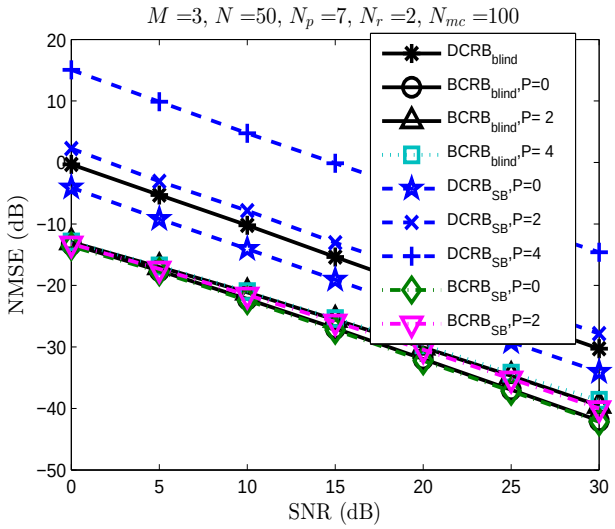


Figure 6.1: NMSE for DCRB_{blind},BCRB_{blind}, DCRB_{SB} and BCRB_{SB}.

Channel's Sparsity impact

In the second experiment, we investigate the consequence of exploiting the sparsity structure on the channel estimation. Figure 6.2 demonstrates that using the sparse structure information gives a significant gain against ignoring that structure, moreover, if there is a prior information such that if different channels overlap in time spread (i.e., have the same common supports' location) or not, would also improve the estimation further. Moreover, in the light of channels' non-zero entries OE case, figure 6.3 shows that by exploiting the sparse structure, the channel estimation is robust to the channel OE scenario.

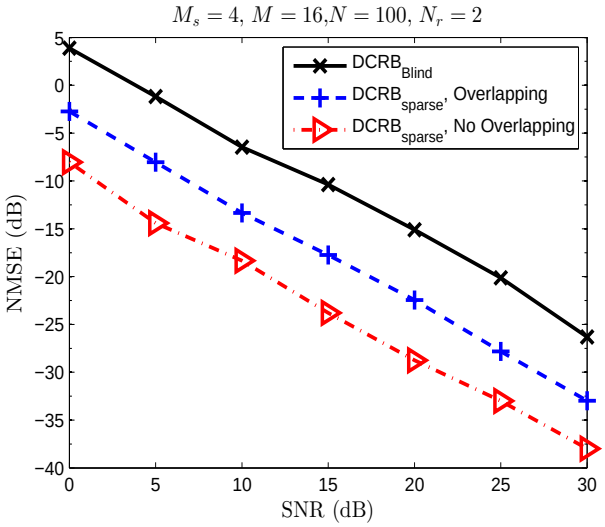


Figure 6.2: Impact of using sparse structure on the CRB, when different channels' coefficients are overlapping and not.

Channel's Structure impact

In the third experiment, a practical scenario which is the specular channel model, was tested. In this experiment, a raised-cosine pulse shaping filter $g(t)$ with 0.25 roll-off factor is used, the array response to a point source vector $\boldsymbol{a}$ is arbitrary

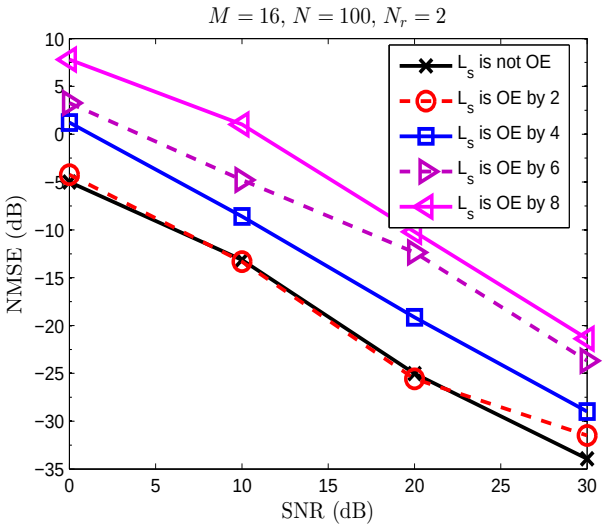


Figure 6.3: Impact of channels' non-zero entries order overestimation in sparse scenario.

generated. Figure 6.4 shows that using the specular embedded structure gives a pronounced gain in the first scenario where the dictionary is assumed to be partially unknown, i.e. τ is unknown, and this gain becomes significant when the dictionary is assumed to be fully known.

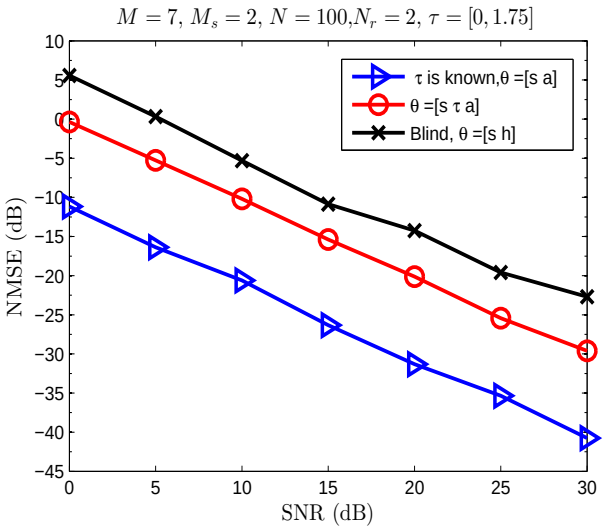
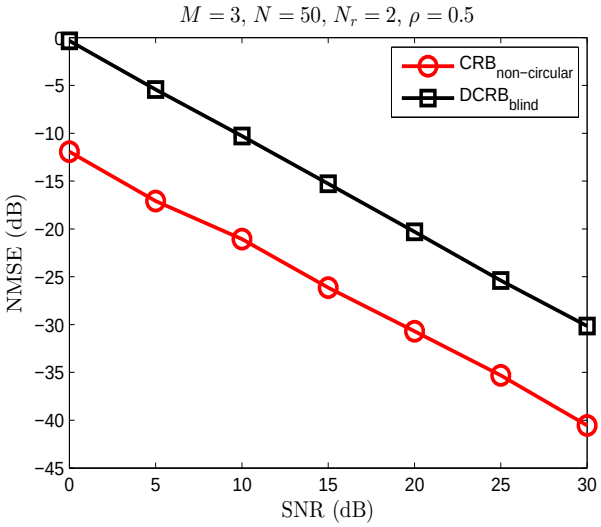


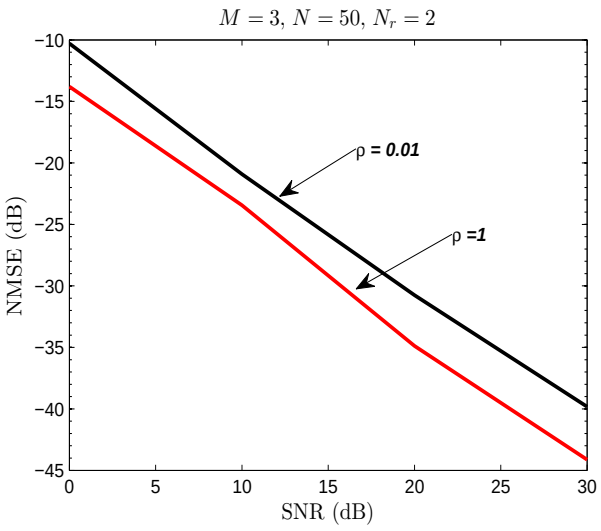
Figure 6.4: Impact of using specular channel structure on the CRB.

Non-circularity impact

In the last experiment, the impact of source non-circularity is presented in figure 6.5. The test conveys two main messages, the first one is that the non-circularity is a gainful side information. Also, the non-circularity degree has a marked impact on the channel estimation performance.



a.



b.

Figure 6.5: a. Impact of using non-circularity of the source, b. Impact of non-circularity factor $|\rho|$ on the CRB

6.7 Conclusion

This chapter addresses and quantifies the impact of exploiting some available side information on the system identification problem. We have derived the CRBs for different scenarii corresponding to the considered side information cases. Sparseness with known dictionary and pilot symbols are shown to be the one that brings the most important gain. In particular, the channel order overestimation problem is mitigated up to certain extent when considering the previous side information.

CHAPTER 7

CONCLUSION & FUTURE WORK

7.1 Conclusion

In this thesis, we try at one hand to fill some existing gaps in the literature, and on the other hand, propose some new competitive and efficient methods. Also, shed the light on some free side information which might improve the channel estimation quality. The motivations behind selecting this topic were the following:

- The shortage of the blind deconvolution and source separation methods which handle the most common type of digital signal modulation, i.e., PSK and QAM.

- The recent studies which demonstrates the superiority of the semi-blind methods over the training based ones. Moreover, the semi-blind topic is no more a theoretical one, there are different applications in the market, such

as optical communication and microwave links, which successfully adopt the semi-blind concept.

- Help in solving the problem of pilot contamination problem which is caused by the pilot propagation especially in some promising wireless system solutions such as Massive MIMO systems.

- There is an increase interest in the blind and semi-blind methods in some emerging topics such as cognitive radio and ad hoc networks whereby sharing the users' information (such as the training sequence) between the primary and secondary is undesirable.

Blind deconvolution for MIMO systems was the first problem targeted in this thesis. The new class of BSS algorithm, proposed in [6], to perform demixing for the instantaneous mixture using Unitary and non-unitary elementary rations, inspired us to generalize these methods to the more general convolutive case. Different schemes were used for this purpose:

- The first one is based on performing equalization to remove the effect of ISI followed with BSS to remove the effect of IUI.

- The second one is to perform full-BSS for the spatio-temporal model which is built from the received signals, hence after, we perform a kind of pairing, to group the separated signals with their delayed copies, and eventually, sorting these copies accordingly.

- The third one is based on optimizing hybrid criteria which penalizes the MM

criterion along with the cross-correlation among the different sources. Again, the used optimization technique is the Givens and Hyperbolic rotations.

- The last method is a deflationary-based one whereby a single source is extracted each time.

These techniques form a basement for possible solutions for extending any BSS method to cover non-instantaneous mixtures, which we believe it constitutes a good reference for researchers.

In addition to the analytic solution to the MM criterion, Givens and Shear rotations so far are the only fast and iterative solution for this problem. At this level, we raised the following question: Is there a better technique or optimization method in term of ease of implementation, performance and computational complexity to perform BSS for non-constant modulus signals?

Accordingly, in Chapter 4, inspired by the well-known fastICA and using the FPO optimization method, we proposed fast and less costly set of algorithms, namely fastMMA algorithms, to perform the BSS. Two variants were developed: the first one is a deflationary-based, while the second is a full-based, whereas the later demixes all the sources at once. Morover, the full-based variant has shown an improved quality of estimation due to relaxing the unitary constraint on the separation matrix. The deflationary based algorithm is extended to cover the convolutive mixtures by employing an appropriate subspace projection. Finally, to boost the convergence speed further, a VSS procedure is provided by deriving an algebraic optimal step size using the well-known exact line search strategy.

Interestingly, in comparison with the existing CMA and MMA algorithms, our proposed algorithms are the best in terms of computational complexity, speed of convergence, ease of derivation and implementation, and separation quality.

The second targeted problem is the BSI. Specifically, we are inspired by the following question: How would the side information improve the channel estimation quality? The answer is considered in both algorithmic and performance evaluation point of view. In Chapter 5, a dual subspace approach is proposed by penalizing the principle subspace of the data covariance matrix from deviation from the Toeplitz structure. By doing so, we show that the channel estimation is significantly improved in the difficult context of weak channel diversity (i.e., a channel with closely spaced roots).

In terms of performance evaluation, using the CRB, different statistical and structural side information has been explored to quantify their impact on the channel estimation quality. In particular, three novel deterministic and statistical CRBs, out of seven, are presented in Chapter 6. These side information has shown a potential improvement if they are used appropriately in the algorithm's design.

7.2 Future Work

It is true that we have explored different topics in this work, also it is true that different directions and open future work are possible. In the ensuing, we list some of these directions which some of them are related to the work presented in this thesis and some are not.

1. In Chapters 3 and 4, we deal with batch methods, whereby the channel is assumed to be constant during the processing data packet. However, in some communication systems which assume mobility, the channel varies within the symbol period. In such scenario, developing adaptive versions of the proposed BSS and BD algorithms would be a good solution.

2. Also, in Chapters 3 and 4, we used the non-constant modulus criterion to develop BSS and BD algorithm for QAM signals, these algorithms are shown to have good performance in case of small and intermediate constellation sizes, i.e., up to 64-QAM. One can use these algorithms and target higher constellation sizes by exploring hybrid cost functions such as MM and Alphabet Matched criteria [see for example [125, 126]] .

3. The optimization methods used in this work can be used to develop new algorithms using some other cost functions such as constant power and kurtosis criterion.

4. So far, we have seen an analytical solution for the CM criterion, namely ACMA. It would be interesting to explore the MM analytical solution, as well.

5. In this thesis, we have also focused on extracting all the sources with giving no priority to any of them. It is of great interest to develop algorithms which can extract specific user's signal (source of interest) with no need to extract all users' signals at once.

6. In Chapter 5, the new proposed SSS method experience a high computational complexity in comparison with the state of the art SS method. Thus, it is so interesting to explore different solutions to reduce its computational cost.

7. The proposed SSS method uses the linear Toeplitz structure to perform channel estimation. Following the same procedure, one can target direct equalization by working on the linear Henkel structure. Henkel structure is a dual presentation of the model (6.3) [see for example [127]].

8. In the context of subspace tracking, one interesting application is to invoke the proposed structure subspace technique along with the subspace tracking in order to estimate the communication channel directly instead of estimating a general principle subspace.

9. In Chapter 6, different CRBs are evaluated for the case of SIMO systems. Similar derivations can be extended to cover the more general MIMO systems.

APPENDIX A

Fourth order polynomial solution

In order to prove that (3.26) is a 4-th order polynomial equation. Let $\mathbf{U}$ and $\mathbf{\Lambda} = diag\left\{[\lambda_1, \lambda_2]\right\}$ be a 2×2 generalized eigenvectors and eigenvalues matrices of the matrix pair $(\mathbf{Q}, \mathbf{J}_G)$, then,

$$\left(\mathbf{Q} + \beta \mathbf{J}_G\right)^{-1} = \mathbf{U}\left(\mathbf{\Lambda} + \beta \mathbf{I}_2\right)\mathbf{U}^{-1}\mathbf{J}_G \tag{7.1}$$

plugging (7.1) into (3.26), lead to

$$\frac{1}{4}\mathbf{q}^T\mathbf{U}(\mathbf{\Lambda} + \beta\mathbf{I}_2)^{-2}\mathbf{U}^{-1}\mathbf{J}_G\mathbf{q} = \mathbf{a}^T(\mathbf{\Lambda} + \beta\mathbf{I}_2)^{-2} = 1 \tag{7.2}$$

where

$$\mathbf{a}^T = \tfrac{1}{2}\mathbf{q}^T\mathbf{U} = \begin{bmatrix} a_1 & a_2 \end{bmatrix},$$
$$\mathbf{b} = \tfrac{1}{2}\mathbf{U}^{-1}\mathbf{J}_G\mathbf{q} = \begin{bmatrix} b_1 & b_2 \end{bmatrix}^T$$

and since

$$\left(\mathbf{\Lambda} + \beta\mathbf{I}_2\right)^{-2} = diag\left\{\begin{bmatrix} (\beta+\lambda_1)^{-2} & (\beta+\lambda_2)^{-2} \end{bmatrix}\right\}$$

then, (7.2) is expressed in a more compact way as

$$\sum_{i=1}^{2} \frac{a_i b_i}{(\beta + \lambda_i)^2} = 1 \tag{7.3}$$

which is equivalent to

$$\prod_{i=1}^{2} \left((\beta + \lambda_i)^2\right) - \left(a_2 b_2 (\beta + \lambda_1)^2 + a_1 b_1 (\beta + \lambda_2)^2\right) = 0 \tag{7.4}$$

Hence, it is clear that (7.4) is a 4-th order polynomial equation of the form

$$P_4(\beta) = c_4 \beta^4 + c_3 \beta^3 + c_2 \beta^2 + c_1 \beta^1 + c_0 = 0$$

with the coefficients

$$\begin{aligned}
c_4 &= 1 \\
c_3 &= 2(\lambda_1 + \lambda_2) \\
c_2 &= 4\lambda_1\lambda_2 - \mathbf{a}^T\,\mathbf{b} + \sum_{i=1}^{2} \lambda_i^2 \\
c_1 &= 2 \sum_{i=1}^{2} \left((\lambda_i - a_i b_i) \sum_{\substack{j=1 \\ j \neq i}}^{2} \lambda_j \right) \\
c_0 &= \lambda_1^2 \lambda_1^2 - \sum_{i=1}^{2} \left(a_i b_i \prod_{\substack{j=1 \\ j \neq i}}^{2} \lambda_j^2 \right)
\end{aligned}$$

As stated earlier, the solutions of this polynomial can be found, at a negligable cost, using the analtical Ferrari or Cardano's formulas [100, 101].

APPENDIX B

Second Givens rotations $\boldsymbol{\varphi}_{p,q}^{a}\left(\theta', -\frac{\pi}{2}\right)$

The following relation are used to extract the real and imaginary parts of the rotation $\mathbf{Z} = \boldsymbol{\varphi}_{p,q}^{a}\left(\theta', -\frac{\pi}{2}\right)\underline{\mathbf{Y}}$

$$z_{p,R}(k) = \frac{z_p(k) + z_p^*(k)}{2}, \quad z_{p,I}(k) = \frac{z_p(k) - z_p^*(k)}{i2} \tag{7.5}$$

using the earlier relations and $\mathbf{Z} = \boldsymbol{\varphi}_{p,q}^{a}\left(\theta', -\frac{\pi}{2}\right)\underline{\mathbf{Y}}$ after setting $\alpha = -\frac{\pi}{2}$. One can easily shows that:

$$\begin{aligned}
z_{p,R}(k) &= \cos(\theta')\underline{y}_{p,R} + \sin(\theta')\underline{y}_{q,I} \\
z_{p,I}(k) &= \cos(\theta)\underline{y}_{p,I} - \sin(\theta')\underline{y}_{q,R} \\
z_{q,R}(k) &= \sin(\theta')\underline{y}_{p,I} + \cos(\theta')\underline{y}_{q,R} \\
z_{q,I}(k) &= -\sin(\theta')\underline{y}_{p,R} + \cos(\theta')\underline{y}_{q,I}
\end{aligned} \tag{7.6}$$

Accordingly,

$$\begin{aligned}
z_{p,R}^2(k) &= \boldsymbol{\Theta}'^T\mathbf{t}'_{pq}(k) + \tfrac{1}{2}\left(\underline{y}_{p,R}^2(k) + \underline{y}_{q,I}^2(k)\right) \\
z_{q,I}^2(k) &= -\boldsymbol{\Theta}'^T\mathbf{t}'_{pq}(k) + \tfrac{1}{2}\left(\underline{y}_{p,R}^2(k) + \underline{y}_{q,I}^2(k)\right) \\
z_{q,R}^2(k) &= \boldsymbol{\Theta}'^T\mathbf{t}'_{qp}(k) + \tfrac{1}{2}\left(\underline{y}_{q,R}^2(k) + \underline{y}_{p,I}^2(k)\right) \\
z_{p,I}^2(k) &= -\boldsymbol{\Theta}'^T\mathbf{t}'_{qp}(k) + \tfrac{1}{2}\left(\underline{y}_{q,R}^2(k) + \underline{y}_{p,I}^2(k)\right)
\end{aligned} \tag{7.7}$$

where

$$\begin{aligned}
\mathbf{\Theta}' &= \begin{bmatrix} \cos(2\theta') & \sin(2\theta') \end{bmatrix}^T, \\
\mathbf{t}'_{pq}(k) &= \begin{bmatrix} \frac{1}{2}\left(\underline{y}^2_{p,R}(k) - \underline{y}^2_{q,I}(k)\right) & \underline{y}_{p,R}(k)\underline{y}_{q,I}(k) \end{bmatrix}^T, \\
\mathbf{t}'_{qp}(k) &= \begin{bmatrix} \frac{1}{2}\left(\underline{y}^2_{q,R}(k) - \underline{y}^2_{p,I}(k)\right) & \underline{y}_{q,R}(k)\underline{y}_{p,I}(k) \end{bmatrix}^T
\end{aligned}$$

using (7.7), and after dropping constant terms that are independent of (θ'), (3.9) can be expressed in terms of (θ') as

$$\begin{aligned}
J_{MM}(\theta') &= \mathbf{\Theta}'^T \mathbf{T}' \mathbf{\Theta}' \\
&= \mathbf{\Theta}'^T \left[2 \sum_{k=1}^{K} \mathbf{t}'_{pq}(k) \mathbf{t}'^T_{pq}(k) + \mathbf{t}'_{qp}(k) \mathbf{t}'^T_{qp}(k) \right] \mathbf{\Theta}'
\end{aligned} \tag{7.8}$$

APPENDIX C

Second Shear rotations $\boldsymbol{\phi}^{a}_{p,q}\left(\gamma', -\frac{\pi}{2}\right)$

Based on the real-complex relations (7.5), one can write the following:

$$\begin{aligned} z_{p,R}(k) &= \cosh(\gamma')\underline{y}_{p,R} + \sinh(\gamma')\underline{y}_{q,I} \\ z_{p,I}(k) &= \cosh(\gamma)\underline{y}_{p,I} - \sinh(\gamma')\underline{y}_{q,R} \\ z_{q,R}(k) &= -\sinh(\gamma')\underline{y}_{p,I} + \cos(\gamma')\underline{y}_{q,R} \\ z_{q,I}(k) &= \sinh(\gamma')\underline{y}_{p,R} + \cosh(\gamma')\underline{y}_{q,I} \end{aligned} \tag{7.9}$$

Accordingly,

$$\begin{aligned} z^2_{p,R}(k) &= \boldsymbol{\Phi}'^{T}\mathbf{u}'_{pq}(k) + \tfrac{1}{2}\left(\underline{y}^2_{p,R}(k) - \underline{y}^2_{q,I}(k)\right) \\ z^2_{q,I}(k) &= \boldsymbol{\Phi}'^{T}\mathbf{u}'_{pq}(k) + \tfrac{1}{2}\left(\underline{y}^2_{p,I}(k) - \underline{y}^2_{q,R}(k)\right) \\ z^2_{q,R}(k) &= \boldsymbol{\Phi}'^{T}\mathbf{u}'_{qp}(k) + \tfrac{1}{2}\left(\underline{y}^2_{q,R}(k) - \underline{y}^2_{p,I}(k)\right) \\ z^2_{p,I}(k) &= \boldsymbol{\Phi}'^{T}\mathbf{u}'_{qp}(k) + \tfrac{1}{2}\left(\underline{y}^2_{q,I}(k) - \underline{y}^2_{p,R}(k)\right) \end{aligned} \tag{7.10}$$

where

$$\begin{aligned} \boldsymbol{\Phi}' &= \begin{bmatrix} \cosh(2\gamma') & \sinh(2\gamma') \end{bmatrix}^T, \\ \mathbf{u}'_{pq}(k) &= \begin{bmatrix} \frac{1}{2}\left(\underline{y}^2_{p,R}(k) + \underline{y}^2_{q,I}(k)\right) & \underline{y}_{p,R}(k)\underline{y}_{q,I}(k) \end{bmatrix}^T, \\ \mathbf{u}'_{qp}(k) &= \begin{bmatrix} \frac{1}{2}\left(\underline{y}^2_{p,I}(k) + \underline{y}^2_{q,R}(k)\right) & -\underline{y}_{q,R}(k)\underline{y}_{p,I}(k) \end{bmatrix}^T \end{aligned}$$

using (7.10), and after dropping constant terms that are independent of (γ'), (3.9) can be expressed in terms of (γ') as

$$\mathcal{J}(\gamma') = \mathbf{\Phi}'^T \boldsymbol{U}' \mathbf{\Phi}' \ + \mathbf{\Phi}'^T \mathbf{u}' \tag{7.11}$$

where

$$\begin{aligned}
\boldsymbol{U}' &= \sum_{k=1}^{K} \mathbf{u}'_{pq,R}(k) \boldsymbol{u}'^T_{pq,R}(k) + \mathbf{u}'_{qp}(k) \boldsymbol{u}'^T_{qp}(k), \\
\boldsymbol{u}' &= -2R \sum_{k=1}^{K} \mathbf{u}'_{pq,R}(k) + \mathbf{u}'_{qp,I}(k),
\end{aligned}$$

APPENDIX D

Step-Size Derivation

The MM criterion given in (4.4) becomes a function of the step-size μ solely, and is expressed by the rational function

$$\begin{aligned}\mathcal{J}(\mu) &= \frac{E[(\boldsymbol{z}_R - \mu\tilde{\mathbf{z}}_R)^4] + E[(\boldsymbol{z}_I - \mu\tilde{\mathbf{z}}_I)^4]}{(\parallel (\dot{\boldsymbol{v}})_{new} \parallel^2)^2} - 2 \\ &= \frac{\mathcal{P}(\mu)}{\mathcal{Q}^2(\mu)} - 2\end{aligned} \tag{7.12}$$

where $\boldsymbol{z}_R = \dot{\boldsymbol{v}}^T\underline{\dot{\mathbf{Y}}}, \tilde{\mathbf{z}}_R = \boldsymbol{g}^T\underline{\dot{\mathbf{Y}}}, \boldsymbol{z}_I = \dot{\boldsymbol{v}}^T\boldsymbol{F}_I^T\underline{\dot{\mathbf{Y}}}, \tilde{\mathbf{z}}_I = \boldsymbol{g}^T\boldsymbol{F}_I^T\underline{\dot{\mathbf{Y}}}, \mathcal{Q} = \parallel (\dot{\boldsymbol{v}})_{new} \parallel^2$. After some straightforward algebraic manipulations, the above polynomials simplify to

$$\mathcal{P}(\mu) = \sum_{i=0}^{4} p_i\mu^i, \quad \mathcal{Q}(\mu) = \sum_{i=0}^{2} q_i\mu^i \tag{7.13}$$

where

$$\begin{aligned} p_0 &= E[\boldsymbol{z}_R^4 + \boldsymbol{z}_I^4], & p_1 &= -4E[\boldsymbol{z}_R^3\tilde{\mathbf{z}}_R + \boldsymbol{z}_I^3\tilde{\mathbf{z}}_I], \\ p_2 &= 6E[\boldsymbol{z}_R^2\tilde{\mathbf{z}}_R^2 + \boldsymbol{z}_I^2\tilde{\mathbf{z}}_I^2], & p_3 &= -4E[\boldsymbol{z}_R\tilde{\mathbf{z}}_R^3 + \boldsymbol{z}_I\tilde{\mathbf{z}}_I^3], \\ p_4 &= E[\tilde{\mathbf{z}}_R^4 + \tilde{\mathbf{z}}_I^4], \end{aligned}$$

and

$$q_0 = \|\dot{\boldsymbol{v}}\|^2, \quad q_1 = -2\boldsymbol{g}^T\dot{\boldsymbol{v}}, \quad q_2 = \|\boldsymbol{g}\|^2$$

Henceforth, the gradient of $\mathcal{J}(\mu)$ in (7.12) w.r.t μ is given by

$$\nabla_\mu \mathcal{J}(\mu) = \frac{\mathcal{Q}(\mu)\mathcal{P}'(\mu) - 2\mathcal{P}(\mu)\mathcal{Q}'(\mu)}{\mathcal{Q}^3(\mu)} = \frac{\mathcal{N}(\mu)}{\mathcal{Q}^3(\mu)} \tag{7.14}$$

By combining (7.13) and (7.14), the numerator $\mathcal{N}(\mu)$ becomes a 4-th order polynomial with the following coefficients:

$$\begin{aligned} &n_0 = q_0 p_1 - 2q_1 p_0, \quad n_1 = 2q_0 p_2 - q_1 p_1 - 4q_2 p_0, \\ &n_2 = 3q_0 p_3 - 3q_2 p_1, \quad n_3 = 4q_0 p_4 + q_1 p_3 - 2q_2 p_2, \\ &n_4 = 2q_1 p_4 - q_2 p_3 \end{aligned} \tag{7.15}$$

The optimal step-size is among the real-valued roots of polynomial $\mathcal{N}(\mu)$. The optimal step size is chosen as the one leading to the minimum value of the cost function in (4.13).

REFERENCES

[1] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[2] R. v. Nee and R. Prasad, *OFDM for wireless multimedia communications*. Artech House, Inc., 2000.

[3] S. Haykin, *Blind deconvolution*. Prentice Hall, 1994.

[4] A. Ladaycia, A. Mokraoui, K. Abed-Meraim, and A. Belouchrani, "What Semi-Blind Channel Estimation Brings In Terms Of Throughput Gain?" in *in Proc. ICSPCS, Australia*, Dec. 2016.

[5] ——, "Performance Analysis of Semi-Blind Channel Estimation in Massive MIMO-OFDM Communications Systems," *submitted to IEEE Transactions on Wirelesss Communications*, 2016.

[6] S. A. W. Shah, K. Abed-Meraim, and T. Y. Al-Naffouri, "Multi-Modulus algorithms using Givens rotations for blind deconvolution of MIMO systems," in *ICASSP*.

[7] A. Ikhlef, R. Iferroujene, A. Boudjellal, K. Abed-Meraim, and A. Belouchrani, "Constant modulus algorithms using hyperbolic Givens rotations," *Signal Processing*, vol. 104, pp. 412–423, 2014.

[8] A. Mesloub, K. Abed-Meraim, and A. Belouchrani, "A new algorithm for complex non-orthogonal joint diagonalization based on shear and givens rotations," *Signal Processing, IEEE Transactions on*, vol. 62, no. 8, pp. 1913–1925, 2014.

[9] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[10] O. Shalvi and E. Weinstein, *Universal methods for blind deconvolution in Blind Deconvolution.* in Blind Deconvolution, S. Haykin, Prentice-Hall, 1994.

[11] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proceedings-IEEE*, vol. 86, pp. 1951–1968, 1998.

[12] S. Haykin, *Adaptive Filter Theory.* Upper Saddle River, NJ: PrenticeHall, 1994.

[13] Y. Sato, "A method of self-recovering equalization for multilevel amplitude-modulation systems," *Communications, IEEE Transactions on*, vol. 23, no. 6, pp. 679–682, 1975.

[14] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp. 1867–1875, 1980.

[15] J. A. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE signal processing magazine*, vol. 13, no. 3, pp. 24–42, 1996.

[16] J. Gomes and V. Barroso, "A super-exponential algorithm for blind fractionally spaced equalization," *IEEE Signal Processing Letters*, vol. 3, no. 10, pp. 283–285, 1996.

[17] C.-Y. Chi, C.-Y. Chen, C.-H. Chen, and C.-C. Feng, "Batch processing algorithms for blind equalization using higher-order statistics," *IEEE Signal Processing Magazine*, vol. 20, no. 1, pp. 25–49, 2003.

[18] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, 1991.

[19] E. de Carvalho and D. Slock, *Universal methods for blind deconvolution in Blind Deconvolution.* in Signal Processing Advances in Wireless and Mobile Communications, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Prentice Hall, 2001.

[20] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications.* Academic press, 2010.

[21] S. S. Haykin, *Unsupervised adaptive filtering: Blind source separation.* Wiley-Interscience, 2000, vol. 1.

[22] R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, R. A. Casas *et al.*, "Blind equalization using the constant modulus criterion: A review," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1927–1950, 1998.

[23] A.-J. van der Veen and A. Leshem, "Constant modulus beamforming," *Chapter*, vol. 6, pp. 299–351, 2005.

[24] A.-J. Van Der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, 1996.

[25] C. B. Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Transactions on Signal Processing*, vol. 48, no. 12, pp. 3508–3519, 2000.

[26] K. N. Oh and Y. O. Chin, "Modified constant modulus algorithm: blind equalization and carrier phase recovery algorithm," in *Communications, 1995. ICC'95 Seattle,'Gateway to Globalization', 1995 IEEE International Conference on*, vol. 1. IEEE, 1995, pp. 498–502.

[27] P. Sansrimahachai, D. Ward, and A. Constantinides, "Blind source separation for BLAST," in *Digital Signal Processing.* IEEE, 2002.

[28] S. Daumont and D. Le Guennec, "An analytical multimodulus algorithm for blind demodulation in a time-varying MIMO channel context," *International journal of digital multimedia broadcasting*, vol. 2010, 2010.

[29] C. B. Papadias, "Unsupervised receiver processing techniques for linear space-time equalization of wideband multiple input/multiple output channels," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 472–482, 2004.

[30] S. Kun and Z. Xudong, "A SE-CMA based blind equalization for MIMO systems," in *Signal Processing, 2004. Proceedings. ICSP'04. 2004 7th International Conference on*, vol. 2. IEEE, 2004, pp. 1674–1677.

[31] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach," *Signal processing*, vol. 45, no. 1, pp. 59–83, 1995.

[32] Z. Ding and T. Nguyen, "Stationary points of a kurtosis maximization algorithm for blind signal separation and antenna beamforming," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1587–1596, 2000.

[33] A.-J. van der Veen, "An adaptive version of the algebraic constant modulus algorithm [blind source separation applications]," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4. IEEE, 2005, pp. iv–873.

[34] A. Ikhlef, K. Abed-Meraim, and D. Le Guennec, "On the constant modulus criterion: a new algorithm," in *Communications (ICC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–5.

[35] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural comp.*, no. 7, pp. 1483–1492, 1997.

[36] 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.

[37] J. Treichler and M. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 420–431, 1985.

[38] R. Gooch and J. Lundell, "The CM array: An adaptive beamformer for constant modulus signals," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86.*, vol. 11. IEEE, 1986, pp. 2523–2526.

[39] A. Boudjellal, K. Abed-Meraim, A. Belouchrani, and P. Ravier, "Adaptive Constant Modulus Algorithm based oncomplex Givens rotations," in *2014 IEEE Workshop on Statistical Signal Processing (SSP)*. IEEE, 2014, pp. 165–168.

[40] C. B. Papadias and A. J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *Signal Processing Letters, IEEE*, vol. 4, no. 6, pp. 178–181, 1997.

[41] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 658–672, 1997.

[42] J. K. Tugnait, L. Tong *et al.*, "Single-user channel estimation and equalization," *IEEE Signal Processing Magazine*, vol. 17, no. 3, pp. 17–28, 2000.

[43] Y. Huang, J. Benesty, and J. Chen, "Using the Pearson correlation coefficient to develop an optimally weighted cross relation based blind SIMO identification algorithm," in *ICASSP 2009*, pp. 3153–3156.

[44] Y. Xiang, L. Yang, D. Peng, and S. Xie, "A second-order blind equalization method robust to ill-conditioned SIMO FIR channels," *Digital Signal Processing*, vol. 32, pp. 57–66, 2014.

[45] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–525, 1995.

[46] W. Kang and B. Champagne, "Subspace-based blind channel estimation: Generalization and performance analysis," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 1151–1162, 2005.

[47] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, 1995.

[48] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 661–672, 1996.

[49] H. H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: algorithms," *IEEE Transactions on Signal Processing*, vol. 45, no. 8, pp. 1919–1930, 1997.

[50] D. Slock, "Bayesian blind and semiblind channel estimation," in *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004.* IEEE, 2004, pp. 417–421.

[51] E. De Carvalho, J. Cioffi, and D. Slock, "Cramer-Rao bounds for blind multichannel estimation," in *Global Telecommunications Conference, 2000. GLOBECOM'00. IEEE*, vol. 2. IEEE, 2000, pp. 1036–1040.

[52] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *Signal Processing Letters, IEEE*, vol. 5, no. 7, pp. 177–179, 1998.

[53] J.-P. Delmas and H. Abeida, "Stochastic Crame´r-Rao bound for noncircular signals with application to DOA estimation," *Signal Processing, IEEE Transactions on*, vol. 52, no. 11, pp. 3192–3199, 2004.

[54] L. Berriche, K. Abed-Meraim, and J.-C. Belfiore, "Cramer-Rao bounds for MIMO channel estimation," in *ICASSP*, 2004.

[55] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *Signal Processing, IEEE Transactions on*, vol. 50, no. 12, pp. 3055–3069, 2002.

[56] S.-M. Omar, D. T. Slock, and O. Bazzi, "A performance of Bayesian semi-blind FIR channel estimation algorithms in SIMO systems," in *Signal Processing Conference, 2011 19th European*. IEEE, 2011, pp. 2219–2223.

[57] E. de Carvalho and D. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, vol. 2. IEEE, 1997, pp. 1624–1628.

[58] S.-M. Omar, D. Slock, and O. Bazzi, "Bayesian semi-blind FIR channel estimation algorithms in SIMO systems," in *Signal Processing Advances in Wireless Communications (SPAWC), 2011 IEEE 12th International Workshop on*. IEEE, 2011, pp. 421–425.

[59] ——, "Bayesian and deterministic CRBs for semi-blind channel estimation in SIMO single carrier cyclic prefix systems," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*. IEEE, 2011, pp. 1682–1686.

[60] E. De Carvalho and D. T. Slock, "Cramer-Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Signal Processing Advances in Wireless Communications, First IEEE Signal Processing Workshop on*. IEEE, 1997, pp. 129–132.

[61] S.-M. Omar, D. Slock, and O. Bazzi, "Further results on Bayesian and deterministic CRBs in the context of blind SIMO channel estimation," in *Telecommunications (ICT), 2012 19th International Conference on*. IEEE, 2012, pp. 1–6.

[62] ——, "Recent insights in the Bayesian and deterministic CRB for blind SIMO channel estimation," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 3549–3552.

[63] L. Tong and Q. Zhao, "Joint order detection and blind channel estimation by least squares smoothing," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2345–2355, 1999.

[64] W. H. Gerstacker and D. P. Taylor, "Blind channel order estimation based on second-order statistics," *IEEE Signal Processing Letters*, vol. 10, no. 2, pp. 39–42, 2003.

[65] J. Via, I. Santamaria, and J. Perez, "Effective channel order estimation based on combined identification/equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3518–3526, 2006.

[66] D. Kotoulas, P. Koukoulas, and N. Kalouptsidis, "Subspace projection based blind channel order estimation of MIMO systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1351–1363, 2006.

[67] S. Karakutuk and T. E. Tuncer, "Channel matrix recursion for blind effective channel order estimation," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1642–1653, 2011.

[68] A. P. Liavas, P. A. Regalia, and J.-P. Delmas, "On the robustness of the linear prediction method for blind channel identification with respect to effective channel undermodeling/overmodeling," *IEEE transactions on signal processing*, vol. 48, no. 5, pp. 1477–1481, 2000.

[69] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *Signal Processing, IEEE Transactions on*, vol. 45, no. 2, pp. 434–444, 1997.

[70] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source-separation using second-order cyclostationary statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 694–701, 2001.

[71] P. D. O'grady, B. A. Pearlmutter, and S. T. Rickard, "Survey of sparse and non-sparse methods in source separation," *International Journal of Imaging Systems and Technology*, vol. 15, no. 1, pp. 18–33, 2005.

[72] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges," in

ESANN'06 proceedings-14th European Symposium on Artificial Neural Networks. d-side publi., 2006, pp. 323–330.

[73] A. T. Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5730–5743, 2013.

[74] S. A. W. Shah, K. Abed-Meraim, and T. Y. Al-Naffouri, "Multi-Modulus algorithms using Givens rotations for blind deconvolution of MIMO systems," in *ICASSP*, 2015.

[75] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 882–895, 2005.

[76] C. H. Frazier and W. D. O'Brien, "Synthetic aperture techniques with a virtual source element," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 45, no. 1, pp. 196–207, 1998.

[77] S. Haykin, J. H. Justice, N. L. Owsley, J. Yen, and A. C. Kak, "Array signal processing," 1985.

[78] E. Ertin and L. C. Potter, "Polarimetric calibration for wideband synthetic aperture radar imaging," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 145, no. 5, pp. 275–280, 1998.

[79] R. Rajagopal and L. C. Potter, "Multivariate MIMO FIR inverses." *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, vol. 12, no. 4, pp. 458–465, 2002.

[80] T. Kailath, *Linear systems*. Prentice-Hall Englewood Cliffs, NJ, 1980, vol. 156.

[81] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 499–511, 1997.

[82] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural computation*, vol. 10, no. 2, pp. 251–276, 1998.

[83] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM journal on matrix analysis and applications*, vol. 17, no. 1, pp. 161–164, 1996.

[84] J. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 2, pp. 459–472, 1983.

[85] S. Abrar and A. K. Nandi, "Blind equalization of square-QAM signals: a multimodulus approach," *IEEE Transactions on Communications*, vol. 58, no. 6, 2010.

[86] Y. Fadlallah, A. Aïssa-El-Bey, K. Abed-Meraim, K. Amis, and R. Pyndiah, "Semi-blind source separation in a multi-user transmission system with in-

terference alignment," *IEEE wireless communications letters*, vol. 2, no. 5, pp. 551–554, 2013.

[87] A. H. Sayed, *Adaptive filters*. John Wiley & Sons, 2011.

[88] ——, *Fundamentals of adaptive filtering*. John Wiley & Sons, 2003.

[89] J.-T. Yuan and K.-D. Tsai, "Analysis of the multimodulus blind equalization algorithm in QAM communication systems," *Communications, IEEE Transactions on*, vol. 53, no. 9, pp. 1427–1431, 2005.

[90] A. Gorokhov, P. Loubaton, and E. Moulines, "Second order blind equalization in multiple input multiple output FIR systems: A weighted least squares approach," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 5. IEEE, 1996, pp. 2415–2418.

[91] Y. Inouye and R.-W. Liu, "A system-theoretic foundation for blind equalization of an FIR MIMO channel system," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 4, pp. 425–436, 2002.

[92] J. Yang, J.-J. Werner, and G. A. Dumont, "The multimodulus blind equalization and its generalized algorithms," *Selected Areas in Communications, IEEE Journal on*, vol. 20, no. 5, pp. 997–1015, 2002.

[93] X.-L. Li and W.-J. Zeng, "Performance analysis and adaptive Newton algorithms of multimodulus blind equalization criterion," *Signal Processing*, vol. 89, no. 11, pp. 2263–2273, 2009.

[94] M. Mizuno and J. Okello, "A high throughput pipelined architecture for blind adaptive equalizer with minimum latency," *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 86, no. 8, pp. 2011–2019, 2003.

[95] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.

[96] H. Gazzah and K. Abed-Meraim, "Blind ZF equalization with controlled delay robust to order over estimation," *Signal processing*, vol. 83, no. 7, pp. 1505–1518, 2003.

[97] K. Abed-Meraim and Y. Hua, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 254–258, 1997.

[98] Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "Structure-Based Subspace Method for Multi-Channel Blind System Identification," *arXiv preprint arXiv:1702.04108*, 2017.

[99] J. Liang and Z. Ding, "Blind MIMO system identification based on cumulant subspace decomposition," *IEEE transactions on signal processing*, vol. 51, no. 6, pp. 1457–1468, 2003.

[100] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C: The Art of Scientific Computing (; Cambridge," 1992.

[101] C. Lanczos, *Applied analysis*. Courier Corporation, 1988.

[102] Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "New Blind Deflation-Based Deconvolution Algorithms Using Givens and Shear Rotations," Accepted in ICC, 2017.

[103] V. Zarzoso and P. Comon, "Blind and semi-blind equalization based on the constant power criterion," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4363–4375, 2005.

[104] A. Hyv *et al.*, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[105] P. Tichavsky, Z. Koldovsky, and E. Oja, "Performance analysis of the FastICA algorithm and Cramer-Rao bounds for linear independent component analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1189–1203, 2006.

[106] N. D. Gaubitch, M. K. Hasan, and P. A. Naylor, "Generalized optimal step-size for blind multichannel LMS system identification," *IEEE Signal Processing Letters*, vol. 13, no. 10, pp. 624–627, 2006.

[107] K. I. Diamantaras and T. Papadimitriou, "An efficient subspace method for the blind identification of multichannel FIR systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5833–5839, 2008.

[108] W. Qiu and Y. Hua, "Performance comparison of three methods for blind channel identification," in *ICASSP 1996*, pp. 2423–2426.

[109] ——, "Performance analysis of the subspace method for blind channel identification," *Signal Processing*, vol. 50, no. 1, pp. 71–81, 1996.

[110] Z. Xu, "Perturbation analysis for subspace decomposition with applications in subspace-based algorithms," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2820–2830, 2002.

[111] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Processing Letters*, vol. 7, no. 3, pp. 60–62, 2000.

[112] C. Avendano, J. Benesty, and D. R. Morgan, "A least squares component normalization approach to blind channel identification," in *ICASSP 1999*, pp. 1797–1800.

[113] F. Lavancier and P. Rochet, "A general procedure to combine estimators," *Computational Statistics & Data Analysis*, vol. 94, pp. 175–192, 2016.

[114] A. K. Jagannatham and B. D. Rao, "Whitening-rotation-based semi-blind MIMO channel estimation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 3, pp. 861–869, 2006.

[115] B. M. Sadler, R. J. Kozick, and T. Moore, "Bounds on MIMO channel estimation and equalization with side information," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 4. IEEE, 2001, pp. 2145–2148.

[116] X. Wautelet, C. Herzet, and L. Vandendorpe, "Cramer Rao bounds for channel estimation with symbol a priori information," in *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications, 2005.* IEEE, 2005, pp. 715–719.

[117] B. Su and K.-H. Tseng, "Cramér-Rao bound for blind channel estimation in cyclic prefixed MIMO-OFDM systems with few received symbols," in *2014 48th Asilomar Conference on Signals, Systems and Computers.* IEEE, 2014, pp. 966–970.

[118] A. K. Jagannatham and B. D. Rao, "Complex constrained CRB and its application to semi-blind MIMO and OFDM channel estimation," in *Sensor Array and Multichannel Signal Processing Workshop Proceedings, 2004.* IEEE, 2004, pp. 397–401.

[119] L. Berriche and K. Abed-Meraim, "Stochastic Cramer-Rao bounds for semi-blind MIMO channel estimation," in *Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on.* IEEE, 2004, pp. 119–122.

[120] K. Abed-Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 694–705, 1997.

[121] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 338–342, 1988.

[122] Q. Mayyala, O. Kukrer, and A. Hocanin, "Recursive inverse basis function (RIBF) algorithm for identification of periodically varying systems," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE, 2012, pp. 919–923.

[123] J. K. Tugnait and W. Luo, "On blind identification of SIMO time-varying channels using second-order statistics," in *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, vol. 1. IEEE, 2001, pp. 747–752.

[124] L. Perros-Meilhac, É. Moulines, K. Abed-Meraim, P. Chevalier, and P. Duhamel, "Blind identification of multipath channels: A parametric subspace approach," *Signal Processing, IEEE Transactions on*, vol. 49, no. 7, pp. 1468–1480, 2001.

[125] A. Labed, T. Chonavel, A. Aïssa-El-Bey, and A. Belouchrani, "Min-norm based alphabet-matching algorithm for adaptive blind equalisation of high-

order QAM signals," *Transactions on emerging telecommunications technologies*, vol. 24, no. 6, pp. 552–556, 2013.

[126] S. A. Shah, K. Abed-Meraim, and T. Y. Al-Naffouri, "Blind Source Separation Algorithms Using Hyperbolic and Givens Rotations for High-Order QAM Constellations," *arXiv preprint arXiv:1506.06650*, 2015.

[127] E. Pité and P. Duhamel, "Bilinear methods for blind channel equalization:(no) local minimum issue," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 4. IEEE, 1998, pp. 2113–2116.

Vitae

- Name: Qadri Mayyala
- Nationality: Palestinian
- Date of Birth: 23.May.1987
- Email: *qmayaleh@hotmail.com*
- Permenant Address: Palestine, Nablus

Journal Publications

1. **Q. Mayyala**, K. Abed-Meraim, and A. Zerguine, "A Class of Multi-Modulus Blind Deconvolution Algorithms Using Hyperbolic and Givens Rotations for MIMO Systems", submitted to *IEEE Transactions on Communications*.

2. **Q. Mayyala**, K. Abed-Meraim, and A. Zerguine,"Fast Multi-Modulus Algorithms", *IEEE Signal Processing Transactions* (**Accepted with Major Revision**).

3. **Q. Mayyala**, K. Abed-Meraim, and A. Zerguine,"Structure-Based Subspace Method for Multi-Channel Blind System Identifications", *IEEE Signal Processing Letters*. (**Accepted**)

Conference Publications

1. Ahmad Khalifi, **Q. Mayyala**, Naveed Iqbal, A. Zerguine, and K. Abed-Meraim "Adaptive Algorithm Based on a New Hyperbolic Sine Cost Function", submitted to *Asilomar*, 2017.

2. U. Mansoor, **Q. Mayyala**, M. Moinuddin, and A. Zerguine, "Quasi-Newton Least-Mean Fourth Adaptive Algorithm", *Proc. of IEEE EUSIPCO*, Kos island, Greece, 2017.

3. **Q. Mayyala**, K. Abed-Meraim, and A. Zerguine, "New Blind Deflation-Based Deconvolution Algorithms Using Givens and Shear Rotations", *Proc. of IEEE ICC*, Paris, France, 2017.

4. **Q. Mayyala**, K. Abed-Meraim, and A. Zerguine, "On the Performance Evaluation of Blind System Identification in Presence of Side Information", *Proc. of IEEE IWCMC*, Valencia, Spain, 2017.

5. M. A. Kousa, A. H. Muqaibel, D. B. Williams, M. T. Alkhodary, and **Q. Mayyala**, "Developing electrical engineering course in an active cooperative learning (ACL) platform", in *International Conference on Interactive Collaborative Learning*. Springer, 2016, pp. 64-74.

6. **Q. Mayyala**, O. Kukrer, and A. Hocanin, "Recursive Inverse Basis Function (RIBF) Algorithm for Identification of Periodically Varying Systems", *European Signal Processing Conf. (EUSIPCO-2012)*, Bucharest, Romania, August 27-31, 2012.

7. **Q. Mayyala**, Aykut Hocanin, and Osman Kukrer, "Two Dimensional FRS-LMS Adaptive Filter (2D FRS-LMS)", *Proc. of IEEE SIU 2012*, Fethiye, Turkey, April 2012.