THÈSE

pour obtenir le grade de

Docteur de l'université Sorbonne Paris Nord

Discipline : "Doctorat de sciences pour l'ingénieur"

présentée et soutenue publiquement par

Ouahbi REKIK

le 08/07/2021

Conception et performance limites pour les futurs récepteurs mobiles

Directeur de thèse : **Prof. Anissa MOKRAOUI**

Co-directeur de thèse : **Prof. Karim ABED-MERAIM**

JURY

Pascal CHEVALIER,	Professeur, CNAM	Président du jury
Philippe CIBLAT,	Professeur, Telecom Paris tech	Rapporteur
Mohamed Nabil EL-KORSO,	MCF-HDR, Université Paris Nanterre	Rapporteur
Adel BELOUCHRANI,	Professeur, Ecole Nationale Polytechnique d'Alger	Examinateur
Gabriel DAUPHIN,	MCF, Université paris XIII dénommée USPN	Examinateur
Anissa MOKRAOUI,	Professeur, Université paris XIII dénommée USPN	Directrice de thèse
Karim ABED-MERAIM,	Professeur, Université d'Orléans	Co-directeur de thèse

It is with my deepest gratitude and appreciation that I dedicate this thesis

To my parents;
To my wife.

To my daughter Rana-Sirine;
To my sons Sirejeddine and Jad;
To my brother and my sister.

for their constant source of love,
support and encouragement.

Acknowledgments

> *None of us got to where we are alone. Whether the assistance we received was obvious or subtle, acknowledging someone's help is a big part of understanding the importance of saying thank you.*
>
> Harvey Mackay.

Time goes by so fast and it has been already four years since the first time I came to France to start a PhD thesis at "Université Paris VIII dénommée Université Sorbonne Paris Nord". Without the guidance of the committee members, the help from my friends and the encouragement of my family, my thesis would not have been possible. Therefore, I would like to thank all people who have contributed in a variety of ways to this dissertation.

First of all, I am so grateful to my country, Algeria, that gave me the opportunity to benefit from a Phd scholarship.

I would like to express my deepest gratitude to my doctoral supervisors Professor Anissa MOKRAOUI and Professor Karim ABED-MERAIM for their guidance and support in several years. It has been a great pleasure for me to work on an interesting PhD project, from which I have a great chance to improve my mathematical skills and work with various researchers from different institutions. I am thankful to them for listening patiently to all my questions, sharing their knowledge and giving me a lot of outstanding advice to overcome numerous obstacles.

Beside my supervisors, I am deeply thankful to Professors Phlippe CIBLAT and Mohamed Nabil EL KORSO, the two reviewers of my dissertation. It is a great honor for me to have my PhD thesis evaluated by such experts. I also want to express my gratitude to the rest of my thesis committee: Professors Pascal CHEVALIER, Adel BELOUCHRANI and Gabriel DAUPHIN, for their insightful comments and encouragements. I appreciate all their hard questions which helped me to widen my research from various aspects.

Moreover, without hesitation, I would like to thank all my friends in L2TI laboratory

(Université Paris VIII dénommée Université Sorbonne Paris Nord) for their encouragement and support. They made my stay and study in this lab more enjoyable. I am happy to share a lot of memorable moments with them.

Thank you very much, everyone !

Résumé

Les systèmes de communications MIMO (Multiple Input Multiple Output) ainsi que leur récente version à grande-échelle, appelée MIMO massive, sont considérés comme des technologies potentielles pour les standards de communication sans fil actuels et futurs, grâce à leurs puissantes capacités d'amélioration des performances. Néanmoins, afin d'exploiter pleinement tous leurs potentiels, une grande attention doit être accordée aux opérations d'identification du système et d'égalisation des canaux de transmission, qui restent une préoccupation actuelle.

Dans ce contexte, la principale contribution de cette thèse s'inscrit dans le cadre de l'identification des systèmes de communication, à travers l'estimation des canaux, ainsi que l'égalisation des canaux via les techniques de séparation de sources, et ce pour des modèles de communications linéaires et non linéaires. Ainsi, en adoptant des approches semi-aveugles, des analyses de performances ainsi que le développement d'algorithmes efficaces sont mis en avant en considérant différents contraintes/problèmes tels que la contamination des pilotes, rencontrés principalement dans les systèmes MIMO massifs, les effets des non-linéarités ainsi que les interférences inter-symboles et inter-utilisateurs .En plus, pour un meilleur gain en performance, l'accent est mise aussi sur l'exploitation des a priori sur les systèmes tels que les séquences d'entraînement (pilotes), la sparcité du canal et la structure de la matrice de données.

Pour ce faire, une analyse des performances limites est réalisée à travers les bornes de Cramèr-Rao (CRB : Cramèr-Rao Bounds); et qui démontre l'efficacité des techniques semi-aveugles pour surmonter le problème de la contamination des pilotes dans les systèmes MIMO-OFDM (MIMO Orthogonal frequency-Division Multiplexing) massifs. Cette efficacité est démontrée pour le cas de cellules synchrones et non-synchrones dans un réseau cellulaire.

Cette étude nous a motivé à adopter les approches semi-aveugles pour l'égalisation des canaux et la détection des données via des techniques de séparation de sources. Une solution basée sur les modules multiples (MM : Multi-Modulus) est proposée pour démixer des mélanges instantanés dans des systèmes MIMO massifs. En l'absence d'informations préalables sur le canal de communication, une optimisation d'une fonction coût semi-aveugle est effectuée de manière itérative avec un pas optimisé. Cette solution est ensuite étendue au cas des systèmes

de communication MIMO-OFDM massifs (modèle convolutif), ou nous estimons une matrice de séparation indépendamment de la sous-porteuse. Les deux solutions présentent des performances intéressantes sous l'effet de la contamination des pilotes.

Aussi, en considérant des a priori sur le système, tels que les pilotes; la parcimonie du canal et la structure de la matrice de données, nous mettons en avant l'estimation conjointe du canal de transmission parcimonial et des données sous forme d'un problème d'optimisation. La solution semi-aveugle proposée est basée sur l'approche d'approximation convexe successive (SCA : Successive Convex Approximation), où l'optimisation est effectuée sur un problème approximatif convexe, plutôt que sur le problème original non convexe, avec convergence garantie vers un point stationnaire.

Ensuite, en considérant des systèmes de communication multicanaux non linéaires, un estimateur de canal basé sur le maximum de vraisemblance (ML : Maximum Likelihood) est proposé. Une version aveugle est proposée en combinant une technique de sous-espace et l'algorithme EM (Expectation-Maximization), avec de nouvelles techniques de suppression d' ambiguïtés. Aussi, une version semi-aveugle est proposée, qui permet de se débarrasser de l'ambiguïté inhérente au traitement aveugle et d'obtenir de meilleures performances d'estimation du canal de transmission.

Enfin, nous initions une étude liée au potentiel de l'approche d'apprentissage profond (DL : Deep Learning) au niveau de la couche physique. Ainsi, la détection de données dans les systèmes non linéaires est traitée comme étant un problème de classification basé sur le DL. Les résultats obtenus montrent que le DL offre un potentiel prometteur pour des applications au niveau de la couche physique.

Mots Clés— MIMO/ MIMO massif, OFDM, CRB, semi-aveugle, méthode sous-espace, algorithme EM, SCA, MMA, Deep Learning

Abstract

Multiple Input Multiple Output (MIMO) communications systems as well as their recent large-scale version, called massive MIMO, are seen as potential technologies for current and future wireless communications standards, thanks to their powerful performance-enhancing capabilities. Nevertheless, in order to fully exploit all their potentials, great attention has to be given to the system identification and communications channel equalization tasks, which remain a current concern.

In this context, the main contribution of this thesis falls into the scope of communications system identification, through channel estimation, as well as channel equalization via source separation techniques, for linear and nonlinear system models. Thus, by adopting semi-blind approaches, performance analysis as well as efficient algorithms development are put forward by considering different constraints/issues such as pilot contamination, encountered mainly in massive MIMO systems, nonlinearities effects as well as inter-symbol and inter-user interference. Furthermore, for a better performance gain, emphasis is also put on the exploitation of priors on the systems such as training sequences (pilots), channel's sparsity, and data matrix structure.

To do so, a performance bounds analysis is carried out through the Cramèr-Rao Bounds (CRB); which demonstrates the effectiveness of semi-blind techniques to overcome the problem of pilot contamination in massive MIMO-OFDM (MIMO Orthogonal frequency-Division Multiplexing) systems. This effectiveness is demonstrated for the case of synchronous and non-synchronous cells in a cellular network.

This study motivated us to adopt semi-blind approaches for channel equalization and data detection via source separation techniques. A Multi-Modulus based (MM) solution is proposed for demixing instantaneous mixtures in massive MIMO systems. With no priors on the communications channel, an optimization of a semi-blind cost function is performed iteratively with an optimized step size. This solution is then extended to the case of massive MIMO-OFDM communications systems (convolutive model), where we estimate a separating matrix independently of the sub-carrier. Both solutions exhibit interesting performance under pilot contamination effect.

Also, by considering priors on the system, such as pilots; channel's sparsity and data matrix

structure, we put forward a joint sparse channel estimation and data recovery as an optimization problem. The proposed semi-blind solution is based on the Successive Convex Approximation (SCA) approach, where the optimization is performed on an approximate convex problem, rather than the original non-convex one, with guaranteed convergence to a stationary point.

Then, by considering nonlinear multichannel communications systems, a Maximum Likelihood (ML)-based channel estimator is proposed. A blind version is proposed by combining a subspace technique and the Expectation-Maximization (EM) algorithm, with new ambiguity removal techniques. Also, a semi-blind version is proposed, which helps getting rid off the inherent ambiguity of blind processing and obtaining better communications channel estimation performance.

Finally, we initiate a study related to the potential of the Deep Learning (DL) approach on the physical layer. Thus, data detection in nonlinear communications systems is treated as a DL-based classification problem. The obtained results show that DL offers promising potential for applications at the physical layer.

Keywords— MIMO/ massive MIMO, OFDM, CRB, semi-blind, subspace method, EM algorithm, SCA, MMA, Deep Learning

Contents

List of Tables

List of Figures

Abbreviations

ADMM	Alternating Direction Method of Multiplier
AWGN	Additive White Gaussian Noise
BCD	Block Coordinate Descent
BPSK	Binary Phase Shift Keying
BS	Base Station
BSS	Blind Source Separation
CG	Circular Gaussian
CMA	Constant Modulus Algorithm
CP	Cyclic Prefix
CRB	Cramér-Rao Bound
CSI	Channel State Information
DFE	Decision Feedback Equalizer
DoA	Direction-of-Arrival
DL	Deep Learning
EM	Expectation Maximization
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
HOS	Higher Order Statistics
i.i.d.	independent and identically distributed
LF	Least Fourth
LS	Least Squares
LS-DF	Least Squares-Decision Feedback
LTE	Long Term Evolution
MIMO	Multiple-Input-Multiple-Output
MISO	Multiple-Input-Single-Output

ML	Maximum Likelihood
MMA	Multi-Modulus Algorithm
MSE	Mean Squared Error
NMSE	Normalized Mean Squared Error
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
OP	Optimization Problem
PARAFAC	PARAllel FACtors decomposition
PSD	Power Spectral Density
QPSK	Quadrature Phase Shift Keying
SB	Semi-Blind
SCA	Successive Convex Approximation
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
SIMO	Single-Input-Multiple-Output
SOS	Second Order Statistics
SS	Subspace
ST-SCA	Soft Thresholding Successive Convex Approximation
TDD	Time Division Duplexing
w.r.t	with respect to
ZC	Zadoff-Chu
ZF	Zero-Forcing

Notations

x	lower case letters for scalars
$\mathbf{x}$	lower case boldface letters for (column) vectors
$\mathbf{X}$	upper case boldface letters for matrices
$\hat{\mathbf{X}}$, $\hat{\mathbf{x}}$	hat is used to denote an estimate
$\mathbf{diag}(\mathbf{x})$	diagonal matrix with entries of $\mathbf{x}$ spread along the diagonal
$\mathbf{diag}(\mathbf{X})$	operator stacking the diagonal of a matrix into a vector
$\mathrm{vec}(\mathbf{X})$	operator stacking the columns of a matrix into a vector
$(.)^T$	transpose operator
$(.)^*$	complex conjugate operator
$(.)^H$	conjugate transpose (Hermitian) operator
$(.)^\#$	pseudo-inverse operator
$tr(.)$	trace of a matrix
$E[.]$	expectation operator
$\Re(.)$, $\Im(.)$	real and imaginary parts of complex variables
$\mathcal{N}(\mu, \mathbf{C})$	Gaussian distribution with mean μ and covariance matrix $\mathbf{C}$
$\mathbf{I}_m$	identity matrix of size $m \times m$
$\mathbf{0}_{a\times b}$	all-zeros matrix of size $a \times b$
$\|.\|^2$	L2 norm
$\otimes$	Kronecker product
$\star$	convolution operator.

Chapter 1

Introduction

> *Creativity requires the courage to let go of certainties .*
>
> Erich Fromm

1.1 Overview

Mobile and wireless technologies have improved substantially [2]; allowing continuous accessibility to the user for situations where installation of physical equipment is not feasible and which require on-the-spot access to information via voice or data. Along, wireless devices become remarkably convenient and affordable, leading to the expansion of wireless communications applications. Public welfare agencies such as police, fire safety, and ambulance services are using extensively mobile computing devices for their daily tasks. Airline staff are gathering information about ticketing, flight scheduling, and luggage using wireless devices. Electronic bills enable consumers to receive and pay their bills using a phone by Wireless Application Protocol (WAP). Hand-held devices used by courier companies such as Federal Express, UPS, and DHL have adopted the wireless and mobile computing technology for parcel tracking, as well as emergency drop or pickups of shipments. Also, besides the importance of transmitting voice, data and video in real time with high accuracy, military leaders are focusing more and more on advanced wireless technologies to help create a digitally networked force, which uses wearable computers and wireless radios to help create composite view of the battlefield [3]. In the background, the same wireless appliances are improving the efficiency of logistics and maintenance personnel.

Moreover, mobile cellular communications have become the most common radio access application for wireless communications in nowadays. Indeed, such an astonishing development of the mobile cellular networks has gone through different technology generations ranging from the first genaration (1G) to the fifth one (5G) so far [4]. While the first generation (1G) of mobile communications refers to the analog mobile radio systems used in 1980s, the second generation (2G) resulted from the appearance of the digital technologies mainly the Global System for Mobile communications (GSM) and the Code Division Multiplexing Access (CDMA) [5] which, with high system capacity and quality of service, have been standardized as 2G technologies and have spread all over the world [6]. Later on appeared the standard GPRS (General Packet Radio System) as 2.5G and the EDGE (Enhanced Data Rates for Global Evolution) norm as 2.75G leading to high throughput. Such a success motivated the development of the third generation (3G) through the norm UMTS (Universal Mobile Telecommunications System) and the protocol HSDPA (High-Speed Downlink Packet Access) considered as 3.5G, which boosted new applications such as internet browsing and audio/video streaming. The fourth Generation (4G), introduced in 2012, aimed at delivering high speed communication with enhanced security to enable high definition mobile TV, video-conferencing and pervasive computing with bandwidths up to 150 Mbps. Long-Term Evolution (LTE) delivered the technological fundament to fulfill the 4G standards. But only the enhanced LTE+ or LTE Advanced (LTE-A) has met the 4G requirements. The milestone technologies of 4G systems were Multiple-Input Multiple-Output (MIMO) and Orthogonal Frequency Division Multiplexing (OFDM) techniques [7]. The 5-th generation (5G) of mobile networks is a new global wireless standard, that enables a new kind of network which is designed to connect virtually everyone and everything together including machines, objects, and devices. Indeed, 5G wireless technologies are meant to deliver higher multi-Gbps (Giga Bit per second) peak data speeds, ultra low latency, more reliability, massive network capacity with increased availability. Among the prime technologies for sustaining the requirements of 5G are the millimeter wave (mmWave) and massive MIMO systems [8].

For the purpose of increasing substantially the data throughput and radio communications reliability, multiple antennas have been used at the transmitter side (Multiple-Input Single-Output : MISO), at the receiver side (Single-Input Multiple Output : SIMO) or at both (MIMO) [9, 10]. Indeed, MIMO architectures allow more degrees of freedom provided by the spatial dimension, that can be exploited to either transmit simultaneously independent data-streams, i.e. spatial multiplexing, or perform multiple transmission of single data stream, i.e. spatial diversity. Furthermore, it has been noticed that, with a higher number of Base Station (BS)

antennas, around few hundreds compared to the classical MIMO systems (8 antennas for the LTE), important gains in spectral and energy efficiencies are obtained. Such systems are called massive MIMO or large-scale MIMO systems [11, 12, 13]. These systems hold promises of boosting system's throughput by 10 times or more while simultaneously serving tens of users in the same time-frequency resources [14], which allow enhancement of both throughput and system capacity in order to satisfy the increasing amount of data exchange and demand for quality of service for the future cellular networks.

For data modulation and multiplexing/demultiplexing, OFDM is a very efficient modulation technique that can achieve very high throughput by transmitting on a great number of carriers simultaneously. It helps improving the system robustness against frequency-selective fading channels by converting the overall channel into a number of parallel flat fading channels [15]. Besides, the OFDM eliminates the inter-symbol interference and inter-carrier interference thanks to the use of a cyclic prefix and an orthogonal transform. Moreover, the combination of MIMO technology with OFDM called MIMO-OFDM systems, has enabled high speed data transmission and broadband multimedia services over wireless links [15, 16]. Although being used successfully for LTE and LTE-Advanced, many variants of OFDM, such as Constant-Envelope OFDM (CE-OFDM) or Generalized FDM (GFDM), are proposed for 5G systems [17] in order to overcome the inherent Peak-to-Average Power Ratio (PAPR) issue.

In order to fully exploit all of the potentials offered by the aforementioned technologies, efficient system identification and channel equalization remain critical tasks. Actually, systems identification aims at estimating the Channel State Information (CSI), which is exploited for coherent detection of the transmitted signals at the receiver side. Whereas, on transmitter side, it is used to design efficient precoding schemes for inter-user interference cancellation. The task of channel equalization is usually performed to compensate the distortion of the transmitted signal imposed by the communications channel, and hence recover the transmitted data. In this context, source separation techniques, used essentially for the multi-user case, is an alternative that aims to detect the transmitted data without direct priors on CSI.

On the other hand, an important interest has been given recently to machine learning and, and perticularly Deep Learning (DL) approaches, for the purpose of obtaining gains over existing physical layer algorithms. In deed, communications is a field of rich expert knowledge focusing on modeling channels of different types, compensating for various hardware imperfections and designing optimal signaling and detection schemes for reliable data transmission. However, there still exist complex communications scenarios that are difficult to describe with tractable

mathematical models and hence, can be treated with leaned systems [18].

1.2 Communications channel estimation

Communications channel refers to the propagation environment between the transmitter and the receiver. Several channel models and channel estimation approaches have been developed and proposed in literature. Basically, two main classes of channel estimation approaches have been adopted. The first class refers to pilot-based channel estimation which is performed by using training sequences, called pilots, that are inserted in the data frames and are known *a priori* by both the transmitter and the receiver. Such an approach is the most commonly used in communications standards [19, 20] for its low computational complexity and high robustness. However, its main drawbacks are the vulnerability to the problem of pilot contamination in massive MIMO systems and the bandwidth waste induced by the use of pilots, since a relatively important number of pilots is needed for accurate CSI. In the case where no pilots are available, blind channel estimation approaches are adopted as the second class. They are fully based on the statistical properties of the unknown transmitted symbols [21, 22]. Such approaches reduce the overhead but still need a large number of data symbols. Nevertheless, each channel estimation class has its own benefits and drawbacks. Generally, the first class (i.e. pilot-based channel estimator) provides a more accurate CSI, at low computational complexity, compared to the blind estimation class. However, the later, in most cases, increases the spectral efficiency compared to the first one. Therefore, it would be advantageous to retain the benefits of the two techniques through the use of semi-blind (SB) approaches [23, 24, 25] which exploit both data and pilots to achieve the desired channel estimation performance.

1.3 Source separation for data detection

Source separation is a main field of research in signal processing, which aims at retrieving a set of statistically independent source signals from a set of observed mixtures. In digital communications, source separation is used for the task of channel equalization, in order to recover the unknown data of different users transmitted through a distorting propagation medium known as communications channel. With no training sequences nor prior knowledge on the channel, Blind Source Separations (BSS) is an efficient alternative that has been widely investigated in the literature[26] [27]. BSS processes the received signal based on *a priori* knowledge about the statistics or the nature of the transmitted signals, through the optimization of an appropriate cost function. However, when training symbols are availabkle, a semi-blind approach can be

adopted through the minimization of a semi-blind cost function [28].

1.4 Deep learning for communications systems

DL-based approaches have been successfully used in many filds espacially in computer vision and natural language processing. Indeed, while it is almost impossible to define a robust algorithm for handwritten digits or objects detection in an image due to difficulties of characterizing real-world images or languages with rigid mathematical models, it is almost trivial to design DL-based techniques that can learn to accomplish such tasks even beyond human levels of accuracy [29]. In communications, on the other hand, researchers have designed channel models and transmit signals that enable straightforward algorithms for symbol detection or system identification. However, for complex communications scenarios that are difficult to describe with tractable mathematical models[18], or for complex techniques/algorithms that are difficult to implement in practice, DL-based techniques are expected to overcome such difficulties and even yield significant improvements over classical approaches. Actually, it is believed that DL applications can be useful and insightful way of fundamentally rethinking the communications system design; and hold promise for performance improvements in the physical layer.

1.5 Thesis purpose and manuscript organization

This thesis is concerned with the channel estimation and data detection, for linear and nonlinear communications systems, by relying on mathematics-based tools then by using deep learning based approaches. More precisely, this thesis focuses on performance analysis and algorithms development when adopting semi-blind approaches; and considering different issues such as pilot contamination and nonlinearities effect. Besides, a focus is also given to the exploitation of priors on the considered communications system for the purpose of performance enhancement.

Indeed, thanks to channel reciprocity property and according to the widely accepted Time Division Duplexing (TDD) protocol, used in MIMO-OFDM and massive MIMO-OFDM systems [30, 31], CSI is estimated only during the uplink transmission, then transmitted to the different users for channel equalization during the downlink phase. Consequently, the BS estimates the CSI by exploiting the known symbols (i.e. pilots), so that the pilots used within the same cell and in the neighboring cells should be mutually orthogonal. However this necessitates a complex cell synchronization and cooperation scheme. In addition, the channel time coherence [32, 33] limits the total number of orthogonal pilots leading to the reuse of the same pilots in many neighboring cells. This phenomenon is called *pilot contamination* [34], which is one of the major

issues of massive MIMO systems that must be addressed because its effect cannot be reduced by increasing the number of BS antennas. In the current thesis, the effectiveness of semi-blind approaches is investigated, through performance limits and algorithms development, to overcome pilot contamination effect.

Also, nonlinear behaviors can be encountered in many practical situations, in which case appropriate (nonlinear) processing is needed, when such nonlinearities are too important to be disregarded [35, 36]. Indeed, because most of real-life systems are inherently nonlinear in nature, nonlinear problems have drawn important interest and extensive attention from engineers, physicists, mathematicians and many other scientists [36]. In communications systems, and due to the presence of nonlinear devices such as power amplifiers and optical equipments [37], communication channels are sometimes corrupted by nonlinear distortions such as nonlinear inter-symbol interference, nonlinear multiple access interference and nonlinear inter-carrier interference. These nonlinear distortions can significantly deteriorate the signal reception, leading to poor system performance. In order to overcome such an issue, nonlinear models are adopted to provide an accurate channel representation which allows the development of efficient signal processing techniques capable of mitigating these nonlinear distortions. Consequently, blind and semi-blind approaches are proposed in this thesis to deal with nonlinear models.

In order to achieve the aforementioned goals, the current manuscript is organized as follows:

Chapter 2 is dedicated to investigate the effectiveness of semi-blind approaches for pilot contamination mitigation, when considering massive MIMO-OFDM communications systems. A performance bound analysis is carried out by using the Cramér-Rao Bound (CRB) as a basic tool for an estimator-independent study. Two scenarios have been considered. Either the cell under test along with the adjacent cells are synchronous or not. For the case of synchronous cells, the analysis demonstrates the possibility to efficiently solve the pilot contamination problem, with semi-blind approaches, when considering a finite alphabet (non Gaussian) communications signal. However, considering only the signal's Second Order Statistics (SOS) is not enough for solving such an issue even if the semi-blind approach is adopted. Moreover, the analysis shows that it is possible to get close to the optimal performance with a semi-blind approach even if the pilots are non-orthogonal as long as they are not fully-coherent. For the asynchronous cells case, it has been demonstrated that the pilot contamination still occurs under small inter-cell delays, but can be strongly mitigated with large inter-cell delays.

Chapter 3 deals with the problem of semi-blind based demixing mixtures when considering, at first instantaneous massive MIMO communications system, then convolutive systems through

massive MIMO-OFDM communications system. In both cases, the issue of pilot contamination is taken into account. To do so, a weighted hybrid cost function is introduced based upon the Multi-Modulus (MM) criterion, applied to unknown data, and the Least Squares (LS) criterion for pilot symbols. A simple but efficient semi-blind block gradient descent procedure is put forward, in which the step size, which globally minimizes the cost function along the search direction, is algebraically computed at each iteration for each user. Besides a reduced computational complexity and an accelerated convergence, simulation results show that the proposed approach allows to mitigate the inherent ambiguity of fully-blind methods, and to withstand to the pilot contamination problem in massive MIMO systems

Chapter 4 aims to propose a semi-blind solution for joint channel estimation and data detection, when considering the sparce nature of the communications channel. In the proposed solution, an optimization problem is formulated then solved by using the successive convex approximation approach. Accordingly, the optimization is performed on an approximate convex problem, rather than the original nonconvex one. By exploiting available data (pilots) and system structure, an iterative procedure is proposed where the channel coefficients and data symbols are updated simultaneously at each iteration. Also an optimized step size, introduced according to line search procedure, is used for convergence improvement with guaranteed convergence to a stationary point. Simulation results show that the proposed solution exhibits fast convergence with very attractive channel and data estimation performance.

In chapter 5, we propose solutions for blind and semi-blind channel estimation of nonlinear multi-channel communications systems. For the system model, two nonlinearities have been considered; a quadratic and a cubic. In the blind case, a first channel estimation solution is proposed based on a subspace approach followed by an appropriate ambiguity removal method. Then, to refine this first estimate, an original maximum likelihood approach is introduced based on the Expectation-Maximization (EM) algorithm. In the semi-blind case, where both data and pilots are available, an extension of this EM-based solution is proposed. Some identifiability results and performance bounds related to the considered models (blind and semi-blind) are provided and discussed. Simulation results show that the proposed solutions exhibit very interesting channel estimation performance, with an attractive convergence speed for the EM-based iterative solution.

The aim of chapter 6 is to propose a DL-based data detection solution for nonlinear MIMO communications systems. To do so, a Neural Network (NN) is built up, trained offline with finite alphabet data, then used for online data detection. With no direct priors about the channel

impulse response nor the transmitted data, the proposed DL-based data detector can deal with the performance degradation that might emerge from nonlinear components. The simulation results show the effectiveness of the proposed solution for different nonlinear model order and with attractive accuracy and data detection performance. Moreover, such a solution is promising to overcome the inherent ambiguity/limitations of classical blind processing.

1.6 List of contributions

Based on the research work presented in this thesis, many papers have been published or submitted for publication to journals and conferences as following:

Journal papers:

1) O. Rekik, A. Ladaycia, A. Mokraoui and K. Abed-Meraim, "CRB-based performance analysis of semi-blind channel estimation for massive MIMO-OFDM systems with pilot contamination", in IET Communications, 2019, vol. 13, no 20, p. 3479-3488.

2) O. Reki, K. Abed-Meraim, M. Nait-Meziane, A. Mokraoui and N. Linh-Trung, "Maximum Likelihood based Identification for Nonlinear Multichannel Communications Systems", in Signal Processing, 2021, p. 108297.

Conference Papers:

1) O. Rekik, A. Ladaycia, K. Abed-Meraim, and A. Mokraoui, "Performance Bounds Analysis for Semi-Blind Channel Estimation with Pilot Contamination in Massive MIMO-OFDM Systems", in 26th European Signal Processing Conference (EUSIPCO), 2018, p. 1267-1271.

2) O. Rekik, A. Mokraoui, A. Ladaycia and K. Abed-Meraim, "Semi-Blind Source Separation based on Multi-Modulus Criterion: Application for Pilot Contamination Mitigation in Massive MIMO Communications Systems", in 19th IEEE International Symposium on Communications and Information Technologies (ISCIT), 2019, p. 31-35.

3) O. Rekik, K. Abed-Meraim and A. Mokraoui, "EM-based Semi-blind Channel Identification for Nonlinear MIMO Systems", accepted in 8th Seminar on Detection systems: Architectures and Technologies (DAT'2020).

4) O. Rekik, K. Abed-Meraim and A. Mokraoui, "Multi-Modulus based Semi-Blind Source Separation for MIMO-OFDM Communications Systems", in 11th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM). 2020. p. 1-5.

5) O. Rekik, K. Abed-Meraim, M. Pesavento and A. Mokraoui, "Semi-blind Sparse Channel Estimation and Data Detection by Successive Convex Approximation", in 21st IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). 2020. p. 1-5.

6) O. Rekik, K. Abed-Meraim, A. Mokraoui and M. Nait-Meziane, "Contribution à l'estimation aveugle du canal de transmission dans les systèmes SIMO non linéaires", in GRETSI 2019.

7) O. Rekik, A. Mokraoui, T. QuynhK. and K. Abed-Meraim, "Performance Bound Analysis of Side Information Effect on MIMO-OFDM Channels Semi-Blind Identification", in ASILOMAR 2021.

CHAPTER 2

Semi-blind channel estimation performance limits for massive MIMO-OFDM systems

Knowledge is the conformity of
the object and the intellect.

Averroes (Ibn Rochd)

Abstract

The aim of this chapter is to investigate, via the Cramér-Rao Bound (CRB) tool, the effectiveness of semi-blind methods for pilot contamination mitigation, when considering MIMO-OFDM communications systems. For synchronous cells, the analysis demonstrated the possibility to efficiently solve the pilot contamination problem with a finite alphabet (non Gaussian) communications signal. However, considering only the signal's Second Order Statistics (SOS) is not enough for solving such an issue even if the semi-blind approach is adopted. Moreover, it has been shown that it is possible to get close to the optimal performance with a semi-blind approach even if the pilots are non-orthogonal as long as they are not fully-coherent. For the asynchronous cells case, it has been demonstrated that the pilot contamination still occurs under small inter-cell delays, but can be strongly mitigated with large inter-cell delays as shown in[1] [38] and [2] [39].

Chapter content

[1] [38] O. Rekik, A. Ladaycia, A. Mokraoui and K. Abed-meraim "CRB-based performance analysis of semi-blind channel estimation for massive MIMO-OFDM systems with pilot contamination", in IET Communications, vol. 13, no 20,p. 3479-3488, 2019.

[2] [39] O. Rekik, A. Ladaycia, K. Abed-meraim and A. Mokraoui, "Performance bounds analysis for semi-blind channel estimation with pilot contamination in massive MIMO-OFDM systems with ", in 26th European Signal Processing Conference (EUSIPCO),2018, pp.1267-1271.

2.1 Introduction

Massive Multiple-Input Multiple-Output (MIMO) is a promising technology for the next generation cellular networks [13]. With a higher number of BS antennas (beyond 100 antennas), compared to the classical MIMO systems, massive MIMO technology has proven its ability to improve the spectral and power efficiency [40, 14]. So that, both throughput and system capacity will be highly enhanced in order to satisfy the increasing amount of data exchange and demand for quality of service for the future cellular networks [41].

In order to fully exploit all of the potentials offered by a massive MIMO system, accurate Channel State Information (CSI) is necessary. It is obtained only during the uplink transmission, thanks to the channel reciprocity property and according to the widely accepted Time Division Duplexing (TDD) protocol [41]. In that case, all users in all cells send their uplink training sequences synchronously which are used, by the BS, to estimate the uplink channels. The traditional methods used to get the CSI rely on the pilot-based channel estimation (e.g. [13]). However, due to the non-orthogonality of the pilot sequences, these methods are severely affected by what is called *pilot contamination* [34], as depicted in Figure 2.1. It is one of the major issues of massive MIMO systems that must be addressed because its effect cannot be reduced by increasing the number of BS antennas.

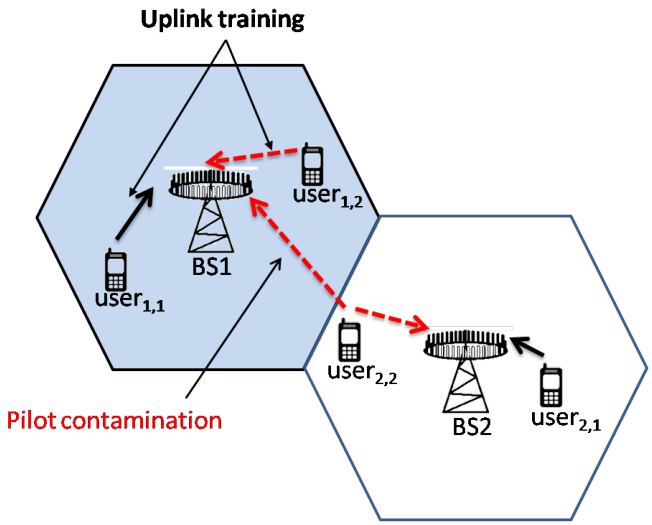


Figure 2.1: *Illustration of pilot contamination in massive MIMO-OFDM systems where* $user_{1,2}$ *and* $user_{2,2}$ *(resp.* $user_{1,1}$ *and* $user_{2,1}$*) share the same training sequence.*

Many pilot contamination mitigation strategies have been proposed [42, 43, 44]. Some of them propose to create more orthogonal pilots by slicing the time and frequency resources [45], however such a choice will lead to a system capacity decrease. Other approaches are based

on suppressing the inter-cell interference by appropriate signal processing techniques, based on statistical information of channel matrices [46, 47]. In such approaches, only a small portion of spatial dimensions is used for data transmission, whereas the unemployed dimensions will be used for suppressing noise and interference. However, many assumptions have to be considered to get statistical information of channel matrices. Instead of depending only on pilot sequences, a data-aided channel estimation has been considered (e.g. [48]), where the decoded data is used for channel estimation. Nonetheless, it requires perfect knowledge of the iner-cell large scale coefficients and it is strongly assumed to have the ability to recover most of data for accurate channel estimation. Some approaches have focused on designing appropriate inter-cell communication protocols and resource allocation [49, 50, 51] in order to allow reusing pilots without inter-cell interference. The counterpart is that the information exchange among cells will add more complexity to the cellular networks.

In recent works, a particular attention has been drawn to blind (e.g. [52, 53]), and semi-blind (e.g [54, 55]) methods. The former is fully based on the statistical properties of the transmitted data, whereas the latter depends on the joint use of pilots and data.

In addition to pilot contamination mitigation techniques, many works have focused on the effect of pilot contamination, in the case of unsynchronized BSs, on the channel estimation performance [56, 57, 58].

Consequently, the focus of this work falls into the scope of performance bounds analysis of semi-blind channel estimation approaches under the effect of pilot contamination in the context of multi-cell massive MIMO-OFDM systems. The motivation for targeting semi-blind techniques is that they allow to retain the advantages of pilot-based and blind-based approaches and hence lead to better estimation accuracy and more robustness against pilot contamination.

In [39], a brief analysis of the semi-blind (SB) channel estimation performance bounds has been initiated in certain simplified scenarios. Here, this draft work is extended to provide a full Cramér-Rao Bound (CRB) based analysis of the pilot contamination effect and how it is mitigated in such SB context. More precisely, the main contributions are as follows:

- Unlike prior works (e.g. [54, 55]) that focused on particular estimators for either the synchronized or the unsynchronized cells cases, the current work is an estimator-independent performance analysis, where the Cramér-Rao Bound (CRB) is derived for the two previous cases when considering the pilot based or the semi-blind channel estimation. It is worth noting that a thorough study has been conducted in [59] where the achievable performance of semi-blind approaches, compared to pilot-based ones, has been quantified for channel

estimation in a single cell MIMO-OFDM system. In the current study, a multi-cell massive MIMO-OFDM system is considered, where the phenomenon of pilot contamination is taken into account and thoroughly investigated.

- A thorough study of the channel estimation non-identifiability caused by the pilot contamination is given, leading to three propositions which describe such a phenomenon.
- The analysis has been carried out by taking into account two types of data statistics: only the Second Order Statistics (SOS) by considering a Gaussian source signal, and Higher Order Statistics (HOS) by using a finite alphabet signal. Besides, two types of pilots have been used: Zadoff-Chu sequences and randomly generated i.i.d. pilots.
- Compared to prior works [56, 58] about the unsynchronized cells case, the influence of the delay between the cell of interest and the neighboring cells, on the pilot contamination problem, has been investigated by differentiating the effect of small delays from the one of large delays.
- In addition to the previous CRB-based theoretical study and due to the heavy computational cost of the FIM derivation, the asynchronous case has been investigated through the use of a Least-Squares Decision Feedback (LS-DF) semi-blind channel estimator. This last analysis demonstrates that large inter-cell delays might be sufficient to mitigate the pilot contamination problem.

Practically, such performance limits analysis helps understanding the pilot contamination effect and can be exploited as benchmark by researchers developing channel estimators for massive MIMO communications systems. Moreover, the different scenarios considered (data models, pilots models and orthogonality levels) can efficiently guide developers of communications systems for the channel estimation task.

2.2 Massive MIMO-OFDM system model

This section presents the adopted massive MIMO-OFDM wireless system model as illustrated in Figure 2.2. It is worth noting that vectors' and matrices' indices represent respectively the cell and the receiver/transmitter that corresponds to the received/transmitted symbols. For the channel taps, the indices indicate respectively the cell, the transmitter and the receiver.

An uplink transmission is considered. The system is composed of N_c cells each having one BS with N_r antennas and N_t randomly located users using each a single antenna.

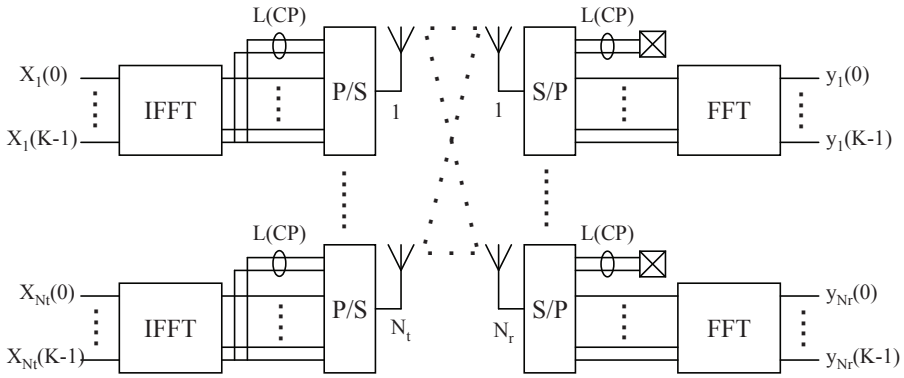


Figure 2.2: *MIMO-OFDM communications system.*

Let's ignore at first the received signals from the adjacent cells. Therefore the received signal, after cyclic prefix removal and FFT, at the r-th BS antenna of the l-th cell, assumed to be a K sub-carriers OFDM signal ($K \times 1$), is given by [60]:

$$\mathbf{y}_{l,r} = \sum_{i=1}^{N_t} \mathbf{F}\, \boldsymbol{\mathcal{T}}(\mathbf{h}_{l,i,r}) \frac{\mathbf{F}^H}{K} \mathbf{x}_{l,i} + \mathbf{v}_{l,r}, \tag{2.1}$$

where K is the OFDM symbol length; $\mathbf{F}$ represents a K-point Fourier matrix; $\mathbf{h}_{l,i,r}$ is a $N \times 1$ vector representing the channel taps between the i-th user, of the l-th cell, and the r-th receive antenna; $\boldsymbol{\mathcal{T}}(\mathbf{h}_{l,i,r})$ is a circulant matrix of dimension $K \times K$ so that its first row is given by $[h_{l,i,r}(0), \mathbf{0}_{1\times K-N}, h_{l,i,r}(N-1), \ldots, h_{l,i,r}(1)]$ while the others are obtained by a simple cyclic shift to the right of the previous one. $\mathbf{x}_{l,i}$ is a vector of size $K \times 1$ which stands for the i-th user OFDM symbol of cell l. $\mathbf{v}_{l,r}$, of size $K \times 1$, is assumed to be an additive white Circulant Gaussian (CG) noise so that $E[\mathbf{v}_{l,r}(k)\mathbf{v}_{l,r}(i)^H] = \sigma^2_{\mathbf{v}_l} \mathbf{I}_K \delta_{ki}$ where $\sigma^2_{\mathbf{v}_l}$ is the noise variance at the l-th cell; δ_{ki} being the Kronecker delta operator.

Using the eigenvalue decomposition of the circulant matrix $\boldsymbol{\mathcal{T}}(\mathbf{h}_{l,i,r})$ given by:

$$\boldsymbol{\mathcal{T}}(\mathbf{h}_{l,i,r}) = \frac{\mathbf{F}^H}{K} \boldsymbol{\lambda}_{l,i,r} \mathbf{F}, \tag{2.2}$$

where $\boldsymbol{\lambda}_{l,i,r}$ is a $K \times K$ diagonal matrix formed by the frequency gains of the channel at the considered subcarriers, i.e. $\boldsymbol{\lambda}_{l,i,r} = diag\{\mathbf{W}\mathbf{h}_{l,i,r}\}$ and $\mathbf{W}$ is formed by the N first columns of $\mathbf{F}$, and by stacking all the data in a single vector form, the received signal, of dimension $N_r K \times 1$, at the l-th BS can be re-expressed as follows:

$$\mathbf{y}_l = \boldsymbol{\lambda}_l \mathbf{x}_l + \mathbf{v}_l, \tag{2.3}$$

where $\mathbf{y}_l = [\mathbf{y}^T_{l,1} \ldots \mathbf{y}^T_{l,N_r}]^T$; $\mathbf{x}_l = [\mathbf{x}^T_{l,1} \ldots \mathbf{x}^T_{l,N_t}]^T$; $\mathbf{v}_l = [\mathbf{v}^T_{l,1} \ldots \mathbf{v}^T_{l,N_r}]^T$; $\boldsymbol{\lambda}_l = [\boldsymbol{\lambda}_{l,1} \ldots \boldsymbol{\lambda}_{l,N_t}]$ with $\boldsymbol{\lambda}_{l,i} = [\boldsymbol{\lambda}_{l,i,1} \ldots \boldsymbol{\lambda}_{l,i,N_r}]^T$.

In order to facilitate the derivation of the CRB w.r.t. $\mathbf{h}_1$, equation (2.3) is rewritten as follows:

$$\mathbf{y}_l = \tilde{\mathbf{X}}_l \mathbf{h}_l + \mathbf{v}_l, \tag{2.4}$$

where $\mathbf{h}_l = [\mathbf{h}_{l,1,1}^T ... \mathbf{h}_{l,N_t,1}^T ..... \mathbf{h}_{l,1,N_r}^T ... \mathbf{h}_{l,N_t,N_r}^T]^T$ is a $N_r N_t N \times 1$ vector; $\tilde{\mathbf{X}}_l = \mathbf{I}_{N_r} \otimes \mathbf{X}_l$ is a $N_r K \times N_r N_t N$ dimensional matrix. where $\mathbf{X}_l = [\mathbf{X}_{l,D_1}\mathbf{W} ... \mathbf{X}_{l,D_{N_t}}\mathbf{W}]$ of size $K \times N_t N$, and $\mathbf{X}_{l,D_i}$ is a $K \times K$ diagonal matrix containing the i-th user symbols, i.e. $\mathbf{X}_{l,D_i} = diag(\mathbf{x}_{l,i})$, and $\otimes$ refers to the Kronecker product.

Now, let's take into account the effect of the neighboring cells on the first one, considered without loss of generality as the interest cell. With the assumption of perfect synchronization between the N_c cells, equation (2.3) becomes:

$$\mathbf{y}_1 = \sum_{l=1}^{N_c} \boldsymbol{\lambda}_l \mathbf{x}_l + \mathbf{v}_1 = \boldsymbol{\lambda}_{tot} \mathbf{x}_{tot} + \mathbf{v}_1, \tag{2.5}$$

where $\boldsymbol{\lambda}_{tot} = [\boldsymbol{\lambda}_1 \dots \boldsymbol{\lambda}_{N_c}]$ and $\mathbf{x}_{tot} = [\mathbf{x}_1^T \dots \mathbf{x}_{Nc}^T]^T$.
Similarly to (2.4), equation (2.5) can be rewritten as follows:

$$\mathbf{y}_1 = \sum_{l=1}^{N_c} \tilde{\mathbf{X}}_l \mathbf{h}_l + \mathbf{v}_1 = \tilde{\mathbf{X}}_{tot} \mathbf{h}_{tot} + \mathbf{v}_1, \tag{2.6}$$

where $\tilde{\mathbf{X}}_{tot} = [\tilde{\mathbf{X}}_1 \dots \tilde{\mathbf{X}}_{N_c}]$ and $\mathbf{h}_{tot} = [\mathbf{h}_1^T \dots \mathbf{h}_{N_c}^T]^T$.

2.3 Effect of pilot contamination with perfectly synchronized cells

In the following section, the effect of pilot contamination on the performance of semi-blind channel estimation approaches is investigated, under the assumption of perfectly synchronized BSs of the different N_c cells. In such a case, and with same pilots in all cells, the worst case of pilot contamination occurs as explained next.

2.3.1 Pilot contamination effect

This subsection discusses the impact of the pilot contamination in a massive MIMO-OFDM system. During the uplink data transmission, the BS has to learn the transmission channel by exploiting the known symbols (i.e. pilots) at the uplink. To adopt this strategy the pilots used within the same cell and in the neighboring cells should be mutually orthogonal. However this necessitates a complex cell synchronization and cooperation scheme. In addition, the channel time coherence [32, 33] limits the total number of orthogonal pilots leading to the reuse of the same pilots in many neighboring cells. The worst case occurs when the same set of pilots is

reused in all N_c adjacent cells. In this situation, equation (2.6) becomes:

$$\mathbf{y}_1 = \sum_{l=1}^{N_c} \tilde{\mathbf{X}}_{1_P} \mathbf{h}_l + \mathbf{v}_1 = \tilde{\mathbf{X}}_{1_P} \sum_{l=1}^{N_c} \mathbf{h}_l + \mathbf{v}_1, \tag{2.7}$$

where $\tilde{\mathbf{X}}_{1_P}$ corresponds to the pilot symbols of the first cell.

To illustrate the pilot contamination effect in that case, the Least Squares (LS) estimate of the first cell channel vector, i.e. $\mathbf{h}_1$, is given by:

$$\hat{\mathbf{h}}_1^{LS} = \tilde{\mathbf{X}}_{1_P}^{\#} \mathbf{y}_1 = \mathbf{h}_1 + \sum_{l=1, l\neq 1}^{N_c} \mathbf{h}_l + \tilde{\mathbf{X}}_{1_P}^{\#} \mathbf{v}_1, \tag{2.8}$$

with $\tilde{\mathbf{X}}_{1_P}^{\#} = (\tilde{\mathbf{X}}_{1_P}^H \tilde{\mathbf{X}}_{1_P})^{-1} \tilde{\mathbf{X}}_{1_P}^H$ is the pseudo inverse of $\tilde{\mathbf{X}}_{1_P}$.

This equation clearly shows that the channel estimate $\hat{\mathbf{h}}_1^{LS}$ is affected by an additional bias corresponding to the sum of channel components of the users sharing the same pilot sequences in different cells. This phenomenon, referred to as pilot contamination, severely degrades the channel estimation performance. To overcome this problem, an alternative solution consists of using a semi-blind channel estimation approach. In the sequel, the potential of this approach is analyzed and discussed through the use of the CRB tool.

2.3.2 Cramér-Rao Bound derivation

Before deriving the CRB for pilot-based and semi-blind channel estimation, it is worthwhile to remind that the CRB expresses a lower bound on the variance of any unbiased estimator. Thus, If $\hat{\boldsymbol{\theta}} = [\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_d]^T$ is an unbiased estimator of $\boldsymbol{\theta}$, then $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \geq \mathrm{CRB}(\boldsymbol{\theta})$ in the sense $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) - \mathrm{CRB}(\boldsymbol{\theta})$ is a positive semi-definite matrix (i.e. with non negative eigenvalues). In particular, this inequality implies that the estimation error variance of parameter $\boldsymbol{\theta}_i$ is lower bounded by the i-th diagonal entry of the CRB matrix, i.e. $\mathrm{var}(\hat{\boldsymbol{\theta}}_i) \geq \mathrm{CRB}(\boldsymbol{\theta})_{i,i}$. In practice, such a tool provides a benchmark for unbiased estimators and alerts us to the physical impossibility of finding an estimator whose variance is less than the theoretical bound. Basically, the CRB is obtained as the inverse of the Fisher Information Matrix (FIM) [61]. The latter is denoted by $\mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ where $\boldsymbol{\theta}$ is the unknown deterministic parameters vector to be estimated. For the complex valued channel taps, the parameters vector $\boldsymbol{\theta}$ is defined as follows:

$$\boldsymbol{\theta} = [\mathbf{h}_{tot}^T \ \ (\mathbf{h}_{tot}^*)^T]^T, \tag{2.9}$$

where, for simplicity, the signal and noise powers are assumed to be known. The FIM, taking into account the pilots and data (that are statistically independent), is then expressed as follows:

$$\mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}^p + \mathbf{J}_{\boldsymbol{\theta}\boldsymbol{\theta}}^d, \tag{2.10}$$

where $\mathbf{J}^{p}_{\boldsymbol{\theta\theta}}$ is the FIM associated to the known pilots while $\mathbf{J}^{d}_{\boldsymbol{\theta\theta}}$ is related to the unknown data.

A block-type pilot arrangement, as described in Figure 2.3, is adopted for this work. In that scheme all sub-carriers are used for pilots within a specific period of time. For a pilot-based channel estimation, N_p pilot symbols will be considered. N_d i.i.d data symbols will be added to the pilots for semi-blind approaches. Both pilots and data are assumed to be OFDM symbols of size K.

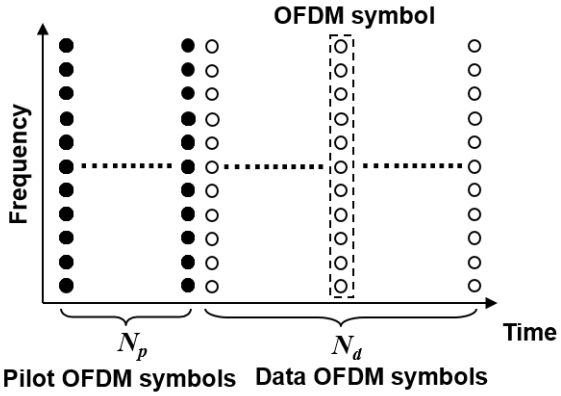


Figure 2.3: *Block-type pilot arrangement*

2.3.2.1 CRB for pilot-based channel estimation

The noise components are assumed to be independent identically distributed (i.i.d.), and only N_p pilots are used for channel estimation. Based on the data model, the pilot-based FIM can be expressed by:

$$\mathbf{J}^{p}_{\boldsymbol{\theta\theta}} = \sum_{i=1}^{N_p} \mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}}, \tag{2.11}$$

with $\mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}}$ is the FIM associated to the p_i-th pilot symbol.

The FIM for a complex parameter has been discussed in [62, 63], thus, it can be shown that the pilot-based FIM is given for the pilot-based channel estimation case by:

$$\mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}} = \begin{pmatrix} \mathbf{J}^{p_i}_{\mathbf{h}_{tot}\mathbf{h}_{tot}} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}^{p_i}_{\mathbf{h}^{*}_{tot}\mathbf{h}^{*}_{tot}} \end{pmatrix}, \tag{2.12}$$

where $\mathbf{J}^{p_i}_{\mathbf{h}^{*}_{tot}\mathbf{h}^{*}_{tot}} = (\mathbf{J}^{p_i}_{\mathbf{h}_{tot}\mathbf{h}_{tot}})^{*}$.

By considering a massive MIMO-OFDM system with N_c cells, the pilot-based FIM associated to the channel vector $\mathbf{h}_{tot}$ is then expressed as follows:

$$\mathbf{J}^{p_i}_{\mathbf{h}_{tot}\mathbf{h}_{tot}} = \frac{\tilde{\mathbf{X}}^{H}_{tot,p_i}\tilde{\mathbf{X}}_{tot,p_i}}{\sigma^2_{\mathbf{v}_1}}, \tag{2.13}$$

which can also be written in a more detailed form:

$$\mathbf{J}^{p_i}_{\mathbf{h}_{tot}\mathbf{h}_{tot}} = \frac{1}{\sigma^2_{\mathbf{v}_1}} \begin{pmatrix} \tilde{\mathbf{X}}^H_{1_{p_i}} \tilde{\mathbf{X}}_{1_{p_i}} & \dots & \tilde{\mathbf{X}}^H_{1_{p_i}} \tilde{\mathbf{X}}_{N_{cp_i}} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{X}}^H_{N_{cp_i}} \tilde{\mathbf{X}}_{1_{p_i}} & \dots & \tilde{\mathbf{X}}^H_{N_{cp_i}} \tilde{\mathbf{X}}_{N_{cp_i}} \end{pmatrix}. \tag{2.14}$$

Ideally, if the pilots of the cells are mutually orthogonal, i.e. $\tilde{\mathbf{X}}^H_{i,p_i} \tilde{\mathbf{X}}_{j,p_i} = \mathbf{0} \ \forall \ i \neq j$, then the FIM becomes a bloc diagonal matrix which is the most favorable case. On the other hand, if the cells share the same set of pilots, i.e. the worst case of pilot contamination, the FIM is then equivalent to:

$$\mathbf{J}^{p_i}_{\mathbf{h}_{tot}\mathbf{h}_{tot}} = \frac{1}{\sigma^2_{\mathbf{v}_1}} \begin{pmatrix} \tilde{\mathbf{X}}^H_{1_{p_i}} \tilde{\mathbf{X}}_{1_{p_i}} & \dots & \tilde{\mathbf{X}}^H_{1_{p_i}} \tilde{\mathbf{X}}_{1_{p_i}} \\ \vdots & \ddots & \\ \tilde{\mathbf{X}}^H_{1_{p_i}} \tilde{\mathbf{X}}_{1_{p_i}} & \dots & \tilde{\mathbf{X}}^H_{1_{p_i}} \tilde{\mathbf{X}}_{1_{p_i}} \end{pmatrix}. \tag{2.15}$$

To compute the CRB, the FIM has to be inverted. However, according to this last equation, $\mathbf{J}^{p_i}_{\boldsymbol{h}_{tot}\boldsymbol{h}_{tot}}$, and consequently $\mathbf{J}_{\boldsymbol{h}_{tot}\boldsymbol{h}_{tot}}$, is not a full rank matrix. In fact, according to proposition1, the kernel of this FIM is of dimension $2(N_c - 1)N_t N_r N$, corresponding to the number of indeterminacies we need to get rid of. In other words, this translates the *non-identifiability* of the channel vector of the interest cell when pilot contamination occurs.

Proposition 1. *The FIM in (11) is a singular matrix and its kernel dimension is* $2(N_c-1)N_tN_rN$ *which corresponds to the number of indeterminacies of the problem (i.e. the number of unknown real channel parameters for the* $N_c - 1$ *neighboring cells).*

Proof:

The FIM kernel dimension corresponds to the number of indeterminacies we need to remove (or equivalently the number of constraints we need to consider) to achieve full identifiability.

In the case of only pilots channel estimation in the presence of pilot contamination, the only parameters vector that can be estimated without bias is $\mathbf{h}_{tot} = \sum_{i=1}^{N_c} \mathbf{h}_i$.

Now, from $\mathbf{h}_{tot}$ one is able to determine every single channel $\mathbf{h}_i, i = 1, ...N_c$ iff $(N_c - 1)$ channel vectors are known (besides $\mathbf{h}_{tot}$). Since each channel vector is complex valued and of size N_tN_rN, this corresponds to $2(N_c-1)N_tN_rN$ unknown real-valued parameters needed for full identifibility. □

2.3.2.2 CRB for semi-blind channel estimation

This section is devoted to the derivation of the CRB for the semi-blind channel estimation for a multi-cell massive MIMO-OFDM system with pilot contamination. Both pilots and data

are taken into account in the derivation of the FIM as shown in equation (2.10). At first, we investigate the performance bounds of the semi-blind scheme when only the Second Order Statistics (SOS) are considered. For that, a Circular Gaussian data model is used. Latter on, we extend this analysis to the case where information based on Higher Order Statistics (HOS) is available. This will be illustrated by using a finite alphabet source signal.

2.3.2.3 Gaussian source signal

As mentioned previously, only the SOS, corresponding to the Gaussian CRB, are considered here. Hence, it assumed assume that the data symbols are i.i.d. circular Gaussian distributed with zero mean and a diagonal covariance matrix composed of the users' transmit powers i.e. $\mathbf{C}_{\mathbf{x_l}} = diag(\sigma^2_{x_{l,i}})$ with $l = 1...N_c$ and $i = 1...N_t$. Under this assumption, the received signal $\mathbf{y}_1$ is circular Gaussian with covariance matrix:

$$\mathbf{C}_{\mathbf{y}_1} = \sum_{l=1}^{N_c} \boldsymbol{\lambda}_l \mathbf{C}_{\mathbf{x}_l} \boldsymbol{\lambda}_l^H + \sigma^2_{\mathbf{v}_1} \mathbf{I}_{KN_r}. \tag{2.16}$$

The data-based FIM can be expressed as follows (e.g.[64], [65]):

$$\mathbf{J}^d_{\mathbf{h}_{tot}\mathbf{h}_{tot}} = \begin{pmatrix} \mathbf{J}^d_{\mathbf{h}_{tot}\mathbf{h}_{tot}} & \mathbf{J}^d_{\mathbf{h}_{tot}\mathbf{h}^*_{tot}} \\ \mathbf{J}^d_{\mathbf{h}^*_{tot}\mathbf{h}_{tot}} & \mathbf{J}^d_{\mathbf{h}^*_{tot}\mathbf{h}^*_{tot}} \end{pmatrix}, \tag{2.17}$$

where $\mathbf{J}^d_{\mathbf{h}_{tot}\mathbf{h}_{tot}}$ is a $(N_c N_r N_t N)$-dimensional matrix with elements $J^d_{h_i h_j}$ given by:

$$J^d_{h_i h_j} = tr\left\{ \mathbf{C}^{-1}_{\mathbf{y}_1} \frac{\partial \mathbf{C}_{\mathbf{y}_1}}{\partial h_i^*} \mathbf{C}^{-1}_{\mathbf{y}_1} \left(\frac{\partial \mathbf{C}_{\mathbf{y}_1}}{\partial h_j^*} \right)^H \right\}. \tag{2.18}$$

The i-th component of the vector $\mathbf{h}_{tot}$ corresponds to the channel tap of indices $\{i_{N_c}, i_{N_t}, i_{N_r}, i_N\}$ associated to the cell, the user, the BS antenna and the time lag of h_i. Based on the results provided in [61], $J^d_{h_i h_j}$ is given by:

$$J^d_{h_i h_j} = (J^d_{h_i^* h_j^*})^* = tr\{ \mathbf{C}^{-1}_{\mathbf{y}_1} \sigma^2_{i_{N_c}, i_{N_t}} \boldsymbol{\lambda}_{i_{N_c}, i_{N_t}} \frac{\partial \boldsymbol{\lambda}^H_{i_{N_c}, i_{N_t}}}{\partial h_i^*} \mathbf{C}^{-1}_{\mathbf{y}_1} \sigma^2_{j_{N_c}, j_{N_t}} \frac{\partial \boldsymbol{\lambda}_{j_{N_c}, j_{N_t}}}{\partial h_j} \boldsymbol{\lambda}^H_{j_{N_c}, j_{N_t}} \} \tag{2.19}$$

and

$$J^d_{h_i h_j^*} = (J^d_{h_i^* h_j})^* = tr\{ \mathbf{C}^{-1}_{\mathbf{y}_1} \sigma^2_{i_{N_c}, i_{N_t}} \boldsymbol{\lambda}_{i_{N_c}, i_{N_t}} \frac{\partial \boldsymbol{\lambda}^H_{i_{N_c}, i_{N_t}}}{\partial h_i^*} \mathbf{C}^{-1}_{\mathbf{y}_1} \sigma^2_{j_{N_c}, j_{N_t}} \boldsymbol{\lambda}_{j_{N_c}, j_{N_t}} \frac{\partial \boldsymbol{\lambda}^H_{j_{N_c}, j_{N_t}}}{\partial h_j^*} \} \tag{2.20}$$

It is important to notice that using a semi-blind estimation method with only the SOS of the received data is not sufficient to alleviate the pilot contamination problem. Indeed, the SOS-SB scheme reduces the number of indeterminacies but does not get rid of all of them. More precisely, we have the following proposition:

Proposition 2. *The FIM in (17) is a singular matrix and, in the case $N_r > N_cN_t$, its kernel dimension is $(N_cN_t)^2$ correponding to the number of indeterminacies in the blind channel estimation case. When considering the SOS-based semi-blind channel estimation, the kernel dimension of the FIM in (10) becomes $((N_c-1)N_t)^2$.*

Proof:

Considering the data only first (i.e. blind context), it is known that if the $N_r \times (N_cN_t)$ channel transfer function is irreductible, then one can estimate the channel parameters using the Second Order Statistics (SOS) up to an $(N_cN_t) \times (N_cN_t)$ unknown constant matrix [66, 67].

Now, since we assumed the source power known, the latter indeterminacy reduces to an unknown $(N_cN_t) \times (N_cN_t)$ unitary matrix, which can be modeled by $(N_cN_t)^2$ free real angle parameters. Somehow, the data SOS allows us to reduce the convolution model into an instantaneous (N_cN_t) dimensional linear mixture model.

Finally, as in the only pilots case, due to the pilot contamination, the only way to complete the channel identification via the pilot use, is to have (know) the space directions of the interfering users of the neighboring cells corresponding to $((N_c-1)N_t)^2$ real parameters to determine. □

2.3.2.4 Finite alphabet source signal

Here, the non Gaussian nature of communications signals is considered through the use of a finite alphabet data model (QPSK). The observed signal at the k-th sub-carrier is given by:

$$\mathbf{y}_{1(k)} = \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}^{\frac{1}{2}} \mathbf{x}_{(k)} + \mathbf{v}_{1(k)} \quad \text{for} \quad k = 1, ..., K, \tag{2.21}$$

where $\boldsymbol{\lambda}_{tot_{(k)}}$ is the k-th Fourier component of $\mathbf{h}_{tot}$; $\mathbf{C}_{\mathbf{x}}$ is a block diagonal matrix formed by users' transmit powers of each cell; $\mathbf{x}_{(k)} = [\mathbf{x}_{1,(k)}^T ... \mathbf{x}_{N_c,(k)}^T]^T$ with $\mathbf{x}_{l,(k)} = [x_{l,1,(k)} ... x_{l,N_t,(k)}]^T$ so that $x_{l,i,(k)}$ for $k = 1...K$ are i.i.d. QPSK symbols with equal probability values.

In this case, the likelihood function is given as a sum of $Q^{N_cN_t}$ ($Q = 4$ for QPSK (4-QAM)) Gaussian pdfs as follows:

$$p(\mathbf{y}_{1(k)}, \boldsymbol{\theta}) = \frac{1}{Q^{N_cN_t}} \sum_{q=1}^{Q^{N_cN_t}} \frac{1}{(\pi\sigma_{\mathbf{v}_1}^2)^{N_r}} e^{-\left\| \frac{\mathbf{y}_{1(k)} - \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}^{\frac{1}{2}} \mathbf{x}_q}{\sigma_{\mathbf{v}_1}^2} \right\|}, \tag{2.22}$$

where $\mathbf{x}_q$ is the q-th realization of $\mathbf{x}_{(k)}$.

Consequently, the data-based FIM is a weighted sum of Gaussian FIMs given by:

$$\mathbf{J}_{\mathbf{h}_{tot}\mathbf{h}_{tot}}^{d}(k) = \frac{1}{\sigma_{\mathbf{v}_1}^2 Q^{N_cN_t}} \sum_{q=1}^{Q^{N_cN_t}} \left(\frac{\partial \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}^{\frac{1}{2}} \mathbf{x}_q}{\partial \mathbf{h}_{tot}^*} \right)^H \left(\frac{\partial \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}^{\frac{1}{2}} \mathbf{x}_q}{\partial \mathbf{h}_{tot}^*} \right). \tag{2.23}$$

To obtain a tractable FIM expression, a realistic approximation for a single cell MIMO-OFDM system was proposed in [59]. This approximation can be easily extended to a multi-cell massive MIMO-OFDM system. To do so, let's express the elements of the data-based FIM:

$$\mathbf{J}^{d}_{h_i h_j}(k) = \frac{1}{\sigma^2_{\mathbf{v}_1} Q^{N_c N_t}} \sum_{q=1}^{Q^{N_c N_t}} \mathbf{x}_q{}^H \left(\frac{\partial \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}{}^{\frac{1}{2}}}{\partial h_i^*} \right)^H \left(\frac{\partial \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}{}^{\frac{1}{2}}}{\partial h_j^*} \right) \mathbf{x}_q. \tag{2.24}$$

$$\mathbf{J}^{d}_{h_i h_j}(k) = \frac{1}{\sigma^2_{\mathbf{v}_1} Q^{N_c N_t}} \sum_{q,m,l} x_q{}^*(m) x_q(l) \boldsymbol{\Gamma}^{i,j}_{m,l} \quad 1 \leq m,l \leq N_c N_t \tag{2.25}$$

where $\boldsymbol{\Gamma}^{i,j} = \left(\frac{\partial \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}{}^{\frac{1}{2}}}{\partial h_i^*} \right)^H \left(\frac{\partial \boldsymbol{\lambda}_{tot_{(k)}} \mathbf{C}_{\mathbf{x}}{}^{\frac{1}{2}}}{\partial h_j^*} \right)$.

Due to normalization and QAM constellations symmetry around zero, we have:

$$\begin{aligned} &\tfrac{1}{Q^{N_t N_c}} \sum_{q=1}^{Q^{N_t N_c}} x_q^*(m)\, x_q(l) = 0 \ \text{ for } m \neq l \\ &\tfrac{1}{Q^{N_t N_c}} \sum_{q=1}^{Q^{N_t N_c}} x_q^*(m)\, x_q(m) = 1 \ \text{ for } m = l \end{aligned} \tag{2.26}$$

Finally, the data-based FIM for the finite alphabet signals (QAM) can be reduced to:

$$\mathbf{J}^{d}_{h_i h_j}(k) = \frac{1}{\sigma^2_{\mathbf{v}}} tr\left\{ \boldsymbol{\Gamma}^{i,j} \right\} \tag{2.27}$$

The total data-based FIM is then obtained as follows:

$$\mathbf{J}^{d}_{\mathbf{h}_{tot}\mathbf{h}_{tot}} = N_d \sum_{k=1}^{K} \mathbf{J}^{d}_{\mathbf{h}_{tot}\mathbf{h}_{tot}}(k), \tag{2.28}$$

where N_d is the total number of data symbols.

Remark: Even though the proposed FIM simplification applies for any symmetric finite alphabet signal, the accuracy of the approximation decreases with the constellation order level and would be valid only for high SNRs in such a case.

Thanks to the implicit higher order statistics information available in this non-Gaussian case, the semi-blind based channel estimation is able to alleviate completely the pilot contamination problem according to the following proposition:

Proposition 3. *The non Gaussian semi-blind FIM as given in (10) is non singular meaning that all indeterminacies have been removed.*

Proof:

For non-Gaussian (communications) signals, the information provided by the Second Order Statistics as well as Higher Order Statistics of the data allows us to identify the channels up to

an unknown $(N_cN_t) \times (N_cN_t)$ diagonal unitary matrix(see for example identifiability results in [68]). This corresponds to N_cN_t unknown real parameters that can be easily estimated through the use of the pilots. □.

In this case, the top-left $(N_rN_tN) \times (N_rN_tN)$ block of the FIM inverse is considered as the CRB for the semi-blind estimation of the first cell channel vector.

2.4 Effect of pilot contamination with unsynchronized cells

This section is devoted to the effect of pilot contamination on the performance of semi-blind channel estimation approaches, when the BSs in the different N_c cells are not synchronized. Such an assumption is more realistic and practical for the cellular network. It is shown in [56, 57], that this desynchronization might help mitigating the pilot contamination problem. Here, this issue is analyzed in detail and it is shown, in particular, that under certain conditions detailed below (small inter-cell delays) the pilot contamination still occurs for asynchronous MIMO-OFDM systems. For large inter-cell delays, the pilot contamination problem might be mitigated where this case is investigated through the CRB derivation, by considering the adjacent cells signal together with the AWGN noise as a colored noise. These results will be experimentally supported by an estimator-dependent study, based on a Least Squares Decision Feedback (LS-DF) estimator.

2.4.1 Small inter-cell delay case

This subsection provides an explanation on why the pilot contamination problem persists when the inter-cell delays are small. Indeed, without lose of generality, consider two time-domain OFDM signals, sent from two adjacent cells using the same pilot sequence, and received at one BS antenna:

$$z_1(t) = h_1(t) * x_{CP,1}(t) \tag{2.29}$$

$$z_2(t) = h_2(t) * x_{CP,1}(t-\tau) = h_2(t-\tau) * x_{CP,1}(t) \tag{2.30}$$

$x_{CP,1}(t)$ being the sequential OFDM signal in the time domain including the Cyclic Prefix (CP), $*$ is the convolution operator and τ is introduced here to model the inter-cell delay. Hence, when the signal from cell 1 is corrupted by the one from cell 2, one observes:

$$z(t) = z_1(t) + z_2(t) = (h_1(t) + h_2(t-\tau)) * x_{CP,1}(t) \tag{2.31}$$

Consequently, if the channel size N and the delay τ are such that:

$$\tau + N \leq L + 1 \tag{2.32}$$

L being the CP length, the model given in equation (2.31) coincides with the one in equation (2.7) which shows that the pilot contamination problem remains unsolved in this case, which might represent a 'rough synchronization' context. This situation is illustrated by the simulation results in Figure 2.12.

2.4.2 Large inter-cell delay case

Now, an effective channel estimation becomes possible in this case (see for example [58]) and to analyze it in the sequel, the pilots from the neighboring cells are considered as interference signal where the interference and the noise term will be modeled as a colored Gaussian signal $\mathbf{v}_{\text{col}}$ independent from the signals of the cell of interest. The noise vector is assumed to be of zero mean and unknown covariance matrix $\mathbf{C}_{\text{v}}$, so that the received signal model becomes:

$$\mathbf{y}_1 = \tilde{\mathbf{X}}_1 \mathbf{h}_1 + \mathbf{v}_{\text{col}}. \tag{2.33}$$

When sending only known pilots, the received signal is so that $\mathbf{y}_1 \sim \mathcal{N}\left(\boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta}) = \tilde{\mathbf{X}}_{1_p}\mathbf{h}_1, \mathbf{C}_{\mathbf{y}_1}(\boldsymbol{\theta}) = \mathbf{C}_{\mathbf{v}}\right)$. The parameters vector to be estimated is expressed as follows:

$$\boldsymbol{\theta} = [\mathbf{h}_1^T \;\; (\mathbf{h}_1^*)^T \;\; \mathrm{q}^T \;\; (\mathrm{q}^*)^T]^T, \tag{2.34}$$

where $\mathbf{h}_1$, of size $N_r N_t N$, is the vector of the channel components of the cell of interest, $\mathbf{q} = [\mathrm{q}_1, \mathrm{q}_2, ...\mathrm{q}_{N_q}]^T$, where $\mathrm{q}_i, i = 1, ...N_q$ are the parameters used to represent the covariance matrix $\mathbf{C}_{\mathbf{v}}$.

According to the complex representation of $\boldsymbol{\theta}$, the global pilot-based FIM is given by:

$$\mathbf{J}^p_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{J}^p_{\mathbf{h}_1\mathbf{h}_1} & \mathbf{J}^p_{\mathbf{h}_1\mathbf{h}_1^*} & \mathbf{J}^p_{\mathbf{h}_1\mathbf{q}} & \mathbf{J}^p_{\mathbf{h}_1\mathbf{q}^*} \\ \mathbf{J}^p_{\mathbf{h}_1^*\mathbf{h}_1} & \mathbf{J}^p_{\mathbf{h}_1^*\mathbf{h}_1^*} & \mathbf{J}^p_{\mathbf{h}_1^*\mathbf{q}} & \mathbf{J}^p_{\mathbf{h}_1^*\mathbf{q}^*} \\ \mathbf{J}^p_{\mathbf{q}\mathbf{h}_1} & \mathbf{J}^p_{\mathbf{q}\mathbf{h}_1^*} & \mathbf{J}^p_{\mathbf{q}\mathbf{q}} & \mathbf{J}^p_{\mathbf{q}\mathbf{q}^*} \\ \mathbf{J}^p_{\mathbf{q}^*\mathbf{h}_1} & \mathbf{J}^p_{\mathbf{q}^*\mathbf{h}_1^*} & \mathbf{J}^p_{\mathbf{q}^*\mathbf{q}} & \mathbf{J}^p_{\mathbf{q}^*\mathbf{q}^*} \end{pmatrix}, \tag{2.35}$$

Their elements are derived according to the general Gaussian CRB derivation model:

$$J^p_{\theta_i\theta_j} = \left\{\frac{\partial \boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \theta_i^*}\right\}^H \mathbf{C}^{-1}_{\mathbf{y}_1}(\boldsymbol{\theta}) \left\{\frac{\partial \boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \theta_j^*}\right\} + tr\left\{\mathbf{C}^{-1}_{\mathbf{y}_1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \theta_i^*} \mathbf{C}^{-1}_{\mathbf{y}_1}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{C}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \theta_j^*}\right)^H\right\}. \tag{2.36}$$

Given equation (2.33), it can be shown that:

$$\frac{\partial \boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \mathbf{h}_1} = \tilde{\mathbf{X}}_{1_p}, \frac{\partial \boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \mathbf{h}_1^*} = \mathbf{0}, \frac{\partial \boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \mathbf{q}} = \mathbf{0}, \frac{\partial \boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \mathbf{q}^*} = \mathbf{0}$$

$$\frac{\partial \boldsymbol{C}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \mathbf{h}_1} = \mathbf{0}, \frac{\partial \boldsymbol{C}_{\mathbf{y}_1}(\boldsymbol{\theta})}{\partial \mathbf{h}_1^*} = \mathbf{0}.$$

Consequently, the FIM, for the pilot-based case, will be expressed as follows:

$$\mathbf{J}^{p}_{\boldsymbol{\theta\theta}} = \begin{pmatrix} \tilde{\mathbf{X}}^{H}_{1_p}\mathbf{C}^{-1}_{\mathbf{v}}\tilde{\mathbf{X}}_{1_p} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{X}}^{T}_{1_p}(\mathbf{C}^{-1}_{\mathbf{v}})^{*}\tilde{\mathbf{X}}^{*}_{1_p} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{J}^{p}_{\mathbf{qq}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{J}^{p}_{\mathbf{q}^{*}\mathbf{q}^{*}} \end{pmatrix}, \tag{2.37}$$

Finally, the pilot-based CRB for the channel parameter vector is given by:

$$\mathrm{CRB}_{\mathrm{OP}} = tr\{(\tilde{\mathbf{X}}^{H}_{1_p}\mathbf{C}^{-1}_{\mathbf{v}}\tilde{\mathbf{X}}_{1_p})^{-1}\}. \tag{2.38}$$

By taking into account the known pilots and the unknown data, the semi-blind FIM is given by equation (2.10). Moreover, by assuming known transmit powers, the vector of parameters to be estimated is given by equation (2.34), whereas the global data-based FIM is given by:

$$\mathbf{J}^{d}_{\boldsymbol{\theta\theta}} = \begin{pmatrix} \mathbf{J}^{d}_{\mathbf{h}_1\mathbf{h}_1} & \mathbf{J}^{d}_{\mathbf{h}_1\mathbf{h}_1^{*}} & \mathbf{J}^{d}_{\mathbf{h}_1\mathbf{q}} & \mathbf{J}^{d}_{\mathbf{h}_1\mathbf{q}^{*}} \\ \mathbf{J}^{d}_{\mathbf{h}_1^{*}\mathbf{h}_1} & \mathbf{J}^{d}_{\mathbf{h}_1^{*}\mathbf{h}_1^{*}} & \mathbf{J}^{d}_{\mathbf{h}_1^{*}\mathbf{q}} & \mathbf{J}^{d}_{\mathbf{h}_1^{*}\mathbf{q}^{*}} \\ \mathbf{J}^{d}_{\mathbf{q}\mathbf{h}_1} & \mathbf{J}^{d}_{\mathbf{q}\mathbf{h}_1^{*}} & \mathbf{J}^{d}_{\mathbf{q}\mathbf{q}} & \mathbf{J}^{d}_{\mathbf{q}\mathbf{q}^{*}} \\ \mathbf{J}^{d}_{\mathbf{q}^{*}\mathbf{h}_1} & \mathbf{J}^{d}_{\mathbf{q}^{*}\mathbf{h}_1^{*}} & \mathbf{J}^{d}_{\mathbf{q}^{*}\mathbf{q}} & \mathbf{J}^{d}_{\mathbf{q}^{*}\mathbf{q}^{*}} \end{pmatrix}, \tag{2.39}$$

Note that the received data signal satisfies $\mathbf{y}_1 \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{y}_1}(\boldsymbol{\theta}) = \mathbf{0},\ \mathbf{C}_{\mathbf{y}_1}(\boldsymbol{\theta}) = \boldsymbol{\lambda}_1\mathbf{C}_{\mathbf{x}_1}\boldsymbol{\lambda}_1^{H} + \mathbf{C}_{\mathbf{v}})$.

Unfortunately, in that case, the off diagonal blocks $\mathbf{J}^{d}_{\mathbf{h}_1\mathbf{q}}$ and $\mathbf{J}^{d}_{\mathbf{h}_1\mathbf{q}^{*}}$ are not equal to zero as in the pilot-based case, and hence the channel estimation CRB depends on the estimation of the vector $\mathbf{q}$. The proper parameterization of the interference plus noise covariance matrix being quite challenging, we propose next to investigate the performance of the semi-blind case in this cell asynchronous context by using an estimator-dependent analysis.

Remark: Note that, the pilots of the cell of interest are known and hence they are considered as deterministic. Therefore, in the case where neighboring cells share the same pilots, the randomness comes only from the unknown data and noise terms. In such a case, the statistical independence assumption is valid. However, the i.i.d. colored Gaussian model remains a limiting approximation. In addition to that, a main difficulty comes from the 'non-synchronization' of the cells which makes the data model of the signals impinging from adjacent cells quite complex (we cannot rely on the simple OFDM model in (3) obtained after CP removal and FFT) and consequently the exact FIM derivation becomes cumbersome and numerically unattractive.

2.4.3 LS-DF estimator performance analysis

The derivation and performance of the Least Squares Decision Feedback (LS-DF) estimator, introduced in [69], are presented in this section. This 'relatively simple' estimator is used to illustrate the semi-blind performance in the different contexts of cell desynchronization, discussed previously.

Traditionally used for data equalization, the LS-DF algorithm is defined as a LS estimator which incorporates a feedback equalizer. During the LS-DF process, the estimated data are re-injected, as a

feedback, to the equalization step in order to enhance the estimation performance of the transmitted data. This process can be iterated several times for more accuracy.

We have exploited this method as a semi-blind channel estimator in [69], where the estimated data are considered as extra "pilots" for the channel taps estimation. The use of LS-DF estimator, as a semi-blind approach, for massive MIMO channel identification is illustrated in Figure 2.4 and resumed by the following steps:

- A pilot-based channel estimation is performed using the conventional LS estimator as follows:

$$\hat{\mathbf{h}}_{op_1} = \left(\tilde{\mathbf{X}}_{1_p}^H \tilde{\mathbf{X}}_{1_p}\right)^{-1} \tilde{\mathbf{X}}_{1_p}^H \mathbf{y}_1, \tag{2.40}$$

  where $\tilde{\mathbf{X}}_{1_p}$ is as defined in equation (2.7).

- A Zero-Forcing (ZF) equalizer is used to estimate the transmitted data, by applying the inverse of the channel frequency response to the received signal as follows:

$$\mathbf{x}_{zf} = \hat{\boldsymbol{\lambda}}^{\#} \mathbf{y}_1, \tag{2.41}$$

  where $\hat{\boldsymbol{\lambda}}$ is obtained from the channel frequency response of $\hat{\mathbf{h}}_{op_1}$ and $\mathbf{x}_{zf}$ is the equalized signal.

- A hard decision is performed on the equalized signal to obtain the estimate of the transmitted signal $\hat{\mathbf{x}}_{d_1}^{op}$.

- A LS estimator is then applied using the new training sequences (i.e. pilots) given by $\mathbf{x}_p = [\tilde{\mathbf{x}}_{1_p}^T \ (\mathbf{C}_{\mathbf{x}_1}^{\frac{1}{2}} \hat{\mathbf{x}}_{d_1}^{op})^T]^T$, where $\mathbf{C}_{\mathbf{x}_1}$ is the known transmit data power matrix introduced in equation(2.21) . This step will lead to the semi-blind channel estimate $\hat{\mathbf{h}}_{sb_1}$.

- A ZF equalizer, followed by a hard decision are performed to obtain the semi-blind estimate of the transmitted data $\hat{\mathbf{x}}_{d_1}^{SB}$.

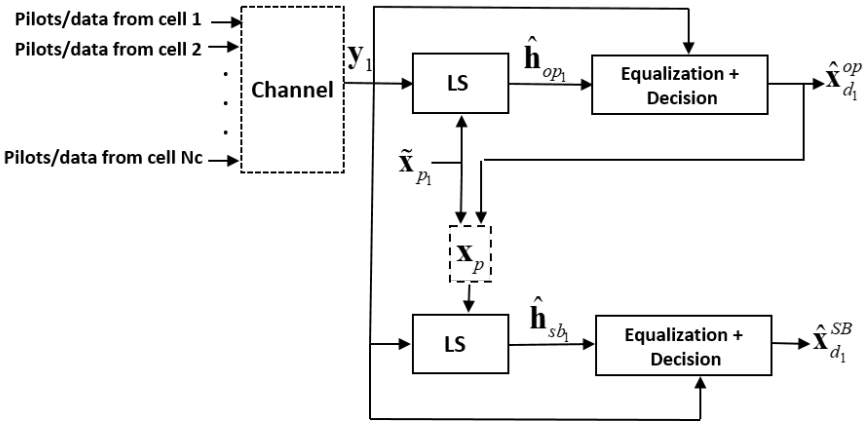


Figure 2.4: *LS-DF semi-blind channel estimation approach.*

The LS channel estimation performance has been widely discussed in literature, where it has been shown that the mean squares error (MSE) of this estimator reaches the CRB_{OP} in a single cell system with

AWGN noise. Therefore the MSE_{OP} is then given by:

$$\text{MSE}_{\text{OP}} = \text{CRB}_{\text{OP}} = \sigma^2_{\mathbf{v}_1} tr\left\{\left(\tilde{\mathbf{X}}^H_{1_p}\tilde{\mathbf{X}}_{1_p}\right)^{-1}\right\}. \tag{2.42}$$

However, in a multi-cell system, the effect of the adjacent cells signals will result in an extra bias term of inter-cell interference which will affect the estimation performance, depending on the inter-cell delay range. In that case, the LS-DF performance would be affected as well as can be seen in Figure 2.12 and Figure 2.13.

2.5 Performance analysis and discussions

In the following section, numerical experiments will be performed to highlight the different results given in the previous sections. The pilots are generated according to Zadoff-Chu sequences [70]. The $N_c N_t N_r N$ channel coefficients are all generated using i.i.d. unit-power, zero-mean, Gaussian distribution. It is important to note that the average Signal to Noise Ratio (SNR) is calculated based on the received signal $\mathbf{y}_1$ given in equations (2.5), i.e. $\text{SNR} = E(\|\boldsymbol{\lambda}_{tot}\mathbf{x}_{tot}\|^2)/E(\|\mathbf{v}_1\|^2) = tr(\boldsymbol{\lambda}_{tot}\mathbf{C}_{\mathbf{x}} \otimes \mathbf{I}_K \boldsymbol{\lambda}^H_{tot})/(N_r K \sigma_v^2)$. Moreover, the differences in users powers reflect their random locations. The different simulation parameters are summarized in Table 2.1, unless otherwise mentioned.

Parameters	Specifications
Number of cells	$N_c = 3$
Number of receive antennas	$N_r = 100$
Number of users per cell	$N_t = 2$
Channel taps	$N = 4$
Number of OFDM sub-carriers	$K = 64$
Number of OFDM pilot symbols	$N_p = 4$
Number of OFDM data symbols	$N_d = 40$
N_c pilot signal powers (dBm)	$P_{x_p} = [23\ 18\ 15]$
$(N_t) \times N_c$ data signal powers (dBm)	$P_{x_d} = [(20\ 18.8),$ $(15.7\ 13.3),\ (11.2\ 9.1)]$

Table 2.1: *Performance bounds analysis simulation parameters.*

2.5.1 Perfectly synchronized BSs

This section discusses the potential of the semi-blind channel estimation approaches, when the worst case of pilot contamination occurs in a massive MIMO-OFDM system.

Experiment 1: Figure 2.5 illustrates the normalized CRB for the channel parameters vector $\mathbf{h}_1$, given by $\frac{tr\{\text{CRB}\}}{\|\mathbf{h}_1\|^2}$, for semi-blind channel estimation (SB) with respect to the SNR for QPSK model as well

as the Gaussian (G) data model using orthogonal pilots. A comparison is made with respect to the pilot-based $\text{CRB}_{\text{OP}}^{\text{O}}$ case, which is the top-left block of the inverse of the FIM given in equation (2.14), using orthogonal (O) intra and inter-cell pilots. Note that $\text{CRB}_{\text{OP}}^{\text{NO}}$ and $\text{CRB}_{\text{SB}}^{\text{G-NO}}$ for the non orthogonal case (when the adjacent cells use the same pilots) are not considered since, as mentioned in sections 2.3.2.1 and 2.3.2.3, the channel parameters vector of the interest cell cannot be identified in that cases. However, such an ambiguity is removed by semi-blind techniques for finite alphabet source signals as illustrated by the plot of $\text{CRB}_{\text{SB}}^{\text{QPSK-NO}}$, which is obtained from the FIM given in equation (2.23) and stands for the semi-blind CRB of a QPSK signal for the worst case of non orthogonal (NO) pilots (i.e. adjacent cells using the same pilots). As can be seen, $\text{CRB}_{\text{SB}}^{\text{QPSK-NO}}$ is almost superposed with $\text{CRB}_{\text{SB}}^{\text{QPSK-O}}$, which denotes the case of orthogonal pilots. Ideally, the latter CRB can be obtained by ignoring the adjacent cells (i.e. $N_c = 1$).

Since the effect of pilot contamination is due to signals from adjacent cells, as described in subsection 2.3.1, Figure 2.6 illustrates the scenario of Figure 2.5 but with a higher number of cells (here the cell of interest is surrounded by six cells). One can observe that the previous results still hold under severe conditions of pilot contamination.

Experiment 2: now, the impact of pilots orthogonality level is investigated through the following metric:

$$\rho = \frac{\left\| \tilde{\mathbf{X}}_{i_P}^H \tilde{\mathbf{X}}_{j_P} \right\|}{\left\| \tilde{\mathbf{X}}_{i_P} \right\| \left\| \tilde{\mathbf{X}}_{j_P} \right\|}, \tag{2.43}$$

where $\|.\|$ is the 2-norm.

Note that $0 \leq \rho \leq 1$, so that $\rho = 0$ corresponds to the perfect orthogonality, whereas $\rho = 1$, corresponding to fully coherent training sequences, stands for the worst case of pilot contamination, i.e. same synchronized pilots.

As can be expected, in the case of non-perfectly orthogonal pilots, the channel vector estimation is slightly degraded but even with a high level of non orthogonality ($\rho = 70\%$ for the SB case and $\rho = 50\%$ for the OP case), the channel estimation for the OP and the Gaussian cases remains possible with relatively good estimation accuracy for moderate and high SNRs as illustrated in Figure 2.7.

Experiment 3: In order to further investigate the impact of the pilots structure on the pilot contamination, i.i.d. Gaussian distributed pilots are considered in this experiment. As given in Figure 2.8, $\text{CRB}_{\text{SB}}^{\text{G-multi-cell}}$ described in subsection 2.3.2.3 (respectively $\text{CRB}_{\text{SB}}^{\text{QPSK-multi-cell}}$ described in subsection 2.3.2.4) is almost superposed to $\text{CRB}_{\text{SB}}^{\text{G-mono-cell}}$ (respectively $\text{CRB}_{\text{SB}}^{\text{QPSK-mono-cell}}$) which indicates that the pilot contamination no longer persists and, thus, it is possible to use only SOS for semi-blind channel estimation. Besides, a pilot-based is now possible as given by $\text{CRB}_{\text{OP}}^{\text{multi-cell}}$ but with a degradation compared to the mono-cell case $\text{CRB}_{\text{OP}}^{\text{mono-cell}}$, given by equation (2.42), due to the interference terms from the neighboring cells. Actually, the independent pilots are different but not perfectly orthogonal which refers to the same results announced in *experiment 2* but given in terms of the orthogonality level.

Experiment 4: By considering the worst scenario of pilot contamination, the effect of the number of OFDM data symbols, i.e. N_d, on the $\text{CRB}_{\text{SB}}^{\text{QPSK-NO}}$, for a given SNR$= 10dB$, is illustrated in Figure 2.9.

It can be observed that, starting by one OFDM data symbol, the BS can successfully identify and estimate the channel components of the interest cell. Moreover, the CRB is significantly lowered with just few tens of OFDM data symbols and almost reaches the performance of the orthogonal case, i.e. $\mathrm{CRB}_{\mathrm{SB}}^{\mathrm{QPSK-O}}$. Such a result matches perfectly with the limited coherence time constraint of massive MIMO systems and helps to reduce the computational cost. As compared to $\mathrm{CRB}_{\mathrm{OP}}^{\mathrm{O}}$ a significant performance gain in favor of the semi-blind method is noticed.

Experiment 5: By considering again the worst case of pilot contamination, the behavior of the CRBs considered in Figure 2.5, with respect to the number of BS antennas, i.e. N_r, is investigated in Figure 2.10. It is easily observed that when N_r increases, which leads also to the increase of the number of channel components to be estimated, the $\mathrm{CRB}_{\mathrm{SB}}^{\mathrm{QPSK}}$ is significantly lowered thanks to the increased receive diversity. Such a result supports the effectiveness of semi-blind techniques for pilot contamination mitigation in the context of massive MIMO-OFDM systems.

Experiment 6: The channel order is often not known with accuracy and needs extra processing for its estimation. Thus, in Figure 2.11 we investigate the behavior of the aforementioned performance when the number of the channel taps is overestimated, i.e. considered equal to its maximum value corresponding to the cyclic prefix size ($N = L$). For illustration purpose, we have considered two cells, each with one user and a BS with $N_r = 10$ antennas. As can be seen from Figure 2.11, the channel order overestimation leads to a performance loss of approximately 6 dB which corresponds to the ratio (in dB) between the overestimated and the exact channel orders.

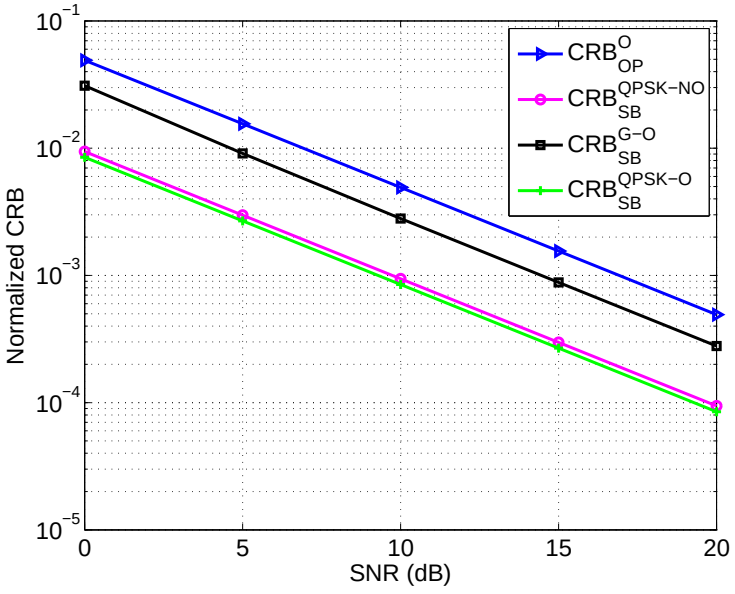


Figure 2.5: *Normalized CRB versus SNR.*

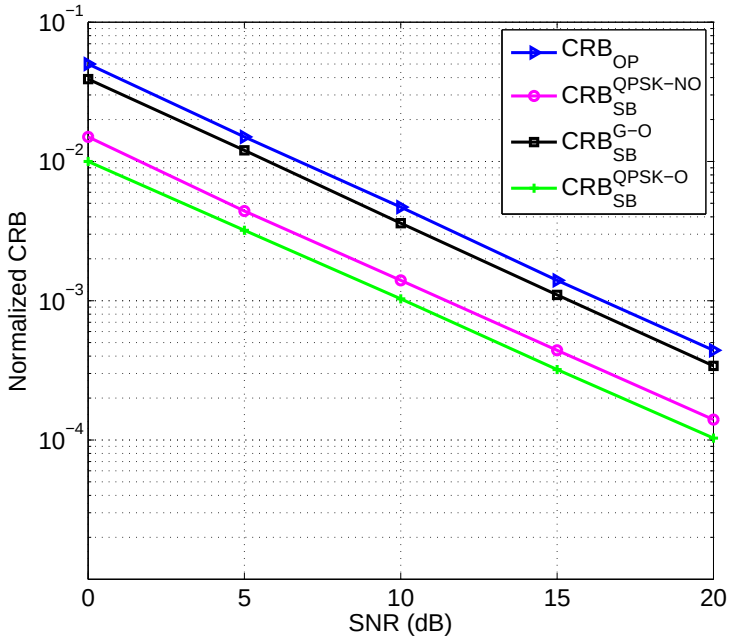


Figure 2.6: *Normalized CRB versus SNR with six adjacent cells.*

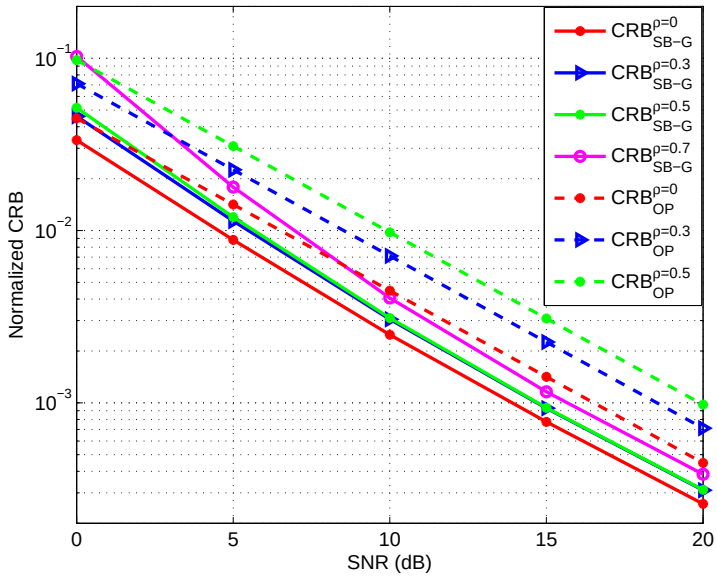


Figure 2.7: *Gaussian CRB versus SNR with different orthogonality levels.*

2.5.2 Unsynchronized BSs

This section investigates the semi-blind channel estimation potential in the case of unsynchronized BSs. The non synchronization was generated by using same pilots in all N_c cells but with different time delays compared to the target cell. Data symbols are assumed to be drawn from a finite alphabet signal (QPSK).

Experiment 7: Figure 2.12 illustrates the NMSE of the LS estimator, as a pilot-based one, and the

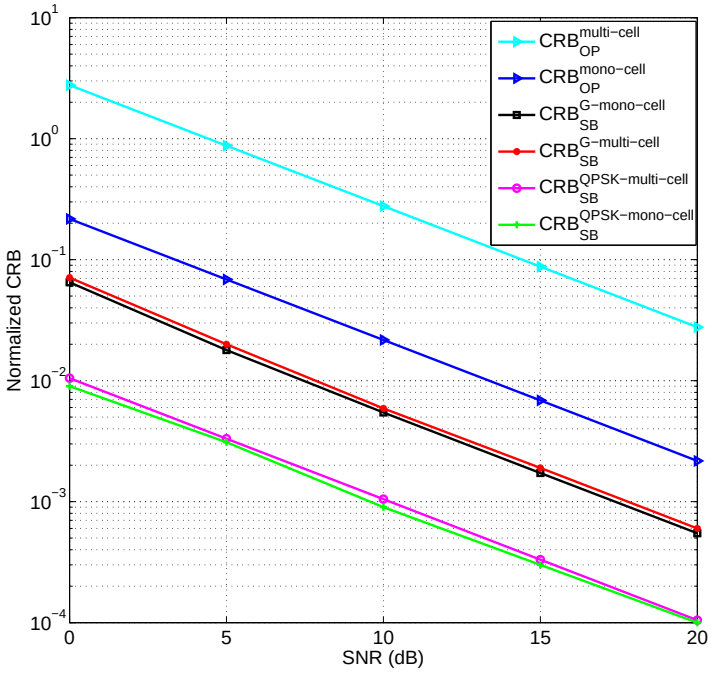


Figure 2.8: *Normalized CRB versus SNR with i.i.d. pilots.*

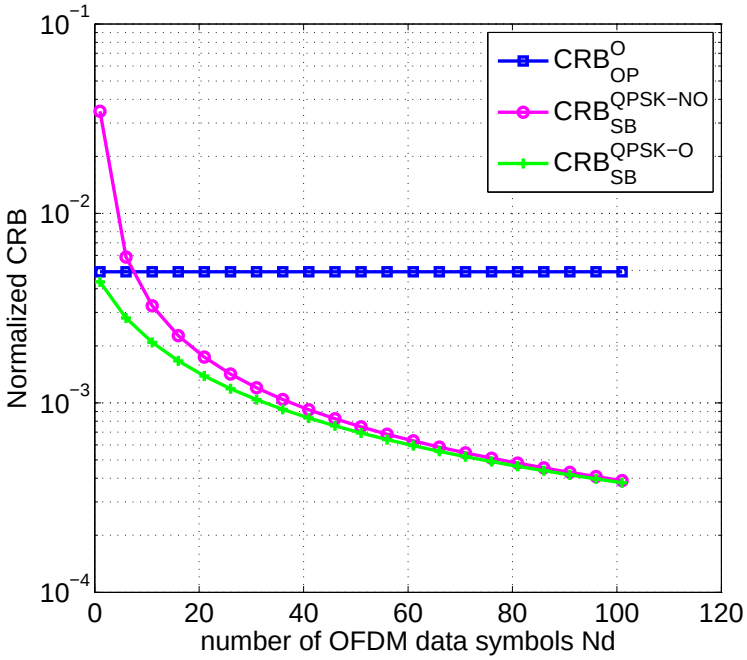


Figure 2.9: *Normalized CRB versus number of OFDM data symbols N_d.*

LS-DF estimator, as a semi-blind approach, with respect to the SNR, in the case of small inter-cell delays. As explained in subsection 2.4.1 and as given in the aforementioned figure, with small inter-cell delays,

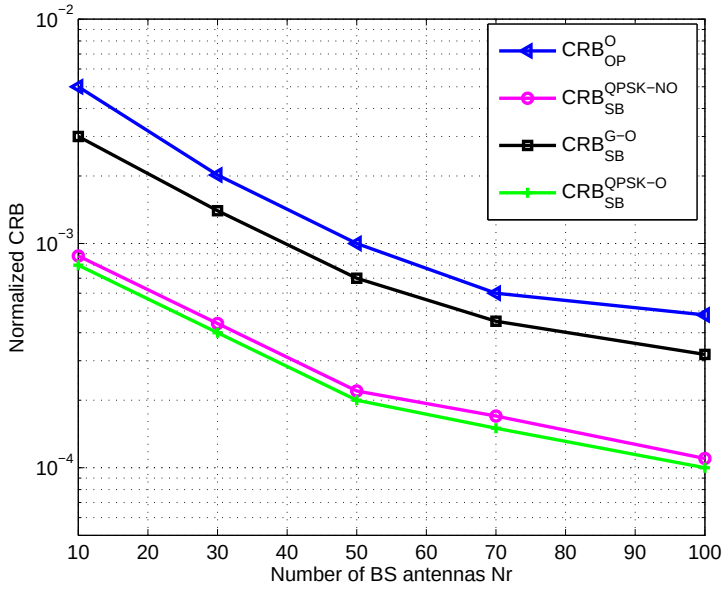


Figure 2.10: *Normalized CRB versus number of BS antennas N_r.*

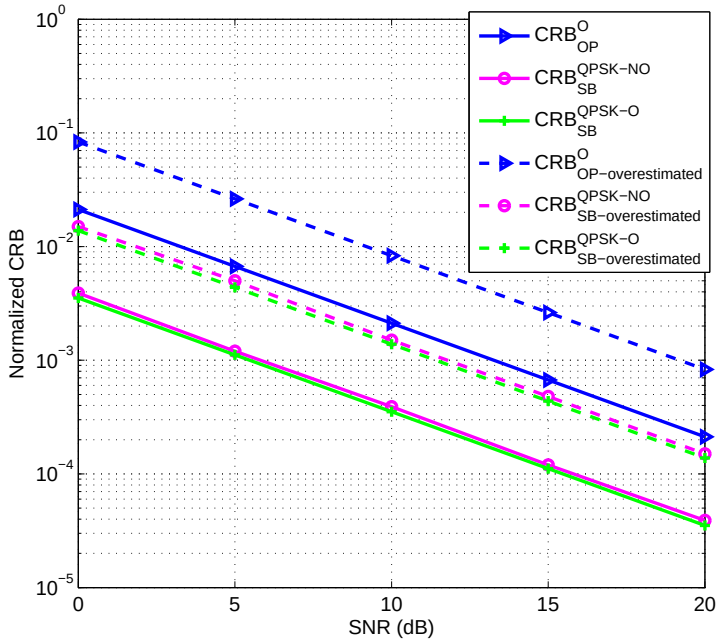


Figure 2.11: *Normalized CRB versus SNR with channel order overestimation.*

the problem of pilot contamination is still unsolved. For comparison, the results corresponding to the mono-cell case are also provided in this figure. In particular, the latter highlights the effectiveness of the LS-DF estimator (which reaches the CRB) in the absence of pilot contamination.

Experiment 8: Figure 2.13 investigates the performance, through the NMSE, of the LS-DF estimator,

as a semi-blind approach, in the case of large inter-cell delays. Compared to the small inter-cell case, the performance is slightly improved in this context. For only pilot case with unsynchronized cells, it is possible to identify and estimate the channel components of the interest cell with a moderate estimation error. As expected, the semi-blind case (SB), described by $\mathbf{h}_{\mathrm{SB-multi-cell}}$, outperforms the only pilot case, but with a degradation compared to the case of perfectly orthogonal pilots $\mathbf{h}_{\mathrm{SB-mono-cell}}$, which has a near optimal performance as it is almost superposed to the lower bound given by $\mathrm{CRB}_{\mathrm{SB}}$.

Experiment 9: Figure 2.14 investigates effect of the pilots' structure on the performance of the LS-DF estimator, by using randomly generated i.i.d. pilots. The performance obtained are similar to those obtained in the case of large inter-cell delays.

Experiment 10: Since the signal power of users in the adjacent cells is usually less than the signal power of users in the cell of interest, the effect of such a parameter, on the LS-DF performance, is investigated in Figure 2.15. We have considered the system parameters of Table 1 but, for each cell the N_t users are given the same power. The superscript stands for the ratio (in percentage) between each neighboring cell users power and the interest cell users power. One can observe for example that, for an interference level of 25% (corresponding approximately to 50% interference level if we add the interference terms of the two neighboring cells), the channel estimation for low and moderate SNRs with the SB approach ($\mathbf{h}^{25}_{\mathrm{SB-multi-cell}}$) is better than the one with the interference-free OP approach ($\mathbf{h}_{\mathrm{OP-mono-cell}}$).

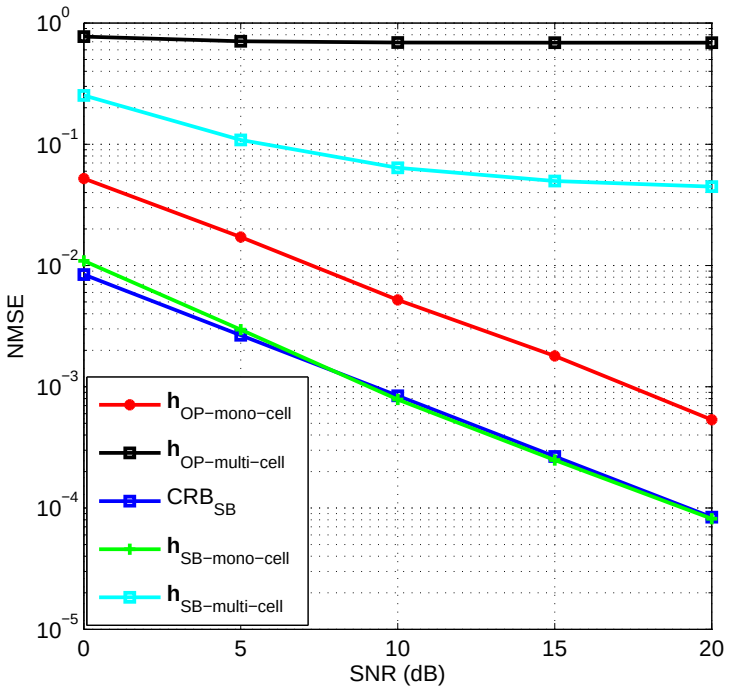


Figure 2.12: *NMSE of LS and LS-DF estimators versus SNR with small inter-cell delays*

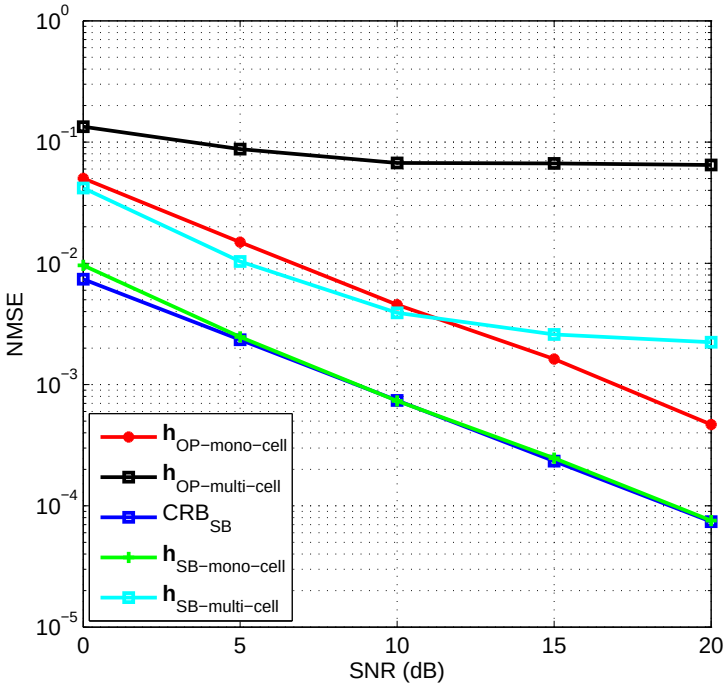


Figure 2.13: *NMSE of LS and LS-DF estimators versus SNR with large inter-cell delays.*

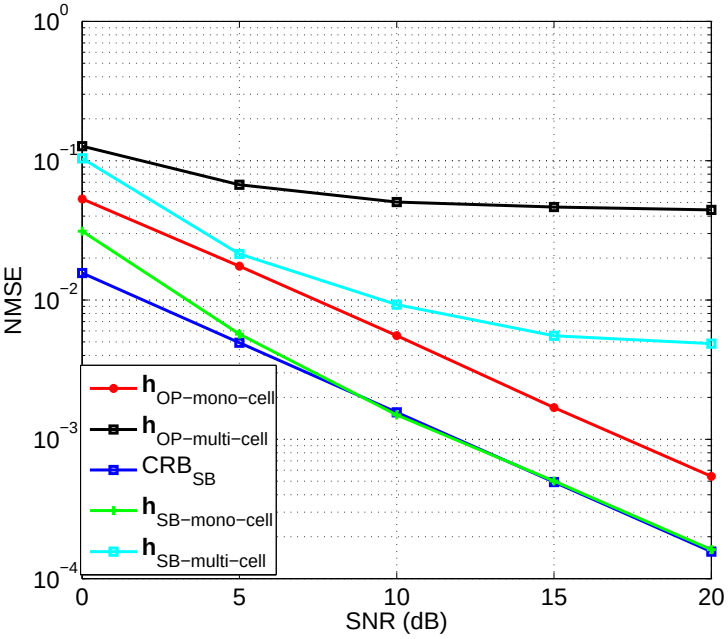


Figure 2.14: *NMSE of LS and LS-DF estimators versus SNR with i.i.d. pilots.*

2.6 Conclusion

This chapter focused on the performance bounds analysis of semi-blind channel estimation approaches, under the effect of pilot contamination, for multi-cell massive MIMO-OFDM systems. An estimator-

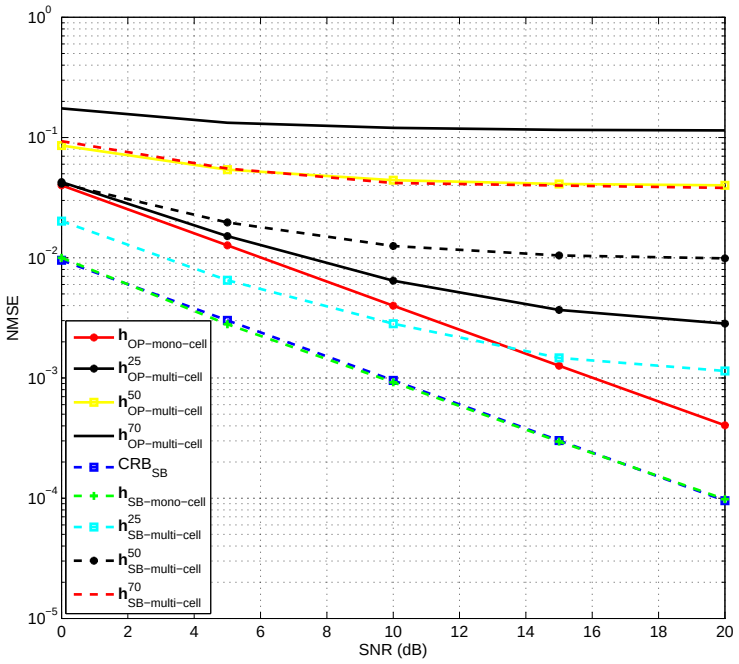


Figure 2.15: *NMSE of LS and LS-DF estimators versus SNR with different data powers.*

independent analysis has been conducted on the basis of the CRB by considering, at first, the worst case of pilot contamination for different data models, then, by taking into account unsynchronized BSs. It has been shown that the pilot contamination issue introduces a non-identifiability of the channel vector of the interest cell, which is not fully solved by considering only SOS, unless using non-perfectly orthogonal pilots, but can be efficiently solved with finite alphabet signals. For unsynchronized BSs with small inter-cell delays, the problem of pilot contamination remains unsolved. However, large inter-cell delays can allow to mitigate the pilot based non-identifiability issue. It is worth pointing out that, under this assumption, the colored Gaussian model is a limiting approximation. Besides, the 'non-synchronization' of the cells makes the data model of the signals impinging from adjacent cells quite complex and consequently the exact FIM derivation becomes cumbersome and numerically unattractive.

CHAPTER 3

Semi-blind multi-modulus based source separation

In theory there is no difference between theory and practice. In practice there is.

Lawrence "Yogui" Berra,1925

Abstract

This chapter presents an efficient semi-blind approach for demixing mixtures when considering, at first instantaneous massive Multiple-Input Multiple-Output (MIMO) communications system, then convolutive systems through massive Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) communications system. In both cases, the issue of pilot contamination will be taken into account. A weighted hybrid cost function is introduced based upon the Multi-Modulus (MM) criterion, applied to unknown data, and the Least Squares (LS) criterion for pilot symbols. A simple but efficient semi-blind block gradient descent procedure is put forward, in which the step size, which globally minimizes the cost function along the search direction, is algebraically computed at each iteration for each user. Besides a reduced computational complexity and an accelerated convergence, simulation results show that the proposed approach allows to mitigate the inherent ambiguity of fully-blind methods, and to withstand to the pilot contamination problem in massive MIMO systems as shown in [1][71] and [2][72].

Chapter content

[1] [71] O. Rekik, A. Mokraoui, A. Ladaycia and Abed-Meraim K. "Semi-Blind Source Separation based on Multi-Modulus Criterion: Application for Pilot Contamination Mitigation in Massive MIMO Communications Systems", in 19th IEEE International Symposium on Communications and Information Technologies (ISCIT), 2019,pp. 31-35.

[2] [72] O. Rekik, K. Abed-Meraim and A. Mokraoui"Multi-Modulus based Semi-Blind Source Separation for MIMO-OFDM Communications Systems", in 11th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM).2020.

3.1 Introduction

Source separation and channel equalization in digital communications aim at recovering the unknown data of the different users transmitted through a distorting propagation medium. In communications systems, with no prior knowledge on the transmitted symbols (i.e. no pilot sequences) nor on the channel state information, Blind Source Separation (BSS) is an efficient approach for data recovering. Indeed many BSS techniques have been proposed in literature (see e.g [26, 73] and references therein), mainly the Independent Component Analysis (ICA), the Sparse Component Analysis (SCA), the Non-Negative matrix Factorization (NMF) and the Bounded Component Analysis (BCA) (see e.g. [74] for more details).

BSS processes the received signal based on a priori knowledge about the statistics or the nature of the transmitted signals, through the optimization of an appropriate cost function. Various BSS cost functions have been proposed in literature (e.g [26] and references therein) depending upon the type of source signals. Among them; the Constant-Modulus (CM) criterion for phase/frequency modulated signals such as PSK/FSK and Multi-Modulus (MM) criterion, where it has been shown that it outperforms the CM one for the case of square QAM constellation [75], which is used in many modern communications systems such as LTE and WiMAX. The optimization of the cost function can be performed through a closed-form solution, such as Analytical Constant Modulus Algorithm (ACMA) [76] and Analytical Constant Power Algorithm (ACPA) [28], when the channel accepts a noiseless AR model and the FIR equalizer is sufficiently long. Otherwise, an iterative solution can be adopted, such as the gradient descent technique, the Newton's method, or Givens and Shear's rotations-based techniques [77].

On the other hand, some pilot sequences are often available in most communications technologies, thus, exploiting this available information should notably improve the source recovering performance of BSS by incorporating a pilot-based Least Squares (LS) criterion in a semi-blind scheme. In particular, this approach is shown to be an efficient solution to the pilot contamination problem in massive MIMO systems (e.g. [13]).

In this context, the focus of current work is to propose a semi-blind source separation technique for instantaneous mixtures and MIMO-OFDM communications systems. The motivation for adopting a SB approach is to make use of available sequences and to avoid any issues that emerge from the BSS such as the inherent ambiguity of blind processing. In particular, this approach is shown to be efficient to the pilot contamination problem in massive MIMO systems (e.g. [13]). In the proposed solution, a hybrid cost function is defined based on the blind MM criterion, for the unknown data, and on the pilot-based LS or Least Fourth (LF) criterion, for the pilots. An iterative-based minimization of the aforementioned cost function is performed through the gradient descent rule, where an optimized step-size procedure is introduced for improving the convergence speed. Moreover, a batch-based full estimation procedure is carried out so that, all sources are separated simultaneously.

3.2 Instantaneous system model

This section is dedicated to the case of instantaneous communications system model.

3.2.1 Communications system model and problem formulation

Consider a MIMO system consisting of N_t sources (transmitters), each having a single antenna, and a receiver equipped with N_r antennas. All sources transmit their signals over the same band of frequencies. Each transmitted source signal is drawn from an M-ary square QAM constellation, then passed through a flat fading channel represented by an unknown mixing matrix $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$, which is assumed to be of full column rank so that $N_r > N_t$. Thus, for the case of instantaneous mixtures, the noisy received signal is given by:

$$\mathbf{y}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{n}(k), \tag{3.1}$$

where $\mathbf{y}(k) = [y_1(k), ..., y_{N_r}(k)]^T$; $\mathbf{s}(k) = [s_1(k), ..., s_{N_t}(k)]^T$ refers to the transmitted signals and $\mathbf{n}(k) = [n_1(k), ..., n_{N_r}(k)]^T$ is an additive white Gaussian noise with a covariance matrix $\sigma_n^2 \mathbf{I}_{N_r}$.

The objective of source separation and channel equalization is to recover the transmitted data symbols, by applying a separation matrix $\mathbf{V} \in \mathbb{C}^{N_r \times N_t}$ to the observed (received) signals as follows: $\mathbf{z}(k) = \mathbf{V}^H \mathbf{y}(k)$. It can be noticed that each row $\mathbf{v}_i$ can extract one source signal. In batch processing approach, N_s samples of the received data are collected before processing, so that the matrix formulation of the problem is given by:

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{N}, \qquad \mathbf{Z} = \mathbf{V}^H \mathbf{Y}, \tag{3.2}$$

where the received signals are given by $\mathbf{Y} \in \mathbb{C}^{N_r \times N_s}$; the transmitted signals are represented by $\mathbf{S} \in \mathbb{C}^{N_t \times N_s}$; the additive white noise is given by $\mathbf{N} \in \mathbb{C}^{N_r \times N_s}$; whereas $\mathbf{Z} \in \mathbb{C}^{N_t \times N_s}$ is the estimated signals. Moreover, in semi-blind approaches, both pilots and data are used, hence, without loss of generality, the pilots are assumed to appear at the beginning of the transmitted frames in a block-type arrangement, thus each frame is formed by N_p pilots followed by N_d data samples, so that $N_s = N_p + N_d$ and $\mathbf{Y} = [\mathbf{Y}_p, \mathbf{Y}_d]$.

3.2.2 Semi-blind source separation

The proposed semi-blind source separation approach is based on the MM criterion for the unknown data and on the LS criterion for the pilots. Indeed to take advantage of both pilots and data, a hybrid cost function denoted $J_{\mathrm{SB}}(\mathbf{V})$ is defined as follows:

$$J_{\mathrm{SB}}(\mathbf{V}) = (1-\alpha) J_{\mathrm{B}}(\mathbf{V}) + \alpha J_{\mathrm{LS}}(\mathbf{v}), \tag{3.3}$$

where $J_{\mathrm{B}}(\mathbf{V})$ stands for a fully blind cost function; $J_{\mathrm{LS}}(\mathbf{V})$ refers to the use of pilots; and α is a real constant, taking values in the interval $[0, 1]$, considered as the weight given to the blind and the training-based parts of the semi-blind cost function.

In what follows, an iterative method based on the gradient descent is adopted to minimize the cost function, given by (3.3), according to:

$$\mathbf{V}_{n+1} = \mathbf{V}_n - \mu \mathbf{G}_n, \qquad n = 0, 1..., \tag{3.4}$$

where $\mathbf{V}_{n+1}$ (respectively $\mathbf{V}_n$) represents the updated (respectively the old) value of the matrix $\mathbf{V}$; μ is a small positive value, called step size, that determines the speed of convergence; and $\mathbf{G}_n$ is the gradient of

the cost function, at the n-th iteration, given by:

$$\mathbf{G}_n = \nabla J_{\mathrm{SB}}(\mathbf{V}_n) = (1-\alpha)\nabla J_{\mathrm{B}}(\mathbf{V}_n) + \alpha \nabla J_{\mathrm{LS}}(\mathbf{V}_n). \tag{3.5}$$

As adopted in [28], the iterative procedure is stopped as soon as:

$$\frac{\|\mathbf{v}_{n+1} - \mathbf{V}_n\|}{\|\mathbf{V}_n\|} < \frac{0.1\mu}{\sqrt{N_s}}. \tag{3.6}$$

Also, a maximum number of iterations can be defined for stopping the iteration process.

In the current work, we adopt a batch processing based on the use of block iterative implementation, as opposed to stochastic algorithms. The latter approaches approximate the gradient by using a one-sample estimate, which leads to dropping the expectation operator. Consequently, these methods generally lead to a slow convergence. By contrast, batch-based methods approximate the gradient from a block of the received samples repeatedly at each iteration. This more precise gradient estimate improves convergence speed and accuracy [78]. Moreover, all the sources are simultaneously estimated, so that the accumulated errors of the deflation-based methods are avoided [79].

3.2.2.1 Multi-Modulus criterion

In the current work, the blind process is based on the MM criterion, which penalizes the deviation of the real and imaginary parts of the equalized signals from the squared constellation shape as follows:

$$J_{\mathrm{MM}}(\mathbf{V}) = \sum_{i=1}^{N_t} \frac{1}{N_d} \sum_{k=N_p+1}^{N_s} [(z_{i,R}^2(k) - R_R)^2 + (z_{i,I}^2(k) - R_I)^2], \tag{3.7}$$

where $z_{i,R} = real(\mathbf{v}_i^H \mathbf{Y}_d)$ (respectively $z_{i,I} = imag(\mathbf{v}_i^H \mathbf{Y}_d)$) is the real (respectively imaginary) part of the (i,k)-th element of the recovered signal; $R_R = E[s_R^4(k)]/E[s_R^2(k)]$ and $R_I = E[s_I^4(k)]/E[s_I^2(k)]$ are the real and imaginary dispersion constants.

The gradient of the MM criterion, for the i-th user, is defined as follows:

$$\nabla J_{\mathrm{MM}}(\mathbf{v}_i) = \frac{1}{N_d} \sum_{k=N_p+1}^{N_s} [\mathbf{y}_d(k) \left((z_{i,R}^2(k) - R_R) z_{i,R}(k) - j(z_{i,I}^2(k) - R_I) z_{i,I}(k) \right)]. \tag{3.8}$$

In what follows, the CM criterion is used for comparison. It is given by [76]:

$$J_{\mathrm{CM}}(\mathbf{V}) = \sum_{i=1}^{N_t} \frac{1}{N_d} \sum_{k=N_p+1}^{N_s} (\mid z_i(k) \mid^2 - R)^2. \tag{3.9}$$

The gradient of the CM criterion, for each user, is defined as follows:

$$\nabla J_{\mathrm{CM}}(\mathbf{v}_i) = \frac{1}{N_d} \sum_{k=N_p+1}^{N_s} [\mathbf{y}_d(k) \left(z_i(k)^* (\mid z_i(k) \mid^2 - R) \right)]. \tag{3.10}$$

Note that the multiplicative constants of the previous gradient formulas are omitted because they are absorbed by the step size as given in (3.4).

3.2.2.2 Pilot-based criterion

In the current work, the use of the pilots is through the LS criterion, which is based on the error between the transmitted pilot symbols and their estimates, according to:

$$J_{\mathrm{LS}}(\mathbf{V}) = \sum_{i=1}^{N_t} \frac{1}{N_p} \sum_{k=1}^{N_p} \mid z_i(k) - s_{p_i}(k) \mid^2, \tag{3.11}$$

where $z_i(k)$ and $s_{p_i}(k)$ stand for the k-th estimated and transmitted pilot sample of the i-th source.

The gradient of the pilot-based LS criterion, for the i-th user, is defined as follows:

$$\nabla J_{\mathrm{LS}}(\mathbf{v}_i) = \frac{1}{N_p} \sum_{k=1}^{N_p} [\mathbf{y}_p(k) ((z_i(k) - s_{p_i}(k))^*)]. \tag{3.12}$$

It is clear that the MM and CM functions are 4-th order ones whereas the LS is a quadratic function. In some works (e.g. [80] and references therein), the 4-th order CM cost function is approximated by a quadratic function. To get a 'homogeneous' hybrid criterion, an alternative pilot-based cost function would be the Least mean Fourth (LF) [81] given by:

$$J_{\mathrm{LF}}(\mathbf{V}) = \sum_{i=1}^{N_t} \frac{1}{N_p} \sum_{k=1}^{N_p} \mid z_i(k) - s_{p_i}(k) \mid^4 . \tag{3.13}$$

As can be seen in the sequel, the latter cost function leads to a slight improvement of the source separation quality.

3.2.3 Optimal step size

Exact line search optimization technique has been successfully used recently for optimizing the step size of the steepest-descent gradient-based algorithms for channel identification/equalization and independent component analysis (e.g. [28] [82]), where the update rule is expressed as given in (3.4) but with a variable step size. Indeed, it can be observed that $J_{SB}(\mathbf{V}_{n+1})$ is a polynomial function of the step size parameter μ, thus, it is possible to perform a steepest descent of the objective function by finding the optimal step size:

$$\mu_{opt} = arg \min_{\mu} J_{SB}(\mathbf{V}_n - \mu \mathbf{G}_n). \tag{3.14}$$

Consequently, μ_{opt} is the appropriate root of the derivative of $J_{SB}(\mathbf{V}_{n+1})$ w.r.t. μ, which is a 3^{rd}-degree polynomial given by:

$$\nabla J_{SB}(\mathbf{V}_{n+1}) = (1-\alpha) p_B(\mu) + \alpha p_{LS}(\mu), \tag{3.15}$$

where

$$p_B(\mu) = p_{MM}(\mu) = (\boldsymbol{\beta}_{3_R} + \boldsymbol{\beta}_{3_I})\mu^3 + (\boldsymbol{\beta}_{2_R} + \boldsymbol{\beta}_{2_I})\mu^2 + (\boldsymbol{\beta}_{1_R} + \boldsymbol{\beta}_{1_I})\mu + \boldsymbol{\beta}_{0_R} + \boldsymbol{\beta}_{0_I}, \tag{3.16}$$

so that

$$\begin{aligned}
\boldsymbol{\beta}_{3_R} &= \frac{2}{N_d}\sum_{k=N_p+1}^{N_s}(\mathbf{a}_R^2), \quad \boldsymbol{\beta}_{2_R} = \frac{3}{N_d}\sum_{k=N_p+1}^{N_s}(\mathbf{a}_R\mathbf{b}_R),\\
\boldsymbol{\beta}_{1_R} &= \frac{2}{N_d}\sum_{k=N_p+1}^{N_s}(\mathbf{a}_R\mathbf{c}_R + \mathbf{b}_R^2),\\
\boldsymbol{\beta}_{0_R} &= \frac{1}{N_d}\sum_{k=N_p+1}^{N_s}(\mathbf{b}_R\mathbf{c}_R),
\end{aligned} \tag{3.17}$$

and

$$\begin{aligned}
&\mathbf{a}_R = |\ \mathbf{G_r}\ |^2, \quad \mathbf{b}_R = -2real\{\mathbf{V_r}\mathbf{G_r}^H\},\\
&\mathbf{c}_R = |\ \mathbf{V_r}\ |^2 - R,
\end{aligned} \tag{3.18}$$

$$\mathbf{G_r} = real\{\mathbf{G}^H\mathbf{Y}_d\} \quad \text{and} \quad \mathbf{V_r} = real\{\mathbf{V}^H\mathbf{Y}_d\}.$$

For $\boldsymbol{\beta}_{x_I}, x = 0,...3$ the same equations used in (3.17) still valid but with $\mathbf{G_I} = imag\{\mathbf{G}^H\mathbf{Y}_d\}$ and $\mathbf{V_I} = imag\{\mathbf{V}^H\mathbf{Y}_d\}$.

Also we have

$$p_{LS}(\mu) = \alpha_1\mu + \alpha_0, \tag{3.19}$$

where

$$\alpha_1 = \frac{1}{N_p}\sum_{k=1}^{N_p}(|\ \mathbf{G}^H\mathbf{Y}_p\ |^2), \tag{3.20}$$

$$\alpha_0 = \frac{-1}{N_p}\sum_{k=1}^{N_p}(real\{\mathbf{G}^H\mathbf{Y}_p^*(\mathbf{Y}_p - \mathbf{S}_p)\}), \tag{3.21}$$

where $\mathbf{S}_p$ contains the pilots of all users.

Remark: For the CM criterion, an optimal step size can be calculated by considering the polynomial:

$$p_{CM}(\mu) = \boldsymbol{\beta}_{3_C}\mu^3 + \boldsymbol{\beta}_{2_C}\mu^2 + \boldsymbol{\beta}_{1_C}\mu + \boldsymbol{\beta}_{0_C}, \tag{3.22}$$

where $\boldsymbol{\beta}_{3_C}, \boldsymbol{\beta}_{2_C}, \boldsymbol{\beta}_{1_C}$ and $\boldsymbol{\beta}_{0_C}$ are calculated in the same way as in (3.17) but by considering $\mathbf{G_C} = \mathbf{G}^H\mathbf{Y}_d$ and $\mathbf{V_C} = \mathbf{V}^H\mathbf{Y}_d$.

Finally, μ_{opt} is chosen as the real-valued root that minimizes the cost function $J_{SB}(\mathbf{V}_{n+1})$.

It is important to note that an optimal step size can be calculated for each user, so that, the upgrading rule can be expressed as follows:

$$\mathbf{V}_{n+1} = \mathbf{V}_n - \mathbf{G}_n diag(\boldsymbol{\mu}), \tag{3.23}$$

where $\boldsymbol{\mu} = [\mu_1, ..., \mu_{N_t}]$ contains all the optimal step size values of all users.

Remark: in the SB cost function, given by equation (3.3), a weighting parameter α has been introduced to determine the influence of the blind or the pilot-based part. Experiments will be given in section 3.5 to illustrate the effect of such parameter on the performance of the proposed approach.

3.2.4 SB-MM source separation under the effect of pilot contamination

This section considers a multi-cell massive MIMO system composed of N_c cells. In such a case, the system model is given by:

$$\mathbf{y}(k) = \sum_{l=1}^{N_c} \mathbf{H}_l \mathbf{s}_l(k) + \mathbf{n}(k), \tag{3.24}$$

where the first cell represents the cell of interest while the others are the interfering neighboring cells.

Traditionally, a channel estimation is performed before recovering the transmitted data during the uplink phase according to the Time Division Duplexing (TDD) protocol. However, pilots of the cell of interest are interfering with pilots of adjacent cells leading to the phenomenon of pilot contamination [13]. Recent work (e.g. [54]) show that semi-blind approaches represent a potential solution to this problem. Thus, we propose to recover the data of users located in the cell of interest by means of the proposed semi-blind source separation method. Interestingly, exploiting the pilots, through the LS or LF criterion included in the semi-blind cost function (i.e. equation (3.3)), allows to recover only the data of the cell of interest, from the observed mixture signal, and removing the inherent ambiguity of the blind-based source separation techniques.

Indeed, massive MIMO systems are characterized by their large size antennas, so that we assume that N_r is larger than the total number of users, i.e. $N_{t_1} + \cdots + N_{t_{N_c}}$, coming from the N_c cells and hence the global system matrix $\mathbf{H}$ is left invertible. In that case, the target separation matrix $\mathbf{V}$ is of size $N_r \times N_{t_1}$.

3.3 Convolutive system model: MIMO-OFDM system model

This section is dedicated to the convolutive model through MIMO-OFDM communications system model.

3.3.1 Communications system model and problem formulation

The system model adopted in this section is similar to that described in chapter 2, where an uplink transmission is considered. Let's consider a BS equipped with N_r antennas and receiving signals from N_t randomly located single-antenna users. Thus, the received signal, after cyclic prefix removal and FFT, at the r-th BS antenna, assumed to be a K sub-carriers OFDM signal, is given by [60]:

$$\mathbf{y}_r = \sum_{i=1}^{N_t} \mathbf{F}\, \boldsymbol{\mathcal{T}}(\mathbf{h}_{i,r}) \frac{\mathbf{F}^H}{K} \mathbf{x}_i + \mathbf{n}_r, \tag{3.25}$$

where $\mathbf{F}$ is a K-point Fourier matrix; $(.)^H$ is the transpose conjugate operator; $\mathbf{h}_{i,r}$ is a $N \times 1$ vector representing the channel coefficients between the i-th user and the r-th receive antenna; $\boldsymbol{\mathcal{T}}(\mathbf{h}_{i,r})$ is a $K \times K$ circulant matrix so that its first row is given by $[h_{i,r}(0), \mathbf{0}_{1\times K-N}, h_{i,r}(N-1), \ldots, h_{i,r}(1)]$ while the remaining rows are obtained by a simple cyclic shift to the right of the first one. $\mathbf{x}_i$ is the i-th user OFDM symbol of size $K \times 1$. $\mathbf{n}_r$, of size $K \times 1$, is assumed to be an additive white Circulant Gaussian (CG) noise so that $E[\mathbf{n}_r(k)\mathbf{n}_r(i)^H] = \sigma_n^2 \mathbf{I}_K \delta_{ki}$ where σ_n^2 is the noise variance, δ_{ki} represents the Kronecker delta operator and $\mathbf{I}_K$ is the identity matrix of size $K \times K$.

By exploiting the eigenvalue decomposition of the circulant matrix $\boldsymbol{\mathcal{T}}(\mathbf{h}_{i,r})$, given by:

$$\boldsymbol{\mathcal{T}}(\mathbf{h}_{i,r}) = \frac{\mathbf{F}^H}{K}\boldsymbol{\lambda}_{i,r}\mathbf{F}, \tag{3.26}$$

where $\boldsymbol{\lambda}_{i,r}$ is a $K \times K$ diagonal matrix formed by the frequency gain of the channel at the considered sub-carriers, i.e. $\boldsymbol{\lambda}_{i,r} = diag\{\mathbf{W}\mathbf{h}_{i,r}\}$ and $\mathbf{W}$ is formed by the first N columns of $\mathbf{F}$, and by stacking all the N_r received symbols in a single vector form, the received $N_rK \times 1$ signal, can be re-expressed as follows:

$$\mathbf{y} = \boldsymbol{\lambda}\mathbf{x} + \mathbf{n}, \tag{3.27}$$

where $\mathbf{y} = [\mathbf{y}_1^T ... \mathbf{y}_{N_r}^T]^T$; $\mathbf{x}_l = [\mathbf{x}_1^T ... \mathbf{x}_{N_t}^T]^T$; $\mathbf{n}_l = [\mathbf{n}_1^T ... \mathbf{n}_{N_r}^T]^T$; $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1 ... \boldsymbol{\lambda}_{N_t}]$ with $\boldsymbol{\lambda}_i = [\boldsymbol{\lambda}_{i,1} ... \boldsymbol{\lambda}_{i,N_r}]^T$.

Moreover, by collecting N_s symbols, one can write the received signal, at the k-th sub-carrier, as follows:

$$\mathbf{Y}_k = \boldsymbol{\lambda}_k\mathbf{X}_k + \mathbf{N}_k, \tag{3.28}$$

where $\mathbf{Y}_k \in \mathbb{C}^{N_r \times N_s}$, $\mathbf{X}_k \in \mathbb{C}^{N_t \times N_s}$, $\mathbf{N}_k \in \mathbb{C}^{N_r \times N_s}$ and $\boldsymbol{\lambda}_k \in \mathbb{C}^{N_r \times N_t}$ which is explicitly given by:

$$\boldsymbol{\lambda}_k = \sum_{l=0}^{N-1} \mathbf{H}(l) exp(-j2\pi \frac{k}{K} l), \tag{3.29}$$

where $\mathbf{H}(l) \in \mathbb{C}^{N_r \times N_t}$ is the l-th matrix coefficient of the considered MIMO transfer function.

In the sequel, the objective is to recover the transmitted data symbols by applying a separation matrix $\mathbf{V}$ to the observed (received) signal. However, the first step is to determine the form of such a matrix.

Basically, the separating matrix is constructed so that $\mathbf{V}(z)\boldsymbol{\lambda}(z) = \mathbf{I}_{N_t}$, where $\mathbf{I}_{N_t}$ is an identity matrix and the elements of $\boldsymbol{\lambda}(z)$ are polynomials of the form $\lambda_{i,r}(z) = \sum_{l=0}^{N-1} h_{i,r}(l) z^{-l}$ for every z. At this step we assume that the conditions for the existence of the inverse of the polynomial matrix $\boldsymbol{\lambda}(z)$ are satisfied, mainly $\boldsymbol{\lambda}(z)$ is irreducible (see [66, 83] for further details). Thus, $\mathbf{V}(z)$ would be a polynomial matrix with elements $v_{r,i}(z) = \sum_{l=0}^{N-1} v_{r,i}(l) z^{-l}$ for every z, which means that, on the unit circle we have $\mathbf{V}(exp(-j2\pi f_k))\boldsymbol{\lambda}(exp(-j2\pi f_k)) = \mathbf{I}_{N_t}$ where $f_k = \frac{k}{K}$. This result is equivalent to saying that $\mathbf{V}_k\boldsymbol{\lambda}_k = \mathbf{I}_{N_t}$ for every sub-carrier k. Consequently, we seek for the separating matrix given, at the k-th sub-carrier, by:

$$\mathbf{V}_k = \sum_{l=0}^{N-1} \mathbf{V}(l) exp(-j2\pi \frac{k}{K} l), \tag{3.30}$$

where $\mathbf{V}(l) \in \mathbb{C}^{N_t \times N_r}$.

Thus, the recovered signal, at the k-th sub-carrier, is given by:

$$\begin{aligned}
\mathbf{Z}_k &= \mathbf{V}_k\mathbf{Y}_k = \sum_{l=0}^{N-1} \mathbf{V}(l) exp(-j2\pi \frac{k}{K} l)\mathbf{Y}_k, \\
&= [\mathbf{V}(0)\, \mathbf{V}(1) \dots \mathbf{V}(N-1)] \begin{bmatrix} \mathbf{I}_{N_r} \\ exp(-j2\pi \frac{k}{K})\mathbf{I}_{N_r} \\ \vdots \\ exp(-j2\pi \frac{k}{K}(N-1))\mathbf{I}_{N_r}, \end{bmatrix} \mathbf{Y}_k, \\
&= [\mathbf{V}(0)\, \mathbf{V}(1) \dots \mathbf{V}(N-1)](\mathbf{w}_k^T \otimes \mathbf{I}_{N_r})\mathbf{Y}_k, \\
&= \mathbf{V}\bar{\mathbf{W}}_k\mathbf{Y}_k,
\end{aligned} \tag{3.31}$$

where $\bar{\mathbf{W}}_k = \mathbf{w}_k^T \otimes \mathbf{I}_{N_r}$, with $\mathbf{w}_k$ being the k-th row of $\mathbf{W}$ and $\mathbf{V} \in \mathbb{C}^{N_t \times N_r N}$.

Indeed, as given in equation (3.31), one needs to estimate only one matrix ($\mathbf{V}$), independently of the sub-carriers, rather than estimating $\bar{\mathbf{W}}_k$. Moreover, it is worth noting that each row $\mathbf{v}_i$ of $\mathbf{V}$ allows to extract one source signal.

Moreover, in SB approaches, both pilots and data are used, hence, w.l.g the pilots are assumed to appear at the beginning of the transmitted frames in a block-type arrangement. Thus each frame is formed by N_p pilots followed by N_d data symbols, so that $N_s = N_p + N_d$ and $\mathbf{Y} = [\mathbf{Y}_p, \mathbf{Y}_d]$.

3.3.2 Proposed Semi-blind source separation

In the sequel, we introduce the proposed solution described in section 3.2.2 for the case of MIMO-OFDM communications system.

3.3.2.1 Multi-Modulus criterion (MM)

For the case of MIMO-OFDM communications system, the same MM criterion used in 3.2.2.1 is used but for every sub-carrier as follows:

$$J_{\mathrm{B}}(\mathbf{V}) = J_{\mathrm{MM}}(\mathbf{V}) = \sum_{k=1}^{K}\sum_{i=1}^{N_t} E[(z_{k,i,R}^2(n) - R_R)^2 + (z_{k,i,I}^2(n) - R_I)^2] \tag{3.32}$$

where E is the expectation operator[3], $z_{k,i,R}(n) = real(\mathbf{v}_i \bar{\mathbf{W}}_k \mathbf{Y}_k(n))$ (resp. $z_{k,i,I}(n) = imag(\mathbf{v}_i \bar{\mathbf{W}}_k \mathbf{Y}_k(n))$) is the real (resp. imaginary) part of the (i,n)-th element of the recovered signal at the k-th sub-carrier. $R_R = E[x_R^4(n)]/E[x_R^2(n)]$ (resp. $R_I = E[x_I^4(n)]/E[x_I^2(n)]$) is the real (resp. imaginary) dispersion constant, which depends on the constellation of the transmitted signal $\mathbf{x}$; and $n = N_p + 1, \ldots N_s$.

The gradient of the MM criterion, for the i-th user, is defined as follows:

$$\nabla J_{\mathrm{MM}}(\mathbf{v}_i) = \sum_{k=1}^{K} 4E[(\bar{\mathbf{W}}_k \mathbf{Y}_k(n))^* \Big((z_{k,i,R}^2(n) - R_R) z_{k,i,R}(n) + j(z_{k,i,I}^2(n) - R_I) z_{i,I}(n)\Big)] \tag{3.33}$$

where $(.)^*$ refers to the complex conjugate.

Also, for benchmarking, the CM criterion [76] is introduced, since it has been widely used in literature. In such a scenario, the CM criterion is given by:

$$J_{\mathrm{CM}}(\mathbf{V}) = \sum_{k=1}^{K}\sum_{i=1}^{N_t} E[(\mid z_{k,i}(n) \mid^2 - R)^2], \tag{3.34}$$

where $z_{k,i}(n) = \mathbf{v}_i \bar{\mathbf{W}}_k \mathbf{Y}_k(n)$ and $R = E[x^4(n)]/E[x^2(n)]$ for $n = N_p + 1, \ldots N_s$.

3.3.2.2 Pilot-based criterion

The pilot-based part, defined through the LS criterion, is given by

$$J_{\mathrm{LS}}(\mathbf{V}) = \sum_{k=1}^{K}\sum_{i=1}^{N_t} E \mid z_{k,i}(n) - xp_{k,i}(n) \mid^2, \tag{3.35}$$

[3] Replaced in practice, by the time averaging operator.

where $z_{k,i}(n)$ and $xp_{k,i}$ stand for the n-th estimated and transmitted pilot symbols of the i-th source at the k-th sub-carrier, $n = 1, ...N_p$.

The gradient of the pilot-based LS criterion, for the i-th user, is given by:

$$\nabla J_{\mathrm{LS}}(\mathbf{V}_i) = 2E[(\bar{\mathbf{W}}_k \mathbf{Y}_k(n))^* \left(z_{k,i}(n) - xp_{k,i}(n)\right)]. \tag{3.36}$$

Here again, one can notice that MM and CM criteria are 4-th order functions whereas the LS-based criterion is a quadratic function. In order to get 'homogeneous' criteria, some works approximated the 4-th order CM cost function by a quadratic one (e.g. [80] and references therein). Thus, for homogeneity purpose, an alternative pilot-based cost function would be the Least mean Fourth (LF) [81], given by:

$$J_{\mathrm{LF}}(\mathbf{V}) = \sum_{k=1}^{K} \sum_{i=1}^{N_t} E \mid z_{k,i}(n) - xp_{k,i}(n) \mid^4 . \tag{3.37}$$

3.4 Multi-cell case and pilot contamination

In [38], authors have investigated the performance bounds of semi-blind channel estimation approaches for multi-cell massive MIMO-OFDM systems (under the effect of pilot contamination). In the current work, the performance of the proposed SB source separation technique is assessed under the effect of pilot contamination (same pilots in all cells). In such a case, a multi-cell massive MIMO system composed of N_c cells is considered. Thus, the system model is given by:

$$\mathbf{Y}_{1,k} = \sum_{i=1}^{N_c} \boldsymbol{\lambda}_{i,k} \mathbf{X}_{i,k} + \mathbf{X}_{1,k}, \tag{3.38}$$

where the first cell is considered as the cell of interest while the others are the interfering neighboring cells.

Mainly, pilot contamination occurs when same pilots are used in different adjacent cells, in which case, the channel estimation using only pilot symbols would fail. Thus, we propose to recover directly the transmitted data without performing channel estimation but rather source extraction of the desired sources (i.e. the users signals of the cell of interest) via our SB algorithm.

3.5 Performance analysis and discussions

This section highlights the performance of the proposed semi-blind source separation and its effectiveness under the effect of pilot contamination, for the case of instantaneous and convolutive communications systems. MM-SB refers to the MM-based SB source separation technique where pilots are used for initialization and in the cost function. However, MM-B stands for a blind cost function but with pilot-based initialization. CM-SB refers to the use of CM criterion for a SB data recovering. LS refers to the use of available pilots, which is calculated using the available training pilots: $\mathbf{V}_{LS} = \mathbf{Y}_p^{\#} \mathbf{S}_p^H$, where $\mathbf{Y}_p^{\#} = (\mathbf{Y}_p \mathbf{Y}_p^H)^{-1} \mathbf{Y}_p$ is the pseudo inverse of the received pilots $\mathbf{Y}_p$. Accordingly, we refer to the LS solution with both data and pilots used for training as the "MMSE lower bound", which is introduced for benchmarking. Except for the experiment of Figure 3.9 and Figure 3.10, pilots and data are drawn

from a 4-QAM modulation, whereas channel coefficients are generated using i.i.d. unit-power, zero-mean, Gaussian distribution. The results are averaged over 200 Monte Carlo runs, and the performance is assessed through an average Symbol Error Rate (SER). Simulation parameters are defined as: $N_t = 4$, $N_r = 10$, $N_d = 100$, $N_p = 1$, $K = 64$, $N = 4$ and $\alpha = 0.5$ and the fixed step size $\mu = 0.01$.

Figure 3.1 illustrates the performance of the blind MM Algorithm (MMA-B), the blind CM Algorithm (CMA-B), the semi-blind MM Algorithm (MMA-SB), and the semi-blind CM Algorithm (CMA-SB) for source separation, in a mono-cell system, with separating matrix randomly initialized. One can notice the enhancement, in terms of SER, of the semi-blind approaches compared to the blind ones. Moreover, better performance is obtained with an optimal step size (MMA-SB-OSS, CMA-SB-OSS) compared to the fixed step size case. By taking into account the nature of the cost function, the MM-based approaches outperform slightly the CM-based ones, since the former takes into account the phase variation.

Since some training pilots are available, an LS-based initialization is used for the separating matrix. In this context the performance of the different approaches are illustrated in Figure 3.2. One can observe that the blind approaches have been clearly enhanced (compared to the results given in Figure 3.1), whereas a local minimum convergence problem is observed for the semi-blind ones with a fixed step size at high SNR. By contrast, one can notice that the performance with optimal step size is virtually independent of the initialization, while dramatically reducing the iterations number as illustrated in Table 3.1. Notice that this reduction compensates the additional consumed time of calculating the optimal step size.

As explained in section 3.35, a LF criterion is adopted to obtain two homogeneous 4-th order cost functions (i.e. the MM and the LF). As shown in Figure 3.3, an enhancement of the semi-blind performance with LF (MMA-SB4 and CMA-SB4) is observed for medium and high SNRs as compared to the case of LS (MMA-SB and CMA-SB).

Remark: note that in figures 1, 2 and 3 the very small values of SER ($<< 10^{-5}$) are not plotted.

Figure 3.4 illustrates the behavior of the semi-blind approaches when increasing the number of pilots, for a given SNR=$5dB$. One can observe that with very few pilots (< 5 in our case), the fully blind case performs better than the semi-blind one as if using too few pilot symbols could 'confuse' the semi-blind approaches and leads to ill convergence. Similar effect has been observed for the semi-blind equalization in [28]. However, for a reasonable number of pilots, the semi-blind methods are able to attain the MMSE lower bound while maintaining good spectral efficiency and an effective data rate.

Figure 3.5 illustrates the behavior of the blind and semi-blind approaches when increasing the number of data symbols N_d (for MMSE lower bound N_d is kept fixed as reference). It can be clearly shown that with higher number of data symbols (around $N_d = 100$), for a given SNR=5dB, the performance of the blind and semi-blind source separation are enhanced. This result allows to use a small pilot/data size ratio for better throughput.

Figure 3.6 illustrates the behavior of the proposed solution under the effect of pilot contamination. One can observe that the LS-based equalizers are severely affected by the problem of pilot contamination (LS-contam) compared to the single cell case (LS). However, such a problem can be alleviated by the semi-

blind approach especially with MM criterion as compared to CM one (MMA-SB-contam, CMA-SB-contam). This result is very promising for pilot contamination mitigation in massive MIMO communications systems.

Figure 3.7 illustrates the performance of the proposed MM-based SB source separation technique with respect to SNR. One can notice the enhancement, in terms of SER, of the SB approaches as compared to the blind and the pilot-based ones. Also, within an OFDM framework, it is possible to perform a pilot-based and SB source separation with only one pilot symbol, which corresponds to K samples for the case of single-carrier framework. As shown in Figure 3.8, an enhancement of the SB performance with LF (MM-SB-4 and CM-SB-4) is observed for medium and high SNRs as compared to the case of LS (MM-SB and CM-SB).

Moreover, using an optimal step size leads to reducing dramatically the iterations number, as illustrated in Table 3.2. It is worth noting that this reduction compensates the additional time consumed for optimal step size calculation.

According to Figure 3.9 and Figure 3.10 , the results obtained in Figure 3.7 are still valid for higher order QAM modulations (16-QAM, 64-QAM). Moreover, one can notice that MM-based processing outperforms the CM-based one for such modulations.

Figure 3.11 investigates the effect of the weighting constant α on the performance of MM-SB. It can be noticed that, taking α values from the interval $[0.45, 0.65]$ allows better performance of the proposed SB approach. Moreover, one can notice that with just a few pilots, interesting performance is obtained with MM-SB. This result illustrates the efficiency of SB techniques for reducing the pilot overhead and improving the throughput of MIMO-OFDM communications systems.

Figure 3.12 investigates the behavior of the proposed solution under the effect of pilot contamination. One can observe that LS-based equalizers are severely affected by the problem of pilot contamination (LS-contam) compared to the single-cell case (LS). However, such a problem can be alleviated by the SB approach especially with MM criterion as compared to CM one (MM-SB-contam, CM-SB-contam). This result is very promising for pilot contamination mitigation in massive MIMO-OFDM communications systems.

Table 3.1: *Average number of iterations for convergence (instantaneous model)*

	B-CMA	SB-CMA	B-MMA	SB-MMA
Fixed Step Size	482	677	616	783
Optimized Step Size	280	392	350	460

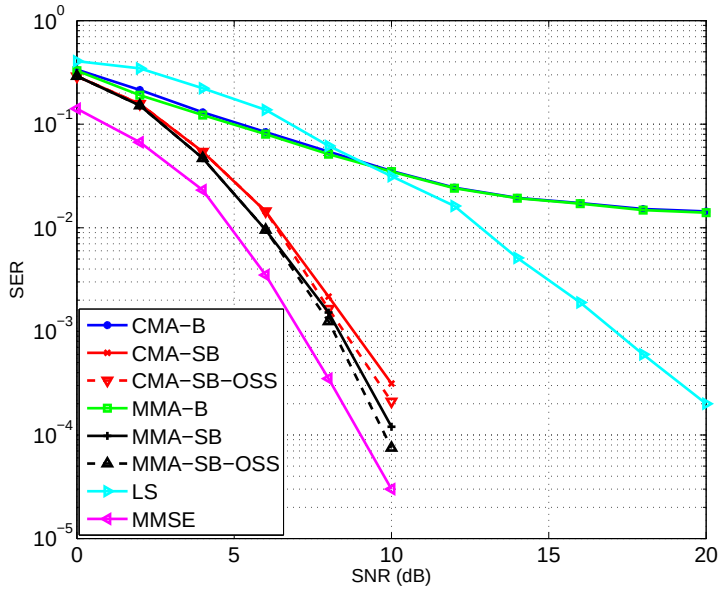


Figure 3.1: *Average SER vs SNR for blind and semi-blind source separation with random initialization* ($N_r = 20, N_t = 4, N_d = 100, N_p = 5$)

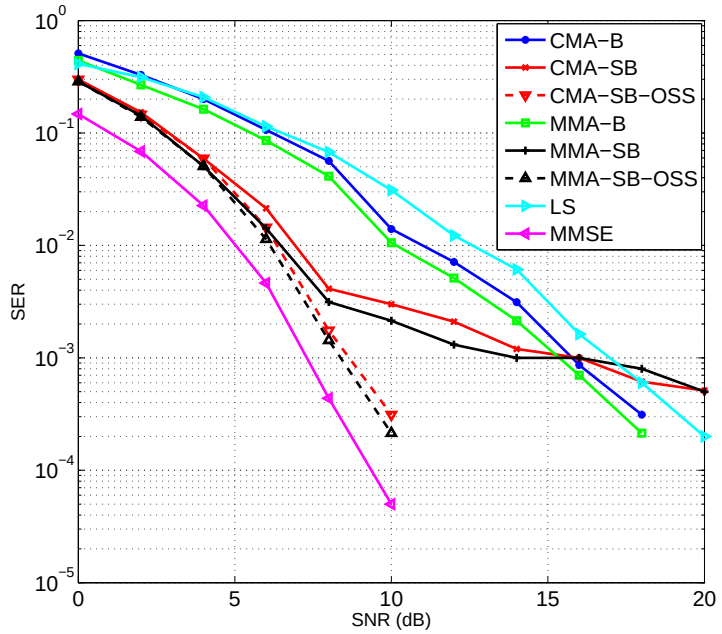


Figure 3.2: *Average SER vs SNR for blind and semi-blind source separation with pilot-based initialization* ($N_r = 20, N_t = 4, N_d = 100, N_p = 5$)

3.6 Conclusion

This chapter proposed a new approach for semi-blind source separation, which can help mitigate the problem of pilot contamination in massive instantaneous and convolutive MIMO communications systems.

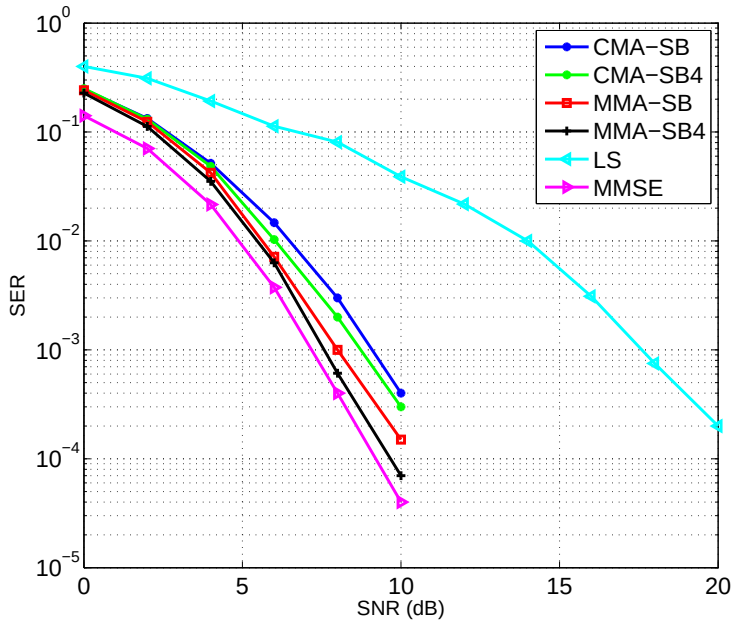


Figure 3.3: *Average SER vs SNR with LS and LF ($N_r = 20, N_t = 4, N_d = 100, N_p = 5$)*

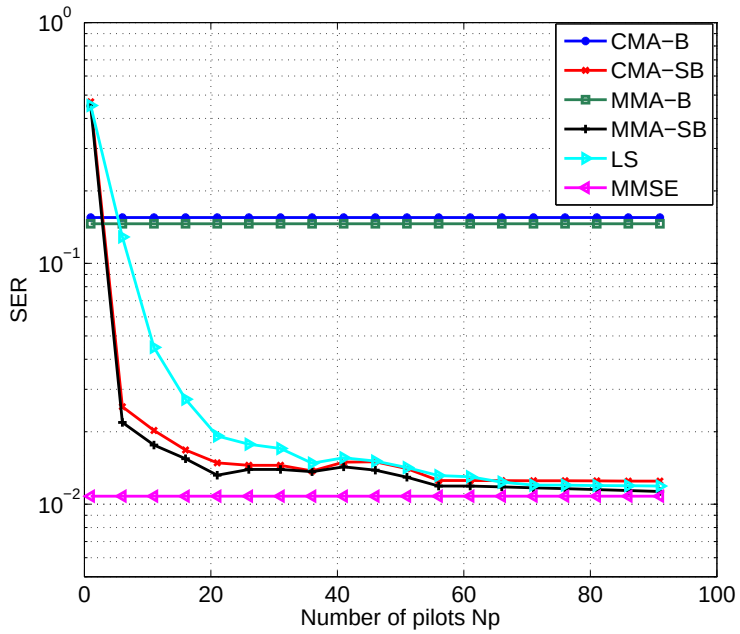


Figure 3.4: *Average SER vs number of pilots N_p for blind and semi-blind source separation ($N_r = 20, N_t = 4, N_d = 100, SNR = 5dB$)*

A hybrid cost function is defined based on the MM criterion for the blind part and the LS or LF criterion for the pilot-based one. A full estimation procedure has been adopted based on the gradient descent rule and an optimized step size. Simulation results have shown that the proposed method exhibits an

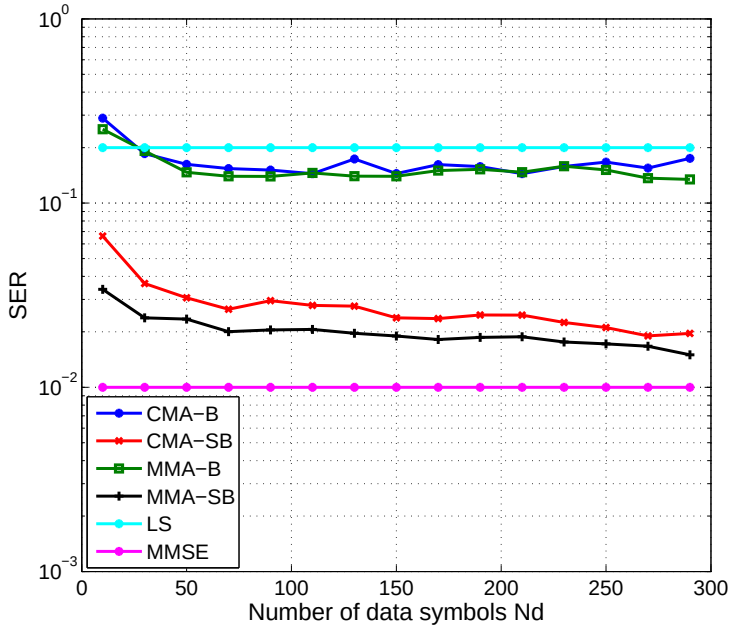


Figure 3.5: *Average SER vs number of data symbols N_d for blind and semi-blind source separation ($N_r = 20, N_t = 4, N_p = 5, SNR = 5dB$)*

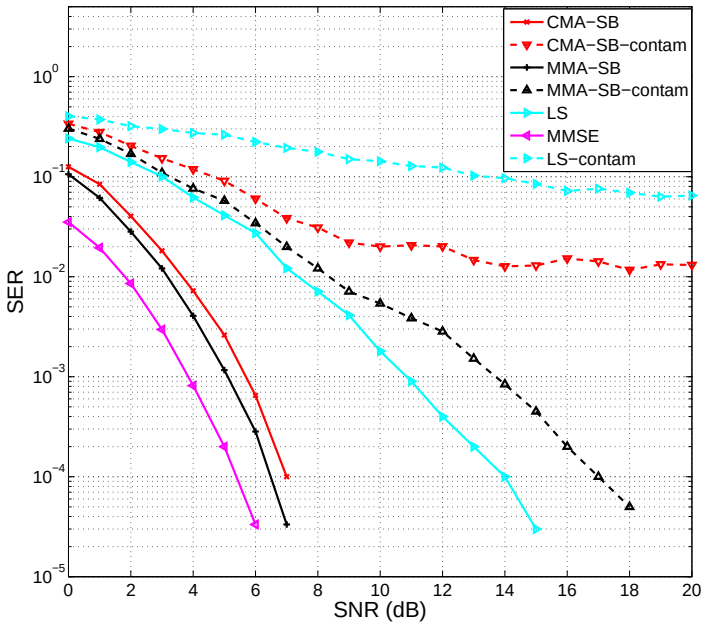


Figure 3.6: *Average SER vs SNR under pilot contamination with asynchronous cells ($N_r = 20, N_t = 2$ per cell$, N_c = 4, N_d = 100, N_p = 5$)*

attractive data recovery accuracy, convergence speed, and a promising source separation (deconvolution) performance under pilot contamination in massive MIMO-OFDM systems.

Table 3.2: *Average number of iterations for convergence (MIMO-OFDM model)*

	MM-SB	CM-SB	MM-B
Fixed Step Size	893	779	713
Optimized Step Size	574	496	472

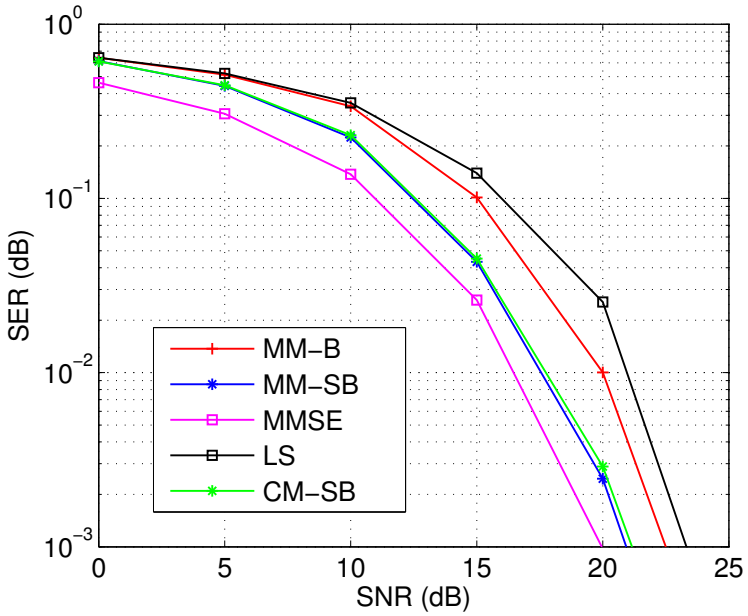


Figure 3.7: *SER vs SNR with LS (MIMO-OFDM model).*

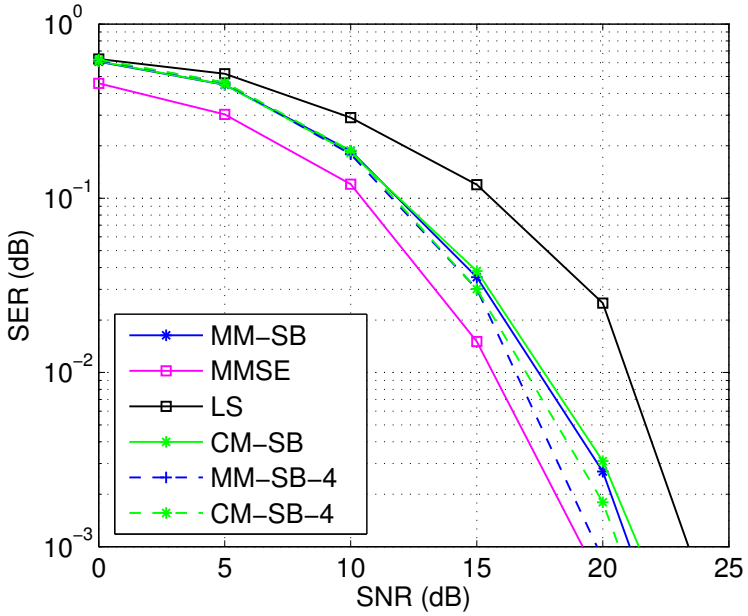


Figure 3.8: *SER vs SNR with LF (MIMO-OFDM model).*

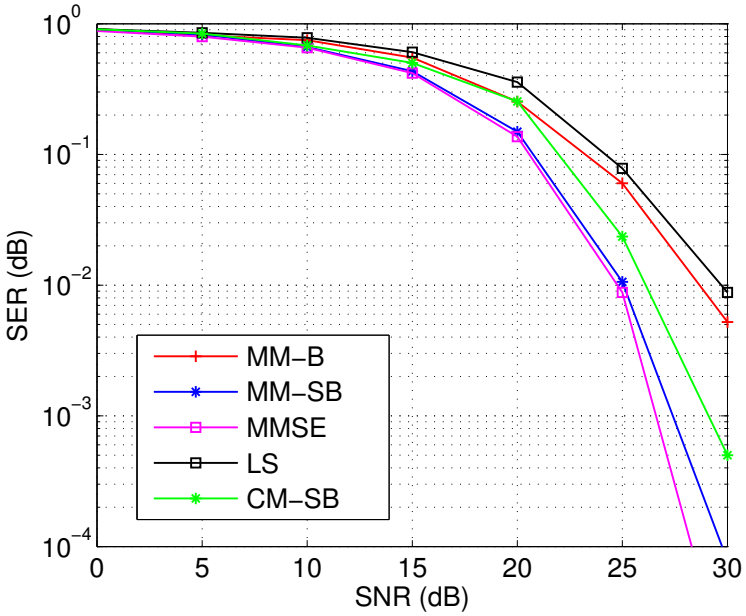


Figure 3.9: *SER vs SNR with 16-QAM (MIMO-OFDM model).*

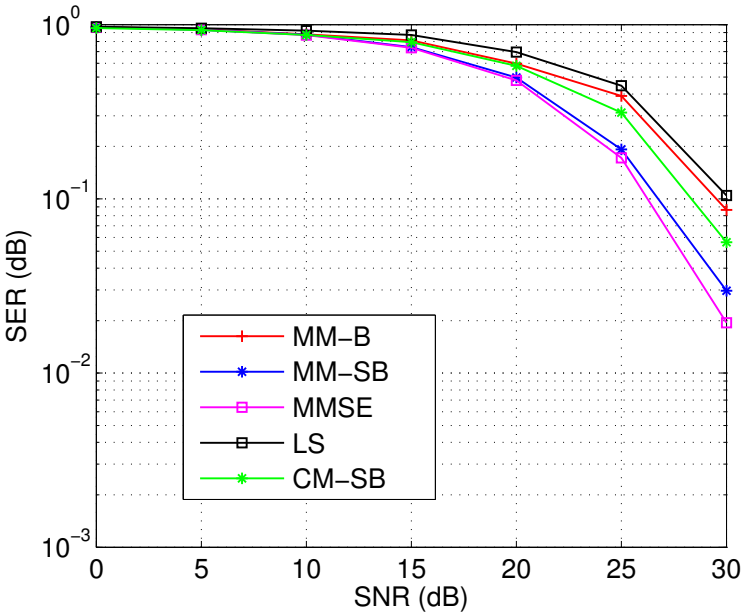


Figure 3.10: *SER vs SNR with 64-QAM (MIMO-OFDM model).*

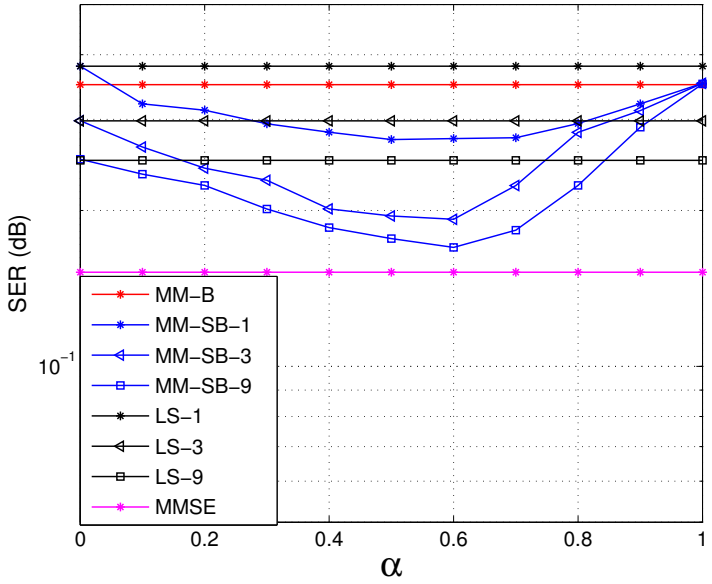


Figure 3.11: *SER vs α with $Np = 1,3,9$ pilot symbols (MIMO-OFDM model).*

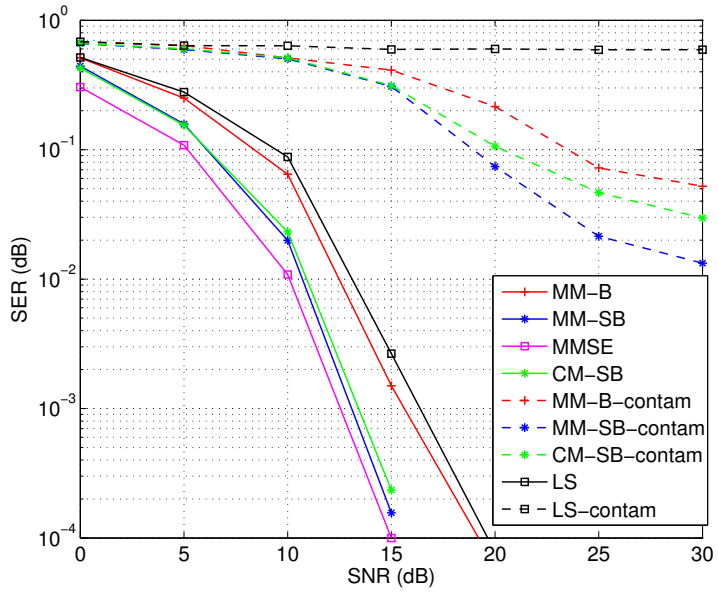


Figure 3.12: *SER vs SNR with pilot contamination ($N_c = 3$, $N_t = 2$) (MIMO-OFDM model).*

CHAPTER 4

Joint channel estimation and data detection with successive convex approximation

Problems are not stop signs, they are guidelines.

Robert H. Schuller.

This work has been done in collaboration with Pof. Marius Pesavento as part of a mobility to Germany (Darmstadt).

Abstract

The aim of this chapter is to propose a semi-blind solution, for joint sparse channel estimation and data detection, based on the successive convex approximation approach. The optimization is performed on an approximate convex problem, rather than the original nonconvex one. By exploiting available data and system structure, an iterative procedure is proposed where the channel coefficients and data symbols are updated simultaneously at each iteration. Also an optimized step size, introduced according to line search procedure, is used for convergence improvement with guaranteed convergence to a stationary point. Simulation results show that the proposed solution exhibits fast convergence with very attractive channel and data estimation performance. This work has been published in in [1][84]

[2] [84] O. Rekik, K. Abed-meraim, M. Pesavento and A. Mokraoui"Semi-blind Sparse Channel Estimation and Data Detection by Successive Convex Approximation", 21st IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2020, pp. 1-5.

Chapter content

4.1 Introduction

Channel estimation and data detection is a key task in wireless communications systems and different approaches have been proposed for joint channel and data estimation (e.g. [85, 86]). Moreover, In many wireless communications, growing experimental studies have showed that many practical channels exhibit sparsity as the delay spread could be very large but the number of distinguishable multi-path delays is usually small. In such a case, sparse channel estimation should be considered (see [87] and references therein).

A solution to perform channel identification and data recovery, by taking into account prior information about the channel and/or the data, would be solving a regression-based Optimization Problem (OP). Many techniques have been proposed for the problem of constrained linear regression, such as the Fast Iterative Soft Thresholding Algorithm (FISTA) [88], the Block Coordinate Descent (BCD) algorithm [89], the Alternating Direction Method of Multiplier (ADMM) [90] or the parallel best-response with exact line search algorithm [91]. In our context, the sequential update algorithm seems to be a good alternative to overcome difficulties arising from the nonconvexity of the problem, however, a major drawback of the sequential update is the induced large delay, because the update of a block variable cannot be performed until its predecessor block variable is updated (BCD algorithm [89]). In such a case, the delay may be very large with big number of data blocks.

Conseqently, an appropriate approach would be the successive convex approximation framework ([92] and references therein), where a sequence of successively refined approximate problems are solved, while preserving the algorithm's convergence to a stationary point of the original OP.

On the other hand, it bears mentioning that, in practice some training sequences, known by the transmitter and the receiver, are usually sent periodically within the wireless network frames besides the unknown data. Thus, the focus of the current work is to propose a solution for semi-blind sparse channel estimation and data recovery by considering a single-carrier Single-Input Multiple-Output (SIMO) system[2]. The motivation for adopting a semi-blind approach is to make use of available sequences and to avoid the different difficulties and issues that emerge from the blind process such as the inherent ambiguity of blind processing [67]. The proposed solution is based on the Successive Convex Approximation (SCA) framework along with the Majorization-Maximization approach [93]. An iterative procedure is performed where channel coefficients and data symbols are estimated simultaneously, at each iteration, with an optimized step size introduced for improving the convergence speed.

4.2 System model

For this work, a SIMO convolutive communications system is considered, where the r-th system output, $r = 1, \cdots, N_r$, is given by:

$$y_r(k) = \sum_{n=0}^{M} h_r(n)s(k-n) + v_r(k), \tag{4.1}$$

[2]This solution can easily be extended to the multi-user case (i.e. convolutive MIMO systems).

where h_r refers to the r-th channel finite impulse response of size $M+1$, $s(k)$ represents the transmitted symbols and $v_r(k)$ is a white Gaussian noise with variance σ_v^2. By considering N_s received samples, one can write (4.1) as follows:

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{V}, \tag{4.2}$$

where $\mathbf{Y}, \mathbf{V} \in \mathbb{C}^{N_r \times N_s}$, $\mathbf{H} \in \mathbb{C}^{N_r \times (M+1)}$, and $\mathbf{S} \in \mathbb{C}^{(M+1) \times N_s}$ is a Toeplitz matrix. Moreover, and without loss of generality, N_p training symbols (pilots) are sent at the beginning of the data frame, so that the transmitted and received symbols are given by $\bar{\mathbf{S}} = [\mathbf{S}_p, \mathbf{S}]$ and $\bar{\mathbf{Y}} = [\mathbf{Y}_p, \mathbf{Y}]$ respectively.

4.3 Proposed channel estimation and data recovery framework

This section is dedicated to the formulation of the appropriate OP describing the aforementioned scenario, then the derivation of the proposed solution.

4.3.1 Problem formulation

In what follows, we consider the problem of joint channel estimation and data recovery, by taking into account the channel sparsity as prior information. Also, a number of data symbols (pilots) are assumed to be known by the transmitter and the receiver. To do so, one can formulate an appropriate OP that incorporates all available model information and requirements. Basically, our OP is composed of the estimation error of a data matching function and regularization (penalty) terms for promoting, in the solution, a certain structure known *a priori*. Hence, the following OP is considered:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 + \mu\|\mathbf{H}\|_1, \tag{4.3}$$

where the l_1 norm enforces the sparsity of the channel response $\mathbf{H}$, with a regularization constant μ, whereas the term $\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2$ can be expressed as the sum of a data (blind) and a pilot-based terms as follows:

$$\begin{aligned} \|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 &= \|[\mathbf{Y}_p, \mathbf{Y}] - \mathbf{H}[\mathbf{S}_p, \mathbf{S}]\|_F^2 \\ &= \|\mathbf{Y}_p - \mathbf{H}\mathbf{S}_p\|_F^2 + \|\mathbf{Y} - \mathbf{H}\mathbf{S}\|_F^2. \end{aligned} \tag{4.4}$$

where $\mathbf{Y}_p$ and $\mathbf{S}_p$ refer respectively to the received and transmitted pilot symbols.

Moreover, given equation (4.1), the data matrix $\mathbf{S}$ has a Teoplitz structure, which can be added as a constraint to our OP (given by (4.3)). To do so, we define the matrix $\mathbf{S}_L$ which is formed by removing the last row and column of $\mathbf{S}$, and the matrix $\mathbf{S}_F$ which is formed by

removing the first row and column of $\mathbf{S}$. In such a case, we have: $\mathbf{S}_L = \mathbf{S}_F$ where:

$$\mathbf{S}_L = \mathbf{J}_{L_1}\mathbf{S}\mathbf{J}_{L_2}, \tag{4.5}$$

$$\mathbf{S}_F = \mathbf{J}_{F_1}\mathbf{S}\mathbf{J}_{F_2}, \tag{4.6}$$

where $\mathbf{J}_{L_1} = [\mathbf{I}_M, \mathbf{0}_{M\times 1}]$, $\mathbf{J}_{L_2} = [\mathbf{I}_{N_s-1}, \mathbf{0}^T_{(N_s-1)\times 1}]^T$, $\mathbf{J}_{F_1} = [\mathbf{0}_{M\times 1}, \mathbf{I}_M]$, $\mathbf{J}_{F_2} = [\mathbf{0}^T_{(N_s-1)\times 1}, \mathbf{I}_{N_s-1}]^T$, $\mathbf{I}$ and $\mathbf{0}$ refer to the identity matrix and all-zero matrix, respectively.

Thus, we can write:

$$\mathbf{S}_L - \mathbf{S}_F = \tilde{\mathbf{M}}\text{vec}(\mathbf{S}), \tag{4.7}$$

where $\tilde{\mathbf{M}} = \mathbf{J}^T_{L_2} \otimes \mathbf{J}_{L_1} - \mathbf{J}^T_{F_2} \otimes \mathbf{J}_{F_1}$ ($\otimes$ being the Kronecker product) and vec(.) denotes the matrix vectorization operator. Consequently, the OP, given by equation (4.3), can be re-expressed as follows:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|^2_F + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|^2_F + \mu\|\mathbf{H}\|_1. \tag{4.8}$$

N.B.: This OP can be easily extended to the multi-user case (i.e. MIMO system) by modifying properly the matrix $\tilde{\mathbf{M}}$.

4.3.2 Proposed solution

To solve the aforementionned OP, the successive convex approximation approach proposed in [92] is adopted. Basically, this approach deals with OP given by:

$$\underset{\mathbf{Z}\in\mathcal{Z}}{\text{minimize}} \ h(\mathbf{Z}) = f(\mathbf{Z}) + g(\mathbf{Z}), \tag{4.9}$$

where $f(.)$ is a smooth nonconvex function, $g(.)$ is a regularization nonsmooth function and $\mathcal{Z}$ is a convex set.

First, an upper bound of the original function h is constructed, by the standard Majorization-Maximization method, then, a convex approximation of this upper bound is defined, based on the standard SCA framework. Finally, the obtained function is minimized so that it has the same optimal points as the original one (see [92] for proofs). Also, a line search based procedure is introduced for calculating an optimal step size.

In the sequel, the following notation is adopted:

$$\mathbf{Z} = (\mathbf{H}, \mathbf{S}), \tag{4.10}$$

$$f(\mathbf{Z}) = \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|^2_F + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|^2_F, \tag{4.11}$$

$$g(\mathbf{Z}) = \mu\|\mathbf{H}\|_1. \tag{4.12}$$

One can notice that $f(\mathbf{H},\mathbf{S})$ is not jointly convex w.r.t. $(\mathbf{H},\mathbf{S})$ but it is individual convex in $\mathbf{H}$ and $\mathbf{S}$. This leads to the best OP approximation: given fixed values $\mathbf{Z}^t=(\mathbf{H}^t,\mathbf{S}^t)$ at iteration t, the original nonconvex function $f(\mathbf{Z})$ is upper bounded by a proximal convex function $\tilde{f}(\mathbf{Z},\mathbf{Z}^t)$ given by:

$$\tilde{f}(\mathbf{Z},\mathbf{Z}^t)=\tilde{f}_H(\mathbf{H},\mathbf{Z}^t)+\tilde{f}_S(\mathbf{S},\mathbf{Z}^t), \tag{4.13}$$

where

$$\tilde{f}_H(\mathbf{H},\mathbf{Z}^t)=f(\mathbf{H},\mathbf{S}^t)=\frac{1}{2}\|\bar{\mathbf{Y}}-\mathbf{H}\bar{\mathbf{S}}^t\|_F^2, \tag{4.14}$$

$$\tilde{f}_S(\mathbf{S},\mathbf{Z}^t)=f(\mathbf{H}^t,\mathbf{S})=\frac{1}{2}\|\mathbf{Y}-\mathbf{H}^t\mathbf{S}\|_F^2+\frac{\lambda}{2}\|\tilde{\mathbf{M}}\mathrm{vec}(\mathbf{S})\|_F^2, \tag{4.15}$$

and $\mathbf{S}^t$ (resp. $\mathbf{H}^t$) refers to a fixed value of the parameter $\mathbf{S}$ (resp. $\mathbf{H}$). In such a case, at iteration t, the approximate problem consists of minimizing:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}}\ \tilde{f}(\mathbf{Z},\mathbf{Z}^t)+g(\mathbf{Z}). \tag{4.16}$$

Since $\tilde{f}(\mathbf{Z},\mathbf{Z}^t)$ is a convex function w.r.t. $\mathbf{Z}$ and $g(\mathbf{Z})$ is convex in $\mathbf{H}$, the approximate problem, given by equation (4.16), is strongly convex and has a unique globally optimal solution, which is denoted by $\mathbb{B}\mathbf{Z}^t=(\mathbb{B}_H\mathbf{Z}^t,\mathbb{B}_S\mathbf{Z}^t)$. Moreover, the approximate problem (4.16) is separable w.r.t. variables $\mathbf{H}$ and $\mathbf{S}$, so that it can be decomposed into smaller problems that can be solved in parallel:

$$\mathbb{B}_H\mathbf{Z}^t = \arg\min_{\mathbf{H}}\ \tilde{f}_H(\mathbf{H},\mathbf{Z}^t)+g(\mathbf{H}), \tag{4.17}$$

$$\mathbb{B}_S\mathbf{Z}^t = \arg\min_{\mathbf{S}}\ \tilde{f}_S(\mathbf{S},\mathbf{Z}^t). \tag{4.18}$$

In order to compute $\mathbb{B}_H\mathbf{Z}^t$ while $g(\mathbf{H})$ is not differentiable, the elements of $\mathbf{H}$ are updated element-wise according to:

$$\mathbb{B}_H\mathbf{Z}^t=\mathrm{diag}(\bar{\mathbf{S}}^t\bar{\mathbf{S}}^{t^H})\mathbb{S}_\mu\left(\mathrm{diag}(\bar{\mathbf{S}}^t\bar{\mathbf{S}}^{t^H})\mathbf{H}^{t^H}-\bar{\mathbf{S}}^t(\bar{\mathbf{S}}^{t^H}\mathbf{H}^H-\bar{\mathbf{Y}}^H)\right), \tag{4.19}$$

where $\mathbb{S}_\mu(\mathbf{X})$ is an element-wise soft-thresholding function so that its complex (i,j)-th element is given by $[\mathrm{real}(X_{i,j})-\mu]^+-[-\mathrm{real}(X_{i,j})-\mu]^+ + j[[\mathrm{imag}(X_{i,j})-\mu]^+-[-\mathrm{imag}(X_{i,j})-\mu]^+]$ with $[x]^+=\max(x,0)$.

On the other hand, we have:

$$\begin{aligned}\mathbb{B}_s\mathbf{Z}^t &= \arg\min_{\mathbf{s}}\ \frac{1}{2}\|\mathbf{y}-\tilde{\mathbf{H}}^t\mathbf{s}\|_2^2+\frac{\lambda}{2}\|\tilde{\mathbf{M}}\mathbf{s}\|_F^2\\ &= (\tilde{\mathbf{H}}^H\tilde{\mathbf{H}}+\lambda\tilde{\mathbf{M}}^H\tilde{\mathbf{M}})^{-1}\tilde{\mathbf{H}}^H\mathbf{y},\end{aligned} \tag{4.20}$$

where $\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}$, $\mathbf{y} = \text{vec}(\mathbf{Y})$ and $\mathbf{s} = \text{vec}(\mathbf{S})$ (or equivalently $\mathbf{S} = \text{unvec}(\mathbf{s})$).

By using these optimal solutions, the variables update is given by:

$$\mathbf{H}^{t+1} = \mathbf{H}^t + \gamma(\mathbb{B}_H \mathbf{Z}^t - \mathbf{H}^t), \tag{4.21}$$

$$\mathbf{S}^{t+1} = \mathbf{S}^t + \gamma(\mathbb{B}_S \mathbf{Z}^t - \mathbf{S}^t), \tag{4.22}$$

where $\gamma \in [0,1]$ is the algorithm's step size (see [92] for more details).

As can be noticed, $\mathbf{H}$ and $\mathbf{S}$ are updated simultaneously at each iteration based only on the old solutions of the approximate problems (4.19) and (4.20), respectively. Also, the approximate problem (4.16) can be solved efficiently because the optimal solutions are provided in analytic expressions.

4.3.3 Optimal step size computation

One can notice that in equations (4.21) and (4.22), the choice of the step size is crucial for the convergence speed and accuracy. Therefore an optimal step size would notably improve such characteristics. For this, a line search can be adopted to obtain an optimal step size value as follows:

$$\gamma_{opt} = \arg \min_{\gamma \in [0,1]} [f(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)) + g(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t))]. \tag{4.23}$$

Although it is a scalar problem, it has no closed form solution due to the non differentiable function g. To overcome this limitation, we use the Jensen's inequality:

$$g(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)) \leq g(\mathbf{Z}^t) + \gamma g(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)). \tag{4.24}$$

One can notice that, the function on the right hand side of (4.24) is differentiable and linear with respect to γ. Hence, a closed form expression, of an approximate optimal step size, is obtained by mimimizing the following polynomial function:

$$\begin{aligned} \gamma_{opt} &= \arg \min_{\gamma \in [0,1]} [f(\mathbf{Z}^t + \gamma(\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)) + \gamma(g(\mathbb{B}_H \mathbf{Z}) - g(\mathbf{H}^t))] \\ &= \arg \min_{\gamma \in [0,1]} \{\frac{1}{4} a\gamma^4 + \frac{1}{3} b\gamma^3 + \frac{1}{2} c\gamma^2 + d\gamma\}, \end{aligned} \tag{4.25}$$

where terms independent of γ are omitted, and:

$$a = 2\|\Delta\mathbf{H}\Delta\mathbf{S}\|_F^2, \tag{4.26}$$

$$b = 3\text{tr}(\text{real}(\Delta\mathbf{H}\Delta\mathbf{S}(\mathbf{H}\Delta\mathbf{S} + \Delta\mathbf{H}\mathbf{S})^H)), \tag{4.27}$$

$$c = 2\text{tr}(\text{real}(\Delta\mathbf{H}\Delta\mathbf{S}(\mathbf{H}\mathbf{S} - \mathbf{Y})^H)) + \|\mathbf{H}\Delta\mathbf{S} + \Delta\mathbf{H}\mathbf{S}\|_2^2 + \lambda\|\tilde{\mathbf{M}}\Delta\mathbf{s}\|_2^2, \tag{4.28}$$

$$d = \text{tr}(\text{real}((\mathbf{H}\Delta\mathbf{S} + \Delta\mathbf{H}\mathbf{S})(\mathbf{H}\mathbf{S} - \mathbf{Y})^H)) + \lambda\mathbf{s}^H\tilde{\mathbf{M}}^H\tilde{\mathbf{M}}\Delta\mathbf{s} + \mu(\|\mathbb{B}_H\mathbf{Z}^t\|_1 - \|\mathbf{H}^t\|_1), \tag{4.29}$$

where tr(.) is the trace of a matrix and $\Delta\mathbf{H} = \mathbb{B}_H\mathbf{Z}^t - \mathbf{H}^t$, $\Delta\mathbf{S} = \mathbb{B}_S\mathbf{Z}^t - \mathbf{S}^t$, and $\Delta\mathbf{s} = \mathbb{B}_s\mathbf{Z}^t - \mathbf{s}^t$.

The proposed solution will be named ST-SCA for Soft Thresholding Successive Convex Approximation algorithm. This method is resumed in **Algorithm** 1.

Algorithm 1 *the proposed ST-SCA algorithm*

Initialization:

1: pilot-based initialization for $\mathbf{H}$, zero-forcing equalization for the initialization of $\mathbf{S}$ and stop criterion ϵ;

Processing:

2: Compute $\mathbb{B}_H\mathbf{Z}^t$ and $\mathbb{B}_S\mathbf{Z}^t$ according to (4.19) and (4.20);

3: Compute the optimal step size according to (4.25);

4: Update $\mathbf{H}$ and $\mathbf{S}$ according to (4.21) and (4.22);

5: While $\mid tr((\mathbb{B}\mathbf{Z}^t - \mathbf{Z}^t)^H \nabla f(\mathbf{Z}^t)) + g(\mathbb{B}\mathbf{Z}^t) - g(\mathbf{Z}^t) \mid \geq \epsilon$ repeat from step 2.

4.3.4 Alternative solutions

In what follows, and for benchmarking, the OP given by equation (4.8) can be optimized by using the widely used ADMM approach. To do so, (4.8) can be reformulated as:

$$\underset{\mathbf{A},\mathbf{B},\mathbf{S}}{minimize} \quad \tfrac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{A}\bar{\mathbf{S}}\|_F^2 + \tfrac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2 + \mu\|\mathbf{B}\|_1, \quad \text{subject to } \mathbf{A} = \mathbf{B}. \tag{4.30}$$

Consequently, the augmented Lagrangian of (4.30) is given by:

$$\mathcal{L}(\mathbf{A},\mathbf{B},\mathbf{S},\mathbf{\Pi}) = \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{A}\bar{\mathbf{S}}\|_F^2 + \frac{\lambda}{2}\|\tilde{\mathbf{M}}\text{vec}(\mathbf{S})\|_F^2 + \mu\|\mathbf{B}\|_1 + tr(\mathbf{\Pi}^H(\mathbf{A}-\mathbf{B})) + \frac{c}{2}\|\mathbf{A}-\mathbf{B}\|_F^2, \tag{4.31}$$

where the matrix $\mathbf{\Pi}$ is the dual variable (the Lagrange multiplier) which adds the constraint to the cost function and c is a positive constant. Within ADMM framework, the variables are updated as follows:

$$\mathbf{A}^{t+1} = \arg\min_{\mathbf{A}} \mathcal{L}(\mathbf{A},\mathbf{B}^t,\mathbf{S}^t,\mathbf{\Pi}^t), \tag{4.32}$$

$$\mathbf{B}^{t+1} = \arg\min_{\mathbf{B}} \mathcal{L}(\mathbf{A}^{t+1},\mathbf{B},\mathbf{S}^t,\mathbf{\Pi}^t), \tag{4.33}$$

$$\mathbf{S}^{t+1} = \arg\min_{\mathbf{S}} \mathcal{L}(\mathbf{A}^{t+1},\mathbf{B}^{t+1},\mathbf{S},\mathbf{\Pi}^t), \tag{4.34}$$

$$\mathbf{\Pi}^{t+1} = \mathbf{\Pi}^t + c(\mathbf{A}^{t+1} - \mathbf{B}^{t+1}). \tag{4.35}$$

The solutions to the above OP are as given by:

$$\mathbf{A}^{t+1} = (\bar{\mathbf{Y}}(\bar{\mathbf{S}}^t)^H + c\mathbf{B}^t - \mathbf{\Pi}^t)(\bar{\mathbf{S}}^t(\bar{\mathbf{S}}^t)^H + c\mathbf{I})^{-1}, \tag{4.36}$$

$$\mathbf{B}^{t+1} = \mathbb{S}_{\frac{\mu}{c}}(\mathbf{A}^{t+1} + \frac{(\mathbf{\Pi}^t)^H}{c}), \tag{4.37}$$

$$\mathbf{s}^{t+1} = ((\tilde{\mathbf{A}}^{t+1})^H\tilde{\mathbf{A}}^{t+1} + \lambda\tilde{\mathbf{M}}^H\tilde{\mathbf{M}})^{-1}(\tilde{\mathbf{A}}^{t+1})^H\mathbf{y}, \tag{4.38}$$

where $\mathbf{s}^{t+1} = \text{vec}(\mathbf{S}^{t+1})$ and $\tilde{\mathbf{A}} = \mathbf{I} \otimes \mathbf{A}$.

Also a BCD-based solution is considered in the sequel (for simulation comparison) for solving OP (4.8) where the channel matrix is updated row-wise, whereas the data symbols are updated by considering the channel's matrix as fixed.

It is important to notice that the numerical cost of the proposed algorithm increases significantly mainly due to the number of the observation samples. To reduce this cost, due mainly to equation (4.20), one should exploit the block-tridiagonal structure of $(\tilde{\mathbf{H}}^H\tilde{\mathbf{H}} + \lambda\tilde{\mathbf{M}}^H\tilde{\mathbf{M}})$ and its quasi block-Toeplitz property for its fast inversion [94]. Another alternative would be to replace the term associated to the data matrix structure by a term associated to the channel matrix. For example, one can use the Cross-Relation (CR) quadratic criterion [95] for blind SIMO channel estimation according to:

$$\underset{\mathbf{H},\mathbf{S}}{\text{minimize}}\ \frac{1}{2}\|\bar{\mathbf{Y}} - \mathbf{H}\bar{\mathbf{S}}\|_F^2 + \frac{\lambda}{2}\mathbf{h}^H\mathbf{Q}_{CR}\mathbf{h} + \mu\|\mathbf{H}\|_1, \tag{4.39}$$

where $\mathbf{h} = \text{vec}(\mathbf{H})$ and $\mathbf{Q}_{CR}$ is the quadratic form (obtained from the observations) associated to the CR method. In the case of large sample sizes, the minimization of the latter cost function is much cheaper than the minimization of (4.8).

4.4 Performance analysis and discussion

This section highlights the performance of the proposed solution for channel estimation and data detection. A pilot-based initialization is used for channel coefficients, which are, in turn, used for zero-forcing equalization to get the initial data values. Also, for comparison, we have used a fully pilot-based channel estimator (i.e. we assume the data symbols known for benchmarking), and a subspace-based (SS) channel and data estimators [66, 96]. The pilots and data symbols are drawn from a 4-QAM modulation, whereas the channel coefficients are generated randomly using i.i.d. unit-power, zero-mean, Gaussian distribution, with some randomly chosen null coefficients (60% of columns in our case) to model the channel sparsity. The results are averaged over 100 Monte Carlo runs, and the performance is assessed through the normalized mean squared error

Parameters	Specifications
Number of transmitters	$N_t = 1$
Number of receive antennas	$N_r = 20$
Number of data symbols	$N_d = 100$
Number of pilot symbols	$N_p = 10$
Channel's taps	$M = 15$
Cost function constants	λ=0.25 $\|\mathbf{Y}\|$ ($\|\mathbf{Y}\|$: spectral norm of $\mathbf{Y}$) μ=1.8 $c = 10^4$ (for ADMM-based algorithm)

Table 4.1: *ST-SCA simulation parameters.*

(NMSE) and an average symbol error rate (SER). Simulation parameters are summarized in TABLE 4.1, unless otherwise mentioned.

Figure 4.1 investigates the performance of channel estimation, in terms of NMSE w.r.t. SNR, of the different techniques described previously. One can notice that the proposed solution ($\mathbf{H}_{ST-SCA}$) outperforms the ADMM-based ($\mathbf{H}_{ADMM}$), the BCD-based ($\mathbf{H}_{BCD}$) and the subspace-based ($\mathbf{H}_{SS}$) solutions, while becoming close to the fully pilot-based one ($\mathbf{H}_{Pilots}$) for high SNRs.

Figure 4.2 assesses the performance of data estimation, in terms of NMSE w.r.t. SNR, of the different techniques described previously. Note that for $\mathbf{S}_{Pilots}$, a zero forcing is applied by using the estimated channel matrix $\mathbf{H}_{Pilots}$. It can be seen that the proposed solution ($\mathbf{S}_{ST-SCA}$) performs better than all other techniques $\mathbf{S}_{ADMM}$, $\mathbf{S}_{BCD}$, $\mathbf{S}_{SS}$ and $\mathbf{S}_{Pilots}$.

By using the estimated data symbols, a hard decision is performed to obtain a 4-QAM symbols. Hence, Figure 4.3 illustrates the obtained SER w.r.t. SNR of the different solutions described previously. One can notice that the results obtained in Figure 4.2 are confirmed here.

Figure 4.4 illustrates the behavior of the cost function, given in (4.8), w.r.t. the number of iterations needed for convergence. One can notice the gain obtained by using an optimal step size illustrated by ST-SCA (optimal step size), compared to the use of a fixed step size ($\gamma = 0.06$) illustrated by ST-SCA (fixed step size). On the other hand, it is shown that a similar behavior is observed for the proposed solution (ST-SCA (optimal step size)) and the BCD-based algorithm, whereas slightly lower number of iterations is observed for the ADMM-based solution. Nevertheless, small number of iterations are needed for convergence for the three techniques, in this context. However, it is worth noting that the choice of the appropriate constant c for the

ADMM-based solution remains a very hard task that influences the algorithm's convergence. Whereas for the BCD-based solution, the sequential update may incur large delays, especially for high dimensions (see TABLE 4.2), which becomes not suitable for real-time processing.

TABLE 4.2 illustrates the CPU time (in seconds) needed for one iteration of the proposed

channel's memory	ST-SCA	BCD	ADMM
M = 15	2.0296	1.0991	1.9881
M = 80	26.1254	30.7365	27.0012

Table 4.2: *CPU time (in seconds) for one iteration.*

algorithm (ST-SCA), the ADMM-based and the BCD-based solutions. In this comparison we considered a brute-force implementation of the algorithms (i.e. without exploiting the close to Toeplitz, block tridiagonal structure of the involved matrices). Compared to the ADMM-based solution, ST-SCA has similar time consumption, but still performs better in terms of channel and data estimation as illustrated in Figures 4.1, 4.2 and 4.3. However, one can notice that the BCD-based solution needs more time for longer channels (e.g. $M = 80$). Note also that the CPU time given in TABLE 4.2 has been calculated for a sequential implementation, and can be further reduced, by around the half, for ST-SCA when using parallel processing or multithreading, since the variable update at iteration $t+1$ depends only on the variables of the t-th iteration.

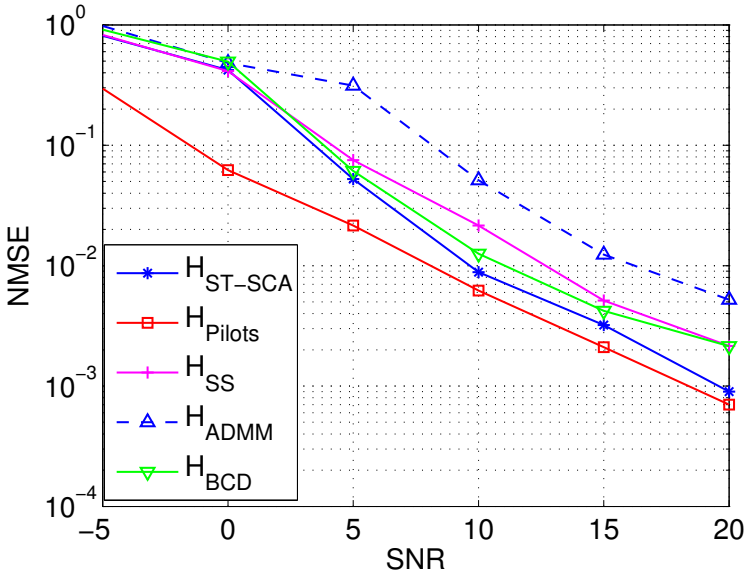


Figure 4.1: *NMSE of channel matrix* **H** *estimate vs. SNR.*

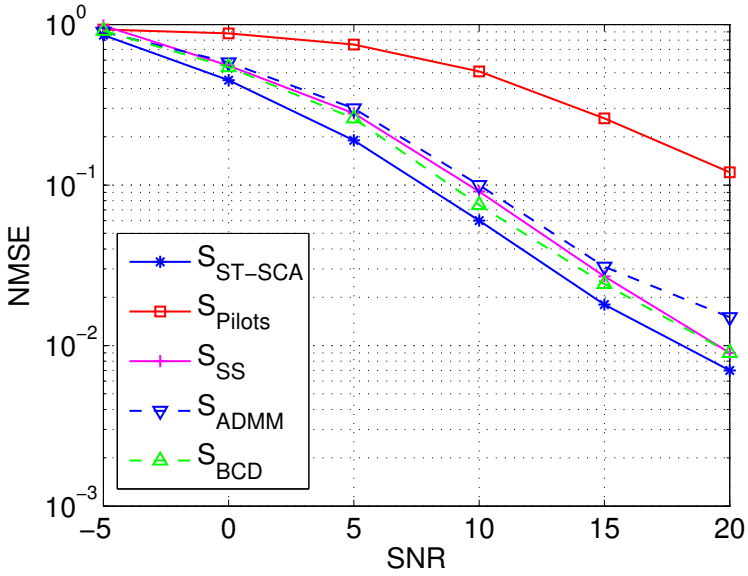


Figure 4.2: *NMSE of data matrix* **S** *estimate vs. SNR.*

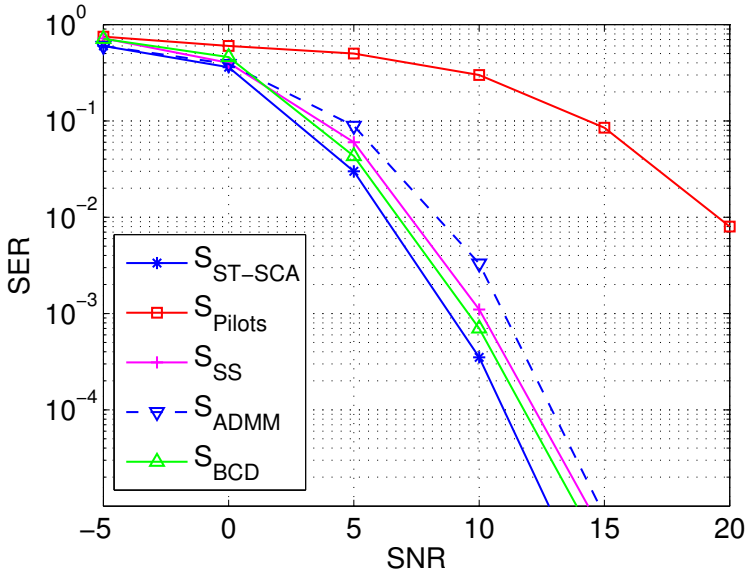


Figure 4.3: *SER vs. SNR.*

4.5 Conclusion

This chapter proposed a semi-blind solution for joint sparse channel estimation and data recovery, by solving an appropriate OP. The proposed solution is based on the successive convex approximation approach, where the optimization is performed on an approximate convex problem, rather than dealing with the original nonconvex one. An OP is formulated, based on available data (pilots) and convolutive system structure, then an iterative procedure is proposed where the

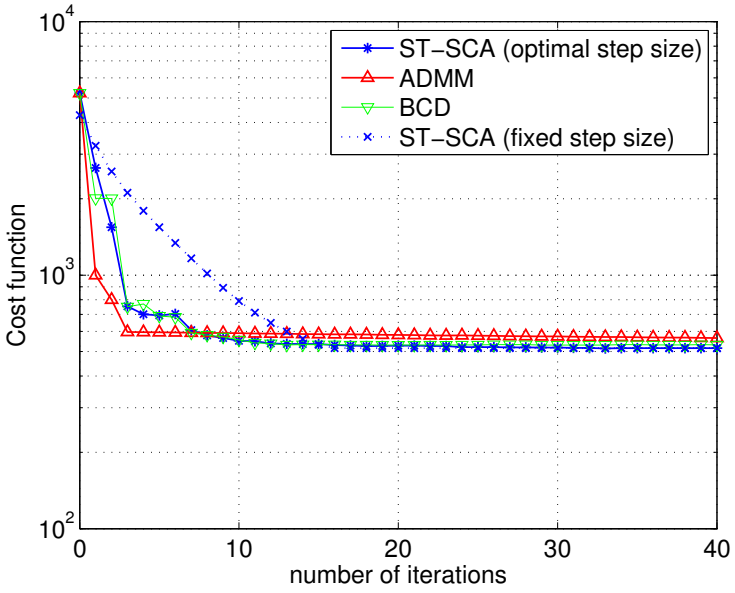


Figure 4.4: *Cost function vs. number of iterations at SNR* $= 10dB$.

channel coefficients and data symbols are updated (estimated) simultaneously, at each iteration. Along the iterations, an optimized step size procedure is introduced for convergence improvement with guaranteed convergence to a stationary point. Simulation results show that ST-SCA outperforms state-of-the-art techniques by exhibiting moderate-complexity, fast-convergence and a promizing channel and data estimation accuracy. Moreover, the adopted approach is suitable for parallel processing or multithreading since all variables are updated simultaneously.

CHAPTER 5

Blind and semi-blind multichannel identification for nonlinear communications systems

The only way of discovering the limits of the possible is to venture a little way past them into the impossible.

Clarke's Second Law.

Abstract

The aim of chapter is to propose solutions for blind and semi-blind channel estimation of nonlinear Single-Input Multiple-Output (SIMO) communications systems. For the system model, two nonlinearities have been considered; a quadratic and a cubic. In the blind case, a first channel estimation solution is proposed based on a subspace approach followed by an appropriate ambiguity removal method. Then, to refine this first estimate, an original maximum likelihood approach is introduced based on the Expectation-Maximization (EM) algorithm. In the semi-blind case, where both data and pilots are available, an extension of this EM-based solution is proposed. Some identifiability results and performance bounds related to the considered models (blind and semi-blind) are provided and discussed. Simulation results show that the proposed solutions exhibit very interesting channel estimation performance, with an attractive convergence speed for the EM-based iterative solution. A part of this work has been published in [1][97].

[1] [97] O. Rekik, K. Abed-meraim, A. Mokraoui and M. Nait-meziane,"Contribution à l'estimation aveugle du canal de transmission dans les systèmes SIMO non linéaires", in GRETSI, 2019.

Chapter content

5.1 Introduction

Nonlinear behaviors can be encountered in many practical situations, in which case appropriate (nonlinear) processing is needed, when such nonlinearities are too important to be disregarded [35, 36]. Indeed, because most of real-life systems are inherently nonlinear in nature, nonlinear problems have drawn important interest and extensive attention from engineers, physicists, mathematicians and many other scientists [36]. In communications systems, and due to the presence of nonlinear devices such as power amplifiers and optical equipments [37], communication channels are sometimes corrupted by nonlinear distortions such as nonlinear inter-symbol interference, nonlinear multiple access interference and nonlinear inter-carrier interference. These nonlinear distortions can significantly deteriorate the signal reception, leading to poor system performance. In order to overcome such an issue, nonlinear models are adopted to provide an accurate channel representation and to allow the development of efficient signal processing techniques capable of mitigating these nonlinear distortions. In the case of system identification, a widely used class of nonlinear models is the class of linear-in-the-parameters models. The input-output relation is essentially nonlinear but the estimation problem is linear with respect to the channel coefficients. Popular examples are polynomial filters, and more particularly Volterra filters [98]. They have been applied in many fields such as, electronic and electrical engineering, mechanical engineering, aeroelasticity problems and control engineering [99]. Indeed, the motivation for adopting these filters is that, they have the ability of modeling the behavior of nonlinear real-life phenomena, especially the ability to capture their "memory" effects [36]; and have mathematical relationship with other nonlinear system models namely the Wiener series, Hammerstein model, Wiener model, Wiener-Hammerstein model (block-oriented nonlinear systems), Taylor series or NARMAX model [99].

For nonlinear system identification, several approaches, most of them based on Volterra filters, have been proposed in the literature. Some works exploited training sequences and are essentially based on Least-Mean-Squares adaptive filters [100], Recursive Least-Squares algorithms [101, 102], and Affine Projection algorithms [103]. Other approaches are fully blind, thus, they seek to determine the system's kernel using the output data only. One could cite the higher order output cumulant-based approach [1], the subspace-based approach [104], the genetic programming-based method using Volterra filter [105] or the tensor-based frameworks in [106, 107]. One can notice that, these methods have been adopted and adapted to nonlinear systems mainly due to their efficiency for the linear case. Consequently, and due to the attractive advantages of the Maximum-Likelihood (ML) approaches (which is used in the current work),

namely the consistency, and the asymptotic efficiency of the estimates, some works proposed ML-based identification techniques of certain nonlinear systems [108, 109]. In these works, an approximation of the complex likelihood function is minimized via modified Gauss-Newton methods assuming the input data to be white Gaussian and a block-structured system model. However, a review of the current literature reveals that an ML solution for the case of nonlinear, finite alphabet, multi-channel communications systems does not exist.

Other works, like those in [110, 111], have considered a Hammerstein model with cascaded nonlinear and linear blocks, where the initialization and the system identification (channel estimation) are performed by firstly estimating the impulse response of the linear filter, which is then used to estimate the nonlinear function parameters. By contrast, in the proposed work, both linear and nonlinear parameters are estimated simultaneously through solutions that fit into the framework of joint channel estimation and data detection.
Also, in [112] a blind nonlinear system identification is proposed based on the parallel factors (PARAFAC) tensor decomposition. However, it is shown that the input signals must satisfy some orthogonality constraints associated with the channel nonlinearities in order to allow the desired PARAFAC decomposition. Hence, a precoding scheme is introduced using temporal redundancy on the signals, which is carried out by imposing some constraints on the symbol transitions.

The aim of the current work is to present ML-like blind and semi-blind channel estimators for Volterra-like nonlinear Single-Input Multiple-Output (SIMO) systems, that can be easily extended to MIMO scenarios. The proposed blind channel estimator combines a subspace-based estimation and an EM-based one. More precisely, firstly we exploits the Second Order Statistics (SOS) using a subspace approach for channel estimation where the nonlinear SIMO system is treated as a linear Multiple-Input Multiple-Output (MIMO) system. A straightforward motivation is that, the use of SOS-based estimators avoids the need of high number of data symbols often required for High Order Statistics based methods, e.g. [1]. Then, unlike many blind-based works (e.g. [104]), we propose also an original method to remove the ambiguity inherent to such a blind approach. Finally, a second estimation is performed based on a maximum likelihood approach, where an iterative optimization is performed using the Expectation-Maximization (EM) algorithm. Indeed, due to its sensitivity to initialization, the EM-based estimator is initialized using the subspace-based one. Note that, efficient and practical initialization for blind EM-based techniques is often missing in the literature. Moreover, within the proposed EM-based framework, one could perform a joint channel estimation and data detection as will be highlighted later in this chapter.

The global scheme of the proposed blind EM-based estimator is given in Figure 5.1, in which the signal at the received antennas is the input for the subspace-based estimator, whereas the estimates of linear and nonlinear channel coefficients represent the output of the EM algorithm after convergence.

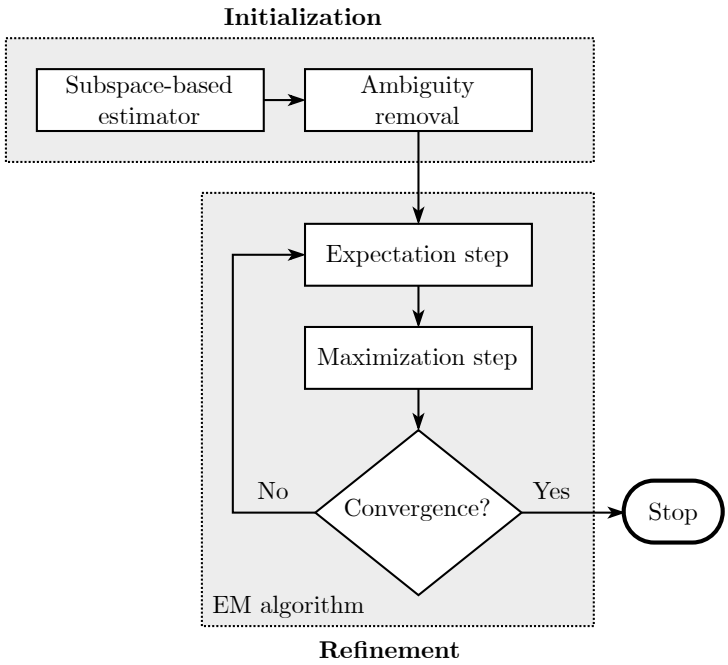


Figure 5.1: *Proposed blind channel estimation scheme.*

To make our method more flexible, and to consider the case where training sequences (pilots) are available, this work is extended to the semi-blind framework where data and pilot symbols are jointly exploited to improve the estimation accuracy and overcome certain limitations of the blind processing. In this case, the initialization of the EM algorithm is performed by exploiting the available pilots.

The proposed blind and semi-blind approaches are supported by some identifiability results and performance bounds related to our context, that allow the reader getting more insights on the problem's identifiability and its inherent performance limits.

5.2 System model

This section details the data model adopted. A nonlinear SIMO system is considered as illustrated in Figure 5.2. It is composed of one single-antenna transmitter and a receiver equipped with $N_r > 2$ antennas. The received signal at the r-th receive antenna, denoted $y_r(k)$ with $1 \leq r \leq N_r$, is

given by:

$$y_r(k) = \sum_{n=0}^{M_{r,\mathrm{L}}} h_{r,\mathrm{L}}(n)u(k-n) + \sum_{n=0}^{M_{r,\mathrm{NL}}} h_{r,\mathrm{NL}}(n)\tilde{u}(k-n) + v_r(k), \tag{5.1}$$

where $h_{r,\mathrm{L}}(n)$ (resp. $h_{r,\mathrm{NL}}(n)$) refers to the elements of the linear (resp. nonlinear), r-th receiver, channel's finite impulse response coefficient vector of size $M_{r,\mathrm{L}}+1$ (resp. $M_{r,\mathrm{NL}}+1$), $u(k)$ is the transmitted (input) symbol sequence assumed to be independent and identically distributed (i.i.d.) complex random variables taking values, with equal probabilities, in a finite alphabet set $\mathcal{A} = \{a_1, a_2, \ldots, a_{2^B}\}$ where B is the number of bits per symbol. $\tilde{u}(k)$ stands for a nonlinear combination of the input signal so that $\tilde{u}(k) = f(u(k), u(k-1), \ldots)$ (f being an appropriate nonlinear function, chosen to accurately model the system's non-linearity) and $v_r(k)$ is a white circular Gaussian noise (uncorrelated from sensor to sensor) with variance σ_v^2.

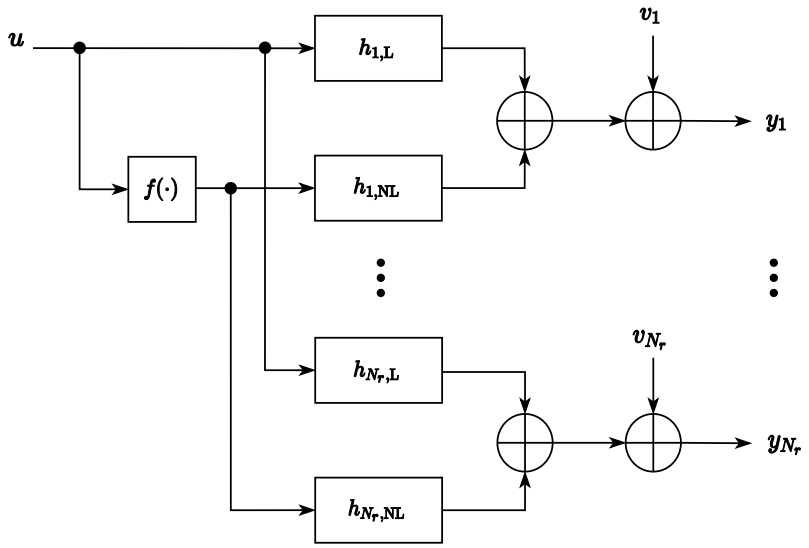


Figure 5.2: *Architecture of the considered SIMO system, where $f(.)$ is a nonlinear function w.r.t. the linear signal* **u**

.

In the sequel, two models for the nonlinear function $\tilde{u}(k)$ will be considered. The first one is a second-order (quadratic) model (e.g., [100, 113]), where $\tilde{u}(k) = u(k)^2$, whereas the second one is a third-order (cubic) model (e.g., [1, 114]) where $\tilde{u}(k) = |u(k)|^2 u(k)$. These models have been used to model real-life nonlinearties, namely those related to power amplifiers and optical devices.

Since the system model is nonlinear with respect to the transmitted signal but linear in regards to the channel coefficients, we propose to treat this nonlinear SIMO model as a linear MIMO

model with two inputs ($u(k)$ and $\tilde{u}(k)$) [115]. For this, let us define the following vectors: $\mathbf{h}_{r,\mathrm{L}} = [h_{r,\mathrm{L}}(0),\ldots,h_{r,\mathrm{L}}(M)]^T, \mathbf{h}_{r,\mathrm{NL}} = [h_{r,\mathrm{NL}}(0),\ldots,h_{r,\mathrm{NL}}(M)]^T, \mathbf{u}(k) = [u(k),u(k-1),\ldots,u(k-M)]^T$ and $\tilde{\mathbf{u}}(k) = [\tilde{u}(k),\tilde{u}(k-1),\ldots,\tilde{u}(k-M)]^T$ where $M = \max(M_{r,\mathrm{L}}, M_{r,\mathrm{NL}})$, $1 \leq r \leq N_r$. The received signal, given by equation (5.1), can be represented as:

$$y_r(k) = \mathbf{h}_r^T \bar{\mathbf{u}}(k) + v_r(k), \tag{5.2}$$

where $\mathbf{h}_r = [\mathbf{h}_{r,\mathrm{L}}^T, \mathbf{h}_{r,\mathrm{NL}}^T]^T$ and $\bar{\mathbf{u}}(k) = [\mathbf{u}(k)^T, \tilde{\mathbf{u}}(k)^T]^T$. Moreover, by considering the N_r receive antennas, one can write:

$$\mathbf{y}(k) = \mathbf{H}\bar{\mathbf{u}}(k) + \mathbf{v}(k), \tag{5.3}$$

where $\mathbf{y}(k) = [y_1(k),\ldots,y_{N_r}(k)]^T$, $\mathbf{H} = [\mathbf{h}_1,\ldots,\mathbf{h}_{N_r}]^T$ and $\mathbf{v}(k) = [v_1(k),\ldots,v_{N_r}(k)]^T$.

The system model, provided by equation (5.3), is considered as a Markov process where the state vector defined as $\mathbf{s}(k) = [u(k-1),\ldots,u(k-M)]^T$ contains M successive symbols. The set of $N = 2^{BM}$ states is denoted $\mathcal{Q} = \{\mathbf{q_1},\ldots,\mathbf{q_N}\}$. The transition vector is defined as $\mathbf{x}_{nm} = [u(k),\ldots,u(k-M)]^T$, containing $(M+1)$ symbols associated with the transition between two connected successive states $\mathbf{q}_n$ and $\mathbf{q}_m$. The set of the $2^{B(M+1)}$ possible transitions is denoted $\mathcal{X}$. For convenience, sequences of observations $\mathbf{y}(n),\ldots,\mathbf{y}(m)$ and states $\mathbf{s}(n),\ldots,\mathbf{s}(m)$ are denoted $\mathbf{Y}_{[n:m]}$ and $\mathbf{S}_{[n:m]}$, respectively. Note that the number of possible transitions and states depends only on the number of linear terms since the nonlinear terms are directly obtained from the former.

In what follows, the parameters to be estimated are the channel coefficients and the noise variance, grouped in a single vector denoted $\boldsymbol{\theta} = [\mathrm{vect}(\mathbf{H}^T)^T, \sigma_v^2]^T$.

5.3 Blind EM-based estimation

This section details the proposed blind channel estimation approach. A subspace-based estimation, for the nonlinear SIMO system, is firstly considered. Then, solutions for ambiguity removal are proposed to get rid of the inherent ambiguity of the blind processing. After that, an EM-based channel estimation is detailed, which helps refining the subspace-based estimation already performed, as illustrated in Figure 5.1. Besides, a data estimation scheme, within the EM framework, is provided.

5.3.1 Subspace-based estimation

Blind subspace-based techniques have been used successfully for channel estimation based on Second Order Statistics (SOS) in the case of linear MIMO systems [66]. In what follows, the

nonlinear SIMO system will be considered as a linear MIMO $(2 \times N_r)$ system, where the term $\tilde{u}(k)$ will be treated as the second source signal. The following assumptions are supposed to hold:

- The polynomial matrix $H(z) = [h_{\mathrm{L}}(z), h_{\mathrm{NL}}(z)] = \sum_{l=0}^{M} \mathbf{H}(l) z^{-l}$ of size $(N_r \times 2)$ is irreducible [66] and its highest order coefficient $\mathbf{H}(M)$ is full column-rank. Its (r,i)-th input is the transfer function given by $h_{r,i}(z) = \sum_{l=0}^{M} h_{r,\mathrm{L}}(l) z^{-l}$ for $i = 1$ and $h_{r,i}(z) = \sum_{l=0}^{M} h_{r,\mathrm{NL}}(l) z^{-l}$ for $i = 2$.
- The 2×2 covariance matrix of the input signal $[u(k), \tilde{u}(k)]^T$ is of full rank.

With the aforementioned assumptions, the subspace channel estimation technique introduced in [66] will be adopted. It allows us to identify the polynomial matrix $H(z)$ up to a constant matrix $\mathbf{Q}$ of size (2×2), which represents the inherent ambiguity of the blind subspace technique.

This technique exploits the received signal's SOS through the use of the covariance matrix of random vector $\mathbf{y}_w(k) = [\mathbf{y}(k)^T, \mathbf{y}(k-1)^T, \ldots, \mathbf{y}(k-N_w)^T]^T$, N_w being the window length[2] assumed sufficiently large $(N_w \geq 2M)$. In this case the channel matrix has a Sylvester block-Toeplitz structure, denoted by $\mathcal{T}_{N_w}(\mathbf{h})$ $(\mathbf{h} = \mathrm{vect}\mathbf{H}^T)$, of size $N_r(N_w+1) \times 2(M+N_w+1)$ where its first block row is given by $[\mathbf{H}(0), \ldots, \mathbf{H}(M), \mathbf{0}_{N_r \times 2}, \ldots, \mathbf{0}_{N_r \times 2}]$.

Under the previous assumptions, the noise variance σ_v^2 is the smallest eigenvalue of the covariance matrix $\mathcal{R}_w$ of $\mathbf{y}_w(k)$. The eigenspace associated with σ_v^2 is referred to as the noise subspace, which is the orthogonal complement of the signal subspace (i.e., range$\mathcal{T}_{N_w}(\mathbf{h})$ where range$\mathbf{X}$ denotes the subspace generated by the column vectors of matrix $\mathbf{X}$) and is denoted $\mathbf{\Pi}_N = \mathrm{range}\mathcal{T}_{N_w}(\mathbf{h})^{\perp}$. The subspace identification method used is ultimately related to the fact that:

$$\mathbf{\Pi}_N^H \mathcal{T}_{N_w}(\mathbf{h}) = \mathbf{0}. \tag{5.4}$$

To take into account the noise effect, one solves equation (5.4) in the Least-Squares (LS) sense, which leads to the following quadratic expression (see [66] for details):

$$\hat{\mathbf{h}}_{ss} = \arg\min_{\mathbf{h}} \ \mathbf{h}^H \mathcal{Q} \mathbf{h}, \tag{5.5}$$

where the subscript "ss" stands for subspace and $\mathcal{Q}$ is a $2N_r(M+1) \times 2N_r(M+1)$ symmetric matrix corresponding to the considered LS cost function.

Note that to avoid degenerate solutions, different constraints on $\mathbf{h}$ can be considered including the unit-norm constraint for which the solution of equation (5.5) is given by the eigenvector associated to the smallest eigenvalue of $\mathcal{Q}$.

[2] This technique processes the signal using windows of length N_w.

This method allows to estimate the polynomial matrix $H(z)$ up to a 2×2 constant matrix $\mathbf{Q}$, i.e., $\hat{H}(z) = H(z)\mathbf{Q}$. In order to use the estimated channels' vector $\hat{\mathbf{h}}_{ss}$ for further processing, one needs to remove the latter matrix ambiguity. In the sequel, solutions are proposed to get rid of $\mathbf{Q}$ by exploiting the nonlinear relation between $u(k)$ and $\tilde{u}(k)$.

5.3.2 Ambiguity removal for quadratic nonlinearity

In this section, it is assumed that $\tilde{u}(k) = u^2(k)$. At first, a channel equalization is performed based on the subspace channel estimate $\hat{\mathbf{h}}$. More precisely, a zero-forcing equalizer of delay M has been used in our simulations. It is given by the $(2M+1)$-th row[3] of the pseudo inverse matrix $\mathcal{T}^{\#}_{N_w}(\hat{\mathbf{h}})$. The obtained signal, in the noiseless case, is given by:

$$\mathbf{x}(k) = \mathbf{Q}^{-1}\mathbf{u}(k) \text{ with } \mathbf{u}(k) = [u(k), u^2(k)]^T, \tag{5.6}$$

The unknown matrix $\mathbf{Q} = (q_{i,j})_{1\leq i,j\leq 2}$ can be seen here as the separating matrix of the previous mixture, characterized (up to a diagonal) by the fact that vector $\mathbf{z}(k) = \mathbf{Q}\mathbf{x}(k)$ verifies $z_2(k) = z_1(k)^2$. Thus, in order to estimate the unknown matrix $\mathbf{Q}$, we propose to minimize the following LS criterion:

$$\sum_{k=1}^{N_s} |z_2(k) - z_1^2(k)|^2 = \|Xq\|^2, \tag{5.7}$$

where $q = [q_{2,1}, q_{2,2}, q_{1,1}^2, q_{1,2}^2, q_{1,1}q_{1,2}]^T$; N_s is the size of the equalized signal and

$$X = \begin{bmatrix} -x_1(1) & -x_2(1) & x_1^2(1) & x_2^2(1) & 2x_1(1)x_2(1) \\ \vdots & & & & \vdots \\ -x_1(N_s) & -x_2(N_s) & x_1^2(N_s) & x_2^2(N_s) & 2x_1(N_s)x_2(N_s) \end{bmatrix}.$$

Vector q is estimated up to a constant. It is proportional to the eigenvector $\mathbf{v} = [v_1, v_2, v_3, v_4, v_5]^T$ associated to the smallest eigenvalue of $\overline{\mathbf{X}}^H\overline{\mathbf{X}}$. The latter is shown to be uniquely identified under some mild assumptions on the input signal as it will be detailed in section 5.5. Consequently, it is possible to estimate $\mathbf{Q}$ as follows:

$$q = \begin{bmatrix} q_{2,1} \\ q_{2,2} \\ q_{1,1}^2 \\ q_{1,2}^2 \\ q_{1,1}q_{1,2} \end{bmatrix} = \alpha \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} \tag{5.8}$$

[3] Note that the odd rows of $\mathcal{T}^{\#}_{N_w}(\hat{\mathbf{h}})$ extract symbols $u(n)$ at different delays while even rows of $\mathcal{T}^{\#}_{N_w}(\hat{\mathbf{h}})$ extract delayed samples of $\tilde{u}(k)$.

where α is an unknown scalar factor. Hence taking into account the structure of q, one can write

$$\mathbf{Q} = \begin{bmatrix} \sqrt{\alpha v_3} & \frac{\sqrt{\alpha} v_5}{\sqrt{v_3}} \\ \alpha v_1 & \alpha v_2 \end{bmatrix} = \sqrt{\alpha} \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \tilde{\mathbf{Q}}. \tag{5.9}$$

The common scalar factor $\sqrt{\alpha}$ can be disregarded due to the scale ambiguity of such blind processing. However, for the diagonal matrix term, scalar $\lambda = \sqrt{\alpha}$ is estimated by substituting the expression (5.9) in the criterion given by equation (5.7), leading to:

$$\lambda = \frac{\sum_{k=1}^{N_s} \tilde{z}_1^2(k) \tilde{z}_2^*(k)}{\sum_{k=1}^{N_s} |\tilde{z}_2(k)|^2}, \tag{5.10}$$

where $\tilde{\mathbf{z}}(k) = \tilde{\mathbf{Q}}\mathbf{x}(k)$.

In the case of blind processing, the source signal (and also the channel coefficients) is estimated up to a constant, which represents the inherent ambiguity of the considered problem. However, it is possible to further reduce this indetermination (unknown phase multiple of $\pi/2$ for Quadrature Amplitude Modulation (QAM) signals) by exploiting the independence of the real and imaginary parts of the transmitted symbols. In our simulations, we have used the phase rotation given in [116].

5.3.3 Ambiguity removal for cubic nonlinearity

In this section, it is assumed that $\tilde{u}(k) = |u(k)|^2 u(k)$. Similarly to the previous case, after channel equalization the obtained signal, in the noiseless case, is expressed as:

$$\mathbf{x}(k) = \mathbf{Q}^{-1}\mathbf{u}(k) \text{ with } \mathbf{u}(k) = [u(k), |u(k)|^2 u(k)]^T, \tag{5.11}$$

To get rid of the unknown mixing matrix $\mathbf{Q}$, one can minimize:

$$\sum_{k=1}^{N_s} |z_2(k) - |z_1(k)|^2 z_1(k)|^2, \tag{5.12}$$

with $\mathbf{z}(k) = \mathbf{Q}\mathbf{x}(k)$. This multivariate optimization problem can be reduced to the search of one complex parameter $b = \bar{q}_{1,2}$, by normalizing $\bar{q}_{1,1} = 1$ (this is possible thanks to the inherent scale ambiguity of the considered problem) and by solving equation (5.12) with respect to $[\bar{q}_{2,1}, \bar{q}_{2,2}]$ in terms of b. Indeed, for a fixed value of b, the criterion in (5.12) reduces to a LS optimization problem with respect to $[\bar{q}_{2,1}, \bar{q}_{2,2}]$, for which a closed-form solution exists:

$$[\bar{q}_{2,1}, \bar{q}_{2,2}] = [1, b](\mathbf{X}\mathbf{X}^{\#}), \tag{5.13}$$

with $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(N_s)]$. Plugging equation (5.13) into equation (5.12) leads to a nonlinear cost function in terms of parameter b that can be solved using numerical optimization techniques.

Note that, as it will be shown in section 5.5, criterion given by equation (5.12) has spurious solutions for small or moderate size constellations (e.g., QAM 4, 16 and 32) in which case an alphabet matching cost function (see [117] and references therein) should be used instead.

5.3.4 EM-based estimation

This section is devoted to the proposed EM-based, maximum likelihood channel estimator for the system model given by equation (5.1).

The EM algorithm is an iterative method aiming at finding maximum likelihood or maximum *a posteriori* estimates of parameters in statistical models, where the model depends on unobserved latent variables. Thus, the sequence of N_s states $\mathbf{S}_{[1:N_s]}$, which are not observed, represent the missing data, whereas the N_s received symbols $\mathbf{Y}_{[1:N_s]}$ stand for the incomplete data (observations). Moreover, the complete-data is given by the sequence $(\mathbf{Y}_{[1:N_s]}, \mathbf{S}_{[1:N_s]})$. Each EM iteration alternates between two steps: E-step and M-step as described below.

5.3.4.1 E-step

The objective of this step is to find the auxiliary function, denoted by $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)})$, which is defined as the conditional expectation of the complete-data log-likelihood, with respect to the conditional distribution of the missing data $\mathbf{S}_{[1:N_s]}$, given the observations $\mathbf{Y}_{[1:N_s]}$ and the current estimated parameter value at the m-th iteration $\boldsymbol{\theta}^{(m)} = \left(\text{vect}(\mathbf{H}^{(\mathbf{m})})^T, \sigma_v^{2(m)}\right)$. Thus, such an auxiliary function can be expressed as:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) \quad = \quad E\left(\log f_\theta\left(\mathbf{Y}_{[1:N_s]}, \mathbf{S}_{[1:N_s]}\right) \middle| \mathbf{Y}_{[1:N_s]}, \boldsymbol{\theta}^{(m)}\right), \tag{5.14}$$

where $E(\cdot)$ refers to the expectation with respect to the distribution of the missing data.

After some straightforward derivations and by ignoring terms that are independent of $\boldsymbol{\theta}$, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)})$ is shown to be proportional to (see [118]):

$$\sum_{\mathbf{x}_{ij} \in \mathcal{X}} \sum_{k=1}^{N_s} \left(-N_r \log(\sigma_v^2) - \|\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{x}}_{ij}\|^2 / \sigma_v^2\right) \gamma_{\theta^{(m)}}(k; i, j), \tag{5.15}$$

where $\bar{\mathbf{x}}_{ij} = [\mathbf{x}_{ij}^T, \tilde{\mathbf{x}}_{ij}^T]^T$ ($\tilde{\mathbf{x}}_{ij}$ being the nonlinear term associated to $\mathbf{x}_{ij}$), and $\gamma_{\theta^{(m)}}(k; i, j) = f_{\theta^{(m)}}\left(\mathbf{s}(k) = \mathbf{q}_i, \mathbf{s}(k+1) = \mathbf{q}_j | \mathbf{Y}_{[1:N_s]}\right)$ represents the posterior probability of the trellis branch $(\mathbf{s}(k) = \mathbf{q}_i, \mathbf{s}(k+1) = \mathbf{q}_j)$ given the observations $\mathbf{Y}_{[1:N_s]}$ and the current estimate of the parameter $\boldsymbol{\theta}^{(m)}$. This probability can be efficiently computed using the forward-backward variables ([119, 120]) denoted by $\alpha_{\theta^{(m)}}(k; i)$ and $\beta_{\theta^{(m)}}(k; j)$. Consequently, by omitting multiplicative scaling factors independent of k, i, j, it can be shown that:

$$\gamma_{\theta^{(m)}}(k; i, j) \quad \propto \quad \alpha_{\theta^{(m)}}(k; i)\ \beta_{\theta^{(m)}}(k+1; j)\ b_{\theta^{(m)}}(k; i, j),$$

where

$$\alpha_{\theta^{(m)}}(k;i) = f_{\theta^{(m)}}(\mathbf{Y}_{[1:k-1]}, \mathbf{s}(k) = \mathbf{q}_i), \tag{5.16}$$

$$\beta_{\theta^{(m)}}(k;j) = f_{\theta^{(m)}}(\mathbf{Y}_{[k:N_s]} \mid \mathbf{s}(k) = \mathbf{q}_j), \tag{5.17}$$

$$b_{\theta^{(m)}}(k,i,j) \propto \left(\sigma_v^{2(m)}\right)^{-N_r} \exp\left(\frac{-\|\mathbf{y}(k) - \mathbf{H}^{(m)}\bar{\mathbf{x}}_{ij}\|^2}{\sigma_v^{2(m)}}\right). \tag{5.18}$$

The forward and backward variables can be evaluated recursively as:

$$\alpha_{\theta^{(m)}}(k+1;i) = \frac{1}{N}\sum_{l\in\mathcal{F}(i)} \alpha_{\theta^{(m)}}(k;l) b_{\theta^{(m)}}(k,l,i), \tag{5.19}$$

$$\beta_{\theta^{(m)}}(k;j) = \frac{1}{N}\sum_{l\in\mathcal{B}(j)} \beta_{\theta^{(m)}}(k+1;l) b_{\theta^{(m)}}(k,j,l), \tag{5.20}$$

where $\mathcal{F}(i)$ (resp. $\mathcal{B}(j)$) denotes the set of states connected to $\mathbf{q_i}$ (resp. $\mathbf{q_j}$) in forward (predecessors) (resp. backward (successors)) directions. Note that, the predecessors of the state $\mathbf{q}_i = \mathbf{s}(k) = [u(k-1),\ldots,u(k-M)]^T$ take the form $\mathbf{s}(k-1) = [u(k-2),\ldots,u(k-M-1)]^T$, whereas its successors are given by $\mathbf{s}(k+1) = [u(k),\ldots,u(k-M+1)]^T$, where the symbols $u(k)$ and $u(k-M-1)$ take values from the set $\mathcal{A}$ with equal probabilities. Hence, the total number of a state's predecessors and successors is 2^B.

5.3.4.2 M-step

The objective of this step is to find the parameter $\boldsymbol{\theta}^{(m+1)}$ that maximizes $Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}\right)$, i.e.,

$$\boldsymbol{\theta}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}\right). \tag{5.21}$$

This process is shown in [61] and [121] to increase the likelihood function, and consequently it leads to the algorithm's convergence to a global maximum point.

Since $Q\left(\boldsymbol{\theta};\boldsymbol{\theta}^{(m)}\right)$ is quadratic in its argument, the maximization step reduces to:

$$\mathbf{H}^{(m+1)} = \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}, \tag{5.22}$$

$$(\sigma_v^2)^{(m+1)} = \frac{1}{N_s N_r}\text{trace}\mathbf{R}_{yy} - \mathbf{H}^{(m+1)}\mathbf{R}_{xy}, \tag{5.23}$$

where $\mathbf{R}_{yy}$ is the autocorrelation matrix of observations, $\mathbf{R}_{xy}$ is the "weighted" cross-correlation matrix between the unobserved symbol transitions and the observations, and $\mathbf{R}_{xx}$ stands for the "weighted" auto-correlation matrix of the unobserved symbol transitions. These matrices are

given by:

$$\mathbf{R}_{yy} = \sum_{k=1}^{N_s} \mathbf{y}(k)\mathbf{y}(k)^H, \tag{5.24}$$

$$\begin{aligned}\mathbf{R}_{xy} &= \mathbf{R}_{yx}^H = \sum_{k=1}^{N_s} \sum_{\mathbf{x}_{ij}\in\mathcal{X}} \bar{\mathbf{x}}_{ij}\mathbf{y}(k)^H \gamma_{\theta^{(m)}}(k;i,j) \\ &= \sum_{k=1}^{N_s} E_{\theta^{(m)}}(\mathbf{x}(k)|\mathbf{Y}_{[1:N_s]})\mathbf{y}(k)^H, \end{aligned} \tag{5.25}$$

$$\begin{aligned}\mathbf{R}_{xx} &= \sum_{k=1}^{N_s} \sum_{\mathbf{x}_{ij}\in\mathcal{X}} \bar{\mathbf{x}}_{ij}\bar{\mathbf{x}}_{ij}^H \gamma_{\theta^{(m)}}(k;i,j) \\ &= \sum_{k=1}^{N_s} E_{\theta^{(m)}}(\mathbf{x}(k)\mathbf{x}(k)^H|\mathbf{Y}_{[1:N_s]}). \end{aligned} \tag{5.26}$$

The iterative procedure can be stopped as soon as:

$$\frac{\left\|\boldsymbol{\theta}^{(m+1)} - \boldsymbol{\theta}^{(m)}\right\|}{\left\|\boldsymbol{\theta}^{(m)}\right\|} < \epsilon, \tag{5.27}$$

for a chosen positive threshold ϵ.

5.3.5 Data detection within EM framework

Several data detection methods can be designed given the value of the estimated parameter obtained at the end of the iterations, denoted by $\boldsymbol{\theta}^{(\infty)}$; the observation sequence $\mathbf{Y}_{[1:N_s]}$ and the trellis diagram of the channel. The optimal criterion retained in the current work is the minimum symbol-error probability [122], which can be easily implemented within an EM framework. Minimizing the symbol-error probability aims at choosing, at each instant k, the data symbol which maximizes the posterior probability of the symbol $u(k)$ given the observations $\mathbf{Y}_{[1:N_s]}$ and the channel parameters $\boldsymbol{\theta}^{(\infty)}$ as follows:

$$\hat{u}(k) = \underset{a_{i_0}\in\mathcal{A}}{\arg\max}\ f_{\boldsymbol{\theta}^{(\infty)}}\left(u(k) = a_{i_0}\middle|\mathbf{Y}_{[1:N_s]};\boldsymbol{\theta}^{(\infty)}\right). \tag{5.28}$$

This quantity may be simply expressed, as a function of the posterior probability $\gamma_{\boldsymbol{\theta}^{(\infty)}}(k;i,j)$ of the trellis branch $(\mathbf{s}(k) = \mathbf{q}_i, \mathbf{s}(k+1) = \mathbf{q}_j)$ given the observations $\mathbf{Y}_{[1:N_s]}$ and $\boldsymbol{\theta}^{(\infty)}$ as:

$$f_{\boldsymbol{\theta}^{(\infty)}}\left(u(k) = a_{i_0}\middle|\mathbf{Y}_{[1:N_s]};\boldsymbol{\theta}^{(\infty)}\right) = \sum_{i,j\in\mathcal{S}(i_0)} \gamma_{\boldsymbol{\theta}^{(\infty)}}(k;i,j), \tag{5.29}$$

where $\mathcal{S}(i_0)$ is the set of all trellis branch values so that $\mathbf{x}_{ij} = [u(k) = a_{i_0}, u(k-1), \ldots, u(k-M)]$. Consequently, the EM-based solutions presented in this work can be further considered as a joint channel estimation and data detection within a maximum likelihood framework.

5.4 Semi-blind EM-based estimation

In most communications systems, some training symbols (pilots) are usually sent periodically within the wireless network frames besides the unknown data. Hence, a Semi-Blind (SB) approach, exploiting pilots, can be adopted in order to take advantage of this available data and reduce the different difficulties and issues related to the blind processing. To do so, and without loss of generality, the transmitted sequences and the received observations are assumed to be composed of N_p pilots and N_d data symbols so that $\mathbf{S}_{[1:N_s]} = [\mathbf{S}_{p_{[1:N_p]}}, \mathbf{S}_{d_{[N_p+1:N_s]}}]$ and $\mathbf{Y}_{[1:N_s]} = [\mathbf{Y}_{p_{[1:N_p]}}, \mathbf{Y}_{d_{[N_p+1:N_s]}}]$ where $N_p + N_d = N_s$ (indices p and d stand for pilot and data, respectively). Moreover, since the initialization can be performed using the available pilot symbols, the constraints on the number of receive antennas $N_r > 2$ as well as channel diversity conditions defined in section 5.2, are no longer required.

In what follows, we describe the E-step and the M-step for the case of EM-based semi-blind framework.

5.4.1 E-step

By considering pilots and data, the auxiliary function, given in equation (5.15) for the blind case, has now an additional term corresponding to the pilot sequence. The new function becomes:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(m)}) \propto \sum_{k=1}^{N_p} \left(-N_r \log(\sigma_v^2) - \frac{\|\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{u}}_p(k)\|^2}{\sigma_v^2} \right) + \sum_{\mathbf{x}_{ij} \in \mathcal{X}} \sum_{k=N_p+1}^{N_s} \left(-N_r \log(\sigma_v^2) - \frac{\|\mathbf{y}(k) - \mathbf{H}\bar{\mathbf{x}}_{ij}\|^2}{\sigma_v^2} \right) \gamma_{\theta^{(m)}}(k;i,j), \tag{5.30}$$

where $\bar{\mathbf{u}}_p(k) = [u_p(k), \ldots, u_p(k-M), \tilde{u}_p(k), \ldots, \tilde{u}_p(k-M)]^T$, of size $2(M+1)$, is composed of linear and nonlinear terms of the pilot signal.

5.4.2 M-step

In the case of semi-blind processing, equations (5.25) and (5.26) become:

$$\begin{aligned} \mathbf{R}_{yx} &= \mathbf{R}_{xy}^H = \sum_{k=1}^{N_p} \mathbf{y}(k) \bar{\mathbf{u}}_p(k)^H + \sum_{k=N_p+1}^{N_s} \sum_{\mathbf{x}_{ij} \in \mathcal{X}} \mathbf{y}(k) \bar{\mathbf{x}}_{ij}^H \gamma_{\theta^{(m)}}(k;i,j), \\ \mathbf{R}_{xx} &= \sum_{k=1}^{N_p} \bar{\mathbf{u}}_p(k) \bar{\mathbf{u}}_p(k)^H + \sum_{k=N_p+1}^{N_s} \sum_{\mathbf{x}_{ij} \in \mathcal{X}} \bar{\mathbf{x}}_{ij} \bar{\mathbf{x}}_{ij}^H \gamma_{\theta^{(m)}}(k;i,j). \end{aligned} \tag{5.31}$$

Remark: A normalization of the posterior probability of the data terms might be performed when considering such semi-blind context. To do so, one can write[4]:

$$\gamma_{\theta^{(m)}}(k;i,j) = \frac{\gamma_{\theta^{(m)}}(k;i,j)}{\sum_{\mathbf{x_{ij}} \in \mathcal{X}} \gamma_{\theta^{(m)}}(k;i,j)}. \tag{5.32}$$

[4]To avoid introducing a new notation, we kept the same expression for the normalized posterior probability $\gamma_{\theta^{(m)}}(k;i,j)$.

5.4.3 Extension to Nonlinear MIMO systems

The previous results can be easily extended to the multi-user case (i.e., MIMO system). To do so, the system model given in equation (5.1), is re-written as follows:

$$y_r(k) = \sum_{i=1}^{N_t}\sum_{n=0}^{M_{i,r,\mathrm{L}}} h_{i,r,\mathrm{L}}(n)u_i(k-n) + \sum_{i=1}^{N_t}\sum_{n=0}^{M_{i,r,\mathrm{NL}}} h_{i,r,\mathrm{NL}}(n)\tilde{u}_i(k-n) + v_r(k), \tag{5.33}$$

where $h_{i,r,\mathrm{L}}$ (resp. $h_{i,r,\mathrm{NL}}$) refers to the linear (resp. nonlinear) channel impulse response between the i-th user and the r-th receive antenna; while $u_i(k)$ represents the transmitted symbol of the i-th user.

Consequently, the system model given by equation (5.2) will be based on the following vectors: $\mathbf{h}_{r,\mathrm{L}} = [h_{1,r,\mathrm{L}}(0),\ldots,h_{1,r,\mathrm{L}}(M),\ldots,h_{N_t,r,\mathrm{L}}(M)]^T$, $\mathbf{h}_{r,\mathrm{NL}} = [h_{1,r,\mathrm{NL}}(0),\ldots,h_{1,r,\mathrm{NL}}(M),\ldots,h_{N_t,r,\mathrm{NL}}(M)]^T$, $\mathbf{u}(k) = [u_1(k),\ldots,u_1(k-M),\ldots,u_{N_t}(k-M)]^T$ and $\tilde{\mathbf{u}}(k) = [\tilde{u}_1(k),\ldots,\tilde{u}_1(k-M),\ldots,\tilde{u}_{N_t}(k-M)]^T$, where $M = \max_{i,r}\{M_{i,r,\mathrm{L}}, M_{i,r,\mathrm{NL}}\}$.

In such a case, the state vector is given by $\mathbf{s}(k) = [u_1(k-1),\ldots,u_1(k-M),\ldots,u_{N_t}(k-1),\ldots,u_{N_t}(k-M)]^T$ containing N_tM symbols, with 2^{BN_tM} possible state values. Whereas, the transition vector is defined by $\mathbf{x}_{nm} = [u_1(k),\ldots,u_1(k-M),\ldots,u_{N_t}(k),\ldots,u_{N_t}(k-M)]^T$ of $N_t(M+1)$ symbols.

By using this new vectors, equations (5.31), (5.31), (5.22) and (5.23) are still valid, leading to an EM-based channel estimation for nonlinear MIMO communications systems.

5.5 Identifiability results and performance bounds

To get more insights on the problem's identifiability and its inherent performance limits, some supplementary results are provided below.

5.5.1 Identifiability results

In the blind context, there exist certain inherent ambiguities with respect to the identification of the channel parameters. In particular, since we are using the subspace method for the initialization of our EM algorithm, we are interested in the SOS-based identifiability. Under the assumption of i.i.d. input symbols considered for the data model, the power spectral density (PSD) of the observed data is expressed as:

$$\mathbf{P}_y(e^{j2\pi f}) = [\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})] \times \mathbf{R}_u [\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})]^H + \sigma_v^2\mathbf{I}, \tag{5.34}$$

where $\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}) = [h_{1,\mathrm{L}}(e^{j2\pi f}),\ldots,h_{M,\mathrm{L}}(e^{j2\pi f})]^T$ (resp. $\mathbf{h}_{\mathrm{NL}}(e^{j2\pi f}) = [h_{1,\mathrm{NL}}(e^{j2\pi f}),\ldots,h_{M,\mathrm{NL}}(e^{j2\pi f})]^T$) is the frequency response of the linear (resp. the nonlinear) channels, while $\mathbf{R}_u$ is the 2×2 covariance matrix (assumed full rank) of $[u(k),\tilde{u}(k)]^T$. As shown in [66], by considering $u(k)$ and $\tilde{u}(k)$ as two different source signals, the SOS allow us to identify $H(z)$ up to a constant matrix $\mathbf{Q}$. Indeed, the PSD can be rewritten as

$$\mathbf{P}_y(e^{j2\pi f}) = \Big([\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})]\mathbf{Q}\Big) \times \mathbf{R}_{u,Q} \Big([\mathbf{h}_{\mathrm{L}}(e^{j2\pi f}), \mathbf{h}_{\mathrm{NL}}(e^{j2\pi f})]\mathbf{Q}\Big)^H + \sigma_v^2\mathbf{I}, \tag{5.35}$$

where $\mathbf{R}_{u,Q} = \mathbf{Q}^{-1}\mathbf{R}_u\mathbf{Q}^{-H}$. Now, to get rid of this ambiguity, we need to use higher order information through the nonlinear cost functions in equations (5.7) and (5.12). The latter help to reduce the ambiguity

from a 2×2 matrix factor to a scalar factor under certain additional assumptions given in the following lemma.

Proposition 4. *For the quadratic nonlinearity case with $\tilde{u}(k) = u^2(k)$, the minimization of the criterion given by equation (5.7) in the large sample size and noiseless case leads to the desired input signal (up to a constant factor c), i.e., $z_1(k) = cu(k)$, if and only if the correlation matrix of vector $\overline{\mathbf{u}}(k) = [1, u(k), u^2(k), u^3(k)]^T$ is full rank, i.e., $E(\overline{\mathbf{u}}(k)\overline{\mathbf{u}}(k)^H) > 0$.*

For the cubic nonlinearity with $\tilde{u}(k) = |u(k)|^2 u(k)$, the minimization of the criterion given by equation (5.12) in the asymptotic and noiseless case leads to the desired input signal (up to a constant factor), if and only if the number of possible modulus values of the input signal, denoted d, satisfies $d > 4$.

Proof. Consider an instantaneous mixture of $u(k)$ and $\tilde{u}(k)$:

$$\begin{bmatrix} z_1(k) \\ z_2(k) \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} u(k) \\ \tilde{u}(k) \end{bmatrix}.$$

For the quadratic nonlinearity case, we would like to prove that criterion given by equation (5.7) is minimum (null in the noiseless case) if and only if (iff):

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & c^2 \end{bmatrix}, \tag{5.36}$$

for a given constant c. We have

$$\begin{aligned} z_2(k) = z_1^2(k) \quad &\Longleftrightarrow \quad u(k)\mathbf{m}^T\overline{\mathbf{u}}(k) = 0 \\ &\Longleftrightarrow \quad \mathbf{m}^T\overline{\mathbf{u}}(k) = 0 \quad \text{since } u(k) \neq 0, \end{aligned}$$

where $\mathbf{m} = [-m_{21}, m_{11}^2 - m_{22}, 2m_{11}m_{12}, m_{12}^2]^T$. Hence, by taking the mean value, $E(|z_2(k) - z_1^2(k)|^2) = 0$ is equivalent to $\mathbf{m}^T E(\overline{\mathbf{u}}(k)\overline{\mathbf{u}}(k)^H)\mathbf{m}^* = 0$. This latter equality has a unique solution $\mathbf{m} = \mathbf{0}$ under the full-rank condition, i.e., $E(\overline{\mathbf{u}}(k)\overline{\mathbf{u}}(k)^H) > 0$. Finally, the vector $\mathbf{m}$ is null iff equation (5.36) holds.

For the cubic nonlinearity, we would like to prove that the criterion given by equation (5.12) is null (in the noiseless case) iff:

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} c & 0 \\ 0 & |c|^2 c \end{bmatrix}, \tag{5.37}$$

for a given constant c. We have:

$$\begin{aligned} z_2(k) = |z_1(k)|^2 z_1(k) \quad &\Longleftrightarrow \quad u(k)m^T\overline{u}(k) = 0 \\ &\Longleftrightarrow \quad m^T\overline{u}(k) = 0 \quad \text{since } u(k) \neq 0, \end{aligned}$$

where $m = [-m_{21}, (m_{11}|m_{11}|^2 - m_{22}), 2m_{11}\text{Real}(m_{11}m_{12}^*) + m_{12}|m_{11}|^2, m_{11}|m_{12}|^2 + 2m_{12}\text{Real}(m_{11}m_{12}^*), m_{12}|m_{12}|^2]^T$ and $\overline{u}(k) = [1,\ |u(k)|^2,\ |u(k)|^4,\ |u(k)|^6,\ |u(k)|^8]^T$. Now, the equation system $m^T\overline{u}(k) = 0$, for all k will have a unique solution $m = 0$, if the Vandermonde-like matrix formed by vectors $\overline{u}(k)$ has a full (row) rank equal to 5. This is the case iff symbols $u(k)$ have at least $d > 4$ different modulus values. Finally, vector m is null iff (5.37) holds. □

For M-QAM modulations, one can easily check that in the case of quadratic nonlinearity, the lemma's condition is met for $M > 4$. However, for the cubic nonlinearity, the lemma's condition is quite restrictive and requires large modulation sizes with $M > 64$. In such a case, when the modulation size is small or moderate (i.e. $M \leq 64$), the ambiguity removal requires the use of another criterion such as the alphabet matching one [117].

Also, note that when the polynomial degrees of the nonlinear and linear channels are not the same, i.e., $\deg(h_{\mathrm{L}}(z)) \neq \deg(h_{\mathrm{NL}}(z))$, the subspace method would identify the channel matrix $H(z)$ up to a certain 2×2 polynomial matrix $Q(z)$ (see [66] for details). In such a case, the proposed ambiguity removal method does not apply and consequently the EM algorithm's initialization might be ineffective leading to potential ill-convergence of the considered blind algorithm. All these issues can be avoided in the semi-blind context where the knowledge of pilot signals can be exploited to initialize our EM algorithm but also to remove the previous blind processing indeterminations. Next, we derive the deterministic[5] Cramér Rao Bound (CRB) relative to the SB context, that will be used later for our algorithm's performance benchmarking.

5.5.2 Deterministic Cramér-Rao Bound (CRB)

Given a parametric statistical model, the CRB provides a lower bound of the error variance for all unbiased estimators of the system's parameter vector. In particular, the Gaussian CRB (G-CRB) represents a lower bound within the class of estimators using only the SOS of the observed data. It is also the least favorable CRB as shown in [123]. In the sequel, we derive the expression of the deterministic G-CRB for our system model.

The data model in equation (5.1) can be rewritten in a more compact way by considering all data samples N_s and all outputs N_r as:

$$\mathbf{y} = \mathcal{H}_{\mathrm{L}}\,\mathbf{u}_{\mathrm{L}} + \mathcal{H}_{\mathrm{NL}}\,\mathbf{u}_{\mathrm{NL}} + \mathbf{v} = \mathcal{U}_{\mathrm{L}}\,\mathbf{h}_{\mathrm{L}} + \mathcal{U}_{\mathrm{NL}}\,\mathbf{h}_{\mathrm{NL}} + \mathbf{v}, \tag{5.38}$$

where $\mathbf{y} = [y_1(0),\ldots,y_1(N-1),\ldots,y_{N_r}(0),\ldots,y_{N_r}(N-1)]^T$ ($N = N_p + N_s$ the total number of transmitted pilot and data symbols), $\mathbf{v} = [v_1(0),\ldots,v_1(N-1),\ldots,v_{N_r}(0),\ldots,v_{N_r}(N-1)]^T$, $\mathbf{u}_{\mathrm{L}} = [u(0),\ldots,u(N-1)]^T$, $\mathbf{u}_{\mathrm{NL}} = [\tilde{u}(0),\ldots,\tilde{u}(N-1)]^T$, $\mathbf{h}_{\mathrm{L}} = [h_{1,\mathrm{L}}(0),\ldots,h_{1,\mathrm{L}}(M),\ldots,h_{N_r,\mathrm{L}}(0),\ldots,h_{N_r,\mathrm{L}}(M)]^T$, $\mathbf{h}_{\mathrm{NL}} = [h_{1,\mathrm{NL}}(0),\ldots,h_{1,\mathrm{NL}}(M),\ldots,h_{N_r,\mathrm{NL}}(0),\ldots,h_{N_r,\mathrm{NL}}(M)]^T$, $\mathcal{H}_{\ddagger} = [\mathbf{H}^{\mathbf{T}}_{\mathbf{1},\ddagger},\ldots,\mathbf{H}^{\mathbf{T}}_{\mathbf{N_r},\ddagger}]^{\mathbf{T}}$, $\mathcal{U}_{\ddagger} = I \otimes \mathbf{U}_{\ddagger}$ ($\ddagger = \mathrm{L}$ or NL) where $\otimes$ denotes the Kronecker product, and $\mathbf{H}_{r,\ddagger}$ and $\mathbf{U}_{\ddagger}$ are defined as ($u_{\ddagger}(k)$ equals $u(k)$ if $\ddagger = \mathrm{L}$ or $\tilde{u}(k)$ if

[5] The input symbols are treated as deterministic unknown parameters.

$\ddagger = \text{NL}$):

$$\mathbf{H}_{r,\ddagger} = \begin{bmatrix} h_{r,\ddagger}(M) & \cdots & h_{r,\ddagger}(0) & & 0 \\ & \ddots & & \ddots & \\ 0 & & h_{r,\ddagger}(M) & \cdots & h_{r,\ddagger}(0) \end{bmatrix}_{N\times(N+M)}, \tag{5.39}$$

$$\mathbf{U}_{\ddagger} = \begin{bmatrix} u_{\ddagger}(0) & \cdots & u_{\ddagger}(-M) \\ & \ddots & \vdots \\ \vdots & & u_{\ddagger}(0) \\ & & \vdots \\ u_{\ddagger}(N_s-1) & \cdots & u_{\ddagger}(N_s-1-M) \end{bmatrix}_{N\times(M+1)}. \tag{5.40}$$

Since we assumed a semi-blind approach, we can write $\mathbf{u}_{\ddagger} = [\mathbf{u}_{\ddagger,p}^T, \mathbf{u}_{\ddagger,d}^T]^T$ where $\mathbf{u}_{\ddagger,p}$ contains pilot-data samples and $\mathbf{u}_{\ddagger,d}$ contains unknown-data samples. We consider the vector of unknown parameters to be $\theta = [\mathbf{h}_{\text{L}}^T, \mathbf{h}_{\text{NL}}^T, \mathbf{u}_d^T]^T$ (of size $2(M+1)N_r + N_s$) where $\mathbf{u}_d = \mathbf{u}_{\text{L},d}$.

The *unconstrained complex* Fisher Information Matrix (FIM) is defined as:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta\theta}} & \mathbf{J}_{\boldsymbol{\theta\theta}^*} \\ \mathbf{J}_{\boldsymbol{\theta}^*\boldsymbol{\theta}} & \mathbf{J}_{\boldsymbol{\theta}^*\boldsymbol{\theta}^*} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta\theta}} & \mathbf{J}_{\boldsymbol{\theta\theta}^*} \\ (\mathbf{J}_{\boldsymbol{\theta\theta}^*})^H & (\mathbf{J}_{\boldsymbol{\theta\theta}})^* \end{bmatrix}, \tag{5.41}$$

where

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = \frac{1}{\sigma_v^2} \left(\frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\theta}} \right)^H \frac{\partial \tilde{\boldsymbol{\mu}}}{\partial \boldsymbol{\theta}}, \tag{5.42}$$

$\tilde{\boldsymbol{\mu}} = [\boldsymbol{\mu}^T, \boldsymbol{\mu}^H]^T$, $\boldsymbol{\mu} = \mathcal{H}_{\text{L},p}\mathbf{u}_p + \mathcal{H}_{\text{L},d}\mathbf{u}_d + \mathcal{H}_{\text{NL},p}\tilde{\mathbf{u}}_p + \mathcal{H}_{\text{NL},d}\tilde{\mathbf{u}}_d$, and $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ ($\mathbf{f} \in \mathbb{C}^{m\times 1}, \mathbf{x} \in \mathbb{C}^{n\times 1}$) denotes the differentiation operator defined as:

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}, \tag{5.43}$$

where the matrix elements represent Wirtinger's derivatives [124]. Note from equation (5.41) that we only need to find $\mathbf{J}_{\boldsymbol{\theta\theta}}$ and $\mathbf{J}_{\boldsymbol{\theta\theta}^*}$ in order to find $\mathbf{J}$. After derivation, we find that:

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = \frac{1}{\sigma_v^2} \begin{bmatrix} \mathcal{U}_{\text{L}}^H\mathcal{U}_{\text{L}} & \mathcal{U}_{\text{L}}^H\mathcal{U}_{\text{NL}} & \mathcal{U}_{\text{L}}^H\boldsymbol{\Lambda} \\ \mathcal{U}_{\text{NL}}^H\mathcal{U}_{\text{L}} & \mathcal{U}_{\text{NL}}^H\mathcal{U}_{\text{NL}} & \mathcal{U}_{\text{NL}}^H\boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}^H\mathcal{U}_{\text{L}} & \boldsymbol{\Lambda}^H\mathcal{U}_{\text{NL}} & \boldsymbol{\Lambda}^H\boldsymbol{\Lambda} + \boldsymbol{\Gamma}^T\boldsymbol{\Gamma}^* \end{bmatrix}, \tag{5.44}$$

$$\mathbf{J}_{\boldsymbol{\theta\theta}^*} = \frac{1}{\sigma_v^2} \begin{bmatrix} 0 & 0 & \mathcal{U}_{\text{L}}^H\boldsymbol{\Gamma} \\ 0 & 0 & \mathcal{U}_{\text{NL}}^H\boldsymbol{\Gamma} \\ \boldsymbol{\Gamma}^T\mathcal{U}_{\text{L}}^* & \boldsymbol{\Gamma}^T\mathcal{U}_{\text{NL}}^* & \boldsymbol{\Lambda}^H\boldsymbol{\Gamma} + (\boldsymbol{\Lambda}^H\boldsymbol{\Gamma})^T \end{bmatrix}, \tag{5.45}$$

where $\boldsymbol{\Lambda} = \mathcal{H}_{\text{L},d} + 2\mathcal{H}_{\text{NL},d}\text{diagabs}\mathbf{u}_d{}^2$, and $\boldsymbol{\Gamma} = \mathcal{H}_{\text{NL},d}\text{diag}\mathbf{u}_d^2$ for the cubic case. For the quadratic case, we found $\mathbf{J}_{\boldsymbol{\theta\theta}^*} = \mathbf{0}, \boldsymbol{\Lambda} = \mathcal{H}_{\text{L},d} + 2\mathcal{H}_{\text{NL},d}\text{diag}\mathbf{u}_d$, and $\boldsymbol{\Gamma} = \mathbf{0}$.

The G-CRB is computed as the inverse[6] of equation (5.41). The G-CRB for the channel parameters is given by the top left submatrix of the global G-CRB.

5.6 Generalization and numerical complexity

This section is dedicated to the extension of our methods to other non-linear models. It provides also details about the computational complexity of our algorithms.

5.6.1 Generalization

The proposed semi-blind estimator can be easily generalized to other, finite memory, nonlinear models, like the one considered in [126]. Indeed, since the initialization is performed by using the pilots, and the state vector, exploited in the EM procedure, depends only on the linear terms, the extension remains possible upon certain derivation adaptations. However, for the blind case, it is feasible only when using 'another' appropriate initialization, since the main difficulty comes from the ambiguity removal of the blind initialization (subspace-based in our case). To do that, one could rely on some algorithms given in the literature (e.g., [126]). In fact, by assuming a linear-in-the-parameters model, a known nonlinear function and a finite impulse response, the proposed techniques can be applied to the following general model (given also in [113]):

$$y_r(k) = \sum_{n=0}^{M_{r,\mathrm{L}}} h_{r,\mathrm{L}}(n)u(k-n) + v_r(k) + \sum_{n=0}^{M_{r,\mathrm{NL}}} \sum_{l=n}^{M_{r,\mathrm{NL}}} h^{(2)}_{r,\mathrm{NL}}(n,l)u(k-n)u(k-l)$$

$$+ \sum_{n}\sum_{l}\sum_{m} h^{(3)}_{r,\mathrm{NL}}(n,l,m)u(k-n)u(k-l)u^*(k-m) + \cdots \tag{5.46}$$

Furthermore, by considering the input-output relation, a block-oriented nonlinear model (e.g., Hammerstein model) can be expressed (approximated) as in (5.46) and hence, treated by the proposed techniques. However in this case, the estimated coefficients would represent products of the block-oriented linear and nonlinear parameters.

5.6.2 Algorithms' complexity discussion

It can be seen that the parameter estimation given by equations (5.22) and (5.23) requires, at each iteration, the calculation of the different conditional expectations given in equations (5.25), (5.26), (5.31) and (5.31). Hence, the global complexity of the proposed techniques is of order $\mathcal{O}(N_{iter}N_s 2^{BN_t(M+1)} 2N_t(M+1)(4N_t(M+1)+2N_r))$, where N_t is the number of transmitters and N_{iter} is the total number of iterations needed for convergence.

It can be noticed that the algorithms' computational complexity is of the same order as the linear case[7].

[6] Note that, in the blind case, the FIM is singular (due to the problem's ambiguities), in which case one needs to rely on the constrained CRB [125].

[7] For our model, there is approximately a factor 2 between the costs of the NL and the L cases.

It is worth pointing out that the computational complexity of our EM-based algorithm is mainly affected by the number of transitions, given by $2^{BN_t(M+1)}$. However, as it will be seen through simulation results, a relatively small number of pilots, data symbols and iterations are needed for convergence.
On the other hand, to reduce the computational complexity, one can apply some approximations for the posterior probability (as in [118] and references therein), use independent symbols through OFDM coding to avoid the forward-backward variables calculation or exploit some approximation/simplification approaches as proposed in [127]. Such complexity reduction is left for future work.

5.7 Performance analysis and discussion

This section provides the performance analysis of the proposed blind and semi-blind channel estimators for the considered nonlinear systems. For benchmarking, we consider a 'full' training-based (fully-pilot) estimator, as done in many works (e.g., [113]), where all transmitted symbols (pilots and data) are assumed known and used to estimate the channel parameters. The estimation performance is evaluated in terms of the Normalized Root-Mean-Square Error (NMSE) given by:

$$\text{NMSE} = \frac{1}{N_{mc}} \sum_{mc=1}^{N_{mc}} \frac{\left\| \hat{\mathbf{h}}_{mc} - \mathbf{h}_{exact} \right\|^2}{\left\| \mathbf{h}_{exact} \right\|^2}, \tag{5.47}$$

where $N_{mc} = 500$ represents the number of independent Monte-Carlo runs used, $\hat{\mathbf{h}}_{mc}$ is the vector of estimated channel parameters at the mc-th run, and $\mathbf{h}_{exact}$ contains the true (exact) values of the channel coefficients. For data detection, the performance is evaluated in terms of the Symbol Error Rate (SER), which is the ratio between the wrongly detected symbols and the total number of transmitted data symbols. The channel coefficients are generated as i.i.d., unit-power, zero-mean (complex) Gaussian random variables, whereas the pilots and data symbols are uniformly randomly drawn from different QAM modulations (specified later for each experiment). The Signal-to-Noise Ratio (SNR) was defined as $\text{SNR} = \|\mathbf{H}\overline{\mathbf{u}}\|_F^2 / \|\mathbf{v}\|_F^2$, where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{H}$ is defined in equation (5.3), and $\overline{\mathbf{u}} = [\overline{\mathbf{u}}(1), \ldots, \overline{\mathbf{u}}(N_s)] \in \mathbb{C}^{2(M+1)\times N_s}$ and $\mathbf{v} = [\mathbf{v}(1), \ldots, \mathbf{v}(N_s)] \in \mathcal{C}^{N_r \times N_s}$ are formed by stacking all values of $\overline{\mathbf{u}}(k)$ and $\mathbf{v}(k)$ (see equation (5.3)), respectively.

In the following, different experiments highlighting different aspects of our estimators are presented. Simulation parameters are summarized in Table 5.1. They are used for all experiments, unless otherwise specified.

Experiment 1: Effect of neglecting the nonlinear terms (Figures 5.3–5.4)
In order to illustrate the effect of ignoring the nonlinear term and considering only a linear model, the system model given by (5.1) is rewritten as follows:

$$y_r(k) = \sum_{n=0}^{M_{r,\text{L}}} h_{r,\text{L}}(n) u(k-n) + \alpha \sum_{n=0}^{M_{r,\text{NL}}} h_{r,\text{NL}}(n) \tilde{u}(k-n) + v_r(k), \tag{5.48}$$

where α determines the weight of the nonlinear term that exists in the assumed underlying model (e.g., $\alpha = 0$: the underlying model is linear, $\alpha = 1$: the weight of the nonlinear term is equal to the linear one).

Table 5.1: *Nonlinear models simulation parameters*

Parameter	Specification
Number of data symbols	$N_s = 100$
Number of pilot symbols	$N_p = 10$
Number of receive antennas	$N_r = 4$
Window length for SS	$N_w = 5$
Order of linear and nonlinear channels	$M = 4$

Figure 5.3 illustrates the NMSE vs. SNR behave when there is a nonlinearity (assuming different weights α indicated in superscript) in the underlying model. One can notice that, depending on the weight of the nonlinear term, the performance of the linear EM-based blind and semi-blind estimators could be highly degraded.

Furthermore, Figure 5.4 illustrates the SER vs. the weight α, for SNR = 5 dB. We notice a clear gain (especially for high α values) obtained by considering a nonlinear model (NL-B, NL-SB) as compared to the case where only linear terms are taken into account (L-B, L-SB).

In the sequel, the nonlinear model will be considered by setting $\alpha = 1$.

Experiment 2: NMSE vs. SNR (Figures. 5.5–5.8)

Figure 5.5 investigates the performance, in terms of NMSE, of the subspace(SS)-based estimator ($\mathrm{h_{SS}}$), and the blind and semi-blind EM-based estimators ($\mathrm{h_{EM-B}}, \mathrm{h_{EM-SB}}$) with respect to SNR for the quadratic nonlinearity considering 4-QAM (Figure 5.5a) and 16-QAM (Figure 5.5b) modulations. These estimators are benchmarked against the semi-blind Gaussian CRB (G-$\mathrm{CRB_{SB}}$) and the fully-pilot-based estimator ($\mathrm{h_{PILOT}}$). The blind EM-based estimator is initialized by the subspace-based one, whereas the semi-blind estimator is initialized by some pilots. One can observe that $\mathrm{h_{SS}}$ presents sub-optimal performance compared to $\mathrm{h_{PILOT}}$ (to all other estimators as well), whereas a significant improvement is observed with $\mathrm{h_{EM-B}}$ and $\mathrm{h_{EM-SB}}$. Moreover, the exploitation of the available pilots enhances the performance as $\mathrm{h_{EM-SB}}$ outperforms $\mathrm{h_{EM-B}}$ and hugs very tightly $\mathrm{h_{PILOT}}$ for moderate and high SNRs. On the other hand, the EM-based estimators are found, interestingly, below the G-$\mathrm{CRB_{SB}}$, which reflects the fact that our channel identification solution outperforms all SOS-based identification methods. This result strongly supports the effectiveness of the aforementioned estimators for non-Gaussian QAM signals.

Also, a comparison has been performed with a cumulant-based technique [1], as illustrated in Figure 5.5a and Figure 5.5b, where, in the context of our scenarios, poor performance has been obtained since such a method requires higher number of symbols for convergence (at least 16000 data symbols as mentioned in [1]).

As mentioned in section 5.4.3, the proposed estimators can be easily applied to MIMO systems, where similar performance to the SIMO case is observed (see Figure 5.6).

From Figure 5.7, one can observe that an important spatial diversity (higher number of receive antennas, as is the case for massive MIMO systems) improves the performance for both $\mathrm{h}_{\mathrm{EM-B}}$ and $\mathrm{h}_{\mathrm{EM-SB}}$ at low SNRs.

In Figure 5.8, a cubic nonlinear model has been used with a 16-QAM modulated input signal. One can note similar performance as provided in Figures. 5.5–5.7; EM-based estimators outperform the SS-based estimator and all SOS-based estimators (represented by the G-$\mathrm{CRB}_{\mathrm{SB}}$) for moderate to high SNRs, and $\mathrm{h}_{\mathrm{EM-SB}}$ outperforms $\mathrm{h}_{\mathrm{EM-B}}$ as it hugs more tightly the $\mathrm{h}_{\mathrm{PILOT}}$ curve.

Note that using a 4-QAM signal modulation with our cubic model will render the model linear, leading to a loss of identifiability. This can be easily seen by writing the 4-QAM sequence as $u(k) = \sqrt{2}e^{j(\frac{\pi}{4}+\eta(k)\frac{\pi}{2})}$ where $\eta(k) \sim \mathcal{U}\{0,1,2,3\}$, then, $\tilde{u}(k) = abs u(k)^2 u(k) = 2u(k)$, which is linear.

Experiment 3: NMSE vs. SNR with different linear and nonlinear channel orders (Figure 5.9)

In many practical situations, the channel order of the linear and nonlinear channels are different. Figure 5.9 illustrates the behavior of two different scenarios considered in *Experiment 1* with a channel order $M_{\mathrm{NL}} = 2$ for the nonlinear channel (the last two channel coefficients out of the previously used five coefficients have been considered null). On can note, particularly, that for high SNRs h_{SS} performs badly affecting the performance of $\mathrm{h}_{\mathrm{EM-B}}$.

Experiment 4: Speed of convergence (number of iterations vs. SNR, Figures. 5.10–5.11)

Figure 5.10 shows the number of iterations needed for convergence for blind (B) and semi-blind (SB) EM-based estimators with respect to SNR. Considering 4-QAM and 16-QAM signal modulations and the considered nonlinear models (Quadratic and Cubic), we observe that for low SNRs ($0-5$ dB), the number of iterations varies according to the signal's model and modulation but is still relatively small (less than 30 at 5 dB). For moderate to high SNRs ($>$ 10 dB), very few iterations (less than 8) are needed. In fact, for high SNRs ($\geq$ 15 dB) only 2 or 3 iterations are needed independently of the signal's model and modulation. Moreover, using a higher number of antennas at the receiver (e.g., as in massive MIMO systems) leads to further reducing the number of iterations needed for convergence, especially at low SNRs, as can be seen from Figure 5.11.

Experiment 5: NMSE vs. number of iterations at fixed SNR (Figure 5.12)

Figure 5.12 illustrates the variation of the NMSE with respect to the number of iterations needed for convergence of the subspace(SS)-based estimator (h_{SS}), and the blind and semi-blind EM-based estimators ($\mathrm{h}_{\mathrm{EM-B}}, \mathrm{h}_{\mathrm{EM-SB}}$) at SNR = 10 dB. Note that $\mathrm{h}_{\mathrm{EM-B}}$ is initialized by h_{SS}, whereas $\mathrm{h}_{\mathrm{EM-SB}}$ is initialized by some pilots such that $\mathrm{h}_{\mathrm{EM-B}}$ and $\mathrm{h}_{\mathrm{EM-SB}}$ start from the same initial point (i.e., no iterations yet). We observe that after two iterations, $\mathrm{h}_{\mathrm{EM-SB}}$ converges to the solution given by the fully-pilot-based estimator ($\mathrm{h}_{\mathrm{PILOT}}$). We also observe that $\mathrm{h}_{\mathrm{EM-B}}$ comes very close to $\mathrm{h}_{\mathrm{PILOT}}$ after being initialized by h_{SS}. This result illustrates better the use of the word "refinement" in Figure 5.1, for the blind EM-based processing and is very interesting with regards to the computational complexity, especially for the nonlinear case where the number of channel coefficients is, in general, higher than the linear case.

Experiment 6: NMSE vs. number of pilots N_p (Figure 5.13)

We consider the quadratic model and 4-QAM modulation and we investigate the impact of the number of pilots N_p on the NMSE of the semi-blind EM-based estimator $\mathrm{h}_{\mathrm{EM-SB}}$, at SNR = 10 dB (Figure 5.13a; h_{SS} and $\mathrm{h}_{\mathrm{EM-B}}$ are included for reference). We observe (Figure 5.13b) that only a slight decrease (around 3×10^{-3}) in the NMSE accompanies the increase in the number of pilots from 1 to 30, which indicates that only few pilots are needed to allow a quasi-optimal semi-blind channel estimation within the EM framework.

Experiment 7: Symbol Error Rate (SER) vs. SNR (Figure 5.14)

Figure 5.14 investigates the performance of the proposed estimators in terms of Symbol Error Rate (SER) with respect to SNR. For blind and semi-blind EM-based estimators (EM-B and EM-SB), a data detection is performed within an EM-based framework as described in section 5.3.5. For the subspace (SS) and the fully-pilots approaches, a zero-forcing is applied using the estimated channel coefficients. It can be noticed that performing data detection, using a ML-based approach leads to a significant performance gain and is part of a joint channel estimation and data detection.

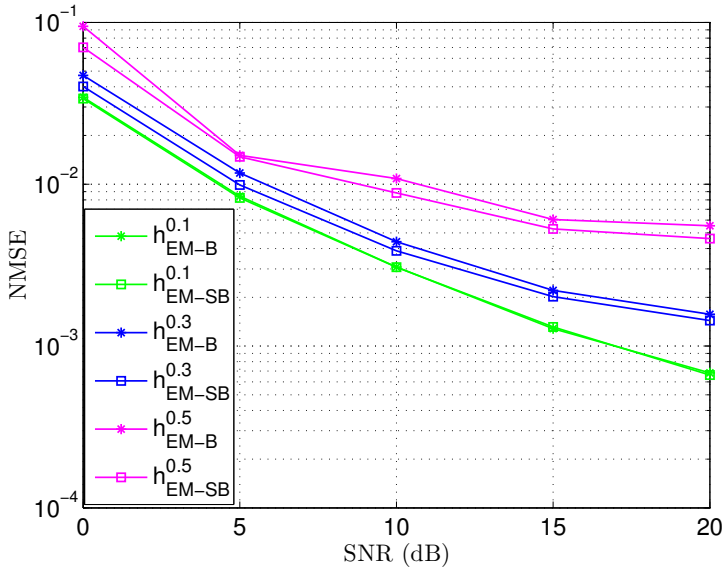


Figure 5.3: *NMSE vs. SNR with different values of the weight α (see (5.48)), indicated in superscript, for the blind and semi-blind 'linear' EM-based estimators.*

5.8 Conclusion

In this chapter, a Maximum Likelihood (ML) solution is presented for the identification of nonlinear multichannel communications systems. The ML criterion is maximized through the Expectation-Maximization (EM) algorithm. In the blind case, the EM algorithm is initialized by the subspace method followed by an original ambiguity removal technique introduced in this work. However, an identifiability study reveals that the success of the initialization step requires some stringent conditions that might not be verified for low order QAM modulations. An alternative solution is proposed, based on the Semi-Blind (SB) approach.

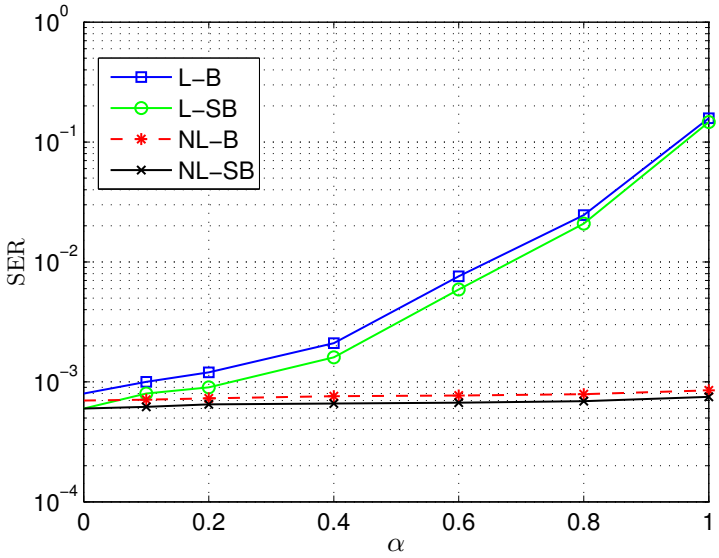
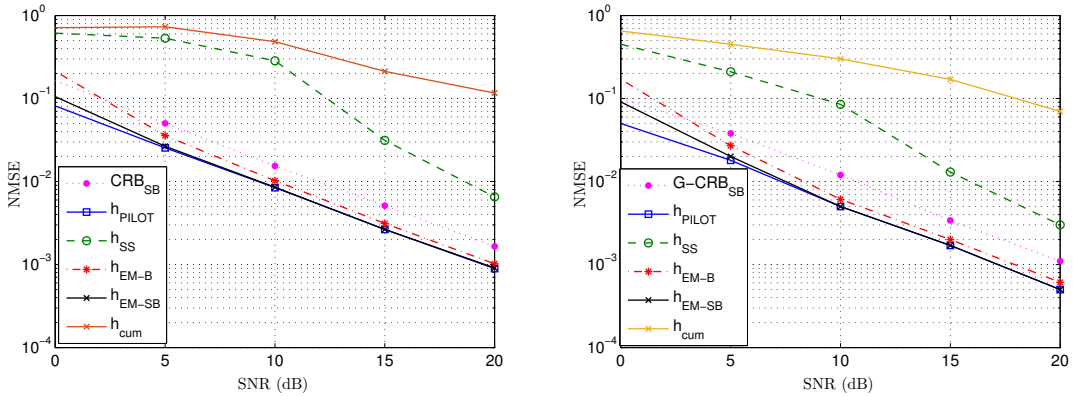


Figure 5.4: *SER vs. α with SNR $= 5$ dB for linear blind and semi-blind (L-B, L-SB) EM-based estimators and for nonlinear (NL-B, NL-SB) ones.*



(a) **(b)**

Figure 5.5: *NMSE vs. SNR for subspace-based estimator (h_{SS}), blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the semi-blind Gaussian CRB (G-CRB$_{SB}$), the cumulant-based technique [1] (h_{cum}) and the fully-pilot-based estimator (h_{PILOT}), in both cases: (a) quadratic model and 4-QAM modulation, and (b) quadratic model and 16-QAM modulation.*

The EM extension to the SB context is first provided followed by simulation experiments that highlight the efficiency of our EM-based method.

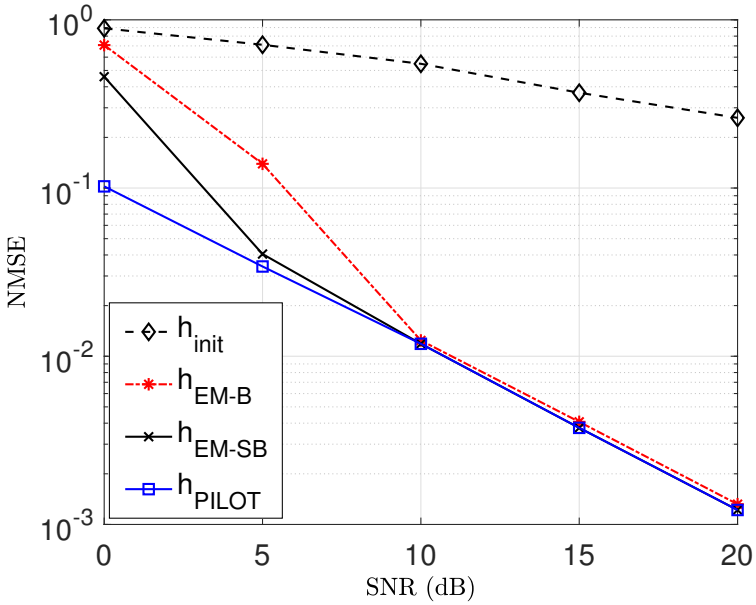


Figure 5.6: *NMSE vs. SNR for blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the fully-pilot-based estimator (h_{PILOT}). We consider a quadratic model, 4-QAM modulation and a MIMO system with $N_t = 2, N_r = 4$, and $M = 2$. h_{init} is included for reference and refers to a pilot-based estimator using available pilot symbols.*

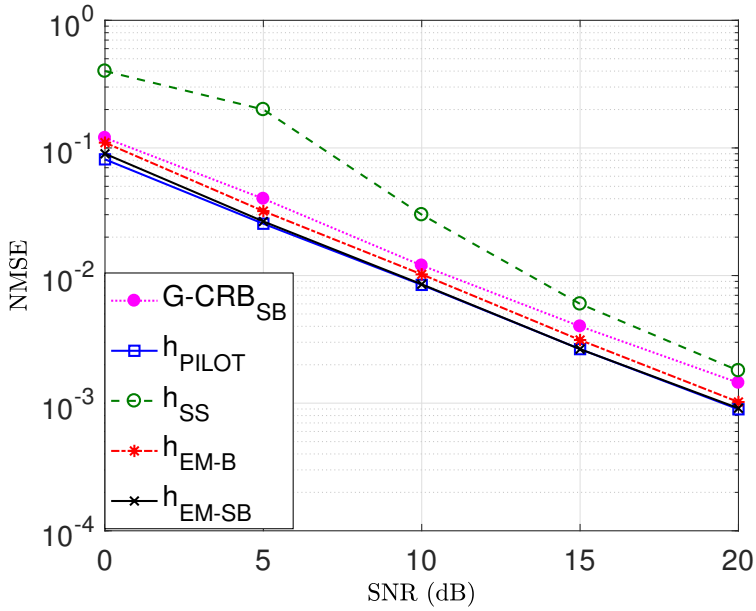


Figure 5.7: *NMSE vs. SNR for subspace-based estimator (h_{SS}), blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the semi-blind Gaussian CRB (G-CRB$_{SB}$) and the fully-pilot-based estimator (h_{PILOT}), in the case of a quadratic model, 4-QAM modulation and $N_r = 9$ receive antennas.*

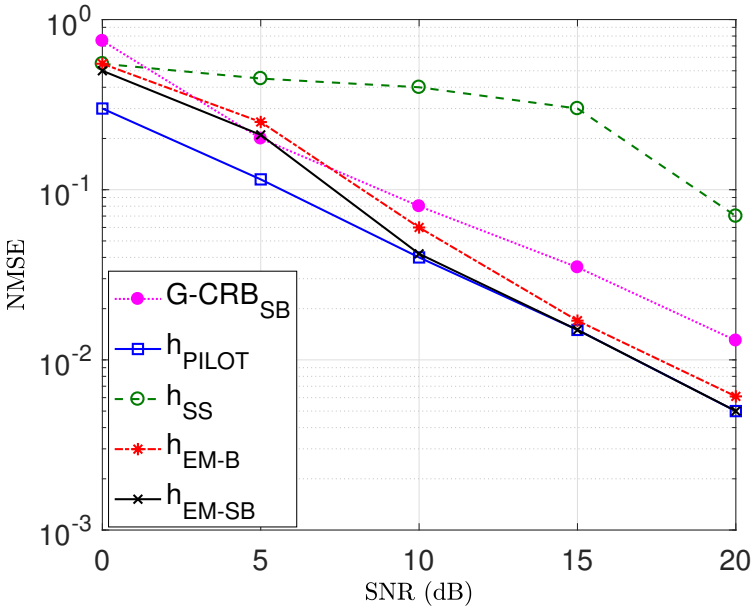


Figure 5.8: *NMSE vs. SNR for subspace-based estimator (h_{SS}), blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the semi-blind Gaussian CRB (G-CRB$_{SB}$) and the fully-pilot-based estimator (h_{PILOT}). We consider a cubic model and 16-QAM modulation.*

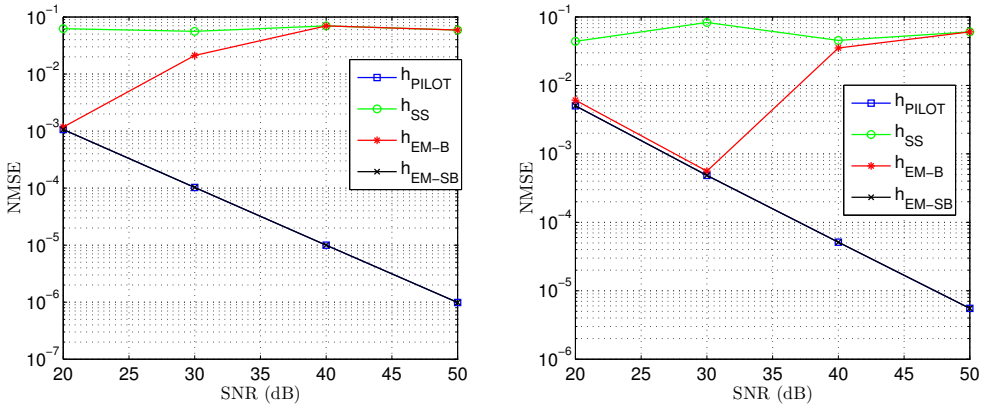


Figure 5.9: *NMSE vs. SNR with $M_{\mathrm{L}} = 4$ and $M_{\mathrm{NL}} = 2$ for subspace-based estimator (h_{SS}), blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the fully-pilot-based estimator (h_{PILOT}). We consider: (a) quadratic model and 4-QAM modulation, and (b) cubic model and 16-QAM modulation.*

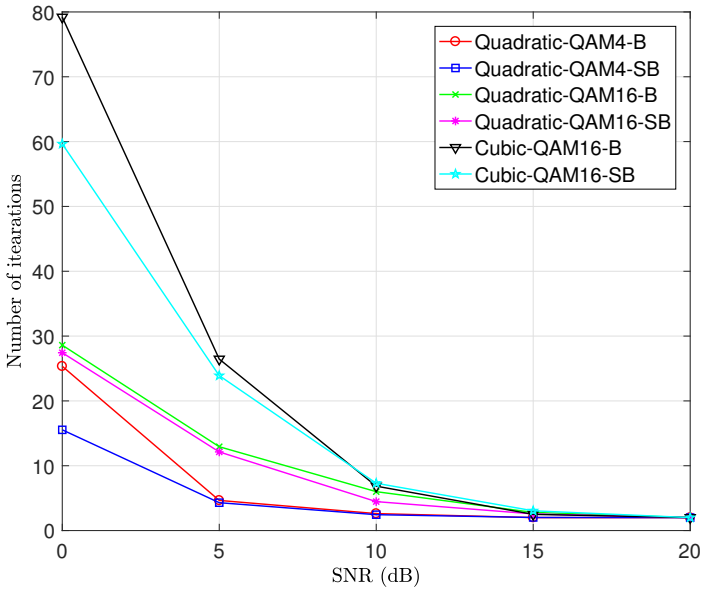


Figure 5.10: *Number of iterations for convergence vs. SNR for blind (B) and semi-blind (SB) EM-based estimators considering different nonlinearities and modulations.*

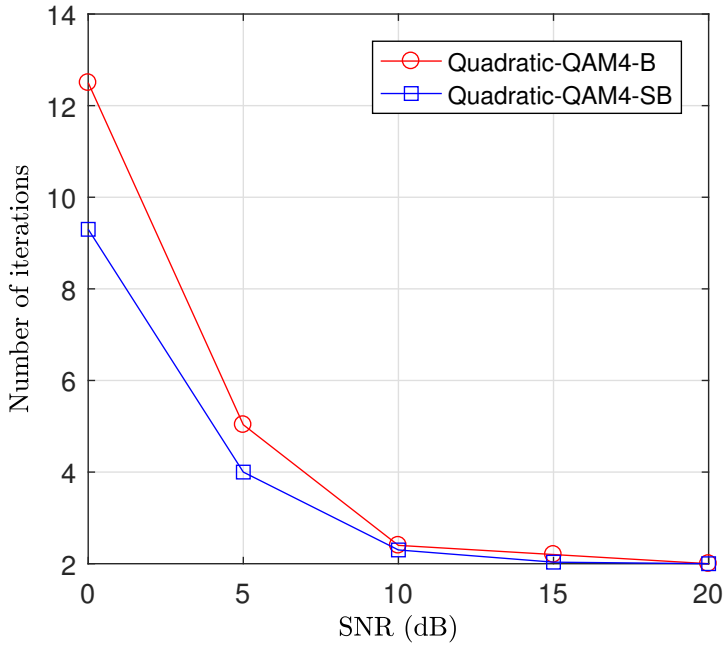


Figure 5.11: *Number of iterations for convergence vs. SNR considering a quadratic model, a 4-QAM modulation, and $N_r = 9$ receive antennas.*

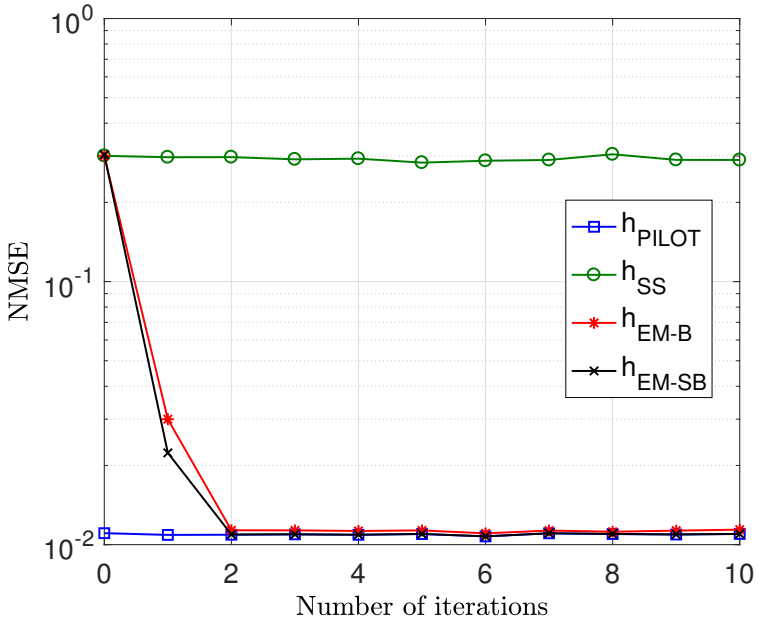
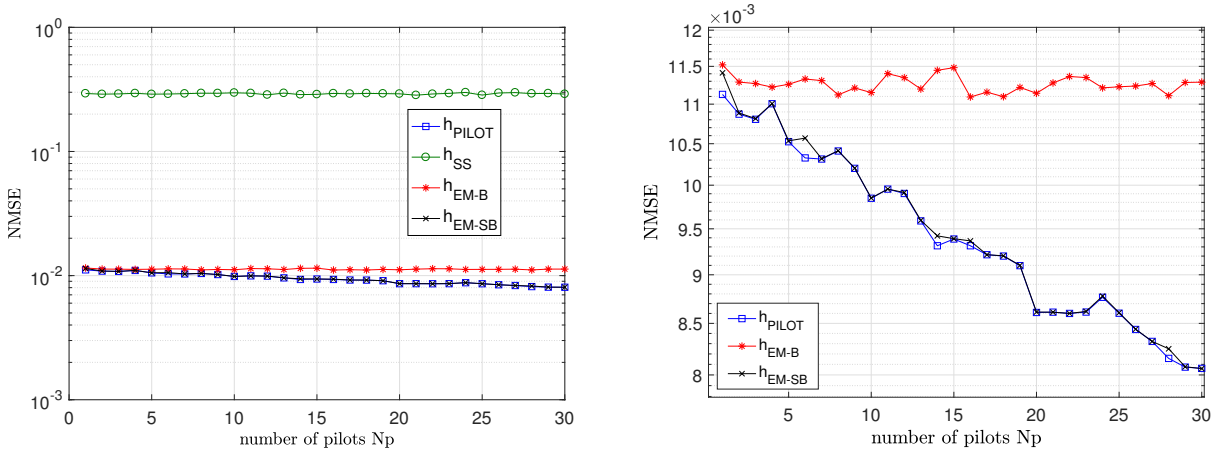


Figure 5.12: *NMSE vs. number of iterations for subspace-based estimator (h_{SS}), blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the fully-pilot-based estimator (h_{PILOT}), in the case of a quadratic model and 4-QAM modulation at SNR = 10 dB.*



(a) NMSE vs. number of pilots.

(b) zoom for NMSE around 10^{-2}.

Figure 5.13: *NMSE vs. number of pilots N_p for subspace-based estimator (h_{SS}), blind (h_{EM-B}) and semi-blind (h_{EM-SB}) EM-based estimators benchmarked against the fully-pilot-based estimator (h_{PILOT}), with a quadratic model and 4-QAM modulation at SNR = 10 dB.*

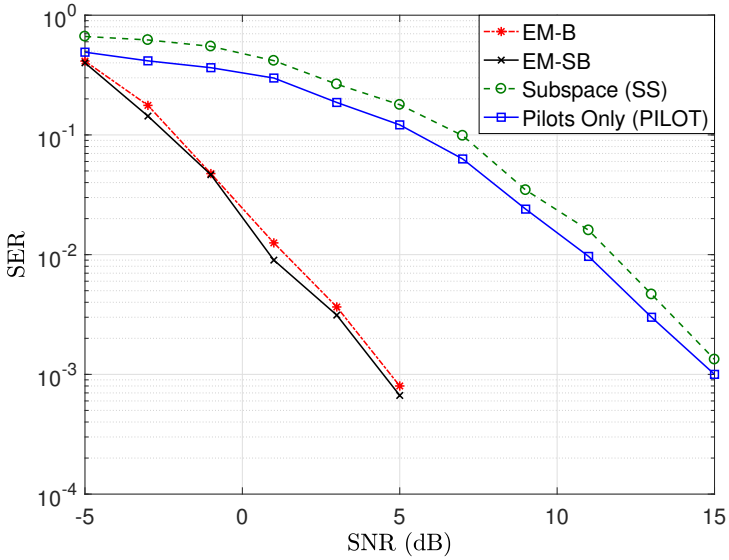


Figure 5.14: *SER vs. SNR for subspace-based estimator (SS), blind (EM-B) and semi-blind (EM-SB) EM-based estimators benchmarked against the fully-pilot-based estimator (PILOT), with a quadratic model and 4-QAM modulation.*

CHAPTER 6

Deep Learning based data detection for nonlinear communications systems

It always seems impossible
until it's done.

Nelson Mandela.

Abstract

The aim of this chapter is to propose Deep Learning (DL) based data detection solution for nonlinear Multiple-Input Multiple-Output (MIMO) communications systems. A Neural Network (NN) is built up, trained offline with finite alphabet data, and then used for online data detection. With no direct priors about the channel impulse response nor the transmitted data, the proposed DL-based data detector can deal with the performance degradation that might emerge from nonlinear components. The simulation results show the effectiveness of the proposed solution for different nonlinear model order and with attractive accuracy and data detection performance. Moreover, such a solution is promising to overcome the inherent ambiguity/limitations of classical blind processing.

Chapter content

6.1 Introduction

DL-based applications have showed recently important potential for the physical layer, when considering a wireless communications system [18]. Indeed, DL has become a new way of fundamentally rethinking the communications system design problem while promising performance improvements in complex communications scenarios, that are difficult to describe with tractable mathematical models, or that have solutions/algorithms with large complexity when implemented in practice. Moreover, open-source DL libraries [18] and readily available specialized hardware along with the astonishing progress of DL in computer vision have stimulated renewed interest in DL applications for communications and networking [128]. Currently, there are essentially two different main approaches of applying DL to the physical layer. Either to improve/augment parts of existing algorithms with DL or to completely replace them. Among the works falling into the first category are [129] and [130] that consider improved belief propagation channel decoding and MIMO detection, respectively. These works are inspired by the idea of deep unfolding [131] of existing iterative algorithms by essentially interpreting each iteration as a set of NN layers. In the second category, one can cite [132], that deals with blind detection for MIMO systems with low-resolution quantization, and [133], which investigates the detection for molecular communications; for which no mathematical channel model exists. The idea of learning to solve complex optimization tasks for wireless resource allocation, such as power control, is investigated in [134].

On the other hand, nonlinear behaviors can be encountered in many practical situations, in which case appropriate (nonlinear) processing is needed, when such nonlinearities are too important to be disregarded [35, 36]. Indeed, because most of real-life systems are inherently nonlinear in nature, nonlinear problems have drawn important interest and extensive attention from engineers, physicists, mathematicians and many other scientists [36]. In terms of machine learning point of view, a DL-based solution is proposed in [135] for Wiener-Hammerstein systems identification in the case of nonlinearities faced in electrical circuits. The solution is based on the Restricted Boltzmann Machine (RBM) along with the Randomized algorithms. Authors in [136] investigated the effectiveness of DL-based approaches for nonlinear system identification when considering soft-robot actuators nonlinearities. A Hammerstein model is used with a heating cooling signal as input for the DL network. In [137], a DL-based solution is proposed for nonlinear systems that describe coupled electrical drivers (two electric motors driving a pulley using a flexible belt). The solution is based on the Bayesian framework by combining deterministic and stochastic layers of recurrent network. However, it can be noticed that DL-based solutions that deal with nonlinear communications system models are still missing in the literature.

Hence, the aim of this chapter is to propose a DL-based solution for data detection, when considering nonlinear MIMO communications systems. By considering a finite alphabet transmitted signal, drawn from QAM modulation, the data detection is formulated as a DL classification problem. The proposed solution is performed blindly, i.e. based only on the observed data and on the trained network. The latter is trained according to a supervised learning fashion, with pairs of generated input-output data according to a nonlinear data model. So far, it is assumed to know the active users and the modulation constellation.

6.2 Communications system model and problem formulation

This section describes the communications system model adopted in this chapter. A MIMO system composed of N_t transmitters and a receiver equipped with N_r antennas is considered. At the receiver, nonlinear components, associated to the transmitted symbols, are assumed to be present. Thus, the instantaneous received signal, according to Volterra nonlinear models [98], is given at the r-th receive antenna by:

$$y_r(k) = \mathbf{h}_{L_r}^T \mathbf{x}(k) + \sum_{i=2}^{P} \mathbf{h}_{i_r}^T \mathbf{x}_i(k) + n(k), \tag{6.1}$$

where k refers to time index, $\mathbf{h}_{L_r}, \mathbf{h}_{i_r} \in \mathbb{C}^{N_t \times 1}$ contain the channel coefficients for linear and nonlinear terms; with P being the nonlinear model order. $\mathbf{x}(k) = [x_1(k), x_2(k), ..., x_{N_t}(k)]^T$ contains the instantaneous users transmitted symbols; whereas $\mathbf{x}_i(k)$ refers to the data nonlinear terms so that $\mathbf{x}_i(k) = [x_1^i, ..., x_{N_t}^i]^T$. $n(k)$ is a white circular Gaussian noise, uncorrelated from sensor to sensor, with zero mean and variance σ_v^2. Hence, by considering all the antennas of the receiver, equation (6.1) can be expressed as follows:

$$\mathbf{y}(k) = \mathbf{H}_L \mathbf{x}(k) + \sum_{i=2}^{P} \mathbf{H}_i \mathbf{x}_i(k) + \mathbf{n}(k), \tag{6.2}$$

where $\mathbf{y}(k) = [y_1(k), y_2(k), ..., y_{N_r}(k)]^T$ is the observed signal, $\mathbf{H}_L = [\mathbf{h}_{L_1}, ..., \mathbf{h}_{L_{N_r}}]^T; \mathbf{H}_i = [\mathbf{h}_{i_1}, ..., \mathbf{h}_{i_{N_r}}]^T \in \mathbb{C}^{N_r \times N_t}$ contain respectively the channel coefficients for the linear and nonlinear terms and $\mathbf{n}(k) = [n_1(k), ..., n_{N_r}(k)]^T$.

The objective of blind data detection is to detect the transmitted symbols $\mathbf{x}(k)$, relying only on the observed signals $\mathbf{y}(k)$ and with no 'strong' priors on the channel impulse response nor the transmitted data.

Many techniques have been proposed in the literature to deal with nonlinear models. However, in such approaches, more complex solutions with higher size data are exploited as will be briefly discussed in section 6.3.

Consequently, in the current work, we propose a solution based on the use of deep learning approach for the problem of data detection, when considering nonlinear data models. Indeed, deep learning based approaches are, in general, applied to solve regression or classification problems. In our case, data detection is treated as a classification problem, where according to the received symbol vector, a well trained neural network is able to assign to it the corresponding transmitted data vector. Actually, such a procedure is somehow similar to a maximum likelihood one, where the network chooses one (the most likely) of the possible realizations of the transmitted signal vector for a given observation.

6.3 Classical approaches for nonlinear systems data detection

Many approaches have been proposed in the literature for the problem of system identification when dealing with nonlinear systems (see e.g. [1, 105, 106, 107]). Few of them have focused on data detection, since it can be performed after channel estimation. Nevertheless, one can notice that, these techniques rely

on new vector representation of data and channel. The system model given in equation (6.2) is re-written as follows:

$$\mathbf{y}(k) = \mathbf{H}\bar{\mathbf{x}}(k) + \mathbf{n}(k), \tag{6.3}$$

where $\mathbf{H} = [\mathbf{H}_L \mathbf{H}_2 ... \mathbf{H}_P]$ and $\bar{\mathbf{x}}(k) = [\mathbf{x}(k)\mathbf{x}_2(k)...\mathbf{x}_P(k)]^T$.
These extended variables are then exploited either with analytical techniques or by iterative procedures which might involve complex operations like matrix inversion, eigenvalue decomposition PARAFAC representation and so on. Furthermore, these techniques are applied online with (in general) some assumptions on data or communications channel. Besides, efficient blind-based data detection techniques require important amount of data to be processed online, while estimating the transmitted signals up to scaling factors and permutation matrix (the blind inherent ambiguity).

For the purpose of estimating the transmitted data while overcoming these constraints, a new deep learning based approach is investigated in the sequel, which aims at competing with the classical approaches, in terms of data detection performance, while providing low online computational complexity.

6.4 Proposed deep learning based data detection

By considering the problem of data detection as a DL-based classification task, the proposed solution is based on a learning phase, performed offline, that results in a Deep Neural Network (DNN), which is in turn, used for online data detection (inference phase) as illustrated in Figure 6.1.

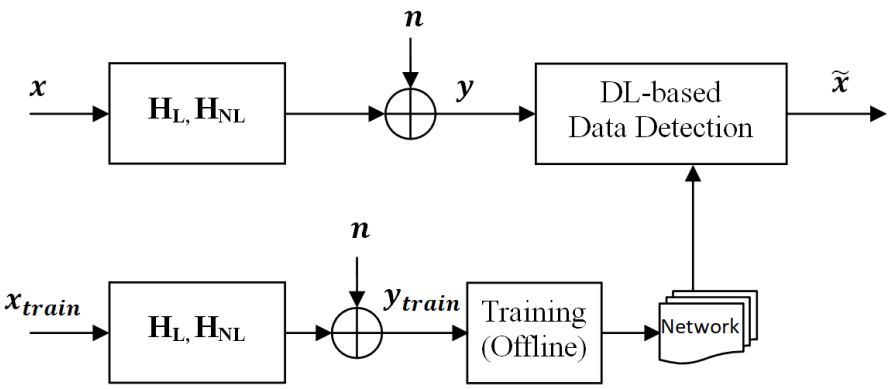


Figure 6.1: *DL based data detection.*

6.4.1 DL network architecture

Generally speaking, Deep Neural Networks (DNNs) are deeper versions of artificial neural networks (ANNs) by increasing the number of hidden layers and using appropriate mathematical tools. Each layer of the network consists of multiple neurons (nodes), each of which has an output that is a function of neurons of its preceding layer and a set of parameters.

For the proposed solution, the adopted DL architecture is given in Figure 6.2. The input layer receives the signal observed at the output of the MIMO system given by $\mathbf{y}(k)$ (as illustrated in Figure 6.1). Since

the available libraries treat only real-valued variables, the size of the input layer is $2 \times N_r$, which results from stacking real and imaginary parts of the observed symbols.

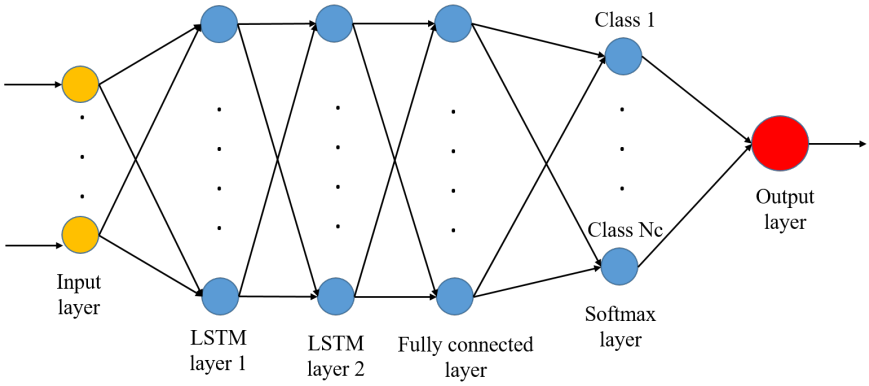


Figure 6.2: *DL network architecture adopted.*

The output layer returns the corresponding class (one node) out of N_c classes, which is then exploited to find the corresponding transmitted symbols $\mathbf{x}(k)$. This is performed through a predefined mapping between the N_c classes and the corresponding N_c possible realizations of the transmitted symbols $\mathbf{x}(k)$. For the hidden layers, the network describes a mapping $f(\mathbf{x}_0;\boldsymbol{\theta}) : \mathbb{R}^{N_0} \rightarrow \mathbb{R}^{N_c}$ of an input vector $\mathbf{x}_0 \in \mathbb{R}^{N_0}$ to an output vector $\mathbf{x}_L \in \mathbb{R}^{N_c}$ through L iterative processing steps:

$$\mathbf{x}_l = f_l(\mathbf{x}_{l-1};\boldsymbol{\theta}_l) \quad l = 1...L, \tag{6.4}$$

where L is the hidden layers ($L = 4$ in our case as given in Figure 6.2) and $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_L]$ being the parameters for all hidden layers; $N_0 = 2 \times N_r$; $\mathbf{x}_l$ refers to the input data at the l-th hidden layer and $f_l(\mathbf{x}_{l-1};\boldsymbol{\theta}_l) \in \mathbb{R}^{N_{l-1}} \rightarrow \mathbb{R}^{N_l}$ stands for the l-th activation function, which depends on the chosen hidden layer. Such a function introduces a non-linearity which is important for the so called expressive power of the NN. Otherwise there would be not much of an advantage of stacking multiple layers on top of each other.

A Long Short Time Memory (LSTM) structure was adopted, which is followed by a fully connected layer then a softmax layer. The motivation behind using LSTM cells is to allow capturing dependencies between time steps in time series and sequence data. The fully connected layer multiplies the input data by a weight matrix and then adds a bias vector; while allowing a size adaptation between LSTM and softmax layers. The later implements the softmax function which assigns a probability to each output class, by maintaining the individual outputs to be within the interval $[0, 1]$, with a sum equals to one. The parameters of hidden layers (i.e. $\boldsymbol{\theta}$) will be estimated during the learning phase.

It is worth noting that the proposed DL architecture, for the considered realistic scenario, is not obvious. Actually, many investigations have focused on the choice and the design of the DL architecture (e.g. number and nature of hidden layers) as well as the training strategy which is tailored by taking into account the nature of transmitted signals. In the sequel, the performance of the proposed DL architecture is confirmed by experimental results.

6.4.2 Learning strategy

The learning phase consists at giving the network enough data $(\mathbf{y}(k), \mathbf{x}(k))$ to be able to assign to each unknown received signal $\mathbf{y}(k)$ the corresponding transmitted symbol vector $\mathbf{x}(k)$ (performing a classification). To do so, the network is provided with all possible cases for the transmitted signal vector $\mathbf{x}(k)$. We consider that every transmitted symbol; for each user $x_i(k)$, $i = 1, ..., N_t$, is taken from a finite alphabet set $\mathcal{A} = \{a_1, ..., a_M\}$ with $M = 2^B$ is the cardinal of $\mathcal{A}$ and B is the number of bits used to code each alphabet. Hence, the total number of possible realizations of $\mathbf{x}(k)$ is $N_c = M^{N_t}$. From deep learning point of view, N_c is the number of classes to which will be assigned each symbol $\mathbf{y}(k)$. Consequently, by taking into account the effect of the communications channel and the noise, the received signal and the original transmitted data are collected as the training data set.

Basically, the learning process aims to minimize, with respect to $\boldsymbol{\theta}$, a loss function. Different loss functions can be used, mainly the mean squared error (MSE) and the Categorical cross-entropy or its variants [138]. In the proposed solution, the network is trained to minimize the difference between the output of the neural network and the known transmitted data class. The difference can be portrayed in several ways. In our experiment settings, we choose the L_2 loss function defined as follows:

$$L_2 = \frac{1}{N_{tot}} \sum_{i=1}^{N_{tot}} (\|\mathbf{x}_i - \tilde{\mathbf{x}}_i\|_2^2), \tag{6.5}$$

where N_{tot} is the size of the training data set; $\mathbf{x}_i$ refers to the transmitted symbol class, whereas $\tilde{\mathbf{x}}_i$ stands for the inferred class.

The minimization of the loss function; and hence the estimation of vector $\boldsymbol{\theta}$; is performed in general iteratively through the Stochastic Gradient Descent (SGD) algorithm. The gradient in SGD can be computed efficiently through the backpropagation algorithm [139]. Also, instead of updating the parameters at each training sample, small batches of the data set can be used within a mini-batch Gradient descent framework. In the proposed solution, the optimization of the loss function is performed through the "Adam" optimization algorithm [140], which is considered as a combination of RMSprop and SGD with momentum. It uses the squared gradients to scale the learning rate as in RMSprop and takes advantage of momentum by using moving average of the gradient instead of gradient itself like SGD with momentum.

When the entire training dataset is passed forward and backward through the neural network only once, this is called one epoch. Many epochs are performed to obtained a well trained network.

6.5 Complexity discussion

One can notice that the main computational complexity of the proposed DL-based solution comes from the learning step, which needs important amount of data and the use of iterative procedures for parameter estimation. However, this constraints concern only the learning phase which is performed offline. For the online step, basic operations are performed; additions and multiplications; through hidden layers to obtain the final decision (output) of the network for given observations. Thus, no need for complex calculation like matrix inversion that consumes time and memory space. Moreover, for an optimal implementation,

one could easily take advantages from the available open-source DL libraries [18] and readily available specialized hardware.

On the other hand, to make the proposed solution more efficient and to alleviate the learning parameters dependency, one could use many prelearned models with different parameters; SNR, number of users, nonlinear model orders or data modulation. Also, it could be advantageous to associate other classification approaches, such as modulation classification as in [141, 142] to the proposed one. Nevertheless, one could notice that with high order modulations and high number of users, the number of classes become very important which may require more complex network. Actually, this issue is left for future work.

6.6 Experimental results and discussion

This section highlights the performance of the proposed deep learning based data detection. Unlike other DL applications (e.g. image classification) no specific data set are available for performing experiments. Thus, synthetic data will be generated and used. The performance evaluation is based on the accuracy, which is widely used for DL-based performance assessment and on the Symbol Error Rate (SER) which is mostly used for data detection performance evaluation.

6.6.1 Experimental settings

A 4×4 ($N_t=4, N_r=4$) MIMO system was considered for simulations. The transmitted symbols are drawn from 4-QAM modulation, thus, the number of the possible realizations for the vector $\mathbf{x}$ is $N_{tot}=4^4=256$. Linear and nonlinear channel coefficients are generated as i.i.d., unit-power, complex Gaussian random variables, so that $\mathbf{H}_L=\mathbf{H}_0+\delta\mathbf{H}$, where $\mathbf{H}_0$ is supposed to be known up to $\delta\mathbf{H}$ which takes different values at each realization according to a Gaussian random process, with zero mean and a standard deviation of 10^{-1} in our experiments. For each possible realization of $\mathbf{x}$ and each SNR value, 1000 realizations are generated, where at each of which, $\delta\mathbf{H}$, $\mathbf{H}_i$ and $\mathbf{n}(k)$ take new values. Consequently, 256000 pairs $(\mathbf{y},\mathbf{x})$ are obtained, where 80% of these pairs are used for training, 10% for validation, whereas 10% are left for testing. It is worth pointing out that the simulation results given latter are based on the testing data, i.e. using the network obtained after training. By contrast to the validation data, which are used during training process for parameters tuning. Also, since most of the available deep learning libraries are based on real-valued variables, each complex variable is represented by concatenating the real and imaginary parts, i.e. $\mathbf{y}_{real}=[real(\mathbf{y}),imag(\mathbf{y})]^T$. Simulation parameters are summarized in Table I.

6.6.2 Simulation results and discussion

Figure 6.3 illustrates the accuracy of the trained network v.s. the SNR; for the case with only linear terms (DL-L), a 3-rd order (DL-NL-3^{rd}) and a 5-th order (DL-NL-5^{th}) nonlinear models, which are benchmarked with the case of perfect knowledge of linear and nonlinear channel coefficients (DL-Perfect-H). One could notice that the accuracy of the trained network is not very affected by the presence of nonlinear terms,

Parameters	Specifications
Number of LSTM hidden layers	2
Number of fully connected hidden layers	1
Number of softmax hidden layers	1
Nodes per hidden layer for LSTM layers	100
Nodes per hidden layer for fully connected and softmax layers	N_{tot}
Input layer size	$2N_r$
Output layer size	1
Number of epochs	2
Mini-batch size	40
Training optimization algorithm	Adam
Learning rate	0.001

Table 6.1: *Simulation parameters.*

and is almost independent of the nonlinear model order. Moreover, on could notice that the obtained performance are acceptable compared to the most favorable case given by (DL-Perfect-H).

Figure 6.4 illustrates the performance of the deep learning based data detection in terms of the symbol error rate (SER). Similar to results of Figure 6.3, very interesting performance is obtained when considering nonlinear terms. Also, it can be noticed that the nonlinear model order does not affect much the obtained performance. Moreover, the proposed solution outperforms the Zero Forcing (ZF) based (ZF-NL-3^{rd}) one even with true channel coefficients, i.e. $\hat{\tilde{\mathbf{x}}}(\mathbf{k}) = \mathbf{H}^{\#}\mathbf{y}(\mathbf{k})$, with $(.)^{\#}$ being the pseudo-inverse operator.

Figure 6.5 assesses the performance in terms of SER when considering different signal levels between linear and nonlinear terms ($P_L - P_{NL}$ in dB), since in practice nonlinear terms have usually less power than linear terms. Also we considered a variable linear channel estimate accuracy, given by $NMSE(H_L) = \frac{\|\mathbf{H}_L - \mathbf{H}_0\|^2}{\|\mathbf{H}_0\|^2}$, which is used during the offline learning phase. Here again, it can be noticed that the proposed solution still presents promising potential under different realistic scenarios.

Figure 6.6 assesses the performance in terms of SER w.r.t. SNR for different values of the receive antennas number. One can notice that the space diversity offers important performance gain for the proposed DL based solution. This result allows to say that such a solution is promising for large scale (or massive) MIMO systems.

Figure 6.7 assesses the performance in terms of SER w.r.t. SNR when considering 4-QAM and 16-QAM modulations. One could notice that the used network architecture provides interesting performance with relatively higher order modulations, 16-QAM in our case.

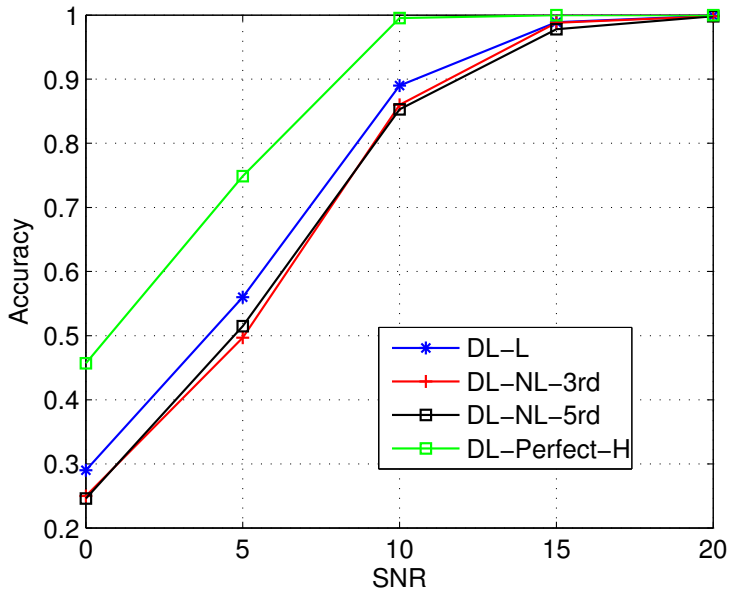


Figure 6.3: *Accuracy v.s. SNR.*

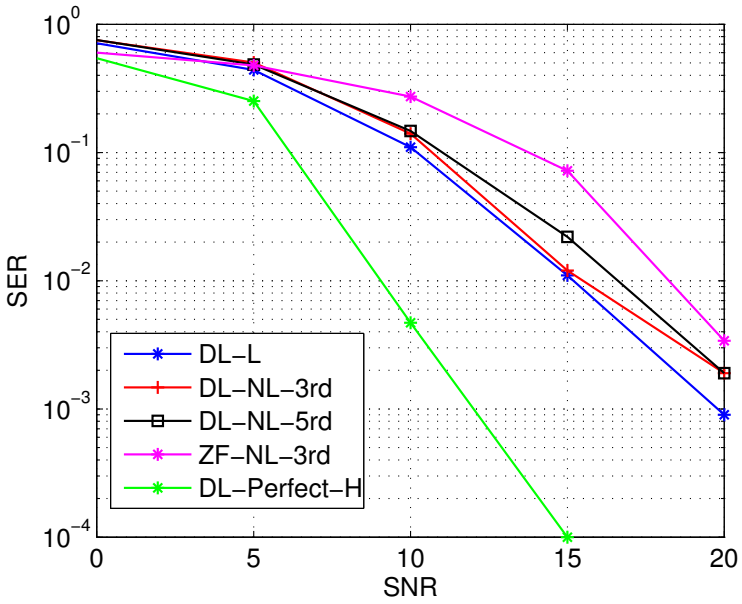


Figure 6.4: *SER v.s. SNR.*

6.7 Conclusion

This chapter proposes a deep learning based data detection solution for nonlinear MIMO systems. A neural network is built up, trained offline, then used for online data detection. With no strong priors about the channel impulse response nor the transmitted data, the proposed DL-based data detector can overcome the performance degradation that emerge from nonlinear components. Simulation results show that the proposed solution can be easily used for high order nonlinear models and offers attractive

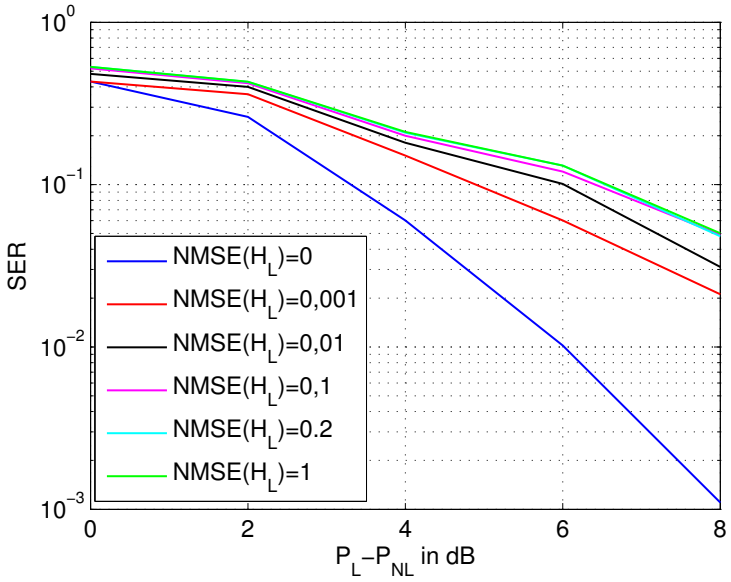


Figure 6.5: *SER vs. power level between linear and nonlinear terms $P_L - P_{NL}$, with variable linear channel estimate accuracy $NMSE(H_L)$ at $SNR = 10dB$.*

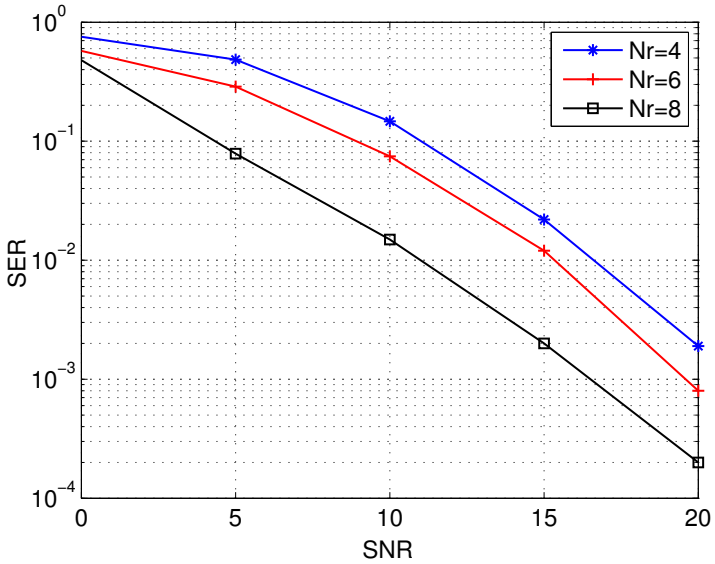


Figure 6.6: *SER v.s. SNR for different number of receive antennas N_r.*

accuracy and data detection performance. Also, such a solution is promising to overcome the inherent ambiguity of classical blind processing.

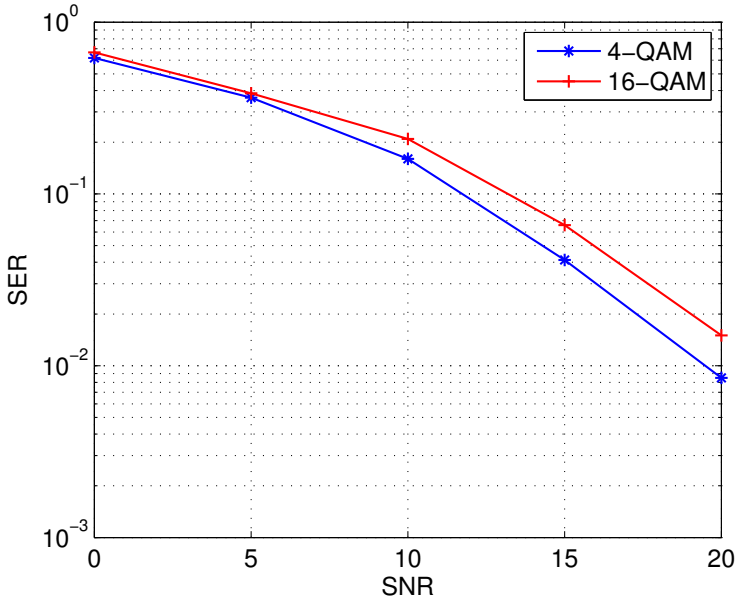


Figure 6.7: *SER v.s. SNR for 4-QAM and 16-QAM constellations, with* $N_t = 2, N_r = 2$ *and* 5*-th order nonlinear model.*

Chapter 7

Conclusion and future work

> *“We ought not to be embarrassed of appreciating the truth and of obtaining it wherever it comes from, even if it comes from races distant and nations different from us. Nothing should be dearer to the seeker of truth than the truth itself, and there is no deterioration of the truth, nor belittling either of one who speaks it or conveys it.”*
>
> Al-Kindi.

Chapter content

7.1 Achieved work

System identification refers generally to the problem of developing empirical mathematical models of systems based on excitation and response measurements, then designing appropriate approaches/algorithms for the model's parameters estimation. It is an important issue in many areas including process control and communications. For wireless communications, channel estimation is of paramount importance for equalization and data detection problems in most wireless communications, especially (massive) MIMO-OFDM systems. Nevertheless, in the case of non availability of CSI, one could adopt source separation approaches as an alternative to detect the transmitted data without direct priors on CSI. Either for channel estimation or source separation, semi-blind techniques are adopted in this thesis, in order to take advantages of blind and pilot-based approaches, while reducing their drawbacks. Besides, the astonishing advances of machine learning approaches, especially deep learning, motivates us to investigate, in the current thesis, the potential of DL-based techniques for the problem of data detection.

Several contributions to semi-blind channel estimation and source separation have been proposed in this thesis, mainly through performance bounds analyses and the development of efficient algorithms by exploiting priors on the system for performance improvement. In the sequel, we briefly summarize the overall thesis work, before listing the main contributions.

For semi-blind communications channel estimation, a performance bound analysis, based on the CRB, has been conducted by considering massive MIMO-OFDM communications systems. The aim of such an analysis is to assess the effectiveness of semi-blind approaches for pilot contamination mitigation, either when the cell under test along with the adjacent cells are synchronous or not. For synchronous cells case, the analysis demonstrates the potential of semi-blind approaches to efficiently overcome the pilot contamination problem, when considering a finite alphabet (non Gaussian) communications signal. However, relying only on the signal's Second Order Statistics (SOS) is not enough for solving such an issue. Nevertheless, it is possible to get close to the optimal performance with a semi-blind approach even if the pilots are non-orthogonal as long as they are not fully-coherent. For the asynchronous cells case, it has been demonstrated that the pilot contamination still occurs under small inter-cell delays, but can be strongly mitigated with large inter-cell delays.

For channel equalization and data detection via source separation approaches, a semi-blind MM-based source separation solution has been proposed for instantaneous mixtures in a massive MIMO communications system, then extended to a MIMO-OFDM communications system, i.e. convolutive case. In the latter case, a separating matrix has been defined independently on the sub-carriers, leading to computational complexity reduction. For both cases, a hybrid cost function is defined based on the MM criterion, for the blind part, and on the Least Squares (LS) criterion for the pilot-based part. The optimization is performed through a simple but efficient semi-blind block gradient descent procedure, with an optimized step size computed at each iteration and for each user. Besides interesting source separation performance and accelerated convergence, the proposed approach is very promising for source recovering when pilot contamination occurs in massive MIMO-OFDM communications systems.

Besides, when taking into account the problem of system identification with data detection, a semi-blind solution for joint channel estimation and data detection, by considering the sparce nature of the communications channel, has been proposed. An optimization problem is formulated then solved by using the successive convex approximation (SCA) approach. Accordingly, the optimization is performed on an approximate convex problem, rather than the original nonconvex one. An iterative procedure has been performed where the channel coefficients and data symbols are estimated simultaneously at each iteration. Also an optimized step size, introduced according to line search procedure, is used for convergence improvement with guaranteed convergence to a stationary point. Simulation results showed that the proposed solution exhibits fast convergence with very attractive channel and data estimation performance.

In order to deal with system identification for multichannel nonlinear communications system, blind and semi-blind solutions have been proposed based on the ML approach. For the blind solution, an original combination of a subspace approach and the EM algorithm is proposed. Basically, an initial channel estimation is performed based on a subspace approach, which is followed by an appropriate ambiguity removal method. Then, to refine this estimate, a maximum likelihood approach is introduced based on the EM algorithm. For the semi-blind case, both pilots and data are used to define the different steps of the EM algorithms. This solutions are supported by some identifiability results and performance bounds analysis, related to the considered models (blind and semi-blind). Basically, the simulation results highlighted the very interesting channel estimation performance and the attractive convergence speed for the EM-based iterative solution for the considered nonlinear communications system.

In this thesis, besides the mathematics-based techniques, a DL-based data detection solution is proposed for nonlinear MIMO communications systems. Basically, data detection is treated as a DL based classification problem, where a Neural Network (NN) has been built up and trained offline with appropriately chosen finite alphabet data. Then, it has been used for online data detection. With no direct priors about the channel impulse response nor the transmitted data, the proposed DL-based solution is shown to deal with the performance degradation that might emerge from nonlinear components. Also, the simulation results showed the effectiveness of the proposed solution for different nonlinear model order and with attractive accuracy and data detection performance. Moreover, such a solution is promising to overcome the inherent ambiguity/limitations of classical blind processing.

7.2 Thesis contributions

The main contributions of this research are listed below:

- Derivation of semi-blind channel estimation CRBs for multi-cell massive MIMO-OFDM system under pilot contamination phenomenon and highlighting the resulting non-identifiability.
- Thorough study of the effectiveness of semi-blind techniques for pilot contamination mitigation in massive MIMO-OFDM systems, by considering SOS, HOS, different pilot structures, synchronous

cells and non-synchronous cells with small and large inter-cell delays.

- Highlighting the effect of pilot's orthogonality level to mitigate pilot contamination.
- For finite alphabet signals, i.e. QPSK data model, a realistic CRB approximation has been given to bypass the high complexity of the exact CRB computation.
- Proposition of a multi-moduls based semi-blind technique for source recovery when considering instantaneous MIMO systems under pilot contamination effect. Then, extending this solution to the convolutive systems, i.e. MIMO-OFDM, where a separation matrix is defined independently on the sub-carrier then estimated.
- Highlighting the importance of optimal step size, for iterative algorithms, to improve convergence speed.
- Putting forward the importance of semi-blind approaches for throughput gain and channel/data estimation performance improvement.
- Proposition of a joint sparse channel estimation and data detection solution based on the successive convex approximation (SCA). This solution deals with an approximate convex problem rather than with the original non convex one.
- Exploiting priors on the system for performance improvement, such as pilots, channel's sparsity and data matrix structure.
- Contribution to blind and semi-blind channel estimation for nonlinear system models by combining a subspace method with the EM-algorithm.
- Proposition of two techniques for blind channel estimation ambiguity removal. The first one concerns the quadratic nonlinearity, whereas the second one is designed for cubic nonlinear models.
- CRB derivation and identifiability study of system identification for nonlinear systems.
- Assessing the potential of a machine learning for communications systems, by a deep learning based solution for data detection; when considering instantaneous nonlinear MIMO systems.

7.3 Future work

The research work related to semi-blind channel estimation and source separation for linear and nonlinear multi-channel communications systems, carried out in this thesis can be extended in several directions. Some recommendations for future work are listed below.

- Extend the semi-blind channel estimation performance analysis, carried out for massive MIMO-OFDM systems, to the stochastic channel model (instead of deterministic one) where Bayesian approach can be considered to evaluate the CRB for a given channel type.

- Investigate the possibility to develop an analytical multi-modulus based solution for source separation.
- Investigate on a solution for computational complexity reduction for the proposed EM-based solutions.
- Propose a DL-based solution for system identification and/or for joint channel estimation and data detection.
- Implement the proposed algorithms in this thesis in a real-world system, such as a video/image transmission system, in order to further evaluate the performance gain in practice.

APPENDIX

Appendix: Performance Bound Analysis of Side Information Effect on MIMO-OFDM Channels Semi-Blind Identification

Success is not final, failure is not fatal: it is the courage to continue that counts.

Winston Churchill.

Abstract

This paper studies the impact of different priors on semi-blind channel estimation performance for Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) communications systems. To do so, for an estimator-independent study, the Cramèr-Rao Bound (CRB) is considered to quantify and analyze the performance limits. The analysis is carried out by considering three cases: (i) no modeling while estimating the channel coefficients in the frequency domain (for each frequency bin); (ii) a finite memory linear time invariant channel model while estimating the channel taps in the time domain; and finally (iii) a specular channel model while estimating the propagation channel parameters, i.e. the fading, the delay and the angle of arrival of each path. In particular, this analysis quantifies the performance gain obtained by properly exploiting side information about the propagation channel.

Chapter content

A.1 Introduction

Channel State Information (CSI) remains a current concern in Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing (MIMO-OFDM) wireless communications systems, since the system's overall performance depends strongly on it [143]. Several channel estimation approaches have been proposed. Some techniques fall into the scope of blind channel estimation which have been extensively studied. These techniques are fully based on the statistical properties of the transmitted data (e.g. [21, 22, 144]). While other techniques rely on some inserted known symbols, called pilots, in the transmitted frames as specified in most communications standards [19, 20]. However all of these approaches have their own merits and weaknesses. In general, pilot-based channel estimators provide an easier and more accurate channel estimation, but leads to decreasing the spectral efficiency and the throughput as compared to the blind methods. Consequently, it would be advantageous to benefit from the two strategies where both data and pilots are exploited through a semi-blind estimation approach [59, 24, 25].

Besides known data or statistical information, some priors (or side information) on the communications system can be available and can therefore affect the performance of the channel estimation.

Consequently, this paper aims to study the impact of different priors, relative to the channel, on a MIMO-OFDM system identification performance when adopting semi-blind approaches. To do so, Cramèr-Rao Bound (CRB) is used to quantify the performance limits independently of the estimator. Furthermore, a comparative study is conducted by considering three cases. The first case concerns the estimation of the channel fading coefficients in the frequency domain (for each frequency bin) disregarding its (finite memory) time domain structure. In the second case, a finite memory linear time-invariant channel model is considered and the channel taps are estimated in the time domain. In the third case, a specular channel model is considered and the propagation channel parameters, i.e. the fading, the delay and the angle of arrival of each path are estimated.

It is worth pointing out that authors in [59] carried out a through study on the CRB derivation when only considering the channel taps estimation in time domain. In the current work; and as an extension of this work, a comparative study of the performance gain is conducted by considering also the CRB derivation for estimating the channel coefficients in the frequency domain as well as the CRB for estimating the propagation channel parameters (i.e. the fading, the delay and the angle of arrival of each path) corresponding to a specular channel model.

A.2 MIMO-OFDM Communications System Model

The MIMO-OFDM communications system considered in this paper is represented by N_t mono-antenna transmitters and a receiver equipped with N_r receive antennas. The transmitted symbols are assumed to be OFDM ones. Each OFDM symbol is composed of K samples extended by the insertion of a Cyclic Prefix (CP) corresponding to the last L samples at the beginning of the OFDM symbol, so that the CP length is assumed to be greater than or equal to the maximum channel delay denoted N (i.e. $N \leq L$). Once removing the CP and taking the K-point FFT of the received OFDM symbols, the received signal

(of size K) at the r-th antenna is given, in time domain, by [60]:

$$\mathbf{y}_r = \sum_{i=1}^{N_t} \mathbf{F}\, \mathbf{T}(\mathbf{h}_{i,r}) \frac{\mathbf{F}^H}{K} \mathbf{x}_i + \mathbf{v}_r, \tag{A.1}$$

where $\mathbf{F}$ stands for the K-point Fourier matrix; $\mathbf{h}_{i,r} = [h_{i,r}(0), \cdots, h_{i,r}(N-1)]^T$ is the $N \times 1$ vector containing the channel taps between the i-th transmitter and the r-th receive antenna; $\mathbf{T}(\mathbf{h}_{i,r})$ is a circulant matrix and $\mathbf{x}_i$ is the i-th OFDM symbol of length K. $\mathbf{v}_r$ is assumed to be an additive white Circular Gaussian noise with $E\left[\mathbf{v}_r(k)\mathbf{v}_r(i)^H\right] = \sigma^2_{\mathbf{v}} \mathbf{I}_K \delta_{ki}$; with $(.)^H$ being the Hermitian operator; $\sigma^2_{\mathbf{v}}$ the noise variance; $\mathbf{I}_K$ the identity matrix of size $K \times K$ and δ_{ki} the Kronecker delta operator.

By exploiting the eigenvalue decomposition of the circulant matrix $\mathbf{T}(\mathbf{h}_{i,r})$ given by: $\mathbf{T}(\mathbf{h}_{i,r}) = \frac{\mathbf{F}^H}{K} diag\left\{\mathbf{W}\mathbf{h}_{i,r}\right\} \mathbf{F}$, where $\mathbf{W}$ is a matrix containing the N first columns of $\mathbf{F}$ and $diag$ is the diagonal matrix composed by its vector argument, equation (A.1) becomes:

$$\mathbf{y}_r = \sum_{i=1}^{N_t} diag\left\{\mathbf{W}\mathbf{h}_{i,r}\right\} \mathbf{x}_i + \mathbf{v}_r. \tag{A.2}$$

By taking into account the N_r receive antennas, the observed signal can be written as:

$$\mathbf{y} = \boldsymbol{\lambda}\mathbf{x} + \mathbf{v}, \tag{A.3}$$

where $\mathbf{y} = \left[\mathbf{y}_1^T \cdots \mathbf{y}_{N_r}^T\right]^T$; $\mathbf{x} = \left[\mathbf{x}_1^T \cdots \mathbf{x}_{N_t}^T\right]^T$; $\mathbf{v} = \left[\mathbf{v}_1^T \cdots \mathbf{v}_{N_r}^T\right]^T$; and $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_{N_t}]$ and $\boldsymbol{\lambda}_i = \left[\boldsymbol{\lambda}_{i,1} \cdots \boldsymbol{\lambda}_{i,N_r}\right]^T$ where $\boldsymbol{\lambda}_{i,r} = diag\left\{\mathbf{W}\mathbf{h}_{i,r}\right\}$.

Note that, equation (A.3) is expressed w.r.t. the channel coefficients in the frequency domain. In order to facilitate the analytical CRB derivations w.r.t. channel coefficients in the time domain, it is rewritten in a most appropriate form. To do so, let's introduce the following notation: $\mathbf{h} = \left[\mathbf{h}_1^T \cdots \mathbf{h}_{N_r}^T\right]^T$ is a vector of size $N_r N_t N \times 1$ where $\mathbf{h}_r = \left[\mathbf{h}_{1,r}^T \cdots \mathbf{h}_{N_t,r}^T\right]^T$; $\mathbf{X}_{D_i} = diag\{\mathbf{x}_i\}$ is a diagonal matrix of size $K \times K$; $\mathbf{X} = \left[\mathbf{X}_{D_1}\mathbf{W} \cdots \mathbf{X}_{D_{N_t}}\mathbf{W}\right]$ of size $K \times NN_t$; and $\tilde{\mathbf{X}} = \mathbf{I}_{N_r} \otimes \mathbf{X}$ is a matrix of size $N_r K \times NN_t N_r$ with $\otimes$ being the Kronecker product. Accordingly, equation (A.3) is rewritten as follows:

$$\mathbf{y} = \tilde{\mathbf{X}}\mathbf{h} + \mathbf{v}. \tag{A.4}$$

Furthermore, some priors on the channel impulse response and/or on the communications system can be available. Hence, by assuming a specular channel model in our case, the communications channel impulse response, of the i-th user, is expressed as a function of the fading, the delay and the Direction Of Arrival (DOA), in time domain, as follows:

$$\mathbf{h}_i(t) = \sum_{l=1}^{M} \bar{h}_{i,l} \mathbf{a}(\alpha_{i,l}) sinc(t - \tau_{i,l}), \tag{A.5}$$

where M is the number of paths for each transmitter, $\bar{h}_{i,l}$ is the complex fading related to the l-th path, $\tau_{i,l}$ being the l-th path delay and $\mathbf{a}(\alpha_{i,l}) = [1\ e^{-j2\pi \frac{d}{\lambda} cos(\alpha_{i,l})} \ldots e^{-j2\pi \frac{d}{\lambda} r_1 cos(\alpha_{i,l})} \ldots e^{-j2\pi \frac{d}{\lambda}(N_r-1)cos(\alpha_{i,l})}]^T$ is the steering vector with $\alpha_{i,l}$ being the corresponding DOA[1]; while λ and d represent respectively the wave length and the distance separating two adjacent receive antennas.

[1] For simplicity, we assumed that the receive antenna corresponds to a uniform linear array.

Without loss of generality, a block-type pilot arrangement is adopted in this paper. In such a scheme all sub-carriers are used for pilots and data within a specific period of time. For a pilot-based channel estimation, N_p pilot symbols will be considered. N_d i.i.d data symbols will be added to the pilots for semi-blind approaches. Both pilots and data are assumed to be OFDM symbols of size K.

A.3 CRB derivation for subcarrier channel coefficients estimation

This section is dedicated to the CRB derivation for subcarrier channel coefficients estimation. Basically, the CRB is obtained as the inverse of the Fisher Information Matrix (FIM) denoted by $\mathbf{J}_{\boldsymbol{\theta\theta}}$ where $\boldsymbol{\theta}$ is the unknown deterministic parameters vector to be estimated [61]. Hence, by taking into account the pilots and data (that are statistically independent) in a semi-blind fashion, the total FIM is expressed as follows:

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = \mathbf{J}^{p}_{\boldsymbol{\theta\theta}} + \mathbf{J}^{d}_{\boldsymbol{\theta\theta}}, \tag{A.6}$$

where $\mathbf{J}^{p}_{\boldsymbol{\theta\theta}}$ is the FIM associated to the known pilots while $\mathbf{J}^{d}_{\boldsymbol{\theta\theta}}$ is related to the unknown data.

Actually, instead of estimating the channel taps, many existing OFDM receivers estimate the subcarrier channel coefficients (i.e. the vector $\boldsymbol{\lambda}_{i,r}$) as if they were 'independent' (see e.g. [145]), by ignoring the relation between these coefficients through the Fourier transform of the channel taps. Therefore, under the assumption of known signal and noise powers[2], the parameters vector to be estimated is given by:

$$\boldsymbol{\theta} = [\boldsymbol{\Lambda}^T \;\; (\boldsymbol{\Lambda}^*)^T]^T, \tag{A.7}$$

where $\boldsymbol{\Lambda} = \left[\tilde{\boldsymbol{\lambda}}^T_{1,1} \cdots \tilde{\boldsymbol{\lambda}}^T_{i,r} \cdots \tilde{\boldsymbol{\lambda}}^T_{N_t,N_r}\right]^T$, which is a $KN_rN_t \times 1$ vector with $\tilde{\boldsymbol{\lambda}}_{i,r} = \mathbf{W}\mathbf{h}_{i,r}$ of size $K \times 1$.

A.3.1 Pilot-based CRB derivation

By considering only pilots and in order to have explicitly the vector $\boldsymbol{\Lambda}$, the system model given by equation (A.3) is re-expressed as:

$$\mathbf{y} = \bar{\mathbf{X}}\boldsymbol{\Lambda} + \mathbf{v}, \tag{A.8}$$

where $\bar{\mathbf{X}} = [\mathbf{I}_{N_r} \otimes \mathbf{x}_1 ... \mathbf{I}_{N_r} \otimes \mathbf{x}_{N_t}]$.

Since the noise is an i.i.d. random process and according to the results in [59], the FIM when considering N_p pilot OFDM symbols can be expressed as follows:

$$\mathbf{J}^{p}_{\boldsymbol{\theta\theta}} = \sum_{i=1}^{N_p} \mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}}, \tag{A.9}$$

where $\mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}}$, which refers to the FIM associated to the i-th pilot OFDM symbol, is given by [146, 147]:

$$\mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}} = E\left\{ \left(\frac{\partial \ln p(\mathbf{y}(i), \boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}^*} \right) \left(\frac{\partial \ln p(\mathbf{y}(i), \boldsymbol{\Lambda})}{\partial \boldsymbol{\theta}^*} \right)^H \right\}, \tag{A.10}$$

with $E(.)$ being the expectation operator and $p(\mathbf{y}(i), \boldsymbol{\Lambda})$ is the probability density function of the received signal given $\boldsymbol{\Lambda}$.

[2] The case of unknown noise and signal powers leads to similar conclusions and is omitted here for simplicity.

One could show that equation (A.10) can be expressed by:

$$\mathbf{J}_{\boldsymbol{\theta\theta}}^{p_i} = \frac{\bar{\mathbf{X}}(i)^H \bar{\mathbf{X}}(i)}{\sigma_{\mathbf{v}}^2}. \tag{A.11}$$

Moreover, based on the FIM for a complex parameter (see [62, 63]), the pilot-based FIM is given by:

$$\mathbf{J}_{\boldsymbol{\theta\theta}}^{p_i} = \begin{pmatrix} \mathbf{J}_{\mathbf{hh}}^{p_i} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{\mathbf{h^*h^*}}^{p_i} \end{pmatrix}, \tag{A.12}$$

where $\mathbf{J}_{\mathbf{h^*h^*}}^{p_i} = (\mathbf{J}_{\mathbf{hh}}^{p_i})^*$. Therefore, the pilot-based CRB is obtained as the inverse of $\mathbf{J}_{\boldsymbol{\theta\theta}}^{p}$.

A.3.2 Semi-blind CRB derivation

For the semi-blind channel estimation case, both pilots and data are taken into account in the derivation of the FIM as shown in equation (A.6). Let's assume that the data symbols are i.i.d. circular Gaussian distributed with zero mean and a diagonal covariance matrix composed of the users' transmit powers i.e. $\mathbf{C}_{\mathbf{x}} = diag\left(\boldsymbol{\sigma}_{\mathbf{x}}^2\right)$ with $\boldsymbol{\sigma}_{\mathbf{x}}^2 \overset{\text{def}}{=} \left[\sigma_{\mathbf{x}_1}^2 \cdots \sigma_{\mathbf{x}_{N_t}}^2\right]^T$ where $\sigma_{\mathbf{x}_i}^2$ denotes the transmit power of the i-th user. Under this assumption, the received signal $\mathbf{y}$ is circular Gaussian with covariance matrix:

$$\mathbf{C}_{\mathbf{y}} = \sum_{i=1}^{N_t} \sigma_{\mathbf{x}_i}^2 \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^H + \sigma_{\mathbf{v}}^2 \mathbf{I}_{KN_r}. \tag{A.13}$$

The total data FIM has the following form:

$$\mathbf{J}_{\boldsymbol{\theta\theta}}^{d} = N_d \begin{bmatrix} \mathbf{J}_{\boldsymbol{\Lambda\Lambda}}^{d} & \mathbf{J}_{\boldsymbol{\Lambda\Lambda}^*}^{d} \\ \mathbf{J}_{\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}}^{d} & \mathbf{J}_{\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^*}^{d} \end{bmatrix}, \tag{A.14}$$

where $\mathbf{J}_{\boldsymbol{\Lambda\Lambda}}^{d}$ is a $(N_r N_t K)$-dimensional matrix with elements $J_{\Lambda_i \Lambda_j}^{d}$ given by:

$$J_{\Lambda_i \Lambda_j}^{d} = tr\left\{ \mathbf{C}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial \Lambda_i^*} \mathbf{C}_{\mathbf{y}}^{-1} \left(\frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial \Lambda_j^*} \right)^H \right\}. \tag{A.15}$$

Once the total FIM $\mathbf{J}_{\boldsymbol{\theta\theta}}$ is obtained, it is inverted to obtain the CRB matrix. Then, the top-left $KN_tN_r \times KN_tN_r$ subblock of the CRB matrix is extracted to deduce the CRB for the subcarrier channel coefficients estimation (i.e. in the frequency domain).

A.4 CRB derivation for channel taps estimation

This section is dedicated to the CRB derivation for channel estimation considering directly the channel coefficients in the time domain. In this case, the complex representation of parameters vector $\boldsymbol{\theta}$ is defined as follows:

$$\boldsymbol{\theta} = [\mathbf{h}^T \ \ (\mathbf{h}^*)^T]^T. \tag{A.16}$$

A.4.1 Pilot-based CRB derivation

In a similar way as in A.3.1, the FIM associated to the i-th pilot OFDM symbol, is given by

$$\mathbf{J}_{\boldsymbol{\theta\theta}}^{p_i} = E\left\{ \left(\frac{\partial \ln p(\mathbf{y}(i), \mathbf{h})}{\partial \boldsymbol{\theta}^*} \right) \left(\frac{\partial \ln p(\mathbf{y}(i), \mathbf{h})}{\partial \boldsymbol{\theta}^*} \right)^H \right\}, \tag{A.17}$$

with $p(\mathbf{y}(i),\mathbf{h})$ being the probability density function of the received signal given $\mathbf{h}$.

According to results in [59], the FIM associated to the i-th pilot OFDM symbol is deduced as:

$$\mathbf{J}^{p_i}_{\boldsymbol{\theta\theta}} = \frac{\tilde{\mathbf{X}}(i)^H \tilde{\mathbf{X}}(i)}{\sigma^2_{\mathbf{v}}}. \tag{A.18}$$

Then, the pilot-based CRB is obtained as the inverse of $\mathbf{J}^{p}_{\boldsymbol{\theta\theta}}$.

A.4.2 Semi-blind CRB derivation

In this section, the derivation of the CRB for the semi-blind channel estimation case is considered. Only the Second Order Statistics (SOS) are considered via a Circular Gaussian data model. More details about the case of Higher Order Statistics (HOS), using a finite alphabet source signal, can be found in [59]. The derivation is similar to that introduced in section A.3.2, however in this case we have:

$$\mathbf{J}^{d}_{\boldsymbol{\theta\theta}} = N_d \begin{bmatrix} \mathbf{J}^{d}_{\mathbf{hh}} & \mathbf{J}^{d}_{\mathbf{hh}^*} \\ \mathbf{J}^{d}_{\mathbf{h}^*\mathbf{h}} & \mathbf{J}^{d}_{\mathbf{h}^*\mathbf{h}^*} \end{bmatrix}, \tag{A.19}$$

where $\mathbf{J}^{d}_{\mathbf{hh}}$ is a $(N_r N_t N)$-dimensional matrix with elements $J^{d}_{h_i h_j}$ given by:

$$J^{d}_{h_i h_j} = tr\left\{ \mathbf{C}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial h_i^*} \mathbf{C}_{\mathbf{y}}^{-1} \left(\frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial h_j^*} \right)^H \right\}, \tag{A.20}$$

with $\frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial h_i^*} = \boldsymbol{\lambda} \mathbf{C}_{\mathbf{x}} \frac{\partial \boldsymbol{\lambda}^H}{\partial h_i^*}$.

Actually, this derivation is performed by specifying for each index $i = 1, \cdots, NN_rN_t$, the corresponding channel tap, receive antenna and user indices i.e. $i_N = 1, \cdots, N$; $i_{N_r} = 1, \cdots, N_r$ and $i_{N_t} = 1, \cdots, N_t$. Therefore, after some simplifications, one could obtain

$$\frac{\partial \mathbf{C}_{\mathbf{y}}}{\partial h_i^*} = \sigma^2_{\mathbf{x}_{i_{N_t}}} \boldsymbol{\lambda}_{i_{N_t}} \frac{\partial \boldsymbol{\lambda}^H_{i_{N_t}}}{\partial h_i^*}. \tag{A.21}$$

Finally, the top-left $NN_tN_r \times NN_tN_r$ subblock of the CRB matrix, obtained by inverting $\mathbf{J}_{\boldsymbol{\theta\theta}}$, refers to the CRB for the channel coefficients estimation.

A.5 CRB derivation for specular channel estimation

This section derives the CRB of semi-blind channel estimation when considering a specular model for the channel impulse response as given in equation (A.5). The vector parameter of size $4NN_rN_t \times 1$ to be estimated is given by:

$$\boldsymbol{\theta} = [\bar{\boldsymbol{h}}^T \ \ (\bar{\boldsymbol{h}}^*)^T \ \ \boldsymbol{\tau}^T \ \ \boldsymbol{\alpha}^T]^T, \tag{A.22}$$

with $\bar{\boldsymbol{h}}, \boldsymbol{\tau}, \boldsymbol{\alpha}$ being vectors of size $NN_rN_t \times 1$ containing respectively the complex fading, the delay and the DOA of channel taps between all users and the receive antennas.

According to the FIM derivation of parameter transformation [61], the FIM in such a case is based on that derived in section A.4. Thus, by denoting $\mathbf{J}^{\mathbf{h}}_{\boldsymbol{\theta\theta}}$ the FIM of the semi-blind channel coefficients estimation, we have:

$$\mathbf{J}_{\boldsymbol{\theta\theta}} = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}^H \mathbf{J}^{\mathbf{h}}_{\boldsymbol{\theta\theta}} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}}, \tag{A.23}$$

where

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} = [\frac{\partial \mathbf{h}}{\partial \bar{\boldsymbol{h}}}, \frac{\partial \mathbf{h}}{\partial \bar{\boldsymbol{h}}^*}, \frac{\partial \mathbf{h}}{\partial \boldsymbol{\tau}}, \frac{\partial \mathbf{h}}{\partial \boldsymbol{\alpha}}], \tag{A.24}$$

with $\mathbf{J}^{\mathbf{h}}_{\boldsymbol{\theta\theta}} \in \mathbb{C}^{NN_rN_t \times NN_rN_t}$ and $\mathbf{J}_{\boldsymbol{\theta\theta}} \in \mathbb{C}^{4NN_rN_t \times 4NN_rN_t}$.

The former derivatives of $\mathbf{h}$ w.r.t. to the propagation parameters are obtained thanks to the formula (A.5). Note that for real parameters ($\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$), the derivation is straightforward. However for complex parameters, i.e. $\mathbf{h}$, one can use either the derivation w.r.t. real and imaginary parts or equivalently, the derivation w.r.t. the complex parameter and its conjugate [148]. Finally, the CRB is obtained by inverting the FIM.

A.6 Simulation results

This section is dedicated to analyze the behavior of the CRB for different scenarios as described previously: the sub-carrier channel coefficient estimation ($\mathrm{CRB}^{\lambda}_{OP}$,$\mathrm{CRB}^{\lambda}_{SB}$), the channel taps estimation ($\mathrm{CRB}^{\mathbf{h}}_{OP}$,$\mathrm{CRB}^{\mathbf{h}}_{SB}$) and the specular channel coefficients estimation ($\mathrm{CRB}^{specular}_{OP}$,$\mathrm{CRB}^{specular}_{SB}$) where OP stands for the pilot-based estimation; whereas SB refers to the semi-blind framework. The pilot symbols are generated according to Zadoff-Chu sequences [70]. The simulation parameters are summarized in table A.1 otherwise mentioned.

Figure A.1 illustrates the behavior of the normalized CRB $\left(\frac{tr\{CRB\}}{\|\mathbf{h}\|^2}\right)$ versus SNR for the three considered scenarios. Adopting a semi-blind framework helps lowering the CRB and hence performing better than pilot-based approaches. One could notice that with only one pilot symbol and few data symbols (40 in our case), the efficiency of the semi-blind framework is well illustrated while preserving a lower overhead. On the other hand, one notices that compared to the frequency domain, estimating directly the channel taps in time domain gives much better performance, which is further enhanced when considering a parametric propagation model for the communications channel (specular representation in our case).

Figure A.2 illustrates the semi-blind CRB behavior of the channel coefficients estimation in frequency ($\mathrm{CRB}^{\lambda}_{\mathrm{SB}}$) and time ($\mathrm{CRB}^{\mathbf{h}}_{\mathrm{SB}}$) domain w.r.t. to the number of pilot symbols N_p. It can be noticed that, in order to reach same performance as the semi-blind CRB for specular channel estimation($\mathrm{CRB}^{specular}_{\mathrm{SB}}$) with one pilot symbol, one needs around 60 pilot symbols when estimating the channel coefficients in time domain and, seemingly, even much more when directly estimating the frequency coefficients.

Figure A.3 assesses the effect of the number of data symbols on the CRBs semi-blind framework. One can notice that with just tens of data symbols, the semi-blind channel estimation is further enhanced, when considering the channel parameters. Also, it is noticed that the CRB curves tends to flatten with high number of data symbols, which indicates that only a reasonable number of data symbols is needed for better channel estimation performance.

Parameters	Specifications
Number of receive antennas	$N_r = 2$
Number of transmitters	$N_t = 2$
Number of channel paths	$M = 4$
Number of OFDM sub-carriers	$K = 64$
Number of OFDM pilot symbols	$N_p = 1$
Number of OFDM data symbols	$N_d = 40$
Pilot signal powers (dBm)	$P_{x_p} = [23\ 13]$
Data signal powers (dBm)	$P_{x_d} = [20\ 18.8]$
Channel Fading	[0.4 0.6 0.1 0.01;0.3 0.9 0.5 0.3];
Channel delay	[0.4 0.6 0.1 0.4;0.3 0.9 0.5 0.1];
DOA	$[\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{6}, \frac{\pi}{8}; \frac{\pi}{3}, \frac{\pi}{4}, \frac{\pi}{7}, \frac{\pi}{8}]$

Table A.1: *Simulation parameters.*

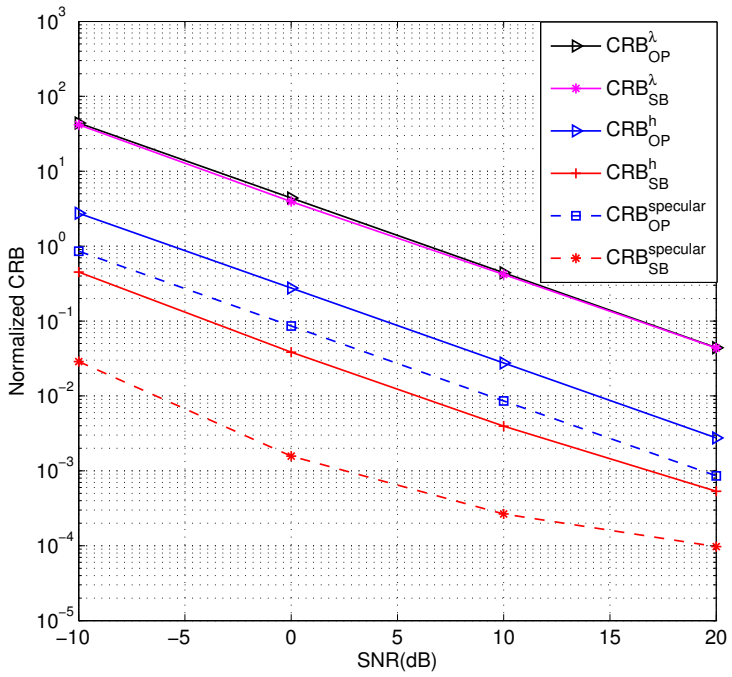


Figure A.1: *Normalized CRB vs SNR.*

A.7 Conclusion

This paper focused on the effect of side information on the performance of semi-blind channel estimation; when considering MIMO-OFDM communications systems. Three scenarios have been investigated,

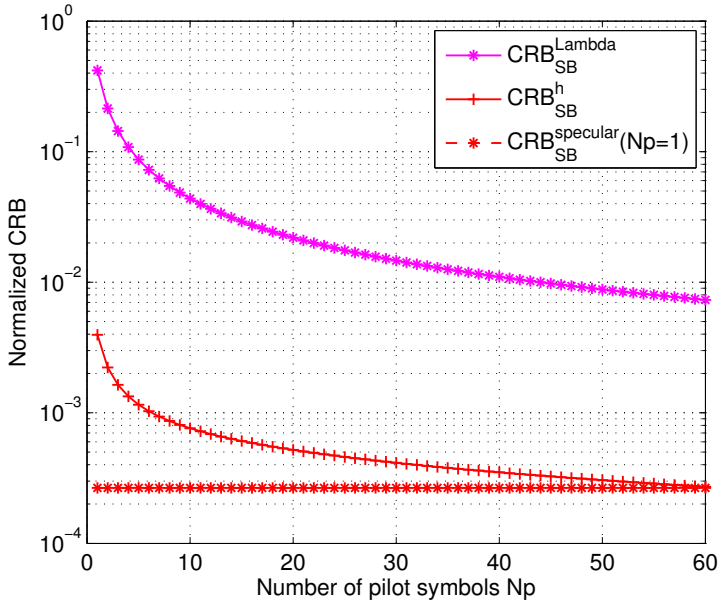


Figure A.2: *Normalized CRB vs number of pilot symbols.*

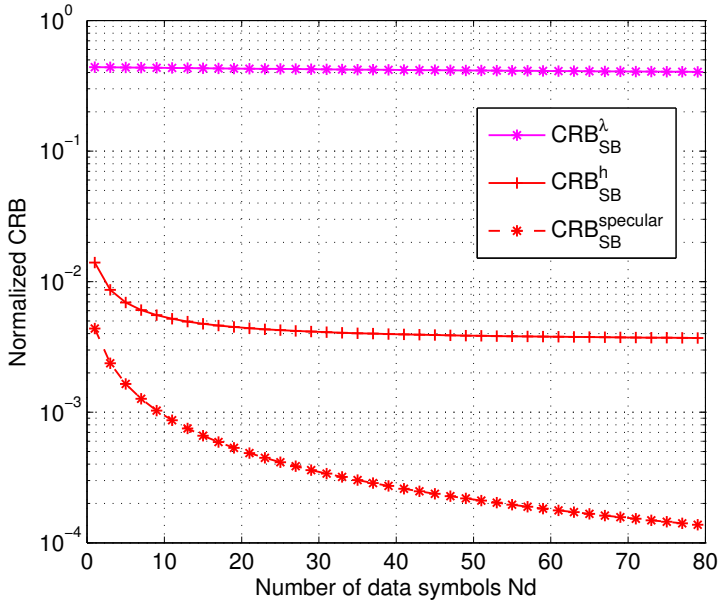


Figure A.3: *Normalized CRB vs number of data symbols.*

estimating: (i) the channel fading coefficients in the frequency domain; (ii) the channel taps in the time domain; and (iii) the propagation channel parameters when considering a specular channel model. To quantify and compare their performance limits, CRBs have been derived. An experimental comparative analysis revealed that estimating the channel taps in time domain, leads to better performance than estimating the frequency response. Furthermore, this performance gain is more enhanced by exploiting priors on the channels (channel model in our case) while preserving a lower overhead.

Bibliography

[1] N. Kalouptsidis and P. Koukoulas, "Blind identification of Volterra-Hammerstein systems," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2777–2787, 2005.

[2] D. P. Agrawal and Q.-A. Zeng, *Introduction to wireless and mobile systems*. Cengage learning, 2015.

[3] R. J. Soar, "Soldier system wireless power and data transmission," May 3 2016, US Patent 9,331,495.

[4] J. A. del Peral-Rosado, R. Raulefs, J. A. López-Salcedo, and G. Seco-Granados, "Survey of cellular mobile radio localization methods: From 1G to 5G," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 1124–1148, 2018.

[5] L. L. Hanzo, L.-L. Yang, E.-L. Kuan, and K. Yen, *Overview of Multicarrier CDMA*, 2004.

[6] G. Gu and G. Peng, "The survey of GSM wireless communication system," in *2010 International Conference on Computer and Information Application*, 2010, pp. 121–124.

[7] F. W. Y. Bouguen, E. Hardouin, *LTE et les réseaux 4G*. Paris, France: Editions Eyrolles, 2013.

[8] R. Vannithamby and S. Talwar, *Massive MIMO Communications*, 2017.

[9] A. J. PAULRAJ, D. A. GORE, R. U. NABAR, and H. BOLCSKEI, "An overview of MIMO communications - a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, 2004.

[10] K. Choi and H. Liu, *MIMO System - Part II: Spatial Multiplexing*, 2016.

[11] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[12] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[13] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[14] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[15] A. Narasimhamurthy, M. Banavar, and C. Tepedelenliouglu, *OFDM Systems for Wireless Communications*, 2010.

[16] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for next-generation wireless systems," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1430–1469, 2007.

[17] "5G today: Modulation technique alternatives, author=Ayanoglu, Ender, booktitle=2016 International Conference on Computing, Networking and Communications (ICNC), pages=1–5, year=2016, organization=IEEE."

[18] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[19] R. S. E. Perahi, *Next Generation Wireless LANS*. New York, USA: Cambridge University Press, 2008.

[20] I.-T. Lu and K.-J. Tsai, "Channel estimation in a proposed IEEE802.11n OFDM MIMO WLAN system," in *IEEE Sarnoff Symposium*, April 2007, pp. 1–5.

[21] S. Noh, Y. Sung, and M. D. Zoltowski, "A new precoder design for blind channel estimation in MIMO-OFDM systems," *IEEE Trans. on Wireless Communications*, vol. 13, no. 12, pp. 7011–7024, 2014.

[22] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. on Information Theory.*, Mar 1994.

[23] A. K. Jagannatham and B. D. Rao, "Whitening-rotation-based semi-blind MIMO channel estimation," *IEEE Trans. on Signal Processing*, vol. 54, no. 3, pp. 861–869, 2006.

[24] K. Liu, J. P. C. L. da Costa, H. C. So, and A. L. F. de Almeida, "Semi-blind receivers for joint symbol and channel estimation in space-time-frequency MIMO-OFDM systems," *IEEE Trans. on Signal Processing*, vol. 61, no. 21, pp. 5444–5457, 2013.

[25] G. Favier and A. L. F. de Almeida, "Tensor space-time-frequency coding with semi-blind receivers for MIMO wireless communication systems," *IEEE Trans. on Signal Processing*, vol. 62, no. 22, pp. 5987–6002, 2014.

[26] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

[27] F. R. Pavan, M. T. Silva, and M. D. Miranda, "A numerically robust blind equalization scheme applied to MIMO communications systems," *Journal of the Franklin Institute*, vol. 355, no. 1, pp. 596–624, 2018.

[28] 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, p. 4363, 2005.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[30] R. C. de Lamare, "Massive MIMO systems: Signal processing challenges and research trends," *arXiv preprint arXiv:1310.7282*, 2013.

[31] N. Hassan and X. Fernando, "Massive MIMO wireless networks: An overview," *Electronics*, vol. 6, no. 3, p. 63, 2017.

[32] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *40th Asilomar Conference on Signals, Systems and Computers (ACSSC)*. IEEE, 2006, pp. 359–363.

[33] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[34] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *International Symposium on Information Theory (ISIT)*. IEEE, 2009, pp. 2184–2188.

[35] V. J. Mathews, "Adaptive polynomial filters," *IEEE Signal Processing Magazine*, vol. 8, no. 3, pp. 10–26, 1991.

[36] J. Schoukens and L. Ljung, "Nonlinear system identification: A user-oriented road map," *IEEE Control Systems Magazine*, vol. 39, no. 6, pp. 28–99, 2019.

[37] A. Amari, O. A. Dobre, R. Venkatesan, O. S. Kumar, P. Ciblat, and Y. Jaouën, "A survey on fiber nonlinearity compensation for 400 Gb/s and beyond optical communication systems," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 3097–3113, 2017.

[38] O. Rekik, A. Ladaycia, A. Mokraoui, and K. Abed-Meraim, "CRB-based performance analysis of semi-blind channel estimation for massive MIMO-OFDM systems with pilot contamination," *IET Communications*, vol. 13, no. 20, pp. 3479–3488, 2019.

[39] O. Rekik, A. Ladaycia, K. Abed-Meraim, and A. Mokraoui, "Performance bounds analysis for semi-blind channel estimation with pilot contamination in massive MIMO-OFDM systems," in *26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1267–1271.

[40] X. Zhang, S. Wu, S. Yu, and J. Liu, "Spectral efficiency of massive MIMO networks with pilot contamination and channel covariance matrix estimation," *IET Communications*, vol. 13, no. 1, pp. 59–65, 2018.

[41] D. C. Araujo, T. Maksymyuk, A. L. de Almeida, T. Maciel, J. C. Mota, and M. Jo, "Massive MIMO: survey and future research topics," *IET Communications*, vol. 10, no. 15, pp. 1938–1946, 2016.

[42] Z. Gong, C. Li, and F. Jiang, "Pilot contamination mitigation strategies in massive MIMO systems," *IET Communications*, vol. 11, no. 16, pp. 2403–2409, 2017.

[43] S. Ali, Z. Chen, and F. Yin, "Eradication of pilot contamination and zero forcing precoding in the multi-cell TDD massive MIMO systems," *IET Communications*, vol. 11, no. 13, pp. 2027–2034, 2017.

[44] J. Zhao, S. Ni, Y. Gong, and Q. Zhang, "Pilot contamination reduction in TDD-based massive MIMO systems," *IET Communications*, 2019.

[45] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 192–201, 2013.

[46] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[47] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, 2013.

[48] J. Ma and L. Ping, "Data-aided channel estimation in large antenna systems," in *International Conference on Communications (ICC)*. IEEE, 2014, pp. 4626–4631.

[49] L. Li, A. Ashikhmin, and T. Marzetta, "Pilot contamination precoding for interference reduction in large scale antenna systems," in *51st Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2013, pp. 226–232.

[50] J. Fan, W. Li, and Y. Zhang, "Pilot contamination mitigation by fractional pilot reuse with threshold optimization in massive MIMO systems," *Digital Signal Processing*, 2018.

[51] M. AlKhaled and E. Alsusa, "Impact of pilot sequence contamination in massive MIMO systems," *IET Communications*, vol. 11, no. 13, pp. 2005–2011, 2017.

[52] T. Peken, G. Vanhoy, and T. Bose, "Blind channel estimation for massive MIMO," *Analog Integrated Circuits and Signal Processing*, vol. 91, no. 2, pp. 257–266, 2017.

[53] E. Amiri, R. Mueller, and W. Gerstacker, "Blind pilot decontamination in massive MIMO by Independent Component Analysis," in *Globecom Workshops*. IEEE, 2017, pp. 1–5.

[54] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 1, pp. 525–536, 2016.

[55] A. S. Alwakeel and A. H. Mehana, "Semi-blind channel estimation and interference alignment for massive MIMO network," in *Global Communications Conference GLOBECOM*. IEEE, 2017, pp. 1–6.

[56] W. A. W. M. Mahyiddin, P. A. Martin, and P. J. Smith, "Performance of synchronized and unsynchronized pilots in finite massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6763–6776, 2015.

[57] A. Pitarokoilis, E. Björnson, and E. G. Larsson, "On the effect of imperfect timing synchronization on pilot contamination," in *International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–6.

[58] X. Zou and H. Jafarkhani, "Asynchronous channel training in multi-cell massive MIMO," *arXiv preprint arXiv:1711.03197*, 2017.

[59] A. Ladaycia, A. Mokraoui, K. Abed-Meraim, and A. Belouchrani, "Performance bounds analysis for semi-blind channel estimation in MIMO-OFDM communications systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 5925–5938, 2017.

[60] Y. S. Cho, J. Kim, W. Y. Yang, and C. G. Kang, *MIMO-OFDM wireless communications with MATLAB*. John Wiley & Sons, 2010.

[61] S. M. Kay, "Fundamentals of statistical signal processing, volume I: Estimation theory (v. 1)," *PTR Prentice-Hall, Englewood Cliffs*, 1993.

[62] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "New results on deterministic Cramér–Rao bounds for real and complex parameters," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1032–1049, 2012.

[63] L. Berriche and K. Abed-Meraim, "Stochastic Cramér-Rao bounds for semiblind MIMO channel estimation," in *IEEE International Symposium on Signal Processing and Information Technology*, 2004, pp. 119–122.

[64] H. Abeida and J.-P. Delmas, "Gaussian Cramer-Rao bound for direction estimation of noncircular signals in unknown noise fields," *IEEE Transactions on Signal Processing*, vol. 53, no. 12, pp. 4610–4618, 2005.

[65] S.-M. Omar, D. T. Slock, and O. Bazzi, "Bayesian and deterministic CRBs for semi-blind channel estimation in SIMO single carrier cyclic prefix systems," in *22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2011, pp. 1682–1686.

[66] 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.

[67] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.

[68] P. Comon, "Independent Component Analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[69] A. Ladaycia, A. Mokraoui, K. Abed-Meraim, and A. Belouchrani, "Toward green communications using semi-blind channel estimation," in *25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 2254–2258.

[70] Y. Bouguen, E. Hardouin, and F.-X. Wolff, *LTE pour les reseaux 4G*. Editions Eyrolles, 2012.

[71] O. Rekik, A. Mokraoui, A. Ladaycia, and K. Abed-Meraim, "Semi-blind source separation based on multi-modulus criterion: Application for pilot contamination mitigation in massive MIMO communications systems," in *19th IEEE International Symposium on Communications and Information Technologies (ISCIT)*, 2019, pp. 31–35.

[72] O. Rekik, K. Abed-Meraim, and A. Mokraoui, "Multi-modulus based semi-blind source separation for MIMO-OFDM communications systems," in *11th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2020, pp. 1–5.

[73] Z. Luo, C. Li, and L. Zhu, "A comprehensive survey on blind source separation for wireless adaptive processing: Principles, perspectives, challenges and new research directions," *IEEE Access*, vol. 6, pp. 66 685–66 708, 2018.

[74] X. Yu, D. Hu, and J. Xu, *Blind source separation: theory and applications*. John Wiley & Sons, 2013.

[75] S. Abrar and A. K. Nandi, "Blind equalization of square-QAM signals: a multimodulus approach," *IEEE Transactions on Communications*, vol. 58, no. 6, 2010.

[76] 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.

[77] S. A. W. Shah, K. Abed-Meraim, and T. Y. Al-Naffouri, "Multi-modulus algorithms using hyperbolic and givens rotations for blind deconvolution of MIMO systems," in *International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 2155–2159.

[78] P. Comon, "Block methods for channel identification and source separation," in *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No. 00EX373)*. IEEE, 2000, pp. 87–92.

[79] N. Fatema, Y. Xiang, and I. Natgunanathan, "Analysis of a semi-blind pilot decontamination method in massive MIMO," in *27th International Telecommunication Networks and Applications Conference (ITNAC)*. IEEE, 2017, pp. 1–6.

[80] X. Hong and S. Chen, "Recursive least squares semi-blind beamforming for MIMO using decision directed adaptation and constant modulus criterion," *International Journal of Automation and Computing*, vol. 14, no. 4, pp. 442–449, 2017.

[81] A. Zerguine, "Convergence and steady-state analysis of the normalized least mean fourth algorithm," *Digital Signal Processing*, vol. 17, no. 1, pp. 17–31, 2007.

[82] 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.

[83] G. D. Forney, Jr, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM Journal on Control*, vol. 13, no. 3, pp. 493–520, 1975.

[84] O. Rekik, K. Abed-Meraim, M. Pesavento, and A. Mokraoui, "Semi-blind sparse channel estimation and data detection by successive convex approximation," in *21st IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.

[85] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint channel estimation and data detection in MIMO-OFDM systems: A sparse Bayesian learning approach," *IEEE Transactions on signal processing*, vol. 63, no. 20, pp. 5369–5382, 2015.

[86] A. Mishra, A. K. Jagannatham, and L. Hanzo, "Sparse Bayesian learning-aided joint sparse channel estimation and ML sequence detection in space-time trellis coded MIMO-OFDM systems," *IEEE Transactions on Communications*, 2019.

[87] R. Prasad, C. R. Murthy, and B. D. Rao, "Joint approximately sparse channel estimation and data detection in OFDM systems using sparse bayesian learning," *IEEE Transactions on Signal Processing*, vol. 62, no. 14, pp. 3591–3603, 2014.

[88] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[89] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, no. 3, pp. 475–494, 2001.

[90] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[91] Y. Yang and M. Pesavento, "A unified successive pseudoconvex approximation framework," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3313–3328, 2017.

[92] Y. Yang, M. Pesavento, S. Chatzinotas, and B. Ottersten, "Successive convex approximation algorithms for sparse signal estimation with nonconvex regularizations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 6, pp. 1286–1302, 2018.

[93] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[94] M. G. Reuter and J. C. Hill, "An efficient, block-by-block algorithm for inverting a block tridiagonal, nearly block Toeplitz matrix," *Computational Science & Discovery*, vol. 5, no. 1, p. 014009, 2012.

[95] 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.

[96] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *IEEE Transactions on signal processing*, vol. 45, no. 11, p. 2714–2723, 1995.

[97] A. M. Rekik Ouahbi, Karim Abed-meraim and M. Nait-Meziane, "Contribution à l'estimation aveugle du canal de transmission dans les systèmes SIMO non linéaires," in *GRETSI*, 2019.

[98] T. Ogunfunmi, *Adaptive nonlinear system identification: the Volterra and Wiener model approaches*. Springer Science & Business Media, 2007.

[99] C. Cheng, Z. Peng, W. Zhang, and G. Meng, "Volterra-series-based nonlinear system modeling and its engineering applications: A state-of-the-art review," *Mechanical Systems and Signal Processing*, vol. 87, pp. 340–364, 2017.

[100] T. Ogunfunmi and S.-L. Chang, "Second-order adaptive Volterra system identification based on discrete nonlinear Wiener model," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 148, no. 1, pp. 21–29, 2001.

[101] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Transactions on Signal Processing*, pp. 3150–3161, 2008.

[102] R. Claser, V. H. Nascimento, and Y. V. Zakharov, "A low-complexity RLS-DCD algorithm for Volterra system identification," in *24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 6–10.

[103] G. L. Sicuranza and A. Carini, "A multichannel hierarchical approach to adaptive Volterra filters employing filtered-X affine projection algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1463–1473, 2005.

[104] J. C. Gómez and E. Baeyens, "Subspace-based identification algorithms for Hammerstein and Wiener models," *European Journal of Control*, vol. 11, no. 2, pp. 127–136, 2005.

[105] L. Yao and C.-C. Lin, "Identification of nonlinear systems by the genetic programming-based Volterra filter," *IET Signal Processing*, vol. 3, no. 2, pp. 93–105, 2009.

[106] T. Bouilloc and G. Favier, "Nonlinear channel modeling and identification using baseband Volterra-PARAFAC models," *Signal Processing*, vol. 92, no. 6, pp. 1492–1498, 2012.

[107] K. Batselier, Z. Chen, and N. Wong, "Tensor Network alternating linear scheme for MIMO Volterra system identification," *Automatica*, vol. 84, pp. 26–35, 2017.

[108] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind Maximum-Likelihood Identification of Wiener Systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3017–3029, 2009.

[109] L. Vanbeylen and R. Pintelon, "Blind Maximum-Likelihood identification of Wiener and Hammerstein nonlinear block structures," in *Block-oriented nonlinear system identification*. Springer, 2010, pp. 273–291.

[110] E.-W. Bai and M. Fu, "A blind approach to Hammerstein model identification," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, pp. 1610–1619, 2002.

[111] L. Vanbeylen, R. Pintelon, and J. Schoukens, "Blind maximum likelihood identification of Hammerstein systems," *Automatica*, vol. 44, no. 12, pp. 3139–3146, 2008.

[112] C. A. R. Fernandes, "Nonlinear MIMO communication systems: Channel estimation and information recovery using Volterra models," Ph.D. dissertation, Université de Nice Sophia Antipolis, 2009.

[113] L. Tao, H. Tan, C. Fang, and N. Chi, "Volterra series based blind equalization for nonlinear distortions in short reach optical CAP system," *Optics Communications*, vol. 381, pp. 240–243, 2016.

[114] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on signal processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[115] G. B. Giannakis and E. Serpedin, "Linear multichannel blind equalizers of nonlinear FIR Volterra channels," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 67–81, 1997.

[116] E. Serpedin, P. Ciblat, G. B. Giannakis, and P. Loubaton, "Performance analysis of blind carrier phase estimators for general QAM constellations," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1816–1823, 2001.

[117] S. A. Wahab Shah, K. Abed-Meraim, and T. Al-Naffouri, "Blind source separation algorithms using hyperbolic and Givens rotations for high-order QAM constellations," *IEEE Transactions on Signal Processing*, vol. 66, no. 7, pp. 1802 – 1816, 2017.

[118] K. Abed-Meraim and E. Moulines, "A maximum likelihood solution to blind identification of multichannel FIR filters," in *7th European Signal Processing Conference (EUSIPCO)*, 1994, pp. 1011–1014.

[119] G. K. Kaleh and R. Vallet, "Joint parameter estimation and symbol detection for linear or nonlinear unknown channels," *IEEE Transactions on Communications*, vol. 42, no. 7, pp. 2406–2413, 1994.

[120] S.-Z. Yu and H. Kobayashi, "An efficient forward-backward algorithm for an explicit-duration hidden markov model," *IEEE Signal Processing Letters*, vol. 10, no. 1, pp. 11–14, 2003.

[121] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[122] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital communication.* Springer Science & Business Media, 2012.

[123] S. Park, E. Serpedin, and K. Qaraqe, "Gaussian assumption: the least favorable but the most useful," *IEEE Signal Processing Magazine*, vol. 30, no. 4, pp. 183 – 186, 2013.

[124] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.

[125] A. K. Jagannatham and B. D. Rao, "Cramer-rao lower bound for constrained complex parameters," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 875–878, 2004.

[126] R. Lopez-Valcarce and S. Dasgupta, "Blind equalization of nonlinear channels from second-order statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3084–3097, 2001.

[127] A. Ladaycia, A. Belouchrani, K. Abed-Meraim, and A. Mokraoui, "Semi-blind MIMO-OFDM channel estimation using expectation maximisation like techniques," *IET Communications*, vol. 13, no. 20, pp. 3452–3462, 2019.

[128] J. Qadir, K.-L. A. Yau, M. A. Imran, Q. Ni, and A. V. Vasilakos, "IEEE access special section editorial: Artificial intelligence enabled networking," *IEEE Access*, vol. 3, pp. 3079–3082, 2015.

[129] E. Nachmani, Y. Be'ery, and D. Burshtein, "Learning to decode linear codes using deep learning," in *54th Annual Allerton Conference on Communication, Control, and Computing (Allerton).* IEEE, 2016, pp. 341–346.

[130] E. Nachmani, E. Marciano, D. Burshtein, and Y. Be'ery, "RNN decoding of linear block codes," *arXiv preprint arXiv:1702.07560*, 2017.

[131] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," *arXiv preprint arXiv:1409.2574*, 2014.

[132] Y.-S. Jeon, S.-N. Hong, and N. Lee, "Supervised-learning-aided communication framework for massive MIMO systems with low-resolution ADCs," *arXiv preprint arXiv:1610.07693*, 2016.

[133] N. Farsad and A. Goldsmith, "Detection algorithms for communication systems using deep learning," *arXiv preprint arXiv:1705.08044*, 2017.

[134] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* IEEE, 2017, pp. 1–6.

[135] E. De la Rosa, W. Yu, and X. Li, "Nonlinear system identification using deep learning and randomized algorithms," in *IEEE International Conference on Information and Automation, 2015.* IEEE, 2015, pp. 274–279.

[136] O. Ogunmolu, X. Gu, S. Jiang, and N. Gans, "Nonlinear systems identification using deep dynamic neural networks," *arXiv preprint arXiv:1610.01439*, 2016.

[137] A. Brusaferri, M. Matteucci, P. Portolani, and S. Spinelli, "Nonlinear system identification using a recurrent network in a Bayesian framework," in *17th IEEE International Conference on Industrial Informatics (INDIN) 2019*, vol. 1. IEEE, 2019, pp. 319–324.

[138] A. Rusiecki, "Trimmed categorical cross-entropy for deep learning with label noise," *Electronics Letters*, vol. 55, no. 6, pp. 319–320, 2019.

[139] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning. vol. 1," 2016.

[140] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[141] F. Meng, P. Chen, L. Wu, and X. Wang, "Automatic modulation classification: A deep learning enabled approach," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 760–10 772, 2018.

[142] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y.-D. Yao, "Modulation classification using convolutional neural network based deep learning model," in *26th Wireless and Optical Communication Conference (WOCC)*. IEEE, 2017, pp. 1–5.

[143] E. Dahlman, S. Parkvall, and J. Skold, *4G, LTE-advanced Pro and the Road to 5G*. Academic Press, 2016.

[144] C. Shin, R. W. Heath, and E. J. Powers, *IEEE Transactions on Vehicular Technology*.

[145] Y. Shen and E. Martinez, "Channel estimation in OFDM systems," 2006.

[146] S.-M. Omar, D. Slock, and O. Bazzi, "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*, Sept 2011, pp. 1682–1686.

[147] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*. Upper Saddle River, New Jersey 07458, USA: Prentice Hall PTR, INC, 1993.

[148] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "New results on deterministic Cramér–Rao bounds for real and complex parameters," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1032–1049, 2011.

Titre : **Conception et performance limites pour les futurs récepteurs mobiles.**

Mots clefs : MIMO massif, Bandes de Cramer-Rao, systèmes non linéaires, apprentissage profond.

Résumé : Les systèmes de communication MIMO (Multiple Input Multiple Output) ainsi que leur récente version à grande-échelle, appelée MIMO massive, sont considérés comme des technologies potentielles pour les standards de communication sans fil actuels et futurs, grâce à leurs puissantes capacités d'amélioration des performances. Néanmoins, afin d'exploiter pleinement leurs potentiels, une grande attention doit être accordée aux opérations d'identification du système et d'égalisation des canaux de transmission, qui restent une préoccupation actuelle. Dans ce contexte, la principale contribution de cette thèse s'inscrit dans le cadre de l'identification des systèmes de communication, à travers l'estimation des canaux, ainsi que l'égalisation des canaux via les techniques de séparation de sources. Ainsi, en adoptant des approches semi-aveugles, des analyses de performances ainsi que le développement d'algorithmes efficaces sont mis en avant en considérant différents contraintes/problèmes telle que la contamination des pilotes, rencontrés principalement dans les systèmes MIMO massifs, les effets des non-linéarités ainsi que les interférences inter-symboles et inter-utilisateurs .En plus, pour un meilleur gain en performance, l'accent est mis aussi sur l'exploitation des a priori sur les systèmes telles que les séquences d'entraînement (pilotes), la sparcité du canal et la structure de la matrice de données.

Title : **Design and performance bounds of future mobile receivers**

Keywords : massive MIMO, Cramer-Rao bounds, nonlinear systems, deep learning.

Abstract : Multiple Input Multiple Output (MIMO) communications systems as well as their recent large-scale version, called massive MIMO, are seen as potential technologies for current and future wireless communications standards, thanks to their powerful performance-enhancing capabilities. Nevertheless, in order to fully exploit all their potentials, great attention has to be given to the system identification and communications channel equalization tasks, which remain a current concern. In this context, the main contribution of this thesis falls into the scope of communications system identification, through channel estimation, as well as channel equalization via source separation techniques, for linear and nonlinear system models. Thus, by adopting semi-blind approaches, performance analysis as well as efficient algorithms development are put forward by considering different constraints/issues such as pilot contamination, encountered mainly in massive MIMO systems, nonlinearities effects as well as inter-symbol and inter-user interference. Furthermore, for a better performance gain, emphasis is also put on the exploitation of priors on the systems such as training sequences (pilots), channel's sparsity, and data matrix structure.

Université Paris XIII Dénommée Université Sorbonne Paris Nord

Laboratoire de Traitement et Transport de l'Information (L2TI)

Villetaneusse, France