UNIVERSITE DE NICE-SOPHIA ANTIPOLIS

ECOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

THESE

pour obtenir le titre de

Docteur en Sciences
de l'Université de Nice-Sophia Antipolis

Mention: Automatique et Traitement de Signal

présentée et soutenu par

Samir Mohamad Omar

Identification de Canal Déterministe et Bayésienne en Aveugle et Semi-Aveugle pour les Communications Sans Fil

Thése dirigée par Prof. Dirk SLOCK (Directeur) and Prof. Oussama BAZZI (Co-Directeur)
soutenue le 7 October 2011

Jury:

Prof. Bernard Fleury, Université Aalborg	Rapporteurs
Prof. Constantinos Papadias, Athens Information Technology (AIT)	Rapporteurs
Prof. Philippe Ciblat, Télécom ParisTech	Examinateur
Prof. Luc Deneire, Universit de Nice Sophia-Antipolis	Examinateur

DISSERTATION

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
from Université Nice Sophia Antipolis

Specialization: Signal Processing

Samir Mohamad Omar

Deterministic and Bayesian Blind and Semi-blind Channel Identification for Wireless Communications

This thesis has been supported by the PACA regional scholarship BDO550.

Defense took place on 7th of October 2011 before a committee composed of:

Reviewers	Prof. Bernard Fleury, Aalborg University
	Prof. Constantinos Papadias, Athens Information Technology (AIT)
Examiners	Prof. Philippe Ciblat, Telecom ParisTech
	Prof. Luc Deneire, University of Nice-Sophia Antipolis
Thesis Supervisor	Prof. Dirk Slock, EURECOM
Thesis Co-supervisor	Prof. Oussama Bazzi, Lebanese University

Abstract

During the last two decades there has been a great interest in blind and semi-blind channel estimation due to the advantages offered by these techniques over training-based ones. The most prominent is the augmentation of the throughput as a result of reducing the length of the training sequence/pilots required to estimate the channel at the receiver. Moreover, semi-blind techniques have the potential to estimate the channel in some situations where the training-based techniques fail. There exists a slew of algorithms that exploit either the second order statistics (SOS) or the higher order statistics (HOS) that have been derived and analyzed in the literature. Recently, this topic has been treated in the context of Space Time Block Coding (STBC), neural networks, multiuser scenario and cognitive radio, to name a few. In the first part of this thesis, we treat the blind channel estimation in the context of SIMO and MIMO cyclic prefix (CP) systems. We propose a novel approach to structure the sample covariance matrix, which in turn leads to a significant enhancement in the estimation quality, even when there is only a single OFDM symbol available at the receiver. On the other hand, we provide an analytical performance analysis of some CP SOS-based algorithms that permits to highlight some features of these algorithms and inspires the derivation of enhanced versions. At the end of this part, we introduce the variational Bayesian approach to carry out the joint ML/MAP estimation of channel and symbols. In the second part, we introduce and elaborate a classical Bayesian approach to estimate the channel and the symbols in the context of blind and semi-blind SIMO systems. As a consequence, six different ML/MAP estimators are derived and their performances are compared numerically by conducting Monte-Carlo simulations. Furthermore, we derive the corresponding Cramer-Rao Bounds (CRBs) for the various scenarios of these estimators. At the end of this part, we propose a novel quasi-Bayesian approach that exploits the knowledge of the power delay profile (PDP) to estimate only part of the channel taps while neglecting the rest. This approach can be applied to various deterministic algorithms that

exist in the literature, allowing their extension to a point that is midway between deterministic and Bayesian approaches. We show by simulations and by deriving the corresponding CRBs how this approach leads to a considerable improvement in the performance of many deterministic algorithms in terms of both Normalized Mean Squared Error (NMSE) and Symbol Error Probability (SER). Finally, in the third part we focus on the performance of Zero-Forcing (ZF) Linear Equalizers (LEs) or Decision-Feedback Equalizers (DFEs) for fading channels when they are based on (semi-)blind channel estimates. Although it has been known that various (semi-)blind channel estimation techniques have a receiver counterpart that is matched in terms of symbol knowledge hypotheses, we show here that these (semi-)blind techniques and corresponding receivers also match in terms of diversity order: the channel becomes (semi-)blindly unidentifiable whenever its corresponding receiver structure goes in outage. In the case of mismatched receiver and (semi-blind) channel estimation technique, the lower diversity order dominates. Various cases of (semi-)blind channel estimation and corresponding receivers are considered in detail.

Résumé

Au cours des deux dernières décennies il y a eu un grand intérêt dans l'estimation aveugle et semi-aveugle de canal en raison des avantages offerts par ces techniques sur celles utilisant des séquences pilote. L'avantage le plus connu est l'augmentation du débit en raison de la réduction de la longueur de la séquence pilote nécessaire pour l'estimation du canal dans le récepteur. Par ailleurs, les techniques semi-aveugles sont capables d'estimer le canal dans certaines situations où l'utilisation des pilotes échoue. Il existe un tas d'algorithmes qui exploitent soit les statistiques du second ordre (SOS) ou les statistiques d'ordre supérieur (HOS) qui ont été développés et analysés dans la littérature. Récemment, ce sujet a été traité dans le contexte du codage block espace-temps (STBC), des réseaux neurones, du scenario multi-utilisateurs, et de la radio cognitive pour n'en citer que quelques-uns. Dans la première partie de cette thèse, nous traitons l'estimation aveugle de canal dans le contexte des systèmes SIMO et MIMO avec préfixe cyclique (CP). Nous proposons une nouvelle approche qui consiste à structurer la réalisation de la matrice de covariance permettant ainsi une amèlioration significative de la qualité d'estimation, même quand seulement un seul symbole OFDM est disponible au niveau du récepteur. D'autre part, nous fournissons une analyse analytique de la performance de certains algorithmes SOS avec préfixe cyclique (CP) qui permet de mettre en évidence certaines caractéristiques de ces algorithmes et inspire la dérivation de versions améliorées. A la fin de cette partie, nous introduisons l'approche variationelle Baysienne pour effectuer l'estimation jointe ML/MAP d'estimation du canal et des symboles. Dans la deuxième partie, nous introduisons et élaborons une approche Baysienne classique pour estimer le canal et les symboles dans le contexte des systèmes SIMO aveugles et semi-aveugles. En conséquence, six différents estimateurs ML / MAP sont dérivés et leurs performances sont comparés numériquement en effectuant des simulations Monte-Carlo. Par ailleurs, nous dérivons les bornes de Cramer-Rao (CRB) correspondant aux différents scénarios de ces estimateurs. A la fin de cette partie, nous

proposons une nouvelle approche quasi Bayesienne qui exploite la connaissance du profil de délai de puissance (PDP) du canal pour estimer une partie des coefficients de canal tout en négligeant les autres. Cette approche peut être mise en œuvre dans diverses algorithmes déterministes qui existent dans la littérature, ce qui permet leur extension vers une catégorie qui se situe à mi-chemin entre les techniques déterministes et Bayesiennes. Nous montrons par des simulations et en dérivant les CRBs correspondant que cette approche conduit à une amélioration considérable de la performance de nombreux algorithmes déterministes tant en termes de l'erreur quadratique moyenne normalisé (NMSE) ou de la probabilité d'erreur sur les symboles (SER). Enfin, dans la troisième partie, nous nous concentrons sur les performances des égaliseurs linéaires et des égaliseurs à retour de décision "Decision Feedback" (DFE) à forçage à zéro (ZF) pour des canaux à évanouissements quand ils sont basés sur l'estimation (semi-) aveugle du canal. Bien qu'il est connu que les diverses techniques d'estimation de canal (semi-) aveugle ont en contrepartie un récepteur qui leur est adapté en termes d'hypothèses sur les connaissances des symboles, nous montrons ici que ces techniques (semi-) aveugles et les récepteurs correspondant concordent aussi en termes de l'ordre de la diversité : le canal devient non-identifiable en (semi-) aveugle dès que le récepteur correspondant subit un évanouissement. Dans le cas d'un récepteur et une technique d'estimation de canal (semi-aveugle) non-concordants, c'est l'ordre inférieur de la diversité qui domine. Différents cas d'estimation (semi-) aveugle de canal et de récepteurs correspondants sont examinés en détail.

Acknowledgements

The story goes back to around 4 years ago, when I got a master degree in mobile communications from Telecom Paris Tech university and decided to pursue my PhD thesis. At that moment, there was an incredibly smart, highly skilled and well reputed professor called Dirk Slock (my thesis supervisor) who gave me an infinite support and did what I thought impossible to be done, in order to overcome all the administrative difficulties that faced us at that time. Later, we spent together couple of years through which we have had a lot of fruitful discussions about many research ideas that have been translated into publications. Thanks Dirk and really I appreciate well your trust and great support. I would like to thank also my co-supervisor, professor Oussama Bazzi for his supervision and guidance during the past few years. I would like to thank also Eurécom for giving me the opportunity of accomplishing my thesis in excellent conditions.

I will always keep wonderful memories of these past years at Eurécom and my stay in France. I had the chance to meet great colleagues and collaborators, but mostly, great friends.

Finally, my parents, my wife, Fatima, my sons, Mohamad and Hadi who suffered a lot during my presence in France, far away from them. I feel lucky having your unlimited love and support throughout my whole life. As a matter of fact, although it is negligible compared to what you have offered to me, I dedicate my PhD to you.

Contents

List of Figures

Acronyms

Here, we list the main acronyms used in this thesis.

ADSL	Asymmetric Digital Subscriber Line
AGMAP	Alternating Gaussian Maximum A posteriori
AWGN	Additive White Gaussian Noise
BCRB	Bayesian Cramér-Rao Bound
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
CIR	Channel Impulse Response.
CP	Cyclic Prefix
CR	Cross Relations
CRB	Cramér-Rao Bound
CSIR	Channel State Information at the Receiver
DCRB	Deterministic Cramér-Rao Bound
DF	Decision Feedback
DFT	Discrete Fourier Transformation
DIQML	Denoised Iterative Quadratic Maximum Likelihood
DML	Deterministic Maximum Likelihood
DMT	Diversity Multiplexing Trade off
Elm	Elimination
EM	Expectation Maximization
EVD	Eigen Value Decomposition
FFT	Fast Fourier Transform
FIM	Fisher Information Matrix
FIR	Finite Impulse Response
FDD-DL	Frequency Domain Duplexing Down-Link
FOT	Fixing One Tap
GML	Gaussian Maximum Likelihood
HOS	Higher Order Statistics

IDFT	Inverse Discrete Fourier Transform
IFFT	Inverse Fast Fourier Transform
i.i.d.	independent and identically distributed
IQML	Iterative Quadratic Maximum Likelihood
Im	Imaginary
ISI	Inter Symbol Interference
LAN	Local Area Network
LE	Linear Equalizer
Lin	Linear
LMMSE	Linear Minimum Mean Squared Error
LS	Least Squares
LTE	Long Term Evolution
MAI	Multiple Access Interference
MAP	Maximum A Posteriori
MFB	Matched Filter Bound
MI	Mutual Information
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
MRC	Maximum Ratio Combining
MSE	Mean Square Error
NMSE	Normalized Mean Squared Error
NSF	Noise Subspace Fitting
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OWSSF	Optimally Weighted Signal Subspace Fitting
pdf	probability density function
PD	Positive Definite
PDP	Power Delay Profile
PQML	Pseudo Quadratic Maximum Likelihood
Pr(O)	Probability Of Outage
PSD	Positive Semi-Definite
QPSK	Quadratic Phase Shift Keying
Re	Real
r.h.s.	right hand side
RMSE	Root Mean Squared Error
Rx	Receiver
SB	Semi Blind
SC	Single Carrier

SCCP	Single Carrier Cyclic Prefix
SCM	Sample Covariance Matrix
SER	Symbol Error Rate
SIMO	Single Input Multiple Output
SNR	Signal-to-Noise Ratio
SIMO	Single Input Multiple Output
SINR	Signal to Interference Noise Ratio
SISO	Single-Input Single-Output
SOS	Second Order Statistics
SRM	Subchannel Response Matching
SSF	Signal Subspace Fitting
SVD	Singular Value Decomposition
Tx	Transmitter
UVVB	Uniform Variational Bayesian Blind
UWSSF	Unweighted Signal Subspace Fitting
VB	Variational Bayesian
VB	Variational Bayesian Blind
w.l.o.g	Without Loss oF Generality
w.r.t.	with respect to
WSSF	Weighted Signal Subspace Fitting
xDSL	x Digital Subscriber Line
ZF	Zero Forcing
ZP	Zero Padding

Notation

$\mathrm{tr}\{\cdot\}$	Trace of the matrix in brackets
$\det\{\cdot\}$	Determinant of the matrix in brackets
$\lvert a\rvert$	Absolute value of a
$\lVert\mathbf{a}\rVert$	Euclidean norm of vector $\mathbf{a}$
$\lVert\mathbf{A}\rVert_F$	Frobenius norm of matrix $\mathbf{A}$
$\hat{(\cdot)}$	An estimate of the quantity in parentheses
$\tilde{(\cdot)}$	The error in the estimate of the quantity in parentheses
$\mathbf{A}^*$	The complex conjugate of matrix $\mathbf{A}$
$\mathbf{A}^H$	The complex conjugate transpose (Hermitian) of matrix $\mathbf{A}$
$\mathbf{A}^T$	The transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	The inverse of matrix $\mathbf{A}$
$\mathbf{A} \geq \mathbf{B}$	means that $\mathbf{A} - \mathbf{B}$ is non-negative definite
$f(\cdot)$	pdf of the continuous random variable in parenthesis
$\Pr(\cdot)$	Probability of the event in parenthesis
$\mathrm{E}\{\cdot\}$	Expected value of the random variable in brackets
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	Circularly symmetric complex Gaussian random vector of mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$
$\mathcal{U}$	Uniform distribution
$\max, \min$	Maximum and minimum
$\sim$	Distributed according to
$\odot$	Hadamard Product
$\otimes$	Kronecker Product

Chapter 1

Introduction

The ever-growing demands of high-speed and high-quality wireless communication services have stimulated research on further promotion of the related technologies, including digital signal processing, antenna, and semiconductor. High-speed wireless communication systems typically require a much larger radio spectrum that may suffer from severe ISI due to the frequency-selective characteristics of the radio channels (including the effects of multipath propagation and limited channel bandwidth). Also, they may suffer from MAI when multiple users share the common radio resources. Mitigation of both the ISI and MAI is accordingly essential to the various types of systems. Familiar examples are single-carrier modulation systems with time-domain equalization, multicarrier modulation systems with frequency-domain equalization, and spread spectrum systems with RAKE reception [1].

1.1 Training Sequence Based Channel Estimation

Conventional designs of receivers that mitigate such distortions require either the knowledge of the channel or the access to the input so that certain training signals can be transmitted. The latter is the case in many, if not most, communication systems design. The transmission of training signals obviously decreases communications throughput although, for time invariant channels, the loss is insignificant because only one training is necessary.

For time varying channels, however, the loss of throughput becomes an issue. For example, in high-frequency (HF) communications, the time used to transmit training signals can be as much as 50 percent of the overall transmission.

1.2 Blind Channel Estimation

The alternative for transmitting training sequence is to depend solely on the received signal to estimate the channel. This is called blind channel estimation. The word blind stems from the fact that there is no cooperation between the transmitter and the receiver since the former does not aid the latter at all to accomplish the channel estimation. At first glance, this estimation problem may not seem tractable. How is it possible to distinguish the signal from the channel when neither is known?

1.2.1 SOS

The essence of blind channel estimation rests on the exploitation of the structure of the channel and the properties of the input. A familiar case is when the input has known probabilistic description, such as distributions and moments. In such a case, the problem of estimating the channel using the output statistics is related to time series analysis. In communications applications, for example, the input signals may have the finite alphabet property, or sometimes exhibit cyclostationarity. This last property was exploited in [2] to demonstrate the possibility of estimating a nonminimum phase channel using only the second-order statistics (SOS), which led to the development of many subspace-based blind channel estimation algorithms. The most popular second-order statistics based estimation techniques suffer from a lack of robustness: channel must satisfy diversity conditions and many blind SOS methods fail when the channel length is overestimated. Furthermore, the blind techniques leave an indeterminacy in the channel or the symbols, a scale or constant phase or a discrete phase factor. This suggests that SOS blind techniques should not be used alone but with some form of additional information. However, the same may also be true for TS based methods, especially when the training sequence is too short for a certain channel length. For an overview of the SOS based methods see [3], [4] and [5].

1.2.2 HOS

Apart from algorithms based on the SOS, there is a family of HOS-based algorithms for blind equalization of single- input, single-output (SISO) channels (i.e., single user case). These statistics, known as cumulants, and their associated Fourier transforms, known as polyspectra, not only reveal amplitude information about a process, but also reveal phase information. This is important, because, as it is well known, second-order statistics (i.e., correlation) are phase blind. Cumulants, on the other hand, are blind to any kind of a Gaussian process, whereas correlation is not; hence, cumulant-based signal processing methods handle colored Gaussian measurement noise automatically, whereas correlation-based methods do not. Consequently, cumulant-based methods boost signal-to-noise ratio when signals are corrupted by Gaussian measurement noise. Higher-order statistics are applicable when we are dealing with non-Gaussian (or, possibly, nonlinear) processes, and, many real world applications are truly non-Gaussian. In the past, due to the lack of analytical tools, we have been forced to treat such applications as though they were Gaussian. With the new results that are being developed, it should be possible to reexamine every application and/or method that has ever made use of second-order statistics, using higher-order statistics, to see if better results can be obtained.

These algorithms can be divided into two classes: implicit and explicit methods [6]. The former, using higher-order moments implicitly, is also known as the Bussgang-type algorithms, which includes the Sato algorithm [7] and the constant modulus algorithm (CMA) (or Godard-2 algorithm) [8] as special cases. In digital communications, the CMA has been a widely applied approach to alleviating the ISI effect induced by telephone, cable, or radio channels [9]. Moreover, its counterparts for blind equalization of multipleinput, multiple-output (MIMO) channels (i.e., multiuser case) have been applied to multiuser detection in DS/CDMA systems, blind beamforming, and source separation in multiple-antenna systems [10]. On the other hand, the explicit SISO methods using HOS include the IFC-based algorithm [11], the super-exponential algorithm (SEA) [12] and polyspectra-based algorithms [13]. The IFC-based algorithm and SEA are suitable for seismic exploration as well as blind equalization of communication channels, and their MIMO counterparts have been also applied in multiuser detection, blind beamforming, and source separation [14], [15], see also [16] and the references therein.

The biggest drawback to-date to the use of polyspectral methods (HOS) is that they require longer data lengths than do correlation-based methods.

Longer data lengths are needed in order to reduce the variance associated with estimating the higher-order statistics from real data using sample averaging techniques. For an overview of HOS please refer to [16] and [17] and the references there in.

1.3 Semi-Blind Channel Estimation

The remedy for the symptoms of illness revealed in blind or training based channel estimation may lie in the semi-blind channel estimation. Hence, the natural question that may raise here, what does semi-blind mean? In order to answer this question, we commence with the definition of the non-blind (training based) and blind channel estimation. Training sequence methods base the parameter estimation only on the received signal containing known symbols and all the other observations, containing (some) unknown symbols, are ignored. On the other hand, blind methods are based on the whole received signal, containing known and unknown symbols, possibly using hypotheses on the statistics of the input symbols, like the fact that they are i.i.d. for example, but no use is made of the knowledge of some input symbols. Now, the purpose of semi–blind methods is to combine both training sequence and blind information (see figure 1.1) and exploit the positive aspects of both techniques. Semi-blind techniques, because they incorporate the information of known symbols, avoid the possible pitfalls of blind methods and with only a few known symbols, any channel, single or multiple, becomes identifiable. Furthermore, exploiting the blind information in addition to the known symbols, allows one to estimate longer channel impulse responses than possible with a certain training sequence length, a feature that is of interest for the application of mobile communications in mountainous areas. For methods based on the second-order moments of the data, one known symbol is sufficient to make any channel identifiable. In addition, semi-blind techniques allow one to use shorter training sequences for a given channel length and desired estimation quality, compared to a training approach. Apart from these robustness considerations, semi-blind techniques appear also very attractive from a performance point of view, as their performance is superior to that of training sequence or blind techniques separately. Semi-blind techniques are particularly promising when TS and blind methods fail separately: the combination of both can be successful in such cases. For an overview of semi-blind channel estimation and the related issues, please refer to [18].

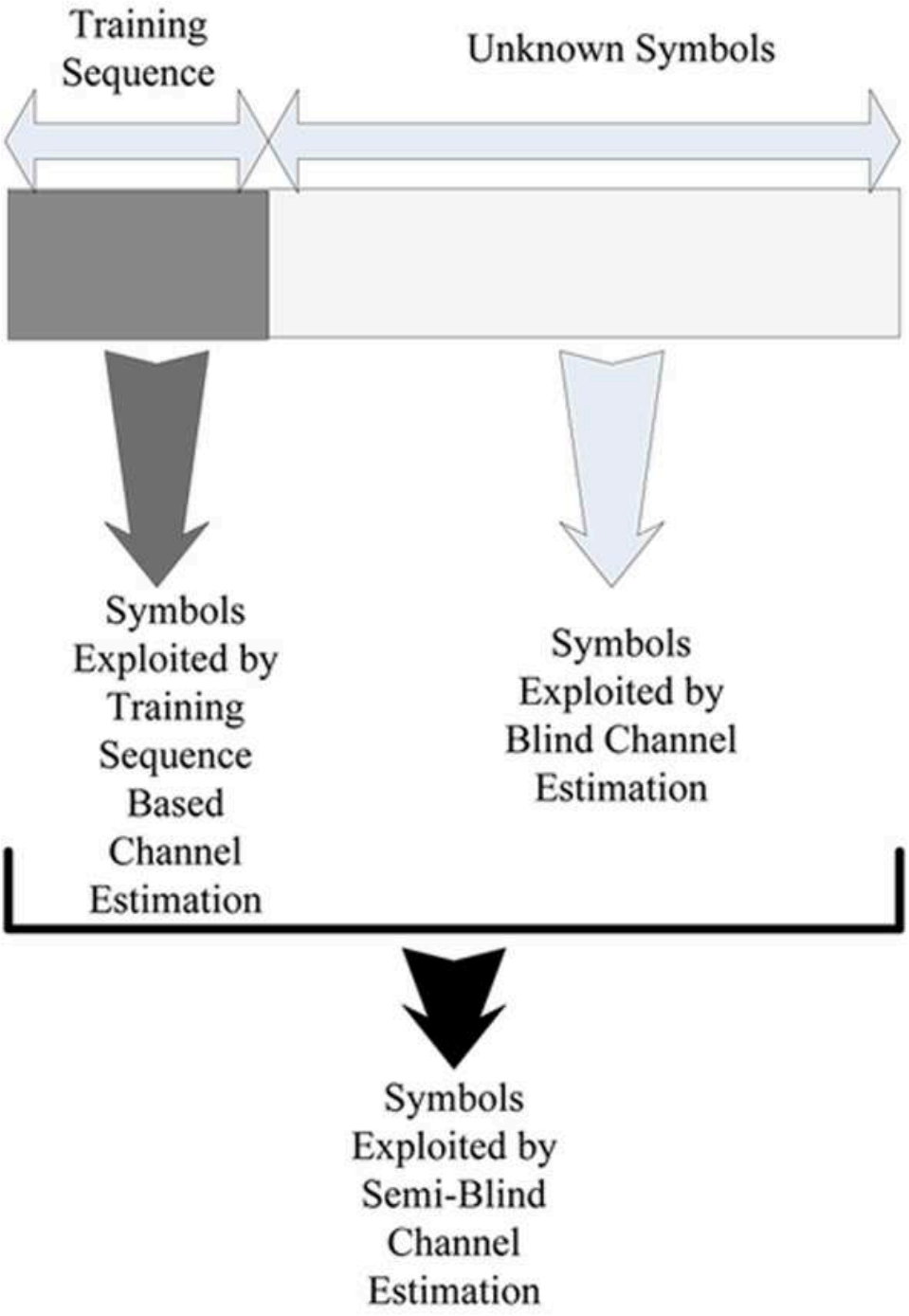

Figure 1.1: Semi-Blind Principle

1.4 The Multichannel Model

We consider here linear modulation (nonlinear modulations such as GMSK can be linearized with good approximation [19], [20]) over a linear channel with additive noise. The received signal after a linear receiver filter is then the convolution of the transmitted symbols with an overall channel impulse response, which is itself the convolution of the transmit shaping filter, the propagation channel and the receiver filter. The communication system is presented in figure 1.2. The overall channel impulse response is modeled as FIR which for multipath propagation in mobile communications appears to be well justified. In mobile communications terminology, this thesis will mostly consider the single-user case; some work has also been done for the multi-user case in which the received signal contains a mixture of multiple users. We describe the FIR multichannel model used throughout the thesis. This multichannel model applies to different cases (see figure 1.3): oversampling w.r.t. the symbol rate of a single received signal [21], [22], [23] or the separation into the real (inphase) and imaginary (quadrature) component of the demodulated received signal if the symbol constellation is real [24], [25]. In the context of mobile digital communications, a third possibility appears in the form of multiple received signals from an array of sensors (figure 1.3, 3rd part). These three sources for multiple channels can also be combined.

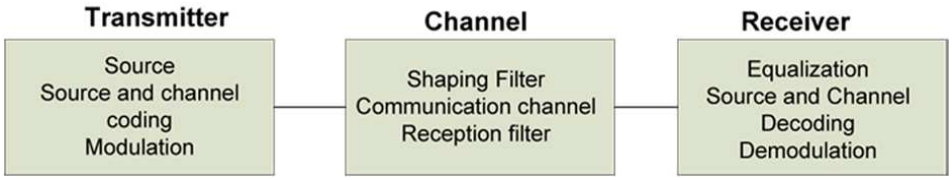

Figure 1.2: Communication System

1.5 Applications of Blind and Semi-blind Signal Processing

It is worth noting that the principle of blind and semi-blind is not limited to the channel estimation field, it has been exploited in many other research areas, among which we will mention the following:

- Wired, Wireless and optical Communications.

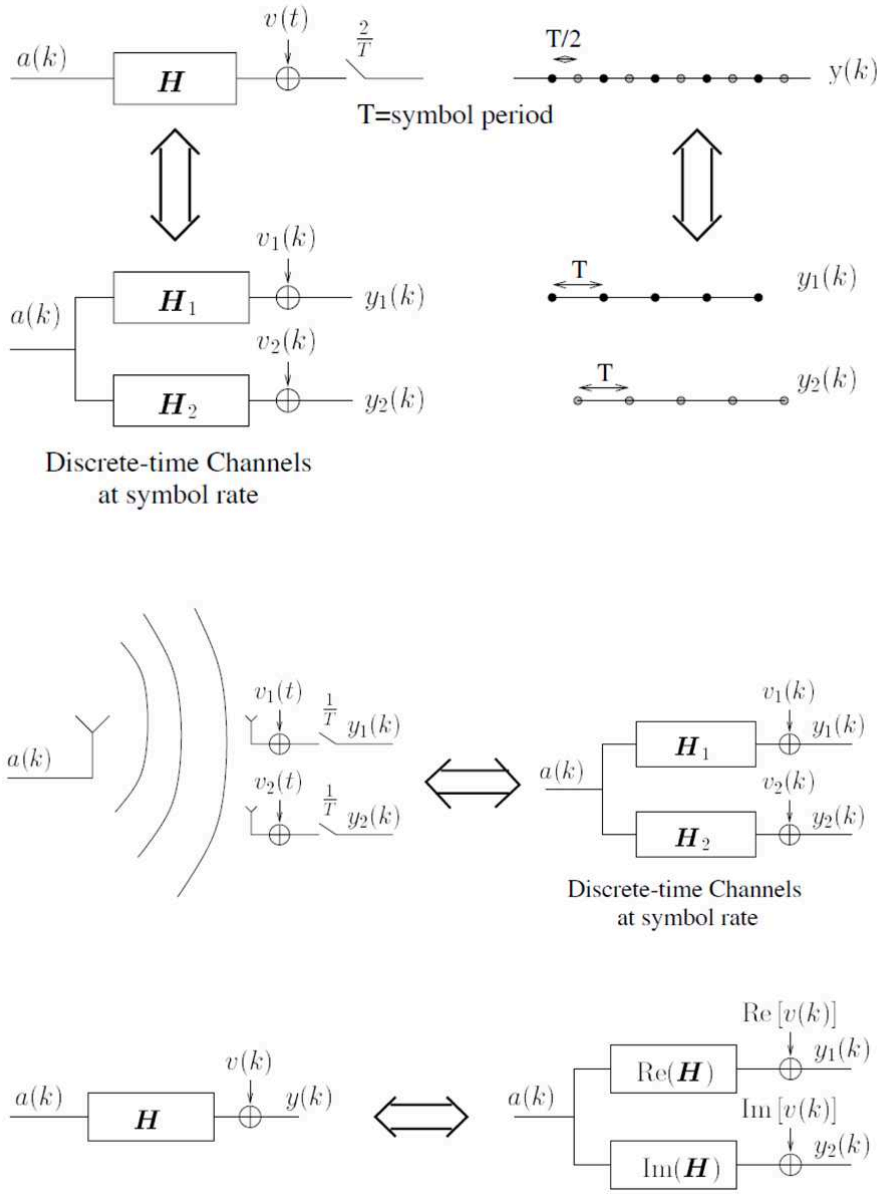

Figure 1.3: Multichannel model: case of oversampling, multiple antennas and separation of inphase and quadrature components when the input symbols are real. Example of a multichannel with 2 subchannels.

- Biomedical (EEG, MEG, ECG).
- Acoustic applications (Sonar, Cocktail-party problem).
- In computer networks where links between terminal and central computers need to be established in an asynchronous way such that, in some instances, training is impossible [26].
- High-Definition Television (HDTV) broadcasting.
- Geophysical data processing.
- Image deblurring applications.

As for the main notations related to the channel estimation problem used through this thesis, we shall present them below:

M	Output Burst Length
M_U	Number of Unknown Symbols
M_K	Number of Known Symbols
L	Channel Length
p	Number of Receiving Antennas
q	Number of Transmitting Antennas
$\boldsymbol{h}$	Variable Channel Vector
$\boldsymbol{h}^o$	True Channel Vector
$\mathcal{T}(h)$	Convolution Channel Matrix in Time Domain
$\mathcal{H}$	Channel Matrix in Frequency Domain
A	Input Symbol Vector in Time Domain
A_K	Vector of Known Symbols in Time Domain
A_U	Vector of Unknown Symbols in Time Domain
X	Input Symbol Vector in Frequency Domain
V	Noise Vector at the Receiver in Time Domain
W	Noise Vector at the Receiver in Frequency Domain
Y	Received Signal Vector in Time Domain
U	Received Signal Vector in Frequency Domain

Table 1.1: Summary of main notations used through this thesis

1.6 Organization of the Thesis

This dissertation is composed of an introduction chapter, three parts and a general conclusion. In the first part, comprising Chapter 2 through 4, we

treat the blind channel estimation in the context of SIMO and MIMO cyclic prefix (CP) systems. In the second part, comprising Chapters 5 through 8, we treat blind and semi-blind channel estimation in the context of SIMO FIR non-CP systems. Finally, in the third part we treat the diversity aspect of blind and semi-blind channel estimation in the context of both CP and non-CP systems. An abstract and an introduction are provided at the beginning of each chapter and a conclusion is drawn at the end of each one. Finally, a general conclusion and some perspectives for future work are presented at the end of this dissertation.

1.6.1 SIMO and MIMO Cyclic Prefix Systems

In chapter 2, which is based on [27], we provide an analytical performance analysis of a weighted and unweighted Signal Subspace Fitting (SSF) algorithm. We also make use of the performance analysis provided in [28] for Denoised Iterative Quadratic Maximum Likelihood (DIQML) and Pseudo Quadratic Maximum Likelihood (PQML) for SIMO FIR channel to derive simple formulas for the SIMO-CP case.

In chapter 3, which is based on [29], see also [30], inspired by the iterative sample covariance matrix (SCM) structure enhancement techniques of Cadzow and others [31], we develop an algorithm to structure the sample block circulant covariance matrix by enforcing two essential properties: rank and FIR structure. The novel enhancement procedure leads to an interesting enhanced SCM, even for the single CP symbol case.

In chapter 4, which is based on [32], we explore a Bayesian approach to (semi-)blind channel estimation, exploiting a priori information on fading channels. In the case of deterministic unknown input symbols, it suffices to augment the classical blind (quadratic) channel criterion with a quadratic criterion reflecting the Rayleigh fading prior. In the case of a Gaussian symbol model the blind criterion is more involved. The joint ML/MAP estimation of channels, deterministic unknown symbols, and channel profile parameters can be conveniently carried out using Variational Bayesian techniques. Variational Bayesian techniques correspond to alternating maximization of a likelihood w.r.t. subsets of parameters, but taking into account the estimation errors on the other parameters. To simplify exposition, we elaborate the details for the case of MIMO OFDM systems.

1.6.2 SIMO FIR Systems

Blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. On the other hand, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. In chapter 5 which is based on [33], we explore a Bayesian approach to blind channel estimation, exploiting a priori information on fading channels. We mainly focus on joint ML/MAP estimation of channels and symbols on one hand, and on ML/MAP estimation of channels with elimination of symbols on the other hand. As a consequence, a unified framework in addition to three new Bayesian estimators are introduced where their performance is compared by simulations to three existing non-Bayesian estimators. In the same context, we provide an insightful discussion of the accurate way of deriving the Bayesian Cramér-Rao bound (BCRB) with an emphasis on its singularity. In the same spirit, we extend in chapter 6, which is based on [34], the framework and the techniques introduced in chapter 5 into the semi-blind case. In chapter 7 which is based on [35], we present the CRBs that exist in the literature and fit to some of the algorithms discussed in chapter 6, and derive the others. In chapter 8 which is based on [36], we introduce in the context of semi-blind channel estimation a new approach that relies on the partial exploitation of the PDP of the channel (assumed known or estimated from the received data) to reduce the channel estimation error. Based on this approach, we have shown that, by neglecting some taps at the tail of the channel that are immersed in noise, the quality of the channel estimation has been improved considerably. The proposed approach has been implemented to a series of deterministic and Bayesian estimators introduced in chapter 6.

1.6.3 Receiver Diversity with Blind and Semi-blind channel Estimates

In chapter 9 which is based on [37] and [38], we analyze the diversity order of MMSE-ZF Linear and Decision-Feedback Equalization for frequency-selective SIMO channels, with the receivers being constructed from matching (semi-)blind channel estimates. The matching is furthermore interpreted here in a strict sense in which both the symbols and the channel get estimated on the basis of the same block of data.

Part I

SIMO and MIMO Cyclic Prefix Systems

Chapter 2

Performance Analysis Of Algorithms Based on SOS

Blind channel estimation problem was a hot topic during the previous decade. Many algorithms were proposed but only few of them were analytically analyzed. In this chapter, we deal with a Cycic Prefix (CP), Single Carrier (SC) or OFDM, SIMO transmission system. At the receiver there exists an FFT block which is responsible for transforming the received signal into the frequency domain. The advantage of CP system is its ability to transform the frequency selective channel into flat fading at each tone. In this chapter, we reintroduce a framework that exploits this fact to easily derive an analytical performance analysis of a category of blind channel estimation algorithms that are based on the second order statistics of the received signal. Based on that framework, we provide an analytical performance analysis of a weighted and unweighted Signal Subspace Fitting (SSF) algorithm by exploiting the analogy in the formulation of the blind channel estimation problem of the flat fading channel at each tone with the sensor array processing. We also make use of the performance analysis for Denoised Iterative Quadratic Maximum Likelihood (DIQML) and Pseudo Quadratic Maximum Likelihood (PQML) for SIMO FIR channel to derive simple formulas for the SIMO-CP case. We also propose an enhanced version of DIQML where the denoising process is performed on a tone basis and derive its performance analytically. On the other hand, we present a performance analysis for Subchannel Response

Matching (SRM) algorithm which is much simpler than the one that already exists in the literature. On the top of that the latter is valid at the whole SNR range while the former becomes cumbersome at low SNR. Finally, all the analytical formulas are verified by simulation.

2.1 Introduction

Initially, the proposed algorithms for blind channel estimation were based on the higher-order statistics of the received signal. The defect of such algorithms is its need to a long data lengths in order to reduce the variance associated with estimating the higher-order statistics from real data using sample averaging techniques. For an overview of algorithms based on higher-order statistics see [17], [16]. However, in most wireless communication systems the channel is varying rapidly so it is very difficult to gather a long data through which the channel doesn't change. In 1991 Tong et al, proposed a new algorithm for channel estimation based solely on the second order statistics of the received signal. Since then, a complete new branch has been established and many new algorithms were introduced. Two types of techniques can be considered, treating the unknown input symbols as either deterministic unknowns or Gaussian white noise. In the first case, the techniques are often based on the subspace structure induced in the data by the multichannel aspect. Many of the deterministic input approaches are also quite sensitive to a number of hypotheses such as correct channel length (filter order) and no channel zeros. In general this means that these blind channel estimates can often become ill-conditioned, when the channel impulse response is tapered (e.g. due to a pulse shape filter) or when the channel is close to having zeros. In fact this means that the blind information on the channel can be substantial, but is only limited to part of the channel. An overview of blind channel estimation techniques can be found in [18] for SIMO systems and in [5] for MIMO systems, see also [3]. Many of the algorithms mentioned above were evaluated by running computer simulations and few were assessed by analytical analysis. In this chapter, we exploit the simple structure of the flat fading channel within the framework of the SIMO cyclic prefix system to derive an analytical performance analysis of a number of algorithms which are based on the 2nd order statistics. It is worthy to note that we proposed this framework initially in [39] see also [29]. The derivations for some algorithms (Signal Subspace Fitting) are made by analogy with sensor array processing presented in [40] while for the rest they are concluded from those of FIR channel derived in [28]. This

chapter is organized as follows: In section 2 we introduce the SIMO CP block TX system, in section 3 we show a generic form of the cost function which is valid for different algorithms presented in this chapter in addition, we introduce the concept of linear parameterization of the noise subspace upon which all the algorithms presented in this chapter are based. In section 4 we develop different versions of SSF algorithm then in section 5 we develop DML algorithm with its derivatives namely, IQML, DIQML, PQML in addition to SRM. In section 6 we propose a framework for performance analysis of SIMO CP systems then we use that framework to derive analytically the performance of all the algorithms discussed in this chapter. Finally, in section 7 we validate all the analytical formulas by simulations.

2.2 SIMO Cyclic Prefix Block TX Systems

Consider a SIMO system with p outputs:

$$\begin{aligned} \underbrace{\mathbf{y}[m]}_{p\times 1} &= \sum_{j=0}^{L-1} \underbrace{\mathbf{h}[j]}_{p\times 1}\,\underbrace{a[m-j]}_{1\times 1} + \underbrace{\mathbf{v}[m]}_{p\times 1} \\ &= \underbrace{H(q)}_{p\times 1}\,\underbrace{a[m]}_{1\times 1} + \underbrace{\mathbf{v}[m]}_{p\times 1} \end{aligned} \tag{2.1}$$

where $H(q) = \sum_{j=0}^{L-1} \mathbf{h}[j]\, q^{-j}$ is the SIMO system transfer function corresponding to the z transform of the impulse response $\mathbf{h}[.]$, $a[.]$ denotes the transmitted symbol and $\mathbf{v}[.]$ denotes a white Gaussian noise vector. Equation (2.1) mixes time domain and z transform domain notations to obtain a compact representation. In $H(q)$, z is replaced by q to emphasize its function as an elementary time advance operator over one sample period. Its inverse corresponds to a delay over one sample period: $q^{-1}\mathbf{a}[m] = \mathbf{a}[m-1]$.

Consider a (OFDM or single-carrier) CP block transmission system with N samples per block. The introduction of a cyclic prefix of K samples means that the last K samples of the current block (corresponding to N samples) are repeated before the actual block. If we assume w.l.o.g. that the current block starts at time 0, then samples $\mathbf{a}[N-K]\cdots\mathbf{a}[N-1]$ are repeated at time instants $-K, \ldots, -1$. This means that the output at sample periods $0, \ldots, N-1$ can be written in matrix form as

$$\begin{bmatrix} \mathbf{y}[0] \\ \vdots \\ \mathbf{y}[N-1] \end{bmatrix} = \mathrm{Y}[0] = \mathrm{H}\,\mathrm{A}[0] + \mathrm{V}[0] \tag{2.2}$$

where the matrix H is not only (block) Toeplitz but even (block) circulant: each row is obtained by a cyclic shift to the right of the previous row. Consider now applying an N-point FFT to both sides of (2.2) at block m:

$$F_{N,p}\mathrm{Y}[m] = F_{N,p}H\,F_N^{-1}\quad F_N\mathrm{A}[m] + F_{N,p}\mathrm{V}[m] \tag{2.3}$$

or with new notations:

$$\mathrm{U}[m] \quad = \quad \mathcal{H} \quad \mathrm{X}[m] \quad + \quad \mathrm{W}[m] \tag{2.4}$$

where $F_{N,p} = F_N \otimes I_p$ (Kronecker product: $A \otimes B = [a_{ij}B]$), F_N is the N-point $N \times N$ DFT matrix, $\mathcal{H} = \mathrm{diag}\,\{\mathbf{h}_0, \ldots, \mathbf{h}_{N-1}\}$ is a block diagonal matrix with diagonal blocks $\mathbf{h}_n = \sum_{l=0}^{L-1} \mathbf{h}[l]\; e^{-j2\pi\frac{1}{N}nl}$, the $p \times 1$ channel transfer function at tone n (frequency $= n/N$ times the sample frequency). In OFDM, the transmitted symbols are in $\mathrm{X}[m]$ and hence are in the frequency domain. The corresponding time domain samples are in $\mathrm{A}[m]$. The OFDM symbol period index is m. In Single-Carrier (SC) CP systems, the transmitted symbols are in $\mathrm{A}[m]$ and hence are in the time domain. The corresponding frequency domain data are in $\mathrm{X}[m]$. The components of V are considered white noise, hence the components of W are white also. At tone (subcarrier) $n \in \{0, \ldots, N-1\}$ we get the following input-output relation

$$\underbrace{\mathbf{u}_n[m]}_{p\times 1} = \underbrace{\mathbf{h}_n}_{p\times 1}\;\underbrace{x_n[m]}_{1\times 1} + \underbrace{\mathbf{w}_n[m]}_{p\times 1} \tag{2.5}$$

where the symbol $x_n[m]$ belongs to some finite alphabet (constellation) in the case of OFDM. It should be noted that in the sequel both k and n are used interchangeably to denote the frequency index (tone), while m is used to denote the time index.

2.3 Some Generalities for CP System Methods

2.3.1 General form of the cost function

The basic idea relies on the fact that to get the cost function or information for the temporal channel response it suffices to sum up the cost functions or information over the tones after transforming back to the time domain. To be a bit more explicit, let $\mathbf{h}$ be the vectorized channel impulse response then there exists transformation matrices G_k (containing DFT portions)such that

$$\mathbf{h}_k \;=\; G_k\,\mathbf{h}\,. \tag{2.6}$$

To be more accurate, G_k is of size $p \times pL$ such that it contains the first pL elements of the kth block row of $F_{N,p}$. Now, if at tone k we have a cost function of the form:

$$\mathcal{F}(\mathbf{h}_k) = \mathbf{h}_k^H \, Q_k \, \mathbf{h}_k \tag{2.7}$$

then this induces a cost function for the overall channel impulse response of the form:

$$\mathcal{F}(\mathbf{h}) = \mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \, Q_k \, G_k \right] \mathbf{h} \tag{2.8}$$

and similarly for Fisher information matrices. So in what follows, we shall concentrate on the cost function for a given tone.

2.3.2 Linear Parameterization of the Noise Subspace

As we shall see later in (2.15 and 2.20), the Deterministic Maximum Likelihood (DML) and the Signal Subspace Fitting (SSF) criteria are highly nonlinear and their direct optimization would require cumbersome optimization techniques. The key to a computationally attractive solution of these problems is a linear parameterization of the noise subspace. We consider here a linear parameterization of the noise subspace in terms of channel coefficients (a parameterization in terms of prediction quantities was also presented in [41]). Let $\mathbf{h}_k^{\perp}(m \times p), (m \geq p-1)$, be such a parametrization: it verifies $\mathbf{h}_k^{\perp}\mathbf{h}_k = 0$ and consequently $P_{\mathbf{h}_k}^{\perp} = P_{\mathbf{h}_k^{\perp H}}$. In the case p = 2 in which the multichannel has 2 subchannels, the obvious choice for $\mathbf{h}_k^{\perp}$ is:

$$\mathbf{h}_k^{\perp} = [-\mathbf{h}_{k,2} \quad \mathbf{h}_{k,1}] \tag{2.9}$$

For a larger number of subchannels, different choices are available [42, 43]. An example is [43],[44]:

$$\mathbf{h}_{k,bal,min}^{\perp} = \begin{bmatrix} -\mathbf{h}_{k,2} & \mathbf{h}_{k,1} & 0 & \cdots & 0 \\ 0 & -\mathbf{h}_{k,3} & \mathbf{h}_{k,2} & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\mathbf{h}_{k,p} & \mathbf{h}_{k,p-1} \\ \mathbf{h}_{k,p} & 0 & \cdots & 0 & -\mathbf{h}_{k,1} \end{bmatrix} \tag{2.10}$$

which has size $(p-\delta_{p,2}) \times p$ where $\delta_{p,2}$ equals one when $p = 2$ and zero elsewhere. This matrix is balanced in the sense that every subchannel appears the same number of times, in this case twice. A balanced $\mathbf{h}_k^{\perp}$ leads to $\mathrm{tr}\left\{\mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}\right\} = \alpha \, ||\mathbf{h}_k||^2$ and $\alpha = 2-\delta_{p,2}$. $\mathbf{h}_k^{\perp}$ appearing in (2.10) is balanced with minimal number of rows $m = p-\delta_{p,2}$.

2.4 Signal Subspace Fitting (SSF)

Let us focus in particular on the signal subspace fitting method, for the (spatiotemporally) white noise case (and assuming spatiotemporally white symbols for simplicity), the eigendecomposition of the covariance matrix of a block of signal in the time domain can in fact be easily computed from the eigendecompositions at each tone. Indeed the covariance matrix is given by:

$$\begin{aligned} R_{\mathrm{YY}} &= \sigma_a^2\, \mathrm{H}\, \mathrm{H}^H + \sigma_v^2\, I_{Np} \\ \Rightarrow\ & F_{N,p} R_{\mathrm{YY}} F_{N,p}^{-1} \\ =\ & \sigma_a^2\, F_{N,p} \mathrm{H} F_N^{-1} F_N \mathrm{H}^H F_{N,p}^{-1} + \sigma_v^2\, F_{N,p} F_{N,p}^{-1} \\ =\ & \sigma_a^2\, \mathcal{H}\, \mathcal{H}^H + \sigma_v^2\, I_{Np} \end{aligned} \tag{2.11}$$

where the matrix $\mathcal{H}\,\mathcal{H}^H$ is block diagonal. Hence the eigenvectors in the time domain are the IDFTs of the eigenvectors at each tone, and the eigenvalues are the same in time or frequency domain. This exact relationship no longer holds for the eigenvectors based on sample covariances in time and frequency domain due to the noise (it remains true in the absence of noise). Nevertheless this relationship encourages us to develop subspace fitting problems in the frequency domain, involving eigendecompositions of N $p \times p$ matrices instead of the eigendecomposition of one $Np \times Np$ matrix. Let $\widehat{\mathrm{E}}$ denote a sample average, then the details of the signal subspace fitting method are:

- $\mathbf{r}_k = \mathrm{E}\, \mathbf{u}_k[n]\mathbf{u}_k^H[n] = \sigma_x^2\, \mathbf{h}_k \mathbf{h}_k^H + \sigma_{w_k}^2 I_p$
  $= V_{\mathcal{S},k} \Lambda_{\mathcal{S},k} V_{\mathcal{S},k}^H + \sigma_{w_k}^2 V_{\mathcal{N},k} V_{\mathcal{N},k}^H$
- $\widehat{\mathbf{r}}_k = \widehat{\mathrm{E}}\, \mathbf{u}_k \mathbf{u}_k^H = \widehat{V}_{\mathcal{S},k} \widehat{\Lambda}_{\mathcal{S},k} \widehat{V}_{\mathcal{S},k}^H + \widehat{V}_{\mathcal{N},k} \widehat{\Lambda}_{\mathcal{N},k} \widehat{V}_{\mathcal{N},k}^H$

Where $V_{\mathcal{S},k}$ and $V_{\mathcal{N},k}$ deonte respectively the signal and the noise subspace at tone k. The basic idea of the SSF is to fit the true signal subspace which is spanned by the column of $\mathbf{h}_k$ to the estimated signal subspace $\widehat{V}_{\mathcal{S},k}$ that we obtain from the sampled spectrum $\widehat{\mathbf{r}}_k$.

$$\min_{\mathbf{h}_k, T}\ ||\widehat{V}_{\mathcal{S},k} - \mathbf{h}_k T||_{\mathcal{W}_k} \tag{2.12}$$

where T is a square transformation matrix and $||X||_{\mathcal{W}} = tr\{X^H \mathcal{W} X\}$. The cost function in (2.12) is separable. In particular, it is quadratic in T. Minimization w.r.t. T leads to $T = \mathbf{h}_k^+ \widehat{V}_{\mathcal{S},k}$ and $\widehat{V}_{\mathcal{S},k} - \mathbf{h}_k \mathbf{h}_k^+ \widehat{V}_{\mathcal{S},k} = P_{\mathbf{h}_k}^{\perp} \widehat{V}_{\mathcal{S},k}$ where $P_{\mathbf{h}_k}^{\perp} = I - P_{\mathbf{h}_k}$.

Hence,

$$\begin{aligned}\min_{\mathbf{h}_k,T} ||\widehat{V}_{\mathcal{S},k} - \mathbf{h}_k T||_{\mathcal{W}_k} &= \min_{\mathbf{h}_k} ||P^{\perp}_{\mathbf{h}_k}\widehat{V}_{\mathcal{S},k}||_{\mathcal{W}_k} \\ &= \min_{\mathbf{h}_k} \operatorname{tr}\left\{P^{\perp}_{\mathbf{h}_k}\widehat{V}_{\mathcal{S},k}\mathcal{W}_k\widehat{V}^H_{\mathcal{S},k}\right\}\end{aligned} \tag{2.13}$$

As mentioned earlier, we attempt to minimize the SSF jointly over all the tones subject to a single constraint to avoid introducing N constraints and to exploit the correlation that exists between the different tones. Therefore, the SSF cost functions takes the following form:

$$\begin{aligned}&\min_{\mathbf{h}} \sum_{k=0}^{N-1} \operatorname{tr}\left\{P^{\perp}_{\mathbf{h}_k}\widehat{V}_{\mathcal{S},k}\mathcal{W}_k\widehat{V}^H_{\mathcal{S},k}\right\} \\ &\min_{\mathbf{h}} \sum_{k=0}^{N-1} \operatorname{tr}\left\{P_{\mathbf{h}_k^{\perp H}}\widehat{V}_{\mathcal{S},k}\mathcal{W}_k\widehat{V}^H_{\mathcal{S},k}\right\} \\ &\min_{\mathbf{h}} \sum_{k=0}^{N-1} \operatorname{tr}\left\{\left(\mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}\right)^{+}\mathbf{h}_k^{\perp}\widehat{V}_{\mathcal{S},k}\mathcal{W}_k\widehat{V}^H_{\mathcal{S},k}\mathbf{h}_k^{\perp H}\right\}\end{aligned} \tag{2.14}$$

Where we have used the result shown in the linear parameterization of the noise subspace section namely, $P^{\perp}_{\mathbf{h}_k} = P_{\mathbf{h}_k^{\perp H}}$. On the other hand, the Moore-Penrose pseudo–inverse needs to be introduced since $\mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}$ is singular for $p > 2$. Let $\mathcal{R}(\mathbf{h}_k) \stackrel{\triangle}{=} \mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}$ and denoting the constant denominator $\mathcal{R}(\mathbf{h}_k) = \mathcal{R}$, then (2.14) becomes:

$$\begin{aligned}&\min_{\mathbf{h}} \sum_{k=0}^{N-1} \operatorname{tr}\left\{\mathcal{R}^{+}\mathbf{h}_k^{\perp}\widehat{V}_{\mathcal{S},k}\mathcal{W}_k\widehat{V}^H_{\mathcal{S},k}\mathbf{h}_k^{\perp H}\right\} \\ &\min_{\mathbf{h}} \sum_{k=0}^{N-1} \mathbf{h}_k^H D_k^{SSF}\mathbf{h}_k \\ &\min_{\mathbf{h}} \mathbf{h}^H\left\{\sum_{k=0}^{N-1} G_k^H D_k^{SSF} G_k\right\}\mathbf{h}\end{aligned} \tag{2.15}$$

We solve (2.15) iteratively in such a way that at each iteration a quadratic problem appears. In iteration (i), the "denominator" $\mathcal{R}$ is computed based on the estimate from the previous iteration/initialization $\widehat{h}^{(i-1)}$ and is considered as constant for the current iteration. Moreover, $\mathbf{h}_k^{\perp}$ being linear in $\mathbf{h}_k$, the criterion (2.15) becomes quadratic. Since we are starting from a cost function per tone then it is not very costly to introduce optimal weighting ($\mathcal{W}_k$) as we shall see in the following sections.

2.5 Deterministic Approach

2.5.1 Deterministic Maximum Likelihood (DML)

The problem is to estimate the deterministic (vector) parameter θ given the probabilistic model of the observation. Specifically, let $f(\mathbf{u}_k/\theta)$ be the probability density function. Given an observation $\mathbf{u}_k$, θ is estimated by maximizing $\max_{\theta} f(\mathbf{u}_k/\theta)$ where $f(\mathbf{u}_k/\theta)$, when viewed as a function of θ is referred to as the likelihood function. When the noise is zero-mean Gaussian with covariance $\sigma^2_{w,k} I_p$, the ML estimates can be obtained by the nonlinear least squares optimization:

$$\min_{\mathbf{h}_k, \mathbf{x}_k} ||\mathbf{u}_k - \mathbf{h}_k \mathbf{x}_k||^2 \tag{2.16}$$

The joint optimization of the likelihood function in both the channel and the symbols is difficult. Fortunately, the observation is linear in both the channel and the symbols. In other words, we have a separable nonlinear LS problem, which allows us to reduce the complexity considerably. The nonlinear LS optimization can be achieved sequentially as:

$$\min_{\mathbf{h}_k} \left\{ \min_{\mathbf{x}_k} ||\mathbf{u}_k - \mathbf{h}_k \mathbf{x}_k||^2 \right\} \tag{2.17}$$

The inner minimization problem yields:

$$\widehat{\mathbf{x}}_k = \frac{1}{||\mathbf{h}_k^2||} \mathbf{h}_k^H \mathbf{u}_k \tag{2.18}$$

Substitute (2.18) in (2.17) we get:

$$\begin{aligned} &\min_{\mathbf{h}_k} \ \mathbf{u}_k^H P^{\perp}_{\mathbf{h}_k} \mathbf{u}_k \\ &\min_{\mathbf{h}_k} \ \mathrm{tr} \left\{ P^{\perp}_{\mathbf{h}_k} \mathbf{u}_k \mathbf{u}_k^H \right\} \\ &\min_{\mathbf{h}_k} \ \mathrm{tr} \left\{ P^{\perp}_{\mathbf{h}_k} \widehat{\mathbf{r}}_k \right\} \end{aligned} \tag{2.19}$$

However, we are interested in estimating the channel impulse response $\mathbf{h}$, hence we gather the cost functions over all the tones:

$$\min_{\mathbf{h}} \sum_{k=0}^{N-1} \mathrm{tr} \left\{ P^{\perp}_{\mathbf{h}_k} \widehat{\mathbf{r}}_k \right\} \tag{2.20}$$

2.5.2 Subchannel Response Matching (SRM):

The Subchannel Response Matching (SRM) algorithm, which was (re)invented four times in [45, 46, 47, 42], is based on a linear parameterization of the noise subspace in terms of the channel coefficients. Using the commutativity of convolution and the linearity of $\mathbf{h}_k^{\perp}$ in $\mathbf{h}_k$, we can write $\mathbf{h}_k^{\perp}\mathbf{u}_k$ as:

$$\mathbf{h}_k^{\perp}\mathbf{u}_k = \mathcal{U}_k\mathbf{h}_k \tag{2.21}$$

where $\mathcal{U}_k$ is a matrix filled with the elements of the observation vector $\mathbf{u}_k$.

In the noiseless case, $\mathbf{u}_k = \mathbf{h}_k\mathbf{x}_k$ and we have $\mathbf{h}_k^{\perp}\mathbf{u}_k = \mathcal{U}_k\mathbf{h}_k = 0$: from this relation, the channel can be uniquely determined up to a scale factor [47, 42], as the unique right singular vector of $\mathcal{U}_k$ corresponding to the singular value zero. When noise is present, $\mathcal{U}_k\mathbf{h}_k \neq 0$ and the SRM criterion is solved in the least–squares sense under the constraint $\|\mathbf{h}_k\| = 1$. Thus, the per tone cost function is given by:

$$\begin{aligned} &\min_{\|\mathbf{h}_k\|=1} \mathbf{h}_k^H\, \mathcal{U}_k^H\mathcal{U}_k\, \mathbf{h}_k \\ &\min_{\|\mathbf{h}_k\|=1} \operatorname{tr}\left\{\mathbf{h}_k^{\perp}\left\{\mathbf{u}_k\mathbf{u}_k^H\right\}\mathbf{h}_k^{\perp H}\right\} \\ &\min_{\|\mathbf{h}_k\|=1} \operatorname{tr}\left\{\mathbf{h}_k^{\perp}\widehat{\mathbf{r}}_k\mathbf{h}_k^{\perp H}\right\} \end{aligned} \tag{2.22}$$

Where by the law of large numbers, asymptotically $\mathbf{u}_k\mathbf{u}_k^H$ can be replaced by its expected value $\mathbf{r}_k$. Practically, $\mathbf{r}_k$ is not available so it is replaced by the sampled spectrum per each tone $\widehat{\mathbf{r}}_k$ which is computed directly from the Fourier transformed version of the received data.

The overall cost function is given by:

$$\min_{\mathbf{h}}\ \mathbf{h}^H\left(\sum_k G_k^H\, \mathrm{E}\left\{\mathcal{U}_k^H\mathcal{U}_k\right\} G_k\right)\mathbf{h} \tag{2.23}$$

The solution is $\mathbf{h} = V_{min}(\sum_k G_k^H\, \mathrm{E}\left\{\mathcal{U}_k^H\mathcal{U}_k\right\} G_k)$, the eigenvector of $\sum_k G_k^H\, \mathrm{E}\ \left\{\mathcal{U}_k^H\mathcal{U}_k\right\} G_k$ corresponding to its smallest eigenvalue. Different choices for the linear parameterization of the noise subspace give different channel estimates. Note that $\mathrm{E}\,\mathbf{h}^H\left(\sum_k G_k^H\left\{\mathcal{U}_k^H\mathcal{U}_k\right\} G_k\right)\mathbf{h} = \left(\sum_k |\mathbf{x}_k|^2\mathbf{h}_k^H\mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp}\mathbf{h}_k\right) + \sigma_w^2 \operatorname{tr}\left\{\sum_k \mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}\right\}$. Since $\operatorname{tr}\left\{\sum_k \mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}\right\}$ is constant, hence it doesn't affect the minimization problem. Therefore, a balanced $\mathbf{h}_k^{\perp}$ yields asymptotically unbiased and consistent channel estimates whereas unbalanced $\mathbf{h}_k^{\perp}$ yields biased and inconsistent estimates. SRM may be viewed as a non-weighted version of the Iterative Quadratic ML (IQML) algorithm described below, and was used in [42] to initialize IQML. We will use it to initialize our algorithms also.

2.5.3 Iterative Quadratic ML (IQML):

Since $P^{\perp}_{\mathbf{h}_k} = P_{\mathbf{h}_k^{\perp H}}$, the DML problem (2.20) can be written as:

$$\min_{\mathbf{h}:\,\|\mathbf{h}\|=1} \sum_k \mathbf{u}_k^H \mathbf{h}_k^{\perp H} \left(\mathbf{h}_k^{\perp} \mathbf{h}_k^{\perp H}\right)^{+} \mathbf{h}_k^{\perp} \mathbf{u}_k \tag{2.24}$$

where again the Moore-Penrose pseudo–inverse needs to be introduced since $\mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}$ is singular for $p > 2$. As discussed in SSF case, the Iterative Quadratic ML (IQML) algorithm [48] solves (2.24) iteratively in such a way that at each iteration a quadratic problem appears. Let $\mathcal{R}(\mathbf{h}_k) \stackrel{\triangle}{=} \mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}$, then (2.24) becomes:

$$\min_{\mathbf{h}:\,\|\mathbf{h}\|=1} \sum_k \mathbf{u}_k^H \mathbf{h}_k^{\perp H}\, \mathcal{R}^{+}(\mathbf{h}_k)\, \mathbf{h}_k^{\perp} \mathbf{u}_k\ . \tag{2.25}$$

In iteration (i) of IQML, the "denominator" $\mathcal{R}(\mathbf{h}_k)$ is computed based on the estimate from the previous iteration/initialization $\widehat{h}^{(i-1)}$ and is considered as constant for the current iteration. Hence, $\mathbf{h}_k^{\perp}$ being linear in $\mathbf{h}_k$, the criterion (2.25) becomes quadratic. Again, denoting the constant denominator $\mathcal{R}(\mathbf{h}_k) = \mathcal{R}$, the IQML criterion can be rewritten as:

$$\min_{\mathbf{h}:\,\|\mathbf{h}\|=1} \mathbf{h}^H \sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^{+} \mathcal{U}_k\, G_k\ \mathbf{h}. \tag{2.26}$$

Under constraint $\|\mathbf{h}\| = 1$, we get $\mathbf{h} = V_{min}(\sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^{+} \mathcal{U}_k\, G_k)$.

In an alternative interpretation, IQML can be viewed as the optimally weighted least-squares version of the SRM least-squares problem: the covariance matrix of the noise contribution in $\mathbf{h}_k^{\perp}\mathbf{u}_k = \mathcal{U}_k \mathbf{h}_k$ is indeed $\sigma_w^2\, \mathcal{R}(\mathbf{h}_k)$ and we use for $\mathbf{h}_k$ in $\mathcal{R}(\mathbf{h}_k)$ the best estimate available.

In the noise-free case, the IQML algorithm behaves very well: the IQML criterion becomes indeed equivalent to:

$$\min_{\mathbf{h}:\,\|\mathbf{h}\|=1} \sum_k |\mathbf{x}_k|^2 \mathbf{h}_k^H \mathbf{h}_k^{\perp H} \mathcal{R}^{+} \mathbf{h}_k^{\perp} \mathbf{h}_k \tag{2.27}$$

As $\mathbf{h}_k^{\perp}\mathbf{h}_k = 0$, $\mathbf{h}^o$ nulls exactly the criterion, regardless of the initialization. At high SNR, a first iteration of IQML gives a consistent estimate of the channel whatever the initialization of $\mathcal{R}(\mathbf{h}_k)$ And it can be proved [42] that a second iteration gives the exact DML estimate.

At low SNR however, the IQML estimate is biased. Indeed, consider the asymptotic situation in which the number of Rx blocks (OFDM or SCCP) M

grows to infinity. By the law of large numbers, the IQML criterion becomes essentially equivalent to its expected value,viz.

$$\begin{aligned}&\tfrac{1}{M}\ \textstyle\sum_k \mathbf{u}_k^H \mathbf{h}_k^{\perp H}\ \mathcal{R}^+\ \mathbf{h}_k^{\perp}\mathbf{u}_k\ =\ \tfrac{1}{M}\,\mathrm{tr}\{\textstyle\sum_k \mathbf{h}_k^{\perp H}\mathcal{R}^+\mathbf{h}_k^{\perp}\,\mathrm{E}(\mathbf{u}_k\mathbf{u}_k^H)\} + \mathcal{O}_p(\tfrac{1}{\sqrt{M}})\\ &=\ \tfrac{1}{M}\ \big[\textstyle\sum_k \mathrm{tr}\{|\mathbf{x}_k|^2\mathbf{h}_k^{\perp H}\mathcal{R}^+\mathbf{h}_k^{\perp}\mathbf{h}_k\mathbf{h}_k^H\} + \sigma_w^2\ \textstyle\sum_k \mathrm{tr}\{\mathbf{h}_k^{\perp H}\mathcal{R}^+\mathbf{h}_k^{\perp}\}\big] + \mathcal{O}_p(\tfrac{1}{\sqrt{M}})\end{aligned} \tag{2.28}$$

The expectation operator E works along the time axis represented by the Rx (OFDM or SCCP) blocks. We conclude from (2.28)that in the presence of noise IQML performs poorly even if initialized by a consistent channel estimate.

In the next section we will show how to "denoise" the IQML criterion: this denoised criterion, solved in the IQML style, will correct the IQML bias and provide a consistent channel estimate.

2.5.4 Denoised IQML (DIQML):

2.5.4.1 Asymptotic Amount of Data

The asymptotic noise contribution to the DML criterion is $\sum_k \sigma_{w,k}^2 \mathrm{tr}\left\{P_{\mathbf{h}_k^{\perp H}}\right\}$ (see (2.28)). The denoising strategy simply consists in removing this asymptotic noise term, or more exactly an estimate of it, $\sum_k \widehat{\sigma}_{w,k}^2 \mathrm{tr}\left\{P_{\mathbf{h}_k^{\perp H}}\right\}$, from the DML criterion which becomes:

$$\begin{aligned}&\min_{\|\mathbf{h}\|=1} \sum_k \mathrm{tr}\left\{P_{\mathbf{h}_k^{\perp H}}\left(\mathbf{u}_k\mathbf{u}_k^H - \widehat{\sigma}_{w,k}^2\, I\right)\right\}\ \Leftrightarrow\\ &\min_{\|\mathbf{h}\|=1}\left\{\sum_k \mathbf{h}_k^H\mathcal{U}_k^H\mathcal{R}^+\mathcal{U}_k\mathbf{h}_k\ -\sum_k \widehat{\sigma}_{w_k}^2 \mathrm{tr}\{\mathbf{h}_k^{\perp H}\mathcal{R}^+\mathbf{h}_k^{\perp}\}\right\}\end{aligned} \tag{2.29}$$

Note that this operation does not change the optimizer of the DML criterion as $\sum_k \widehat{\sigma}_{w,k}^2 \mathrm{tr}\,\{P_{\mathbf{h}_k^{\perp H}}\} = p\ \sum_k \widehat{\sigma}_{w,k}^2$ is constant w.r.t. $\mathbf{h}$. We take $\widehat{\sigma}_{w,k}^2$ to be a consistent estimate of the noise variance at tone k.

Where the denoised DML criterion (2.29) is now solved in the IQML way: considering $\mathcal{R}(\mathbf{h}_k) = \mathcal{R}$ as constant, the optimization problem becomes again quadratic in $\mathbf{h}$:

$$\begin{aligned}&\min_{\|\mathbf{h}\|=1} \sum_k \mathbf{h}_k^H\mathcal{U}_k^H\mathcal{R}^+\mathcal{U}_k\mathbf{h}_k - \sum_k \widehat{\sigma}_{w_k}^2 \mathbf{h}_k^H\,\widehat{\mathrm{B}}_k\mathbf{h}_k\\ &\min_{\|\mathbf{h}\|=1} \mathbf{h}^H\left\{\sum_k G_k^H\,\widehat{A}_k G_k - \sum_k \widehat{\sigma}_{w_k}^2 G_k^H\,\widehat{\mathrm{B}}_k G_k\right\}\mathbf{h}\end{aligned} \tag{2.30}$$

where $\widehat{A}_k = \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k$ and the matrix $\widehat{\mathrm{B}}_k(\mathbf{h}_k)$ is such that $\mathbf{h}_k^{\text{“}H} \widehat{\mathrm{B}}_k \mathbf{h}_k^{'} = \mathrm{tr}\{\mathbf{h}_k^{\text{“}\perp H} \mathcal{R}^+(\mathbf{h}_k) \mathbf{h}_k^{\text{‘}\perp}\}$.

If in (2.30) the noise is considered white so its variance $\widehat{\sigma}_{w_k}^2$ is constant over all the tones. Hence, $\widehat{\sigma}_{w_k}^2$ is replaced by $\widehat{\sigma}_w^2$ and we call this method of denoising a global denoising. Hence,

$$\min_{\mathbf{h}} \ \mathbf{h}^H \left(\widehat{A} - \widehat{\sigma}_w^2 \, \widehat{\mathrm{B}} \right) \mathbf{h} \tag{2.31}$$

Where $\widehat{A} = \sum_k G_k^H \, \widehat{A}_k G_k$ and $\widehat{\mathrm{B}} = \sum_k G_k^H \, \widehat{\mathrm{B}}_k G_k$.

Asymptotically in the number of data, DIQML is globally convergent. Indeed, asymptotically it is essentially equivalent to the denoised criterion:

$$\frac{1}{M} \mathbf{h}^H \left\{ \sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k - \widehat{\sigma_w^2} \sum_k G_k^H \, \widehat{\mathrm{B}}_k G_k \right\} \mathbf{h} \ = \\ \frac{1}{M} \mathbf{h}^H \sum_k G_k^H \mathbf{h}_k^{\perp H} \mathcal{R}^+ \mathbf{h}_k^{\perp} G_k \ \mathbf{h} + \ \mathcal{O}_p(\frac{1}{\sqrt{M}}) \tag{2.32}$$

if $\widehat{\sigma_w^2} - \sigma_w^2 = \mathcal{O}_p(\frac{1}{\sqrt{M}})$. The denoised criterion (the first term of the RHS of (2.32)) corresponds to the IQML criterion in the noiseless case and hence leads to $\mathbf{h} = \alpha \, \mathbf{h}^o$ for some scaling factor α, under the identifiability conditions of SRM. One iteration of DIQML hence yields an estimate $\mathbf{h} = \alpha \, \mathbf{h}^o + \mathcal{O}_p(\frac{1}{\sqrt{M}})$. So the DIQML algorithm behaves asymptotically at any SNR like the IQML algorithm behaves at high SNR:

- the first iteration gives a consistent estimate of the channel,
- this behavior holds whatever the initialization.

The second iteration gives asymptotically the global minimizer of DIQML. Unlike in the high SNR IQML case though, this global minimizer at an arbitrary SNR is not the DML minimizer [28]. As the SNR increases, the difference between DIQML and IQML disappears and we have global convergence, to the DML solution.

2.5.4.2 Finite Amount of Data

The choice of $\widehat{\sigma_w^2}$ turns out to be crucial. In practice, with large but finite amount of data (No. of Rx CP blocks) M, and the true noise variance value, the central matrix $\mathcal{Q} = \sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k - \widehat{\sigma_w^2} \sum_k G_k^H \, \widehat{\mathrm{B}}_k G_k$ in (2.30) is indefinite, and the minimization problem is no longer well posed.

The solution in this case would be $V_{min}(Q)$ corresponding to the smallest eigenvalue $\lambda_{min}(Q)$, which is negative. Simulations have shown that performance does not improve upon IQML in this case. The central matrix $\mathcal{Q}$ should be constrained to be positive semi-definite [28].

For the consistent estimate of σ_w^2, we choose here a certain λ that renders $\mathcal{Q} = \mathcal{Q}(\mathcal{R}^+) = \sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k - \lambda \sum_k G_k^H \widehat{\mathrm{B}}_k G_k$ exactly positive semi-definite with one singularity. The DIQML criterion becomes:

$$\begin{array}{l} \min\limits_{\|\mathbf{h}\|=1,\lambda} \mathbf{h}^H \left\{ \sum\limits_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k - \lambda \sum\limits_k G_k^H \widehat{\mathrm{B}}_k G_k \right\} \mathbf{h} \\ \min\limits_{\|\mathbf{h}\|=1,\lambda} \mathbf{h}^H \left(\widehat{A} - \lambda \widehat{\mathrm{B}} \right) \mathbf{h} \end{array} \tag{2.33}$$

with the constraint that $\mathcal{Q}$ be positive semi-definite. The solution is:
$\lambda = \lambda_{min}(\sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k, \sum_k G_k^H \widehat{\mathrm{B}}_k G_k)$, the minimal generalized eigenvalue of $\sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k$ and $\sum_k G_k^H \widehat{\mathrm{B}}_k G_k$, and
$\mathbf{h} = V_{min}(\sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k, \sum_k G_k^H \widehat{\mathrm{B}}_k G_k)$, the corresponding generalized eigenvector.

Asymptotically, the DIQML criterion (2.33) becomes

$$\begin{array}{l} \frac{1}{M} \mathbf{h}^H \left\{ \sum_k G_k^H \mathcal{U}_k^H \mathcal{R}^+ \mathcal{U}_k G_k - \lambda \sum_k G_k^H \widehat{\mathrm{B}}_k G_k \right\} \mathbf{h} = \\ \frac{1}{M} \mathbf{h}^H \sum_k G_k^H \mathbf{h}_k^{\perp} \mathcal{R}^+ \mathbf{h}_k^{\perp H} G_k \mathbf{h} + \frac{1}{M} (\sigma_w^2 - \lambda) \mathbf{h}^H \sum_k G_k^H \widehat{\mathrm{B}}_k G_k \mathbf{h} + \mathcal{O}_p(\frac{1}{\sqrt{M}}) \ . \end{array} \tag{2.34}$$

Optimization w.r.t. λ, subject to the non-negativity constraint, leads to $\lambda = \sigma_w^2 + \mathcal{O}_p(\frac{1}{\sqrt{M}})$, regardless of channel initialization (in $\mathcal{R}$ and $\widehat{\mathrm{B}}_k$), and the criterion (2.33) in $\mathbf{h}$ and λ becomes equivalent to the criterion (2.32) in $\mathbf{h}$. Hence, asymptotic global convergence applies for $\mathbf{h}$ and for λ (to σ_w^2), with the same properties as mentioned earlier (independently from the initialization).

On the other hand, if the noise is colored then we should denoise the IQML cost function on a tone by tone basis as shown in (2.30). We call this method a denoising per tone. Moreover, as we will show later in our simulations, even if the noise is white it is better to denoise on a tone by tone basis. One way to interpret this result is by noting that $\widehat{\sigma}_{w_k}^2$ guarantees that $\widehat{A}_k - \widehat{\sigma}_{w_k}^2 \widehat{\mathrm{B}}_k$ to be positive semi-definite whatever the tone k is. On the contrary, $\widehat{\sigma}_w^2$ can't guarantee this condition to be fullfilled for all the tones. As our simulations show, the fulfillment of this condition is crucial in making DIQML outperforms IQML.

The solution for denoising on a tone by tone basis is given by:
$\mathbf{h} = V_{min} \left(\sum_k G_k^H \widehat{A}_k G_k - \sum_k \widehat{\sigma}_{w_k}^2 G_k^H \widehat{\mathrm{B}}_k G_k \right)$ where $\widehat{\sigma}_{w_k}^2 = \lambda_{min}(\widehat{A}_k, \widehat{\mathrm{B}}_k)$,

the minimal generalized eigenvalue of $\widehat{A}_k$ and $\widehat{\mathrm{B}}_k$. Of course, one can now go further in denoising and replace $\widehat{\mathbf{r}}_k - \widehat{\sigma}^2_{w_k} I_p$ by its pure signal subspace part.

2.5.5 Pseudo Quadratic Maximum Likelihood (PQML)

The cost function of DML per tone is given by (2.20):

$$\begin{array}{l} \min\limits_{\mathbf{h}_k} \mathrm{tr}\left\{ R^+ \mathbf{h}_k^{\perp} \widehat{\mathbf{r}}_k \mathbf{h}_k^{\perp H} \right\} \\ \min\limits_{\mathbf{h}_k} \mathbf{u}_k^H \mathbf{h}_k^{\perp H} R^+ \mathbf{h}_k^{\perp} \mathbf{u}_k \end{array} \tag{2.35}$$

The principle of PQML has been introduced in the context of sinusoids in noise estimation [49] and then applied to DML channel estimation in [50]. The gradient of the DML cost function may be arranged as $\mathcal{P}(\mathbf{h}_k)\,\mathbf{h}_k$, where $\mathcal{P}(\mathbf{h}_k)$ is ideally a positive semi-definite matrix with a one-dimensional nullspace. The DML estimate satisfies

$$\mathcal{P}(\mathbf{h}_k)\ \mathbf{h}_k = 0, \tag{2.36}$$

which is solved for $\mathbf{h}_k$ under the constraint $\|\mathbf{h}_k\| = 1$. The DML gradient in (2.36) is the same as the gradient of the (pseudo-)quadratic cost function $\mathbf{h}_k^H\, \mathcal{P}(\widehat{\mathbf{h}}_k)\, \mathbf{h}_k$ evaluated at $\widehat{\mathbf{h}}_k = \mathbf{h}_k$. The PQML strategy is now the following. At iteration (i), $\mathcal{P}(\widehat{\mathbf{h}}_k^{(i-1)}) \geq 0$ is fixed. The problem $\min_{\mathbf{h}_k:\, \|\mathbf{h}_k\|=1} \mathbf{h}_k^H\, \mathcal{P}(\widehat{\mathbf{h}}_k^{(i-1)})\, \mathbf{h}_k$ is quadratic and its solution is $\widehat{\mathbf{h}}_k^{(i)} = V_{min}(\mathcal{P}(\widehat{\mathbf{h}}_k^{(i-1)}))$. This solution is used to reevaluate $\mathcal{P}(\mathbf{h}_k)$ and further iterations may be performed.

The difficulty consists in defining the right $\mathcal{P}(\mathbf{h}_k)$ in the DML gradient, especially with the positive semi-definiteness constraint. In general, and in particular for the DML problem at hand, the choice for $\mathcal{P}(\mathbf{h}_k)$ is indeed not unique. The gradient of the DML cost function consists of two terms (here we write the gradient w.r.t. $\mathbf{h}_{k,j}$, which is also component j of the gradient w.r.t. $\mathbf{h}_k$):

$$\left(\mathcal{P}(\mathbf{h}_k)\underline{\mathbf{h}_k}\right)(j) = \mathbf{u}_k^H \frac{\partial \mathbf{h}_k^{\perp H}}{\partial \mathbf{h}_{k,j}} R^+ \underline{\mathbf{h}_k}^{\perp} \mathbf{u}_k - \mathbf{u}_k^H \mathbf{h}_k^{\perp H} R^+ \underline{\mathbf{h}_k}^{\perp} \frac{\partial \mathbf{h}_k^{\perp H}}{\partial \mathbf{h}_{k,j}} R^+ \mathbf{h}_k^{\perp} \mathbf{u}_k \tag{2.37}$$

In each iteration, $\mathcal{P}(\mathbf{h}_k)$ will be considered as constant. The question now is which factors $\mathbf{h}_k$ should be considered as variable and which instances of $\mathbf{h}_k$ are considered as part of $\mathcal{P}(\mathbf{h}_k)$. $\underline{\mathbf{h}_k}$ in (2.37) designates those instances of $\mathbf{h}_k$ that we consider as variable (on which minimization will be done) and $\mathbf{h}_k$

designates those instances of $\mathbf{h}_k$ that are considered as part of the constant $\mathcal{P}(\mathbf{h}_k)$. The first term of $\mathcal{P}(\mathbf{h}_k)\underline{\mathbf{h}_k}$ is $\mathcal{U}_k^H\mathcal{R}^+(\mathbf{h}_k)\mathcal{U}_k\underline{\mathbf{h}_k}$, which is the IQML gradient, and the second term is $\overline{\mathcal{B}^H}(\mathbf{h}_k)\mathcal{B}(\mathbf{h}_k)\underline{\mathbf{h}_k}$, with $\mathbf{u}_k^H\mathbf{h}_k^{\perp H}\mathcal{R}^+(\mathbf{h}_k)\underline{\mathbf{h}_k^{\perp}} = \underline{\mathbf{h}_k}^T\mathcal{B}^T(\mathbf{h}_k)$ (note that $\mathbf{u}_k^H P_{\mathbf{h}_k^{\perp H}}\mathbf{u}_k = \left(\mathbf{u}_k^H P_{\mathbf{h}_k^{\perp H}}\mathbf{u}_k\right)^*$). Then $\mathcal{P}(\mathbf{h}_k)$ has the following form:

$$\mathcal{P}(\mathbf{h}_k) = \mathcal{U}_k^H R^+\mathcal{U}_k - \mathcal{B}^H(\mathbf{h}_k)\mathcal{B}(\mathbf{h}_k) \tag{2.38}$$

The second term of $\frac{1}{M}\mathcal{P}(\mathbf{h}_k)$ asymptotically tends to its expected value by the law of large numbers. The matrix $\mathcal{P}(\mathbf{h}_k)$ is indefinite for a finite number of receiving blocks M, and applying the PQML strategy directly will not work. In [50], $\mathbf{h}_k$ is chosen as the eigenvector corresponding to the smallest eigenvalue magnitude of $\mathcal{P}(\mathbf{h}_k)$; it gives poor performance except at high SNR.

PQML is closely related to DIQML as the first term of (2.30) and (2.38) are the same and $\mathrm{E}(\mathcal{B}^H(\mathbf{h}_k^o)\mathcal{B}(\mathbf{h}_k^o)) = \sigma_{w,k}^2\,\widehat{\mathrm{B}}_k(\mathbf{h}_k^o)$. By analogy with DIQML for which $\mathcal{Q}$ was also indefinite for finite M if an arbitrary $\widehat{\sigma_{w,k}^2}$ were to be used, we introduce a variable λ_k such that $\mathcal{U}_k^H\mathcal{R}^+\mathcal{U}_k - \lambda_k\,\mathcal{B}_k^H\mathcal{B}_k$ is exactly positive semi-definite. PQML then becomes the following minimization problem:

$$\min_{\|\mathbf{h}_k\|=1,\lambda} \mathbf{h}_k^H\left\{\mathcal{U}_k^H\mathcal{R}^+\mathcal{U}_k - \lambda_k\,\mathcal{B}_k^H\mathcal{B}_k\right\}\mathbf{h}_k \tag{2.39}$$

with a semi-definite positivity constraint on the central matrix. However, the overall cost function is as follows:

$$\min_{||h||=1,\lambda} \mathbf{h}^H\left\{\sum_k G_k^H\mathcal{U}_k^H R^+\mathcal{U}_k G_k - \lambda\sum_k G_k^H\mathcal{B}_k^H\mathcal{B}_k G_k\right\}\mathbf{h} \tag{2.40}$$

The solution is again
$\mathbf{h} = V_{min}(\sum_k G_k^H\mathcal{U}_k^H R^+\mathcal{U}_k G_k, \sum_k G_k^H\mathcal{B}_k^H\mathcal{B}_k G_k)$ corresponding to:
$\lambda = \lambda_{min}(\sum_k G_k^H\mathcal{U}_k^H R^+\mathcal{U}_k G_k, \sum_k G_k^H\mathcal{B}_k^H\mathcal{B}_k G_k)$. Asymptotically for a consistent initialization, there is global convergence for $\mathbf{h}$, as described previously, as well as for λ ($\rightarrow 1$). However, for a finite amount of data, and for an arbitrary $\mathbf{h}$,

$$\begin{aligned}\lambda &= \lambda_{min}(\textstyle\sum_k G_k^H\mathcal{U}_k^H R^+\mathcal{U}_k G_k, \sum_k G_k^H\mathcal{B}_k^H\mathcal{B}_k G_k)\\ &= \min_{\widehat{\mathbf{h}}} \frac{\widehat{\mathbf{h}}^H\sum_k G_k^H\mathcal{U}_k^H R^+\mathcal{U}_k G_k\widehat{\mathbf{h}}}{\widehat{\mathbf{h}}^H\sum_k G_k^H\mathcal{B}_k^H\mathcal{B}_k G_k\widehat{\mathbf{h}}} \leq \frac{\mathbf{h}^H\sum_k G_k^H\mathcal{U}_k^H R^+\mathcal{U}_k G_k\mathbf{h}}{\mathbf{h}^H\sum_k G_k^H\mathcal{B}_k^H\mathcal{B}_k G_k\mathbf{h}} = 1\end{aligned} \tag{2.41}$$

which means that using $\lambda = 1$, as in the original PQML algorithm, systematically leads to an indefinite $\mathcal{P}(\mathbf{h})$. On the other hand, we can proceed

with denoising per tone as in the case of DIQML such that the cost function becomes:

$$\min_{||h||=1,\lambda} \mathbf{h}^H \left\{ \sum_k G_k^H \mathcal{U}_k^H R^+ \mathcal{U}_k G_k - \sum_k \lambda_k G_k^H \mathcal{B}_k^H \mathcal{B}_k G_k \right\} \mathbf{h} \tag{2.42}$$

Where $\lambda_k = \lambda_{min}(\mathcal{U}_k^H R^+ \mathcal{U}_k, \mathcal{B}_k^H \mathcal{B}_k)$, the minimal generalized eigenvalue of $\mathcal{U}_k^H R^+ \mathcal{U}_k$ and $\mathcal{B}_k^H \mathcal{B}_k$ and $\mathbf{h}$ is given by $V_{min}\left(\sum_k G_k^H \mathcal{U}_k^H R^+ \mathcal{U}_k G_k - \sum_k \lambda_k G_k^H \mathcal{B}_k^H \mathcal{B}_k G_k\right)$. However, as we will see later, the asymptotic performance of PQML with per tone denoisng or global denoising are equivalent. Moreover, for a finite amount of data, PQML with global denoising outperforms that with per tone denoising.

2.6 Performance Analysis

2.6.1 Cramer-Rao Bound (CRB)

Before we proceed in the derivations of the performance of different algorithms mentioned earlier we need to derive a reference or a lower bound for the estimation process at our hands. It is well known that CRB is one of the famous bounds used extensively by the statisticians as a lower bound for any unbiased estimation process. To derive a formula for CRB for our case we start with the derivation of the FIM per tone k for the deterministic case then we can deduce the overall FIM. It is well known that FIM for such a case is given by:

$$J_{\theta_k \theta_k} = \mathrm{E}_{\mathbf{u}_k/\theta_k} \left(\frac{\partial Lnf(\mathbf{u}_k/\theta_k)}{\partial \theta_k^*} \right) \left(\frac{\partial Lnf(\mathbf{u}_k/\theta_k)}{\partial \theta_k^*} \right)^H \tag{2.43}$$

Where $\theta_k = \begin{bmatrix} \mathbf{x}_k^H & \mathbf{h}_k^H \end{bmatrix}^H$ Knowing that $f(\mathbf{u}_k/\theta_k) = \frac{1}{(\pi\sigma_{w,k}^2)^p} exp\left(-\frac{1}{\sigma_{w,k}^2} ||\mathbf{u}_k - \mathbf{h}_k\mathbf{x}_k||^2 \right)$, we get:

$$\frac{\partial Lnf(\mathbf{u}_k/\theta_k)}{\partial \mathbf{h}_k^*} = \frac{1}{\sigma_{w,k}^2} \mathbf{x}_k^* \left(\mathbf{u}_k - \mathbf{x}_k \mathbf{h}_k \right) \tag{2.44}$$

$$\frac{\partial Lnf(\mathbf{u}_k/\theta_k)}{\partial \mathbf{x}_k^*} = \frac{1}{\sigma_{w,k}^2} \mathbf{h}_k^H \left(\mathbf{u}_k - \mathbf{x}_k \mathbf{h}_k \right) \tag{2.45}$$

Hence we get:

$$\begin{aligned} J_{\theta_k \theta_k} &= \frac{1}{\sigma_{w,k}^2} \begin{bmatrix} \mathbf{h}_k^H \\ \mathbf{x}_k^* I_p \end{bmatrix} \begin{bmatrix} \mathbf{h}_k & \mathbf{x}_k I_p \end{bmatrix} \\ J_{\theta_k \theta_k} &= \begin{bmatrix} J_{\mathbf{x}_k \mathbf{x}_k} & J_{\mathbf{x}_k \mathbf{h}_k} \\ J_{\mathbf{h}_k \mathbf{x}_k} & J_{\mathbf{h}_k \mathbf{h}_k} \end{bmatrix} \end{aligned} \tag{2.46}$$

Using Schur's complement we get:

$$\begin{aligned} J_{\mathbf{h}_k\mathbf{h}_k}(\theta) &= J_{\mathbf{h}_k\mathbf{h}_k} - J_{\mathbf{h}_k\mathbf{x}_k} J_{\mathbf{x}_k\mathbf{x}_k}^{-1} J_{\mathbf{x}_k\mathbf{h}_k} \\ J_{\mathbf{h}_k\mathbf{h}_k}(\theta) &= \frac{|\mathbf{x}_k|^2}{\sigma_{w,k}^2}\left(I_p - \frac{\mathbf{h}_k\mathbf{h}_k^H}{||\mathbf{h}_k||^2}\right) \\ J_{\mathbf{h}_k\mathbf{h}_k}(\theta) &= \frac{|\mathbf{x}_k|^2}{\sigma_{w,k}^2} P_{\mathbf{h}_k}^{\perp} \end{aligned} \tag{2.47}$$

As we mentioned earlier in section III, the overall FIM can be obtained from the FIMs at different tones by:

$$J_{\mathbf{hh}} = \sum_k \frac{|\mathbf{x}_k|^2}{\sigma_{w,k}^2} G_k^H P_{\mathbf{h}_k}^{\perp} G_k \tag{2.48}$$

Hence, the deterministic CRB under the constraints that will be shown in the next section can be obtained [51] as the pseudo inverse of $J_{\mathbf{hh}}$:

$$CRB_{det} = J_{\mathbf{hh}}^{+} = \left\{ \sum_k \frac{|\mathbf{x}_k|^2}{\sigma_{w,k}^2} G_k^H P_{\mathbf{h}_k}^{\perp} G_k \right\}^{+} \tag{2.49}$$

2.6.2 Fixing the ambiguity by constraints

The channel can only be estimated blindly up to a scale factor. To make the estimation problem well posed, constraints on the channel need to be introduced to fix its unidentifiable components. We compute the asymptotic performance under the following constraints:

(1) a unit norm constraint:

$$||\mathbf{h}||^2 = 1 \tag{2.50}$$

which allows to fix the norm of the channel to unity. This constraint appears naturally when $\mathbf{h}$ is obtained as an eigenvector corresponds to the minimum eigenvalue of the cost function Q in (2.8).

(2) An additional constraint is necessary to adjust the scalar ambiguity:

$$\mathbf{h}^{oH}\mathbf{h}_{Lin} = \mathbf{h}^{oH}\mathbf{h}^{o} \tag{2.51}$$

Where $\mathbf{h}^o$ denotes the true channel impulse response, $\mathbf{h}$ denotes the estimated channel that we obtain from (2.8) subject to (2.50) and $\mathbf{h}_{Lin} = \beta\mathbf{h}$ stands for the estimated channel after imposing the linear constraint in (2.51)with β designates the scalar ambiguity. The solution we get from (2.51) is :

$$\mathbf{h}_{Lin} = \frac{\mathbf{h}^{oH}\mathbf{h}^{o}}{\mathbf{h}^{oH}\mathbf{h}}\mathbf{h}. \tag{2.52}$$

However, sometimes it is possible to find the optimal solution by minimizing (2.8) subject directly to the constraint in (2.51), hence we obtain: $\min_{\mathbf{h}^{oH}\mathbf{h}_{Lin}=\mathbf{h}^{oH}\mathbf{h}^o} \mathbf{h}^H\, Q\, \mathbf{h}$. Applying the Lagrange multiplier, the solution is:

$$\mathbf{h}_{Lin} = \frac{\mathbf{h}^{oH}\mathbf{h}^o}{\mathbf{h}^{oH}Q^{-1}\mathbf{h}^o}Q^{-1}\mathbf{h}^o \tag{2.53}$$

When Q is singular like in the case of DIQML and PQML, the solution in (2.53) can't be applied and consequently we refer to the solution in (2.52). The particular constraints in (2.50, 2.51) were chosen to characterize the asymptotic performance because they yield the minimal MSE, $\mathrm{E}\|\mathbf{h}-\mathbf{h}^o\|^2$, for a minimal number of independent constraints [51].

2.6.3 Framework for performance analysis

The constraints (2.50), (2.51) are of the form $\mathcal{K}(\mathbf{h}) = 0$, $\mathcal{K} : \mathrm{R}^{pL} \rightarrow \mathrm{R}$. We denote by $\mathcal{M}_{\mathbf{h}^o}$ the tangent subspace to the constraint set at the point $\mathbf{h}^o$:

$$\mathcal{M}_{\mathbf{h}^o} = \left\{ Z \in \mathrm{R}^{pL}\ ;\ \left(\frac{\partial \mathcal{K}^H(\mathbf{h}^o)}{\partial \mathbf{h}^*}\right)^H Z = 0 \right\} . \tag{2.54}$$

For the constraints (2.50), (2.51), we get $\dfrac{\partial \mathcal{K}^H(\mathbf{h}^o)}{\partial \mathbf{h}^*} = \mathbf{h}^o$ For the asymptotic performance, any constraint set that leads to the same tangent subspace $\mathcal{M}_{\mathbf{h}^o}$ is equivalent. Let $\mathcal{V}_R^o$ be a matrix whose columns form an orthonormal basis of $\mathcal{M}_{\mathbf{h}_R^o}$. Then locally we can write $\Delta\mathbf{h} = \mathcal{V}^o\,\theta$ where θ are the unconstrained parameter variations. A Taylor series expansion of $\mathcal{F}(\mathbf{h})$ at $\mathbf{h}^o$ in terms of θ gives

$$\mathcal{F}(\mathbf{h}) = \mathcal{F}(\mathbf{h}^o) + \theta^H \mathcal{V}^{o\,H}\frac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*} + \left[\frac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*}\right]^H \mathcal{V}^o\theta + \theta^H \mathcal{V}^{o\,H}\frac{\partial^2 \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^2}\mathcal{V}^o\theta + \mathcal{O}(||\theta||^3) . \tag{2.55}$$

Optimization of (2.55) up to second order w.r.t. θ gives for $\Delta\mathbf{h} = \mathcal{V}^o\,\theta$:

$$\Delta\mathbf{h} = \mathcal{V}^o \left(\mathcal{V}^{o\,H}\frac{\partial}{\partial \mathbf{h}^*}\left(\frac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*}\right)^H \mathcal{V}^o \right)^{-1} \mathcal{V}^{o\,H}\frac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*} \quad . \tag{2.56}$$

For the constraints (2.50), (2.51) or equivalent, the columns of $\mathcal{V}^o$ form a basis for the orthogonal complement of $\mathbf{h}^o$. We shall also require

$$\begin{cases} J_1 &= \mathrm{E}\left(\dfrac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*}\right)\left(\dfrac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*}\right)^H \\ J_2 &= \mathrm{E}\dfrac{\partial}{\partial \mathbf{h}^*}\left(\dfrac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*}\right)^H . \end{cases} \tag{2.57}$$

Note that if $\mathcal{F}$ would have been the log likelihood function, then $J_1 = -J_2$, but this equality does not hold here. We now obtain

$$\begin{cases} \tilde{\mathbf{h}} & = & \mathcal{V}^o \left(\mathcal{V}^{o\,H} J_2 \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} \dfrac{\partial \mathcal{F}(\mathbf{h}^o)}{\partial \mathbf{h}^*} + \mathcal{O}_p(\frac{1}{M}) \\ C_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} & = & \mathcal{V}^o \left(\mathcal{V}^{o\,H} J_2 \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} J_1 \mathcal{V}^o \left(\mathcal{V}^{o\,H} J_2 \mathcal{V}^o\right)^{-1} \mathcal{V}^{o\,H} + o(\frac{1}{M}) \\ C_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} & = & \left(\mathcal{V}^o \mathcal{V}^{o\,H} J_2 \mathcal{V}^o \mathcal{V}^{o\,H}\right)^{+} J_1 \left(\mathcal{V}^o \mathcal{V}^{o\,H} J_2 \mathcal{V}^o \mathcal{V}^{o\,H}\right)^{+} + o(\frac{1}{M}) \\ C_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} & = & \left(P^{\perp}_{\mathbf{h}^o} J_2 P^{\perp}_{\mathbf{h}^o}\right)^{+} J_1 \left(P^{\perp}_{\mathbf{h}^o} J_2 P^{\perp}_{\mathbf{h}^o}\right)^{+} \end{cases} \tag{2.58}$$

where $\mathcal{V}^o \mathcal{V}^{o\,H} = P^{\perp}_F$, $F = \frac{\partial \mathcal{K}^H(\mathbf{h}^o)}{\partial \mathbf{h}^*}$, $\mathcal{K}(\mathbf{h}) = \mathbf{h}^{oH}\mathbf{h} - \mathbf{h}^{oH}\mathbf{h}^o = 0$. Hence, $F = \mathbf{h}^o$ and $P^{\perp}_F = P^{\perp}_{\mathbf{h}^o}$.

We know from (2.8) that the over all cost function can be written as: $\mathcal{F}(\mathbf{h}) = \mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \, Q_k \, G_k\right] \mathbf{h}$.
Hence the gradient of the cost function is given by:
$\mathcal{F}'(\mathbf{h}) = \frac{\partial \mathcal{F}(\mathbf{h})}{\partial \mathbf{h}^*} = \sum_k G_k^H Q_k G_k \mathbf{h} = \sum_k G_k^H Q_k \mathbf{h}_k = \sum_k G_k^H \mathcal{F}'_k(\mathbf{h}_k)$, where $\mathcal{F}'_k(\mathbf{h}_k)$ is the gradient of the cost function per tone k, $\mathcal{F}(\mathbf{h}_k)$.

Thus the covariance matrix of the gradient is given by:

$$\begin{aligned} J_1 &= \sum_k \sum_n G_k^H \, \mathrm{E}\left\{Q_k \mathbf{h}_k \mathbf{h}_n^H Q_n\right\} G_n \\ &= \sum_k \sum_n G_k^H \, \mathrm{E}\left\{\mathcal{F}'_k(\mathbf{h}_k) \mathcal{F}'^{H}_n(\mathbf{h}_n)\right\} G_n \\ &= \sum_k \sum_n G_k^H J_{1,k,n} G_n \end{aligned} \tag{2.59}$$

On the other hand, the hessian matrix of the cost function is given by:

$$\begin{aligned} J_2 &= \mathrm{E}\left\{\mathcal{F}''(\mathbf{h})\right\} \\ &= \mathrm{E}\left\{\frac{\partial \mathcal{F}'^{H}(\mathbf{h})}{\partial \mathbf{h}^*}\right\} \\ &= \sum_k G_k^H \, \mathrm{E}\left\{Q_k\right\} G_k \\ &= \sum G_k \, \mathrm{E}\left\{\mathcal{F}''_k(\mathbf{h}_k)\right\} G_k^H \\ &= \sum G_k J_{2,k} G_k^H \end{aligned} \tag{2.60}$$

Where $\mathcal{F}''_k(\mathbf{h}_k)$ is the hessian of the cost function per tone k, $\mathcal{F}_k(\mathbf{h}_k)$. In the sequel, we will focus on the derivation of J_1 and J_2 for different channel estimation algorithms.

2.6.4 Performance Analysis of SSF

In [40] Ottersten et al derived formulas for J_1 and J_2 for signal subspace fitting method within the frame work of sensor array processing. In their derivation they considered the parameters to be estimated as real unknowns.

In our case, the paramters to be estimated are the channel coefficients which are complex. However, we can readily play around this problem by splitting the complex coefficients into their real and imaginary parts and treat them all as real parameters. Once we get J_1 and J_2 for the real case, we can deduce their corresponding formulas for the complex case as we shall see later. Moreover, we consider first that the channel is flat to be consistent with [40] hence all the aforementioned tasks are done first at a tone level then we exploit the relations deriverd in (2.59) and (2.60) to extend the results to a frequency selective case. Hence, our goal is to derive $J_{1,k}$ and $J_{2,k}$ for the channel coefficients per tone k by making an analogy with the sensor array processing problem then J_1 and J_2 follows directly. Hence by analogy with (50) and (51) in [40] we can write:

$$J_{2,k}^{real}(i,j) = 2Re\left[tr\left\{\Delta\mathbf{h}_{k,j}^{H} P_{\mathbf{h}_k^o}^{\perp} \Delta\mathbf{h}_{k,i} \mathbf{h}_k^{o+} V_{\mathcal{S},k} \mathcal{W}_k V_{\mathcal{S},k}^{H} \mathbf{h}_k^{o+H}\right\}\right] \tag{2.61}$$

$$J_{1,k}^{real}(i,j) = 2\sigma_{w,k}^2 Re\left[tr\left\{\Delta\mathbf{h}_{k,j}^{H} P_{\mathbf{h}_k^o}^{\perp} \Delta\mathbf{h}_{k,i} \mathbf{h}_k^{o+} V_{\mathcal{S},k} \mathcal{W}_k \Lambda_{\mathcal{S},k} \widetilde{\Lambda}_{\mathcal{S},k}^{-2} \mathcal{W}_k V_{\mathcal{S},k}^{H} \mathbf{h}_k^{o+H}\right\}\right] \tag{2.62}$$

Where $\Delta\mathbf{h}_{k,j} = \frac{\partial \mathbf{h}_k^o}{\bar{\mathbf{h}}_{k,j}}$, $\bar{\mathbf{h}}_{k,j}$ is the jth element of $\bar{\mathbf{h}}_k$, $\bar{\mathbf{h}}_k = [\mathbf{h}_k^{orT}, \mathbf{h}_k^{osT}]^T$ is a $(2p \times 1)$ element vector containing the real part $\mathbf{h}_k^{or} = Re\,[\mathbf{h}_k^o]$, and the imaginary part $\mathbf{h}_k^{os} = Im\,[\mathbf{h}_k^o]$. Finally, $\mathcal{W}_k$ is the weight used per tone k for signal subspace fitting method.

Let $\mathcal{C}_2(\mathbf{h}_k^o) = \mathbf{h}_k^{o+} V_{\mathcal{S},k} \mathcal{W}_k V_{\mathcal{S},k}^H \mathbf{h}_k^{o+H}$ in $J_{2,k}^{real}$ and $\mathcal{C}_1(\mathbf{h}_k^o) = \mathbf{h}_k^{o+} V_{\mathcal{S},k} \mathcal{W}_k \Lambda_{\mathcal{S},k}$ $\widetilde{\Lambda}_{\mathcal{S},k}^{-2} \mathcal{W}_k V_{\mathcal{S},k}^H \mathbf{h}_k^{o+H}$ in $J_{1,k}^{real}$, As we will see later, both $\mathcal{C}_2(\mathbf{h}_k^o)$ and $\mathcal{C}_1(\mathbf{h}_k^o)$ yield real scalar values. Moreover, knowing that i and j varies between 1 and 2p and let $\mathcal{P}_k(i,j) = 2Re\left[tr\left\{\Delta\mathbf{h}_{k,j}^H P_{\mathbf{h}_k^o}^{\perp} \Delta\mathbf{h}_{k,i}\right\}\right]$ then the matrix $\mathcal{P}_k(\mathbf{h}_k^o)$ has 4 states as follows:

$$\begin{array}{llll}
i \le p \ , \ j \le p & \mathcal{P}_k^{rr}(\mathbf{h}_k^o) & = & \frac{2}{||\mathbf{h}_k^o||^2}\left[||\mathbf{h}_k^o||^2 I_p - \mathbf{h}_k^{or}\mathbf{h}_k^{orT} - \mathbf{h}_k^{os}\mathbf{h}_k^{osT}\right] \\
i > p \ , \ j \le p & \mathcal{P}_k^{sr}(\mathbf{h}_k^o) & = & \frac{2}{||\mathbf{h}_k^o||^2}\left[\mathbf{h}_k^{or}\mathbf{h}_k^{osT} - \mathbf{h}_k^{os}\mathbf{h}_k^{orT}\right] \\
i \le p \ , \ j > p & \mathcal{P}_k^{rs}(\mathbf{h}_k^o) & = & \frac{2}{||\mathbf{h}_k^o||^2}\left[\mathbf{h}_k^{os}\mathbf{h}_k^{orT} - \mathbf{h}_k^{or}\mathbf{h}_k^{osT}\right] \\
i > p \ , \ j > p & \mathcal{P}_k^{ss}(\mathbf{h}_k^o) & = & \frac{2}{||\mathbf{h}_k^o||^2}\left[||\mathbf{h}_k^o||^2 I_p - \mathbf{h}_k^{or}\mathbf{h}_k^{orT} - \mathbf{h}_k^{os}\mathbf{h}_k^{osT}\right]
\end{array} \tag{2.63}$$

Now to transform our formulas from real to complex we exploit the following relationship:

$$\begin{array}{rcl} \mathcal{P}_k^{complex}(\mathbf{h}_k^o) & = & \frac{1}{4}\left[I_p \quad jI_p\right] \left[\begin{array}{cc} \mathcal{P}_k^{rr}(\mathbf{h}_k^o) & \mathcal{P}_k^{rs}(\mathbf{h}_k^o) \\ \mathcal{P}_k^{sr}(\mathbf{h}_k^o) & \mathcal{P}_k^{ss}(\mathbf{h}_k^o) \end{array}\right] \left[\begin{array}{c} I_p \\ -jI_p \end{array}\right] \\ \mathcal{P}_k^{complex}(\mathbf{h}_k^o) & = & \left(I_p - \frac{\mathbf{h}_k^o \mathbf{h}_k^{oH}}{||\mathbf{h}_k^o||^2}\right) \\ \mathcal{P}_k^{complex}(\mathbf{h}_k^o) & = & P_{\mathbf{h}_k^o}^{\perp} \end{array} \tag{2.64}$$

On the other hand, we know that $\mathcal{V}_{\mathcal{S},k} = \frac{\mathbf{h}_k^o}{||\mathbf{h}_k^o||}$, $\Lambda_{\mathcal{S},k} = \lambda_{\mathcal{S},k} = |\mathbf{x}_k|^2||\mathbf{h}_k^o||^2 + \sigma_{w,k}^2$, $\widetilde{\Lambda}_{\mathcal{S},k} = \lambda_{\mathcal{S},k} - \sigma_{w,k}^2 = |\mathbf{x}_k|^2||\mathbf{h}_k^o||^2$. Therefore, we are now ready to conclude J_2 and J_1 for all versions of SSF methods namely, UnWeighted Signal Subspace Fitting (UWSSF), Weighted Signal Subspace Fitting (WSSF), Optimally Weighted Signal Subspace Fitting (OWSSF).

2.6.4.1 Case of UWSSF

In this method, the weighting matrix $\mathcal{W}_k$ is scalar and it is replaced by a unity 1, $\mathcal{C}_2(\mathbf{h}_k^o) = \frac{1}{||\mathbf{h}_k^o||^2}$ and $\mathcal{C}_1(\mathbf{h}_k^o) = \frac{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2 + \sigma_{w,k}^2}{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^6}$. Making use of (2.61), (2.62), (2.64) and the above mentioned relations, we conclude that:

$$J_2^{UWSSF} = \sum_k \frac{1}{||\mathbf{h}_k^o||^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \tag{2.65}$$

$$J_1^{UWSSF} = \sum_k \frac{(|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2 + \sigma_{w,k}^2)\sigma_{w,k}^2}{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^6} \; G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \tag{2.66}$$

Hence, the error covariance matrix of UWSSF is given by:

$$\begin{array}{rcl} C_{\tilde{\mathbf{h}}} & = & \left(P_{\mathbf{h}^o}^{\perp} J_2^{UWSSF} P_{\mathbf{h}^o}^{\perp}\right)^{+} J_1^{UWSSF} \left(P_{\mathbf{h}^o}^{\perp} J_2^{UWSSF} P_{\mathbf{h}^o}^{\perp}\right)^{+} \\ C_{\tilde{\mathbf{h}}} & = & J_2^{UWSSF+} \; J_1^{UWSSF} \; J_2^{+UWSSF} \end{array} \tag{2.67}$$

Where we have used the fact that in this case $\left(P_{\mathbf{h}^o}^{\perp} J_2^{UWSSF} P_{\mathbf{h}^o}^{\perp}\right) = J_2^{UWSSF}$. This is obvious since J_2^{UWSSF} has already a singularity in the direction $\mathbf{h}^o$. This fact will be reused in the sequel for both WSSF and OWSSF.

2.6.4.2 Case of WSSF

The scalar weight is given by [40]: $\mathcal{W}_k = \widetilde{\Lambda}_{\mathcal{S},k} = \Lambda_{\mathcal{S},k} - \sigma_{w,k}^2 I = \lambda_{\mathcal{S},k} - \sigma_{w,k}^2 = |\mathbf{x}_k|^2||\mathbf{h}_k^o||^2$, $\mathcal{C}_2(\mathbf{h}_k^o) = |\mathbf{x}_k|^2$ and $\mathcal{C}_1(\mathbf{h}_k^o) = \frac{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2 + \sigma_{w,k}^2}{||\mathbf{h}_k^o||^2}$. Again making use of (2.61), (2.62), (2.64) and the above mentioned relations, we conclude that:

$$J_2^{WSSF} = \sum_k |\mathbf{x}_k|^2 G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \tag{2.68}$$

$$\begin{aligned} J_1^{WSSF} &= \textstyle\sum_k \frac{(|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2+\sigma_{w,k}^2)\sigma_{w,k}^2}{||\mathbf{h}_k^o||^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \\ &= \sigma_w^2 J_2 + \sigma_w^4 \textstyle\sum_k \frac{1}{||\mathbf{h}_k^o||^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \\ &= \sigma_w^2 J_2 + \sigma_w^4 \, J^{'} \end{aligned} \tag{2.69}$$

Where in the last two equalities in (2.69) we have assumed that the noise is temporally white so $\sigma_{w,k}^2 = \sigma_w^2 \quad \forall \, k$. Hence, the error covariance matrix of WSSF is given by:

$$\begin{aligned} C_{\tilde{\mathbf{h}}} &= \left(P_{\mathbf{h}^o}^{\perp} J_2 P_{\mathbf{h}^o}^{\perp}\right)^+ J_1 \left(P_{\mathbf{h}^o}^{\perp} J_2 P_{\mathbf{h}^o}^{\perp}\right)^+ \\ C_{\tilde{\mathbf{h}}} &= J_2^+ \, J_1 \, J_2^+ \\ C_{\tilde{\mathbf{h}}} &= J_2^+ (\sigma_w^2 J_2 + \sigma_w^4 \, J^{'}) J_2^+ \\ C_{\tilde{\mathbf{h}}} &= \sigma_w^2 J_2^+ \; + \; \sigma_w^4 J_2^+ J^{'} J_2^+ \\ &= CRB + \sigma_w^4 J_2^+ J^{'} J_2^+ \end{aligned} \tag{2.70}$$

However, in [40] it was proved that SSF with this weighting achieves asymptotically the same performance as DML. Moreover, it is that at very high SNR, the second term in (2.70) is negligible compared to the first term, hence $C_{\tilde{\mathbf{h}}}^{WSSF} \longrightarrow CRB_{det}$. This means that WSSF as well as DML achieves asymptotically CRB_{det}.

2.6.4.3 Case of OWSSF

The scalar weight is given by [40]: $\mathcal{W}_k = \widetilde{\Lambda}_{\mathcal{S},k}^2 \Lambda_{\mathcal{S},k}^{-1} = \widetilde{\lambda}_{\mathcal{S},k}^2 \lambda_{\mathcal{S},k}^{-1} = \frac{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^4}{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2+\sigma_{w,k}^2}$, $\mathcal{C}_2(\mathbf{h}_k^o) = \frac{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^2}{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2+\sigma_{w,k}^2}$ and $\mathcal{C}_1(\mathbf{h}_k^o) = \frac{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^2}{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2+\sigma_{w,k}^2}$. Note that $\mathcal{C}_1(\mathbf{h}_k^o) = \mathcal{C}_2(\mathbf{h}_k^o)$. Again making use of (2.61), (2.62), (2.64) and the above mentioned relations and on the top of that we assume that the noise is temporally white, we conclude that:

$$J_2^{OWSSF} = \sum_k \frac{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^2}{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2 + \sigma_w^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \tag{2.71}$$

$$J_1^{OWSSF} = \sum_k \frac{|\mathbf{x}_k|^4||\mathbf{h}_k^o||^2\sigma_w^2}{|\mathbf{x}_k|^2||\mathbf{h}_k^o||^2 + \sigma_w^2} \, G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \tag{2.72}$$

It is obvious that $J_1^{OWSSF} = \sigma_w^2 J_2^{OWSSF}$, hence,

$$\begin{aligned} C_{\tilde{\mathbf{h}}} &= \sigma_w^2 J_2^{+} \\ &= \sigma_w^2 \left(\sum_k \frac{|\mathbf{x}_k|^4 ||\mathbf{h}_k^o||^2}{|\mathbf{x}_k|^2 ||\mathbf{h}_k^o||^2 + \sigma_w^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \right)^{+} \end{aligned} \tag{2.73}$$

We have used here the fact that $P_{\mathbf{h}^o}^{\perp} J_2^{OWSSF} P_{\mathbf{h}^o}^{\perp} = J_2^{OWSSF}$. This relation is also true for DIQML and PQML as we will see later.
Remark: It is worthy to note that with this scalar weighting, SSF attains the stochastic CRB [40]. However, at very high SNR the term $\left(|\mathbf{x}_k|^2 ||\mathbf{h}_k^o||^2 + \sigma_w^2\right) \rightarrow |\mathbf{x}_k|^2 ||\mathbf{h}_k^o||^2$ in (2.73) hence, $C_{\tilde{\mathbf{h}}}^{OWSSF} \rightarrow CRB_{det} = CRB_{sto}$.

Another important remark is that for flat fading channel (single tap channel) all SSF methods (UWSSF, WSSF and OWSSF) yield the same performance as shown below:

$$C_{\tilde{\mathbf{h}}} = \sigma_w^2 \left(\frac{|\mathbf{x}_k|^2 ||\mathbf{h}_k^o||^2 + \sigma_w^2}{|\mathbf{x}_k|^4 ||\mathbf{h}_k^o||^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \right)^{+} \tag{2.74}$$

This means that the importance of weighting starts to appear once we have frequency selective fading channel.

2.6.5 Performance Analysis of SRM

The performance of SRM which is sometimes called cross relation has been investigated in [52]. The methodology of that analysis is similar to the case of subspace fitting presented in [40] since it is based on the asymptotic covariance matrix of the eigenvectors of R_{YY}. However, as we can readily notice from the simulation section in [52], the analytical performance analysis provided over there is only valid for moderate and high SNR. However, at low SNR where the simulation plots start to level off due to the constraint used to fix the scalar ambiguity, the analytical ones don't level off at all and consequently, it is no more valid at that SNR range. The shortcomings of the previously mentioned analysis encourage us to proceed in our work in order to provide a different way of performance analysis which is much more simpler and on the top of that which is valid over the the whole SNR range. To proceed with our analysis we start with the overall cost function of SRM which is given by (2.23):

$$\min_{\mathbf{h}} \; \mathbf{h}^H \left(\sum_k G_k^H \, \mathrm{E}\left\{ \mathcal{U}_k^H \mathcal{U}_k \right\} G_k \right) \mathbf{h} \tag{2.75}$$

Again, we are interested in computing J_1 and J_2 based on (2.59) and (2.60) respectively. Hence, $J_2 = \sum_k G_k^H J_{2,k} G_k$ and $J_{2,k} = \mathrm{E}Q_k = \mathrm{E}\mathcal{U}_k^H \mathcal{U}_k$ hence,

$\mathrm{E}\left\{\mathcal{U}_k^H\mathcal{U}_k\right\}$ is required to be computed. $J_1 = \sum_k \sum_n G_k^H J_{1,k,n} G_n$ hence, $J_{1,k,n} = \mathrm{E}\left\{Q_k \mathbf{h}_k^o \mathbf{h}_n^{oH} Q_n\right\}$ is also required. To compute $\mathrm{E}\left\{\mathcal{U}_k^H\mathcal{U}_k\right\}$, we will start with $\mathrm{E}\left\{\mathbf{h}_k^H\mathcal{U}_k^H\mathcal{U}_k\mathbf{h}_k\right\}$.

$$\begin{aligned} &= \mathrm{E}\left\{\mathbf{u}_k^H \mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp}\mathbf{u}_k\right\} \\ &= \mathrm{E}\,\mathrm{tr}\left\{\mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp}\mathbf{u}_k\mathbf{u}_k^H\right\} \\ &= \mathrm{E}\,\mathrm{tr}\left\{\mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp}\left(\mathbf{h}_k^o\mathbf{x}_k + \mathbf{w}_k\right)\left(\mathbf{x}_k^*\mathbf{h}_k^{oH} + \mathbf{w}_k^H\right)\right\} \\ &= |\mathbf{x}_k|^2 \mathrm{tr}\left\{\mathbf{h}_k^{\perp}\mathbf{h}_k^o\mathbf{h}_k^{oH}\mathbf{h}_k^{\perp H}\right\} + \sigma_w^2 \mathrm{tr}\left\{\mathbf{h}_k^{\perp H}\mathbf{h}_k^{\perp}\right\} \\ &= |\mathbf{x}_k|^2 \mathbf{h}_k^H\left\{\mathbf{h}_k^{o\perp H}\mathbf{h}_k^{o\perp}\right\}\mathbf{h}_k + \sigma_w^2\alpha||\mathbf{h}_k||^2 \\ &= |\mathbf{x}_k|^2 \mathbf{h}_k^H D_k^{(1)}\mathbf{h}_k + \sigma_w^2\alpha||\mathbf{h}_k||^2 \end{aligned} \tag{2.76}$$

Where $\alpha = 2 - \delta_{p,2}$. Therefore, $J_{2,k} = |\mathbf{x}_k|^2 D_k^{(1)} + \alpha\sigma_w^2 I$ and consequently:

$$J_2^{SRM} = \sum_k G_k^H \left(|\mathbf{x}_k|^2 D_k^{(1)} + \alpha\sigma_w^2 I\right) G_k \tag{2.77}$$

It is worthy to note here that unlike different versions of SSF, $P_{\mathbf{h}^o}^{\perp} J_2^{SRM} P_{\mathbf{h}^o}^{\perp} \neq J_2^{SRM}$. Now to compute $\mathrm{E}\left\{Q_k\mathbf{h}_k^o\mathbf{h}_n^{oH}Q_n\right\}$ we will start with $\mathrm{E}\left\{\mathbf{h}_k^H Q_k \mathbf{h}_k^o \mathbf{h}_n^{oH} Q_n \mathbf{h}_n\right\}$

$$\begin{aligned} &= \mathrm{E}\left\{\mathbf{h}_k^H\mathcal{U}_k^H\mathcal{U}_k\mathbf{h}_k^o\mathbf{h}_n^{oH}\mathcal{U}_n^H\mathcal{U}_n\mathbf{h}_n\right\} \\ &= \mathrm{E}\left\{\mathbf{u}_k^H\mathbf{h}_k^{\perp H}\mathbf{h}_k^{o\perp}\mathbf{u}_k\mathbf{u}_n^H\mathbf{h}_n^{o\perp H}\mathbf{h}_n^{\perp}\mathbf{u}_n\right\} \end{aligned} \tag{2.78}$$

Substituting $\mathbf{u}_k$ by $\mathbf{h}_k\mathbf{x}_k + \mathbf{w}_k$ and let $C_k = \mathbf{h}_k^{\perp H}\mathbf{h}_k^{o\perp}$, $C_n = \mathbf{h}_n^{o\perp H}\mathbf{h}_n^{\perp}$ and knowing that $\mathbf{h}_k^{o\perp}\mathbf{h}_k^o = 0$ then (2.78) yields:

$$\begin{aligned} &\underbrace{\mathrm{E}\left\{\mathbf{x}_k^*\mathbf{h}_k^H C_k\mathbf{w}_k\mathbf{w}_n^H C_n^H\mathbf{h}_n\mathbf{a}_n\right\}}_{P1} + \underbrace{\mathrm{E}\left\{\mathbf{w}_k^H C_k\mathbf{w}_k\mathbf{w}_n^H C_n^H\mathbf{w}_n\right\}}_{P2} + \\ &\underbrace{\mathrm{E}\left\{\mathbf{x}_k^*\mathbf{h}_k^H C_k\mathbf{w}_k\mathbf{w}_n^H C_n^H\mathbf{w}_n\right\}}_{P3} + \underbrace{\mathrm{E}\left\{\mathbf{w}_k^H C_k\mathbf{w}_k\mathbf{w}_n^H C_n^H\mathbf{h}_n\mathbf{a}_n\right\}}_{P4} \end{aligned} \tag{2.79}$$

.

Knowing that the noise is temporally-spatially uncorrelated, we have $\mathrm{E}\mathbf{w}_k\mathbf{w}_n^H = 0$ $(k \neq n)$, $\mathrm{E}\mathbf{w}_k\mathbf{w}_n^H = \sigma_w^2 I$ $(k = n)$ then: P1 vainshes for $k \neq n$, while for $k = n$ it becomes as follows:

$$\begin{aligned} P1 &= \sigma_w^2|\mathbf{x}_k|^2 \mathrm{tr}\left\{\mathbf{h}_k^{\perp}\mathbf{h}_k^o\mathbf{h}_k^{oH}\mathbf{h}_k^{\perp H}\mathbf{h}_k^{o\perp}\mathbf{h}_k^{o\perp H}\right\} \\ &= \sigma_w^2|\mathbf{x}_k|^2\mathbf{h}_k^H\left\{\mathbf{h}_k^{o\perp H}\mathbf{h}_k^{o\perp}\mathbf{h}_k^{o\perp H}\mathbf{h}_k^{o\perp}\right\}\mathbf{h}_k \\ &= \sigma_w^2|\mathbf{x}_k|^2\mathbf{h}_k^H D_k^{(2)}\mathbf{h}_k \end{aligned} \tag{2.80}$$

P2 is developed as follows using the formula of the mean of quartic forms of a Guassian variable [53]:

$$\begin{aligned} &= tr\left\{\sigma_w^4 C_k C_n\right\} + \sigma_w^4 \mathrm{tr}\left(C_k\right) tr\left(C_n^H\right) \\ &= \sigma_w^4 \mathrm{tr}\left\{\mathbf{h}_k^{\perp H}\mathbf{h}_k^{o\perp}\mathbf{h}_n^{o\perp H}\mathbf{h}_n^{\perp}\right\} + \sigma_w^4 tr\left\{\mathbf{h}_k^{\perp H}\mathbf{h}_k^{o\perp}\right\}\mathrm{tr}\left\{\mathbf{h}_n^{o\perp H}\mathbf{h}_n^{\perp}\right\} \\ &= \sigma_w^4\mathbf{h}_k^H D_{k,n}^{(3)}\mathbf{h}_n + \sigma_w^4\alpha^2\mathbf{h}_k^H\mathbf{h}_k^o\mathbf{h}_n^{oH}\mathbf{h}_n \end{aligned} \tag{2.81}$$

On the other hand, P3 and P4 vanishes since the 3rd order moment of the Gaussian noise is null. Thus, $\mathrm{E}\left\{\mathbf{h}_k^H Q_k \mathbf{h}_k^o \mathbf{h}_n^{oH} Q_n \mathbf{h}_n\right\}$ yields: $|\mathbf{x}_k|^2\sigma_w^2 \mathbf{h}_k^H D_k^{(2)} \mathbf{h}_k + \sigma_w^4 \mathbf{h}_k^H D_k^{(3)} \mathbf{h}_k + \sigma_w^4 \alpha^2 \mathbf{h}_k^H \mathbf{h}_k^o \mathbf{h}_n^{oH} \mathbf{h}_n$.
Consequently, $J_{1,k,n} = |\mathbf{x}_k|^2\sigma_w^2 D_k^{(2)} + \sigma_w^4 D_{k,n}^{(3)} + \sigma_w^4 \alpha^2 \mathbf{h}_k^o \mathbf{h}_n^{oH}$.
Now J_1^{SRM} can be written as follows:

$$J_1^{SRM} = \sum_k G_k^H \left(|\mathbf{x}_k|^2\sigma_w^2 D_k^{(2)} + \sigma_w^4 D_{k,n}^{(3)}\right) G_k + \sum_k \sum_n G_k^H \left(\sigma_w^4 \alpha^2 \mathbf{h}_k^o \mathbf{h}_n^{oH}\right) G_n \tag{2.82}$$

2.6.6 Performance Analysis of DIQML

In [28] the performance of DIQML for FIR channel is derived, hence, we can easily deduce the performance for CP (SCCP or OFDM) channel. The error covariance matrix for channel impulse response is given by (36) in [28] as follows:

$$C_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}^{DIQML} = CRB + CRB(B - \frac{B\mathbf{h}^o\mathbf{h}^{oH}B}{\mathbf{h}^{oH}B\mathbf{h}^o})CRB \tag{2.83}$$

B is defined similarly as $\widehat{B}$ in the DIQML section with $\mathbf{h}$ being replaced by $\mathbf{h}^o$. Substitute (2.49) in (2.83) and use (2.58) we get:

$$J_2^{DIQML} = \sum_k |\mathbf{x}_k|^2 G_k^H (I - \frac{\mathbf{h}_k^o \mathbf{h}_k^{oH}}{||\mathbf{h}_k^o||^2}) G_k \tag{2.84}$$

$$J_1^{DIQML,Global} = \sigma_w^2 J_2^{DIQML} + \sigma_w^4 (B - \frac{B\mathbf{h}^o\mathbf{h}^{oH}B}{\mathbf{h}^{oH}B\mathbf{h}^o}) \tag{2.85}$$

On the other hand, for denoising per tone we start with $J_{2,k}$ and $J_{1,k}$ which can be deduced directly from (36) in [28] for the case of one channel tap. Afterwards, J_2 and J_1 will be concluded. Hence from (2.83) we get:

$$C_{\tilde{\mathbf{h}}_k\tilde{\mathbf{h}}_k}^{DIQML} = CRB_k + CRB_k(B_k - \frac{B_k\mathbf{h}_k^o\mathbf{h}_k^{oH}B_k}{\mathbf{h}_k^{oH}B\mathbf{h}_k^o})CRB_k \tag{2.86}$$

Where from (2.49) $CRB_k = \frac{\sigma_{w,k}^2}{|\mathbf{x}_k|^2} P_{\mathbf{h}_k^o}^{\perp +}$.

Therefore, $J_{2,k} = |\mathbf{x}_k|^2 P_{\mathbf{h}_k^o}^{\perp}$ and consequently J_2 is the same as for global denoising. However, for J_1 we have:

$$\begin{aligned} J_1^{DIQML,pertone} &= \textstyle\sum_k G_k^H J_{1,k} G_k \\ J_{1,k} &= \sigma_{w,k}^2 J_{2,k} + \sigma_{w,k}^4 (B_k - \tfrac{B_k\mathbf{h}_k^o\mathbf{h}_k^{oH}B_k}{\mathbf{h}_k^{oH}B_k\mathbf{h}_k^o}) \end{aligned} \tag{2.87}$$

2.6.7 Performance Analysis of PQML

In [28] the performance of PQML for FIR channel is derived. By analogy, from (37) in [28] we can deduce the performance for CP (SCCP or OFDM) channel as follows:

$$J_2^{PQML} = \sum_k |\mathbf{x}_k|^2 G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k \tag{2.88}$$

$$J_1^{PQML} = \sigma_w^2 J_2^{PQML} + \sigma_w^4 D^{''} \tag{2.89}$$

Where $D^{''} = \sum_k G_k^H D_k^{''} G_k$ and $D_k^{''}(i,j) = \mathrm{tr}\left\{ \frac{\partial \mathbf{h}_k^{\perp H}}{\partial \mathbf{h}_{k,i}} P_{\mathbf{h}_k^o}^{\perp} \frac{\partial \mathbf{h}_k^{\perp}}{\partial \mathbf{h}_{k,j}} (\mathbf{h}_k^{oH} \mathbf{h}_k^o)^{-1} \right\}$. Thus $D_k^{''} = \frac{1}{||\mathbf{h}_k^o||^2} \left(I - \frac{\mathbf{h}_k^o \mathbf{h}_k^{oH}}{||\mathbf{h}_k^o||^2} \right) = J_k^{'}$. Hence $D^{''} = J^{'}$, where $J^{'}$ has been stated earlier in the WSSF case. It is worthy to note here that for spatio-temporally white noise unlike DIQML both versions of PQML (global denoising and per tone denoising) attain the same performance. Now we proceed with the error covariance matrix we get:

$$\begin{array}{rcl} C_{\tilde{\mathbf{h}}} & = & J_2^{PQML^+} \left[\sigma_w^2 J_2^{PQML^+} \sigma_w^4 D^{''} \right] J_2^{PQML^+} \\ C_{\tilde{\mathbf{h}}} & = & \sigma_w^2 J_2^{PQML^+} \; + \; \sigma_w^4 J_2^{PQML^+} J^{'} J_2^{PQML^+} \end{array} \tag{2.90}$$

Moreover, comparing (2.88) and (2.49) we have $CRB = \sigma_w^2 J_2^{PQML^+}$ hence,

$$C_{\tilde{\mathbf{h}}}^{PQML} = CRB + CRB \; J^{'} \; CRB = C_{\tilde{\mathbf{h}}}^{WSSF} = C_{\tilde{\mathbf{h}}}^{DML} \tag{2.91}$$

We conclude that asymptotically both PQML and WSSF attains the same performance as DML. Another important remark is that DML doesn't attain CRB even asymptotically (in the number of data). However, as SNR gets higher and higher then $CRB \; J^{'} \; CRB$, the second term in (2.91) which is proportional to σ_w^4 becomes negligible compared to the first term and consequently $C_{\tilde{\mathbf{h}}}^{DML} \longrightarrow CRB$. Moreover, it should be noted that DML and all its approximations has an important drawback. Since the symbols are considered as unknown parameters, it follows that the dimension of the parameter vector grows without bound with increasing N or M. For this reason consistent estimation of all model parameters is impossible. More precisely, the DML estimate of $\mathbf{h}$ is consistent, whereas the estimate of A is inconsistent. To conclude the analysis part, we show in Table 1 J_2 and J_1 of all algorithms.

	J_2	J_1
UWSSF	$\sum_k \frac{1}{\|\|\mathbf{h}_k^o\|\|^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$	$\sum_k \frac{(\|\mathbf{x}_k\|^2\|\|\mathbf{h}_k^o\|\|^2+\sigma_w^2)\sigma_w^2}{\|\mathbf{x}_k\|^4\|\|\mathbf{h}_k^o\|\|^6} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$
WSSF	$\sum_k \|\mathbf{x}_k\|^2 G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$	$\sigma_w^2 J_2^{WSSF} + \sigma_w^4 \sum_k \frac{1}{\|\|\mathbf{h}_k^o\|\|^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$
OWSSF	$\sum_k \frac{\|\mathbf{x}_k\|^4\|\|\mathbf{h}_k^o\|\|^2}{\|\mathbf{x}_k\|^2\|\|\mathbf{h}_k^o\|\|^2+\sigma_w^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$	$\sigma_w^2 J_2^{OWSSF}$
SRM	$\sum_k G_k^H \left(\|\mathbf{x}_k\|^2 D_k^{(1)} + \alpha\sigma_w^2 I\right) G_k$	$\sum_k G_k^H \left(\|\mathbf{x}_k\|^2\sigma_w^2 D_k^{(2)} + \sigma_w^4 D_k^{(3)}\right) G_k$ $+ \sum_k \sum_n G_k^H \left(\sigma_w^4 \alpha^2 \mathbf{h}_k^o \mathbf{h}_n^{oH}\right) G_n$
DIQML,Global	$\sum_k \|\mathbf{x}_k\|^2 G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$	$\sigma_w^2 J_2^{DIQML} + \sigma_w^4 (B - \frac{B\mathbf{h}^o\mathbf{h}^{oH}B}{\mathbf{h}^{oH}B\mathbf{h}^o})$
DIQML,per tone	$\sum_k \|\mathbf{x}_k\|^2 G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$	$\sum_k G_k^H \left\{\sigma_{w,k}^2 J_{2,k} + \sigma_{w,k}^4 (B_k - \frac{B_k\mathbf{h}_k^o\mathbf{h}_k^{oH}B_k}{\mathbf{h}_k^{oH}B_k\mathbf{h}_k^o})\right\} G_k$
PQML	$\sum_k \|\mathbf{x}_k\|^2 G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$	$\sigma_w^2 J_2^{PQML} + \sigma_w^4 \sum_k \frac{1}{\|\|\mathbf{h}_k^o\|\|^2} G_k^H P_{\mathbf{h}_k^o}^{\perp} G_k$

Table 2.1: A summary of performance analysis of all algorithms

2.7 Simulations

In all the simulations we run in the sequel, except when otherwise stated, we use the fading channel shown below which is composed of five taps and corresponds to 3 antennas at the receiver. On the other hand, each (OFDM or single-carrier) CP block transmission system is composed of either 128 or 32 tones and the symbols are drawn from a 16 QAM constellation. As for the analytical NMSE, it is computed as follows: $NMSE = \frac{\mathrm{tr}(C_{\tilde{\mathbf{h}}})}{||\mathbf{h}_o||^2}$ while that of simulation is given by: $NMSE = \frac{||\mathbf{h}^o - \mathbf{h}||^2}{||\mathbf{h}_o||^2}$. All the algorithms are initialized by SRM where the scalar ambiguity that results from the SRM is fixed using the solution stated in (2.52). However, in what follows the scalar ambiguity of all versions of SSF is fixed using the solution in (2.53) while for DIQML and PQML we use the solution in (2.52) because Q is singular. On the other hand, when M receiving blocks are used at the receiver to estimate the channel then the analytical formulas of both CRB and $C_{\tilde{\mathbf{h}}}$ are divided by M.

We start the simulation part by plotting in Fig. 2.1 the analytical NMSE of different SSF algorithms versus the SNR. The analytical NMSE is computed as shown above where $C_{\tilde{\mathbf{h}}}$ for USSF, WSSF and OWSSF algorithms are obtained from (2.67, 2.70 and 2.73) respectively. It can be observed that UWSSF attains inferior performance compared to the other versions of SSF. However, OWSSF outperforms WSSF algorithm only at very low SNR, while both attain CRB_{det} at high SNR. In Fig. 2.2 we show how the simulation results (markers) match the analytical ones (solid lines) for different SSF. In this case we run 500 Monte-Carlo simulations with different noise realizations. Knowing that the analytical results are valid only asymptotically, hence in the simulation we consider the channel to be fixed over

10 transmission blocks in order to be able to use a large data amount. It is clear from the Fig. that the simulated and analytical results for different versions of SSF are congruent except at low SNR where it seems we need to gather more data amount. In Fig. 2.3 we show the variation of NMSE of different SSF algorithms versus the number of iterations at SNR = 10 dB. We use 100 Rx blocks to estimate the channel and the results are averaged over 500 Monte-Carlo simulations. We can easily notice from the Fig. that the performance of UWSSF is inferior to that of SRM while both WSSF and OWSSF achieve the same performance and it is superior to that of SRM. However, for all SSF algorithms only one or two iterations are required for convergence to occur. Moreover, the performance of the simulation is only fraction of dB away from that of the analytical one. In Fig. 2.4 we show the variation of NMSE of different versions of DIQML and PQML algorithms versus the number of iterations at SNR = 10 dB. We use 10 Rx blocks to estimate the channel and the results are averaged over 100 Monte-Carlo simulations. Concerning DIQML, it is obvious that DIQML with denoising per tone outperforms DIQML with global denoising as explained previously. However for PQML with a finite amount of data the Fig. shows that PQML with global denoising outperforms that with denoising per tone. On the other hand, it is clear that one or two iterations are sufficient for both DIQML and PQML to converge to a position just fraction of dB away from the analytical performance. Moreover, since PQML attains asymptotically the same performance as DML we label the analytical result for PQML as DML. In Fig. 2.5 we repeat the same simulation but this time we increase the number of transmission blocks over which the channel considered fixed to 100. Doing so, we note that both DIQML and PQML attain exactly their corresponding analytical performance with one iteration. Moreover, It is worthy to note that the analytical performance of DIQML with per tone denoising as well as that of both versions of PQML attain that of the DML. To shed light on the superiority of denoising per tone over the global denoising for DIQML we show in Fig. 2.6 the simulated NMSE versus SNR for both versions of DIQML and PQML where the noise is temporally colored spatially white. The results are averaged over 100 Monte-Carlo simulations where in each simulation we use different Rayleigh channel realization. We can notice from the Figure how DIQML with per tone denoising outperforms considerably DIQML with global denoising at moderate SNR. Moreover, it is also clear that the performance of DIQML with per tone denoising is equivalent to that of PQML and once again we notice that there is no difference in this case between PQML with global or per tone denoising. In Fig. 2.7 we show how the analytical performance of

SRM (solid line) matches the simulation (markers) along the whole range of SNR. It should be noted here that the scalar ambiguities of SRM and IQML are fixed using the solution in (2.53) and as a consequence of that the NMSE levels off at low SNR. We average the result over 100 Monte-Carlo simulation where in each one we use different Rayleigh channel realization. We can also observe that SRM outperforms considerably IQML at low and moderate SNR while their performance are congruent at very high SNR.

$$\mathbf{h} = \begin{bmatrix} 0.814 + 1.096i & -1.027 - 0.389i & -0.257 - 0.832i & 1.165 + 0.696i & 0.486 - 0.018i \\ -0.713 + 0.659i & -1.178 - 0.057i & -0.044 + 1.355i & -0.802 - 0.444i & -0.445 + 0.187i \\ 0.966 + 0.103i & 0.478 + 0.724i & -0.943 + 0.645i & -1.202 + 0.321i & 1.523 + 0.244i \end{bmatrix} \tag{2.92}$$

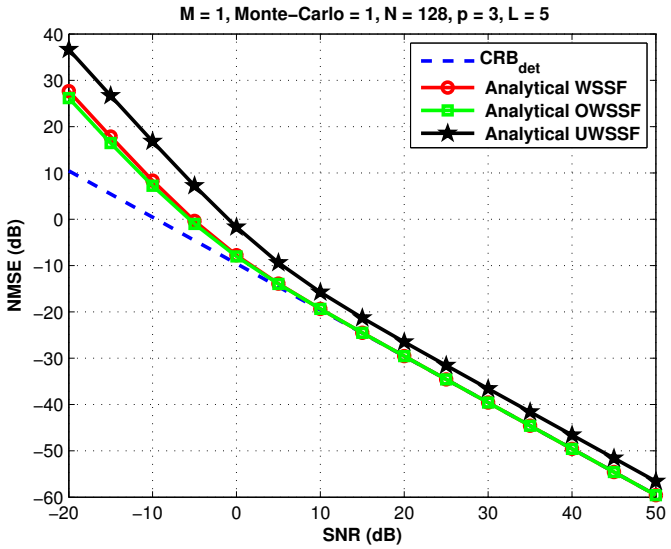

Figure 2.1: Analytical performance for different SSF algorithms.

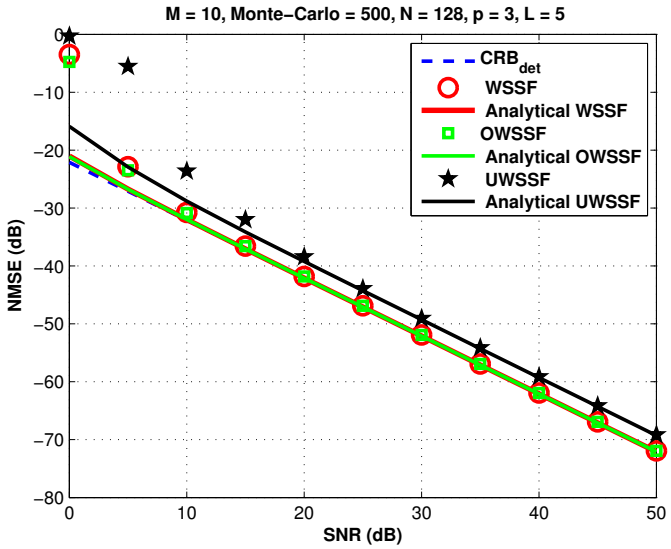

Figure 2.2: Analytical and simulated performance for different SSF algorithms.

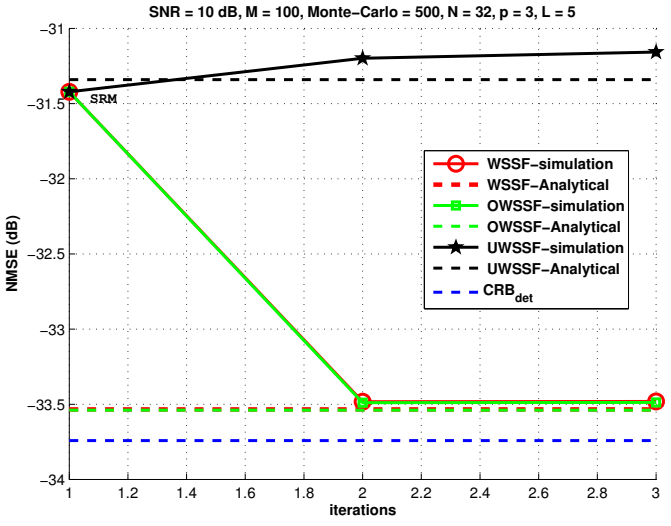

Figure 2.3: NMSE for SSF algorithms at different iterations

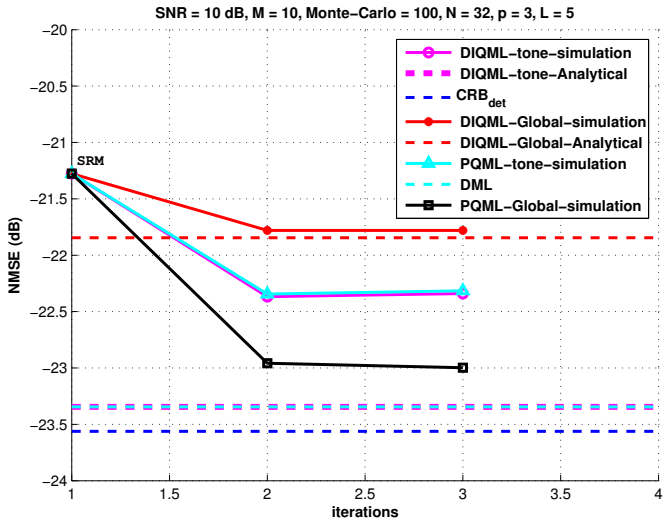

Figure 2.4: NMSE for different versions of DIQML and PQML algorithms at different iterations with small amount of data.

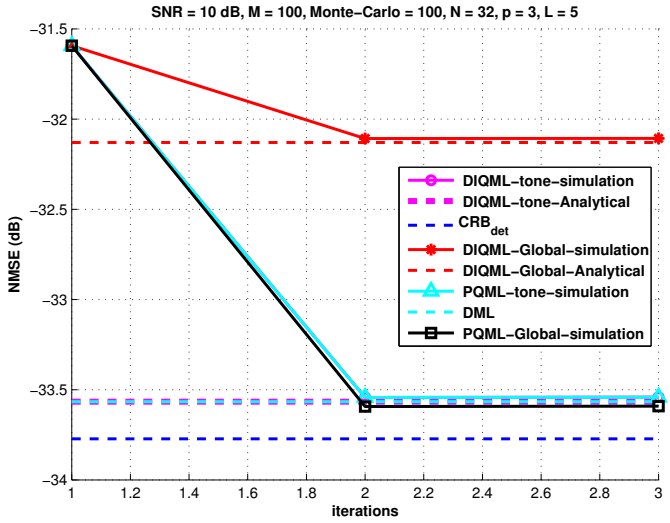

Figure 2.5: NMSE for different versions of DIQML and PQML algorithms at different iterations with large amount of data.

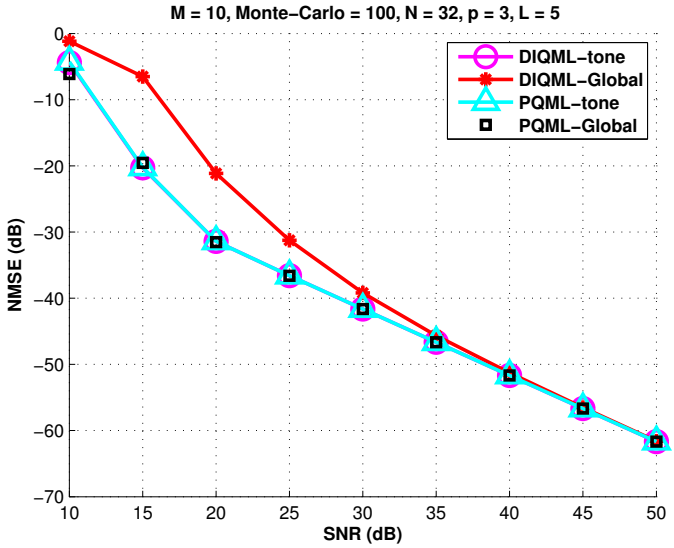

Figure 2.6: NMSE versus SNR for different versions of DIQML and PQML algorithms with spatially white temporally colored noise.

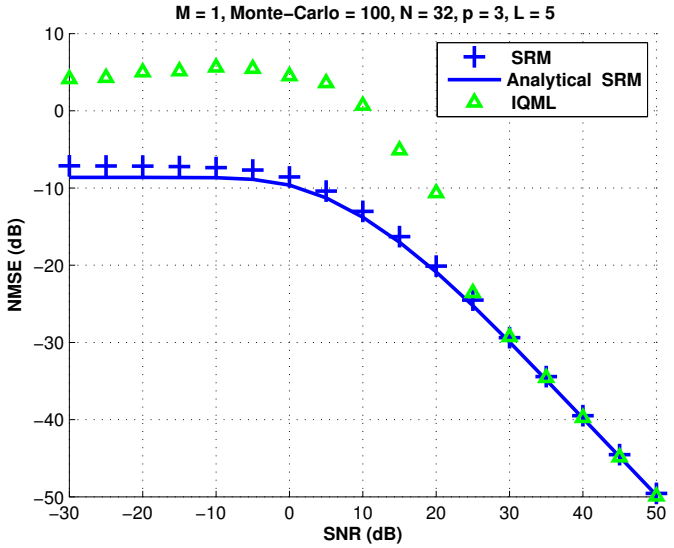

Figure 2.7: Comparison of analytical and simulated SRM with IQML.

Chapter 3

Spatio-Temporal Sample Covariance Matrix Enhancement

Multichannel aspect allows the introduction of blind channel estimation techniques. Most existing such techniques for frequency-selective channels are quite complex. In this chapter, we consider the blind channel estimation problem for Single Input Multi Output (SIMO) cyclic prefix (CP) systems. We have shown [39] that blind channel estimation becomes computationally much more attractive and more straight forward to analyze in terms of performance in CP systems. Inspired by the iterative sample covariance matrix (SCM) structure enhancement techniques of Cadzow and others [31], we develop in this chapter an algorithm to structure the sample block circulant covariance matrix by enforcing two essential properties: rank and FIR structure. These two properties are exhibited by the true covariance matrix in the case of FIR SIMO channels with spatially white noise and CP transmission. The novel enhancement procedure leads to an interesting enhanced SCM, even for the single CP symbol case. A simulation study for some classical channel estimators that depend on the SCM (with and without structuring) is presented, indicating that structuring allows for a considerable performance gain in terms of the channel normalized mean square error (NMSE) over a wide SNR range.

3.1 Introduction

A wealth of blind channel estimation techniques have been introduced for spatio-temporal channels over the past decade, based on the singularity of the received signal power spectral density matrix [4]. This singularity can be exploited to separate the white noise contribution. The main problem characteristic in fact that allows channel identifiability is the minimum phase characteristic of the Single-Input Multiple-Output (SIMO) or MIMO matrix channel transfer function of the spatio-temporal channel. Spatio-temporal channels arise in mobile communications when multiple antennas or polarizations or beams are used at the receiver. Physical multi-channels can also arise in xDSL systems when the receiver has access to a complete cable bundle. Other problem formulations that lead to multi-channel models are the use of oversampling at the receiver or the decoupling of inphase and inquadrature components when real symbols get modulated or the reception of multiple signal copies in ARQ protocols. A variety of blind symbol/channel estimation strategies can be developed depending on the amount of a priori information that gets formulated on the unknown symbols. In general, the less structure that gets exploited about the symbol alphabet, the less problems that tend to be encountered with local minima. Of course, more estimation accuracy is obtained by exploiting more information. A reasonable strategy is hence to exploit a progressive range of algorithms exploiting increasing a priori information levels. The algorithm at the next level can be initialized with the estimate obtained at the previous level of a priori information. The memory introduced by a convolutive channel leads to the requirement of having to treat all available data in a contiguous observation interval in one shot if no suboptimality is allowed. This leads to problem formulations with large convolution matrices, large covariance matrices and high complexity. Attempts have been made by our own group to introduce asymptotic approximations, by approximating large Toeplitz convolution matrices by circulant matrices, to allow transformation to the frequency domain, or by others by introducing approximate DFT operations. Cyclic prefixes have been introduced in a number of existing systems such as OFDM systems for ADSL and wireless LANs. Recently, Orthogonal Frequency Division Multiple Access (OFDMA) has been adopted as a multiple access scheme for the Frequency Domain Duplexing Down-Link (FDD-DL)in LTE (Long Term Evolution) systems. The introduction of a cyclic prefix renders the transformation to the frequency domain clean and exact even for a finite data length. The resulting algorithmic simplifications will be detailed for a number of classical blind channel estimation methods.

Furthermore, the same framework can be used to analyze the performance of the algorithms and the algorithmic simplifications also translate into much simplified performance expressions, which allow a direct and insightful analytical performance comparison between a number of algorithms.

This chapter is organized as follows. Section 2 starts with the description of the basic baseband SIMO-cyclic prefix model. In section 3 we develope a unified framework for cyclic prefix system channel estimators. Section 4 defines some classical blind channel estimators within the framework introduced in section 3. The algorithm for structuring the covariance matrix is developed in section 5. In section 6, we provide the experimental results and finally a conclusion is drawn in Section 7.

3.2 SIMO Cyclic Prefix Block TX Systems

We are using the same model used in chapter 2, section 2.2.

3.3 Frequency domain Framework for CIR Estimation

In what follows, we shall see that for various methods, we get a cost function or information at each tone for the channel response at that tone, and to get the cost function or information for the temporal channel response, it suffices to sum up the cost functions or informations over the tones after transforming back to the time domain. Now, if we follow the same steps mentioned in chapter 2, section 2.3 we will end up again with 2.8 which we will restate here:

$$\mathcal{F}(\mathbf{h}) = \mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \, Q_k \, G_k \right] \mathbf{h} \; = \; \mathbf{h}^H \, Q \, \mathbf{h} \tag{3.1}$$

and similarly for Fisher information matrices (FIM) as we will see in the sequel. So in what follows, we shall concentrate on the cost function for a given tone.

3.4 Blind SIMO Channel Estimation

3.4.1 Subchannel Response Matching (SRM)

This algorithm has been elaborated in detail in chapter 2, section 2.5.2. it is based on a linear parametrization of the noise subspace in terms of the

channel coefficients [54] so that $P^{\perp}_{\mathbf{h}_k} = P_{\mathbf{h}_k^{\perp H}}$ where $\mathbf{h}_k^{\perp}$ is given by 2.10. In the noiseless case, $\mathbf{u}_k = \mathbf{h}_k x_k$ and we have $\mathbf{h}_k^{\perp}\mathbf{u}_k = \mathcal{U}_k\mathbf{h}_k = 0$. Based on this relation the channel at tone k can be uniquely determined up to a scale factor [47],[42], as the unique right singular vector of $\mathcal{U}_k$ corresponding to the singular value zero. When noise is present, $\mathcal{U}_k\mathbf{h}_k \neq 0$ and the SRM criterion is solved in the least-squares sense $||\mathbf{h}_k^{\perp}\mathbf{u}_k||_2^2 = tr\{\mathbf{h}_k^{\perp}\mathbf{u}_k\mathbf{u}_k^H\mathbf{h}_k^{\perp H}\}$. By the law of large numbers, asymptotically this criterion can be replaced by its expected value: $tr\{\mathbf{h}_k^{\perp}S_{\mathbf{u_k u_k}}\mathbf{h}_k^{\perp H}\}$ where $S_{\mathbf{u_k u_k}} = \mathrm{E}\mathbf{u_k u_k^H}$. Practically, $S_{\mathbf{u_k u_k}}$ is not available so it is replaced by the sampled spectrum per each tone $\hat{S}_{\mathbf{u_k u_k}}$ which is computed directly from the Fourier transformed version of the received data as we will show in the next section. Moreover, the SRM criterion can be written in the form shown in (2.8) where $Q_k = \sum_{i=1}^{p} D_i\hat{S}^*_{\mathbf{u_k u_k}}D_i^H$ where $\mathcal{D}_{i+1} = \mathcal{C}\mathcal{D}_i\mathcal{C}$.

$$\mathcal{D}_{\mathbf{1}} = \begin{bmatrix} 0 & 1 & 0 & \cdots \\ -1 & 0 & \cdots & \\ 0 & \vdots & \ddots & \\ \vdots & & & \end{bmatrix}. \tag{3.2}$$

$$\mathcal{C} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & 0 & 1 & 0 \end{bmatrix}. \tag{3.3}$$

Then, we attempt to minimize the sum of the SRM criteria (cost functions) over all tones jointly to obtain an estimate of the channel impulse response $\mathbf{h}$. The SRM cost function is shown below:

$$\min_h \mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \left\{\sum_{i=1}^{p} D_i\hat{S}^*_{\mathbf{u_k u_k}}D_i^H\right\} G_k\right] \mathbf{h} \tag{3.4}$$

We denote the matrix between the braces in (3.4) by Q_{SRM}, it has the size of $pL \times pL$. The estimated channel impulse response $\hat{\mathbf{h}}_{SRM}$ that is obtained by solving (3.4) is the eigen vector that corresponds to the minimum eigen value of Q_{SRM}. This solution has a scalar ambiguity that can be solved by forcing a least square constraint as follows: $\min_{\alpha} ||\mathbf{h}^o - \alpha\hat{\mathbf{h}}_{SRM}||^2$. This yields the following solution:

$$\hat{\hat{\mathbf{h}}}_{SRM} = \frac{\hat{\mathbf{h}}_{SRM}^H\mathbf{h}^o}{||\hat{\mathbf{h}}_{SRM}||^2}\hat{\mathbf{h}}_{SRM} \tag{3.5}$$

3.4.2 Noise Subspace Fitting (NSF)

The sampled spectrum per each tone $\hat{S}_{\mathbf{u_k u_k}}$ can be decomposed into signal and noise subspace contributions:

$$\begin{aligned} \hat{S}_{\mathbf{u_k u_k}} &= \hat{S}_{\mathbf{u}_k,\mathcal{S}} + \hat{S}_{\mathbf{u}_k,\mathcal{N}} \\ &= \hat{E}_{s,k}\hat{\Lambda}_{s,k}\hat{E}_{s,k}^H + \hat{E}_{n,k}\hat{\Lambda}_{n,k}\hat{E}_{n,k}^H \end{aligned} \tag{3.6}$$

The basic idea of the NSF is to fit the estimated noise subspace that we obtain from the sampled spectrum to the true noise subspace which is spanned by the columns of $\mathbf{h}_k^{\perp H}$.

$$\min_{h_k,T} \; ||\mathbf{h}_k^{\perp H} - \hat{E}_{n,k}T||_F \tag{3.7}$$

where $||X||_F = tr\{X^H X\}$ and T is a square transformation matrix. This criterion differs from the original subspace fitting strategy proposed in [55], which would propose $\min\limits_{h_k,T} \; ||\hat{E}_{n,k} - \mathbf{h}_k^{\perp H}T||_F$ as criterion. We propose (3.7) because it leads to a simpler optimization problem. Both approaches can be made to be equivalent by the introduction of column space weighting. The cost function in (3.7) is separable. In particular, it is quadratic in T. Minimization w.r.t. T leads to $T = \hat{E}_{n,k}^H \mathbf{h}_k^{\perp H}$ and $\mathbf{h}_k^{\perp H} - \hat{E}_{n,k}\hat{E}_{n,k}^H\mathbf{h}_k^{\perp} = P_{\hat{E}_{n,k}}^{\perp}\mathbf{h}_k^{\perp H}$ where $P_{\hat{E}_{n,k}}^{\perp} = I - P_{\hat{E}_{n,k}} = P_{\hat{E}_{s,k}}$ and $P_{\hat{E}_{n,k}}, P_{\hat{E}_{s,k}}$ denote respectively the projection matrix on the noise subspace ($\hat{E}_{n,k}$) and the signal subspace ($\hat{E}_{s,k}$).

Hence,

$$\begin{aligned} \min_{h_k,T} \; ||\mathbf{h}_k^{\perp H} - \hat{E}_{n,k}T||_F &= \min_{h_k} \; ||P_{\hat{E}_{s,k}}\mathbf{h}_k^{\perp H}||_F \\ &= \min_{h_k} \; tr\{\mathbf{h}_k^{\perp}\hat{E}_{s,k}\hat{E}_{s,k}^H\mathbf{h}_k^{\perp H}\} \end{aligned} \tag{3.8}$$

Similar to the case of SRM, the NSF criterion can be written in the form shown in (2.8) where $Q_k = \sum_{i=1}^{p} D_i\hat{E}_{s,k}\hat{E}_{s,k}^H D_i^H$ and D_i is the same as for the SRM criterion. Again, we attempt to minimize the NSF jointly over all the tones subject to the least square constraint to avoid introducing N constraints and to exploit the correlation that exists between the different tones. Therefore, the NSF cost functions takes the following form:

$$\min_{h} \mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \left\{\sum_{i=1}^{p} D_i\hat{E}_{s,k}\hat{E}_{s,k}^H D_i^H\right\} G_k\right] \mathbf{h} \tag{3.9}$$

Following the same discussion as in case of SRM we get the following solution:

$$\hat{\hat{\mathbf{h}}}_{NSF} = \frac{\hat{\mathbf{h}}_{NSF}^{H}\mathbf{h}^{o}}{||\hat{\mathbf{h}}_{NSF}||^{2}}\hat{\mathbf{h}}_{NSF} \tag{3.10}$$

The substanial computational power saving offered by our farmework namely, working per tones instead of working in the time domain, is elucidated by remarking that we perform Eigen Value Decomposition (EVD) of N matrices $\hat{S}_{\mathbf{u_k u_k}}$ each of size $(p \times p)$, while working in the time domain requires the EVD of a huge matrix $\hat{S}_{YY}$ of size $(pN \times pN)$. Knowing that the number of operations required to perform the EVD of a matrix is proportional to the cubic of its size and the number of tones N for some systems (eg. LTE downlink) may reach to 2048, then the great advantage of our framework in terms of computational power saving becomes evident.

3.5 Block Toeplitz Covariance Matrix Enhancement

Here we go back to sample covariance refinements suggested by Cadzow in the eighties [31] and which we tried to exploit to enhance the dereverberation of the acoustic channel [30]. The idea is to iteratively reinforce several structural properties, the reinforcement of which consists of a projection onto a convex set. The iterations then converge to the joint reinforcement of all properties. Theoretically, the matrix valued vector signal spectrum is of the form

$$S_{\mathbf{uu}}(z) = \mathbf{h}(z)\, S_{xx}(z)\, \mathbf{h}^{\dagger}(z) + S_{\mathbf{ww}}(z) \tag{3.11}$$

where $.^{\dagger}$ denotes paraconjugate which is defined as $h^{\dagger}(z) = h(1/z^{*})^{H}$ and $S_{\mathbf{ww}}(z) = \sigma_w^2\, I$ is the white noise spectrum. Saking for the simplicity of notations, we omit the index k in $\mathbf{S_{uu}}$. The signal part of the spectrum, $\mathbf{h}(z)\, S_{xx}(z)\, \mathbf{h}^{\dagger}(z)$ is singular, not because of spectral poverty as in the SISO case, but because of limited rank in the matrix dimension. In the SISO case, a stationary signal covariance matrix can only be singular if the signal consists of a number of (complex) sinusoids, with their number being smaller than the covariance matrix dimension. Singularity in the SIMO case has nothing to do with spectral poverty but with matrix singularity of the matrix spectrum at every frequency.

Inspired by [31], (3.11) suggests the following procedure. First we start

with the sample spectrum at each tone:

$$\begin{aligned} &\widehat{S}_{\mathbf{uu}}(z_n) = \tfrac{1}{M}\textstyle\sum_{m=1}^{M} \mathbf{u}_n[m]\mathbf{u}_n^H[m] \ , \\ &n = 0, \ldots, N-1 \end{aligned} \tag{3.12}$$

where M is the number of OFDM symbols over which we compute the sample spectrum and $z_n = e^{j2\pi n/N}$, with the following properties: $\widehat{S}_{\mathbf{u}}^{\dagger}(z_n) = \widehat{S}_{\mathbf{u}}^{H}(z_n)$ (Hermitian transpose).

Now, at each frequency bin n, $S_{\mathbf{uu}}(z_n)$ is of the form

$$\begin{aligned} S_{\mathbf{uu}}(z_n) &= S_{\mathbf{u},\mathcal{S}}(z_n) + S_{\mathbf{u},\mathcal{N}}(z_n) \\ &= \mathbf{h}(z_n)\, S_x(z_n)\, \mathbf{h}^{\dagger}(z_n) + \sigma_w^2\, I_p \\ &= V_{max,n}\, (\lambda_{max,n} - \sigma_w^2)\, V_{max,n}^H + \sigma_w^2\, I_p \end{aligned} \tag{3.13}$$

where $S_{\mathbf{u},\mathcal{S}}(z_n)$, $S_{\mathbf{u},\mathcal{N}}(z_n)$ are the signal and noise components of $S_{\mathbf{u}}(z_n)$, and $\lambda_{max,n}$ and $V_{max,n}$ are its maximum eigenvalue and corresponding eigenvector. Now, the $\widehat{S}_{\mathbf{u}}(z_n)$ can be forced to the closest (in Frobenius norm) matrix of the form in (3.13) by computing its spatial eigen decomposition. Let $\widehat{\lambda}_{1,n} \geq \widehat{\lambda}_{2,n} \geq \cdots \geq \widehat{\lambda}_{p,n}$ be its eigenvalues, hence $\widehat{\lambda}_{max,n} = \widehat{\lambda}_{1,n}$, $\widehat{V}_{max,n} = \widehat{V}_{1,n}$. Then we get $\widehat{S}_{\mathbf{u}}(z_n) = \widehat{S}_{\mathbf{u},\mathcal{S}}(z_n) + \widehat{S}_{\mathbf{u},\mathcal{N}}(z_n) = \widehat{V}_{max,n}\, (\widehat{\lambda}_{max,n} - \widehat{\sigma}_v^2)\, \widehat{V}_{max,n}^H + \widehat{\sigma}_w^2\, I_p$ with $\widehat{\sigma}_w^2 = \frac{1}{N(p-1)} \sum_{n=0}^{N-1} \sum_{i=2}^{p} \widehat{\lambda}_{i,n}$ due to the spatio-temporal white noise assumption. Note that in fact at every frequency bin only $\lambda_{max,n}$ and $V_{max,n}$ need to be computed since $\sum_{i=2}^{M} \widehat{\lambda}_{i,n} = \mathrm{tr}\{\widehat{S}_{\mathbf{u}}(z_n)\} - \widehat{\lambda}_{max,n}$. Since the noise spectrum $\widehat{S}_{\mathbf{u},\mathcal{N}}(z_n) = \widehat{\sigma}_w^2\, I_M$ is fairly simple, there is no further structure to be imposed. The signal spectrum $\widehat{S}_{\mathbf{u},\mathcal{S}}(z_n) = \widehat{V}_{max,n}\, (\widehat{\lambda}_{max,n} - \widehat{\sigma}_w^2)\, \widehat{V}_{max,n}^H$ on the other hand, it is supposed to be spectrum of a FIR correlation sequence. This FIR characteristic can be imposed by windowing in the time domain. The resulting signal spectrum $\widehat{S}_{\mathbf{u},\mathcal{S}}(z_n)$ then undergoes IFFT to obtain the corresponding matrix correlation sequence. The frequency-wise rank structure enforcement will have destroyed the FIR character of the correlation sequence, which can then simply be enforced in the time domain by proper windowing (without forgetting the symmetry structure of the first block column of the block circulant matrix). The operations of eigen structure enforcement in frequency domain and FIR structure enforcement in the time domain can then be iterated untill convergence. Typically a few iterations suffice. We are now ready to state the following iterative process:

1. Compute the matrix spectrum $\widehat{S}_{\mathbf{u}}(z_n)$ at each frequency bin as illustrated in (3.12).

2. Compute the eigen decomposition of the spectrum $\widehat{S}_{\mathbf{u}}(z_n)$ at each frequency bin $n = 0, 1, \ldots, N-1$. Determine the noise variance $\widehat{\sigma}_w^2 = \frac{1}{N(p-1)} \sum_{n=0}^{N-1} \sum_{i=2}^{p} \widehat{\lambda}_{i,n}$ and the signal part of the spectrum $\widehat{S}_{\mathbf{u},\mathcal{S}}(z_n) = \widehat{V}_{max,n} \left(\widehat{\lambda}_{max,n} - \widehat{\sigma}_w^2\right) \widehat{V}_{max,n}^H$.

3. Compute the channel correlations:

$$\begin{bmatrix} \widehat{r}_{\mathbf{u}}(0) \\ \widehat{r}_{\mathbf{u}}(1) \\ \vdots \\ \widehat{r}_{\mathbf{u}}^H(1) \end{bmatrix} = \frac{1}{N}(F_N^* \otimes I_M) \begin{bmatrix} \widehat{S}_{\mathbf{u}}(z_0) \\ \widehat{S}_{\mathbf{u}}(z_1) \\ \vdots \\ \widehat{S}_{\mathbf{u}}(z_{N-1}) \end{bmatrix} \tag{3.14}$$

   Put the correlations outside the range $n \in \{0, 1, \ldots, L-1\}$ to zero to obtain the Hermitian of the following block row
   $[\widehat{r}_{\mathbf{u}}(0)\ \widehat{r}_{\mathbf{u}}^H(1) \cdots \widehat{r}_{\mathbf{u}}^H(L-1)\ 0\ \cdots\ 0\ \widehat{r}_{\mathbf{u}}(L-1) \cdots \widehat{r}_{\mathbf{u}}(1)]$.

4. Compute the spectrum of the thus windowed correlation sequence

$$\begin{bmatrix} \widehat{S}_{\mathbf{u}}(z_0) \\ \widehat{S}_{\mathbf{u}}(z_1) \\ \vdots \\ \widehat{S}_{\mathbf{u}}(z_{N-1}) \end{bmatrix} = (F_N \otimes I_p) \begin{bmatrix} \widehat{r}_{\mathbf{u}}(0) \\ \widehat{r}_{\mathbf{u}}(1) \\ \vdots \\ \widehat{r}_{\mathbf{u}}^H(1) \end{bmatrix} \tag{3.15}$$

   Go back to step 2 untill convergence. Note that the IFFTs and FFTs in (3.14) and (3.15) can be carried out efficiently in Matlab by reshaping the $N \times 1$ vectors of $p \times p$ blocks into $N \times p^2$ matrices.

After convergence, we make use of the refined spectrum we get at step (4) to get an enhanced channel impulse response estimation within the framework described in the previous section.

3.6 Experimental Results

We run our simulations within the framework of a SIMO-OFDM system where each OFDM symbol is composed of 128 tones. The performance of the different deterministic channel estimators (structured and non structured) are evaluated by means of the NMSE vs. SNR. The NMSE is defined as $\frac{||\mathbf{h}^0 - \hat{\mathbf{h}}||^2}{||\mathbf{h}^o||^2}$ and the SNR is defined as $\frac{\sum_{k=0}^{N-1} ||\mathbf{h}_k||^2 \sigma_{x_k}^2}{\sigma_w^2 pN}$. The symbols are drawn from QPSK constellation and the NMSE is averaged over 10000 Monte-Carlo runs of the noise, symbols and the channel. We consider Rayleigh fading

channel realizations where each one is composed of five i.i.d. channel coefficients. It is worthy to note that for SNR less than 20 dB the algorithm always converges typically after three or four iterations while at higher SNR the convergence is guaranteed at no more than ten iterations. However, we consider that convergence is achieved when the following condition is fulfilled: $\frac{\widehat{\sigma}^2_{w,i}-\widehat{\sigma}^2_{w,i-1}}{\widehat{\sigma}^2_{w,i}} \leq 0.1$ where i denotes the number of the current iteration at which the convergence is checked. Figure 3.1 shows the performance of both SRM and NSF with and without structuring where three antennas have been utilized at the receiver and the sampled spectrum has been computed from just one OFDM symbol. We remark that SRM yields better performance than NSF. This is due to the fact that when we work with one OFDM symbol then SRM is a weighted version of NSF with the weight being the largest eigenvalue of the sampled spectrum at each tone. However, when structuring is used, both estimators show at least 3 dB gain even at very high SNR. It is also obvious that after structuring the performance of both estimators is congruent whatever the SNR is. To elaborate more the advantage of our structuring algorithm we plot in Figure 3.2 the BER versus SNR where we have used the estimated channels by various algorithms to equalize the received signal using MMSE equalizer and a hard decision decoding to extract the received bits. This result shows that our algorithm outperforms the non-structured ones by more than 2 dB at BER = 10^{-2}.

3.7 Conclusions

To sum up, we have shown in this chapter the capability to exploit the classical blind deterministic channel estimators with a great computational power saving within the cyclic prefix systems. This is accomplished by minimizing the sum of the cost functions at different tones instead of minimizing the ordinary cost function in the time domain. Moreover, we propose a spatio-temporal based algorithm to enhance the sample covariance matrix upon which a class of well-known estimators rely. The enhancement is achieved by enforcing both the rank and the FIR structure properties. The numerical simulations show that the proposed algorithm has the potential to provide a 5 dB gain (in terms of NMSE) at low to moderate SNR while it still has the capability to provide a noticeable gain at high SNR.

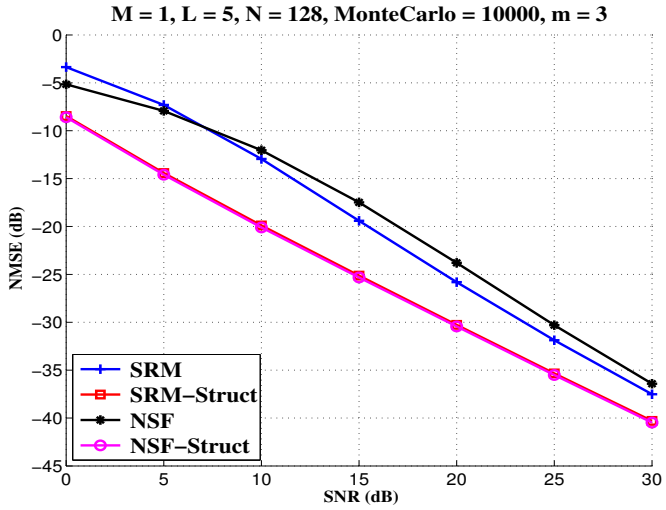

Figure 3.1: The NMSE versus SNR for structured and non-structured estimators.

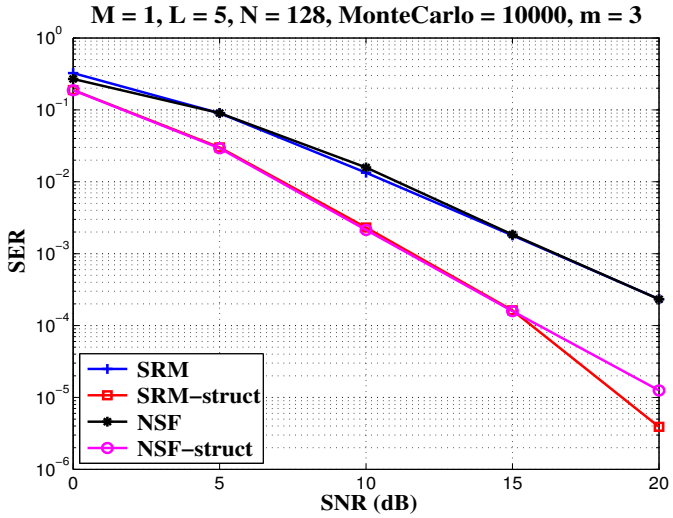

Figure 3.2: The BER versus SNR for structured and non-structured estimators.

Chapter 4

Variational Bayesian Blind and Semi-blind Channel Estimation

Blind and semi-blind channel estimation is a topic that enjoyed explosive developments throughout the nineties, and then came to a standstill, probably because of perceived unsatisfactory performance. Blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. Such blind channel estimates, especially those based on subspaces in the data, are often only partial and ill-conditioned. On the other hand, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. In recent years, such prior information on the channel has started to get exploited in pilot-based channel estimation, since often the pure pilot-based (deterministic) channel estimate is of limited quality due to limited pilots. In this chapter we explore a Bayesian approach to (semi-)blind channel estimation, exploiting a priori information on fading channels. In the case of deterministic unknown input symbols, it suffices to augment the classical blind (quadratic) channel criterion with a quadratic criterion reflecting the Rayleigh fading prior. In the case of a Gaussian symbol model the blind criterion is more involved. The joint ML/MAP estimation of chan-

nels, deterministic unknown symbols, and channel profile parameters can be conveniently carried out using Variational Bayesian techniques. Variational Bayesian techniques correspond to alternating maximization of a likelihood w.r.t. subsets of parameters, but taking into account the estimation errors on the other parameters. To simplify exposition, we elaborate the details for the case of MIMO OFDM systems.

4.1 Introduction

Blind and semi-blind channel estimation techniques have been developed and are usually evaluated for a given channel realization, i.e. with a deterministic channel model, see [18] for an overview of such techniques. Such blind channel estimates, especially those based on subspaces in the data, are often only partial and ill-conditioned. Indeed, only part of the channel is blindly identifiable, especially in the case of MIMO channels. The type of blind channel estimation techniques we are mostly referring to here involve an FIR multichannel and are typically based on the second-order statistics of the received signal. Two types of techniques can be considered, treating the unknown input symbols as either deterministic unknowns or Gaussian white noise. In the first case, the techniques are often based on the subspace structure induced in the data by the multichannel aspect. The part of the channel that can be identified blindly is larger in the Gaussian input model case than in the deterministic input model case, but is in any case incomplete. Many of the deterministic input approaches are also quite sensitive to a number of hypotheses such as correct channel length (filter order) and no channel zeros. In general this means that these blind channel estimates can often become ill-conditioned, when the channel impulse response is tapered (e.g. due to a pulse shape filter) or when the channel is close to having zeros. In fact this means that the blind information on the channel can be substantial, but is limited to only part of the channel.

An overview of blind channel estimation techniques can be found in [18] for SIMO systems and in [5] for MIMO systems. Specific blind channel estimation techniques for Cyclic Prefix systems, as will be considered here, were introduced in [39], see also [29]. The concept of Bayesian blind channel estimation was introduced in [56], with in particular some considerations on identifiability issues where-as in this chapter we focus on algorithms.

4.2 MIMO Cyclic Prefix Block TX Systems

Consider a MIMO system with q inputs x_l, $p > q$ outputs y_i per (symbol/sample) period. This is considered as an extension of the model used in chapter 2, section 3.2 from SISO to MIMO.

$$\begin{aligned} \underbrace{\mathbf{y}[m]}_{p\times 1} &= \sum_{l=1}^{q}\sum_{j=0}^{L} \underbrace{\mathbf{h}^l[j]}_{p\times 1}\,\underbrace{a_l[m-j]}_{1\times 1} + \underbrace{\mathbf{v}[m]}_{p\times 1} \\ &= \sum_{j=0}^{L} \underbrace{\mathbf{h}[j]}_{p\times q}\,\underbrace{\mathbf{a}[m-j]}_{q\times 1} + \underbrace{\mathbf{v}[m]}_{p\times 1} = \underbrace{H(q)}_{p\times q}\,\underbrace{\mathbf{a}[m]}_{q\times 1} + \underbrace{\mathbf{v}[m]}_{p\times 1} \end{aligned} \tag{4.1}$$

where $H(q) = \sum_{j=0}^{L} \mathbf{h}[j]\, q^{-j}$ is the MIMO system transfer function corresponding to the z transform of the impulse response $\mathbf{h}[.]$. Equation (4.1) mixes time domain and z transform domain notations to obtain a compact representation. In $H(q)$, z is replaced by q (not to be confused with the number of transmit antennas) to emphasize its function as an elementary time advance operator over one sample period. Its inverse corresponds to a delay over one sample period: $q^{-1}\mathbf{a}[n] = \mathbf{a}[n-1]$.

Consider a (OFDM or single-carrier) CP block transmission system with N samples per block. The introduction of a cyclic prefix of K samples means that the last K samples of the current block (corresponding to N samples) are repeated before the actual block. If we assume w.l.o.g. that the current block starts at time 0, then samples $\mathbf{a}[N-K]\cdots\mathbf{a}[N-1]$ are repeated at time instants $-K, \ldots, -1$. This means that the output at sample periods $0, \ldots, N-1$ can be written in matrix form as

$$\begin{bmatrix} \mathbf{y}[0] \\ \cdots \\ \mathbf{y}[N-1] \end{bmatrix} = \mathrm{Y}[0] = \mathrm{H}\ \mathrm{A}[0] + \mathrm{V}[0] \tag{4.2}$$

where the matrix H is not only (block) Toeplitz but even (block) circulant: each row is obtained by a cyclic shift to the right of the previous row. Consider now applying an N-point FFT to both sides of (4.2) at block m:

$$F_{N,p}\mathrm{Y}[m] = F_{N,p} H\, F_{N,q}^{-1}\ \ F_{N,q}\mathrm{A}[m] + F_{N,p}\mathrm{V}[m] \tag{4.3}$$

or with new notations:

$$\mathrm{U}[m] \quad = \quad \mathcal{H} \qquad \mathrm{X}[m] \quad + \quad \mathrm{W}[m] \tag{4.4}$$

where $F_{N,p} = F_N \otimes I_p$ (Kronecker product: $A \otimes B = [a_{ij}B]$), F_N is the N-point $N \times N$ DFT matrix, $\mathcal{H} = \text{diag}\{\mathbf{h}_0, \ldots, \mathbf{h}_{N-1}\}$ is a block diagonal matrix with diagonal blocks $\mathbf{h}_k = \sum_{l=0}^{L} \mathbf{h}[l]\ e^{-j2\pi \frac{1}{N} kl}$, the $p \times q$ channel transfer function at tone k (frequency $= k/N$ times the sample frequency). In OFDM, the transmitted symbols are in A$[m]$ and hence are in the frequency domain. The corresponding time domain samples are in X$[m]$. The OFDM symbol period index is m. In Single-Carrier (SC) CP systems, the transmitted symbols are in A$[m]$ and hence are in the time domain. The corresponding frequency domain data are in X$[m]$. The components of V are considered white noise, hence the components of W are white also. At tone (subcarrier) $n \in \{0, \ldots, N-1\}$ we get the following input-output relation

$$\underbrace{\mathbf{u}_n[m]}_{p\times 1} = \underbrace{\mathbf{h}_n}_{p\times q}\ \underbrace{\mathbf{x}_n[m]}_{q\times 1} + \underbrace{\mathbf{w}_n[m]}_{p\times 1} \tag{4.5}$$

where the symbol $x_n[m]$ belongs to some finite alphabet (constellation) in the case of OFDM.

4.3 Some Generalities for CP System Methods

The facts stated in chapter 2, section 3.2 are still applicable here. There is only one necessary modification namely, the channel per tone in this chapter is a matrix instead of being vector. Hence, it needs to be vectorized before commencing in the development of the our algorithms. To be a bit more explicit, let $\overline{\mathbf{h}}_k = \text{vec}\,(\mathbf{h}_k)$ and let $\mathbf{h}$ be the vectorized channel impulse response, i.e. $\mathbf{h} = \text{vec}\,([\mathbf{h}^H[0] \cdots \mathbf{h}^H[L]]^H)$. Then there exists transformation matrices G_k (containing DFT portions) such that

$$\overline{\mathbf{h}}_k \ = \ G_k\, \mathbf{h}\,. \tag{4.6}$$

Now, if at tone k we have a cost function of the form

$$\overline{\mathbf{h}}_k^H\ Q_k\ \overline{\mathbf{h}}_k \tag{4.7}$$

then this induces a cost function for the overall channel impulse response of the form

$$\mathbf{h}^H\ \left[\sum_{k=0}^{N-1} G_k^H\ Q_k\ G_k\right]\ \mathbf{h}\ =\ \mathbf{h}^H\ Q\ \mathbf{h} \tag{4.8}$$

and similarly for Fisher information matrices. So in what follows, we shall concentrate on the cost function for a given tone.

4.4 Bayesian Blind with Deterministic Symbols

Assume the Rayleigh fading channel has a prior distribution $\mathbf{h} \sim \mathcal{CN}(0, C^o_{\mathbf{h}})$, then a Bayesian blind criterion can be obtained straightforwardly by augmenting a classical blind criterion as follows

$$\mathbf{h}^H \, \frac{1}{\sigma_v^2 \, \|\mathbf{h}\|^2} \, Q \, \mathbf{h} + \mathbf{h}^H \, (C^o_{\mathbf{h}})^{-1} \, \mathbf{h} \tag{4.9}$$

which still remains (pseudo) quadratic in $\mathbf{h}$ ($\|\mathbf{h}\|^2$ refers to a separate estimate of $\|\mathbf{h}\|^2$). (4.9) would correspond to joint ML for the symbols and MAP for the channel if Q corresponds to one of the DML versions. Under a unit norm constraint, the minimization of (4.9) leads to
$\mathbf{h} = \|\mathbf{h}\| \, V_{min}(\frac{1}{\sigma_v^2 \, \|\mathbf{h}\|^2} \, Q + (C^o_{\mathbf{h}})^{-1})$ (which may need to be solved iteratively if Q depends on $\mathbf{h}$).

4.5 Gaussian Symbols Approaches

In Gaussian ML (GML), since both symbols and noise are Gaussian, the received signal is Gaussian with a channel-dependent covariance matrix. This leads to the GML likelihood, in which the symbols are eliminated. Alternatively, it is quite straightforward to add the Bayesian Rayleigh channel prior to the likelihood for the joint estimation of channel and Gaussian symbols, leading to joint MAP estimation of channel and symbols:

$$\frac{1}{\sigma_w^2} \|\mathrm{U} - \mathcal{H} \, \mathrm{X}\|^2 + \frac{1}{\sigma_a^2} \|\mathrm{X}\|^2 + \mathbf{h}^H \, (C^o_{\mathbf{h}})^{-1} \, \mathbf{h} \tag{4.10}$$

which is quadratic in $\mathbf{h}$ for fixed X, or in X for fixed $\mathbf{h}$. Knowing that (4.2) can be written in another form as follows:

$$\mathrm{Y} \; = \; \mathrm{H} \, \mathrm{A} + \mathrm{V} = \mathcal{A}\mathbf{h} + \mathrm{V} \tag{4.11}$$

where $\mathcal{A} = \mathcal{A}' \otimes \mathbf{I_p}$ and $\mathcal{A}'$ is a circulant matrix filled with the elements (vectors) of A. Then after some manipulations we get the following solutions:

$$\hat{\mathbf{X}} = \left(\widehat{\mathcal{H}}^H \widehat{\mathcal{H}} + \frac{\sigma_w^2}{\sigma_a^2} I \right)^{-1} \widehat{\mathcal{H}}^H \mathrm{U} \tag{4.12}$$

$$\hat{\mathbf{h}} = \left(\widehat{\mathcal{A}}^H \widehat{\mathcal{A}} + \sigma_w^2 (C^o_{\mathbf{h}})^{-1} \right)^{-1} \widehat{\mathcal{A}}^H F^H_{N,p} \mathrm{U} \tag{4.13}$$

Hence the Alternating GMAPGMAP (AGMAPGMAP) algorithm, by iteratively minimizing for $\mathbf{h}$ or X. Note that, although the (non-Gaussian) joint

posterior density of **h** and X would be difficult to compute, the joint posterior is not required to obtain their MAP estimate, which is fairly simple to compute (at least when done iteratively as suggested here).

4.6 Variational Bayesian Techniques

A recent tutorial on Variational Bayesian (VB) estimation techniques can be found in [57]. VB provides an approximate technique to determine the posterior probability density function (pdf) of the quantities to be estimated. Let θ denote the vector of all quantities to be estimated, including parameters and possibly signals (e.g. the "hidden variables" in EM terminology). And Y denotes the measurements. In many problems, the joint posterior pdf $f(\theta|Y)$ can be complicated to determine. Consider now a partition of θ into K subgroups of quantities that will get estimated per subgroup $\theta = \{\theta_k,\ k = 1, \ldots, K\}$. The idea of VB is to approximate $f(\theta|Y)$ by a product form $g(\theta|Y) = \prod_{k=1}^{K} g(\theta_k|Y)$ where the $g(\theta_k|Y)$ in general will differ from the true marginal pdfs $f(\theta_k|Y)$. The $g(\theta_k|Y)$ are determined by minimizing the Kullback-Leibler distance between $\prod_{k=1}^{K} g(\theta_k|Y)$ and $f(\theta|Y)$. This leads to the following implicit relations

$$\ln g(\theta_k|Y) = \mathrm{E}_{g(\theta_{\bar{k}}|Y)} \ln f(Y, \theta_k, \theta_{\bar{k}}) \ , k = 1, \ldots, K \tag{4.14}$$

where $\theta_{\bar{k}}$ is θ minus θ_k, hence $\theta = \{\theta_k, \theta_{\bar{k}}\}$. In practice, (4.14) needs to be solved iteratively by consecutively sweeping through $k = 1, \ldots, K$, at all times using for $g(\theta_{\bar{k}}|Y)$ the latest version available. This iterative process can be shown to converge montonically. Typically, when $f(Y|\theta)$ and the prior $f(\theta)$ are exponential family pdfs (typically Gaussian), then one can see from (4.14) that $g(\theta_k|Y)$ will also be of the exponential family. VB techniques have mainly been introduced to deal with hierarchical signal models: signals that contain parameters with a prior that depends itself on parameters with a prior etc. However, VB techniques can be useful in simpler problems also.

Note that Variational Bayesian techniques can also be applied in the presence of deterministic unknowns θ_D. There are two ways to think about deterministic unknowns:

(i) as truly deterministic, with prior $f(\theta_D) = \delta(\theta_D - \theta_D^o)$ where θ_D^o is the unknown true value of θ_D; in other words, $\theta_D \sim \mathcal{N}(\theta_D^o, C_{\theta_D}^o)$ where $C_{\theta_D}^o = 0\, I$.

(ii) as random with no prior information, hence $\theta_D \sim \mathcal{N}(\theta_D^o, C_{\theta_D}^o)$ where $C_{\theta_D}^o = \infty\, I$ and θ_D^o is unimportant (e.g. $\theta_D^o = 0$).

In case (i), VB becomes EM [57], in which case during the iterations the deterministic parameters are simply substituted by their current estimate. Case (ii) is closer to the VB spirit. If $\theta = \{\theta_D, \theta_S\}$ where θ_S are the stochastic parameters, then it suffices to replace $f(Y, \theta)$ in (4.14) by $f(Y, \theta_S|\theta_D) = f(Y|\theta)\, f(\theta_S)$. In this case also for the deterministic parameters not only their current estimates (posterior means) are accounted for but also their estimation error.

4.7 Variational Bayesian Blind Channel Estimation

We shall focus on the MIMO OFDM case with Rayleigh fading FIR channel and Gaussian symbol model. In this case $Y = \mathrm{U}$ and $\theta = \{\mathrm{X}, \mathbf{h}\}$. When applying the VB technique, $g(\mathrm{X}|\mathrm{U})$ factors as $g(\mathrm{X}|\mathrm{U}) = \prod_{n=1}^{N} g(\mathbf{x}_n|\mathrm{U})$. We have

$$\begin{aligned} \ln f(\mathbf{u}_n, \mathbf{x}_n, \mathbf{h}_n) = \mathrm{c^t} + \tfrac{1}{\sigma_w^2}\{ & -\mathbf{u}_n^H \mathbf{h}_n \mathbf{x}_n \\ & -\mathbf{x}_n^H \mathbf{h}_n^H \mathbf{u}_n + \mathbf{x}_n^H (\mathbf{h}_n^H \mathbf{h}_n + \tfrac{\sigma_w^2}{\sigma_x^2} I_q)\mathbf{x}_n\} \end{aligned} \tag{4.15}$$

where $\mathrm{c^t}$ denotes a constant. With (4.14) we hence get $g(\mathbf{x}_n|\mathrm{U}) \sim \mathcal{CN}(m_{\mathbf{x}_n}, C_{\mathbf{x}_n})$, where

$$\begin{aligned} C_{\mathbf{x}_n} &= \left(\tfrac{1}{\sigma_w^2}\mathrm{trb}\{R_{\overline{\mathbf{h}}_n}^T\} + \tfrac{1}{\sigma_x^2} I_q\right)^{-1} \\ m_{\mathbf{x}_n} &= \tfrac{1}{\sigma_w^2} C_{\mathbf{x}_n} m_{\mathbf{h}_n}^H \mathbf{u}_n \end{aligned} \tag{4.16}$$

where trb of a block matrix denotes a matrix obtained by taking trace of its blocks (e.g. $\mathbf{h}_n^H \mathbf{h}_n = \mathrm{trb}\{(\overline{\mathbf{h}}_n \overline{\mathbf{h}}_n^H)^T\}$),
$\overline{\mathbf{h}}_n = \mathrm{vec}\ (\mathbf{h}_n)$, $m_{\overline{\mathbf{h}}_n} = G_n m_{\mathbf{h}}$, $C_{\overline{\mathbf{h}}_n} = G_n\, C_{\mathbf{h}}\, G_n^H$ and $m_{\mathbf{h}_n} = \mathrm{unvec}\ \{m_{\overline{\mathbf{h}}_n}\}$. In general, $R_{\mathbf{x}} = m_{\mathbf{x}} m_{\mathbf{x}}^H + C_{\mathbf{x}}$. The estimation of the symbols can be seen to correspond to the output of a MMSE linear equalizer in which the channel is not just replaced by its estimate, but the channel estimation error is taken into account also. On the other hand,

$$\begin{aligned} \ln f(\mathrm{U}, \mathrm{X}, \mathbf{h}) = \mathrm{c^t} - \textstyle\sum_n \{ & \tfrac{1}{\sigma_w^2} \|\mathbf{u}_n - (\mathbf{x}_n^T \otimes I_p) G_n \mathbf{h}\|^2 \\ & + \mathbf{h}^H G_n^H (G_n C_{\mathbf{h}}^o G_n^H)^{-1} G_n \mathbf{h}\} \end{aligned} \tag{4.17}$$

using e.g. $\mathbf{h}_n \mathbf{x}_n = (\mathbf{x}_n^T \otimes I_p) G_n \mathbf{h}$.
Hence with (4.14), $g(\mathbf{h}|\mathrm{U}) \sim \mathcal{CN}(m_{\mathbf{h}}, C_{\mathbf{h}})$ where

$$\begin{aligned} C_{\mathbf{h}} &= \left(\textstyle\sum_n G_n^H \{\tfrac{1}{\sigma_w^2}\left(R_{\mathbf{x}_n}^* \otimes I_p\right)\} G_n + (C_{\mathbf{h}}^o)^{-1}\right)^{-1} \\ m_{\mathbf{h}} &= C_{\mathbf{h}}\ \tfrac{1}{\sigma_w^2} \textstyle\sum_n G_n^H (m_{\mathbf{x}_n}^* \otimes I_p) \mathbf{u}_n \end{aligned} \tag{4.18}$$

The estimation of the channel can be seen to correspond to a Bayesian MMSE estimate, using all symbols as pilots, but taking into account that they have estimation error also. Upon convergence, the posterior means $m_{\mathbf{x}_n}$ and $m_{\mathbf{h}}$ are both MAP and MMSE estimates (due to Gaussianity) according to $g(\theta|Y) = g(\mathbf{h}|\mathrm{U}) \prod_n g(\mathbf{x}_n|\mathrm{U})$. However, they are neither MAP nor MMSE estimates according to the true posterior $f(\theta|Y)$. But it is intuitively clear that they should be reasonable approximations of the true MMSE estimates, which contrasts with the true MAP estimates of the AGMAPGMAP algorithm.

Remarks

The case of deterministic symbols can be handled also by just setting $\sigma_x^2 = \infty$ in (4.15), (4.16).

The extension to semi-blind, with some symbols being pilots, hence being known exactly, is immediate. Their error covariance matrix will remain zero and their mean equals the pilot value.

Extensions of the methods presented can be considered to incorporate the estimation of e.g. $C_{\mathbf{h}}^o$ (or a set of parameters parameterizing $C_{\mathbf{h}}^o$), which would be especially meaningful in the scenario in which multiple realizations of $\mathbf{h}$ with the same $C_{\mathbf{h}}^o$ need to be estimated (e.g. in a sequence of OFDM symbols).

4.8 Identifiability Considerations

Consider the joint estimation of the transmitted symbols in time domain A and the collective channel impulse response coefficients $\mathbf{h}$, making up together the parameter vector $\boldsymbol{\theta} = \left[\mathrm{A}^H \ \mathbf{h}^H\right]^H$. Then the Fisher Information matrix (FIM) on the basis of the data Y in (4.11) alone is

$$J_{\mathrm{Y}} = \frac{1}{\sigma_v^2} \left[\mathrm{H}\ \mathcal{A}\right]^H \left[\mathrm{H}\ \mathcal{A}\right] \ . \tag{4.19}$$

$\boldsymbol{\theta}$ is unidentifiable since indeed for a $q \times q$ mixing filter $\boldsymbol{\psi}(z)$, we have $H(q)\,\mathbf{a}[m] = (H(q)\boldsymbol{\psi}(q))\,(\boldsymbol{\psi}^{-1}(q)\,\mathbf{a}[m]) = \widetilde{H}(q)\,\widetilde{\mathbf{a}}[m]$. Hence it is impossible to distinguish $H(q)$ from $\widetilde{H}(q)$. If the delay spread of $H(q)$ is known and/or there are border conditions on the transmitted signal, then the frequency-selective mixture $\boldsymbol{\psi}(z)$ becomes a frequency-flat $\boldsymbol{\psi}$. In this case, J_{Y} has a null space which is the column space of $\left[\mathrm{A}^H \ -\mathbf{h}^H\right]^H$. Indeed $\left[\mathrm{H}\ \mathcal{A}\right] \left[\mathrm{A}^H \ -\mathbf{h}^H\right]^H = 0$. So the multiplicative ambiguity $\boldsymbol{\psi}$ translates into an (additive) singularity in the FIM.

In the case of Gaussian white symbols, the prior information on A translates into an additional FIM

$$J_{\mathrm{A}} = \frac{1}{\sigma_a^2} \left[\begin{array}{cc} I & 0 \\ 0 & 0 \end{array} \right] \tag{4.20}$$

so at this point the overall FIM is $J_{\boldsymbol{\theta}} = J_{\mathrm{Y}} + J_{\mathrm{A}}$ which will have become non-singular. This would indicate identifiability. The ambiguity in this case is indeed reduced from an unconstrained $\boldsymbol{\psi}$ to a unitary $\boldsymbol{\psi}$. However, there is still ambiguity and hence unidentifiability. Actually, the proper treatment with Gaussian symbols does not allow presenting the FIM in the compact complex form presented here. In fact, $\boldsymbol{\theta}$ needs to be doubled in size by considering separately its real and imaginary components and the associated FIM needs to be considered, in order to see the FIM nullspace corresponding to a unitary ambiguity matrix.

When now furthermore (or alternatively) a Gaussian prior for the channel $\mathbf{h}$ is considered, then the FIM for $\boldsymbol{\theta}$ gets augmented with

$$J_{\mathbf{h}} = \left[\begin{array}{cc} 0 & 0 \\ 0 & (C_{\mathbf{h}}^o)^{-1} \end{array} \right] \tag{4.21}$$

which will again render the overall FIM $J_{\boldsymbol{\theta}} = J_{\mathrm{Y}} + J_{\mathbf{h}}(+J_{\mathrm{A}})$ nonsingular. So it would seem that the addition of prior information with an identical non-zero Power Delay Profile (PDP) for each of the antenna pair channels (corresponding to a nonsingular diagonal $C_{\mathbf{h}}^o$) renders $J_{\boldsymbol{\theta}}$ nonsingular and hence leads to (channel) identifiability. However this is not necessarily the case. In the case of Gaussian white symbols, and a unitary ambiguity matrix $\boldsymbol{\psi}$, if $C_{\mathbf{h}}^o$ is such that $(\boldsymbol{\psi} \otimes I_{p(L+1)})^H C_{\mathbf{h}}^o (\boldsymbol{\psi} \otimes I_{p(L+1)}) = C_{\mathbf{h}}^o$ (in which case the channel prior is insensitive to a unitary mixture), then still the Bayesian blind problem remains unidentifiable. The above condition occurs if $C_{\mathbf{h}}^o$ is of the form $C_{\mathbf{h}}^o = I_q \otimes C$ for any square matrix C of size $p(L+1)$. Hence the regularization of the blind channel estimation problem via prior information is a tricky issue due to the multiplicative nature of the ambiguity.

4.9 Simulations

We simulate in this section both Bayesian and Variational Bayesian Blind (VBB) channel estimation techniques based on (4.12), (4.13) and (4.16), (4.18) appearing above. Moreover we simulate also a version of the Variational Bayesian approach where the channel parameters are deterministic unknowns, treated as random with no prior information, so $C_{\mathbf{h}}^o = \infty I$ in

(4.18). We shall refer to this approach as Uniformed VBB (UVBB). In each MonteCarlo simulation we generate a Rayleigh fading channel with exponentially decaying power delay profile (PDP) for the channel between each transmitting and receiving antenna pair as follows: e^{-wn} where $n = 0 : L$ and $w = 2$ normally. Hence, $C^o_{\mathbf{h}}$ is the diagonal matrix $C^o_{\mathbf{h}} = I_q \otimes C \otimes I_p$ where $C = \text{ diag}\{e^{-wn},\, n = 0 : L\}$. As for the symbols, we generate i.i.d. Gaussian symbols (which are hence i.i.d. Gaussian both in time and frequency domain). The performance of the different Bayesian channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The per receive antenna SNR is $\text{SNR} = \frac{\sigma_a^2 \text{tr}\{C^o_{\mathbf{h}}\}}{p\,\sigma_v^2}$.

The NMSE is defined as $\frac{avg\,||\mathbf{h}-\widehat{\widehat{\mathbf{h}}}||^2}{avg\,||\mathbf{h}||^2}$ where $\widehat{\widehat{\mathbf{h}}} = \widehat{\mathbf{h}}\boldsymbol{\psi}$ is the channel estimate adjusted for blind channel estimation ambiguities. As we assume the channel length known here, ψ represents an instantaneous mixing matrix of size $(q \times q)$. The mixing matrix $\boldsymbol{\psi}$ can be obtained by minimizing the Frobenius norm of the following matrix error: $\min\limits_{\boldsymbol{\psi}} ||\mathbf{h}' - \widehat{\mathbf{h}}'||_F^2$ where $\mathbf{h}' = (\mathbf{h}^H[0] \cdots \mathbf{h}^H[L])^H$ and $\widehat{\mathbf{h}}' = (\widehat{\mathbf{h}}^H[0] \cdots \widehat{\mathbf{h}}^H[L])^H$. For an unconstrained mixture, we get $\boldsymbol{\psi} = (\widehat{\mathbf{h}}'^H \widehat{\mathbf{h}}')^{-1} \widehat{\mathbf{h}}'^H \mathbf{h}' = U \Sigma V^H$ where the last expression represents the SVD of the resulting $\boldsymbol{\psi}$. In the case that $\boldsymbol{\psi}$ gets constrained to be a unitary matrix, the solution is $\boldsymbol{\psi} = UV^H$, see [5].

Both (B and VBB) algorithms are initialized by using (for $m_{\overline{\mathbf{h}}_n}$) noisy perturbations of the true channels. In the first iteration of (4.16) we use $R_{\overline{\mathbf{h}}_n} = m_{\overline{\mathbf{h}}_n} m^H_{\overline{\mathbf{h}}_n}$, hence $C_{\overline{\mathbf{h}}_n} = 0$. In Figure 4.1 we can notice how close the performance of both the Variational Bayesian and the Bayesian algorithms is since both fully exploit the prior information that exists about the channel and the symbols. However, we can notice that the UVBB method (with + marker, also called "Deterministic" in the legend) lags behind the normal Variational Bayesian (with * marker) where the prior information is taken into consideration. This is an expected result since the more information we exploit the better performance we get. However, at higher SNR the performance of the deterministic blind algorithm converges to that of the Bayesian blind algorithms. Also this result is expected since at very high SNR the contribution of prior information becomes negligible.

Whereas Fig. 4.1 uses a unitary $\boldsymbol{\psi}$, Fig. 4.2 uses an unconstrained $\boldsymbol{\psi}$, which leads to reduced NMSE since more prior information is exploited. At least, Fig. 4.1 shows that the exploitation of the white character of the symbols as we do here leads to a reduced unidentifiability of $\boldsymbol{\psi}$ to just a unitary mixing matrix.

In Fig. 4.3 the OFDM block length N gets increased from 20 (as in the

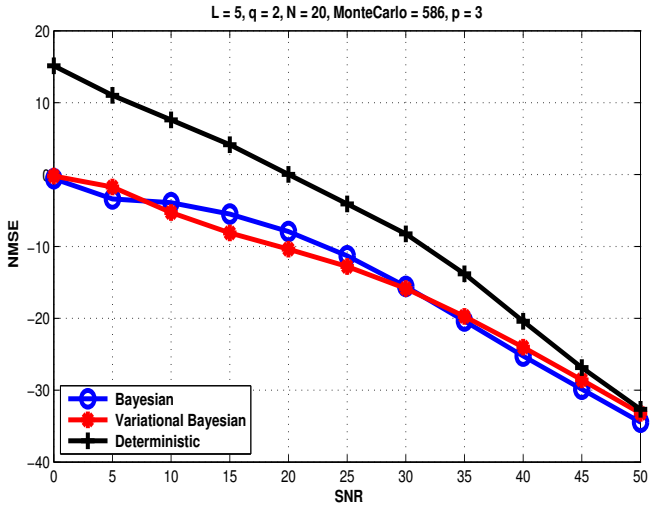

Figure 4.1: NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 20$, unitary ψ.

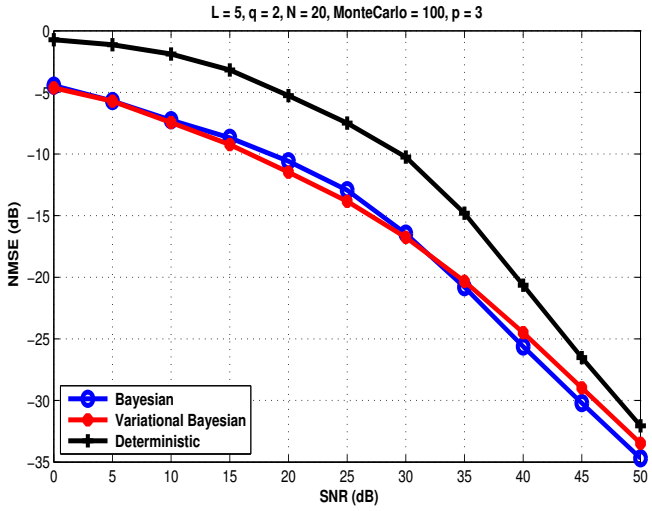

Figure 4.2: NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 20$, unconstrained ψ.

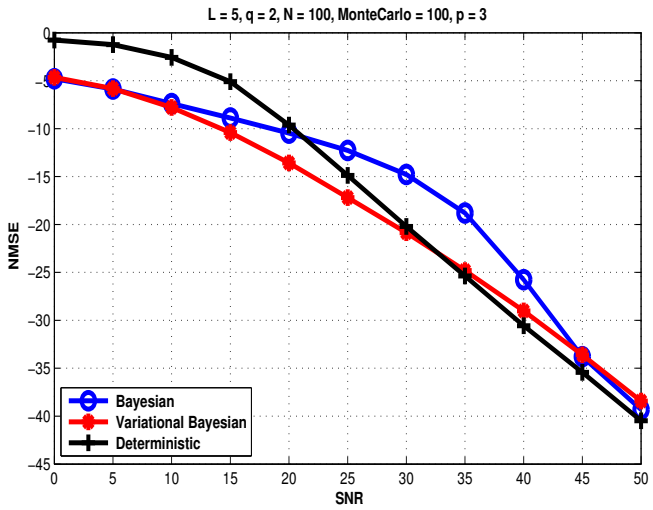

Figure 4.3: NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 100$, unconstrained $\boldsymbol{\psi}$.

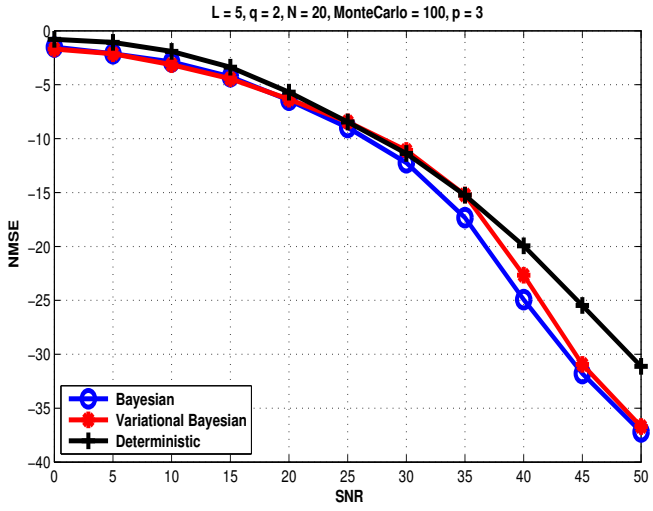

Figure 4.4: NMSE vs. SNR for B, VBB, UVBB algorithms, for $N = 20$, unconstrained $\boldsymbol{\psi}$, $w = 0.5$ in PDP.

previous two figures) to 100. The result is that the prior information introduced by a Bayesian approach only helps at low SNR, as could be expected. The other noticeable effect is that the Variational approach outperforms the non-Variational version over a wide SNR range.

In Fig. 4.4 finally, an exponential PDP with much shorter time constant ($w = 0.5$) is used, as compared to $w = 2$ in the previous three figures. The result is that the prior information only helps at very high SNR.

Part II

SIMO FIR Systems

Chapter 5

Bayesian Blind FIR Channel Estimation Algorithms

Blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. On the other hand, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. In this chapter, we explore a Bayesian approach to blind channel estimation, exploiting a priori information on fading channels. We mainly focus on joint ML/MAP estimation of channels and symbols on one hand, and on ML/MAP estimation of channels with elimination of symbols on the other hand. As a consequence, a unified framework in addition to three new Bayesian estimators are introduced where their performance is compared by simulations to three existing non-Bayesian estimators. In the same context, we provide an insightful discussion of the accurate way of deriving the Bayesian Cramer Rao bound (BCRB) with an emphasis on its singularity.

5.1 Introduction

In the context of blind channel estimation, our main goal is to estimate the channel at the receiver and feed it to a certain equalizer (Linear, Decision Feedback ..) to detect the symbols. To accomplish this task blindly at the

receiver we should exploit every piece of information related to any element of the transmission system. Moreover, sometimes assumptions are made and consequently the accuracy of the estimated parameters depends on how close those assumptions are to the reality. We will focus in this chapter on the second order statistics and specifically on the maximum likelihood (ML) and/or maximum a posteriori methods (MAP). Two approaches exist in the literature on how to tackle the problem. The first approach is based on the fact that the symbols are considered as deterministic unknowns to be jointly estimated with the channel. Such algorithm is called Deterministic (or conditional) Maximum Likelihood (DML) method [58]. The second approach is based on treating the symbols as random quantities with known prior information to be eliminated or jointly estimated. When the symbols are eliminated, the method is called Gaussian (or unconditional) Maximum Likelihood (GML) [59] see also [60] for its implementation in sensor array processing. While when they are jointly estimated, the method is called GMAP-ML [61]. This is because we use maximum a posteriori (MAP) for symbols and ML for channels and noise variance. Furthermore, in all these approaches the channel was considered as deterministic unknown however, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. The concept of Bayesian blind channel estimation was introduced in [56], with in particular some considerations on identifiability issues. However, in chapter 4 we discussed briefly some classical Bayesian algorithms and introduced the concept of variational Bayesian in the context of MIMO OFDM. Apart from the variational Bayesian techniques, we develop in this chapter classical Bayesian algorithms that treat the channel as random with known prior information rather than deterministic. Once the channel is treated as random, we are within the framework of Bayesian blind channel estimation and there are three cases to be handled. In the first case the symbols are considered as deterministic unknowns to be jointly estimated with the channel. We call this method as ML-GMAP, for a similar reasoning as discussed above. In the second case, the symbols are again to be jointly estimated with the channel but this time they are considered as random with known prior Gaussian distribution. We call this method GMAP-GMAP. In the third case, the symbols are again random with known prior Gaussian distribution but they are going to be eliminated rather than estimated. We call this method GMAP-Elm-GMAP where "Elm" stands for elimination of symbols. Consequently, in section 3 we revisit three already existing deterministic estimators and develop three new Bayesian ones. Therefore, with the introduction of the Bayesian blind

channel estimation algorithms the picture is broadened considerably and to sum up we depict the current picture in Table 1.

Algorithm	Symbols	Channel	Elm of Sym	Novel
ML-ML (DML)	Det	Det	No	No
GMAP-ML	Gaussian	Det	No	No
GMAP-Elm-ML	Gaussian	Det	Yes	No
ML-GMAP	Det	Gaussian	No	Yes
GMAP-GMAP	Gaussian	Gaussian	No	Yes
GMAP-Elm-GMAP	Gaussian	Gaussian	Yes	Yes

Table 5.1: Summary of all algorithms

5.2 SIMO FIR Tx System Model

In blind channel identification, a multichannel framework can be obtained from oversampling a received signal and leads to a Single Input Multiple Output (SIMO) vector channel representation. The multiple FIR channels we obtain in this representation can also be obtained from multiple received signals from an array of antennas or from a combination of both. To further develop the case of oversampling, consider a linear digital modulation over a linear channel with additive noise so that the received signal $y(t)$ has the following form

$$y(t) \;=\; \sum_{k} h(t-kT)a(k) \;+\; v(t). \tag{5.1}$$

In (5.1) $a(k)$ are the transmitted symbols, T is the symbol period and $h(t)$ is the channel impulse response. The channel is assumed to be FIR with length LT. If the received signal is oversampled at the rate $\frac{p}{T}$ (or if p different samples of the received signal are captured by p sensors every T seconds, or a combination of both), the discrete input-output relationship can be written as:

$$\boldsymbol{y}(k) = \sum_{i=0}^{L-1} \boldsymbol{h}(i)a(k-i) + \boldsymbol{v}(k) = \boldsymbol{H}A_L(k) + \boldsymbol{v}(k) \tag{5.2}$$

where $\boldsymbol{y}(k) = [y_1^H(k)\cdots y_p^H(k)]^H, \boldsymbol{h}(i) = \left[h_1^H(i)\cdots h_p^H(i)\right]^H$, with $h_p(i)$ denotes the ith tap of the pth receiving antenna, $\boldsymbol{v}(k) = [v_1^H(k)\cdots v_p^H(k)]^H$,

$\boldsymbol{H} = [\boldsymbol{h}(L-1)\cdots\boldsymbol{h}(0)]$, $A_L(k) = \left[a(k-L+1)^H\cdots a(k)^H\right]^H$ and superscript ^{H} denotes Hermitian transpose. Let $\mathbf{H}(z) = \sum_{i=0}^{L-1}\boldsymbol{h}(i)z^{-i} = [\mathrm{H}_1^H(z)\cdots\mathrm{H}_p^H(z)]^H$ be the SIMO channel transfer function, and $\boldsymbol{h} = \left[\boldsymbol{h}^H(L-1)\cdots\boldsymbol{h}^H(0)\right]^H$. Consider additive independent white Gaussian circular noise $\boldsymbol{v}(k)$ with $r_{\boldsymbol{v}\boldsymbol{v}}(k-i) = \mathrm{E}\,\boldsymbol{v}(k)\boldsymbol{v}(i)^H = \sigma_v^2 I_p\,\delta_{ki}$. Assume we receive M samples:

$$\boldsymbol{Y}_M(k) \;=\; \mathcal{T}_M(\boldsymbol{h})\,A_{M+L-1}(k) + \boldsymbol{V}_M(k) \tag{5.3}$$

where $\boldsymbol{Y}_M(k) = [\boldsymbol{y}^H(k-M+1)\cdots\boldsymbol{y}^H(k)]^H$ and similarly for $\boldsymbol{V}_M(k)$, and $\mathcal{T}_M(\boldsymbol{h})$ is a block Toeplitz matrix with M block rows and $[\boldsymbol{H}\;\; 0_{p\times(M-1)}]$ as first block row. We shall simplify the notation in (5.3) with $k = M-1$ to

$$\boldsymbol{Y} \;=\; \mathcal{T}(\boldsymbol{h})\,A + \boldsymbol{V} = \mathcal{A}\boldsymbol{h} + \boldsymbol{V}\;. \tag{5.4}$$

where $\mathcal{A}$ is a block Toeplitz matrix filled with the elements of A. We assume that $pM > M+L-1$ in which case the channel convolution matrix $\mathcal{T}(\boldsymbol{h})$ has more rows than columns. If the $\mathrm{H}_i(z)$, $i = 1,\ldots,p$ have no zeros in common, then $\mathcal{T}(\boldsymbol{h})$ has full column rank.

5.3 A Unified Framework for different Algorithms

As we have shown in Table 1 there are six possible estimators that can be classified into two categories. In the first category the subject of the estimators is to estimate the channel and the symbols jointly by making some assumptions on the channel and the symbols. If we denote by θ the unknown parameters to be estimated then it is given by:

$$\theta = [A^H, \boldsymbol{h}^H]^H \tag{5.5}$$

The likelihood function is given by:

$$f(Y,\theta) = f(Y/\theta)f(\theta) \tag{5.6}$$

where $f(\theta)$ stands for the probability density function (pdf) of θ, $f(Y,\theta)$ stands for the joint probability density function of Y and θ and $f(Y/\theta)$ stands for the pdf of Y conditioned on θ is given or known. Once we substitute θ in (5.6) by its elements we get:

$$f(Y,A,\boldsymbol{h}) = f(Y/A,\boldsymbol{h})f(A)f(\boldsymbol{h}) \tag{5.7}$$

Since the symbols and the channel are a priori independent of each other we can write $f(\theta) = f(A)f(\boldsymbol{h})$. Of course on the basis of how we treat the

symbols and the channel both $f(A)$ and $f(\boldsymbol{h})$ differ from one estimator to another as we shall see in the sequel. Knowing that the cost function for the estimator is derived by maximizing the log-likelihood function, hence we apply the log function on both sides of (5.7) to get:

$$\ln[f(Y, A, \boldsymbol{h})] = \ln[f(Y/A, \boldsymbol{h})] + \ln[f(A)] + \ln[f(\boldsymbol{h})] \tag{5.8}$$

However, in the second category the subject of the estimators is to estimate the channel and the noise variance only while the symbols are supposed to be eliminated during the estimation process. Thus

$$\theta = [\boldsymbol{h}^H, \sigma_v^2]^H \tag{5.9}$$

Once we substitute θ in (5.6) by its elements we get:

$$f(Y, \boldsymbol{h}, \sigma_v^2) = f(Y/\boldsymbol{h}, \sigma_v^2) f(\boldsymbol{h}) f(\sigma_v^2) \tag{5.10}$$

Again, since the cost function for the estimator is derived by maximizing the log-likelihood function, hence we apply the log function on both sides of (5.10) to get:

$$\ln[f(Y, \boldsymbol{h}, \sigma_v^2)] = \ln[f(Y/\boldsymbol{h}, \sigma_v^2)] + \ln[f(\boldsymbol{h})] + \ln[f(\sigma_v^2)] \tag{5.11}$$

We will develop in the following sections the cost functions of all the estimators that belong to both categories and provide a closed form formula for both the estimated channel and symbols where it is possible. It is worthy to note here that since the channel is treated as random rather than deterministic in some of the above mentioned estimators (ML-GMAP, GMAP-GMAP, GMAP-Elm-GMAP) in both categories, these estimators are considered as an example of the Bayesian blind channel estimation.

5.3.1 ML-ML (DML)

We start with ML-ML or what is called DML in the literature [58]. In this case both the symbols and the channel are considered as deterministic unknowns to be estimated. Hence it belongs to the first category and consequently the log-likelihood function is given by (5.8). Moreover, since both are deterministic we have $f(\boldsymbol{h}) = \boldsymbol{h}^o \delta(\boldsymbol{h} - \boldsymbol{h}^o)$ and $f(A) = A^o \delta(A - A^o)$ where $\boldsymbol{h}^o$ and A^o represent respectively the true values of the channel and the symbols. It is obvious that the pdfs of both symbols and channel have no influence on the maximization of (5.8). Hence,

we can derive the cost function by maximizing $\ln[f(Y/A, \boldsymbol{h})]$ directly where $f(Y/A, \boldsymbol{h}) = \frac{1}{(\pi\sigma_v^2)^{Mp}} exp[-\frac{1}{\sigma_v^2}(Y - \mathcal{T}(h)A)^H(Y - \mathcal{T}(h)A)]$. Thus the cost function is given by:

$$\min_{A,h} \; ||Y - \mathcal{T}(h)A||^2 \tag{5.12}$$

The joint optimization of this cost function in both the channel ($\boldsymbol{h}$) and the symbols (A) is difficult. Fortunately, the observation is linear in both the channel and the symbols. In other words, we have a separable nonlinear LS problem, which allows us to reduce the complexity considerably. The nonlinear LS optimization can be done by iterating between minimization with respect to A and $\boldsymbol{h}$. The alternating minimization strategy is as follows:

1. Initialization $\widehat{\boldsymbol{h}}^0$.

2. Iteration (i+1):

   - Minimization w.r.t. A; $\boldsymbol{h} = \widehat{\boldsymbol{h}}^{(i)}$:
     $\min_{A} \; ||Y - \mathcal{T}(\hat{h}^{(i)})A||^2$

     $$\Rightarrow \quad \widehat{A}^{(i+1)} = (\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}(\widehat{\boldsymbol{h}}^{(i)}))^{-1}\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y} \tag{5.13}$$

   - Minimization w.r.t. $\boldsymbol{h}$; $A = \widehat{A}^{(i+1)}$:
     $\min_{h} \; ||Y - \mathcal{T}(h)\,\widehat{A}^{(i+1)}||^2$

     $$\Rightarrow \quad \widehat{\boldsymbol{h}}^{(i+1)} = (\widehat{\mathcal{A}}^{(i+1)^H}\widehat{\mathcal{A}}^{(i+1)})^{-1}\widehat{\mathcal{A}}^{(i+1)^H}\boldsymbol{Y} \tag{5.14}$$

     $\widehat{\mathcal{A}}^{(i+1)}$ is constructed from $A^{(i+1)}$.

3. Repeat step 1 until $(\,\widehat{A}^{(i+1)}, \widehat{\boldsymbol{h}}^{(i+1)}) \approx (\,\widehat{A}^{(i)}, \widehat{\boldsymbol{h}}^{(i)})$.

We can prove that at each iteration of blind ML-ML, the cost function decreases until a fixed point is reached; and again with good initialization, ML-ML converges to the global minimum. It can be observed that both the symbols and the channel estimates are the outputs of a Zero-Forcing (ZF) linear equalizer. This estimator is sensitive to common zeros that may exist between channels at the receiver. Once it happens, $(\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}))$ becomes rank deficient and consequently the symbols estimates in (5.13) blows up.

5.3.2 GMAP-ML

In this estimator [59],[61] we treat the symbols as random with Gaussian distribution while the channel is considered deterministic to be jointly estimated with the symbols. This estimator also belongs to the first category, thus the log-likelihood function is given by (5.8). Moreover, $f(A) = \frac{1}{(\pi\sigma_a^2)^{M+L-1}} exp[-\frac{A^H A}{\sigma_a^2}]$ and $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h} - \boldsymbol{h}^o)$. It is obvious here that $\ln[f(\boldsymbol{h})]$ can be omitted without affecting the maximization of the log-likelihood function in (5.8). Hence, the cost function is given by:

$$\min_{A,h} \frac{1}{\sigma_v^2}||Y - \mathcal{T}(h)A||^2 + \frac{||A||^2}{\sigma_a^2} \tag{5.15}$$

Following the same methodology used in ML-ML estimator we get:

$$\widehat{A}^{(i+1)} = (\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}(\widehat{\boldsymbol{h}}^{(i)}) + \frac{\sigma_v^2}{\sigma_a^2} I_{pL})^{-1}\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y}. \tag{5.16}$$

$$\widehat{\boldsymbol{h}}^{(i+1)} = (\widehat{\mathcal{A}}^{(i+1)^H}\widehat{\mathcal{A}}^{(i+1)})^{-1}\widehat{\mathcal{A}}^{(i+1)^H}\boldsymbol{Y} \tag{5.17}$$

It can be observed that the symbols estimates are the outputs of a linear MMSE equalizer, while the channel estimates as in the case of ML-ML are the outputs of a ZF linear equalizer. This feature makes this estimator less sensitive to the common zeros problem that ML-ML suffers. But still this estimator suffers from another problem namely, when $\mathcal{A}^H\mathcal{A}$ in (5.17) becomes rank deficient. In spite of the fact that we consider here the symbols to have a Gaussian distribution but in reality we know that the symbols belong to a discrete linear digital modulation scheme for e.g. BPSK and they are uniformly distributed. Hence, if we have a sequence of alternating symbols +1 , -1 then $\mathcal{A}^H\mathcal{A}$ in (5.17) becomes rank deficient and consequently (5.17) blows up.

5.3.3 GMAP-Elm-ML (GML)

This estimator belongs to the second category [59], hence we are interested in estimating the channel and the variance of the noise only while the symbols are supposed to be eliminated during the estimation process. Furthermore, the log-likelihood function is given by (5.11) where we consider the channel and the noise variance to be deterministic while the symbols have a Gaussian distribution. Here again, $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ have no influence on maximizing (5.11). Substituting $f(Y/\boldsymbol{h}, \sigma_v^2) = \frac{1}{(\pi)^{Mp}|R_{YY}|} exp[-Y^H R_{YY}^{-1} Y)]$

where $R_{YY} = E \ \boldsymbol{Y}\boldsymbol{Y}^H = \sigma_a^2 \mathcal{T}(\boldsymbol{h})\mathcal{T}(\boldsymbol{h})^H + \sigma_v^2 I_{Mp}$ in (5.11) after omitting $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ we get:

$$\min_{h,\sigma_v^2} \ \ln|R_{YY}| + \mathrm{tr}(R_{YY}^{-1}\hat{R}_{YY}) \tag{5.18}$$

This cost function can be minimized by resorting to the method of scoring [62]. This method consists in an approximation of the Newton-Raphson algorithm which finds an estimate $\theta(i)$ at iteration i from $\theta(i-1)$, the estimate at iteration $i-1$, as:

$$\theta^{(i)} = \theta^{(i-1)} - \mu \left[\mathcal{F}^{''}|_{\theta^{(i-1)}}\right]^{-1} \ \mathcal{F}^{'}|_{\theta^{(i-1)}} \tag{5.19}$$

where $\mathcal{F}(\theta)$ is the cost function in (5.18), $\mathcal{F}^{''}$ is the Hessian, $\mathcal{F}^{'}$ is the gradient of the cost function and μ is the step length that should be appropriately chosen to guarantee convergence to a local minimum. The method of scoring approximates the Hessian by its expected value, which is here the Gaussian Fisher Information Matrix (FIM). This approximation is justified by the law of large numbers as the number of data is generally large. In our case, the FIM is singular, and as a consequence formula (5.19) cannot be applied directly so we take the Moore-Penrose pseudo inverse of the FIM.

It should be noted that in order to minimize the above cost function we should resort to splitting θ into its real and imaginary parts, hence we get θ_R then we compute $\mathcal{F}^{''}_{\theta_R\theta_R}$ from $\mathcal{F}^{''}_{\theta\theta}$ and $\mathcal{F}^{''}_{\theta\theta^*}$ as we will see in chapter 7 section 7.4.2. As for $\mathcal{F}^{'}(\theta_R)$, it can be computed from $\mathcal{F}^{'}(\theta)$ where the latter is given by:

$$\mathcal{F}^{'}(\theta_j) = \frac{\partial \mathcal{F}(\theta)}{\partial \theta_j^*} = \mathrm{tr}\left(R_{YY}^{-1}\frac{\partial R_{YY}}{\partial \theta_j^*}\left(I - R_{YY}^{-1}\widehat{R}_{YY}\right)\right) \tag{5.20}$$

where θ_j represents the jth element of θ. Hence, we can write:

$$\mathcal{F}^{'}(\theta_R) = 2\left[Re\left(\mathcal{F}^{'}(\theta)\right)^H, Im\left(\mathcal{F}^{'}(\theta)\right)^H\right]^H \tag{5.21}$$

On the other hand, it can be easily shown that this estimator is less sensitive to the common zeros problem. In fact by applying the matrix inversion lemma we can readily prove that $R_{YY}^{-1} = \frac{1}{\sigma_v^2}[I - \mathcal{T}(\boldsymbol{h})(\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2}I_{pL})^{-1}\mathcal{T}^H(\boldsymbol{h})]$.Therefore, even if $\mathcal{T}^H(\boldsymbol{h})\mathcal{T}(\boldsymbol{h})$ is rank deficient, the cost function may not blow up thanks to the regularization parameter $\frac{\sigma_v^2}{\sigma_a^2}I_{pL}$ introduced by the prior information on the symbols.

5.3.4 ML-GMAP

This is our first novel estimator in this chapter, where we introduce the concept of blind Bayesian channel estimation by treating the channel as random with Gaussian distribution $f(\boldsymbol{h}) = \frac{1}{(\pi)^{pL}|C_h^o|} exp[-\boldsymbol{h}^H C_h^{o-1}\boldsymbol{h}]$. However, the symbols are considered as deterministic unknowns to be jointly estimated with the channel hence this estimator belongs to the first category where the log-likelihood function is given by (5.8). Moreover, here again $\ln[f(A)]$ has no effect on maximizing (5.8) so it can be omitted. Therefore, the cost function is given by:

$$\min_{A,h} \quad \frac{1}{\sigma_v^2}||Y - \mathcal{T}(h)A||^2 + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{5.22}$$

Once again here, following the same methodology used in ML-ML estimator we get:

$$\widehat{A}^{(i+1)} = (\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}(\widehat{\boldsymbol{h}}^{(i)}))^{-1}\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y} \tag{5.23}$$

$$\widehat{\boldsymbol{h}}^{(i+1)} = (\widehat{\mathcal{A}}^{(i+1)^H}\widehat{\mathcal{A}}^{(i+1)} + \sigma_v^2 C_h^{o-1})^{-1}\widehat{\mathcal{A}}^{(i+1)^H}\boldsymbol{Y} \tag{5.24}$$

It can be observed that the symbols estimates as in the case of ML-ML are the outputs of a ZF equalizer while the channel estimates are the outputs of a MMSE linear equalizer. Consequently, this estimator is sensitive to the common zeros problem while no matter whatever the sequence of symbols is, the channel estimate in (5.24) always exists thanks to the regularization parameter ($\sigma_v^2 C_h^{o^{-1}}$) introduced by the the prior information on the channel.

5.3.5 GMAP-GMAP

This is our second novel estimator in this chapter, where both the channels and the symbols are assumed random with Gaussian distribution and are supposed to be estimated jointly. Hence, this estimator in its turn belongs to the first category and its log-likelihood is given by (5.8). By substituting the terms in (5.8) by their corresponding functions, we deduce the cost function as follows:

$$\min_{A,\boldsymbol{h}} \frac{1}{\sigma_v^2}||Y - \mathcal{T}(h)A||^2 + \frac{||A||^2}{\sigma_a^2} + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{5.25}$$

Also here, following the same methodology used in ML-ML estimator we get:

$$\widehat{A}^{(i+1)} = (\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}(\widehat{\boldsymbol{h}}^{(i)}) + \frac{\sigma_v^2}{\sigma_a^2} I_{pL})^{-1}\mathcal{T}^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y} \tag{5.26}$$

$$\widehat{\boldsymbol{h}}^{(i+1)} = (\widehat{\mathcal{A}}^{(i+1)^H}\widehat{\mathcal{A}}^{(i+1)} + \sigma_v^2 C_h^{o-1})^{-1}\widehat{\mathcal{A}}^{(i+1)^H}\boldsymbol{Y} \tag{5.27}$$

It can be observed that both the symbols estimates and the channel estimates are the outputs of a MMSE linear equalizer. Therefore, no matter if the channels have zeros in common or the sequence of the symbols take any form, both (5.27) and (5.26) always exist.

5.3.6 GMAP-Elm-GMAP

This is our third novel estimator and it belongs to the second category since the symbols are supposed to be eliminated. It can be considered as an extension to GMAP-Elm-ML by exploiting the prior information that exists about the channel. Its log-likelihood function is given by (5.11) but this time $\ln[f(\boldsymbol{h})]$ can't be omitted. Substituting the terms in (5.11) by their corresponding functions we get the cost function as follows:

$$\min_{\boldsymbol{h},\sigma_v^2} \ \ln|R_{YY}| + \mathrm{tr}(R_{YY}^{-1}\hat{R}_{YY}) + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{5.28}$$

This cost function can be minimized using the scoring method discussed in case of GMAP-Elm-ML estimator. The same formulas stated in that section are still valid with some modifications as shown below:

$$\mathcal{F}_{\theta\theta}^{''\,GMAP-Elm-GMAP} = \mathcal{F}_{\theta\theta}^{''\,GMAP-Elm-ML} + C_h^{o^{-1}} \tag{5.29}$$

$$\mathcal{F}_{\theta}^{'\,GMAP-Elm-GMAP} = \mathcal{F}_{\theta}^{'\,GMAP-Elm-ML} + C_h^{o^{-1}}\ \boldsymbol{h} \tag{5.30}$$

However, $\mathcal{F}_{\theta\theta^*}^{''\,GMAP-Elm-GMAP} = \mathcal{F}_{\theta\theta^*}^{''\,GMAP-Elm-ML}$.

Moreover, following the same explanation developed over there, we can notice that this estimator is insensitive neither to the common zeros problem nor to the case of symbols having special sequences.

5.4 Bayesian Cramér-Rao Bound (BCRB)

It is well known that the channel can be estimated blindly up to a scalar ambiguity $\rho e^{j\phi}$ where ρ stands for the amplitude and ϕ stands for the phase. In [63] the Bayesian CRB in the context of cooperative OFDM was derived

where the authors claimed (section III-B) that the knowledge of the prior information of the channel eliminates the ambiguity of the blind channel estimation. We will show in this short discussion that the prior information of the channel doesn't provide any information about the phase while it provides only a very limited information about the amplitude. Consequently, the ambiguity is not totally removed and the singularity persists. From the pdf of the channel shown before, we can easily notice that the prior Fisher information Matrix (FIM) is given by C_h^{o-1}. Usually the total FIM is the sum of the prior FIM and the FIM of the data. The latter is singular while the former has usually a full rank. Hence, the total FIM has a full rank. At the first glance this will lead to the same conclusion that was drawn in [63] namely, the prior information eliminates the blind channel ambiguity. However, a closer look at the problem will prove that this result is inaccurate at all.

Suppose we have a channel $\boldsymbol{h}^{'} = \rho e^{j\phi}\boldsymbol{h}$ then the FIM of this channel is given by $\frac{1}{\rho^2}C_h^{0-1}$ where we note that ϕ has been completely absorbed. This result shows that the prior FIM has no capability to provide any piece of information regarding the phase. If so, then the question is why the prior FIM has a full rank and doesn't admit any singularity? In order to answer this question and show that the prior FIM is singular we should reparametrize the problem between our hands. Moreover, we should also resort to splitting the complex channel parameters into their real and imaginary parts. When we accomplish the two previous steps and derive the FIM for the new reparametrized prior we will find it singular for sure. To commence with this task, lets take the first tap of the channel as a common factor we get $\boldsymbol{h} = \rho e^{j\phi}\boldsymbol{h}^{'}$ where $\boldsymbol{h}^{'} = [1\ \bar{\boldsymbol{h}}^H]^H$. Denote by $\theta = [\bar{\boldsymbol{h}}^{rT}, \bar{\boldsymbol{h}}^{sT}, \rho, \phi]$ the set of parameters to be estimated where $\bar{\boldsymbol{h}}^{rT}$ and $\bar{\boldsymbol{h}}^{sT}$ denotes respectively the real and the imaginary parts of $\bar{\boldsymbol{h}}$. Due to the lack of space we will not go into the detailed derivation nevertheless we will show below the resulting prior FIM (2pL x 2pL)which is given by:

$$FIM_{prior} = 2\begin{bmatrix} C_h^o(1,1)\bar{C}_h^{o-1} & 0 & 0 & 0 \\ 0 & C_h^o(1,1)\bar{C}_h^{o-1} & 0 & 0 \\ 0 & 0 & C_h^{o-1}(1,1) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{5.31}$$

where $C_h^o(1,1)$ denotes the element that lies in the first row and first column of C_h^o and $\bar{C}_h^{o-1}$ can be obtained from C_h^{o-1} by omitting the first row and the first column. It is evident now that the prior FIM admits one singularity that

corresponds to the phase and it provides only the variance of the ambiguous amplitude $C_h^o(1,1)$ and not the amplitude itself. Hence, this information is considered limited and incomplete. Now to pursue the derivation of the BCRB we should play the same game with the FIM of the data. Doing so, we can show that the latter admits two singularities, one corresponds to the amplitude and the other corresponds to the phase. However, the total FIM which is the sum of the prior and the data FIMs will admit only one singularity that corresponds to the phase. This is because the prior FIM ameliorate only the singularity that corresponds to the amplitude which results from the FIM of the data. Therefore, the prior FIM only contributes to fix one singularity while it has no means to deal with the other. As a consequence, the resulting BCRB which is defined as the inverse of the total FIM is still singular and needs an additional constraint to fix the phase ambiguity.

5.5 Simulations

In this section we try to shed light by means of MonteCarlo simulations on the advantages of blind Bayesian compared to the blind non-Bayesian channel estimation. In each MonteCarlo simulation we generate a Rayleigh fading channel with exponentially decaying power delay profile (PDP)(assumed known) for the channel between each transmitting and receiving antenna pair as follows: e^{-wn} where $n = 0 : L-1$ and w is a constant that controls how steep is the decaying of the PDP. Hence, $C_{\mathbf{h}}^o$ is the diagonal matrix $C_{\mathbf{h}}^o = I_p \otimes C$ where $C = \text{diag}\{e^{-wn},\, n = 0 : L-1\}$. As for the symbols, we generate random 8PSK symbols to reflect the real world case. The performance of the different channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The SNR is defined as: $\text{SNR} = \frac{||\mathcal{T}(h)A||^2}{pM\,\sigma_v^2}$ while The NMSE is defined as $\frac{avg\,||\boldsymbol{h}-\hat{\hat{\boldsymbol{h}}}||^2}{avg\,||\boldsymbol{h}||^2}$ where $\hat{\hat{\boldsymbol{h}}} = \frac{\hat{\boldsymbol{h}}^H \boldsymbol{h}}{||\hat{\boldsymbol{h}}||^2}\boldsymbol{h}$ is the channel estimate adjusted by the least squares constraint to fix the scalar ambiguity that results from the blind channel estimation. All the simulations are initialized by the Subchannel Response Matching (SRM) estimate [47]. In Figure 5.1, we take a look at the considerable gain (4dB) offered by both GMAP-GMAP and GMAP-Elm-GMAP over GMAP-ML at high SNR. Moreover, these two novel Bayesian estimators that we have introduced have the potential to exceed even GMAP-Elm-ML by couple of dBs. This emphasizes the indispensable role of exploiting the prior information of the channel in enhancing the estimation quality at the receiver. It is worthy

to note that in simulating GMAP-Elm-GMAP and GMAP-Elm-ML in Figure 5.1, we consider σ_v^2 to be known hence, we only estimate the channel. This permits us to make a fair comparison between joint estimation of the channel and the symbols (GMAP-GMAP, GMAP-ML) and the estimation of the channel with marginalization of the symbols (GMAP-Elm-GMAP, GMAP-Elm-ML). Taking a close look at Figure 5.1 shows that, in such a scenario, the estimation of the channel with marginalization of the nuisance parameters (symbols) outperforms a little bit the joint estimation of the channel and the symbols. This holds true whatever is the assumption made for the channel namely, deterministic or Bayesian although it is more evident in the deterministic case. In the same spirit, we notice that in Figure 5.2 ML-GMAP where the channel is treated as Bayesian outperforms ML-ML where the channel is treated as deterministic by around 5 dB. However, in Figure 5.3 we show the number of iterations required for each algorithm to converge at SNR = 10 dB. We notice that the algorithms based on the scoring method (GMAP-Elm-ML, GMAP-Elm-GMAP) is faster than those based on alternating between minimization of channel and symbols (ML-ML, ML-GMAP, GMAP-ML, GMAP-GMAP). The former needs tens of iterations while the latter needs hundreds and even thousands of iterations to converge.

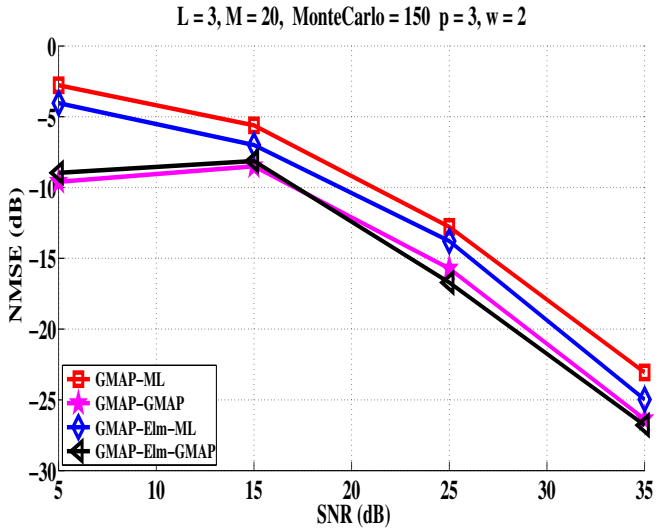

Figure 5.1: NMSE vs. SNR for different estimators.

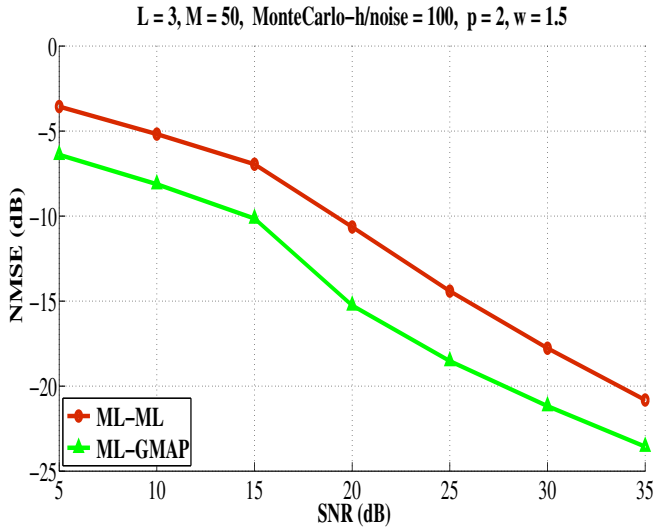

Figure 5.2: NMSE vs. SNR for ML-ML and ML-GMAP.

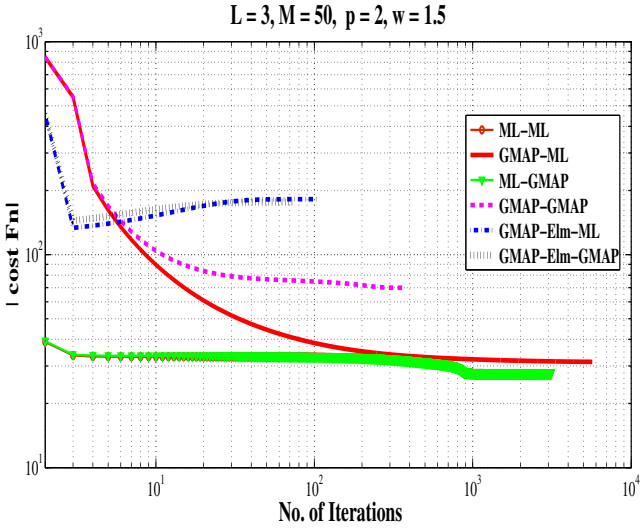

Figure 5.3: The No. of iterations required for convergence for different estimators.

5.6 Conclusion

After we have introduced previously [56] the concept of blind Bayesian channel estimation without providing any specific algorithm, the main message of this chapter is to prove that there is yet a classical way to implement the blind Bayesian channel estimation apart form the variational Bayesian techniques introduced in chapter 4. This concept has been shown by augmenting the cost functions of some ML/MAP estimators that already exist in the literature. Moreover, the novel Bayesian estimators that we have derived in this chapter show a considerable performance gain compared to the deterministic ones. Another aspect that has been addressed in this chapter is the limited contribution of the prior information of the channel in fixing the ambiguities that result from the blind channel estimation problem. We have derived the reparametrized prior FIM of the channel showing that it is singular and is not capable of providing any information related to the ambiguous phase while it provides a limited information to fix the amplitude ambiguity.

Chapter 6

Bayesian Semi-Blind FIR Channel Estimation Algorithms

When the transmission scenario includes a training sequence or pilots, semi-blind channel estimation techniques have shown a tendency to fully exploit the information available from the received signal if they are correctly implemented. This feature leads semi-blind channel estimation performance to exceed that of the schemes based on the blind part or the training sequence part only. Moreover, in some situations they can estimate the channel when the other techniques fail. Semi-blind channel estimation techniques were developed and usually evaluated for a given channel realization, i.e. with a deterministic channel model. On the other hand, as we have indicated in chapter 5, in wireless communications the channel is typically modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. This valuable information plays a dispensable role in enhancing the channel estimation quality if exploited properly. In the spirit of what we have introduced in chapter 5 in the context of blind SIMO channel estimation, we explore in this chapter a Bayesian approach to semi-blind channel estimation. This will be achieved by exploiting the a priori information on fading channels. We mainly focus on semi-blind joint ML/MAP estimation of channels and symbols on one

hand, and on semi-blind ML/MAP estimation of channels with elimination of symbols on the other hand. As a consequence, a unified framework along with three novel semi-blind Bayesian estimators are introduced whose performance is compared by simulations to three, one extended and another two already existing semi-blind non-Bayesian estimators.

6.1 Introduction

Traditionally, the transmitter sends some known information to the receiver to aid the latter in estimating the channel. However, in wireless communication the channel varies rapidly with time and as a consequence more training sequence/pilots are required. This process wastes a lot of bandwidth as a result of augmenting the transmission rate to maintain the throughput. In the last two decades a new branch of channel estimation has emerged focusing on accomplishing this task blindly i.e. without the need for a training sequence. Nevertheless, most wireless standards that have evolved during that period are still relying on the training sequence/pilots to estimate the channel. This is due probably to the unsatisfactory results of the blind channel estimation algorithms. On the other hand, there are some powerful channel estimation algorithms that take the advantage of both aforementioned techniques have been also developed during the same era. These are known as semi-blind where a superior performance is achieved although few training sequence/pilots are transmitted [64], [65], [66]. We will focus in this chapter on the semi-blind maximum likelihood (ML) and/or maximum a posteriori methods (MAP) [67],[68]. Two approaches exist in the literature on how to tackle the problem. The first approach is based on the fact that the unknown symbols are considered as deterministic to be jointly estimated with the channel. Such algorithm is called Semi-Blind Deterministic (or conditional) Maximum Likelihood (SB-DML) method [69]. The second approach is based on treating the unknown symbols as random quantities with known prior information to be eliminated or jointly estimated. When the unknown symbols are eliminated, the method is called Semi-Blind Gaussian Maximum Likelihood (SB-GML) [70], see also [71]. While when they are jointly estimated, we call this method SB-GMAP-ML. This is because we use semi-blind maximum a posteriori (MAP) for unknown symbols and semi-blind ML for channels and noise variance. The corresponding blind algorithm appeared first in [61]. Furthermore, in all these approaches the channel was considered as deterministic unknown however, in chapter 4 we discussed briefly some classical Bayesian algorithms and introduced the con-

cept of variational Bayesian in the context of MIMO OFDM. Apart from the variational Bayesian techniques, we have also developed in chapter 5 some classical Bayesian blind channel estimation algorithms in the context of SIMO transmission systems. In this chapter, we extend the work done in chapter 5 to the case of semi-blind. Once the channel is treated as random in the presence of training sequence/pilots, we are within the framework of Bayesian semi-blind channel estimation and there are are three cases to be handled. In the first case, the unknown symbols are considered as deterministic to be jointly estimated with channel. We call this method as SB-ML-GMAP, for a similar reasoning discussed above. In the second case, the unknown symbols are again to be jointly estimated with the channel but this time they are considered as random with known prior Gaussian distribution. We call this method SB-GMAP-GMAP. In the third case, the unknown symbols are again random with known prior Gaussian distribution but they are going to be eliminated rather than estimated. We call this method SB-GMAP-Elm-GMAP where Elm stands for elimination of symbols to distinguish it from SB-GMAP-GMAP where both the unknown symbols and the channel are jointly estimated. Consequently, in section III we revisit two already existing deterministic estimators and develop novel ones, one deterministic and three Bayesian. Therefore, with the introduction of the Bayesian semi-blind channel estimation algorithms, the picture is broadened considerably and to sum up we depict the current picture in Table 6.1.

Algorithm	**Unknown Sym**	**Channel**	**Elm of Sym**	**Novel**
SB-ML-ML	Det	Det	No	No
SB-GMAP-ML	Gauss	Det	No	Yes
SB-GMAP-Elm-ML	Gauss	Det	Yes	No
SB-ML-GMAP	Det	Gauss	No	Yes
SB-GMAP-GMAP	Gauss	Gauss	No	Yes
SB-GMAP-Elm-GMAP	Gauss	Gauss	Yes	Yes

Table 6.1: Summary of all algorithms

6.2 SIMO FIR Tx System Model

We will use the same model stated in chapter 5, section 5.2. More specifically we will start from equation (5.4) which is rewritten below with the addition

of the notations of known and unknown symbols:

$$\begin{aligned} \boldsymbol{Y} = \mathcal{T}(\boldsymbol{h})\, A + \boldsymbol{V} &= \mathcal{T}_K(\boldsymbol{h})\, A_K + \mathcal{T}_U(\boldsymbol{h})\, A_U + \boldsymbol{V} \\ &= \mathcal{A}_K \boldsymbol{h} + \mathcal{A}_U \boldsymbol{h} + \boldsymbol{V}\ . \end{aligned} \tag{6.1}$$

Where $\mathcal{T}_K(\boldsymbol{h})$ and $\mathcal{T}_U(\boldsymbol{h})$ represent respectively the portions of $\mathcal{T}(\boldsymbol{h})$ that correspond to A_k (M_K known symbols) and A_U (M_U unknown symbols), see Figure 6.1. On the other hand, $\mathcal{A}$ is a block Toeplitz matrix filled with the elements of A while $\mathcal{A}_K$ and $\mathcal{A}_U$ are block Toeplitz matrices filled with the elements of A_K and A_U respectively. It is worthy to note that the way we split the received data, as in Figure 6.1, doesn't permit to fully exploit the information available, nevertheless it is still capable of showing the superiority of semi-blind on one hand, and allows for pursuing an analytical performance analysis. On the other hand, though the formulation here is explained for the time domain, it is actual general enough to allow handling also OFDM, SC-CP, SC-ZP etc. And in the case of OFDM, $\boldsymbol{Y}$ is composed of $\boldsymbol{Y}_K$ and $\boldsymbol{Y}_U$ only.

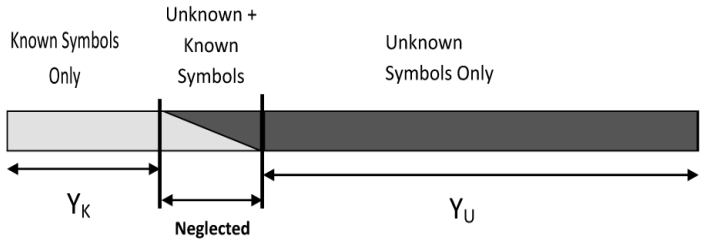

Figure 6.1: Splitting the received signal into two parts containing only pure known and unknown symbols.

6.3 A Unified Framework for different Algorithms

As we have shown in Table 6.1 there are six possible estimators that can be classified into two categories. In the first category the subject of the estimators is to estimate the channel and the unknown symbols jointly by making some assumptions on the channel and the unknowns symbols. If we denote by θ the unknown parameters to be estimated then it is given by:

$$\theta = [A_U^H, \boldsymbol{h}^H]^H \tag{6.2}$$

Following the same discussion shown in chapter 5, section 5.3 we get:

$$\ln[f(Y, A_U, \boldsymbol{h})] = \ln[f(Y/A_U, \boldsymbol{h})] + \ln[f(A_U)] + \ln[f(\boldsymbol{h})] \tag{6.3}$$

However, in the second category the subject of the estimators is to estimate the channel and the noise variance only, while the unknown symbols are supposed to be eliminated during the estimation process. Thus

$$\theta = [\boldsymbol{h}^H, \sigma_v^2]^H \tag{6.4}$$

Once again, following the steps stated in chapter 5, section 5.3 we get:

$$\ln[f(Y, \boldsymbol{h}, \sigma_v^2)] = \ln[f(Y/\boldsymbol{h}, \sigma_v^2)] + \ln[f(\boldsymbol{h})] + \ln[f(\sigma_v^2)] \tag{6.5}$$

We will develop in the following sections the cost functions of all the estimators that belong to both categories and provide a closed form formula for both the estimated channel and symbols where it is possible. It is worthy to note here that since the channel is treated as random rather than deterministic in some of the above mentioned estimators (SB-ML-GMAP, SB-GMAP-GMAP, SB-GMAP-Elm-GMAP) in both categories, these estimators are considered as an example of the Bayesian semi-blind channel estimation.

6.3.1 SB-ML-ML (SB-DML)

We start with SB-ML-ML or what is called SB-DML in the literature [69]. In this case both the unknown symbols and the channel are considered as deterministic unknowns to be estimated. Hence it belongs to the first category and consequently the joint probability density function is given by (6.3). Moreover, since both are deterministic we have $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h} - \boldsymbol{h}^o)$ and $f(A_U) = A_U^o\delta(A_U - A_U^o)$ where $\boldsymbol{h}^o$ and A^o represent respectively the true values of the channel and the symbols. It is obvious that the the pdfs of both the unknown symbols and the channel have no influence on the maximization of (6.3). Hence, we can derive the cost function by maximizing $\ln[f(Y/A_U, \boldsymbol{h})]$ directly where $f(Y/A_U, \boldsymbol{h}) = \frac{1}{(\pi\sigma_v^2)^{MP}} exp[-\frac{1}{\sigma_v^2}(Y - \mathcal{T}(h)A)^H(Y - \mathcal{T}(h)A)]$. Thus, the cost function is given by:

$$\min_{A_U,h} \; ||Y - \mathcal{T}(h)A||^2 = \min_{A_U,\boldsymbol{h}} \; ||Y - \mathcal{T}_K(h)A_K - \mathcal{T}_U(h)A_U||^2 \tag{6.6}$$

However, our model allows for a simplification of this cost function because Y_K and Y_U are decoupled in terms of noise. On the other hand, the

drawback of this model as we indicated before is that it leads to a suboptimal solution because we are neglecting the part that contains both known and unknown symbols. Hence, (6.6) can be written as:

$$\min_{A_U,h} \quad \{||Y_K - \mathcal{T}_K(h)A_K||^2 + ||Y_U - \mathcal{T}_U(h)A_U||^2\} \tag{6.7}$$

The joint optimization of this cost function in both the channel ($\boldsymbol{h}$) and the symbols (A_U) is difficult. Fortunately, the observation is linear in both the channel and the symbols. In other words, we have a separable nonlinear LS problem, which allows us to reduce the complexity considerably. The nonlinear LS optimization can be done by iterating between minimization with respect to A_U and $\boldsymbol{h}$. The alternating minimization strategy is as follows:

1. Initialization $\widehat{\boldsymbol{h}}^0$.

2. Iteration (i+1):

   - Minimization w.r.t. $A_U; \boldsymbol{h} = \widehat{\boldsymbol{h}}^{(i)}$:
     $\min_{A_U} \quad \left\{||Y_K - \mathcal{T}_K(\widehat{h}^{(i)})A_K||^2 + ||Y_U - \mathcal{T}_U(\widehat{h}^{(i)})A_U||^2\right\}$

     $$\Rightarrow \quad \widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}_U(\widehat{\boldsymbol{h}}^{(i)}))^{-1}\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y}_U \tag{6.8}$$

   - Minimization w.r.t. $\boldsymbol{h}; A_U = \widehat{A}_U^{(i+1)}$:
     $\min_{h} \quad \left\{||Y_K - \mathcal{T}_K(h)A_K||^2 + ||Y_U - \mathcal{T}_U(h)\,\widehat{A}_U^{(i+1)}||^2\right\}$

     $$\Rightarrow \quad \widehat{\boldsymbol{h}}^{(i+1)} = (\mathcal{A}_K^H\mathcal{A}_K + \widehat{\mathcal{A}}_U^{(i+1)^H}\widehat{\mathcal{A}}_U^{(i+1)})^{-1}(\mathcal{A}_K^H\boldsymbol{Y}_K + \widehat{\mathcal{A}}_U^{(i+1)^H}\boldsymbol{Y}_U) \tag{6.9}$$

     $\widehat{\mathcal{A}}_U^{(i+1)}$ is constructed from $\widehat{A}_U^{(i+1)}$.

3. Repeat step 1 until $(\widehat{A}_U^{(i+1)}, \widehat{\boldsymbol{h}}^{(i+1)}) \approx (\widehat{A}_U^{(i)}, \widehat{\boldsymbol{h}}^{(i)})$.

It can be observed that both the unknown symbols and the channel estimates are the outputs of a Decision Feedback (DF) Zero-Forcing (ZF) linear equalizer.

6.3.2 SB-GMAP-ML

This is the first novel estimator we introduce in this chapter. It is considered as an extension of the corresponding blind one proposed in [61], [59] and stated in chapter 5. In this estimator we treat the unknown symbols as random with Gaussian distribution while the channel is considered deterministic to be jointly estimated with the unknown symbols. This estimator also belongs to the first category, thus the joint probability density function is given by (6.3). Moreover, $f(A_U) = \frac{1}{(\pi\sigma_a^2)^{M_U}} exp[-\frac{A_U^H A_U}{\sigma_a^2}]$ and $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h}-\boldsymbol{h}^o)$. It is obvious here that $\ln[f(\boldsymbol{h})]$ can be omitted without affecting the maximization of the joint probability density function in (6.3). Hence, the cost function is given by:

$$\min_{A_U,h} \frac{1}{\sigma_v^2}||Y_K - \mathcal{T}_K(h)A_K||^2 + \frac{1}{\sigma_v^2}||Y_U - \mathcal{T}_U(h)A_U||^2 + \frac{||A_U||^2}{\sigma_a^2} \qquad (6.10)$$

Following the same methodology used in SB-ML-ML estimator, we get:

$$\widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}_U(\widehat{\boldsymbol{h}}^{(i)}) + \frac{\sigma_v^2}{\sigma_a^2} I_{M+L-1-M_K})^{-1}\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y}_U \qquad (6.11)$$

$$\widehat{\boldsymbol{h}}^{(i+1)} = (\mathcal{A}_K^H\mathcal{A}_K + \widehat{\mathcal{A}}_U^{(i+1)^H}\widehat{\mathcal{A}}_U^{(i+1)})^{-1}(\mathcal{A}_K^H\boldsymbol{Y}_K + \widehat{\mathcal{A}}_U^{(i+1)^H}\boldsymbol{Y}_U) \qquad (6.12)$$

It can be observed that the symbols estimates are the outputs of a Decision Feedback (DF) MMSE equalizer, while the channel estimates as in the case of SB-ML-ML are the outputs of a DF-ZF linear equalizer.

6.3.3 SB-GMAP-Elm-ML (SB-GML)

This estimator belongs to the second category [70], hence we are interested in estimating the channel and the variance of the noise only while the unknown symbols are supposed to be eliminated during the estimation process. Furthermore, the joint probability density function is given by (6.5) where we consider the channel and the noise variance to be deterministic while the unknown symbols have a Gaussian distribution. Here again, $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ have no influence on maximizing (6.5). Substituting $f(Y_U/\boldsymbol{h},\sigma_v^2) = \frac{1}{(\pi)^{(M-M_K)p}|R_{Y_UY_U}|} exp[-Y_U^H R_{Y_UY_U}^{-1} Y_U)]$ Where $R_{Y_UY_U} = E\ \boldsymbol{Y}_U\boldsymbol{Y}_U^H = \sigma_a^2\mathcal{T}_U(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h})^H + \sigma_v^2 I_{(M-M_K)p}$ in (6.5) after omitting $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ we get:

$$\min_{\boldsymbol{h},\sigma_v^2} \ \frac{1}{\sigma_v^2}||Y_K - \mathcal{T}_K(h)A_K||^2 + \ln|R_{Y_UY_U}| + \mathrm{tr}(R_{Y_UY_U}^{-1}\hat{R}_{Y_UY_U}) \tag{6.13}$$

This cost function can be minimized by resorting to the method of scoring [62]. This method consists in an approximation of the Newton-Raphson algorithm which finds an estimate $\theta(i)$ at iteration i from $\theta(i-1)$, the estimate at iteration $i-1$, as:

$$\theta^{(i)} = \theta^{(i-1)} - \mu\left[\mathcal{F}^{''}|_{\theta^{(i-1)}}\right]^{-1} \ \mathcal{F}^{'}|_{\theta^{(i-1)}} \tag{6.14}$$

where $\mathcal{F}(\theta)$ is the cost function in (6.13), $\mathcal{F}^{''}$ is the Hessian, $\mathcal{F}^{'}$ is the gradient of the cost function nad μ is the step length that should be appropriately chosen to guarantee convergence to a local minimum. The method of scoring approximates the Hessian by its expected value, which is here the Gaussian Fisher Information Matrix (FIM). This approximation is justified by the law of large numbers as the number of data is generally large.

6.3.4 SB-ML-GMAP

This estimator is novel, where we introduce the concept of semi-blind Bayesian channel estimation by treating the channel as random with Gaussian distribution $f(\boldsymbol{h}) = \frac{1}{(\pi)^{pL}|C_h^o|}exp[-\boldsymbol{h}^H C_h^{o-1}\boldsymbol{h}]$. However, the unknown symbols are considered as deterministic to be jointly estimated with the channel hence, this estimator belongs to the first category where the joint probability density function is given by (6.3). Moreover, here again $\ln[f(A_U)]$ has no effect on maximizing (6.3) so it can be omitted. Therefore, the cost function is given by:

$$\min_{A_U,\boldsymbol{h}} \ \frac{1}{\sigma_v^2}||Y_K - \mathcal{T}_K(h)A_K||^2 + \frac{1}{\sigma_v^2}||Y_U - \mathcal{T}_U(h)A_U||^2 + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{6.15}$$

Once again here, following the same methodology used in SB-ML-ML estimator we get:

$$\widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)})\mathcal{T}_U(\widehat{\boldsymbol{h}}^{(i)}))^{-1}\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)})\boldsymbol{Y}_U \tag{6.16}$$

$$\widehat{\boldsymbol{h}}^{(i+1)} = (\mathcal{A}_K^H \mathcal{A}_K + \widehat{\mathcal{A}}_U^{(i+1)^H} \widehat{\mathcal{A}}_U^{(i+1)} + \sigma_v^2 C_h^{o^{-1}})^{-1} (\mathcal{A}_K^H \boldsymbol{Y}_K + \widehat{\mathcal{A}}_U^{(i+1)^H} \boldsymbol{Y}_U) \quad (6.17)$$

It can be observed that the symbols estimates, as in the case of ML-ML, are the outputs of a DF-ZF equalizer, while the channel estimates are the outputs of a DF-MMSE linear equalizer.

6.3.5 SB-GMAP-GMAP

This estimator is also novel where both the channels and the unknown symbols are assumed random with Gaussian distribution and are supposed to be estimated jointly. Hence, this estimator in its turn belongs to the first category and its joint probability density is given by (6.3). By substituting the terms in (6.3) by their corresponding functions we deduce the cost function as follows:

$$\min_{A_U, \boldsymbol{h}} \frac{1}{\sigma_v^2} ||Y_K - \mathcal{T}_K(h) A_K||^2 + \frac{1}{\sigma_v^2} ||Y_U - \mathcal{T}_U(h) A_U||^2 + \frac{||A_U||^2}{\sigma_a^2} + \boldsymbol{h}^H C_h^{o^{-1}} \boldsymbol{h} \quad (6.18)$$

Also here, following the same methodology used in SB-ML-ML estimator we get:

$$\widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)}) \mathcal{T}_U(\widehat{\boldsymbol{h}}^{(i)}) + \frac{\sigma_v^2}{\sigma_a^2} I_{M+L-1-M_K})^{-1} \mathcal{T}_U^H(\widehat{\boldsymbol{h}}^{(i)}) \boldsymbol{Y}_U \quad (6.19)$$

$$\widehat{\boldsymbol{h}}^{(i+1)} = (\mathcal{A}_K^H \mathcal{A}_K + \widehat{\mathcal{A}}_U^{(i+1)^H} \widehat{\mathcal{A}}_U^{(i+1)} + \sigma_v^2 C_h^{o^{-1}})^{-1} (\mathcal{A}_K^H \boldsymbol{Y}_K + \widehat{\mathcal{A}}_U^{(i+1)^H} \boldsymbol{Y}_U) \quad (6.20)$$

It can be observed that both the symbols estimates and the channel estimates are the outputs of a DF-MMSE linear equalizer.

6.3.6 SB-GMAP-Elm-GMAP

This is the last novel estimator we are going to introduce in this chapter. It belongs to the second category since the symbols are supposed to be eliminated. It can be considered as an extension of SB-GMAP-Elm-ML by

exploiting the prior information exists about the channel. Its joint probability density function is given by (6.5) but this time $\ln[f(\boldsymbol{h})]$ can't be omitted. Substituting the terms in (6.5) by their corresponding functions we get the cost function as follows:

$$\min_{\boldsymbol{h},\sigma_v^2} \frac{1}{\sigma_v^2}||Y_K - \mathcal{T}_K(h)A_K||^2 + \ln|R_{Y_UY_U}| + \operatorname{tr}(R_{Y_UY_U}^{-1}\hat{R}_{Y_UY_U}) + \boldsymbol{h}^H C_h^{o-1}\boldsymbol{h} \tag{6.21}$$

This cost function can be minimized using the scoring method discussed in case of SB-GMAP-Elm-ML estimator.

6.4 Simulations

In this section we try to shed light by means of MonteCarlo simulations on the advantages of semi-blind Bayesian compared to the semi-blind non-Bayesian channel estimation. In each MonteCarlo simulation we generate a Rayleigh fading channel with exponentially decaying power delay profile (PDP) for the channel between each transmitting and receiving antenna pair as follows: e^{-wn} where $n = 0 : N-1$ and $w = 1$ except otherwise stated. Hence, $C_{\mathbf{h}}^o$ is the diagonal matrix $C_{\mathbf{h}}^o = I_p \otimes C$ where $C = \operatorname{diag}\{e^{-wn},\, n = 0 : N-1\}$. As for the symbols, we generate random 8PSK symbols to reflect the real world case. The performance of the different channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The SNR is defined as: $\text{SNR} = \frac{||\mathcal{T}(h)A||^2}{pM\,\sigma_v^2}$. The NMSE is defined as $\frac{avg\,||\boldsymbol{h}-\hat{\boldsymbol{h}}||^2}{avg\,||\boldsymbol{h}||^2}$. All the simulations are initialized by the Subchannel Response Matching (SRM) estimate [47] where the scalar ambiguity of the latter has been fixed by a least squares constraint. In Figure 6.2 we compare the performance of SB-ML-ML estimator with SB-ML-GMAP, we can notice how the latter exceeds the former by around 5 dB along the medium SNR range while their performances are congruent at high SNR. In Figure 6.3 we take a look at the considerable gain (7 dB) offered by SB-GMAP-GMAP over SB-GMAP-ML along the medium SNR range. However, in Figure 6.4 both SB-GMAP-Elm-ML and SB-GMAP-Elm-GMAP are plotted on the same figure. Also in this case, where the symbols are eliminated, we can obviously observe the indispensable role that the prior information about the channel plays in enhancing the estimation quality at the receiver ($\approx$ 5 dB gain), especially at low and medium SNR.

On the other hand, we have noticed in Figures 6.2, 6.3 and 6.4 that at high SNR, the Bayesian estimators are no more able to outperform the deterministic ones. In order to highlight the NMSE of different deterministic and Bayesian estimators at high SNR, we plot in Figures 6.5, 6.6 and 6.7 the NMSE versus the number of iterations at SNR = 30 dB. However, the algorithms are initialized by the true channel and not by the estimated one as we did previously. Nevertheless, we can readily observe that even if the estimators are initialized by the true channel, the gain offered by the Bayesian estimators over the deterministic ones are almost negligible.

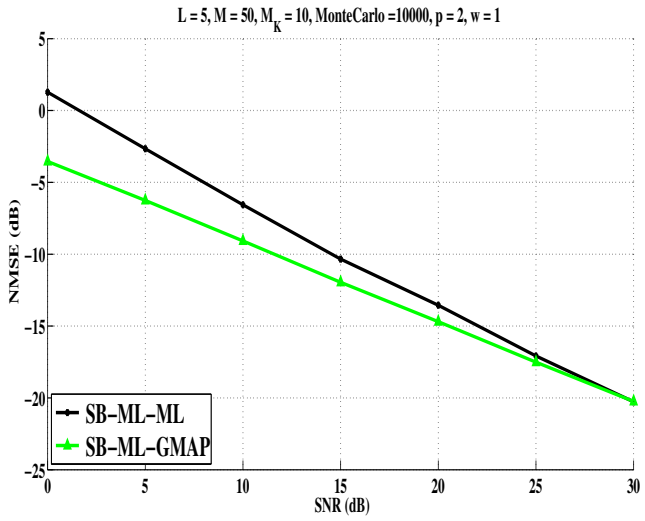

Figure 6.2: NMSE vs. SNR for SB-ML-ML and SB-ML-GMAP.

6.5 Conclusion

We have shown in this chapter that there is still a considerable room to enhance the performance of semi-blind channel estimation. This is true when we take into consideration the a priori information that exists about the channel power delay profile. Hence, we have introduced in this chapter the concept of Bayesian semi-blind channel estimation and proposed three new Bayesian semi-blind estimators. On the other hand, we have also extended one existing deterministic blind estimator to the semi-blind case. Furthermore, we have presented a unified framework from which we can derive both deterministic and Bayesian estimators. As our simulations show, the pro-

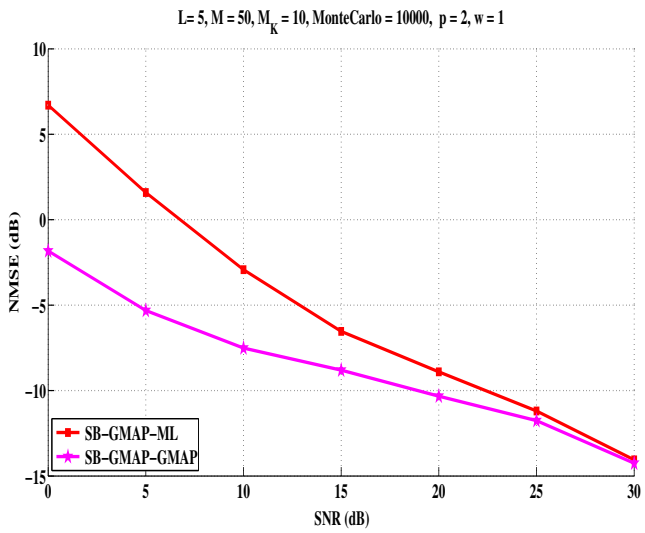

Figure 6.3: NMSE vs. SNR for SB-GMAP-ML and SB-GMAP-GMAP.

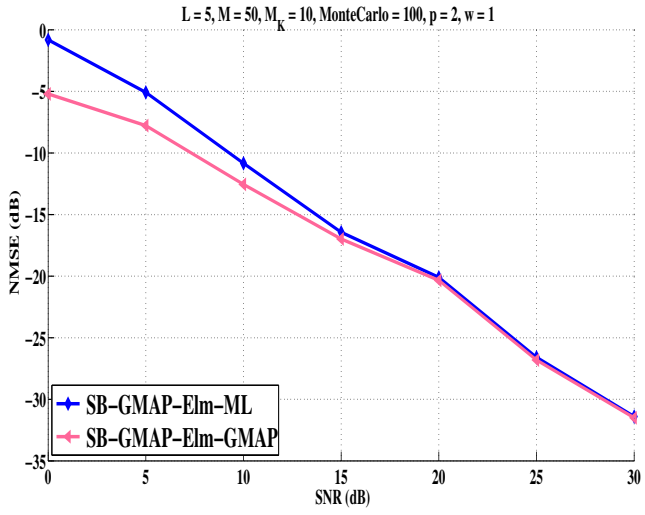

Figure 6.4: NMSE vs. SNR for SB-GMAP-Elm-ML and SB-GMAP-Elm-GMAP.

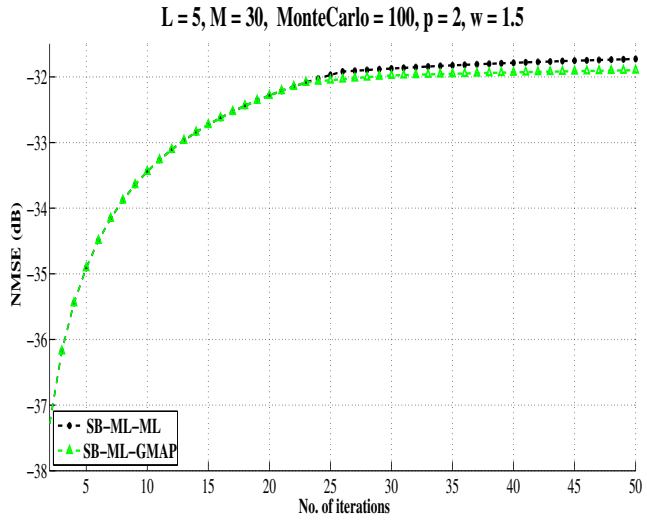

Figure 6.5: NMSE vs. No. of iterations for SB-ML-ML and SB-ML-GMAP.

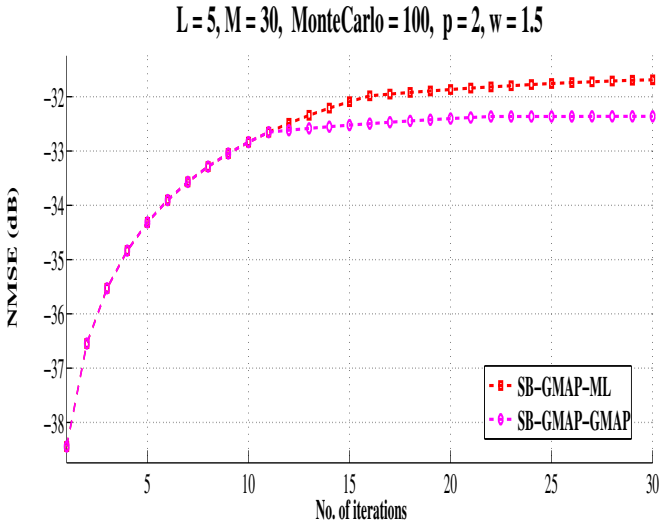

Figure 6.6: NMSE vs. No. of iterations for SB-GMAP-ML and SB-GMAP-GMAP.

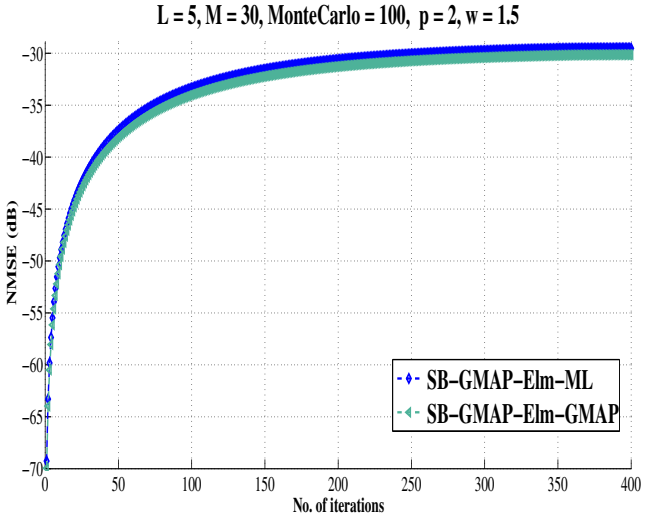

Figure 6.7: NMSE vs. No. of iterations for SB-GMAP-Elm-ML and SB-GMAP-Elm-GMAP.

posed Bayesian estimators have a superior performance compared to the deterministic ones. The main disadvantage of the algorithms introduced in this chapter is that they require a large number of iterations to converge. However, we believe that this topic could be further investigated to provide more practical algorithms.

Chapter 7

Bayesian and Deterministic CRBs for Semi-Blind Channel Estimation

Traditionally, the performance of different semi-blind channel estimation algorithms has been assessed and compared to a certain lower bound. One of these famous lower bounds that has been extensively used in the literature is the Cramér-Rao Bound (CRB). Depending on how we treat the symbols and the channel, different versions of CRB have been derived. There are two possible cases on how to treat the symbols and/or the channel namely, deterministic unknowns or random. Moreover, the symbols are either jointly estimated with the channel or eliminated. In other words, we have six different cases to be handled. In this chapter we present the CRBs that exist in the literature and fit to some of these cases and derive the others in the context of SIMO FIR system. On the top of that we present a unified framework that permits to derive all versions of CRBs in a concrete manner. All the derived CRBs are validated numerically by conducting limited Monte-Carlo simulations.

7.1 Introduction

We have indicated in chapter 6 that most wireless standards that have been evolved recently are still relying on the training sequence/pilots to estimate the channel. Moreover, we have indicated that the reasonable justification for this trend is due probably to the unsatisfactory results of the blind channel estimation algorithms. On the other hand, we have also indicated in chapter 6 that there are some powerful channel estimation algorithms that take advantage of both training sequence-based and blind techniques known as semi-blind. Furthermore, we have derived in that chapter many algorithms that lie in this category. As usual the performance of these algorithms are lower bounded. We will focus in this chapter on the most famous lower bound used by the statisticians namely, the CRB. Different versions of semi-blind CRBs are shown and derived in the sequel. Basically, there are two approaches on how to tackle the problem of semi-blind channel estimation depending on how we treat the transmitted symbols. The first approach is based on jointly estimating the symbols with the channel while the second approach is based on estimating the channel and marginalizing the symbols. Moreover, in the first approach we have the choice to consider the channel and/or the symbols as either deterministic unknowns or random with known probability density function (pdf). Hence, there are four methodologies to jointly estimate the channel and the symbols. However, in the second approach we can only marginalize the symbols if we consider them as random with known pdf regardless of how we treat the channel. Therefore, there are only two methodologies to estimate the channel while marginalizing the symbols. Overall we have six cases to be handled. It should be noted that treating the channel as random rather than deterministic in the context of blind and semi-blind channel estimation has been introduced in [56] and we have developed it in chapters 5 and 6. Once the channel is treated as random, we are within the framework of Bayesian semi-blind channel estimation. In chapter 6, the Bayesian and the deterministic algorithms are evaluated by running Monte-Carlo simulations. In this chapter, we will derive the lower bounds that correspond to the different algorithms elaborated in chapter 6. This chapter is organized as follows: In section 2 we develop the SIMO FIR transmission system model, while in section 3 we show a general framework that permits the derivation of the different CRBs that belong to the two approaches stated above. In section 4 we make use of the framework developed in section 3 to derive the different CRBs. In section 5 we show a summary of the CRBs and in section 6 we conduct some Monte-Carlo simulations to pictorially compare different CRBs with their

corresponding algorithms. Finally, in section 7 we draw some conclusions.

7.2 SIMO FIR Tx System Model

We will use the same model introduced in chapter 5, section 5.2. More specifically, we will start from equation (5.4) which is rewritten below with the addition of the notations of known and unknown symbols. However, it is worthy noting that the way by which we split the received signal into known and unknown parts is different from that used in chapter 6. In chapter 6, we neglected the part of the received signal that contains both known and unknown symbols. The goal was to develop algorithms that can be analyzed analytically. However, in this chapter we are interested in deriving the CRBs that correspond to the optimal case. By optimality we mean exploiting every piece of information that exists in both the received signal and in the prior information that we do have.

$$\begin{aligned} \boldsymbol{Y} = \mathcal{T}(\boldsymbol{h})\, A + \boldsymbol{V} &= \mathcal{T}_K(\boldsymbol{h})\, A_K + \mathcal{T}_U(\boldsymbol{h})\, A_U + \boldsymbol{V} \\ &= \mathcal{A}_K \boldsymbol{h} + \mathcal{A}_U \boldsymbol{h} + \boldsymbol{V}\ . \end{aligned} \tag{7.1}$$

Where $\mathcal{T}_K(\boldsymbol{h})$ and $\mathcal{T}_U(\boldsymbol{h})$ represent respectively the portions of $\mathcal{T}(\boldsymbol{h})$ that correspond to A_k (M_K known symbols) and A_U (M_U unknown symbols), see (7.2). Here we assume for simplicity that the known symbols are gathered at the beginning of the block. On the other hand, $\mathcal{A}$ is a block circulant matrix filled with the elements of A while $\mathcal{A}_K$ and $\mathcal{A}_U$ are block circulant matrices filled with the elements of A_K and A_U respectively.

$$\mathcal{T}(\boldsymbol{h}) = \left[\begin{array}{c|c} \mathcal{T}_K(\boldsymbol{h}) & \mathcal{T}_U(\boldsymbol{h}) \end{array}\right] \tag{7.2}$$

7.3 A Unified Framework for different CRBs

As we have stated before, there are six possible cases that can be classified into two categories. In the first category the subject of the estimators is to estimate the channel and the unknown symbols are estimated jointly by making some assumptions on the channel and the unknowns symbols. It is worthy to note that in this category the estimation of the channel and symbols from one side and the noise variance estimation from the other side are decoupled. In other words, the estimation of the noise variance has no effect on the estimation of both the channel and the symbols. Hence, the

estimation of the noise variance is excluded in this category. If we denote by θ the unknown parameters to be estimated then it is given by:

$$\theta = [A_U^H, \boldsymbol{h}^H]^H \tag{7.3}$$

The framework that we have introduced in chapter 5, section 5.3 is valid. Hence, equation 6.3 is still applicable and we will restate it here below:

$$\ln[f(Y, A_U, \boldsymbol{h})] = \ln[f(Y/A_U, \boldsymbol{h})] + \ln[f(A_U)] + \ln[f(\boldsymbol{h})] \tag{7.4}$$

Now, let J represents the Fisher Information matrix (FIM), it is given by [72]:

$$\begin{aligned} J_{\theta\theta} &= \mathrm{E}\left(\frac{\partial \ln[f(Y, A_U, \boldsymbol{h})]}{\partial \theta^*}\right)\left(\frac{\partial \ln[f(Y, A_U, \boldsymbol{h})]}{\partial \theta^*}\right)^H \\ &= -\mathrm{E}\,\frac{\partial}{\partial \theta^*}\left(\frac{\partial \ln[f(Y, A_U, \boldsymbol{h})]}{\partial \theta^*}\right)^H \end{aligned} \tag{7.5}$$

As we shall observe later, since we are treating complex parameters we also need, besides $J_{\theta\theta}$, $J_{\theta\theta^*}$ which is defined by:

$$\begin{aligned} J_{\theta\theta^*} &= \mathrm{E}\left(\frac{\partial \ln[f(Y, A_U, \boldsymbol{h})]}{\partial \theta^*}\right)\left(\frac{\partial \ln[f(Y, A_U, \boldsymbol{h})]}{\partial \theta}\right)^H \\ &= -\mathrm{E}\,\frac{\partial}{\partial \theta}\left(\frac{\partial \ln[f(Y, A_U, \boldsymbol{h})]}{\partial \theta^*}\right)^H \end{aligned} \tag{7.6}$$

When $J_{\theta\theta^*} \neq 0$ we shall resort to θ_R defined below:

$$\theta_R = \begin{bmatrix} Re(\theta) \\ Im(\theta) \end{bmatrix} = \mathcal{M}\begin{bmatrix} \theta \\ \theta^* \end{bmatrix}, \mathcal{M} = \frac{1}{2}\begin{bmatrix} I & I \\ -jI & jI \end{bmatrix} \tag{7.7}$$

Knowing that $J_{\theta\theta} = J^*_{\theta^*\theta^*}$ and $J_{\theta\theta^*} = J^*_{\theta^*\theta}$ then (7.7) yields:

$$J_{\theta_R\theta_R} = \mathcal{M}\begin{bmatrix} J_{\theta\theta} & J_{\theta\theta^*} \\ J^*_{\theta\theta^*} & J^*_{\theta\theta} \end{bmatrix}\mathcal{M}^H \tag{7.8}$$

On the other side, when $J_{\theta\theta^*} = 0$, $J_{\theta_R\theta_R}$ can be deduced totally from $J_{\theta\theta}$. This holds true for all the cases where we jointly estimate the channel and the symbols as we shall notice later. Under some assumptions and regularity conditions [73], the error covariance matrix of an unbiased channel estimator $\widehat{\boldsymbol{h}}(Y)$, which is defined as:

$$C(\widehat{h}) = \mathrm{E}\left\{\left[\widehat{\boldsymbol{h}}(Y) - \boldsymbol{h}\right]\left[\widehat{\boldsymbol{h}}(Y) - \boldsymbol{h}\right]^H\right\} \tag{7.9}$$

satisfies the following inequality:

$$C(\widehat{\boldsymbol{h}}) \geq \{J_{\theta_R\theta_R}\}^{-1} \stackrel{\triangle}{=} CRB \tag{7.10}$$

We usually focus on comparing the Mean Square Error, MSE = tr $\left\{C(\widehat{\boldsymbol{h}})\right\}$ to the minimum error variance which is defined by tr $\{CRB\}$ where tr stands for the trace of a matrix.

However, in the second category the channel and the noise variance are the only parameters to be estimated while the symbols are supposed to be marginalized during the estimation process. Here we can't exclude the estimation of the noise variance because it is coupled to the estimation of the channel. Thus,

$$\theta = [\boldsymbol{h}^H, \sigma_v^2]^H \tag{7.11}$$

Once again, if we follow the steps mentioned in chapter 5, section 5.3 we will arrive at the formula 6.5 restated below:

$$\ln[f(Y, \boldsymbol{h}, \sigma_v^2)] = \ln[f(Y/\boldsymbol{h}, \sigma_v^2)] + \ln[f(\boldsymbol{h})] + \ln[f(\sigma_v^2)] \tag{7.12}$$

As for FIM, both (7.5) and (7.6) are still applicable where only θ is redefined as in (7.11).

7.4 Derivations of Different CRBs

We shall develop in this section the CRBs of all the cases that belong to both categories and provide a closed-form formula where it is possible. This will be done by exploiting the framework introduced in the previous section. To commence with this task, we shall explain the way by which we call the different CRBs. First of all, to differentiate between the CRBs that correspond to the deterministic and Bayesian channels we call them respectively DCRB and BCRB. However, to differentiate between CRBs where we treat the symbols as deterministic and random we use respectively CRB$_{det}$ and CRB$_{sto}$. On the other hand, to differentiate between joint estimation and marginalization we use respectively CRB$_{joint}$ and CRB$_{marg}$.

7.4.1 DCRB$_{det,joint}$

In this lower bound [72] both the unknown symbols and the channel are considered as deterministic unknowns to be estimated. Hence it belongs to the first category and consequently the joint pdf is given by (7.4). Moreover, since both are deterministic we have $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h} - \boldsymbol{h}^o)$ and $f(A_U) =$

$A_U^o\delta(A_U - A_U^o)$ where $\boldsymbol{h}^o$ and A^o represent respectively the true values of the channel and the symbols. It can be easily noticed that the pdfs of both the unknown symbols and the channel have no effect on the computation of the FIM. Hence, $\ln[f(Y, A_U, \boldsymbol{h})]$ is replaced by $\ln[f(Y/A_U, \boldsymbol{h})]$ in (7.5) where $f(Y/A_U, \boldsymbol{h}) = \frac{1}{(\pi\sigma_v^2)^{Mp}} exp[-\frac{1}{\sigma_v^2}(Y - \mathcal{T}(h)A)^H(Y - \mathcal{T}(h)A)]$. After a little treatment (7.5) yields:

$$J_{\theta\theta} = \frac{1}{\sigma_v^2}\begin{bmatrix} \mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}) & \mathcal{T}_U^H(\boldsymbol{h})\mathcal{A} \\ \mathcal{A}^H\mathcal{T}_U(\boldsymbol{h}) & \mathcal{A}^H\mathcal{A} \end{bmatrix} \tag{7.13}$$

With a little manipulation we can easily show that $J_{\theta\theta^*} = 0$. Hence, by applying the Schur's complement on (7.13) we get:

$$DCRB_{det,joint} = J_{\boldsymbol{hh}}^{-1} = \sigma_v^2\left(\mathcal{A}^H P^{\perp}_{\mathcal{T}_U(\boldsymbol{h})}\mathcal{A}\right)^{-1} \tag{7.14}$$

Where $P^{\perp}_{\mathcal{T}_U(\boldsymbol{h})} = I - P_{\mathcal{T}_U(\boldsymbol{h})}$ and $P_{\mathcal{T}_U(\boldsymbol{h})} = \mathcal{T}_U(\boldsymbol{h})(\mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}))^{-1}\mathcal{T}_U^H(\boldsymbol{h})$ is the projection matrix on $\mathcal{T}_U(\boldsymbol{h})$.

7.4.2 $\mathbf{DCRB}_{sto,joint}$

The corresponding blind CRB appeared first in [61]. In this novel lower bound (see [74] for a profound analysis of its blind counterpart) we consider the unknown symbols as random with Gaussian distribution while the channel is considered deterministic to be jointly estimated with the unknown symbols. This estimator also belongs to the first category, thus the joint pdf is given by (7.4). Moreover, $f(A_U) = \frac{1}{(\pi\sigma_a^2)^{M+L-1-M_K}} exp[-\frac{A_U^H A_U}{\sigma_a^2}]$ and $f(\boldsymbol{h}) = \boldsymbol{h}^o\delta(\boldsymbol{h} - \boldsymbol{h}^o)$. It is obvious here that $\ln[f(\boldsymbol{h})]$ can be omitted without affecting the computation of FIM. Hence, (7.5) yields:

$$J_{\theta\theta} = \mathrm{E}_{Y,A_U/\boldsymbol{h}}\ \frac{1}{\sigma_v^2}\begin{bmatrix} \mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2}I_{M_U} & \mathcal{T}_U^H(\boldsymbol{h})\mathcal{A} \\ \mathcal{A}^H\mathcal{T}_U(\boldsymbol{h}) & \mathcal{A}^H\mathcal{A} \end{bmatrix} \tag{7.15}$$

Denoting $\mathrm{E}_A\{\mathcal{A}\} = \mathcal{A}_K^{'}$ and $\mathrm{E}_A\left\{\mathcal{A}^H\mathcal{A}\right\} = C_K$ where $C_k = \mathcal{A}_K^{'H}\mathcal{A}_K^{'} + M_U\sigma_a^2 I_{pL}$ and noting that $J_{\theta\theta^*=0}$, then by applying the Schur's complement on (7.15) we get:

$$DCRB_{sto,joint} = J_{\boldsymbol{hh}}^{-1} = \sigma_v^2 \left(C_K - \mathcal{A}_K^{'H}\mathcal{T}_U(\boldsymbol{h})[\mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2}I]^{-1}\mathcal{T}_U^H(\boldsymbol{h})\mathcal{A}_K^{'}\right)^{-1} \tag{7.16}$$

7.4.3 DCRB$_{sto,marg}$

This lower bound [72], [75] belongs to the second category, hence we are interested in estimating the channel and the variance of the noise only while the unknown symbols are supposed to be eliminated during the estimation process. Furthermore, the joint pdf is given by (7.12) where we consider the channel and the noise variance to be deterministic while the unknown symbols have a Gaussian distribution. Here again, $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ have no influence on computing the FIM. Substituting $f(Y/\boldsymbol{h},\sigma_v^2) = \frac{1}{(\pi)^{(M_P}|C_{YY}|}exp[-(Y-m_Y)^H C_{YY}^{-1}(Y-m_Y))]$ where $m_Y = \mathcal{T}_K(\boldsymbol{h})A_K$ and $C_{YY} = E\ (\boldsymbol{Y}-m_Y)(\boldsymbol{Y}-m_Y)^H = \sigma_a^2\mathcal{T}_U(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h})^H + \sigma_v^2 I_{M_{UP}}$ in (7.12) after omitting $\ln[f(\boldsymbol{h})]$ and $\ln[f(\sigma_v^2)]$ then with a little manipulation (7.5) and (7.6) yield:

$$\begin{aligned} &J_{\theta\theta}^{sto}(i,j) = \\ &\quad \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial\theta_i^*}C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial\theta_j^*}\right)^H\right\} + \left[\mathcal{A}_K^H C_{YY}^{-1}\mathcal{A}_K\right]_{i,j} \\ &J_{\theta\theta^*}^{sto}(i,j) = \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial\theta_i^*}C_{YY}^{-1}\left(\frac{\partial C_{YY}}{\partial\theta_j^*}\right)\right\} \end{aligned} \tag{7.17}$$

Where $[B]_{i,j}$ denotes the element that lies in the ith row and jth column of matrix B. We have used in the derivation of (7.17) the following facts: $\frac{\partial C_{YY}}{\partial \boldsymbol{h}_i^*} = \sigma_a^2\mathcal{T}_U(\boldsymbol{h})\mathcal{T}_U(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{h}_i^*})^H$, $\frac{\partial C_{YY}}{\partial\sigma_v^2} = \frac{1}{2}$, $\frac{\partial ln|C_{YY}|}{\partial\theta_i^*} = \mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial\theta_i^*}\right\}$ and $\frac{\partial}{\partial\theta_i^*}\mathrm{tr}\left\{C_{YY}^{-1}\right\} = -\mathrm{tr}\left\{C_{YY}^{-1}\frac{\partial C_{YY}}{\partial\theta_i^*}C_{YY}^{-1}\right\}$. Once we compute both $J_{\theta\theta}$ and $J_{\theta\theta^*}$ from (7.17), we substitute them in (7.8) to compute $J_{\theta_R\theta_R}$. Consequently, by using Schur's complement we can extract easily $J_{\boldsymbol{hh}}$ from $J_{\theta_R\theta_R}$ then $DCRB_{sto,marg} = J_{\boldsymbol{hh}}^{-1}$ follows directly.

7.4.4 BCRB$_{sto,joint}$

In this lower bound [76] (see also [63] for its application in cooperative-OFDM system) both the channels and the unknown symbols are assumed random with Gaussian distribution and are supposed to be estimated jointly. Hence, this lower bound in its turn belongs to the first category and its joint pdf is given by (7.4). By substituting the terms in (7.4) by their

corresponding functions we deduce the corresponding FIM as follows:

$$J_{\theta\theta} = \frac{1}{\sigma_v^2} \, \mathrm{E}_{Y,\boldsymbol{h},A} \left[\begin{array}{cc} \mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2} I_{M_U} & \mathcal{T}_U^H(\boldsymbol{h})\mathcal{A} \\ \mathcal{A}^H\mathcal{T}_U(\boldsymbol{h}) & \mathcal{A}^H\mathcal{A} + \sigma_v^2 C_h^{0-1} \end{array} \right] \tag{7.18}$$

Assuming that both the channel and the symbol distributions have a zero mean as stated above we get:

$$J_{\theta\theta} = \frac{1}{\sigma_v^2} \left[\begin{array}{cc} \mathrm{E}_{Y,\boldsymbol{h},A} \left\{ \mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}) + \frac{\sigma_v^2}{\sigma_a^2} I_{M_U} \right\} & 0 \\ 0 & C_K + \sigma_v^2 C_h^{0-1} \end{array} \right] \tag{7.19}$$

The corresponding CRB for the channel can be readily extracted from (7.19) as follows:

$$BCRB_{sto,joint} = \sigma_v^2 \left(C_K + \sigma_v^2 C_h^{0-1} \right)^{-1} \tag{7.20}$$

It is obvious that this CRB is independent of the number of training symbols used. Moreover, $\mathrm{BCRB}_{sto,joint}$ is a block diagonal matrix which means that the estimation of the channel and the symbols are decoupled. Of course, this is not true in case of semi-blind channel estimation except if all the transmitted symbols are known but in that case we are no more estimating the symbols. As a consequence, this CRB is considered to be too optimistic.

7.4.5 $\mathrm{BCRB}_{det,joint}$

This lower bound called Bayesian CRB for deterministic symbols $\mathrm{BCRB}_{det,joint}$ is novel. However, it is considered as a variation of $\mathrm{BCRB}_{sto,joint}$ that has been derived in the previous section. The main difference with $\mathrm{BCRB}_{sto,joint}$ is that we consider the symbols here to be deterministic unknowns while there we consider them to be random with Gaussian distribution. Hence, with this lower bound we introduce the concept of semi-blind Bayesian CRB for channel estimation by treating the channel as random with Gaussian distribution $f(\boldsymbol{h}) = \frac{1}{(\pi)^{pL}|C_h^o|} exp[-\boldsymbol{h}^H C_h^{o-1}\boldsymbol{h}]$. However, the unknown symbols are considered as deterministic to be jointly estimated with the channel

hence, this estimator belongs to the first category where the joint pdf is given by (7.4). Moreover, here again $\ln[f(A_U)]$ has no effect on computing FIM so it can be omitted. Therefore, (7.5) yields:

$$J_{\theta\theta} = \mathrm{E}_{Y,\boldsymbol{h}/A}\, \frac{1}{\sigma_v^2} \begin{bmatrix} \mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}) & \mathcal{T}_U^H(\boldsymbol{h})\mathcal{A} \\ \mathcal{A}^H\mathcal{T}_U(\boldsymbol{h}) & \mathcal{A}^H\mathcal{A} + \sigma_v^2 C_h^{0-1} \end{bmatrix} \tag{7.21}$$

Assuming that the channel distribution has a zero mean as stated above we get:

$$J_{\theta\theta} = \frac{1}{\sigma_v^2} \begin{bmatrix} \mathrm{E}_{Y,\boldsymbol{h}/A}\, \left\{\mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h})\right\} & 0 \\ 0 & \mathcal{A}^H\mathcal{A} + \sigma_v^2 C_h^{0-1} \end{bmatrix} \tag{7.22}$$

the corresponding CRB for the channel can be readily extracted from (7.22) as follows:

$$BCRB_{det,joint} = \sigma_v^2 \left(\mathcal{A}^H\mathcal{A} + \sigma_v^2 C_h^{0-1}\right)^{-1} \tag{7.23}$$

This CRB is also too optimistic for the same reasons discussed in the case of $\mathrm{BCRB}_{sto,joint}$.

7.4.6 $\mathbf{BCRB}_{sto,marg}$

This lower bound called Bayesian CRB for stochastic symbols ($\mathrm{BCRB}_{sto,marg}$) is by its turn novel. It belongs to the second category since the symbols are supposed to be eliminated. It can be considered as an extension to $DCRB_{sto,marg}$ by exploiting the prior information that exists about the channel. The joint pdf is given by (7.12) but this time $\ln[f(\boldsymbol{h})]$ can't be omitted. Substituting the terms in (7.12) by their corresponding functions and following the same steps mentioned in $\mathrm{DCRB}_{sto,marg}$ section we get:

$$\begin{cases} J_{\theta\theta} &= \mathrm{E}_h\, \left\{J_{\theta\theta}^{sto}\right\} + \begin{bmatrix} C_h^{0-1} & 0 \\ 0 & 0 \end{bmatrix} \\ J_{\theta\theta^*} &= \mathrm{E}_h\, \left\{J_{\theta\theta^*}^{sto}\right\} \end{cases} \tag{7.24}$$

Now we can resort to (7.8) to compute $J_{\theta_R\theta_R}$. Consequently, by using Schur's complement we can extract easily $J_{\boldsymbol{hh}}$ from $J_{\theta_R\theta_R}$ then $BCRB_{sto,marg} = J_{\boldsymbol{hh}}^{-1}$ follows directly.

7.5 Summary

Therefore, with the extension of some existing semi-blind CRBs and deriving novel ones, the picture is broadened considerably and to sum up we depict the current picture in Table 1. On the other hand, since some CRBs correspond to one realization of the channel and/or the symbols while others correspond to random channel and/or symbols then these CRBs, in their current form, are not suitable to be compared together. This problem can be overcome readily by computing the expectation of the CRBs that correspond to one realization of the channel and/or the symbols. Hence, in the simulation section we are going to compare the following: $\mathrm{E}_h \mathrm{E}_{A_U} \{DCRB_{det,joint}\}$, $\mathrm{E}_h \{DCRB_{sto,joint}\}$, $\mathrm{E}_h \{DCRB_{sto,marg}\}$, $\mathrm{E}_{A_U} \{BCRB_{det,joint}\}$, $BCRB_{sto,joint}$ $BCRB_{sto,marg}$. However, due to the difficulties that we face when we try to carry on the expectation operator in some situations, we are going to run Monte-Carlo simulations to perform the averaging over the ensemble of the symbols and/or the channel realizations.

CRB Type	Unknown Sym	Channel	Elm of Sym	Novel
$DCRB_{det,joint}$	Det	Det	No	No
$DCRB_{sto,joint}$	Gauss	Det	No	Yes
$DCRB_{sto,marg}$	Gauss	Det	Yes	No
$BCRB_{sto,joint}$	Gauss	Gauss	No	No
$BCRB_{det,joint}$	Det	Gauss	No	Yes
$BCRB_{sto,marg}$	Gauss	Gauss	Yes	Yes

Table 7.1: Summary of CRBs

7.6 Simulations

In this section we plot the different CRBs to verify some of their aspects that we mentioned in the chapter. In each Monte-Carlo simulation we generate different realizations of the channel, the symbols and the noise. As for the channel, we generate a Rayleigh fading channel with exponentially decaying power delay profile (PDP) as follows: e^{-wn} where $n = 0 : N-1$ and $w = 2$. Hence, $C_{\mathbf{h}}^{o}$ is the diagonal matrix $C_{\mathbf{h}}^{o} = I_p \otimes C$ where $C = \mathrm{diag}\{e^{-wn}, n = 0 : N-1\}$. As for the symbols, we generate random QPSK symbols to reflect the real world case. The performance of the different CRBs is evaluated by means of the Normalized MSE (NMSE) vs.

SNR. The SNR is defined as: SNR $= \frac{||\mathcal{T}(h)A||^2}{pM\,\sigma_v^2}$. The NMSE is defined as $\frac{avg\,\mathrm{tr}(CRB)}{avg\,||\boldsymbol{h}||^2}$ where avg stands for average. In Figure 7.1 we plot the NMSE of all the CRBs that have been derived in this chapter versus the number of iterations at SNR = 10 dB. To validate our comments about the looseness of the Bayesian CRBs elaborated in this chapter, we plot in the same figure the results of the algorithms derived in chapter 6 which correspond to these CRBs. To be more specific the algorithms SB-ML-ML, SB-ML-GMAP, SB-GMAP-ML, SB-GMAP-GMAP, SB-GMAP-Elm-ML and SB-GMAP-Elm-GMAP are lower bounded respectively by $\mathrm{DCRB}_{det,joint}$, $\mathrm{BCRB}_{det,joint}$, $\mathrm{DCRB}_{sto,joint}$, $\mathrm{BCRB}_{sto,joint}$, $\mathrm{DCRB}_{sto,marg}$ and $\mathrm{BCRB}_{sto,marg}$. Well, at this moderate SNR it is clear that none of the algorithms attain its corresponding CRB. This holds true also at high SNR except for $\mathrm{DCRB}_{det,joint}$ and $\mathrm{DCRB}_{sto,marg}$ which can be attained asymptotically in SNR by SB-ML-ML and SB-GMAP-Elm-ML respectively. Apart from the fact that they are loose, we can observe that all the Bayesian CRBs and $\mathrm{DCRB}_{sto,joint}$ are so close to each other. In Figure 7.2, we show how both $\mathrm{DCRB}_{det,joint}$ and $\mathrm{DCRB}_{sto,marg}$ are almost attained by their corresponding estimators at SNR = 30 dB. Furthermore, we can readily observe that at high SNR, the cost function of SB-ML-ML that has been minimized by the alternating minimization technique (see chapter 6 converges faster than that of SB-GMAP-Elm-ML where we use the scoring method to minimize its cost function. Generally speaking, we have observed that the scoring method is faster than the alternating minimization at low SNR, while at high SNR the latter is faster.

7.7 Conclusion

We introduced in chapter 5 the concept of blind Bayesian channel estimation and extended it in chapter 6 to the semi-blind case, by proposing a bunch of useful algorithms. In this chapter, we have presented a framework that permits to derive a complete set of CRBs that correspond to the various Deterministic and Bayesian cases. Some of these algorithms already exist in the literature and the others are novel. The main conclusion that can be drawn is that the Bayesian Cramér-Rao Bound is loose and there is a need for another lower bound which is tighter. This result is valid regardless of how we treat the symbols namely, deterministic or random and it is even valid when we marginalize the symbols. Furthermore, this result extends also to $\mathrm{DCRB}_{sto,joint}$ which corresponds to joint estimation of deterministic channel and random symbols. Hence, not only Bayesian CRBs but also

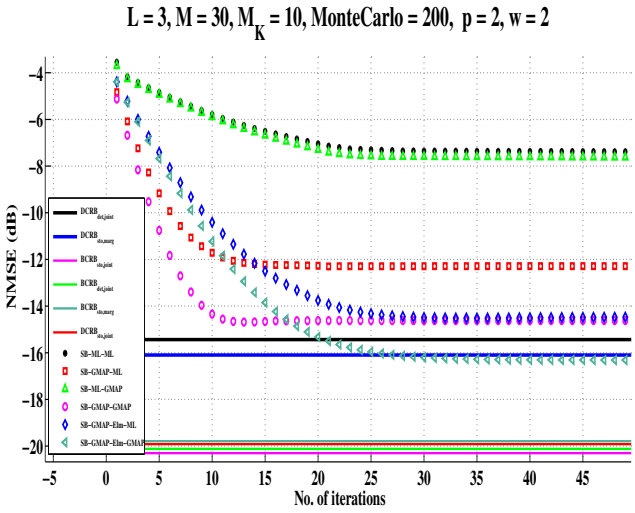

Figure 7.1: NMSE vs. No.of iterations of all CRBs at SNR = 10 dB

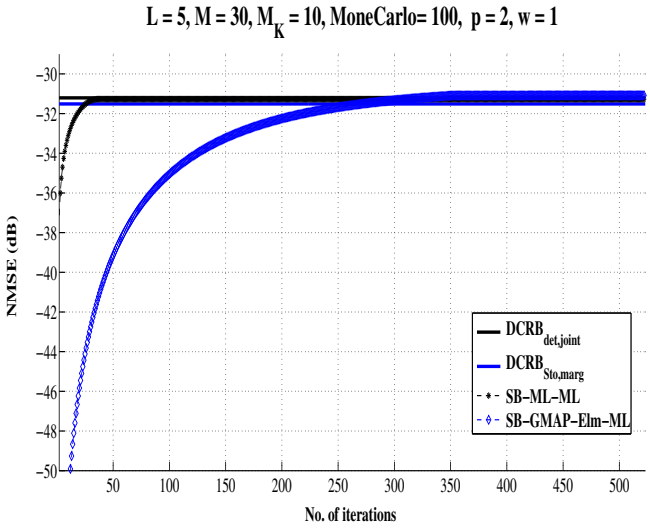

Figure 7.2: NMSE vs. No.of iterations of $\text{DCRB}_{det,joint}$ and $\text{DCRB}_{sto,marg}$ at SNR = 30 dB

some deterministic CRBs requires tighter alternatives. This point is under investigation and is subject for further research.

Chapter 8

Quasi-Bayesian Semi-Blind FIR Channel Approximation Algorithms

In this chapter, we are going to develop new methodology on how to exploit the channel prior information. Specifically, we shall exploit partially the knowledge of the power delay profile to determine the number of taps at the tail of the channel that can be neglected during the estimation process. Although we shall implement this methodology to the same collection of ML/MAP algorithms introduced in chapter 6, but this methodology has a sufficient flexibility to be implemented to all deterministic algorithms in the literature that don't require iterations for eg. SRM, signal subspace, etc. This methodology extends those deterministic algorithms into an intermediate point between deterministic and Bayesian. Hence, we call this novel framework quasi-Bayesian. We shall prove by simulations the advantages of these quasi-Bayesian algorithms in terms of improving the NMSE of the channel and the probability of error of the received symbols. Moreover, The CRBs have been derived for two quasi-Bayesian algorithms and compared to their deterministic counterparts showing how the former is lower than the latter.

8.1 Introduction

In wireless communications the channel power delay profile (pdp) is usually assumed to be exponentially decaying. This fact is based on many measurement campaigns conducted so far and specially for some emerging wireless standards [77]. Although when this pdp is known, this usually constitutes only a partial knowledge of the channel prior information, nevertheless it constitutes sometimes the full knowledge. This is true for instance, when the different channel taps are considered as i.i.d. hence, the channel prior information is represented by a block diagonal matrix. The exploitation of this channel prior information in the context of semi-blind channel estimation transfers us from deterministic into Bayesian framework. We have derived in chapter 6 some useful Bayesian algorithms and we have shown their capability to outperform their deterministic counterparts. The main drawback of those algorithms are their need to a large number of iterations before convergence is achieved. Furthermore, the Bayesian (semi-)blind channel estimation algorithms that are introduced in [34] and [33] exploit perfectly the knowledge of the channel a priori information to enhance the channel estimation quality. In this chapter, we are exploring an approach that exploits partially the knowledge of the Power Delay Profile (PDP) to enhance the estimation of a part of the channel while neglecting totally the remaining part. It is worth noting that this approach is not restricted to Bayesian algorithms but can rather be implemented to any existing deterministic algorithm. By doing so, we are extending those deterministic algorithms to a point in the middle between deterministic and Bayesian, hence we can classify them as quasi-Bayesian algorithms. The question that may be raised here, is there still a room to enhance the estimation quality of the Bayesian algorithms? Moreover, sometimes the estimation of the channel is required by itself, to be used in the beamforming for instance, while in some other cases it constitutes only one step toward another ultimate goal, the detection of symbols. Hence, one may wonder what are the consequences of neglecting a part of the channel on the detection of the symbols. In the following sections, we will try to elaborate our approach and answer these questions.

8.2 SIMO FIR Tx System Model

We will use the same model introduced in chapter 5, section 5.2. More specifically, we will start from equation (5.4) which is rewritten below with

the addition of the notations of known and unknown symbols.

$$\begin{aligned} \boldsymbol{Y} = \mathcal{T}(\boldsymbol{h})\, A + \boldsymbol{V} &= \mathcal{T}_K(\boldsymbol{h})\, A_K + \mathcal{T}_U(\boldsymbol{h})\, A_U + \boldsymbol{V} \\ &= \mathcal{A}_K \boldsymbol{h} + \mathcal{A}_U \boldsymbol{h} + \boldsymbol{V}\ . \end{aligned} \tag{8.1}$$

Where $\mathcal{T}_K(\boldsymbol{h})$ and $\mathcal{T}_U(\boldsymbol{h})$ represent respectively the portions of $\mathcal{T}(\boldsymbol{h})$ that correspond to A_k (M_K known symbols) and A_U (M_U unknown symbols).

$$\mathcal{T}(\boldsymbol{h}) = \left[\begin{array}{c|c} \mathcal{T}_K(\boldsymbol{h}) & \mathcal{T}_U(\boldsymbol{h}) \end{array}\right] \tag{8.2}$$

Here we assume for simplicity that the known symbols are gathered at the beginning of the block. On the other hand, $\mathcal{A}$ is a block Toeplitz matrix filled with the elements of A while $\mathcal{A}_K$ and $\mathcal{A}_U$ are block Toeplitz matrices filled with the elements of A_K and A_U respectively.

8.3 Channel Approximation

In this section we introduce the concept of approximating the channel by neglecting some taps at the tail during the estimation process. The neglect is justified by the fact that the estimation of those taps will introduce an estimation error that exceeds the approximation error. In fact, this is true if the power of the channel approximation error times the input is below the noise power at some finite SNR. In this case, the approximation error does not count and we have an approximated channel whose length varies with SNR. However, in order to make this channel approximation, we need to have a certain finite (and small) covariance of the part of the channel that we are going to neglect (approximation error). Hence, in a way or another this looks like a Bayesian approach. In a deterministic model, we don't indeed have any prior information about the channel that would allow to make such an approximation.

Assume that $\boldsymbol{h}$ can be split into two parts, the approximated part $\boldsymbol{h}_a$ ($pL_a \times 1$) and the neglected part $\boldsymbol{h}_n$ ($pL_n \times 1$) where $L = L_a + L_n$. Hence, we can write $\boldsymbol{h} = [\boldsymbol{h}_a^H \quad \boldsymbol{h}_n^H]^H$. The number of the neglected taps L_n should be upper bounded by $\min(L-1, M_K)$. When the length of the training sequence is greater than the number of the channel taps, the interpretation of this bound is that the approximated channel should be composed of at least one tap. Hence, the maximum number of taps that can be neglected is $L-1$. However, when the number of the channel taps is greater than the training sequence, and apart from the identifiability issues that may raise

here, the number of the neglected channel taps should not exceed M_K. This is due to the fact that every further neglected tap will lead to a one symbol loss. Actually, the size of $\mathcal{T}(\widehat{\boldsymbol{h}})$ is $pM \times (M + L - 1)$, however, when some taps are neglected, the size of the estimated channel matrix is reduced to $pM \times (M + L_a - 1)$. This reduction in the number of columns leads in general to a reduction in the number of the detected symbols. On the other hand, we are treating a semi-blind scenario where we assume that the training sequence (which fortunately there is no need to be detected) is gathered at the beginning of the block. This permits a margin of M_K symbols, at the beginning of the block, to be skipped in the detection process. On the contrary, in the blind scenario there is no allowable margin and consequently, every neglected tap will lead to a one symbol loss. This puts a severe limitation for implementing this approach in the blind scenario. Fortunately, this is no more true in the cyclic prefix case. Taking a close look at the structure of the FIR cyclic prefix channel matrix, shows that there is obviously no symbol loss due to the neglected taps. This is true because the estimated channel matrix has a size $pM \times M$ which is independent of the number of taps. In fact this feature makes our approach more attractive in the context of cyclic prefix systems. To illustrate the procedure by which the neglected channel length is determined, we start with the description of the channel model used throughout this chapter. In fact we consider a Rayleigh fading channel with exponentially decaying PDP for the channel between each transmitting and receiving antenna pair as follows: e^{-wn} where $n = 0 : L - 1$ and w is a constant that controls how fast the decaying is. Hence, if we denote by $C^o_{\mathbf{h}}$ the channel covariance matrix, which is diagonal in this case because the taps are independent, then $C^o_{\mathbf{h}} = I_p \otimes C$ where $C = \text{ diag}\,\{e^{-wn},\, n = 0 : L - 1\}$. Assume that the PDP and the variance of the noise are known (in practice they are estimated from the received signal), we start searching from the tail, for the maximum number of taps whose power times the power of the symbols is less than the variance of the noise. Mathematically, this can be written as:

$$\max_i \;\; \sigma_a^2 \sum_{j=0}^{i} C(L-j, L-j) \qquad 0 \le i \le min(L-2, M_K - 1) \qquad (8.3)$$

The above maximization is done subject to the following constraint: $\sigma_a^2 \sum_{j=0}^{i} C$ $(L-j, L-j) \le \sigma_v^2$. If we cannot find i that fulfills the above constraint, this means that we can't neglect any part of the channel. Otherwise, the last $i+1$ taps in the tail of the channel can be neglected and consequently the length of the neglected part is $(i+1) \times p$. Now, we may reformulate the

model in (8.1) as follows:

$$\begin{aligned} \boldsymbol{Y} &= \mathcal{T}(\boldsymbol{h})\, A + \boldsymbol{V} \\ &= \underbrace{\mathcal{T}(\boldsymbol{h}_a)}_{Mp\times(M+L_a-1)} A_a + \underbrace{\mathcal{T}(\boldsymbol{h}_n)}_{Mp\times(M+L_n-1)} A_n + V \\ &= \mathcal{T}(\boldsymbol{h}_a)\, A_a + Z \\ &= \mathcal{A}_a \boldsymbol{h}_a + \boldsymbol{Z}. \end{aligned} \tag{8.4}$$

where $\mathcal{T}(\boldsymbol{h}_a)$ and $\mathcal{T}(\boldsymbol{h}_n)$ are Toeplitz matrices containing the elements of $\boldsymbol{h}_a$ and $\boldsymbol{h}_n$ respectively. On the other hand, A_a constitutes the last $(M+L_a-1)$ elements of A while A_n constitutes the first $(M+L_n-1)$ elements of A. Finally, $Z = \mathcal{T}(\boldsymbol{h}_n)\, A_n + V$ is in general a spatially and temporally colored Gaussian noise with covariance R_{ZZ}. It should be noted that R_{ZZ} varies from one estimator to another, depending on how we treat A_n as we will see later. However, $\boldsymbol{h}_n$ is going to be treated always as random with Gaussian distribution.

To treat the semi-blind case correctly, we have to split the approximated channel in its turn into two parts. These two parts correspond respectively to the known and the unknown symbols in analogy to what we have done in (8.1). Hence we can write:

$$\begin{aligned} Y &= \mathcal{T}_K(\boldsymbol{h}_a)\, A_{K,a} + \mathcal{T}_U(\boldsymbol{h}_a)\, A_U + Z \\ &= \mathcal{A}_{K,a}\boldsymbol{h}_a + \mathcal{A}_{U,a}\boldsymbol{h}_a + Z \end{aligned} \tag{8.5}$$

where $\mathcal{T}_K(\boldsymbol{h}_a)$ and $\mathcal{T}_U(\boldsymbol{h}_a)$ contain the first $(M_K - L_n)$ and the last $(M + L - 1 - M_K)$ columns of $\mathcal{T}(\boldsymbol{h}_a)$ respectively. Similarly, $A_{K,a}$ and A_U contain the first $(M_K - L_n)$ and the last $(M + L - 1 - M_K)$ elements of A_a. Finally, $\mathcal{A}_{K,a}$ and $\mathcal{A}_{U,a}$ are Toeplitz matrices filled with the elements of $A_{K,a}$ and A_U respectively. It is worth noting that only $A_{K,a}$ undergoes a change compared to A_K in (8.1) while A_U remains unchanged. This is true thanks to the upper bound on the length of the neglected channel imposed in (8.3).

8.4 Enhanced Estimators

In [34] we have introduced a general framework that permitted the derivation of three Bayesian semi-blind channel estimators and another three deterministic ones. Among those estimators, there were four that jointly estimate the channel and the symbols while the remaining two were based on estimating the channel and marginalizing the symbols. In the following sections, we will show a slight variation of those estimators relying on the channel approximation approach that was introduced in the previous section. On the

other hand, there is an important difference between the model we stated in (8.1) and that used in [34] namely, in the latter we neglected a part of the received signal that contains both known and unknown symbols, whereas in this chapter we are using an optimal model that allows a proper exploitation of the training sequence and the blind part of the received data.

8.4.1 SB-ML-ML (SB-DML)

We start with SB-ML-ML or what is called SB-DML in the literature [69]. In this case, both the unknown symbols and the approximated channel are considered as deterministic unknowns to be estimated. Thus, the cost function is given by:

$$\begin{aligned} &\min_{A_U,h_a} \quad ||Y-\mathcal{T}(h_a)A||^2_{R^{-1}_{ZZ}} = \\ &\min_{A_U,h_a} \quad ||Y-\mathcal{T}_K(h_a)A_{K,a}-\mathcal{T}_U(h_a)A_U||^2_{R^{-1}_{ZZ}} \end{aligned} \tag{8.6}$$

The nonlinear LS optimization can be done by iterating between minimization with respect to A_U and $\boldsymbol{h}$. The alternating minimization strategy is as follows:

1. Initialization $\widehat{\boldsymbol{h}}_a^0$.

2. Iteration (i+1):

   - Minimization w.r.t. $A_U; \boldsymbol{h}_a = \widehat{\boldsymbol{h}}_a^{(i)}$:
     $$\min_{A_U} \quad \left\{||Y_K-\mathcal{T}_K(\widehat{h}_a^{(i)})A_K||^2+||Y_U-\mathcal{T}_U(\widehat{h}_a^{(i)})A_U||^2\right\}$$
     $$\Rightarrow \quad \widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R^{-1}_{ZZ}\mathcal{T}_U(\widehat{\boldsymbol{h}}_a^{(i)}))^{-1}\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R^{-1}_{ZZ}(\boldsymbol{Y}-\mathcal{T}_K(\widehat{\boldsymbol{h}}_a^{(i)})A_{K,a}) \tag{8.7}$$
   - Minimization w.r.t. $\boldsymbol{h}_A; A_U = \widehat{A}_U^{(i+1)}$:
     $$\min_{h_a} \quad \left\{||Y_K-\mathcal{T}_K(h_a)A_K||^2+||Y_U-\mathcal{T}_U(h_a)\,\widehat{A}_U^{(i+1)}||^2\right\}$$
     $$\Rightarrow \quad \widehat{\boldsymbol{h}_a}^{(i+1)} = (\mathcal{A}_a^{(i+1)^H}R^{-1}_{ZZ}\mathcal{A}_a^{(i+1)})^{-1}\mathcal{A}_a^{(i+1)^H}R^{-1}_{ZZ}\boldsymbol{Y} \tag{8.8}$$
     $\widehat{\mathcal{A}}_a^{(i+1)}$ is constructed from $A_{k,a}$ and $\widehat{A}_U^{(i+1)}$.

3. Repeat step 1 until $(\widehat{A}_U^{(i+1)}, \widehat{\boldsymbol{h}}_a^{(i+1)}) \approx (\widehat{A}_U^{(i)}, \widehat{\boldsymbol{h}}_a^{(i)})$.

where $R_{ZZ} = \mathcal{A}_n h_n h_n^H \mathcal{A}_n^H + \sigma_v^2 I$. As we may notice here, we need to know both h_n and A_n in order to compute R_{ZZ}. However, this estimator as well as the other estimators we are going to develop in the sequel, need to be properly initialized to converge to the global minimum. In our case, we choose to initialize all the estimators by the semi-blind Subchannel Response Matching (SRM) estimate [78]. Hence, we use the SRM estimated values of h_n and A_n to compute R_{ZZ}. The same principle is going to be implemented whenever it is necessary in the sequel.

On the other hand, we can derive the deterministic CRB that represents a lower bound for this estimator as shown in [35]. Doing so we get:

$$DCRB_{det,joint}^{app} = \left\{\mathcal{A}_a^H R_{ZZ}^{-1}\left[R_{ZZ} - \mathcal{T}_U(\boldsymbol{h}_a)\left(\mathcal{T}_U(\boldsymbol{h}_a)^H R_{ZZ}^{-1}\mathcal{T}_U(\boldsymbol{h}_a)\right)^{-1}\mathcal{T}_u(\boldsymbol{h}_a)^H\right]R_{ZZ}^{-1}\mathcal{A}_a\right\}^{-1} \tag{8.9}$$

At high SNR, there is no room to neglect any tap and we have $N_a = L$ hence $\boldsymbol{h}_a = \boldsymbol{h}$ and consequently $\boldsymbol{h}_n$ disappears completely. As a result $R_{ZZ} = \sigma_v^2 I$. Substituting this result in (8.9), we get the same formula derived in [35] for the $DCRB_{det,joint}$.

$$DCRB_{det,joint} = \sigma_v^2 \left(\mathcal{A}^H P_{\mathcal{T}_U(\mathbf{h})}^{\perp}\mathcal{A}\right)^{-1} \tag{8.10}$$

Where $P_{\mathcal{T}_U(\boldsymbol{h})}^{\perp} = I - P_{\mathcal{T}_U(\boldsymbol{h})}$ and $P_{\mathcal{T}_U(\boldsymbol{h})} = \mathcal{T}_U(\boldsymbol{h})(\mathcal{T}_U^H(\boldsymbol{h})\mathcal{T}_U(\boldsymbol{h}))^{-1}\mathcal{T}_U^H(\boldsymbol{h})$ is the projection matrix on $\mathcal{T}_U(\boldsymbol{h})$. However, at low SNR there are usually many taps at the tail of the channel that are immersed in the noise. As a consequence, they can be neglected without having any negative effect on the detection of the symbols at the receiver. On the contrary, as we will see in the simulation section, neglecting these taps enhances the detection quality at the receiver. Hence, $\boldsymbol{h}$ is approximated by $\boldsymbol{h}_a$ and there is a term that depends on $\boldsymbol{h}_n$, and that appears in R_{ZZ}. At a sufficient low SNR, $\sigma_v^2 I$ dominates R_{ZZ} so we can neglect the term that depends on $\boldsymbol{h}_n$. Substitute this result in (8.9) we get:

$$DCRB_{det,joint}^{app} \cong \sigma_v^2 \left(\mathcal{A}_a^H P_{\mathcal{T}_U(\boldsymbol{h}_a)}^{\perp}\mathcal{A}_a\right)^{-1} \tag{8.11}$$

To prove that the approximation approach, we propose in this chapter, enhances the channel estimation quality at the receiver, we compare the CRB in (8.11) with a part of the CRB matrix stated in [35] namely,

the part that corresponds to the approximated channel. Let's call this part $\widetilde{DCRB}_{det,joint}$. It is composed of the first pL_a rows and columns of $DCRB_{det,joint}$.

Knowing that the CRB is the inverse of the Fisher Information Matrix (FIM), let $\widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h})$ denotes the FIM of the first L_a taps of the channel when we estimate not only those taps but also the remaining N_n taps and $FIM_{\mathbf{h}_a\mathbf{h}_a}$ denotes the FIM of the approximated channel when we are interested only in estimating the first N_a taps. Now, we can write $\widetilde{DCRB}_{det,joint} = \widetilde{FIM}^{-1}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h})$ and $DCRB^{app}_{det,joint} = FIM^{-1}_{\mathbf{h}_a\mathbf{h}_a}$. On the other hand, it is well known that the FIM of $\boldsymbol{h}$ can be decomposed into four parts corresponding to different combinations of $\boldsymbol{h}_a$ and $\boldsymbol{h}_n$. In order to extract the $\widetilde{FIM}_{\boldsymbol{h}_a\boldsymbol{h}_a}(\boldsymbol{h})$ from these FIMs, we apply the Schur's complement so we get: $\widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h}) = FIM_{\mathbf{h}_a\mathbf{h}_a} - FIM_{\mathbf{h}_a\mathbf{h}_n}FIM^{-1}_{\mathbf{h}_n\mathbf{h}_n}FIM_{\mathbf{h}_n\mathbf{h}_a}$ Since $FIM_{\mathbf{h}_a\mathbf{h}_n}FIM^{-1}_{\mathbf{h}_n\mathbf{h}_n}FIM_{\mathbf{h}_n\mathbf{h}_a} \geq 0$ i.e. a positive semi-definite matrix, we infer that $\widetilde{FIM}_{\mathbf{h}_a\mathbf{h}_a}(\boldsymbol{h}) \leq FIM_{\mathbf{h}_a\mathbf{h}_a}$ and consequently $\widetilde{DCRB}_{det,joint} \geq DCRB^{app}_{det,joint}$. This result shows that our approximation approach leads to an enhancement in the channel estimation quality. This result has been confirmed also by numerical simulations as will be shown later.

8.4.2 SB-GMAP-ML

This estimator is considered as an extension of the corresponding blind one proposed in [61], [59]. In this estimator we treat the unknown symbols as random with Gaussian distribution, while the approximated channel is considered deterministic to be jointly estimated with the unknown symbols. Hence, the cost function is given by:

$$\min_{A_U,h_a} \; ||Y - \mathcal{T}_K(h_a)A_{K,a} - \mathcal{T}_U(h_a)A_U||^2_{R^{-1}_{ZZ}} + \frac{||A_U||^2}{\sigma_a^2}$$

Following the same methodology used in SB-ML-ML estimator we get:

$$\widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R^{-1}_{ZZ}\mathcal{T}_U(\widehat{\boldsymbol{h}}_a^{(i)}) + \frac{1}{\sigma_a^2}\,I)^{-1}\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R^{-1}_{ZZ}(\boldsymbol{Y} - \mathcal{T}_K(\widehat{\boldsymbol{h}}_a^{(i)})A_{K,a}) \tag{8.12}$$

$$\widehat{\boldsymbol{h}_a}^{(i+1)} = (\mathcal{A}_a^{(i+1)^H}R^{-1}_{ZZ}\mathcal{A}_a^{(i+1)})^{-1}\mathcal{A}_a^{(i+1)^H}R^{-1}_{ZZ}\boldsymbol{Y} \tag{8.13}$$

where $R_{ZZ} = \sigma_a^2\mathcal{T}(\boldsymbol{h}_n)\mathcal{T}(\boldsymbol{h}_n)^H + \sigma_v^2\,I$. It is worth noting that we treat A_n as random with Gaussian distribution although it contains some known

symbols. This approximation is justified by the fact that usually the number of the known symbols is small compared to the unknown symbols.

8.4.3 SB-GMAP-Elm-ML (SB-GML)

In this estimator we are only interested in estimating the approximated channel and the variance of the noise, while the unknown symbols are supposed to be eliminated during the estimation process. Hence, $\theta = [\boldsymbol{h}_a^H, \sigma_v^2]^H$. Furthermore, we consider the channel and the noise variance to be deterministic while the unknown symbols have a Gaussian distribution. Hence, the cost function is given by:

$$\min_{\boldsymbol{h}_a, \sigma_v^2} \ln|C_{YY}| + (Y - \mathcal{T}_K(\boldsymbol{h}_a)A_{k,a})^H C_{YY}^{-1}(Y - \mathcal{T}_K(\boldsymbol{h}_a)A_{k,a}) \tag{8.14}$$

where $C_{YY} = E\ (\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h}_a)A_a)(\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h}_a)A_a)^H = \sigma_a^2 \mathcal{T}_U(\boldsymbol{h}_a)\mathcal{T}_U(\boldsymbol{h}_a)^H + R_{ZZ}$ and $R_{ZZ} = \left(\sigma_a^2 \mathrm{tr}(C_{h_n}^o) + \sigma_v^2\right) I$. This cost function can be minimized by resorting to the method of scoring ([62] see also [34]).

As for deriving the CRB that corresponds to this estimator, we can follow the same methodology used in [35]. Doing so we get these formulas:

$$\begin{aligned} &J_{\theta\theta}^{sto}(i,j) = \\ &\quad \mathrm{tr}\left\{ C_{YY}^{-1} \tfrac{\partial C_{YY}}{\partial \theta_i^*} C_{YY}^{-1} \left(\tfrac{\partial C_{YY}}{\partial \theta_j^*}\right)^H \right\} + \left[\mathcal{A}_{K,a}^H C_{YY}^{-1} \mathcal{A}_{K,a}\right]_{i,j} \\ &J_{\theta\theta^*}^{sto}(i,j) = \mathrm{tr}\left\{ C_{YY}^{-1} \frac{\partial C_{YY}}{\partial \theta_i^*} C_{YY}^{-1} \left(\frac{\partial C_{YY}}{\partial \theta_j^*}\right) \right\} \end{aligned} \tag{8.15}$$

where $\frac{\partial C_{YY}}{\partial \boldsymbol{h}_{a,i}^*} = \sigma_a^2 \mathcal{T}_U(\boldsymbol{h}_a)\mathcal{T}_U(\frac{\partial \boldsymbol{h}_a}{\partial \boldsymbol{h}_{a,i}^*})^H$ and $\frac{\partial C_{YY}}{\partial \sigma_v^2} = \frac{1}{2}$. Once we compute both $J_{\theta\theta}$ and $J_{\theta\theta^*}$ from (8.15), we substitute them in ([35], eq(13)) to compute $J_{\theta_R\theta_R}$ where $\theta_R = [Re(\theta)^T\ Im(\theta)^T]^T$, Re and Im denotes Real and Imaginary respectively. Consequently, by using Schur's complement we can extract easily $J_{\boldsymbol{h}_a\boldsymbol{h}_a}$ from $J_{\theta_R\theta_R}$ then $DCRB_{sto,marg} = J_{\boldsymbol{hh}}^{-1}$ follows directly. Following the same discussion elaborated in the SB-ML-ML section, we can show that $DCRB_{sto,marg}^{app}$ is lower than $\widetilde{DCRB}_{sto,marg}$ that can be drawn from $DCRB_{sto,marg}$ ([35],eq 23) by taking the first pL_a rows and columns.

8.4.4 SB-ML-GMAP

This estimator is Bayesian since we treat the approximated channel as random with Gaussian distribution. However, the unknown symbols are con-

sidered as deterministic to be jointly estimated with the channel. Therefore, the cost function is given by:

$$\min_{A_U, \boldsymbol{h}_a} \quad ||Y - \mathcal{T}_K(h_a)A_{K,a} - \mathcal{T}_U(h_a)A_U||^2_{R_{ZZ}^{-1}} + \boldsymbol{h}_a^H C_{h_a}^{o^{-1}} \boldsymbol{h}_a \tag{8.16}$$

$$\widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R_{ZZ}^{-1}\mathcal{T}_U(\widehat{\boldsymbol{h}}_a^{(i)}))^{-1}\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R_{ZZ}^{-1}(\boldsymbol{Y} - \mathcal{T}_K(\widehat{\boldsymbol{h}}_a^{(i)})A_{K,a}) \tag{8.17}$$

$$\widehat{\boldsymbol{h}_a}^{(i+1)} = (\mathcal{A}_a^{(i+1)^H} R_{ZZ}^{-1} \mathcal{A}_a^{(i+1)} + C_{h_a}^{o^{-1}})^{-1} \mathcal{A}_a^{(i+1)^H} R_{ZZ}^{-1} \boldsymbol{Y} \tag{8.18}$$

where $R_{ZZ} = \mathcal{A}_n C_{h_n}^o \mathcal{A}_n^H + \sigma_v^2 I$ and $C_{h_n}^o$ is the part of C_h^o that corresponds to $\boldsymbol{h}_n$.

8.4.5 SB-GMAP-GMAP

In this estimator both the approximated channel and the unknown symbols are assumed random with Gaussian distribution. Moreover, they are supposed to be estimated jointly. The cost function is given by:

$$\min_{A_U, h_a} \quad ||Y - \mathcal{T}_K(h_a)A_{K,a} - \mathcal{T}_U(h_a)A_U||^2_{R_{ZZ}^{-1}} + \boldsymbol{h}_a^H C_{h_a}^{o^{-1}} \boldsymbol{h}_a + \frac{1}{\sigma_a^2}||A_U||^2$$

Also here, following the same methodology used in SB-ML-ML estimator we get:

$$\widehat{\boldsymbol{h}_a}^{(i+1)} = (\mathcal{A}_a^{(i+1)^H} R_{ZZ}^{-1} \mathcal{A}_a^{(i+1)} C_{h_a}^{o^{-1}})^{-1} \mathcal{A}_a^{(i+1)^H} R_{ZZ}^{-1} \boldsymbol{Y} \tag{8.19}$$

$$\widehat{A_U}^{(i+1)} = (\mathcal{T}_U^H(\widehat{\boldsymbol{h}}_a^{(i)})R_{ZZ}^{-1}\mathcal{T}_U(\widehat{\boldsymbol{h}}_a^{(i)}) + \frac{1}{\sigma_a^2} I)^{-1}\mathcal{T}_U^H (\widehat{\boldsymbol{h}}_a^{(i)})R_{ZZ}^{-1}(\boldsymbol{Y} - \mathcal{T}_K(\widehat{\boldsymbol{h}}_a^{(i)})A_{K,a}) \tag{8.20}$$

where $R_{ZZ} = \left(\sigma_a^2 \mathrm{tr}(C_{h_n}^o) + \sigma_v^2\right) I$.

8.4.6 SB-GMAP-Elm-GMAP

As in the case of SB-GMAP-Elm-ML, in this estimator the symbols are supposed to be eliminated. Hence, it can be considered as an extension of

SB-GMAP-Elm-ML by exploiting the prior information that exists about the channel. The corresponding cost function is given by:

$$\min_{\boldsymbol{h}_a,\sigma_v^2} \ln|C_{YY}| + (Y - \mathcal{T}_K(\boldsymbol{h}_a)A_{k,a})^H C_{YY}^{-1}(Y - \mathcal{T}_K(\boldsymbol{h}_a)A_{k,a}) + \boldsymbol{h}_a^H C_{h_a}^{o^{-1}} \boldsymbol{h}_a \tag{8.21}$$

This cost function can also be minimized using the scoring method.

8.5 Simulations

In this section, we show, by means of MonteCarlo simulations, how our approach for approximating the channel leads to a superior performance compared to classical techniques. In each MonteCarlo simulation we generate a Rayleigh fading channel as discussed previously while for the symbols, we generate random 8PSK symbols to reflect the real world case. The performance of the different channel estimators is evaluated by means of the Normalized MSE (NMSE) vs. SNR. The SNR is defined as: SNR $= \frac{||\mathcal{T}(h)A||^2}{pM\,\sigma_v^2}$. The NMSE is defined as $\frac{avg\,||\boldsymbol{h}_a - \hat{\boldsymbol{h}}_a||^2}{avg\,||\boldsymbol{h}_a||^2}$. All the simulations are initialized by the semi-blind Subchannel Response Matching (SRM) estimate [78]. In Figure 8.1 we compare the performance of the SB-SRM and the SB-ML-ML estimators with their enhanced counterparts proposed in this chapter. We can notice how the SB-SRM based on our approach, (SB-SRM-Approx), exceeds its counterpart (SB-SRM) by more than 7 dB at low SNR and by couple of dBs at moderate SNR. However, no more enhancement is possible at very high SNR because, as explained before, no taps can be neglected at this SNR. As for our SB-ML-ML-Approx, we can notice from the same figure that only an enhancement of 2.5 dB is possible at low SNR compared to its counterpart SB-ML-ML, while this advantage diminishes as SNR increases. On the other hand, we plot on the same figure both the $\widetilde{DCRB}_{det,joint}$ and $DCRB^{app}_{det,joint}$. We notice that the latter exceeds the former by around 5 dB at low SNR which means that there is a considerable room to enhance the estimators that treat the channel and the symbols as deterministic. However, we notice that SB-SRM-Approx, and not SB-ML-ML-Approx succeeds well in taking advantage of our approach and fills the gap between both CRBs. The result is somehow surprising because SB-SRM-Approx is considered as a non-weighted version of SB-ML-ML-Approx. Finally, it is obvious that our approach leads the SB-ML-ML-Approx to almost attain the $\widetilde{DCRB}_{det,joint}$.

It is well known that the latter is only attainable by SB-ML-ML asymptotically in SNR while it is not attainable asymptotically in the number of data. In Figure 8.2 we compare SB-GMAP-Elm-ML with our proposed counterpart. It is clear that the gain offered by our approach (6 dB) at low to moderate SNR is tremendous. Also on the same figure, we plot both $\widetilde{DCRB}_{sto,marg}$ and $DCRB^{app}_{sto,marg}$. Once again, we can notice that our approach leads to a lower bound (2 dB). It is interesting to note here also that our approach leads SB-GMAP-Elm-Ml-Approx to attain $\widetilde{DCRB}_{sto,marg}$ which is not attainable by SB-GMAP-Elm-ML. In Figure 8.3 we can observe once again the great enhancement (5 dB) obtained by our approximation approach at low SNR, specially in the SB-ML-GMAP case whereas in the SB-GMAP-ML case the gain is around 2 dB. In Figure 8.4 we show numerically that SB-GMAP-GMAP (which jointly estimate the channel and the symbols) and SB-GMAP-Elm-GMAP (which estimates the channel and marginalizes the symbols) are perfect in the sense that our approach for approximating the channel is not capable of enhancing their performance.

In all the simulations we have conducted up till now the PDP is assumed to be known perfectly. However, in Figure 8.5 we estimate the PDP from the received data and we apply our approach using the SB-SRM algorithm. We compare the enhancement obtained by our approach relying on the estimated PDP against the perfect PDP. We observe that although the improvement degrades when we use the estimated PDP but the reduction in the NMSE compared to the traditional SB-SRM is still interesting. At last, to prove that our approach leads also to an enhancement in the probability of symbol error (Pe) (an MMSE equalizer is used), we plot in Figure 8.6 the Pe for SB-SRM and an enhanced version of it based on our channel approximation approach. We can readily observe the considerable gain (2 dB) offered by our approach at medium and low SNR.

8.6 Conclusion

We have introduced in the context of semi-blind channel estimation a new approach that relies on the partial exploitation of the PDP of the channel (assumed known or estimated from the received data) to reduce the channel estimation error. Based on this approach, we have shown that, by neglecting some taps at the tail of the channel that are immersed in noise, the quality of the channel estimation has been improved considerably. The proposed approach has been implemented to a series of deterministic and Bayesian estimators introduced previously. We have shown by numerical simulations

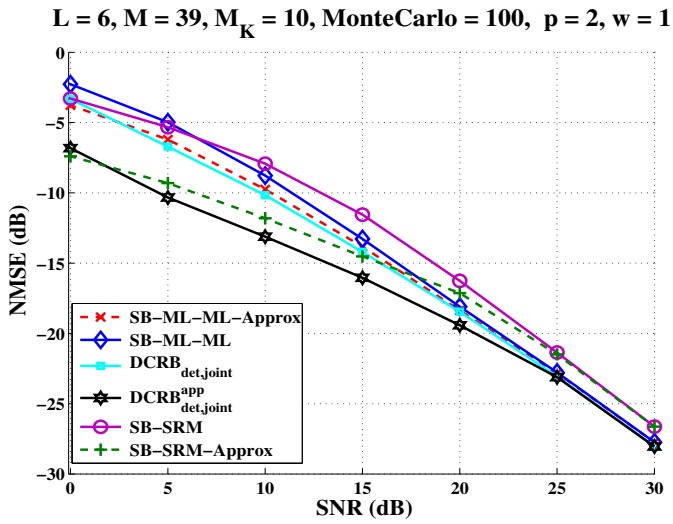

Figure 8.1: NMSE vs. SNR for SB-SRM, SB-ML-ML, $DCRB_{det,joint}$ and $DCRB^{app}_{det,joint}$.

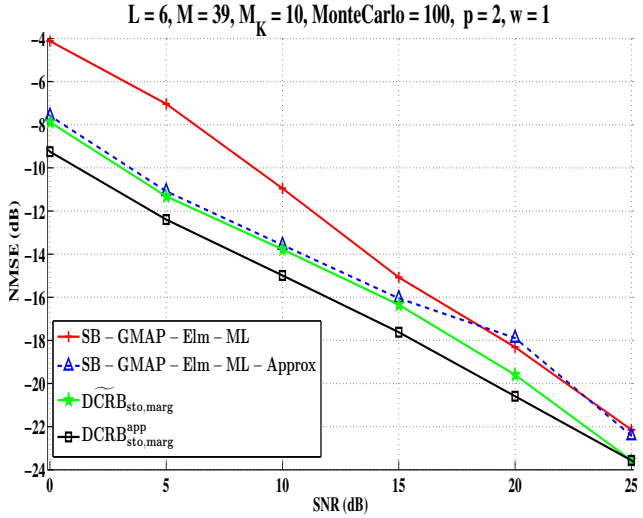

Figure 8.2: NMSE vs. SNR for SB-GMAP-Elm-ML, $DCRB_{sto,joint}$ and $DCRB^{app}_{sto,marg}$.

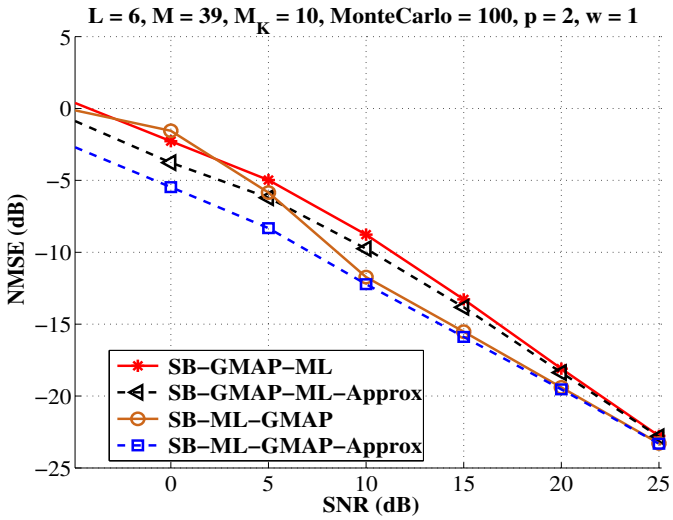

Figure 8.3: NMSE vs. SNR for SB-ML-GMAP and SB-GMAP-ML.

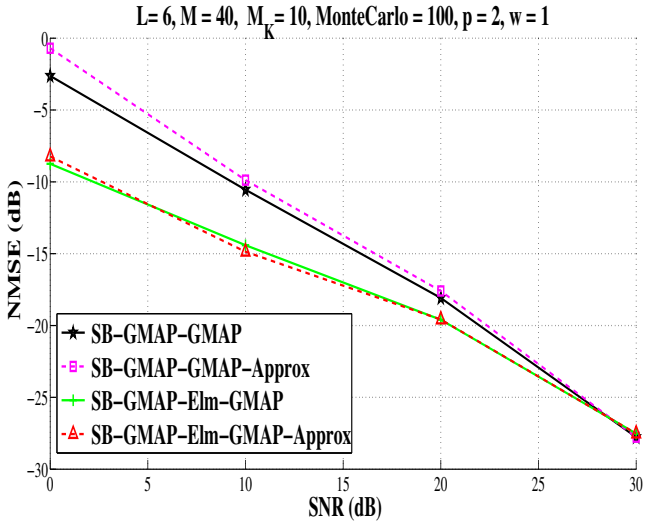

Figure 8.4: NMSE vs. SNR for SB-GMAP-GMAP and SB-GMAP-Elm-GMAP.

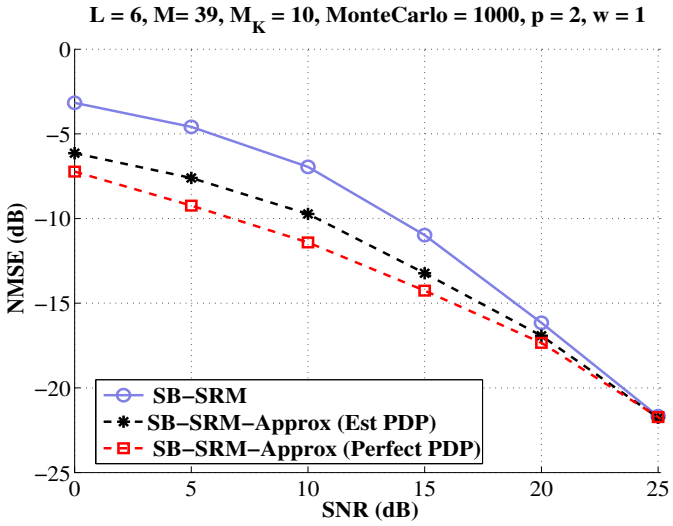

Figure 8.5: NMSE vs. SNR for SB-SRM with perfect and estimated PDP.

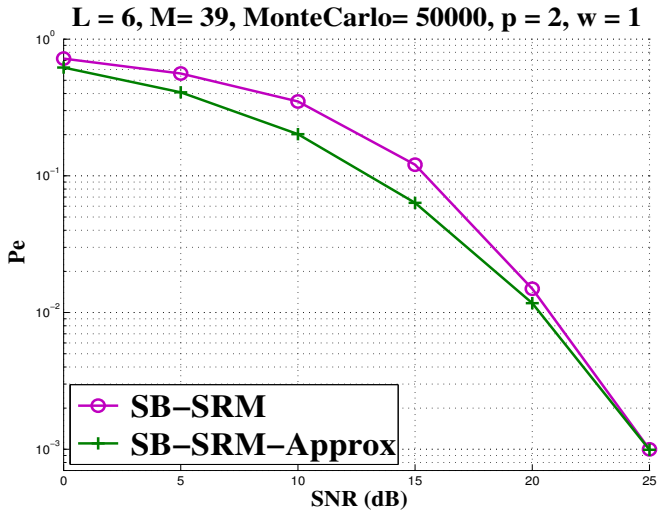

Figure 8.6: Probability of error vs. SNR for SB-SRM and its enhanced counterpart.

that there is a great enhancement in the NMSE over a wide range of SNR. Moreover, we have shown analytically that the CRBs of two of the proposed estimators are lower than their corresponding ones that exist in the literature. Finally, we have shown also numerically that not only the NMSE of the channel has been improved but also the probability of error of the detected symbols. On the other hand, our simulations show that there is no room left to enhance the estimators that take full advantage of the prior information about the channel and the symbols. This fact has been reflected in both SB-GMAP-GMAP and SB-GMAP-Elm-GMAP performance where our approach has not succeeded to reduce their NMSE.

Part III

Diversity For Different Systems

Chapter 9

Receiver Diversity With Blind And Semi-Blind Channel Estimates

Traditionally, the performance of blind SIMO channel estimates has been characterized in a deterministic fashion, by identifying those channel realizations that are not blindly identifiable. In this chapter, we focus instead on the performance of Zero-Forcing (ZF) Linear Equalizers (LEs) or Decision-Feedback Equalizers (DFEs) for fading channels when they are based on (semi-)blind channel estimates. Although it has been known that various (semi-)blind channel estimation techniques have a receiver counterpart that is matched in terms of symbol knowledge hypotheses, we show here that these (semi-)blind techniques and corresponding receivers also match in terms of diversity order: the channel becomes (semi-)blindly unidentifiable whenever its corresponding receiver structure goes in outage. In the case of mismatched receiver and (semi-blind) channel estimation technique, the lower diversity order dominates. Various cases of (semi-)blind channel estimation and corresponding receivers are considered in detail. To be complete however, the actual combination of receiver and (semi-)blind channel estimation lowers somewhat the diversity order w.r.t. the ideal picture.

9.1 Introduction

Consider a linear modulation scheme and single-carrier transmission over a Single Input Multiple Output (SIMO) linear channel with additive white noise. The multiple (subchannel) outputs will be mainly thought of as corresponding to multiple antennas. After a receive (Rx[1]) filter (possibly noise whitening), we sample the Rx signal to obtain a discrete-time system at the symbol rate[2]. When stacking the samples corresponding to multiple Rx antennas in column vectors, the discrete-time communication system is described by

$$\underbrace{\boldsymbol{y}_m}_{p\times 1} = \underbrace{\boldsymbol{h}[q]}_{p\times 1}\,\underbrace{\boldsymbol{a}_m}_{1\times 1} + \underbrace{\boldsymbol{v}_m}_{p\times 1} \tag{9.1}$$

where m is the symbol (sample) period index, p is the number of Rx antennas. The noise power spectral density matrix is $S_{\boldsymbol{vv}}(z) = \sigma_v^2\, I_p$, q^{-1} is the unit sample delay operator: $q^{-1}\, a_m = a_{m-1}$, and $\boldsymbol{h}[z] = \sum_{i=0}^{L-1} \boldsymbol{h}_i\, z^{-i}$ is the SIMO channel transfer function in the z domain. The channel delay spread is L symbol periods. In the Fourier domain we get the vector transfer function $\boldsymbol{h}(f) = \boldsymbol{h}[e^{j2\pi f}]$.

We introduce the vector containing the SIMO impulse response coefficients $\mathbf{h}^\dagger[z] = \mathbf{h}^H[1/z^*]$, it denotes the paraconjugate (matched filter). Note that $\mathbf{h}^\dagger[e^{j2\pi f}] = \mathbf{h}^H(f)$. $\mathbf{h} = [\mathbf{h}_0^T \cdots \mathbf{h}_{L-1}^T]^T$. Assume the energy normalization $\mathrm{tr}\{R_{\mathbf{hh}}\} = p$ with $R_{\mathbf{hh}} = \mathrm{E}\{\mathbf{h}\mathbf{h}^H\}$. By default we shall assume the i.i.d. complex Gaussian channel model: $\mathbf{h} \sim \mathcal{CN}(0, \frac{1}{L} I_{pL})$ so that spatio-temporal diversity of order $p\,(L)$ is available (which is the case from the moment $R_{\mathbf{hh}}$ is nonsingular). The average SNR per Rx antenna is $\rho = \sigma_a^2/\sigma_v^2$.

Whereas in non-fading channels, the probability of error P_e decreases exponentially with SNR, for a given symbol constellation, in fading channels the probability of error taking channel statistics into account behaves as $P_e \sim \rho^{-d}$ for large SNR ρ, where d is the diversity order. Also, at high SNR, the P_e is dominated by the outage probability P_o and has the same diversity order for a well-designed system. If the data rate R is adapted with SNR such that we get a normalized rate $r = \lim\limits_{\rho\to\infty} \dfrac{R}{\rho} \in [0, 1]$, then the diversity becomes $d(r)$ [79]. For all ZF Rx's considered in this chapter, we get the following Diversity-Multiplexing Trade off (DMT): $d(r) = d(0)(1-r)$.

[1] In this chapter, "Rx" stands for "receive" or "receiver" or "reception" etc., and similarly for "Tx" and "transmit", ...

[2] In the case of additional oversampling with integer factor w.r.t. the symbol rate, the Rx dimension would get multiplied by the oversampling factor.

Hence it suffices to limit the diversity analysis to the fixed rate R case with diversity $d(0) = d$.

In practice also the Linear Equalizer (LE) is often used because of low detection complexity. Also in practice, for both LE and DFE, only a limited degree of non-causality (delay) can be used and the filters are usually of finite length (FIR). Analytical investigations into the diversity for SISO with LEs are much more recent, see [80],[81] for linearly precoded OFDM and [82] for Single-Carrier with Cyclic Prefix (SC-CP). The DMT for various forms of LE and DFE with frequency-selective SIMO channels is investigated in [83]. In [80], it was shown that the introduction of redundant linear precoding in OFDM allows a MMSE-ZF linear block receiver to regain full diversity in the SISO (or SIMO) case. For instance Zero Padding (ZP) introduces redundancy in the time (delay) dimension which allows a LE of inter-symbol interference (ISI) to maintain full diversity: every input symbol can be recovered linearly unless the whole channel impulse response becomes zero. In all the references mentioned above the channel was assumed to be perfectly known at the Rx and in some cases at the Tx too. However, practical receivers must estimate the channel, thereby incurring estimation error that needs to be accounted for in the performance analysis. In [37] we treated the effect of blind channel estimation on the diversity of ZF-LE within the context of SIMO Tx system. However, we focused there more on the effect of the constraint usually used to handle the ambiguity that results from blind channel estimation. In [84] the effect of channel estimation error on the performance of the Viterbi equalizer is studied in a SIMO framework. In [85] the bit-error rate (BER) performance of multilevel quadrature amplitude modulation with pilot-symbol-assisted modulation channel estimation in static and Rayleigh fading channels is derived, both for single branch reception and maximal ratio combining diversity receiver systems. However, in [86] it is shown that the(practical) ML channel estimator preserves the diversity order of MRC (Maximum Ratio Combining), see also [87] for a more thorough analysis.

In this chapter we assume the channel to be estimated at the Rx using blind and semi-blind deterministic algorithms. We investigate the effect of the resulting channel estimation error on the diversity achieved by the corresponding equalizers (matched to the channel estimation hypotheses).

9.2 Outage Analysis of Suboptimal Receiver SINRs

A perfect outage occurs when SINR = 0. For the Matched Filter Bound (MFB) this can only occur if $\mathbf{h} = 0$. For a suboptimal Rx however, the SINR = SINR($\mathbf{h}$) can vanish for any $\mathbf{h}$ on the *Outage Manifold* $\mathcal{M} = \{\mathbf{h} : \text{SINR}(\mathbf{h}) = 0\}$. At fixed rate R, the diversity order is the codimension of (the tangent subspace of) the outage manifold, assuming this codimension is constant almost everywhere and assuming a channel distribution with finite positive density everywhere (e.g. Gaussian with non-singular convariance matrix). For example, for the MFB (which only depends on $\mathbf{h}$) the outage manifold is the origin, the codimension of which is the total size of $\mathbf{h}$. The codimension is the (minimum) number of complex constraints imposed on the complex elements of $\mathbf{h}$ by putting SINR($\mathbf{h}$) = 0. Some care has to be excercised with complex numbers. Valid complex constraints (which imply two real constraints) are such that their number becomes equal to the number of real constraints if the channel coefficients were to be real. A constraint on a coefficient magnitude however, which is in principle only one real constraint, counts as a valid complex constraint (at least if the channel coefficient distributions are insensitive to phase changes). For ZF equalizers, consideration of the outage manifold is suffucient. For MMSE equalizers however, a more complete analysis is required. An actual outage occurs whenever the rate exceeds the capacity, $\log(1 + \text{SINR}) < R$, which occurs when $\mathbf{h}$ lies in the *Outage Shell*, a (thin) shell containing the outage manifold. The thickness of this shell shrinks as the rate increases and depends also on the regularization appearing in MMSE equalizers.

9.3 Blind (B) and Semi-Blind (SB) Channel Estimation and Matched ZF Equalization

Consider a block Tx system with Rx signal in the time domain [18]

$$\boldsymbol{Y} = \mathcal{T}(\boldsymbol{h})\ \boldsymbol{A} + \boldsymbol{V} = \mathcal{T}_K(\boldsymbol{h})\ \boldsymbol{A}_K + \mathcal{T}_U(\boldsymbol{h})\ \boldsymbol{A}_U + \boldsymbol{V} = \boldsymbol{\mathcal{A}}\mathbf{h} + \boldsymbol{V} \qquad (9.2)$$

where $\boldsymbol{A}$ is the vector of Tx symbols, containing possibly known symbols $\boldsymbol{A}_K$ (training/pilots, semi-blind case) and unknown symbols $\boldsymbol{A}_U$ (actual data, i.i.d. with variance σ_a^2). $\mathcal{T}(\boldsymbol{h})$ is the channel convolution matrix of which the part $\mathcal{T}_K(\boldsymbol{h})$ is affected by $\boldsymbol{A}_K$ and the part $\mathcal{T}_U(\boldsymbol{h})$ is affected by $\boldsymbol{A}_U$. Due to the commutativity of convolution, $\mathcal{T}(\boldsymbol{h})\,\boldsymbol{A} = \boldsymbol{\mathcal{A}}\,\mathbf{h}$ in which $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}(\boldsymbol{A}_K, \boldsymbol{A}_U)$ and $\mathbf{h}$ contains the vectorized channel impulse response

coefficients. $\boldsymbol{V}$ is the AWGN with variance σ_v^2. Even though we shall investigate the diversity of receivers due to fading channels, for (semi-)blind channel estimation purposes, the channel $\mathbf{h}$ is considered a deterministic unknown. In the (semi-)blind techniques considered here, also $\boldsymbol{A}_U$ is considered a deterministic unknown. $\mathcal{T}_U(\boldsymbol{h})$ and $\boldsymbol{\mathcal{A}}$ are assumed to have full column rank w.p. 1 when $\mathbf{h}$ and $\boldsymbol{A}$ would be considered random.

Maximum likelihood (ML) estimation of $\mathbf{h}$ (with $\boldsymbol{A}_U$ as nuisance parameters) leads to the least-squares cost function [88]

$$\min_{\mathbf{h},\boldsymbol{A}_U} \|\boldsymbol{Y} - \mathcal{T}(\boldsymbol{h})\boldsymbol{A}\|^2 \,. \tag{9.3}$$

As this cost function is separable [88], we can first optimize w.r.t. $\boldsymbol{A}_U$, which leads to

$$\widehat{\boldsymbol{A}}_U = (\mathcal{T}_U(\boldsymbol{h})^H \mathcal{T}_U(\boldsymbol{h}))^{-1} \mathcal{T}_U(\boldsymbol{h})^H (\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h})\, \boldsymbol{A}_K) \,. \tag{9.4}$$

In the semi-blind case ($\boldsymbol{A}_K \neq 0$), this is a particular form of a MMSE-ZF block DFE, with feedback only from the known symbols $\boldsymbol{A}_K$. Here, the diversity of a DFE will only get analyzed with a matched semi-blind channel estimate, in which the feedback symbols play the role of pilots. In the blind case, $\boldsymbol{A}_K = 0\,,\ \mathcal{T}_U(\boldsymbol{h}) = \mathcal{T}(\boldsymbol{h})$ and (9.4) corresponds to a MMSE-ZF block LE. The ML (semi-)blind channel estimate is obtained by minimizing (9.3) after having plugged in (9.4), leading to

$$\widehat{\mathbf{h}} = \arg\min_{\mathbf{h}} \|P^{\perp}_{\mathcal{T}_U(\boldsymbol{h})}(\boldsymbol{Y} - \mathcal{T}_K(\boldsymbol{h})\boldsymbol{A}_K\|^2 \tag{9.5}$$

where we introduced the projection matrices $P^{\perp}_{\mathcal{T}(\boldsymbol{h})} = I - P_{\mathcal{T}(\boldsymbol{h})}$ and $P_{\mathcal{T}(\boldsymbol{h})} = \mathcal{T}(\boldsymbol{h})(\mathcal{T}(\boldsymbol{h})^H\mathcal{T}(\boldsymbol{h}))^{-1}\mathcal{T}(\boldsymbol{h})^H$. Note that the Rx diversity with (semi-)blind channel estimate to be considered here is not restricted to only ML channel estimates however; any other (semi-)blind method that exploits the same information will lead to similar diversity results.

The Fisher Information Matrix (FIM) for the joint estimation of $\boldsymbol{\theta} = [\boldsymbol{A}_U^H\ \mathbf{h}^H]^H$ is

$$FIM^{SB}_{joint} = \frac{1}{\sigma_v^2}\,[\mathcal{T}_U(\boldsymbol{h})\ \ \boldsymbol{\mathcal{A}}]^H\,[\mathcal{T}_U(\boldsymbol{h})\ \ \boldsymbol{\mathcal{A}}] \,. \tag{9.6}$$

The marginal Cramer-Rao Bound (CRB) for $\boldsymbol{A}_U$ (treating $\mathbf{h}$ as nuisance parameters) is

$$CRB^{SB}_{\boldsymbol{A}_U} = \sigma_v^2 \left(\mathcal{T}_U(\boldsymbol{h})^H P^{\perp}_{\boldsymbol{\mathcal{A}}} \mathcal{T}_U(\boldsymbol{h})\right)^{-1} \tag{9.7}$$

while for $\mathbf{h}$ (treating $\boldsymbol{A}_U$ as nuisance parameters), it is

$$CRB_{\mathbf{h}}^{SB} = \sigma_v^2 \left(\boldsymbol{\mathcal{A}}^H P^{\perp}_{\mathcal{T}_U(\boldsymbol{h})} \boldsymbol{\mathcal{A}} \right)^{-1} \tag{9.8}$$

in which the inverses become pseudo-inverses in the blind case or in the semi-blind case with insufficient pilots [89]. On the other hand, if the channel is known (full Channel State Information at the Rx (CSIR)), the CRB for $\boldsymbol{A}_U$ becomes

$$CRB_{\boldsymbol{A}_U}^{CSIR} = \sigma_v^2 \left(\mathcal{T}_U(\boldsymbol{h})^H \mathcal{T}_U(\boldsymbol{h}) \right)^{-1} \ . \tag{9.9}$$

The CRB for symbol k in $\boldsymbol{A}_U$ provides a lower bound on the symbol estimation (reception) error variance, which leads to an SINR upper bound

$$\mathrm{SINR}_k = \frac{\sigma_a^2}{(CRB_{\boldsymbol{A}_U})_{k,k}} \tag{9.10}$$

In the case of full CSIR, this is not an upper bound but the correct SINR. In the (semi-)blind case, the bound becomes tight at high SNR, which is the regime of interest for diversity analysis. Now, we get $\mathrm{SINR}_k^{CSIR} = 0$ whenever $\mathcal{T}_U(\boldsymbol{h})$ loses full column rank, in which case $\mathcal{T}_U(\boldsymbol{h})^H \mathcal{T}_U(\boldsymbol{h})$ becomes singular. The number of constraints that this loss of column rank imposes on $\mathbf{h}$ will be the diversity order. This diversity order will be considered in detail for various cases in the further sections.

Now, considering SINR_k^{SB} (see (9.8), (9.9) also), we get $\mathrm{SINR}_k^{SB} = 0$ whenever $\mathrm{SINR}_k^{CSIR} = 0$. Hence the diversity order of the Rx with (semi-)blind channel estimate will be at most that of the Rx with full CSIR. The Rx signal dimension reduction due to the projection $P^{\perp}_{\boldsymbol{\mathcal{A}}}$ on the noise subspace leads to some reduction in diversity order. Note that due to the randomness of $\boldsymbol{A}$, the orientation of the subspaces considered is random. Due to this randomness, the effect of this reduction should become negligible whenever the relative effect of this dimension reduction becomes negligible, namely whenever the ratio of channel delay spread over block length becomes small.

In the paradigm of matched (semi-)blind channel estimate and Rx considered so far, the channel estimation and the data reception are based on the same data block. However, simulations show that the diversity to be analyzed does not change when the channel estimation and data reception are performed on disjoint data blocks, where the Rx for one data block is constructed with the channel estimate from a different data block (see the next section also). This would indicate that the diversity effect of the (semi-)blind channel estimate dominates.

9.4 General Treatment of the Case of Non-Matched Receivers

The channel impulse response $\boldsymbol{h}$ can be decomposed into its estimate $\widehat{\boldsymbol{h}}$ and its estimation error $\widetilde{\boldsymbol{h}}$: $\boldsymbol{h} = \widehat{\boldsymbol{h}} + \widetilde{\boldsymbol{h}}$. In the (semi-)blind case, $\widehat{\boldsymbol{h}}$ represents the channel estimate in which possible ambiguities have been resolved. This channel decomposition leads to the following signal model

$$\boldsymbol{Y} = \mathcal{T}(\widehat{\boldsymbol{h}})\ \boldsymbol{A} + \mathcal{T}(\widetilde{\boldsymbol{h}})\ \boldsymbol{A} + \boldsymbol{V} = \mathcal{T}_K(\widehat{\boldsymbol{h}})\ \boldsymbol{A}_K + \mathcal{T}_U(\widehat{\boldsymbol{h}})\ \boldsymbol{A}_U + \boldsymbol{Z} \tag{9.11}$$

where $\boldsymbol{Z} = \mathcal{T}(\widetilde{\boldsymbol{h}})\ \boldsymbol{A} + \boldsymbol{V} = \boldsymbol{\mathcal{A}}\widetilde{\boldsymbol{h}} + \boldsymbol{V}$ has covariance matrix $R_{\boldsymbol{ZZ}} = \mathrm{E}_{\boldsymbol{A}_U} \boldsymbol{\mathcal{A}} R_{\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}} \boldsymbol{\mathcal{A}}^H + \sigma_v^2 I$ (if we assume that the channel estimate is obtained from data independent of the $\boldsymbol{Y}$ considered here, to make $\widetilde{\boldsymbol{h}}$ and $\boldsymbol{V}$ independent). If we treat $\boldsymbol{Z}$ as Gaussian noise that is independent of $\widehat{\boldsymbol{h}}$ and $\boldsymbol{A}_U$, then we get a capacity (or mutual information (MI)) lower bound (that is fairly tight). The correlations in $R_{\boldsymbol{ZZ}}$ depend on the correlations $R_{\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}}$ in the channel estimation error, but may get suppressed by the averaging over $\boldsymbol{A}_U$, depending on the structure of $\boldsymbol{\mathcal{A}}$. As far as the independence of $\boldsymbol{Z}$ and $\boldsymbol{A}_U$ is concerned, this independence is correct if we estimate the channel from one Rx block and use that channel to detect the symbols in a different Rx block (with independent data). In any case, considering outage probability, the MI lower bound leads to a diversity order upper bound.

Whereas the considerations so far pave the way to consider arbitrary Rx structures, in what follows we shall again focus on matched Rx structures (but applied to different data blocks). Thus, a MMSE-ZF ($\delta = 0$) or MMSE ($\delta = 1$) LE/DFE output is obtained as

$$\widehat{\boldsymbol{A}}_U = (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H R_{\boldsymbol{ZZ}}^{-1} \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta\, \sigma_a^{-2} I)^{-1} \mathcal{T}_U(\widehat{\boldsymbol{h}})^H R_{\boldsymbol{ZZ}}^{-1} (\boldsymbol{Y} - \mathcal{T}_K(\widehat{\boldsymbol{h}})\, \boldsymbol{A}_K) \tag{9.12}$$

with resulting error covariance matrix

$$R_{\widetilde{\boldsymbol{A}}_U \widetilde{\boldsymbol{A}}_U} = (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H R_{\boldsymbol{ZZ}}^{-1} \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta\, \sigma_a^{-2} I)^{-1}\ . \tag{9.13}$$

At least, this expression becomes correct at high SNR, where we can limit the expression to first order terms in σ_v^2 and where $\mathcal{T}(\widetilde{\boldsymbol{h}})\boldsymbol{A}$ and $\boldsymbol{A}_U$ become decorrelated as $\widetilde{\boldsymbol{h}}$ becomes linear in the noise. The resulting SINR for symbol k in $\boldsymbol{A}_U$ is:

$$\mathrm{SINR}_k^{MMRx} = \frac{\sigma_a^2}{\left(R_{\widetilde{\boldsymbol{A}}_U \widetilde{\boldsymbol{A}}_U}\right)_{k,k}} - \delta\ . \tag{9.14}$$

where MMRx stands for either MMSE-ZF or MMSE depending on the value of δ. Practically, $R_{\boldsymbol{ZZ}}$ is not known because it depends on the true channel through $R_{\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}} = R_{\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}}(\boldsymbol{h})$. However, at high SNR, one can equivalently use $R_{\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}}(\widehat{\boldsymbol{h}})$. A different complication rises when the channel gets estimated and the Rx symbols get detected from the same Rx signal block. In that case the expression for $R_{\boldsymbol{ZZ}}$ needs to be modified in order to account for the correlation between $\widetilde{\boldsymbol{h}}$ and $\boldsymbol{V}$. Finally, one has to admit that accounting for $R_{\boldsymbol{ZZ}}$ in the Rx as in (9.12) complicates the Rx quite a bit. To avoid all these complications, one could consider the simplified Rx

$$\widehat{\boldsymbol{A}}_U^s = (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta \, \frac{\sigma_v^2}{\sigma_a^2} I)^{-1} \mathcal{T}_U(\widehat{\boldsymbol{h}})^H (\boldsymbol{Y} - \mathcal{T}_K(\widehat{\boldsymbol{h}}) \, \boldsymbol{A}_K) \tag{9.15}$$

which corresponds to ignoring $\mathcal{T}(\widetilde{\boldsymbol{h}}) \; \boldsymbol{A}$ and hence using $R_{\boldsymbol{ZZ}} = R_{\boldsymbol{VV}} = \sigma_v^2 I$. Now further neglecting $\mathcal{T}(\widetilde{\boldsymbol{h}}) \; \boldsymbol{A}$ leads to a symbol estimation error covariance matrix lower bound $R^s_{\widetilde{\boldsymbol{A}}_U \widetilde{\boldsymbol{A}}_U} = \sigma_v^2 (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta \, \frac{\sigma_v^2}{\sigma_a^2} I)^{-1}$ and to a corresponding SINR upper bound

$$\mathrm{SINR}_k^{MMRxs} = \frac{\sigma_a^2}{\left(R^s_{\widetilde{\boldsymbol{A}}_U \widetilde{\boldsymbol{A}}_U} \right)_{k,k}} - \delta \; . \tag{9.16}$$

A perhaps more accurate approximation would be $R^s_{\widetilde{\boldsymbol{A}}_U \widetilde{\boldsymbol{A}}_U} = (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta \frac{\sigma_v^2}{\sigma_a^2} I)^{-1} (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H R_{\boldsymbol{ZZ}} \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta \frac{\sigma_v^4}{\sigma_a^2} I) (\mathcal{T}_U(\widehat{\boldsymbol{h}})^H \mathcal{T}_U(\widehat{\boldsymbol{h}}) + \delta \frac{\sigma_v^2}{\sigma_a^2} I)^{-1}$. Our simulations show that these approximate equalizers (9.15) achieve the same diversity order as those of (9.12), either in terms of outage using the SINR in (9.16) (with either of the two approximate expressions for $R^s_{\widetilde{\boldsymbol{A}}_U \widetilde{\boldsymbol{A}}_U}$) or (9.14), or in terms of probability of error of these Rxs with QAM transmission. Indeed, according to the various expressions for SINR_k^{MMRx}, an outage should occur whenever $\widehat{\mathbf{H}}_U$ loses full column rank and/or $R_{\boldsymbol{ZZ}}$ explodes (because $R_{\widetilde{\boldsymbol{h}}\widetilde{\boldsymbol{h}}}$ explodes). In the simulations shown in [37], we worked with (9.12) except for the case of FIR.

9.5 Fixing the Scalar Ambiguity in the Blind Case

The blind channel estimate $\widehat{\boldsymbol{h}}$ can only be determined up to a scalar α and to make it comparable to the true channel (or to use it in a Rx), this ambiguity needs to be fixed to obtain the final estimate $\widehat{\widehat{\mathbf{h}}} = \widehat{\boldsymbol{h}} \, \alpha$. As we shall see (see

[37] also), the way by which we resolve the scalar ambiguity has a major effect on the diversity achieved by the receiver. In this chapter we deal with three different constraints namely, Linear (Lin) constraint, Least-Squares (LSq) constraint and Fixing one-tap (FOT) constraint. Admittedly, these fixings are rather theoretical. In practice, one needs to consider differential modulation (see [37]) or a semi-blind approach.

9.5.1 Linear (Lin) Constraint

Generally, the cost function of any blind deterministic channel estimation can be represented by $\widehat{\boldsymbol{h}}^H \mathbf{Q}\, \widehat{\boldsymbol{h}}$ where possibly $\mathbf{Q} = \mathbf{Q}(\widehat{\boldsymbol{h}})$. To resolve the scalar ambiguity we can minimize this cost function subject to a linear constraint as follows: $\min_{\boldsymbol{h}^H \widehat{\widehat{\mathbf{h}}} = \widehat{\boldsymbol{h}}^H \boldsymbol{h}} ||\widehat{\widehat{\mathbf{h}}}^H \mathbf{Q}\, \widehat{\widehat{\mathbf{h}}}||^2$. Applying the Lagrange multiplier we get:

$$\widehat{\widehat{\mathbf{h}}} = \frac{\boldsymbol{h}^H \boldsymbol{h}}{\boldsymbol{h}^H \mathbf{Q}^{-1} \boldsymbol{h}} \mathbf{Q}^{-1} \boldsymbol{h} \,. \tag{9.17}$$

This constraint yields $\widetilde{\widehat{\mathbf{h}}} \perp \boldsymbol{h}$ and leads to the minimal CRB. Normally, the CRB is defined as the inverse of the FIM while for a singular FIM with the linear constraint considered here, the corresponding CRB is the pseudo-inverse of the FIM [51].

9.5.2 Least-Squares (LSq) Constraint

In this case the minimization process is done in two steps. First: $\min_{||\widehat{\boldsymbol{h}}||=1} \widehat{\boldsymbol{h}}^H \mathbf{Q}\, \widehat{\boldsymbol{h}}$ to get $\widehat{\boldsymbol{h}} = V_{min}(\mathbf{Q})$, where V_{min} represents the eigenvector that corresponds to the minimum eigenvalue. Then the scalar ambiguity is resolved by least squares as follows: $\min_{\alpha} ||\boldsymbol{h} - \alpha \widehat{\mathbf{h}}||^2$. After some manipulation we get the following solution:

$$\widehat{\widehat{\mathbf{h}}} = \frac{\widehat{\boldsymbol{h}}^H \boldsymbol{h}}{||\widehat{\boldsymbol{h}}||^2} \widehat{\boldsymbol{h}} = P_{\widehat{\boldsymbol{h}}} \boldsymbol{h} \tag{9.18}$$

so that $\widetilde{\widehat{\boldsymbol{h}}} \perp \widehat{\widehat{\boldsymbol{h}}}$ which is a well known feature of LS estimation. Also with this constraint, the corresponding CRB is the pseudo-inverse of the FIM. As a result, both the Linear and Least-Squares constraints lead to the same diversity order. Either of these constraints will be assumed in the further discussion of diversity in the blind case.

9.5.3 Fixing One Tap (FOT) Constraint

Now we minimize the cost function by considering wlog. that the first tap of the channel on the first Rx antenna is known: $\boldsymbol{e}_1^H \boldsymbol{h} = 1$ with $\boldsymbol{e}_1^H = [1\,0\cdots 0]$, $\min_{\boldsymbol{e}_1^H \widehat{\widehat{\boldsymbol{h}}}=1} \widehat{\widehat{\boldsymbol{h}}}^H \mathbf{Q}\, \widehat{\widehat{\boldsymbol{h}}}$. Applying the Lagrange multiplier we get:

$$\widehat{\widehat{\boldsymbol{h}}} = \mathbf{Q}^{-1}\boldsymbol{e}_1 \, \frac{\boldsymbol{e}_1^H \mathbf{h}}{\boldsymbol{e}_1^H \mathbf{Q}^{-1}\boldsymbol{e}_1} \, . \tag{9.19}$$

It is obvious from (9.19) that for $\widehat{\widehat{\boldsymbol{h}}}$ to vanish it is sufficient that $\boldsymbol{e}_1^H \mathbf{h}$ gets very small. Hence the diversity achieved is one regardless of the Rx used: $d^{FOT} = 1$. This may in part explain the bad performance of blind channel estimation algorithms using this constraint.

9.6 ZF Equalization in Single Carrier Cyclic Prefix (SC-CP) Systems

The diversity of LE for SC-CP systems has been studied in [82] for the SISO case with i.i.d. Gaussian channel elements, fixed rate R and block size $N = L$. The LE DMT for SIMO SC-CP systems appears in [83]. Consider a block of N symbol periods preceded by a cyclic prefix (CP) of length L (as a result of the CP insertion, actual rates are reduced by a factor $\frac{N}{N+L}$, which is ignored here in what follows). The channel input-output relation over one block can be written as

$$\boldsymbol{Y} = \mathcal{T}(\boldsymbol{h})\, \boldsymbol{A} + \boldsymbol{V} = \mathcal{A}\boldsymbol{h} + \boldsymbol{V}; \tag{9.20}$$

where $\boldsymbol{Y} = \boldsymbol{Y}_k = [\boldsymbol{y}_k^T \;\; \boldsymbol{y}_{k+1}^T \cdots \boldsymbol{y}_{k+N-1}^T]^T$ etc. $\mathcal{T}(\boldsymbol{h})$ is a banded block-circulant matrix (see (13) in [83]) and $\mathcal{A} = \mathcal{A}' \otimes \mathbf{I_p}$ where $\mathcal{A}'$ is a toeplitz matrix filled with the elements of $\boldsymbol{A}$. Now apply an N-point DFT (with matrix F_N) to each subchannel received signal, then we get

$$\underbrace{F_{N,p}\boldsymbol{Y}}_{\boldsymbol{U}} = \underbrace{F_{N,p}\mathcal{T}(\boldsymbol{h})\, F_N^{-1}}_{\mathcal{H}} \;\; \underbrace{F_N\, \boldsymbol{A}}_{\boldsymbol{X}} + \underbrace{F_{N,p}\boldsymbol{V}}_{\boldsymbol{W}} \tag{9.21}$$

where $F_{N,n} = F_N \otimes I_p$ (Kronecker product: $A \otimes B = [a_{ij}B]$), $\mathcal{H} = \text{blockdiag}\,\{\boldsymbol{h}_0, \ldots, \boldsymbol{h}_{N-1})\}$ with $\boldsymbol{h}_n = \boldsymbol{h}(f_n)$, the $p \times 1$ channel transfer function at tone n: $f_n = \frac{n}{N}$, at which we have

$$\mathbf{u}_n \;=\; \boldsymbol{h}_n\, \boldsymbol{x}_n + \mathbf{w}_n \, . \tag{9.22}$$

The $\boldsymbol{x}_n$ components are i.i.d. and independent of the i.i.d. $\mathbf{w}_n$ components with $\sigma_x^2 = N\,\sigma_a^2$, $\sigma_w^2 = N\,\sigma_v^2$.

9.6.1 Blind Channel Estimation

The Rx matched to blind channel estimation is the ZF LE. In the case of full CSIR, the SINR is given by (9.9), (9.10). In this case $\mathcal{T}_U(\boldsymbol{h}) = \mathcal{T}(\boldsymbol{h})$ is block circulant and loses column rank when $\boldsymbol{h}_n = 0$, i.e. when there is a complete fade on one of the tones, which represents p constraints on $\boldsymbol{h}$. So in this case simultaneously the ZF LE fades and the channel becomes unidentifiable. Hence, the full CSIR diversity is p. In the case of the LE with blind channel estimate, we need to consider (9.7), (9.10). As mentioned earlier, the combination of the blind channel estimate in the LE Rx leads to a reduction in the Rx dimension and hence some diversity loss due to $P^{\perp}_{\boldsymbol{\mathcal{A}}}$. As a result we can state that

$$d^{B-ZF}_{SC-CP} \leq d^{CSIR-ZF}_{SC-CP} = p \tag{9.23}$$

where the inequality becomes an equality as $\dfrac{L}{N} \to 0$. In the case of full CSIR, the SINR is identical for all symbols in the block. The SINR becomes position dependent in the blind case. We have investigated via simulations the dependence of the diversity order on the symbol position but did not find any. Also replacing the per symbol MSE by an average over the block led to the same diversity.

9.6.2 Semi-Blind Channel Estimation

We consider here M_K consecutive pilot symbols in the time domain. For the symbol following the M pilots, the block DFE Rx configuration is exactly that of a classical DFE with feedback length M_K. It has been shown in [83] that the diversity for such a full CSIR DFE is $d = p(1 + \min\{M_K, L-1\})$. Hence we conclude

$$d^{B-ZF}_{SC-CP} \leq d^{SB-ZF}_{SC-CP} \leq d^{CSIR-ZF}_{SC-CP} = p(1 + \min\{M_K, L-1\}) \; . \tag{9.24}$$

In this case the inequality is not only due to channel estimate-Rx coupling for a finite block length as in the blind case, but possibly also depends on the distribution of the M_K pilots over the block, as simulations reveal (see further).

9.7 ZF Equalization in OFDM Systems

Whereas for SC-CP the Tx symbols are $\boldsymbol{A}$ in time domain, in OFDM the symbols are in $\boldsymbol{X}$ in frequency domain. The same block processing formulas

remain valid, if considered in frequency domain. In OFDM, the channel is flat at every tone and transmission at different tones is decoupled. As a result, we get for the blind case

$$d_{OFDM}^{B-ZF} \leq d_{OFDM}^{CSIR-ZF} = p \; . \tag{9.25}$$

In the semi-blind case, pilots are now placed in the frequency domain. If we introduce $\mathcal{H}\,\boldsymbol{X} = \mathcal{X}\boldsymbol{h}$ then we get

$$CRB_{\boldsymbol{X}_U}^{SB} = \sigma_w^2 \left(\mathcal{H}_U^H P_{\mathcal{X}}^{\perp} \mathcal{H}_U \right)^{-1} \tag{9.26}$$

where $\mathcal{H}_U$ is obtained from $\mathcal{H}$ by eliminating the columns corresponding to the pilot positions. The pilots have no incidence on reception, only on channel estimation. As a result we get

$$d_{OFDM}^{B-ZF} \leq d_{OFDM}^{SB-ZF} \leq d_{OFDM}^{CSIR-ZF} = p \; . \tag{9.27}$$

9.8 ZF FIR/Non-CP Equalization

For time domain FIR equalization of length N, the block signal Tx model can be derived from the SC-CP case in (9.20) by considering a SC-CP block length of $N{+}L$ and removing the L first Rx samples in the block. $\mathcal{T}(\boldsymbol{h})$ is now replaced by a $Np \times (N{+}L{-}1)$ banded block Toeplitz matrix $\overline{\mathcal{T}(\boldsymbol{h})}$ which can be obtained from the block circulant $\mathcal{T}(\boldsymbol{h})$ by removing the $L-1$ top block rows, and $\mathbf{A}$ is replaced by $\overline{\mathbf{A}}$ containing N+L-1 symbols. For a ZF FIR LE with full CSIR and $N > L-1$, it was shown in [83] that the diversity is $d = p-1$. There is a diversity order loss of 1 compared to the SC-CP case because now the LE SINR fades or the channel becomes blindly unidentifiable whenever $\boldsymbol{h}[z]$ has a zero anywhere in the z-plane, as opposed to a zero at N discrete points on the unit circle as for the SC-CP case. So the constraint on $\boldsymbol{h}$ is that the $p{-}1$ other subchannels have a zero equal to a (any) zero of the first subchannel, which is $p{-}1$ constraints. As a result, we get for the FIR ZF LE with matching blind channel estimate

$$d_{FIR}^{B-ZF} \leq d_{FIR}^{CSIR-ZF} = p{-}1 \; . \tag{9.28}$$

For the semi-blind case, we can expect similarly $d_{FIR}^{SB-ZF} \leq d_{SC-CP}^{SB-ZF}$.

9.9 Simulations

The probability of error for Tx symbols drawn from an 8PSK constellation is simulated, averaged over 10^6 Monte-Carlo runs of AWGN noise, symbols

and i.i.d. Rayleigh fading channel realizations. We consider Tx block lenth $N = 20$, $p = 3$ Rx antennas, and channel memory $L - 1 = 1$. For the case of comparable outage probability Pr(O) considerations, we assume a rate $R = \frac{N}{N+L-1} \log(K)$, where $K = 8$ is the constellation size. In Fig. 9.1 we have simulated Pr(O) (via (9.7), (9.10)) for the SC-CP and FIR scenarios with blind channel estimation. It is obvious that in both scenarios $d \approx 2$. This result confirms our interpretation that blind channel estimation leads to a loss in the diversity order of a ZF-LE. In Fig. 9.2 we simulate all the Tx scenarios considered in this chapter but this time with semi-blind channel estimation. We assume that $M_K = 4$ training symbols (pilots) are inserted at the beginning of each block. We observe that SC-CP attains a full diversity order of $d = 6$ while OFDM achieves $d = 3$ (full spatial diversity) and the Non-CP case achieves just 2.5 (less than full spatial diversity). However, in Fig. 9.3 we reduce M_K to 2 (SC-CP and OFDM), but $M_K = 3$ for Non-CP (1 at the beginning of each block and 2 at the end). For SC-CP we get $d = 4$ while for OFDM $d = 3$ only. This result reveals the effect of the length of the training sequence (pilots) used on the diversity order achieved. On the other hand, for the Non-CP case with the distributed 3 training symbols, we get $d = 4$ which is higher than in the previous simulation where 4 training symbols were inserted at the beginning of the block. This shows the necessity to distribute the training sequence over both edges of the Tx block to achieve higher diversity orders for the Non-CP case.

9.10 Conclusions

In this chapter we have analyzed the diversity order of MMSE-ZF Linear and Decision-Feedback Equalization for frequency-selective SIMO channels, with the receivers being constructed from matching (semi-)blind channel estimates. The matching is furthermore interpreted here in a strict sense in which both the symbols and the channel get estimated on the basis of the same block of data. We have seen that matching leads essentially to the same diversity order for the receivers considered, built from (semi-)blindly estimated channels or from the true channel. For finite block lengths, the combination of receivers and channel estimates leads to some diversity reduction that requires further investigation. The effect of the positioning of pilot symbols also requires further investigation, as also the analysis of non-matching scenarios.

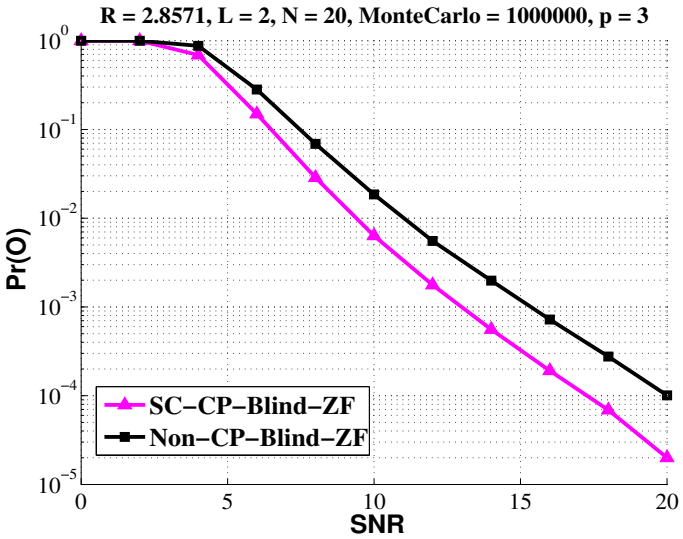

Figure 9.1: Probability of outage vs. SNR for SC-CP and FIR (Non-CP) Tx scenarios with blind channel estimation and ZF-LE.

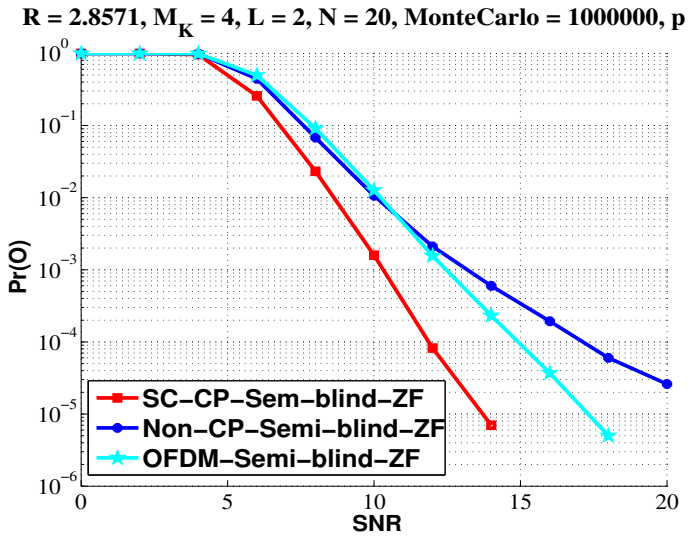

Figure 9.2: Probability of outage vs. SNR for SC-CP, OFDM and FIR (Non-CP) Tx scenarios with semi-blind channel estimation and ZF block DFE with 4 pilots at the start of the Tx block.

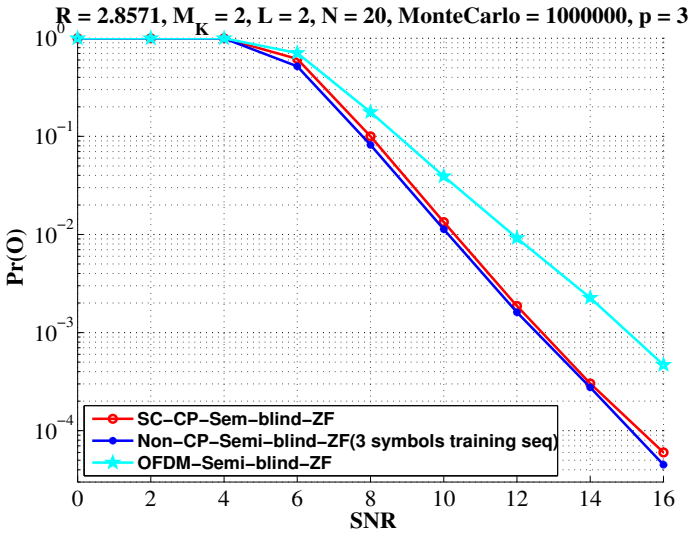

Figure 9.3: Probability of outage vs. SNR for SC-CP, OFDM and FIR (Non-CP) Tx scenarios with semi-blind channel estimation and ZF-DFE where 3 pilots are used for the case of Non-CP and 2 pilots for SC-CP or OFDM.

Chapter 10

Conclusions and Future Work

10.1 Conclusions

Although the (semi-)blind channel estimation topic has been investigated intensively during the past two decades, there are still numerous research papers that appear from time to time trying to adapt the existing techniques to some emerging wireless standards. Furthermore, some of these papers try to fill the gap and to handle the subjects that are considered untouched. In this thesis we have tried at one hand to introduce some novel algorithms that improve the channel estimation quality, and on the other hand we have tried to derive an analytical performance analysis for some other algorithms that already exist in the literature. Moreover, we have tried to address some aspects of the (semi-)blind channel estimation that have never been treated before like diversity. The motivations that encourage us to choose this topic are many folds:

- The superiority of the performance of the semi-blind channel estimation algorithms over the training based ones.
- The burden of sending a training sequence in the training-based channel estimation is high compared to the blind case, where training sequence is not required at all.

- The evolution in the signal processing hardware (DSPs, FPGAs, etc..) in the recent years permits the implementation of some sophisticated algorithms that were difficult or impossible to be implemented before.
- The (semi-)blind channel estimation topic is no more a theoretical topic. There are many applications in the market nowadays, from optical communications to microwave links, where (semi-)blind channel estimation constitutes a major block in the receiver.
- There is a new interest in some topics like cognitive radio and ad-hoc networks where the exchange of information or the cooperation between the different users is either prohibited (primary user and secondary user) or should be minimized to the minimum. In such systems, the transmission of pilots/training sequence is not possible.

We shall try to summarize in the sequel our main contributions throughout this thesis:

- The advantage of CP system is its ability to transform the frequency selective channel into flat fading at each tone. We have reintroduced in chapter 2 a framework that exploits this fact to easily derive an analytical performance analysis of a weighted and unweighted Signal Subspace Fitting (SSF), DIQML, PQML and SRM.
- We have also proposed in chapter 2 an enhanced version of DIQML where the denoising process is performed on a tone basis and derive its performance analytically.
- In chapter 3, by minimizing the sum of the cost functions at different tones (frequency domain) instead of minimizing the ordinary cost function in the time domain, we have shown that there is a great computational power saving.
- In chapter 3, we have proposed a spatio-temporal based algorithm to enhance the sample covariance matrix upon which a class of well-known estimators rely.
- In chapter 4, we have introduced the concept of variational Bayesian in the context of MIMO OFDM systems. To the best of our knowledge, this was the first trial to exploit this technique in the context of (semi-)blind channel estimation. This technique has shown a great tendency to fully exploit the available channel prior information to yield better channel estimates.

- In chapters 5 and 6, we have introduced the concept of classical Bayesian blind and semi-blind channel estimation in terms of some useful algorithms. A framework that allows the derivation of six deterministic and Bayesian algorithms has been presented. Some of these algorithms jointly estimate the channel and the symbols whereas the others estimate the channel and marginalize the symbols.

- In chapter 7, the CRBs that correspond to the different deterministic and Bayesian algorithms presented in chapter 6 have been derived. Three out of six of these CRBs are novel.

- In chapter 8, we have tried to extend the traditional deterministic algorithms into an intermediate point between deterministic and Bayesian. Hence, a new framework called quasi-Bayesian has been established. This framework permits to enhance the deterministic channel estimates noticeably. This enhancement has been achieved by neglecting some taps at the tail of the channel during the estimation process. The knowledge/estimation of the pdp is a must in this framework.

- In chapter 9, in the context of blind and semi-blind channel estimation, we have shown that for finite block lengths, the combination of receivers and channel estimates leads to some diversity reduction.

10.2 Future Work

It is true that we have tried to deal with as many aspects of the topics treated in this thesis as we can, but it is true also that there are still many questions unanswered. We shall present in the following what we do believe are the most important ideas. Nevertheless, we believe that not all these ideas are of the same difficulty level. Perhaps some are doable and the others seem more challenging.

In chapter 3 we proposed a spatio-temporal based algorithm to enhance the sample covariance matrix upon which a class of well-known estimators rely. This algorithm seems promising; however, the enhancement has been proven just by simulations. Hence, an analytical performance analysis will lead to a better understanding of the algorithm and the factors that affect much the enhancement.

In chapter 4 we introduced the concept of VB in the context of blind MIMO OFDM systems. We have seen that this technique allows for the exploitation of the prior information that exists about the channel effectively.

Once again here, the improvement in the channel estimation has been shown by conducting numerical simulations. Hence, an analytical performance analysis is a must if one needs to take a close look at this kind of algorithms.

The channel in wireless communications has almost been modeled as Rayleigh fading, i.e. with a Gaussian (prior) distribution expressing variances of and correlations between channel coefficients. Nevertheless, most of the people who worked in the (semi-) blind channel estimation field ignored this prior information that exists about the channel. Once this information gets exploited, the channel estimation problem is transferred from the deterministic to the Bayesian framework. Moreover, we have shown in chapters 5 and 6 that working within a Bayesian blind and Bayesian semi-blind framework leads to a noticeable improvement in the quality of channel estimation. However, this was only done by conducting numerical experiments. What is more interesting in this area is to perform an analytical performance analysis for the Bayesian algorithms presented in this dissertation. This would shed more light on the pros and cons of Bayesian algorithms and will allow us to know under which circumstances the latter outperforms the deterministic ones. The main obstacle that may hinder the accomplishment of this task is the biasedness that stems from exploiting the channel and/or the symbols prior information. Traditionally, this task is done by expanding the cost function of the algorithm under investigation using Taylor's series expansion. However, trying to perform this job classically does not permit taking the effect of the bias into consideration in the ultimate expression of the error covariance matrix. This point needs further research and investigation.

On the other hand, we have derived in chapter 7 the different deterministic and Bayesian CRBs that correspond to the different algorithms elaborated in chapters 6. However, we have noticed that most of these CRBs are not tight enough and consequently, there is a need to derive tighter lower bounds. The key point in this derivation is to consider the lower bound for one channel realization then averaging over the different realizations that we do have. However to do so, we need a formula for a biased CRB. Fortunately, this formula has been derived in the literature since a long time ago but the difficulty rises here while we try to compute the bias. There are many ideas on how to play around this problem that should be checked. Hence, this topic is also subject to further research.

In chapter 9 we have analyzed the diversity order of MMSE-ZF Linear and Decision-Feedback Equalization for frequency-selective SIMO channels, with the receivers being constructed from matching (semi-)blind channel estimates. For finite block lengths, the combination of receivers and channel estimates leads to some diversity reduction that requires further investiga-

tion. The effect of the positioning of pilot symbols also requires further investigation, as also the analysis of non-matching scenarios.

Bibliography

[1] J. Proakis, *Digital Communications 4th edition.* Englewood Cliffs, NJ: McGraw-Hill, 2000.

[2] O. Shalvi and E. Weinstein, *Universal methods for blind deconvolution.* in Blind Deconvolution, S. Haykin, Ed. Englewood Cliffs: Prentice-Hall, NJ, 1994.

[3] L. Tong and S. Perreau, "Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[4] D. Slock, "From Sinusoids in Noise to Blind Deconvolution in Communications," in Communications, Computation, Control and Signal Processing: a Tribute to Thomas Kailath, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Kluwer Academic Publishers, 1997.

[5] D. Slock and A. Medles, "Blind and Semiblind MIMO Channel Estimation," in *Space-Time Wireless Systems, From Array Processing to MIMO Communications*, H. Bölcskei, C. P. D.G̃esbert, and A.-J. van der Veen, Eds. Cambridge University Press, 2006.

[6] S. Haykin, *Adaptive Filter Theory.* Upper Saddle River, NJ: Prentice-Hall, 1996.

[7] Y. Sato, "A Method of Self–Recovering Equalization for Multilevel Amplitude Modulation Systems," *IEEE Transactions on Communications*, vol. 23, pp. 679–682, June 1975.

[8] D. Godard, "Self recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, vol. 28, pp. 1867–1875, Nov. 1980.

[9] J. Treichler, I. Fijalkow, and C. Johnson, "Fractionally spaced equalizers: How long should they really be?" *IEEE Transactions on signal processing*, vol. 13, pp. 65–81, May 1996.

[10] C. Xu, G. Feng, and K. Kwak, "A modified constrained constant modulus approach to blind adaptive multiuser detection," *IEEE Transactions on communications*, vol. 49, pp. 1642–1648, Sept. 2001.

[11] J. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Mag.*, vol. 13, pp. 24–42, May 1996.

[12] J. Gomes and V. Barroso, "A super-exponential algorithm for blind fractionally spaced equalization," *IEEE Signal Processing Lett.*, vol. 3, no. 5, pp. 283–285, Oct. 1996.

[13] A. Bessios and C. Nikias, "POTEA: The power cepstrum and tricoherence equalization algorithm," *IEEE Trans Commun.*, vol. 43, pp. 2667–2671, Nov. 1995.

[14] C.-H. Chen, "Blind multi-channel equalization and two-dimensional system identification using higher-order statistics," Ph.D. dissertation, Dept. Elect. Eng., National Tsing Hua Univ., Hsinchu, Taiwan, 2001.

[15] J. Gorman and A. Hero, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Transactions on Signal Processing*, vol. 45, pp. 268–271, Jan. 1997.

[16] C. Chi, C. Chen, C. Chen, and C. Feng, "Batch Processing Algorithms for Blind Equalization Using Higher-Order Statistics," *IEEE Signal Processing Magazine*, Jan 2003.

[17] J. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications," *Proceedings of the IEEE*, vol. 70, no. 3, March 1991.

[18] E. de Carvalho and D. Slock, "Semi–Blind Methods for FIR Multichannel Estimation," in *Signal Processing Advances in Wireless & Mobile Communications*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Prentice Hall, 2001.

[19] P. Laurent, "Exact and Approximate Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses (AMP)," *IEEE Trans. Communications*, vol. 34, February 1986.

[20] Y. Yoon, R. Kohno, and H. Imai, "A Spread-Spectrum Multiaccess System with Cochannel Interference Cancellation for Multipath Fading Channels," *IEEE Journal on Selected Areas in Communications*, vol. 11, no. 7, pp. 1067–1073, Sept. 1993.

[21] L. Tong, G. Xu, and T. Kailath, "A New Approach to Blind Identification and Equalization of Multipath Channels," in *Proc. of the 25th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1991, pp. 856–860.

[22] D. Slock, "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction," in *Proc. ICASSP 94 Conf.*, Adelaide, Australia, April 1994.

[23] D. Slock and C. Papadias, "Blind Fractionally-Spaced Equalization Based on Cyclostationarity," in *Proc. Vehicular Technology Conf.*, Stockholm, Sweden, June 1994.

[24] M. Kristensson, B. Ottersten, and D. Slock, "Blind Subspace Identification of a BPSK Communication Channel," in *Proc. of the 30th Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, Nov. 1996.

[25] A. van der Veen, "Analytical Method for Blind Binary Signal Separation," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1078–1082, April 1997.

[26] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Transactions on Communications*, pp. 1867–1875, Nov 1980.

[27] S.-M. Omar, O. Bazzi, and D. Slock, "Performance Analysis Of Blind FIR Channel Estimation Algorithms In SIMO Cyclic Prefix Systems," *To be submitted to IEEE trans. on signal processing*, 2011.

[28] E. de Carvalho, S.-M. Omar, and D. Slock, "Performance and Complexity Analysis of Blind FIR Channel Identification Algorithms Based on Deterministic Maximum Likelihood in SIMO Systems," *Submitted to circuits, systems and signal processing journal*, 2011.

[29] S.-M. Omar and D. Slock, "Structured Spatio-Temporal Sample Covariance Matrix Enhancement with Application to Blind Channel Estimation in Cyclic Prefix Systems," in *Proc. IEEE Int'l Workshop on*

Signal Processing Advances in Wireless Comm's (SPAWC), Perugia, Italy, June 2009.

[30] ——, "Singular Block Toeplitz Matrix Approximation and Application to Multi-Microphone Speech Dereverberation," in *Proc. 10th IEEE International Workshop on MultiMedia Signal Processing (MMSP)*, Queensland, Australia, Oct. 2008.

[31] J. Cadzow, "Signal Enhancement – A Composite Property Mapping Algorithm," *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 36, no. 1, pp. 49–62, jan 1988.

[32] S.-M. Omar and D. Slock, "Variational Bayesian Blind and Semiblind Channel Estimation," in *Proc. IEEE Int'l Symposium on Communications, Control and Signal Processing (ISCCSP)*, Limassol, Cyprus, March 2010.

[33] S.-M. Omar, D. Slock, and O. Bazzi, "Bayesian Blind FIR Channel Estimation Algorithms in SIMO Systems," in *Proc. IEEE Int'l Workshop on Statistical Signal Processing (SSP)*, Nice, France, June 2011.

[34] ——, "Bayesian Semi-Blind FIR Channel Estimation Algorithms in SIMO Systems," in *Proc. IEEE Int'l Workshop on Signal Processing Advances in Wireless Comm's (SPAWC)*, San Francisco, USA, June 2011.

[35] ——, "Bayesian and Deterministic CRBs for Semi-Blind Channel Estimation in SIMO Single Carrier Cyclic Prefix Systems," in *Proc. IEEE PIMRC*, Toronto, Canada, Sep. 2011.

[36] ——, "A performance of bayesian semi-blind FIR channel estimation algorithms in SIMO systems," in *EUSIPCO 2011, 19th European Signal Processing Conference*, Barcelona, Spain, Sep. 2011.

[37] S.-M. Omar, O. Bazzi, and D. Slock, "Receiver diversity with blind FIR SIMO channel estimates," in *Proc. IEEE ICASSP*, Dallas, US, March 2010.

[38] S.-M. Omar, D. Slock, and O. Bazzi, "Receiver Diversity With Blind And Semi-Blind FIR SIMO Channel Estimates," in *Proc. IEEE Int'l Symposium on Communications, Control and Signal Processing (ISCCSP)*, Limassol, Cyprus, March 2010.

[39] D. Slock, "Blind FIR Channel Estimation in Multichannel Cyclic Prefix Systems," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, July 2004.

[40] M. Viberg and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[41] D. Slock, "Blind Joint Equalization of Multiple Synchronous Mobile Users Using Oversampling and/or Multiple Antennas," in *Proc. 28th Asilomar Conference on Signal, Systems & Computers*, Pacific Grove, CA, Nov. 1994.

[42] Y. Hua, "Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels," *IEEE Transactions on Signal Processing*, vol. 44, no. 3, pp. 661–672, March 1996.

[43] J. Ayadi and D. Slock, "Cramér-Rao Bounds and Methods for Knowledge Based Estimation of Multiple FIR Channels," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.

[44] ——, "On Linear Channel-based Noise Subspace Parameterizations for Blind Multichannel Identification," in *Signal Processing Advances in Wireless Communications (SPAWC)*, March 2001.

[45] M. Gürelli and C. Nikias, "A New Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals," in *Proc. ICASSP*, 1993, pp. 448–451.

[46] L. Baccala and S. Roy, "A New Time-Domain Blind Identification Method," *IEEE Signal Processing Letters*, vol. 1, no. 6, pp. 89–91, June 1994.

[47] G. Xu, H. Liu, L. Tong, and T. Kailath, "A Least Squares Approach to Blind Channel Identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[48] Y. Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE Transactions on Acoustics, Speech, Signal Processing*, vol. 35, no. 10, pp. 1081–1089, Oct. 1986.

[49] M. Osborne and G. Smyth, "A Modified Prony Algorithm for Fitting Functions Defined by Difference Equations," *SIAM J. Sci. Stat. Comput.*, vol. 12, no. 2, pp. 362–382, 1991.

[50] G. Harikumar and Y. Bresler, "Analysis and Comparative Evaluation of Techniques for Multichannel Blind Deconvolution," in *Proc. 8th IEEE Sig. Proc. Workshop Statistical Signal and Array Proc.*, Corfu, Greece, June 1996, pp. 332–335.

[51] E. de Carvalho, J. Cioffi, and D. Slock, "Cramér–Rao Bounds for Blind Multichannel Estimation," in *Proc. GLOBECOM 2000 Conf.*, San Francisco, USA, December 2000.

[52] H. Yang and Y. Hua, "On performance of the cross relation method for blind channel identification," *Signal Processing*, vol. 62, no. 2, pp. 187–205, Oct. 1997.

[53] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Version: February 10, 2007.

[54] D. S. J.Ayadi, "On Linear Channel-based Noise Subspace Parameterizations for Blind Multichannel Identification," in *Proc. IEEE Signal Processing Advances in Wireless Communications*, Taiwan , China, March 2001.

[55] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–526, Feb. 1995.

[56] D. Slock, "Bayesian Blind and Semiblind Channel Estimation," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, July 2004.

[57] D. Tzikas, A. Likas, and N. Galatsanos, "The Variational Approximation for Bayesian Inference, Life After the EM Algorithm," *IEEE Signal Processing Magazine*, Nov. 2008.

[58] D. Slock, "Subspace Techniques in Blind Mobile Radio Channel Identification and Equalization using Fractional Spacing and/or Multiple Antennas," in *Proc. 3rd International Workshop on SVD and Signal Processing*, Leuven, Belgium, Aug. 22-25 1994.

[59] E. de Carvalho and D. Slock, "Maximum-Likelihood FIR Multi-Channel Estimation with Gaussian Prior for the Symbols," in *Proc. ICASSP 97 Conf.*, Munich, Germany, April 1997.

[60] A. Renaux, P. Forster, E. Boyer, and P. Larzabal, "Unconditional Maximum Likelihood Performance at Finite Number of Samples and High Signal-to-Noise Ratio," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 2358 – 2364, 2007.

[61] D. Slock and C. Papadias, "Further Results on Blind Identification and Equalization of Multiple FIR Channels," in *Proc. ICASSP 95 Conf.*, Detroit, Michigan, May 1995.

[62] P. . Kay, W. . Murray, and M. . Wright, *Practical Optimization*. London: Academic press, 1981.

[63] W. Wei, J. Xin, C. Yueming, and W. X. ;, "Bayesian Cramer-rao Bound for channel estimation in cooperative OFDM," in *IEEE WCSP*, Suzhou, China, Oct. 2010.

[64] A. Jagannatham and B. Rao, "Constrained ML Algorithms For Semi-Blind MIMO Channel Estimation," in *IEEE Globecom 2004*, Dallas, Texas, USA, Dec 2004.

[65] W. Y. Y. C. Y. Xun;, "Semi-blind Channel Estimation for OFDM Systems," in *IEEE Proc. VTC*, Seattle, USA, May 2006.

[66] W.-C. Huang, C.-H. Pan, C.-P. Li, and H.-J. Li, "Subspace-Based Semi-Blind Channel Estimation in Uplink OFDMA Systems," *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 58–65, Jan. 2010.

[67] G. K. D. Raphaeli, "Maximum likelihood semi-blind equalization of doubly selective channels," in *ISCCSP*, Malta, March 2008.

[68] H. Cirpan and M. Tsatsanis, "Stochastic maximum likelihood methods for semi-blind channel estimation," *IEEE Signal Processing Letters*, vol. 5, pp. 21–24, 1998.

[69] J. Ayadi, E. de Carvalho, and D. Slock, "Blind and Semi-Blind Maximum Likelihood Methods for FIR Multichannel Identification," in *Proc. ICASSP 98 Conf.*, Seattle, USA, May 1998.

[70] E. de Carvalho and D. Slock, "Semi-Blind Maximum-Likelihood Estimation with Gaussian Prior for the Symbols using Soft Decisions," in

48th Annual Vehicular Technology Conference, Ottawa, Canada, May 1998.

[71] L. Berriche and K. Abed-Meraim, "Semi-Blind Stochastic Maximum Likelihood for Frequency Selective MIMO Channels," in *PIMRC*, Berlin, Germany, Sep. 2005.

[72] E. de Carvalho and D. Slock, "Cramér-Rao Bounds for Semi-blind, Blind and Training Sequence based Channel Estimation," in *Proc. SPAWC 97 Conf.*, Paris, France, April 1997.

[73] E. Weinstein and A. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Trans. Inform. Theory*, vol. 34, no. 34, p. 338342, Mar. 1988.

[74] Y. Noam and H. Messer, "Notes on the Tightness of the Hybrid CramérRao Lower Bound," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2074–2084, June 2009.

[75] Y. Zi and C. Yue-ming, "Cramer-Rao bound for blind, semi-blind and non-blind channel estimation in OFDM systems," in *IEEE ISCIT*, Beijing, China, Oct. 2005.

[76] M. Dong and L. Tong, "Optimal Design and Placement of Pilot Symbols for Channel Estimation," *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec. 2002.

[77] D. Cassioli, M. Win, and A. Molisch, "A statistical model for the UWB indoor channel," in *Proc. IEEE 53rd Vehicular Technology Conference (VTC)*, Rhodes, Greece, May 2001.

[78] E. Carvalho and D. Slock, "Deterministic quadratic semi-blind FIR multichannel estimation : algorithms and performance," in *IEEE International conference on acoustic, speech and signal (ICASSP)*, Istanbul, Turkey, June 2000.

[79] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Info. Theory*, vol. 49, no. 5, pp. 1073 – 1096, May 2003.

[80] C. Tepedelenlioglu, "Maximum Multipath Diversity with Linear Equalization in Precoded OFDM Systems," *IEEE Trans. Info. Theory*, vol. 50, no. 1, pp. 232 – 235, Jan 2004.

[81] N. Prasad, L. Venturino, X. Wang, and M. Madihian, "Diversity–Multiplexing Trade–off Analysis of OFDM Systems with Linear Detectors," in *Proc. IEEE Globecom*, Washington, DC, Nov 2007.

[82] A. Hedayat, A. Nosratinia, and N. Al-Dhahir, "Outage Probability and Diversity Order of Linear Equalizers in Frequency–Selective Fading Channels," in *Proc. 38th Asilomar Conf. on SSC*, CA, USA, Nov 2004.

[83] D. Slock, "Diversity and Coding Gain of Linear and Decision-Feedback Equalizers for Frequency-Selective SIMO Channels," in *Proc. IEEE Int'l Symp. Info. Theory (ISIT)*, Seattle, USA, July 2006.

[84] A. Gorokhov, "On the Performance of the Viterbi Equalizer in the Presence of Channel Estimation Errors," *IEEE Signal Processing Letters*, vol. 5, no. 12, pp. 321–324, Dec. 1998.

[85] N. B. L.C̃ao, "Exact Error-Rate Analysis of Diversity 16-QAM With Channel Estimation Error," *IEEE Trans. Communications*, vol. 52, no. 6, pp. 1019–1029, Jun. 2004.

[86] M. C. W.M.G̃ifford, M.Z.W̃in, "Diversity With Practical Channel Estimation," *IEEE Trans. Communications*, vol. 4, no. 4, pp. 1935 – 1947, July 2005.

[87] N. B. L.C̃ao, "Closed-Form BER Results for MRC Diversity With Channel Estimation Errors in Ricean Fading Channels," *IEEE Trans. Communications*, vol. 4, no. 4, pp. 1440–1447, July 2005.

[88] E. de Carvalho and D. Slock, "Cramer–rao bounds for semi-blind and training sequence based channel estimation," in *Proc. IEEE SPAWC*, Paris, France, April 1997.

[89] ——, "Blind and Semi–Blind FIR Multichannel Estimation: Identifiability Conditions," *IEEE Transactions on Signal Processing*, vol. 52, pp. 1053–1064, April 2004.